AGPH Books

EDITION

2021

Advances in Cloud Computing Security
Techniques and Applications

Edited By

Dr. Stuti Asthana

Dr. Rakesh Kumar Bhujade

Prof. Ghanshyam Prasad Dubey

Advances in Cloud Computing Security

Techniques and Applications

Dr. Stuti Asthana

Independent Researcher and Analyst

Dr. Rakesh Kumar Bhujade

Head of the Information Technology Department in Government Polytechnic, Daman

Prof. Ghanshyam Prasad Dubey

Associate Professor, Sagar Institute of Science and Technology, Gandhi Nagar, Bhopal.

AG PH Books

2021

First Edition: 2021

ISBN: 978-81-955340-6-7

Published by:

Academic Guru Publishing House,

B-6, Shopping Complex, Ground Floor, Hoshangabad Rd, behind Indian Oil Petrol Pump, Vidya Nagar, Bhopal, Madhya Pradesh 462026

Website: https://www.agphbooks.com

About the Book

Cloud computing is a technical and social reality today; at the same time, it is an emerging technology. At this time one can only speculate how the infrastructure for this new paradigm will evolve and what applications will migrate to it. The economic, social, ethical, and legal implications of this shift in technology, whereby users rely on services provided by large data centers and store private data and software on systems they do not control, are likely to be significant.

In the last several years, many books have been published on cloud computing. Each book has attempted to present some element of the topic for a particular audience. In this book, 15 chapters are written by various researchers. And each chapter reflects the different aspect of cloud computing. It is an edited book and many topics in this book are unique to this book and are based on published information that is both current and timely.

Preface

Cloud computing has become a great solution for providing a flexible, on-demand, and dynamically scalable computing infrastructure for many applications. Cloud computing also presents a significant technology trend, and it is already obvious that it is reshaping information technology processes and the IT marketplace.

Cloud computing is a movement started sometime during the middle of the first decade of the new millennium. The movement is motivated by the idea that information processing can be done more efficiently on large farms of computing and storage systems accessible via the Internet. In this book we attempt to sift through the large volume of information and dissect the main ideas related to cloud computing.

The appeal of cloud computing is that it offers scalable and elastic computing and storage services. The resources used for these services can be metered and the users can be charged only for the resources they use. Cloud computing is a business reality today as increasing numbers of organizations are adopting this paradigm.

This book attempts to provide a snapshot of the state of the art of a dynamic field likely to experience significant developments in the near future. This is an edited book; comprises 15 chapters written by different researchers.

- Editorial Team

Advances in Cloud Computing Security: Techniques and Applications

(ISBN: 978-81-955340-6-7)

CONTENT

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

A Survey on Intelligent Data Analysis: Issues and Challenges

Mousami V. Munot[1*], Kaustubh V. Sakhare[2]

[1]*Associate Professor, Department of Electronics & Telecommunication Engg., SCTR's Pune Institute of Computer Technology (PICT), Pune*

[2]*Technical Specialist, Lear Corporation, Pune*

Abstract

(IDA) that is Intelligent data analysis is a new topic that combines several disciplines, particularly AI as well as statistics, to analyse data sets automatically or semiautomatically in a variety of real-world applications. All three of these areas are mutually beneficial: Several statistical procedures depend on computers, especially for big data sets, yet computational power alone cannot replace statistical expertise. There has been a rise in an intelligent data analysis system. It is goal of such a work to address a broad variety of issues that might arise when analysing data, as well as to provide solutions. A real-world instance of such a risk assessment of the level crossing data is used to analyse a few of such issues and ideas.

Keywords: Data analysis, data mining, risk assessment of level crossing, rule extraction, neural networks, rule induction

1. INTRODUCTION

It is the activity of utilising AI as well as machine learning to evaluate and turn large datasets into an intelligent data insight that can subsequently be utilised to enhance services as well as investments that is known as data intelligence. Better business processes may be developed via the use of data

* ISBN No. 978-81-955340-6-7

the intelligence tools as well as strategies that assist decision makers well comprehend the data they've acquired.

There are five primary components to the data-driven intelligence process: descriptive data, prescriptive information, diagnostic information, as well as predictive information. These fields are concerned with figuring out how to make sense of data, coming up with new ideas, resolving problems, and looking back at the past to make predictions about the future. Cybersecurity, banking, health, as well as insurance, as well as law enforcement, are some of the most pressing fields in need of data intelligence. Using intelligent data capture technologies, printed documents or photographs may be transformed into useful data in various fields.

Business intelligence relies on the use of intelligent data as a foundational component. As a result of intelligent data processing, big datasets can be restructured into the valuable information which is relevant to the business performance, allowing organisations to understand patterns, make the informed choices as well as adapt to the new information; as well as advanced analytics can be incorporated to improve visualisations of prescriptive as well as predictive analytics in order to better understand the data as well as make better decisions.

It is a multidisciplinary topic of research that focuses on the extraction of usable information from data using methods from a wide range of domains, including AI which is Artificial Intelligence, elevated performance computing, pattern recognition, as well as statistics. Firms like, Strategic Data Intelligence, Data Visualization Intelligence and Global Data Intelligence provide platforms as well as solutions for data intelligence. The Data intelligence companies like these included.

Increased demand and supply for more advanced IDA approaches have been spurred by the explosive growth in the amount of real-time data generated by online, multimedia, as well as electronic commerce activities. The fundamental notion of analysing enormous volumes of data with detailed descriptions is both attractive and straightforward, but it is a substantial challenge and difficulty to implement in practise. To make the most of data acquired from certain huge and complicated sources, there has to be a plan in place.

In other words, IDA extract value from the data by finding out rules as well as knowledge from it. Although it's impossible to count the precise no. of IDA methods, their evolving patterns, which include (1) algorithm principle, (2) dataset size, (3) dataset type may be summarised.

2. ALGORITHM PRINCIPLE

The evolution of IDA's algorithm concept has showed a progression from basic to sophisticated. On the basis of probability theory or even Euclidean distance dependent similarity theory, earlier IDA algorithms were developed IDA principles became increasingly complicated over time when computational intelligence was included.

2.1. Probability Based Algorithm

The IDA methods depending on the probability theory are often used for classification and grouping because of the property of probability theory altogether. Prior probability as well as posteriori probability are used in the Naive Bayes Classifier (NBC) to categorise sample data. Classification is performed using C4.S, which calculates the sampling data's entropy gain, whereas clustering is carried out using that is Expectation Maximization (EM), which seeks to find the parameters' greatest probability estimates. Because they are easy to implement as well as perform well, IDA algorithms that is probability-based have become a popular choice.

[2] use the auxiliary feature approach to perform a 2nd feature selection following NBC to improve the accuracy of NBC with in large text categorization.

[3] combine the NBC with the Decision Tree in order to improve classification accuracy as well as reduce the phenomena of over-fitting.

[4] By using stochastic processes in feature selection step of NBC's, (RSNB) that is Randomly Selected Naive Bayes method avoids the local optimum difficulties that plague classic NBC.

[5] During plant monitoring, use EM towards the challenge of detecting change points for the multivariate data.

[6] To perform fault diagnosis based on data, they suggest a hybrid EM technique depending on forward and backward Kalman filtering.

[7] Boolean factor analysis based on High-dimensional may be done using two new EM methods.

2.2. Euclidean Distance Based Algorithm

Inside the context of the n-dimensional dataset, Euclidean distance between various components may be used to measure the similarity in between them in context of dataset. Euclidean distance IDA techniques that focus on finding cluster centres by minimising total number of mean-square errors, such as k-Means as well as (k-NN) which is k-nearest-neighbour algorithms, are also popular choices. With SVM model, the data is represented as points in the space, which are then remapped to the higher-dimensional space. This creates a gap as large as possible between the data points in each category, resulting in a distinct separation between them.

[8] Breast cancer tumours may be diagnosed using a combination of k-means clustering and a support vector machine (SVM).

[9] Reduce sensitivity to an initial cluster centre while increasing the capacity to cope with scattered data by modifying k-Means technique with just an evolutionary approach, [10] The repetitious training for the continuous input conditions may be eliminated by using a quick k-Means algorithm to the graphic processing.

3. DATA ANALYSIS TASKS AND TECHNIQUES

For example, a user's purpose may be to characterise the entire data set or to construct linkages between the subsets of the patterns within data set [11]. Purpose of Predictive modelling is to create predictions depending over the data's basic properties. There are various preset classes and real-valued prediction variables that may be used to model data. Predictive modelling may be performed using whatever supervised machine learning technique which builds a model based on past or current data. The model is taught how to correctly forecast the future by being provided a list of previously established facts with right replies. A few of the methodologies used to map discrete-valued sets of variables include, decision trees, even neural networks, K-nearest neighbour classifier, Bayesian classifiers, reasoning based on case, genetic programming, fuzzy sets, as well as rough sets. There are a number of strategies for mappings regular-valued target variables using regression, even neural networks, including radial basis functions. Clustering is a technique used to create hierarchies of events by grouping together those having similar features. It is possible to accomplish clustering using any of the unsupervised machine learning technique that does not have a preconceived set of data categories. There are certain pre-existing facts that model is given, through it produces categories of the data with comparable features. Methods for clustering include partitioning, even hierarchical as well as density-depend algorithms; including model-based algorithms. [12].

Using link analysis, one may discover the intrinsic connections between data points. This objective is accomplished by the identification of associations, the discovery of sequential patterns, and related activities involving the discovery of temporal sequences [11]. By anticipating the correlation of elements that might otherwise be obscure, such activities reveal samples as well as patterns. Counting all potential combinations of objects is the basis for link analysis approaches. Apriori and its variants are among the most often used algorithms. [13].

4. CHALLENGES FACING THE IDA IN BIG DATA ENVIRONMENT

People's need to get more out of their data has been reached historic heights in the big data era, making it difficult for IDA to keep up. There are four ways wherein big data environment presents problems to the IDA, that may be summarised as follows: (1) data management, (2) data collecting, (3) data analysis, and (4) application pattern are all aspects of the big data process.

a. Big Data Management

Massive data management technologies, such as (HDFS) that is Hadoop Distributed File System [14], are already available and quite mature. A good large data management system, on the other hand, does more than just store information correctly. There are three key issues in the big data management: managing life cycle of data, securing data, and managing costs.

- **Data Life Cycle Management**

The most difficult part of managing the life cycle of data is determining how long a piece of data should be kept in storage. The conventional wisdom is that data life cycle concludes with the completion of the data analysis. Despite this, data life cycle management is indeed no more a straightforward matter in the context of big data. Users' perspectives vary even while working with the same dataset, resulting in varying amounts of value gleaned from the information. Data lifecycles are also varied because of this. A medical record of patient , for instance, comes to an end whenever the patient is well enough to be discharged. Medical records, on the other hand, may be an invaluable source of information for a clinician interested in learning about a history of allergies of family of the Patient. In contrast, for just an epidemiologist looking into a specific pandemic, large medical records integration is essential and useful. Case studies like this one demonstrate how diverse data life cycle management may result in varied data value extractions.

b. Data Security Management

Another issue that the IDA must deal with is the complete lifecycle of data security management. Individuals in the big data environment are indeed worried about their data's privacy as well as security. Personal, corporate, as well as national secrets may all be compromised at any point in the data lifecycle with such an inclusive environment. Data encryption might protect data to a certain degree, but it can also slow down data processing if encryption is too complicated. Data fuzzification, in addition to encryption, is another option for data security, albeit it has the potential to damage data. For having data security maintenance, it is the critical to optimise the efficiency of data security IDA.

c. Cost Management

Another issue that the IDA must deal with is how much money it spends on operations. The expense of strengthening the IDA's performance should be kept in check in order to ensure sustainable development. Dispersed data analysis, for instance, speeds up data processing whereas increasing the cost of hardware as well as network transmission; implementing various life cycle management strengthens value extracted from the data; complicated data encryption strengthens data security during increasing the computation complexity; and a variety of other examples. In order to achieve a balance between value as well as cost, the primary goal of the cost management is being to reduce both the explicit but also hidden costs.

d. Data Collection

In additional increase in size, data with in the big data environment have demonstrated qualities such as multi-source as well as heterogeneous. It is difficult for the IDA to gather data from several diverse sources and combine or preprocess such a large amount of heterogeneous data.

- **Fusion of Multi-source Heterogeneous**

Data In big data environment, data sources can be ubiquitous, which gives rise to the data heterogeneousness. Apart from traditional numerical data, data also include text, images, sound, and other electrical signals. The key point is to develop a integrate analysis algorithm in a novel big data analysis framework to transform all the multi-mode heterogeneous data into a uniform format that can be dealt with by IDA.

- **Pre-processing of Messy Data**

The noise and redundancy contained in raw data may greatly influence IDA's speed, accuracy, and robustness [15]. The preprocessing of data is necessary. But in big data environment, traditional data pre-processing technologies cannot reach the real-time demand of the applications. Hence, the key point of the pre-processing of messy data is to directly do feature selection to the raw data with noise and redundancy, and find out the features that matter most, so that the real-time changing of analysis demand can be followed.

e. Data Analysis

Data analysis difficulties such as the identification of features for distributed data analysis, an unbalanced dataset, as well as big data modelling persist despite the advances achieved by IDA methods.

- Asymmetrical Dataset Feature Selection

Data preparation has picked the most essential characteristics from such multi-source heterogeneous dataset, yet this selection may result in an unbalanced dataset. The minority in an unbalanced dataset are always ignored as noises by traditional IDA techniques. However, in other situations, such as fault diagnosis, the knowledge contained in these outliers may be quite significant. For this reason, it is vital to build specialised feature selection methods for unbalanced datasets.

f. Application Pattern

There is a limited number of ways to use IDA in the traditional sense. Patterns for such cross-platform applications.

Data interchange standards and display of complicated data face additional hurdles in the context of big data.

- **Exchanging Data Standard**

As IDA's application pattern evolves, so does the amount of data that may move across various platforms. It is challenging for such cross-platform IDA programme to unify data storage formats as well as data structures during the data transmission. The expenses of standardising data storage formats as well as architectures are also increasing. RosettaNet, for example, is indeed a standard for industrial data transmission, but it hasn't been widely used because of issues with universality as well as usability. By defining the data exchange standard having strong relevance, universality as well as usability, it is possible to even further boost data mobility inside a large-scale data environment.

- **Visualization**

Data that is difficult to analyse In a (DSS) decision support system, data visualisation may provide a straightforward as well as consumer-friendly man-machine interface. Inside a big data context, correlations between data grow increasingly complicated because of the rising size, dimensionality, data sources, as well as heterogeneity of such data. The decision maker might well be better able to understand the IDA findings but also make more informed judgments if the data is shown. In spite of this, IDA products may become better known as a result of the display of complicated data.

5. CONCLUSION

There are several stages to data analysis: issue conceptualization; data quality assurance; model creation; interpretation as well as post-processing. An investigation of intelligent data analysis has been conducted in this study. Problems in actual applications are the primary source of the issues, thus solutions must be tailored to the situation at hand. It is essential that new data models as well as lDAs be customised for individual applications in order to extract the maximum amount of value and information that can be put to use. IDA researchers must engage more with industry and mix actual applications as well as theoretical investigations in order to address the issues that may arise in future.

REFERENCES

1. R. Nayak, Data Mining for Web-Enabled Electronic Business Applications, to be published in Architectural Issues of Web-Enabled Electronic Business, Shi Nansi Ed., Idea Publishing Group, April 2002.
2. Zhang, Wei, and Feng Gao, "An Improvement to Naive Bayes for Text Classification," Procedia Engineering, vol. 15, pp. 2160-2164, 2011.
3. Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A., and Strachan, R., "Hybrid decision tree and naive Bayes classifiers for multi-class classification tasks," Expert Systems with Applications, vol. 41(4), pp. 1937-1946,2014.
4. Liangxiao Jiang, Zhihua Cai, Harry Zhang, and Dianhong Wang, "Not so greedy: Randomly Selected Naive Bayes," Expert Systems with Applications, vol. 39(12), pp. 11022-11028,2012.
5. Keshavarz, M., and Huang, B., "Bayesian and Expectation Maximization methods for multivariate change point detection," Computers & Chemical Engineering, vol. 60, pp. 339-353,2014.
6. Mahmoud, M. S., and Khalid, H. M., "Expectation maximization approach to data-based fault diagnostics," Information Sciences, vol. 235, pp. 80-96,2013.

7. Frolov, A. A., Husek, D., and Polyakov, P. Y., "Two Expectation-Maximization algorithms for Boolean Factor Analysis," Neurocomputing, vol. 130, pp. 83-97,2013

8. Zheng, B., Yoon, S. W., and Lam, S. S., "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," Expert Systems with Applications, vol. 41(4), pp. 1476-1482,2014.

9. M.e. Naldi, and RJ.G.B. Campello, "Evolutionary k-means for distributed data sets," Neurocomputing, vol. 127, pp. 30-42,2014.

10. Lin, e. H., Chen, e. e., Lee, H. L., and Liao, 1. R., "Fast K-means algorithm based on a level histogram for image retrieval," Expert Systems with Applications, vol. 41(7), pp. 3276-3283,2014.

11. P. Cabena, P. Hadjinian, R. Stadler, J. Verhees & A. Zanasi, Discovering Data Mining from Concept to Implementation, Prentice Hall PTR, 1997.

12. J. Han & M. Kamber, Mastering Data Mining, San Francisco: Morgan Kaufmann, 2001.

13. R. Agrawal & R. Srikant, Fast Algorithms for Mining Association Rules, IBM Research Report RJ9839, IBM Almaden Research Center, 1994.

14. Kambatla, K., Kollias, G., Kumar, V., and Grama, A., "Trends in big data analytics," Journal of Parallel and Distributed Computing, in press.

15. Kwon, 0., and Sim, 1. M., "Effects of data set features on the performances of classification algorithms," Expert Systems with Applications, vol. 40(5), pp. 1847-1857,2013.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

A Survey on Knowledge representation learning

Dr. Pallavi S. Deshpande[1*]

[1]*Associate Professor, Department of Electronics and Telecommunication,Bharati Vidyapeeth (Deemed to be University)College of Engineering,Pune*

Abstract

Knowledge representation is among AI's most basic questions as well as one of its most essential notions. Extension data mining, on the other hand, generates transformable information, that expands knowledge base required for the generation of an extension strategy. Utilizing finite automata, we now have a novel way to describe knowledge in the context of the data mining. This paper offer a brief review of the structure and techniques that have been developed in this different discipline.

Keywords: Data Warehouse, Data Mining, Clustering, Data Integration, Pattern evaluation, Knowledge representation, Retrospective tool.

1. INTRODUCTION

In the past, knowledge representation techniques in Artificial Intelligence (AI) provided powerful mechanisms for the hierarchical storage of knowledge and to manipulate them with deductive inference. Such systems used intelligent rules to perform effective clustering and classification of documents so that it is possible to perform fast retrieval. However, the storage techniques were used only in knowledge-based systems but not in database systems. Hence, the natural language sentences used in e-Learning scenario are to be stored in database systems. In such a scenario, suitable indexing techniques are not provided for cloud databases. Therefore, the knowledge representation techniques from AI can be used for logical representation of data and a corresponding mapping to the physical database can be made to store the data in the database systems.

[*] ISBN No. 978-81-955340-6-7

It is widely accepted in cognitive research that there are indeed two primary methods for representing knowledge: First, there is also symbolic approach, that also uses nodes to represent concepts but also arcs to show relationships in between concepts to represents knowledge visibly; second, there's also the descriptive logic approach, where it uses more formal language to construct categorization definitions but also algorithms to decide on relationships in between the concepts [1]. A conventional notion of knowledge representation is used by both methods, in that concepts are defined using a set of qualities that are either adequate or essential [2]. When it comes to building large-scaled knowledge bases, these approaches have proved effective, but there are substantial issues that prevent their practical applications to the agents functioning in a social context. Among the issues are:

- Costly and time-consuming idea acquisition
- Difficulty in defining and identifying features
- A lack of expressive capacity to adequately reflect the circumstances

Representation of the knowledge in artificial intelligence is called Knowledge Representation (KR). An intelligent agent's ideas, intentions, as well as judgements are examined to see whether they can be articulated in a way that can be used by machines. Modeling intelligent behaviour for just an agent is a major goal of (KR) Knowledge Representation.

Using a method known as Knowledge Representation and Reasoning that is (KR, KRR), the information from the actual world is represented so that a computer can comprehend and use it for solving real-world issues, such as conversing with humans in natural language. Rather than just storing information, AI's approach to the knowledge representation enables a computer to draw on that material and develop the same level of reasoning abilities as a person. [3].

AI must be able to represent a diverse array of knowledge:

- Objects
- Events
- Performance
- Facts
- Meta-Knowledge
- Knowledge-base

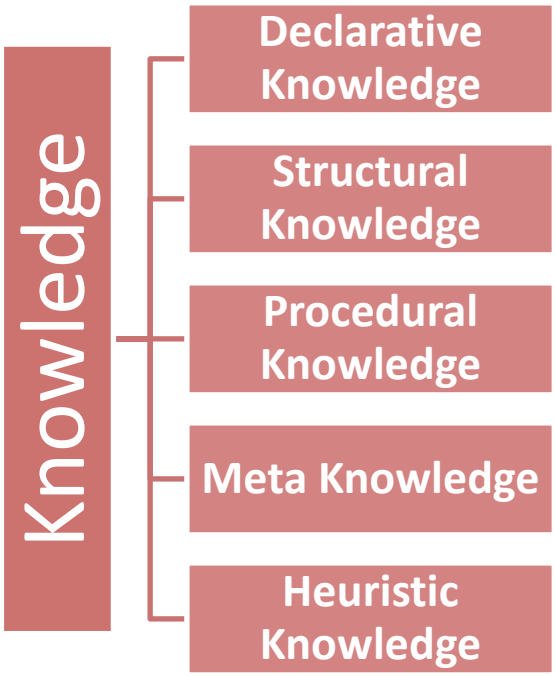

Figure 1: Different Types of Knowledge

- **Declarative Knowledge** – It is a declarative phrase that incorporates ideas, facts, as well as objects.
- **Structural Knowledge** – Knowledge of the link between ideas and things is a fundamental problem-solving skill.
- **Procedural Knowledge** – Having a working knowledge of how to accomplish a task is the responsibility of this term.
- **Meta Knowledge** – In those other words, "meta knowledge" refers to information about the other sorts of information.
- **Heuristic Knowledge** – This demonstrates a certain level of expertise in the sector or subject matter.

2. The relation between knowledge and intelligence:

Intelligence and the creation of artificial intelligence both rely heavily on knowledge of the actual world. In order for AI agents to seem clever, they must have access to knowledge. Only by having some prior information or experience of the input can an agent correctly act on it [4].

How would you respond if you were confronted with a stranger who spoke a language we do not really understand? The same holds true for agents' intelligent behaviour.

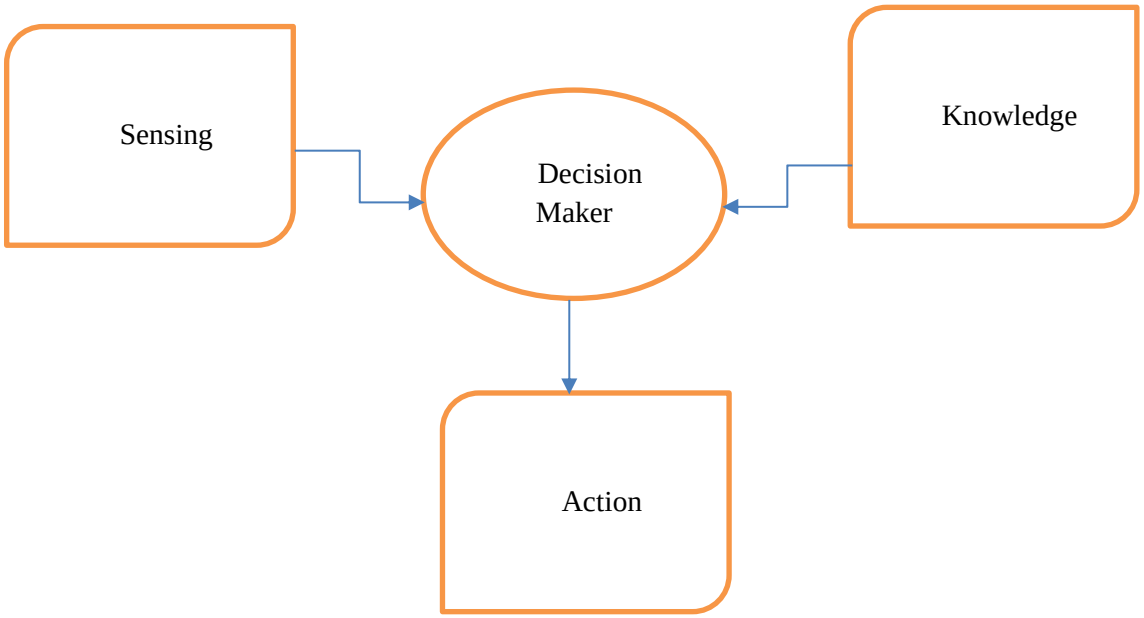

Figure 2: Example of intelligent behaviour

As seen in the flowchart above, the decision maker makes decisions based on information gleaned from environmental sensors and experience. However, intelligent behaviour cann't be shown if indeed the knowledge component is absent.

3. Life Cycle Of Knowledge Management

There are normally eight phases in the life cycle of a knowledge management project. Starting with the production of knowledge and ending with its use [5]. The eight distinct phases of information management, as well as a tracking function, are all required for managing information, even whether something multimedia for the marketing or even the heuristics for decision-making. Knowledge Management life cycle is made up of these phases:

1. Knowledge creation or acquisition
2. Knowledge modification
3. Immediate use
4. Archiving

5. Transfer
6. Translation/repurposing
7. User access
8. Disposal

The idea of "knowledge" is undergoing significant transformation in light of the fast growth of the knowledge-based economy. The first step in any discussion about knowledge management is to understand what the term "knowledge" really means. Facts about the concepts of the "knowledge" that have been widely accepted include: the significance of awareness throughout an acquisition of knowledge, inseparability of the knowledge as well as its carrier, the distinction among knowledge as well as data, and the preconditioned nature of other forms of knowledge, among others. As seen in Table no. 1, each degree of knowledge has an own hierarchical structure. In the first layer of a hierarchy, we have data. Objects or even events may be described in terms of their data. They don't have a clear goal in mind. Sometimes, data isn't included within the activity at all. Aside from fixing a specific issue, data collecting is sometimes carried out without regard. The information has been at the very second level of hierarchy. In contrast to data, information has a more precise purpose wherein the data are integrated with process of it's own processing in order to achieve. Questions like 'where,' 'what,' 'when,' 'who,' and 'how' all start with "what," "who," and so on.

Table 1: Levels of Knowledge

Level Example	**Execution of the order on employment**	**Purpose**
Data	There are many things to consider, such as employee's name, date of hiring, educational background, department, and position.	Nothing of interest to learn here.
Information		Understanding - what
Knowledge	Request towards data base of the young workers employed in last year (identity of employee, date of hiring, qualification, division, job)	Understanding -how To find out why
Wisdom	Upon being hired, the order is automatically fulfilled.	Nothing of interest to learn here.

Knowledge is found at the apex of the food chain. The "how" issue is answered through training as well as the dissemination of knowledge. Rather than just being a collection of facts, knowledge is a state of mind that includes the capacity to synthesise information from many sources and organise those efforts in order to achieve a certain goal. A "why" inquiry is indeed fourth degree of knowledge in hierarchy. Wisdom may be defined as "meta-knowledge," example, "knowledge of the knowledge" itself. In the cognitive psychology, "wisdom" has a broader definition. These human skills are often connected with the wisdom:

- Achieving a successful outcome by balancing the many interests;

- To integrate new information by considering opposing viewpoints
- To recognise your own errors and limits in knowledge;
- To identify and articulate the difficulties (as the consequence of the creative thinking).

After moments of optimism and predictions of quick advancement, AI as just a scientific subject has experienced periods of stagnation or e decline whenever the findings acquired began to sound unconvincing as well as investment inside the research activity significantly fell. When artificial intelligence (AI) approaches were put to use within real economy inside 1970s, the rise of expert systems as well as knowledge-based systems (KBS) was clearly linked to "success stories."

The distinction among "information" as well as "knowledge" is critical in the notion of the knowledge representation, which is the basis for developing KBS (see Table no.-I). Facts concerning the precise characteristics of things, events, as well as processes are known as data. The term "raw data" refers to the information that a computer collects, saves, plus processes. Data is stored in databases (DBs) at KBS, which are known for their large volume and cheap cost per information unit. Information is a tool for elucidating the truth for the general public (data). "But what's it? How does it work? ", these are the kind of questions that information provides answers to. Querying databases for information is one instance of how data may be accessed.

In the context of an application domain, knowledge is described as the objective rules that enable experts to create and solve issues. Because knowledge is utilised to make decisions, it is indeed an understanding of how one should behave. Practice as well as professional experience are the means through which one gains knowledge. Knowledge may also be defined as "meta-data" or even "data about data," which is well-structured kind of data. Knowledge bases (KBs) are being used to store an information, which have a relatively modest volume as well as a high cost.

4. Artificial intelligence impact in knowledge management

The effect of AI on (KM) which is Knowledge Management is wide-ranging, but we can say that AI has such an impact over the availability and speed of real-time processing of the big data, the ease with which knowledge could be discovered, the improvement of the customer relationships, the maintenance of current content throughout knowledge bases, as well as the improvement of company management. Big data may now be accessed and processed in real time thanks to AI. Artificial Intelligence (AI) is increasingly being used to handle the massive amounts of data that are now being exchanged between people and divisions of a company, both locally, nationwide, and worldwide, and to swiftly turn them into meaningful information. Unstructured data like text, photographs, videos, and other visual representations may be analysed by AI to uncover new patterns and insights. Natural language processing, even Semantic search including machine learning are some of AI tools that make it simpler for workers to get information they need. Artificial intelligence (AI) makes it easier to find information

in a company's vast knowledge base. Employees may search company's knowledge base utilizing natural language thanks to semantic search as well as natural language processing in the AI. Machine learning tracks search phrases as well as user behaviour throughout the time to make predictions well about kinds of things people are searching for in workplace. By linking data from many sources, AI makes information more accessible. Artificial intelligence aids workers who are unsure of where to seek for information. Workers may link and aggregate knowledge across many systems using AI-powered software, allowing all the employees to access information regardless of the where it is housed. Customer service may be greatly improved with use of AI, which can create and use fresh information about customers' behaviour as well as demands. Search engine optimization (SEO), to enhance organic search performance, marketing for content, to personalise content output for every user as well as, generally, to adapt towards the needs of user depending on learning from the previous behaviour, are instances of AI have been used in the Chatbots, the mapping possible solutions gathered from past conversations. Archetypes with in data may be recognised by AI, allowing for better segmentation as well as handling of customers.[6].

Utilizing machine learning technology, AI aids workers in locating the most current information in company's knowledge base. User-rated information would no longer be sent by AI as well as will instead be replaced with content that better satisfies user intent. When AI is utilised to facilitate or enhance managerial activities, it may strengthen overall management of organisation. Artificial intelligence (AI) models that self-learn enable businesses to swiftly adapt following changes within patterns of an internal or external data as well as underlying economic circumstances. When you utilise AI to maximise value of your data as well as go from the predictive analytics towards prescriptive analytics, you can make better as well as quicker choices [6]. The following are some examples of AI-powered toolkits: [7] considers 1) Evie for the meeting scheduling, 2) Aiden AI for the growth empowerment, 3) Datasine for the email content personalization, 4) Attest for the strategy validation, 5) Unito for the work maintenance, 6) Freeagent for financial and accounting simplification, as well as 7) Bizplan for business planning perfection] as the most important.

Artificial Intelligence has been shown to have a positive influence on the management of the knowledge in a variety of ways. "Artificial intelligence has already brought a huge contribution to the excellence and effectiveness of KM in term of thinking and problem-solving methods, but also through knowledge acquisition, modelling and processing, decision support systems, intelligent tutors, planning, scheduling and optimization systems"[8].

Conclusion

The advancement of information technology has made it possible for us not just to collect data but also to transmit it to appropriate people in real time and error-free. Such raw facts are transformed into information and subsequently into knowledge via a series of intermediary processes in the Knowledge Management strategies. Experts in their fields play an important part in this. Predictions and forecasts

based on the knowledge management inputs are generated by artificial intelligence programmes. These results are put to good use in a wide range of fields, from medical diagnosis to stock market forecasting. The quality of artificial intelligence (AI) systems' output is directly related to the quality of its knowledge management. For even more precise estimates of any of the system, particularly business-related systems that already have a lot of data, information technology must be used to develop these principles. Researchers in relevant domains have been expected to do so.

References

1. S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach. Pearson Education, 2003.
2. Z. Kunda, Social cognition: Making sense of people. MIT Press, 1999, pp. xi, 602.
3. Mylopoulos, John, and Hector Levesque. "An overview of knowledge representation." *GWAI-83* (1983): 143-157.
4. Rothberg, Helen, and G. Scott Erickson. From knowledge to intelligence. Routledge, 2007.
5. Mishra, Brojo Kishore, and Sushanta Kumar Das. "AI Techniques in Knowledge Management." (2011).
6. Akerkar, R. “Artificial Intelligence for Business. Cham: Springer” (2019)..
7. Posser, D. “Seven AI-Powered Tools To Help Start-ups Grow.” Retrieved from https://www.forbes.com/sites/davidprosser/2019/07/26/seven-ai-powered-tools-to-helpstart-ups-grow/. (2019).
8. Mercier-Laurent, E. (2015). Artificial intelligence for successful Kflow. In IFIP International Workshop on Artificial Intelligence for Knowledge Management (pp. 149-165). Cham: Springer.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

A Survey on Big Data: Technologies, Trends and Tools

Dr. Sarika A. Panwar[1*], Dr. Pallavi S. Deshpande[2]

[1]*Assistant Professor, Department of Electronics and Telecommunication Engineering, AISSMS Institute of Information Technology, Pune-01*

[2]*Department of Electronics and Telecommunication Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune-43*

Abstract

Data sets which are too huge or complicated for typical data processing tools, such as relational databases, are referred to as "big data." From the outset, big data has been at the core of companies such as Ebay, Google, LinkedIn, and Fb. Massive as well as complicated data sets, including social media analytics as well as data management skills, as well as real-time data, are included within the collection. The complexity of big data necessitates the development of new methods, algorithms, and analytics for their management and analysis, as well as for their value creation and information extraction. The primary goal of this article is to provide an overview of the current status of Big Data research. In addition, we'll talk about the latest in technology as well as tools, as well as potential problems and emerging trends.

Keywords: Big data, Big Data Quality, Big Data Quality Dimensions, Big Data Analysis.

1. INTRODUCTION

Semi-structured, Structured, as well as unstructured data all fall under the umbrella of "big data." A structured form of such data refers to data that has been properly formatted or tabulated [1]. The data that includes both text and pictures fall underneath the semi-structured category. Unstructured data is any data that does not follow a predetermined structure, such as text, photos, or videos. Databases

* ISBN No. 978-81-955340-6-7

designed for the relational data cannot handle the billions of records that this sort of data includes. As a result, Big Data Analysts must use additional tools and methods to do this [2]. As a result, dealing with tools as well as approaches is a highly challenging work for the Big Data Analysts.

Instances of the Big Data in an action may be seen here:

- In order to get a better understanding of their customers' habits, preferences, as well as perceptions, firms and retailers are using social media such as Facebook as well as Twitter.
- As equipment degrades, manufacturers may track the tiniest variations in the vibration data to determine when to replace or repair it. If it's replaced too early, money is wasted, and if it's replaced too late, a costly work halt result.
- Social media is being used by manufacturers in such a different way from marketers: They are looking for difficulties regarding warranty support prior they can become public.
- Governments around the country, states, and cities are making data available to public so that citizens may build innovative apps which can better serve the people.
- In order to produce more relevant as well as intelligent offerings, financial institutions are using data extracted from client interactions to divide their consumers into the finely calibrated categories.
- Advertisers as well as marketing firms are keeping tabs on social media to assess that house insurance applications can indeed be handled promptly, as well as which ones require a face-to-face visit to verify their authenticity.
- Retail companies are well engaging brand champions, altering the perspective of brand adversaries, or even allowing passionate consumers to pitch their items. •. Using social media has made all of this possible.
- Hospitals use medical data as well as patient records to forecast which patients seem to be likely to return within the next few months following release. So, the hospital is able to avoid another expensive hospital stay.
- Web-based firms are creating information products which aggregate client data in order to provide more enticing suggestions and more effective discount programmes.
- In addition to analysing sales of ticket , sports clubs are now employing big data to monitor team strategy.

Fig. 1 shows the four stages of such big data processing.

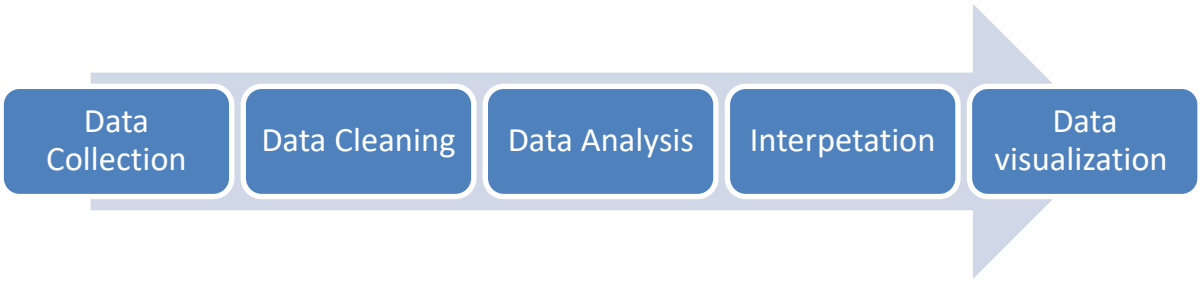

Figure1: Four stage process of Data Mining

- **Data Collection**: Collecting data from various sources should be based on the specifications you've established. This information may be gleaned through a wide range of resources, such as surveys as well as interviews as well as direct observation. For the analysis, it is important to arrange the data you have gathered.
- **Data Cleaning:** It's time to go through the data you've gathered to see what you can utilise. During this step, you'll be removing any empty spaces, duplicated records, as well as other typographical problems. The data must be cleaned up before it can be used for any further investigation.
- **Data Analysis.** Data analysis software as well as some other tools are being used to assist you comprehend the data as well as the draw conclusions. Xls, Python language, R,Rapid Miner, Looker,Chartio, Redash, Metabase,and Microsoft Power BI are among the data analysis tools.
- **Data Interpretation:** Now that you've gotten your outcomes, it's time to analyse them and choose the best next steps depending on what you've learned.
- **Data Visualization:** The term "data visualisation" is indeed a fancy way of stating "graphically present the information in a manner that others can read as well as comprehend." "A variety of tools are at your disposal, including charts, maps, graphs and bullet points. In order to get significant insights, visualisation lets you to compare datasets as well as discover the links between them.

2. BIG DATA MANAGEMENT

Inside the Big Data-related project, how should it be managed and developed? What kind of design should we use to keep track of all the different parts of Big Data? The structure of Big Data should be linked with firm's support infrastructure. Data is being generated by a variety of sources, many of which are unreliable, loud, and dirty. Here, we'll quickly touch on Hadoop, as well as the other management tools, in just this part.

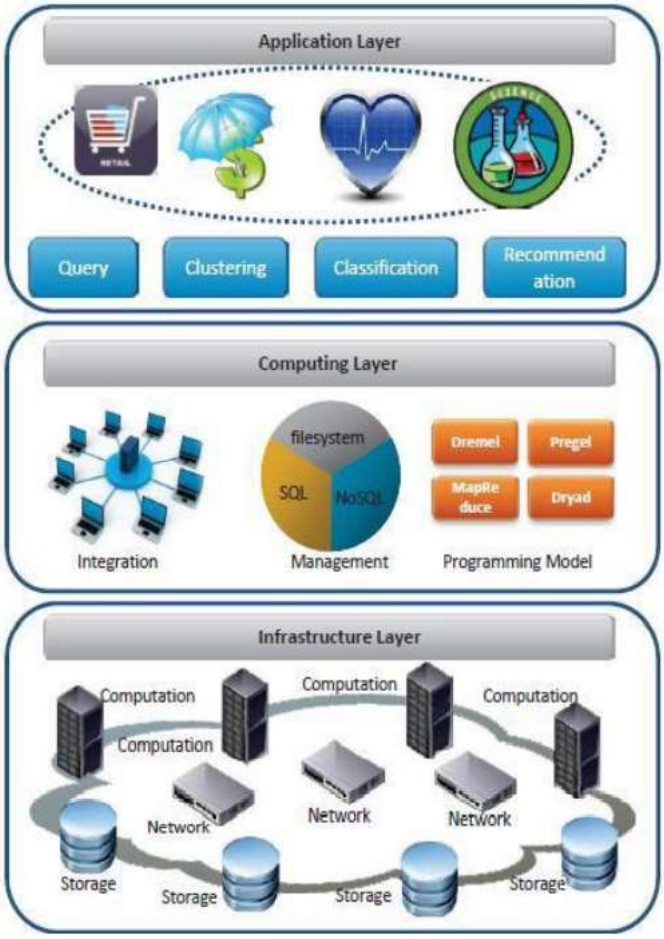

Figure 2: A layered Big Data Architecture

Interactions between systems, data storage, including network devices external towards the system are all handled by the infrastructure layer, which also responds to requests for the data retrieval through other levels like computing. The physical layer of the Big Data is another name for this layer. An abstraction of such underlying data layer is provided through this Mid Layer, which facilitates access to as well as retrieval of data. Distributed storage devices are organised by indexing data in just this layer. Multiple repositories are used to arrange data into chunks. The analytics or even application layer includes the necessary tools and methodologies, as well as the logic needed to provide domain-specific analytics. Another name for this layer might be "Logical layer."

Storage management has a wide variety of tools as well as approaches to choose from. Simple DB, Google Big Table, MemcacheDB, NoSQL are some of the options. [4].

2.1. Hadoop

For the search engine initiative, Doug Cutting and Mike Cafarella initiated a project that would index approximately 1 billion pages. A Google File System, or GFS, was first established in 2003 by the search giant. Later that year, Google released Map Reduce architecture that served as the basis for the Hadoop platform. MapReduce as well as HDFC are at the heart of the (Hadoop Distributed File System) Hadoop system. Let us take a quick look at this Hadoop component in this part.

Hadoop's Design Principles It comprises mostly of four parts.

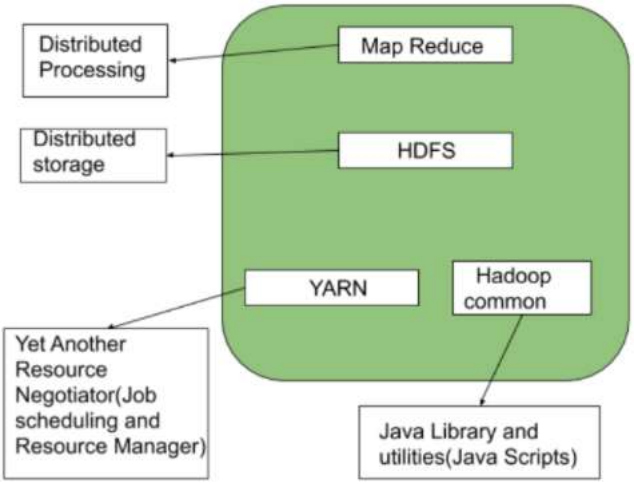

Figure 3. Hadoop Architecture with 4 components

- HDFS (Hadoop distributed File System)
- MapReduce
- YARN (Yet Another Resource Framework)
- Common Utilities or Hadoop Common

A. HDFS:

HDFS is indeed the Java-based file system for large-scale commodity server clusters which enables scalable as well as dependable data storage. There are two sorts of nodes within a cluster. There is indeed a master node just at top of the tree, which is the name node. It is also possible to have a data node acting as a slave node. The default block size for HDFS is 64MB. In order to analyse enormous volumes of data simultaneously, these files are duplicated in multiples.

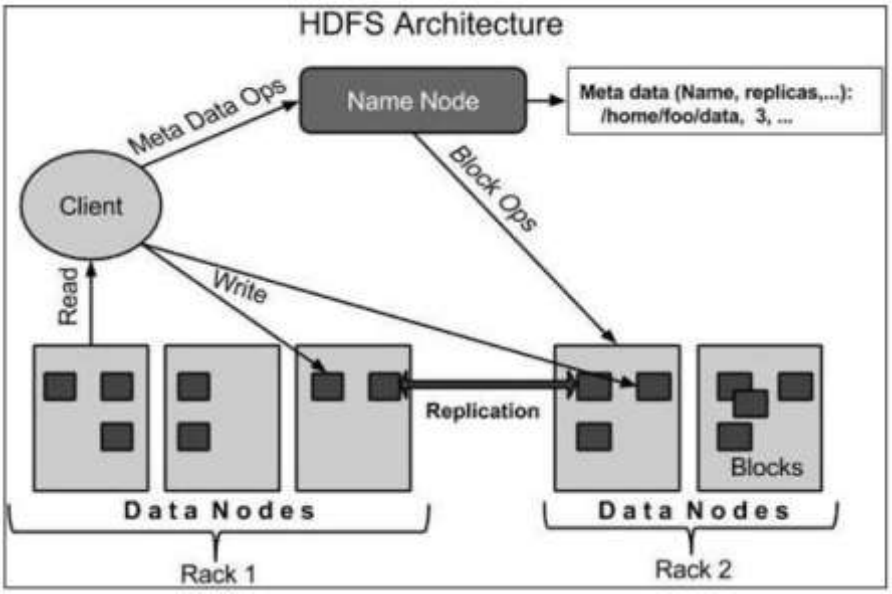

Figure 4. HDFS Architecture [5]

B. MapReduce

Java-based Map Reduce is indeed a distributed computing programming methodology. It's a method of preparing something for consumption or use. Map as well as Reduce are key components of Map Reduce algorithm. The Hadoop program's Map that Reduces function really consists of two discrete and different operations. In the first place, there's the map task, that takes a collection of data as well as turns it into the new set of data under which such individual pieces are broken down further in the tuples (key or the value pairs). The result of a map is used as input for a reduction operation, which merges the tuples into the smaller set. Map and reduce jobs are always executed sequentially, in accordance with name Map Reduce.

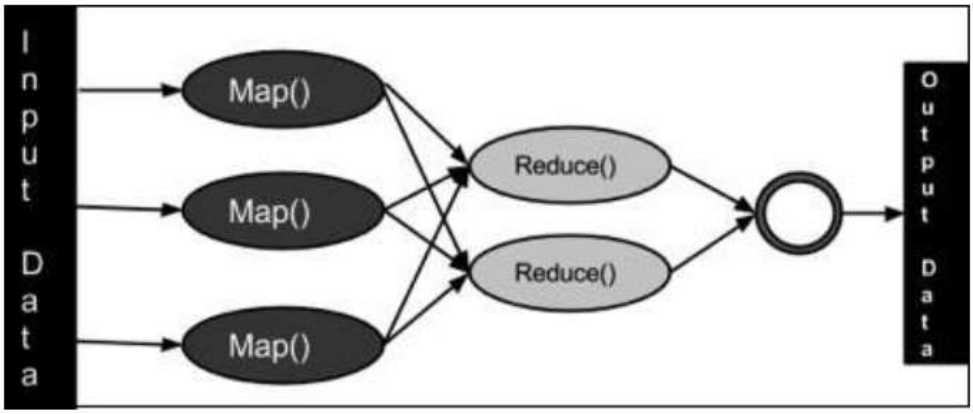

Figure 5. MapReduce Architecture [6]

C. YARN

It stands such Yet another Resource Negotiator, and it was introduced within Hadoop version 2 as a way to manage clusters. Big data applications use YARN, which is today regarded as a large-scale, widely dispersed operating system. It is referred to through Apache as a new resource management. Apps run on resources provided by YARN Infrastructure, such as CPUs and RAM that may be accessed by the applications themselves. A resource manager as well as the node manager are two sub-parts of the manager. The master is indeed Resource Manager (that is one for each cluster). They understand where slaves are (Rack Awareness) as well as how much they have at their disposal. The Resource Scheduler, which determines how resources are assigned, is perhaps the most critical of the many services it provides. The infrastructure is the master of the Node Managers (of which there are several in a cluster). Whenever it begins, the Resource Manager is notified. The Resource Manager receives a pulse from it on a regular basis.

Features of YARN

- Multi-Tenancy
- Scalability
- Cluster-Utilization
- Compatibility

D. Common Utilities or Hadoop Common

Inside the context of Hadoop cluster, "common utilities" refers to the java library as well as Java files, or even "java scripts," which we require to run all of the other's components in the system. YARN, HDFS, as well as Map Reduce all rely on these tools to operate the cluster. Because hardware failure is so prevalent among Hadoop clusters, the problem must be primarily managed in software through the Hadoop Framework, according to Hadoop Common.

3. BIG DATA APPLICATION

These properties may be derived from the vast amounts of data that can be analysed via Big Data Analysis. An overview of Big Data use cases is provided in the section.

a) **Text Data Analysis**

Text mining is yet another term for the text analytics. To get more information about clients fro the unstructured data sources, text analytics may be used. Computational linguistics, Machine learning and statistics all play a role in the text analytic projects. With the use of text analytics, companies can turn massive amounts of the human-generated text into digestible summaries that aid in fact-based

decision-making. If you want to anticipate the stock market, for instance, you may use the text analytics for extracting information from the financial news. (NLP) Natural Language Processing is used in the majority of text mining approaches (NLP). Text analysis, interpretation, and generation are all possible with NLP. Certain NLP-based text mining approaches have been implemented, such as the extraction of data, the creation of topic models, the summarising of text, the clustering, classification the answering of questions, and the gathering of opinions.

b) Social Media Analysis

There are a variety of approaches used to study social networks, that are networks which are made up of persons or organisations that are interdependent in some way, whether it's via friendship, a shared interest or monetary transactions. Nodes and linkages connect each other within a social network. A network of nodes (actors) as well as connections (relationships between nodes) forms the basis of the social structure, which may be shown as a network diagram. The actor, relation, as well as the network are indeed the three main building blocks of social network systems. Categories of social media services include link-based as well as content-based analyses [8]. It's always been the goal of link-based structural analysis to focus on link the prediction, community discovery, the development of social networks, and social impact analysis and many more.

c) Mobile Data Analysis

It is the process of analysing and acting on user behaviour data in order to increase, engagement, user retention and conversions in mobile applications. A new field of study has opened up as a result of advancements in the wireless sensor, or even mobile communication, and the stream processing technologies. In addition to health as well as business-related applications, there are a number of additional Big Data application areas, like surveillance monitoring, weather forecasting, as well as the multimedia data analysis.

4. BIGDATA CHALLENGES

Big Data's heterogeneity, size, complexity, timeliness, and privacy issues inhibit development at all stages of such data pipeline which might provide value. Currently, we are forced to make judgments on the ad hoc basis about which data to preserve and which to discard, as well as on how to consistently save data we do keep with appropriate metadata, due to data tsunami. For example, unlike pictures as well as video, which are stored and shown in organised formats, most data today also isn't inherently structured. For semantic material, though, you should instead search for keywords. The major challenge is to organise this data so that it can be analysed in the future. In the context of other data, the value of a piece of information increases exponentially. A key source of added worth comes from data integration. Today, the vast majority of data is produced digitally, giving us the possibility and the task of both influencing the development of new data as well as automatically linking previously-made data to make new connections. Other core issues include data analysis, organisation, recovery, and modelling. Due to

the intricacy of such data which has to be processed, as well as the limited scale of an original method, data analysis is often a bottleneck in many applications. Finally, non-technical domain specialists are essential to deriving actionable Knowledge from the data.

CONCLUSION

Various technologies for dealing with huge data have been examined in such research. Utilizing HDFS Hadoop distributed data storage as a framework, this article explores Big Data architecture and explains its many components. The mining sector, for example, finds a lot of value in the big data. Big data isn't a luxury for an industry that conducts billions of dollars in commerce every year, it's a need. Overall system as well as application performance was the primary focus of this article, that included a review of several Big Data architectures and their handling approaches for dealing with large amounts of data from multiple sources. Big data has had a significant impact on the corporate sector. It is possible to exploit big data in a variety of ways, and in ways that many people have never considered before. The mining sector, for example, finds a lot of value in the big data. Big data isn't a luxury for the industry that conducts billions of dollars in commerce every year, it's a need. The algorithms for effectively and swiftly mining large amounts of data are always being improved by researchers. In addition, a look at some of the analytics as well as management tools being provided.

References

1. Arti Chandani, M. M. (2015). Banking On Big Data: A Case Study. ARPN Journal Of Engineering And Applied Sciences , 4.
2. Basvanth Reddy, P. B. (2016). Weather Prediction Based on Big Data Using Hadoop Map Reduce Technique. International Journal of Advanced Research in Computer and Communication Engineering , 5.
3. Keshav Sanse, Meena Sharma, (2015).Clustering methods for Big data analysis,(IJARCET) Volume 4 Issue 3, March.
4. Min Chen Shiwen Mao Yunhao Liu, (2014). Big Data: A Survey,Mobile Networks and Application 19:171:209.
5. Sharma, Sugam, et al. (2014): "A brief review on leading big data models." Data Science Journal 14-041.
6. Jena, Bibhudutta, et al. (2016) "Improvising name node performance by aggregator aided HADOOP framework." 2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). IEEE,

7. Chung, W. (2014). BizPro: Extracting and categorizing business intelligence fac-tors from textual news articles. International Journal of Information Management,34(2), 272284.
8. Aggarwal CC (2011) An introduction to social network data analytics. Springer.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

A Review of Challenges and Security Risks of Cloud Computing

Dr. Algubelly Yashwanth Reddy[1*], Dr. P. Hasitha Reddy[2]

[1]*Head of the Department, Computer Science and Engineering, Sree Dattha Group of Institutions, Hyderabad*

[2]*Assistant Professor, Computer Science and Engineering, Sree Dattha Group of Institutions, Hyderabad*

Abstract

Because of its capacity to cut processing expenses, cloud computing is now one of the most hotly debated topics in the field today. The most intriguing and alluring technology of the present day is that which provides its customers with the ability to request services online. Because cloud computing stores as well as disseminates data in an environment, security has emerged as the primary impediment to the widespread use of the cloud computing settings. Many people utilise cloud to store our personal data, which necessitates a high level of the data storage security over media. When data is being uploaded to a cloud server, security is a key issue. The security dangers and problems of cloud computing are examined in just this review article, which also examines the security standards for cloud computing. Accordingly, the main purpose of this analysis is to categorise cloud computing security threats and issues.

Keywords: Cloud, security, solutions, challenges, risks.

1. INTRODUCTION

Computer services in the information technology such as platforms, infrastructure, or even applications may be set up and accessed through internet under the umbrella term "Cloud Computing. It is the large scaled distributed infrastructure in that shared resources are virtualized as well as services given to customers are dispersed in the terms of such virtual machines, deployment

* ISBN No. 978-81-955340-6-7

environments, or software. As a result, it's clear that cloud services may be dynamically scaled based on the current workloads as well as needs. According to the amount of resources that are consumed, the payment is paid based on that usage. One definition of 'cloud' is a'remote data centre,' as per [1]. There are two ways to look at it. First one is the use of a web browser to access information as well as data resources through the Internet. Secondly, the computer resources are paid for on a per-use basis. When you hear about "cloud computing," you're most likely thinking about the (NIST) National Institute of Standards and Technology since “a model for enabling universal, suitable, on-demand network access to share pool of configurable computing resources (e.g., servers, networks, applications, storage, and services) that can be rapidly released and provisioned with minimal management effort or service provider interaction” [2].

1.1. TYPES OF CLOUD

The range of the cloud computing architectures is vast, as well as the customer may pick which one best suits their needs and budget. It's basically as follows when it comes to cloud architecture [3].

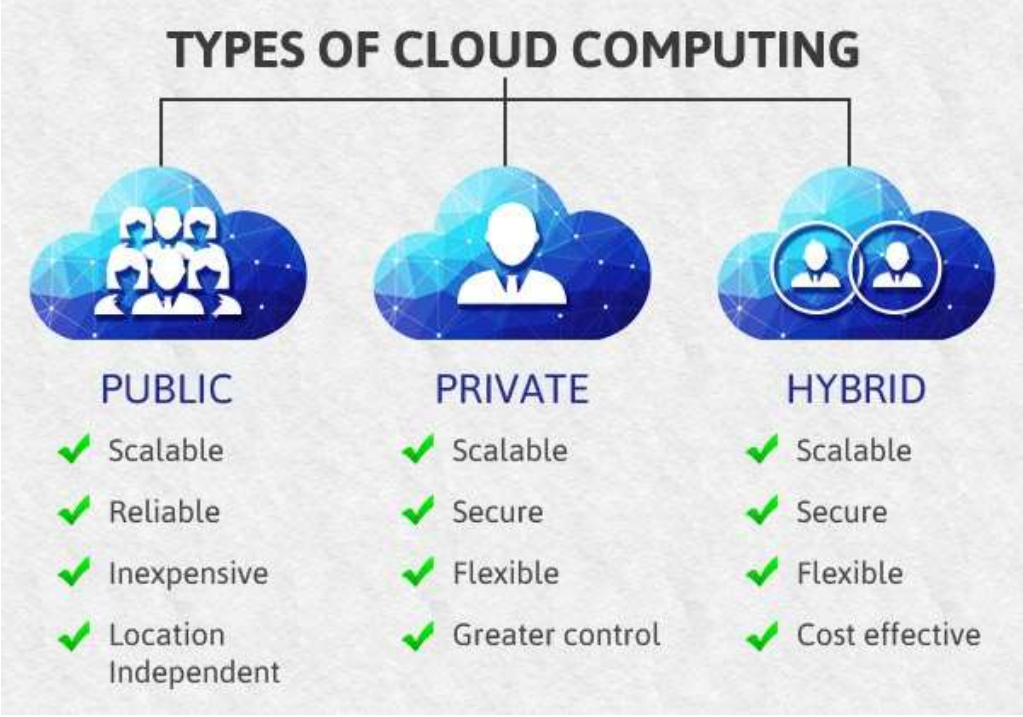

Figure1: Types of cloud [4]

- **Public Cloud:** Pay-as-you-go cloud computing is a sort of the cloud computing architecture that is designed for the customer in a pay-per-use paradigm. Google App Engine, or even Microsoft Azure and Amazon Web Services seem to be instances of this kind of service.
- **Private Cloud:** Designed for vital infrastructure as well as private businesses and organisations,

this sort of the cloud computing architecture. The cloud computing environment like this one is not open to the general public. A private cloud is indeed an instance of such a data centre owned by a private company or a government agency.

- **Community Cloud:** Third parties may use it to construct new apps as well as platforms for new services, and then it involves a wide range of diverse parties.
- **Hybrid Cloud:** There are public as well as private clouds in just this cloud computing structure. (NIST) that we called as National Institute for Standards and Technology defines it as "a hybrid cloud that has a combination of public and private clouds bound together by either standardized or proprietary technology that enables data and application portability" [2].

2. TYPES OF CLOUD SECURITY

- **Identity Security:** In terms of privacy as well as work, it is referred to as a technique. " allows the authenticate people to retrieve the resources at the appropriatet time and for the good objectivess" [5]. It protects the privacy as well as security of data and applications while increasing their availability to authenticated individuals.
- **Information security:** Regardless of whether such data is encoded, transferred, processed, or the deposited, business practises must be in place to ensure that the data is protected at all times. [6].
- **Network Security:** The safety of a computer network is a precondition for its operation. Defending the current network infrastructure against unvarified individuals, violation, adjustment, breakdown, deterioration, or incorrect distribution is part of this process [7]. Because to issues at network level, web system's capacity as well as responsiveness might suffer, as well as overall latency.
- **Software Security:** It is necessary to start with the idea of the programme and work through the design and execution phases to build a security analysis procedure. In order to provide the highest degree of software security, each of these procedures is dependent on the others [8]. Even though complexity of such software development varies widely, everything requires a security assurance.
- **Infrastructure Security:** In order to verify the company's operations, it is essential that perhaps underlying infrastructure be secure. [9] Elements also need to be maintained apart. Users of the network may avoid having accessibility easily to the memory drivers or cryptographic codes by separating modules from the administration.

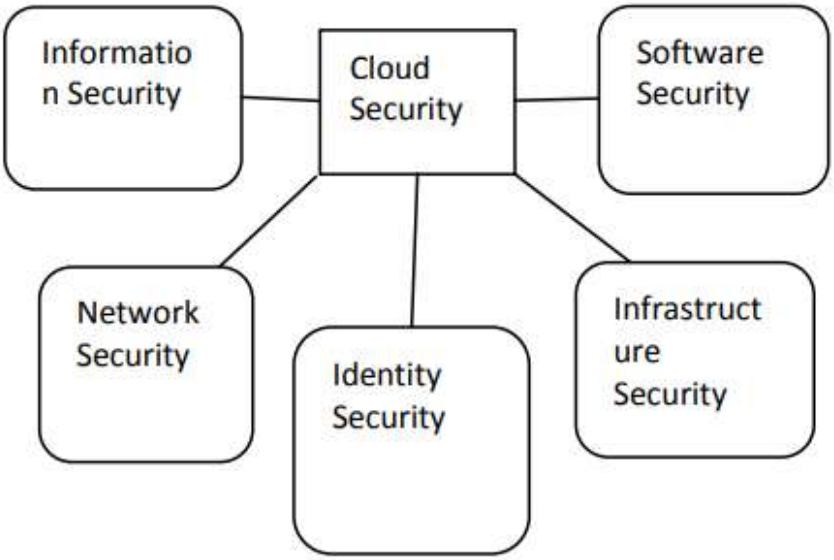

Figure 2: Types of Cloud Security.

Solutions: Cloud computing's security has been beefed up with the addition of a new encryption technique. The most common forms of encryption are attribute-based encryption, including homomorphic encryption, as well as symmetric encryption. Text security cypher or even key code are two options for attribute encryption. Encrypted messages and secret numbers billing encoded text that a customer leaves behind for the decryption are explained here. It is possible to decrypt encrypted material using the homomorphic encryption in CC. In order to safeguard sensitive data, it is necessary to use primitive cryptography in symmetric encryption. In order to provide high data protection, several types of encryption may be augmented with active solutions.

3. ATTACKS AND THREATS IN THE CLOUD SECURITY

The following is a list of cloud security assaults as well as threats:

- **Account and service hijacking:** Considered to be among the most dangerous dangers, Attackers target the web service hosted by the cloud service provider, as well as then install its control software inside cloud service provider infrastructure so that they may take control of such web service once it has been compromised. [12].
- **Backdoor Channel attacks:** This form of attack occurs in IaaS, whenever it delivers an effective customer's high penetration just on VM’s or even Hypervisor level. This may impair service availability as well as data privacy [13].
- **Cross site scripting attacks**: XSS is another name for it. It is among the most potent web

application security vulnerabilities that has been discovered to date. JavaScript is among the most often utilised programming languages in this kind of attack [12].

- **Cloud malware injection attack:** Injecting malware, the macules programme, or even a virtual machine into cloud infrastructure is among the most dangerous things you can do to your cloud computing security [14].
- **Denial of Service attacks:** When consumers try to use service after it has been compromised, it would not be accessible. They'll receive a 404 which is Not Found error [12] as a result. [12].

4. CLOUD COMPUTING CHALLENGES

- **Access controls:** Service providers should be concerned since it might expose user data as well as provide hackers access to an organization's infrastructure. [15].
- **Accounting:** In order to sustain network administration, it's an important factor to consider while installing cloud computing services [16].
- **Compliance:** The approaches of the compliance management are not well supported by cloud computing. Data security as well as privacy might be compromised as a result of this [16].
- **Cross-Organizational Security Management:** In the cloud computing, achieving and maintaining security as well as SLA compliance is a major difficulty. Cloud computing security can't be accomplished without the cooperation of several organisations [17].
- **Extensibility and Shared Responsibilities:** Cloud computing security is a topic that has to be addressed by both service providers as well as end users. Until recently, there has been no clear picture of how the cloud computing's security duties would be met. [18].
- **Private Cloud:** Due to the phrase "private cloud" meaning "on-premises," the working environment is supposed to be the same as in conventional computing. Through the use of virtualization technologies for computing resources, the computing resources are virtually extendable or de-extendable depending on the user’s needs. This will give accessibility to shared resources for the entire departments in the organization. However, this has not been fully implemented in a wide range in the organizations. In other words, it is a halfway step to be implemented by the public cloud services [1].

Despite the fact that the concept of cloud computing has only just emerged. Cloud computing research is still in its infancy. Numerous concerns remain unresolved, as well as new difficulties continue to crop up across all industries on a daily basis. The follows are some cloud computing research issues to keep an eye on.

- Service level agreement (SLA)
- Cloud data management and security
- Data Encryption
- Virtual machines migration

- Access controls
- Multi-tenancy
- Reliability and availability of services.

5. FUTURE ADVANCES IN THE FIELD OF CLOUD COMPUTING

A concept termed as Automation is taking hold in IT business during the next five years. AI as well as ML are likely to play an important part inside automation process within next five years, at least. Programming-related professions in the IT sector are likely to become less common as automation advances. To illustrate the point, let's take a look at a hypothetical situation. When a computer replaces human brain in logic-building process, we can envisage the harm that will be done in terms of the INTRUSION. It takes a reasonable length of the time for a typical programmer to conduct (or) complete the specified task. Considering today's technology as well as the knowledge it gained via machine learning, a machine can do the identical task in a matter of seconds. Traditional IDS systems might be ineffective in this case because of the increased automation. As a result, there is a pressing need to bolster existing security measures, such as fire walls as well as IDS.

6. CONCLUSION AND FUTURE WORK

Using cloud computing is enticing because of its adaptability, effectiveness, usefulness, as well as cost-savings features. It's the newest and most promising technology, but it's also vulnerable to a variety of threats. Data security challenges and methods for dealing with them are discussed here in order to deal with the CC risks.

Recently, both the business and academic worlds have taken a keen interest in the cloud computing, which many see as a foundation for the contemporary societies of future. Economic improvements as well as cost savings may be achieved by using cloud computing. Empowering cloud computing qualities are sought by governments, organisations as well as enterprises. There are still many security as well as privacy concerns that must be addressed before cloud computing can be widely used in the next several years. The most recent threats and difficulties associated with cloud computing have indeed been explored in this study. Developing a security model that takes into account the dangers and difficulties that have been highlighted will be the focus of our next research.

References

1. Kim, W., et al. Adoption issues for cloud computing. in Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia. (2009). ACM.

2. Mell, P. and T. Grance, Draft nist working definition of cloud computing. (2009)
3. Subbiah, M., D.S.S. Muthukumaran, And D. Ramkumar, Enhanced Survey And Proposal To Secure The Data In Cloud Computing Environment. International Journal Of Engineering Science, (2013) .
4. Ali, Alwesabi, Almutewekel Abdullah, and Okba Kazar. "Implementation of cloud computing approach based On mobile agents." International Journal of Computer and Information Technology 2.06 (2013): 2279-0764.
5. S. Hajra et al., "DRECON: DPA Resistant Encryption by Construction," Springer, 2014, pp. 420–439.
6. A. Tripathi and A. Mishra, "Cloud computing security considerations," IEEE Intl.Conference on 90Signal Processing, Communications and Computing (ICSPCC), 2011, pp. 1–5
7. "SANS Institute: Network Security Resources." [Online]. Available: 91https://www.sans.org/network-security/. [Accessed: 16 Feb. 2017]
8. M. Al Morsy, J. Grundy, and I. Müller, "An Analysis of the Cloud Computing Security Problem,"APSEC Cloud W., Nov.2010.
9. K. M. Khan and Q. Malluhi, "Establishing Trust in Cloud Computing," IT Prof., vol. 12 (5), Sept. 2010, pp. 20–27.
10. N. Attrapadung, J. Herranz, F. Laguillaumie, B. Libert, E. de Panafieu, and C. Ràfols, "Attribute-based encryption schemes with constant-size ciphertexts," Theor. Comput. Sci., vol. 422, Mar.2012, pp. 15–38.
11. Younis, M. and K. Kifayat, Secure cloud computing for critical infrastructure: A survey. Liverpool John Moores University, United Kingdom, Tech. Rep, (2013).
12. Modi, C., et al., A survey of intrusion detection techniques in cloud. Journal of Network and Computer Applications, 36(1) (2013). 42-57.
13. SUBBIAH, M., D.S.S. MUTHUKUMARAN, and D. RAMKUMAR, Enhanced Survey and Proposal to secure the data in Cloud Computing Environment. International Journal of Engineering Science, (2013) 5.
14. Chou, T.-S., Security threats on cloud computing vulnerabilities. International Journal of Computer Science & Information Technology, 5(3) (2013) 79.
15. Zissis, D. and D. Lekkas, Addressing cloud computing security issues. Future Generation computer systems, 28(3) (2012) 583-592.
16. Moreno-Vozmediano, R., R.S. Montero, and I.M. Llorente, Key challenges in cloud computing: Enabling the future internet of services. Internet Computing, IEEE, 17(4) (2013) 18-25.
17. Khalil, I.M., A. Khreishah, and M. Azeem, Cloud computing security: a survey. Computers, 3(1) (2014) 1-35.
18. Zhang, L., et al., Cloud manufacturing: a new manufacturing paradigm. Enterprise Information Systems, 8(2) (2014) 167-187.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

A Survey of Issues and Approaches in Mobile Cloud Computing

B. Muthu Kumar[1*]

[1]*Professor, School of Computing and Information Technology, REVA University, Bengaluru, Karnataka, India, muthu122@gmail.com*

Abstract

(MCC) Mobile cloud computing refers to the provision of the cloud computing services inside a mobile context. MCC utilises both mobile networks as well as cloud computing to provide mobile users with the best possible experience. Mobile devices don't need to have strong configurations such as CPU speed, storage capabilities, and so on, since all data as well as difficult computing modules may be processed in the clouds under mobile cloud computing. An overview of MCC applications' needs and current solutions is presented in this study. We also cover architecture, applications, main features, security challenges, advantages and disadvantages, and maybe a solution for mobile cloud computing.

Keywords: Mobile Cloud Computing (MCC), Mobile environment, Mobile networks.

1. INTRODUCTION

A decade ago, the number of people using mobile phones was a fraction of what it is now. Those who use mobile devices have full access to the data and must find a method to properly use the applications on their phones regardless of the time or even storage constraints, as an instance. Despite this, mobile devices still have, storage, computational battery, bandwidth, as well as power limitations, even if the rise is strong. They also need to be less connected as well as less secure, at least compared to whatever we are used to with fixed equipment. This is the basic need. Such constraints may be overcome with the use of the current LTE mobile network standards as well as cloud augmentation that is presently used in

* ISBN No. 978-81-955340-6-7

newer mobile apps. In order to function on a mobile phone, such apps often incorporate activities such as computer vision, image analysis, facial identification, an optical character recognition, as well as augmented reality.

These three technologies work together to deliver high-quality commutating resources that may be used by both mobile users as well as cloud computing providers. When mobile devices as well as the cloud computing are combined, MCC creates a whole new infrastructure. It refers to a system beyond mobile device wherein data is stored and processed. When it comes to computing-intensive operations and storing massive volumes of data, the cloud takes the lead. It's clear that the quick rise of the (MC) mobile computing [4] is a significant breakthrough in the field of technology. Mobile computing devices, on other hand, are plagued by several resource as well as communication issues (– for example, battery life, storage capacity, as well as bandwidth) (for example, mobility as well as security) [5].

The followed criteria may be used to categorise the many uses of the mobile cloud services:

- **Sensing capability**: It is possible to use a smartphone as a sensor. Blood pressure, Humidity, as well as temperature are just a few of the numerous variables that sensors may detect. Sensor data may be uploaded to the cloud at such a later time. Users from all around the world may get their hands on this data thanks to the cloud.
- **Maintaining privacy and security of user data:** When using a mobile cloud, even a user may determine what information they want to keep private as well as what the information they want to make it public.
- **Data storage and reliability:** If such storage device fails, mobile cloud will back up such data as well as secure users' information.
- **Security of personal information:** Virtual computers in cloud provide better protection for the personal information thanks to the secured search engine.
- **Health monitoring:** It is possible to store and send data about one's health to cloud utilizing mobile devices equipped with sensors. Mostly in cloud, health centres may provide customers advice on how to keep their health in check. For the health monitoring, mobile devices may be utilised as sensors, too.
- **Sensing as a service:** For example, the mobile cloud offers a service that delivers platform, infrastructure as well as software as a subscription. As a result, a user may effortlessly switch between several programmes without having to worry about their compatibility.

Mobile cloud computing has indeed become a prominent study subject in scientific as well as industrial sectors because of the key application models inside the Internet age. It's growing more and more popular with each passing day. There are several mobile cloud computing apps available to consumers, including Google's Maps, Gmail, including Navigation systems for the Mobile, Voice Search

as well as other services. It is clear from a Juniper Research report that perhaps market for the cloud-based mobile apps is $9.5 billion for both consumers and businesses [6]. It's really the primary goal of cloud computing to offer a wide range of services as well as software through the Internet, expanding storage capacity, decreasing costs, automating systems, and divorcing service delivery again from such underlying technology. Fig . 1 illustrates mobile cloud computing infrastructure.

1.1. Benefits of Mobile Cloud Computing

1. Businesses save money by using mobile cloud computing.
2. Since of their mobility, they are able to carry out their duties more quickly and effectively.
3. Cloud customers use their mobile devices to investigate new capabilities that aren't available on their computers.
4. Using mobile cloud computing web services, developers may reach a wider audience.
5. It's possible that other network service providers may enter this market.

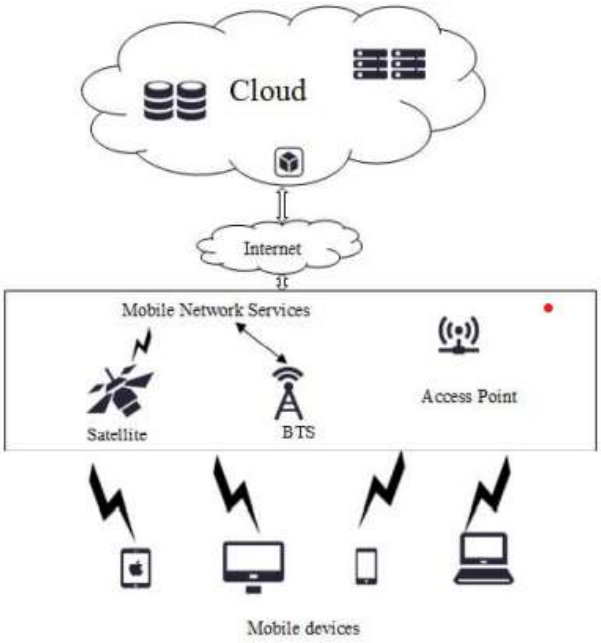

Figure 1: Architecture of Mobile Cloud Computing

- **Mobile Network** Cellular Service Providers In order to connect to mobile networks, mobile devices (such as smartphones, iPads, laptops, tablets, and etc) need satellites, BTSs whose full form is Base Transceiver Stations, or even access points. The operational interface in between mobile device as well as network operator is controlled by mobile network.

- **Internet Service** Internet Service Internet links mobile network to cloud through the Internet. Mobile users may utilise internet to connect to cloud and get resources that they require. Cloud services may be accessed wirelessly or through a 3G or even 4G telecommunications network.
- **Cloud service** Service in the cloud Service-oriented cloud architecture has three layers: SaaS that is Software as a Service, PaaS that is Platform as a Service, as well as IaaS Infrastructure as a Service (Infrastructure as a Service). Virtualization systems like VMs are offered by IaaS. (Virtual Machines). Such infrastructure services may be tailored to meet the needs of the user at any given time. To create, test, as well as deploy applications, or even PaaS offers a platform. SaaS provides access to applications as well as a database.

2. MOBILE CLOUD COMPUTING CHARACTERISTIC

Since mobile as well as cloud computing are combined in this form, mobile cloud computing exhibits both the advantages as well as disadvantages of each.

- **Entertainment**: Smartphones, which can run a wide range of apps, now account for the majority of such mobile phone market. There are a variety of ways that smartphone users may have fun while they're doing other things.
- **Communication**: The primary purpose of mobile devices like cellphones is to well communicate with everyone. Users from across the globe may connect with one other using social software in addition to phone calls.
- **Movability**: This is a fundamental feature of the mobile computing. Just on bus or even at KFC, users may get work done or even get the information they need.
- **On-demand self-service** Users may request computational resources (server time as well as network storage) as well as have them delivered to them on demand. It also makes it simpler to respond to the changes in demand within real time.
- **Resource pooling:** Cloud service providers have a vast pool of computer power at their disposal. Infrastructure as well as network facilities are included in this resource pool, which gives computing power that has never before been seen.
- **Rapid elasticity:** Some services may be automatically deployed and withdrawn based on demand from users. As a result of this feature, cloud customers may design as well as deploy services fast.
- **Cloud computing**: Using an internet connection,even users may access cloud resources as well as services.

- **Virtualization:** One of the fundamental technologies involves virtualization. In an attempt to dynamically supply resources and save costs, cloud providers may use virtualization to combine resources and virtualize them.
- **Broad network access:** A wide range of client platforms, including tablets, mobile phones, workstations as well as so on, may be used to access the cloud through the Internet. Users and suppliers alike are able to the monitor and regulate the number of resources that are being used.

I. SECURING INFORMATION ON THE CLOUD

Individuals and businesses can use cloud to store large amounts of data or even applications. Moreover, the authentication, integrity and digital rights of such data or even application must be protected while processing.

- **Authentication:** A slew of different methods of authentication have been floated in an effort to keep mobile users' access to the cloud computing data safe. A few organisations adhere to open standards as well as encourage use of a variety of different forms of authentication. Such as showing how to use log-in IDs, PIN codes and passwords to gain access.
- **Integrity:** Data stored on the cloud network should be protected by all mobile cloud users. Authentication and verification are required for all access. Distinct approaches are indeed being presented to ensure the integrity of one's cloud-based information. Using an instance, evey cloud user's information is labelled or initialised to them, and they are only ones who can move or even delete information from the cloud.
- **Digital Rights Management**: There has been a steady rise in the number of people who illegally download different digital material such as images, music, and video and e-book applications. The above digital contents have been protected from an illegal access by numerous methods, including provision of an encryption as well as decryption keys. There should be some sort of code or even decoding platform in order to access like digital content over mobile devices.

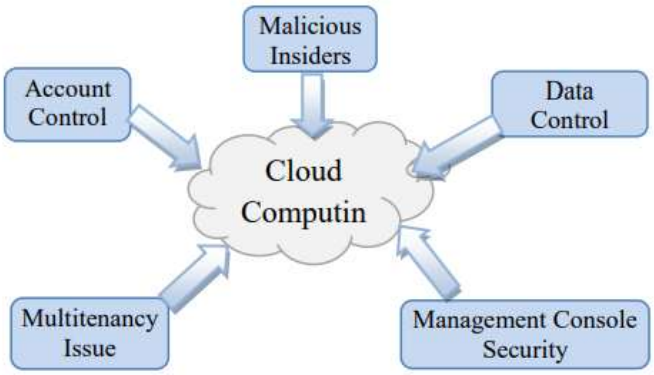

Figure 2: Categorization of threats in MCC

3. MOBILE CLOUD COMPUTING - CURRENT TRENDS

3G is perhaps the most popular mobile technology in use today, and 3G-enabled mobile computers are being developed.

- **3G**: When the (ITU) International Telecommunications Union released the (IMT-2000) that is called International Mobile Telecommunications-2000 standards in the year 2000, 3G or 3rd generation mobile was introduced. Wide area wireless phone, even video calls, including mobile Internet as well as TV in the mobile environment are only some of the features of the action services.
- **Global Positioning System (GPS):** A space-based system of navigation, (GPS) the Global Locate System, may be used in any weather conditions, anywhere near the Earth wherein there is a clear line of sight to at least four GPS planetoids. Additionally, civil as well as commercial buyers throughout the globe now have more demanding options thanks to the GPS initiative. Improved worldwide air traffic control, resistance, including location services rely on GPS.
- **Long Term Evolution (LTE):** For the mobile phones as well as data depots, LTE is indeed a high-speed data transmission protocol. As a result of innovative tone approaches, it has capacity as well as speed of GSM or EDGE as well as UMTS/HSPA networks but is more efficient. It has to do with the usage of 4th Generation (4G) mobile technology.
- **5G** For the first time, 5G cellular communications technology goes above and beyond previous generations of the mobile technologies. To put it another way, 5G isn't merely next generation of the mobile communications. Instead, 5G technology is quite diverse. In past, systems had been shaped more through the possibilities offered by newer technologies. New 5G technologies has indeed been pushed by specific applications. A wide range of purposes, including audio communications, remote control, large video downloads, including low data rate capabilities such as remote sensors as well as Internet of the Things, have propelled the development of 5G.
- **WiMAX:** To increase data speeds from 30 to 40 megabits per second, WiMAX that is the "Worldwide Interoperability for Microwave Access " is indeed a wireless communication test method that can now provide data rates of up to 1 gigabit per second across a fixed location. A component of 4G (4th Generation) wireless communication, or 4th generation. With such a signal radius of around 50 kilometres, WiMAX offers the metropolitan area network that outperforms the traditional Wi-Fi LAN's 30 metre wireless range. Cable-modem as well as DSL connections may be replaced by WiMAX because of its ability to transmit data quickly, however the capacity should be shared among several users, which results in slower speeds in most cases. The "final mile" is where WiMAX goods and services are also most probable to occur by consumers. For ISPs but also carriers, WiMAX means not having to utilise physical cabling (copper, even cable, and so on.) to link clients' premises towards the Internet.

4. MOBILE CLOUD COMPUTING LIMITATION

By using mobile cloud computing, subscribers will be able to quickly and easily access their cloud-based files, applications, and other data. Because mobile devices as well as wireless networks are so unique, they provide the most significant obstacle to cloud computing on the go. All of these difficulties make using the app more difficult than it would be on a stationary cloud device. When it comes to the cloud computing, overall quality of the wireless connectivity, as well as the cloud computing's support for mobile devices, are all key considerations to keep in mind. Listed below are the most common issues and the solutions for the mobile cloud computing:

Security and Privacy in the Cloud: Increasingly, security as well as privacy have taken centre stage in the mobile cloud computing. The organization's private data as well as secret information will inevitably be exposed when it establishes a cloud-based architecture. That cloud service provider will then be in charge of managing, safeguarding, and storing them. Before making a choice, all available options should be considered. As a result, consumers may be reluctant to hand up their personal information to the third party.

Low Bandwidth: In the mobile cloud context, bandwidth is indeed a critical problem since mobile network resources are substantially fewer than those of conventional networks. As a result, P2P Media Streaming may be used to distribute tiny amounts of bandwidth among subscribers in the very same region that are watching the same movie. Each user may transmit or exchange portions of same material with a second user that use this process, leading to an enhancement inside quality of the sent information, particularly for the video transmission.

Prone to Attack: Information stored in cloud seems to be more exposed to external hack assaults as well as threats. There is no such thing as total security just on internet. Just on internet, hackers including malevolent users are continuously on the lookout for the opportunity to steal sensitive data as well as information.

Limited Control and Flexibility: Because all apps including services are hosted in distant or even the third-party virtual environments, consumers have minimal control over how hardware as well as software work together. Because distant software is now being utilised for the mobile cloud computing, it often lacks the functionalities of a local programme.

Dependency and Vendor Lock: In Inherent in the mobile cloud computing's reliance on an ISP is a significant drawback. If a customer wishes to change providers, the process of transferring significant amounts of data through one provider to another may be very time consuming and frustrating. Choosing a vendor is a really important decision, which is just one of the many reasons why.

Increased Vulnerability: Cloud-based solutions that deal with privacy as well as security are a more attractive target for the hackers as well as other bad actors since they are all available to public internet. Several of the greatest names on the internet are the victims of significant attacks and security lapses. Nothing about the internet can be guaranteed to be completely safe.

The Mobile cloud computing problems may be reduced using the following techniques: Utilizing regional data centres, wireless networks may benefit from increased bandwidth as well as more mobile-friendly content over the web.

- Deploy an application processing node just at "edge" of the cloud computing to speed up data transmission.
- (DIC) that is Data-intensive computing and energy-intensive computing, including virus detection on the mobile devices, may be duplicated to the cloud utilising virtualization including imaging technologies.
- Dynamically improve cloud application push as well as division through the mobile terminals.

5. CONCLUSION

It is the primary goal of mobile cloud computing for empowering mobile user, independent of mobile device's resource limits, by offering a smooth and the rich functionality. Even while mobile cloud computing is indeed still in its infancy, it has the potential to become the dominant paradigm for mobile applications in the future. An introduction to the mobile computing, followed by a discussion of its evolution and the direction the technology may go in the future, all while navigating the many classes as well as security concerns that may arise. Mobile computing has made it possible to capture video including audio while on the move. Movies, academics, and talkative content may all be followed with ease. One may acquire all the fun they desire when browsing the internet for the flood data thanks to the advancement and display of the high-speed data ally at such an extravagant cost. The internet provides access to a wide range of entertainment options, including news, movies, and films. This wasn't created prior to the rise of mobile computing.

An overview of the MCC's architecture as well as features is presented in this publication. MCC is sometimes characterised in literature as a huge difficulty. Open research topics in this domain and associated techniques are presented in just this taxonomy, concentrating on the low bandwidth, the computational offloading and the heterogeneity and etc. We'll look at ways to enhance resource allocation in the MCC environments, like how to distribute as well as offload tasks based on quality of service (QoS) profiles as well as cost functions.

REFERENCES

1. Abolfazli, S., Sanaei, Z., Ahmed, E., Gani, A., &Buyya, R. 2014. Cloud-based augmentation for mobile devices: motivation, taxonomies, and open challenges. Communications Surveys & Tutorials, IEEE, 16(1), 337-368.

2. Liu, F., Shu, P., Jin, H., Ding, L., Yu, J., Niu, D., & Li, B. 2013. Gearing resource-poor mobile devices with powerful clouds: architectures, challenges, and applications. Wireless Communications, IEEE, 20(3), 14-22.
3. Sarah Perez, August 4, 2009, Why cloud computing is the future of mobile, http://www.readwriteweb.com/archives/why_cloud_com puting_is_the_future_of_mobile.php, Retrieved on February 2015.
4. Satyanarayanan M. 2010. Proceedings of the 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond (MCS).
5. Tavel, P. 2007. Modeling and Simulation Design. AK Peters Ltd.
6. S. Perez, Mobile cloud computing: $9.5 billion by 2014, http://exoplanet.eu/ catalog.php, 2010.
7. G. Deepak and B. S. Pradeep. "Challenging Issues and Limitations of Mobile Computing." International Journal of Computer Technology & Applications, vol. 03, Feb. 2012.
8. P. Yadav and N. Chaudhary. "Mobile Computing the Future of Digital Era." International Journal of Computer Applications, pp. 23-27, 2015.
9. P. Mell and T. Grance. "The NIST definition of cloud computing." , Sep. 2011.
10. Bahar, A. N., Habib, M. A., & Islam, M. M. 2013. Security architecture for mobile cloud computing. International Journal, 3(3), 2305-1493.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

A Review on Big Data: Privacy and Security Challenges

Dr. Rohit Kumar[1*]

[1]*Assistant Professor, IT, Institute of Management Studies Noida*

Abstract

A significant rise in data is being generated due to factors such as the fast expansion and dissemination of the network services, the mobile devices, but also internet users. Almost every business is attempting to deal with the massive amounts of data that are being generated. The concept of big data has starting to gain traction. Traditional apps can't handle huge data because it's hard to store as well as analyse, so it poses serious privacy as well as security concerns. As a result, this article explores the scope of big data and reviews the most recent studies on security as well as privacy issues surrounding it. We'll talk about the problems and the things that impact security. Privacy-preserving methods are also examined and expanded on in this paper.

Keywords: big data; Hadoop security; cloud security; monitoring; auditing; key management; anonymization.

1. INTRODUCTION

The term "big data" is a relatively new one in the field of information technology, and it refers to vast amounts of data which are now possible for collecting, storing, including managing, as well as analysing [1].

It is important to note that in [2], the term "Big Data" refers to data with a high volume, speed, and variety that must be processed in new ways to improve decision making as well as the process optimization. Data sets can be referred to as "Big Data" if their capture, analysis, storage, filtering and visualisation is beyond the capabilities of current or even such traditional technologies. As it's discussed

* ISBN No. 978-81-955340-6-7

through The Economist in [4], "Managed well, the data can be used to unlock new sources if economic value, provide fresh insights into science and hold governments to accounts".

With the use of security solutions such as network monitoring, event management, as well as security information [5], the big data helps in preserving security concerns. The usage of cryptographic techniques, data authenticity, security of the stored data, access management, and monitoring of the real-time data are just a few of challenges that big data faces when it comes to security [6]. Big data can only be used effectively if privacy as well as security concerns are addressed. Privacy, integrity, as well as availability are such three of the most important security concepts. When it comes to protecting user information, security may be described as the ability to monitor and protect user-specific access information against unauthorised disclosure, change or destruction [6]. Controls based on the operational and technological elements may provide security. In other words, the right to the privacy is the ability of the individual to limit the disclosure of personal information about themselves. Policies and procedures may be used to ensure privacy.

2. FEATURES OF BIG DATA

Volumes, velocities, varieties, truthfulness, and value are the five vs of Big Data characterization.

1) **Volume (data in rest):**There are two main characteristics of big data: volume (information in the rest) and scalability.
2) **Variety (data in many forms):**There are three sorts of data in the world: internal, an external, and a combination of the two
    - Clearly defined (information ffrom the relational databases)
    - Structured in a semi-structured manner (website logs, the social media feeds, the sensor data, email and so on)
    - A lack of structure (videos, images, audios and so on.)
3) **Velocity (data in motion):** "The production rate of data is notably high. The increase in data means that data should be analyzed more swiftly" [7]. Indeed it is not just the velocity of incoming data that is the issue it is possible to stream fast moving data into bulk storage for later batch processing.
4) **Veracity:**An issue of veracity is degree to which data are uncertain or inaccurate.
5) **Value:**It's really value of knowledge that can be gleaned from it.

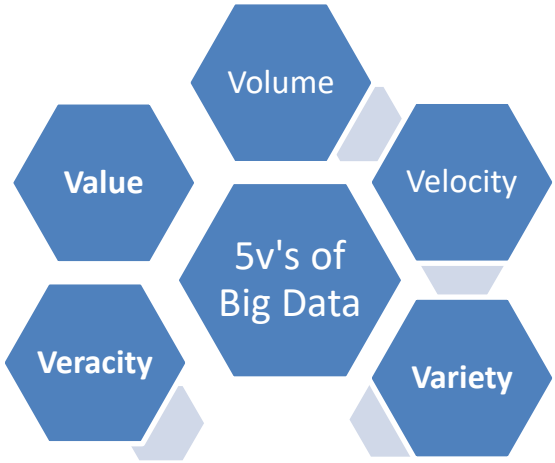

Figure 1: V's of big data

3. CHALLENGES OF BIG DATA

New value intuitions are expected to be gleaned from accessible data sources when Big Data is used. These include issues like the heterogeneity, the data life cycle management as well as processing, the data scalability as well as the security including privacy as well as data visualisation:

A. Heterogeneity

There are essentially endless dispersed data sources, which contributes to Big Data's variety [8]. Images, videos, audios and sensors are all examples of data diversity [9]. Because of such difficulties in converting structured as well as semi-structured information into homogeneous data, handling heterogeneous data types as well as sources is a critical future problem [10]. Metadata is also part of Big Data. The smartphone's metadata can tell you where a picture was taken, when it was shot, and even what kind of camera it used. The problem of dealing with a variety of metadata types is likewise a difficult one.[11].

B. Data Processing

Pre-processing may be necessary before the actual examination of data obtained from multiple sources. Data reutilization, data reorganisation, and even data exhaust [12] may be used to process redundant data for future study. It is possible that inaccurate data was created during the data mining in an attempt to improve mining results.

C. Data Life Cycle Management

There is no end to the data life cycle in the Big Data. Dataset values might well be interpreted in a variety of ways by various users. When a patient's recovery is complete, the patient's health record is no longer useful from the patient's viewpoint, but it might be useful from the doctor's or even researcher's perspective [9, 12]. Because of this, it is necessary to revisit how data lifecycles are judged and defined.

D. Security and Privacy

Conventional protection methods won't work for Big Data due of its unique properties. Because of this, the widespread usage of the Big Data in everyday operations raises security concerns. Outsourcing sensitive information also raises concerns about data security. Data capture or even data storage of the personal sensitive data must be done in a manner that protects the privacy of such data. [13], [8-9], [12].

E. Scalability

Volume has a direct bearing on scalability. Storage scalability refers to the system's capacity to accommodate growing data volumes in an appropriate way [8]. As the phrase "Big Data" implies, it necessitates an ever-increasing number of data to be processed.

4. BIG DATA SECURITY AND PRIVACY APPROACHES

Security as well as privacy are not adequately protected when working with large amounts of data. When it comes to encrypting data or accessing it, access restrictions, firewalls, the transport layer security, and even the anonymized data may all be compromised [14]. This is why improved methodologies and technologies are being created to safeguard, monitor as well as audit big data activities from an infrastructures, applications as well as data perspective. Based on previous research on these topics, this article has grouped security and privacy concerns for big data into the following five titles: cloud security, Hadoop security; monitoring as well as auditing; key management as well as anonymization (Figure. 2). To summarise the research, we created Table I, which lists the studies' aims, methods, and results.

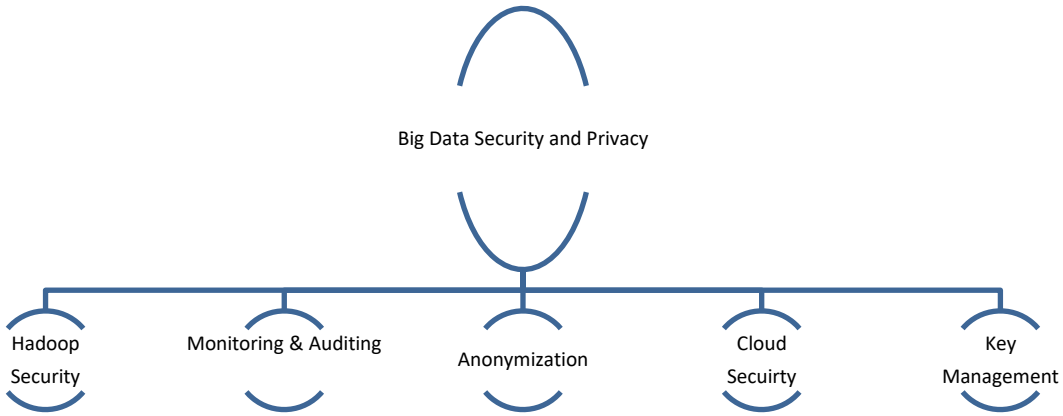

Figure 2: Big data security and privacy categorization

a) Hadoop Security

Hadoop wasn't really designed with security in mind when it was first built as the distributed process platform. It's designed to work in safe surroundings. The development of security measures for Hadoop has accelerated in response to its growing popularity. In addition, academics are starting to pay attention to it. To prevent hackers from accessing all of such data stored within Hadoop system, two solutions were presented [15]. The name node, that is part of HDFS and controls data nodes, has created the trust mechanism between the user and the node. For accessing a name node, the user should authenticate using this technique. Firstly, the user provides the hash function, as well as then the name node generates a hash function of its own including compares them. Access to the system is granted if the comparison results are valid. One of the hashing algorithms is employed in this phase to authenticate the data. In order to prevent hackers from accessing the whole database, random encryption algorithms including , AES, RSA, Rijndael and RC6 have been utilised. This technique uses MapReduce to do the encryption or decryption operation. Ultimately, twitter stream is used to test such two methods and show how to manage security concerns.

b) Monitoring and Auditing

Gathering and studying network events in order to detect breaches is the goal of network security monitoring. It is the systematic and quantifiable security strategy to apply several approaches for security audits. Active security relies heavily on both of these factors.. On the overall, network traffic detection as well as prevention techniques are quite tough. There is a solution to this issue by analysing the DNS traffic, the IP flow records, the HTTP traffic, and the honeypot data [16]. Data correlation techniques are used to store and analyse data from several sources in a dispersed manner. In order to determine if the domain name, packet, or even flow is malicious, 3- probability metrics have indeed been computed. An alarm is triggered in the detection system or the procedure is terminated mostly by the prevention system based on just such score. Using open-source big data platforms, Shark and Spark outperform

Hadoop, Hive, as well as Pig in terms of speed and consistency when it comes to analysing electronic payment data from a corporation.

c) Anonymization

Data mining for analytics raises serious issues about privacy. When it comes to protecting personal information (PII), it's becoming tougher. Agreement between the firm and person must be based on rules to eliminate privacy problems. De-identification and anonymization of personal information are critical steps in the data transmission process [17]. However, the company's algorithms and artificial intelligence analyses may reveal the individuals identify. This analysis's conclusions have the potential to lead to unethical situations.

d) Cloud Security

Due to factors like on-demand service,the pooling of resources, and flexibility, cloud computing has become a suitable setting for big data [6]. It is now common practise to employ the cloud computing. The cloud, on the other hand, is vulnerable to both old and new dangers. In today's world, the cloud data storage is indeed a major issue. As a result, service provider should take certain measures to protect its customers. As a result, a safe method for managing and sharing large amounts of data on the cloud has indeed been developed [18]. Authentication, Security measures such as decryption, as well as compression are included in order to protect large amounts of data. For the authorised user, email as well as password authentication was utilised. In order to ensure the safety of the data, it has indeed been compressed as well as encrypted. In the event of a natural catastrophe, the company employs 3 backup servers. In such servers, encrypted data has indeed been saved. The secret key was used to decode encrypted data in event of a server failure.

e) Key Management

Another important data security concern is the generation and distribution of keys between servers as well as users. Fast as well as dynamic authentication procedures, on the other hand, may be proposed for large data centres. The PairHand protocol was developed for authentication in the mobile or stationary data centres in [19] using a tiered approach to the quantum cryptography, data-reading, the front-end, , quantum-key-processing-management, as well as application layers all fall under this umbrella term. Both key search operations as well as passive assaults have been decreased by this approach.

The big data services are made up of a number of different groups, each of which requires a safe way to exchange group keys. As a result, a new protocol-based o the Diffie-Hellman key agreement as well as a linear secret sharing method has indeed been proposed instead of the current protocols [20]. In order to protect the system, protocol ensures freshness of key, authentication of key, and secrecy of key.

CONCLUSION

New signs of scientific revolt are on the horizon as we move into an age of big data that represents next frontier for the revolution, competition, as well as efficiency. There are many issues with Big Data, and we've covered them all in such article, from heterogeneity to management of the data life cycle to data processing to privacy as well as security. Security including privacy problems like data privacy as well as key management have also been thoroughly examined. We have examined them in-depth. In order to address the above-mentioned security including privacy concerns, a number of commendable efforts have been made. Future discussions on big data security, privacy, as well as safety will need new methodologies as well as technologies for the human-computer interactions or even the improvement of current ones for the more accurate outcomes. Research in these areas will, however, need careful attention and effort on the part of academics. We believe that our extensive poll will assist to build better security as well as privacy solutions for Big Data, which is in its infancy.

REFRENCES

1. Johns Hopkins, “Big data custodianship in global society”, SAIS Review of international Affairs, Volume 34, Number 1, Winter-Spring 2014, pp. 109-116 (Article).
2. C. Philip Chen and C. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", Information Sciences, vol. 275, 2014, pp. 314-347.
3. D. Terzi, R. Terzi and S. Sagiroglu, "A survey on security and privacy issues in big data", 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST), 2015, pp. 202-207.
4. E. Bertino, "Big Data - Security and Privacy", 2015 IEEE International Congress on Big Data, 2015, pp. 757-761.
5. A. A. Cardenas, P. K. Manadhata, and S. P. Rajan, “Big data analytics for security,”IEEE Security & Privacy, vol. 11, no. 6, pp. 74–76, 2013.
6. D. S. Terzi, R. Terzi, and S. Sagiroglu, “A survey on security and privacy issues in big data,” in2015 10th International Conference for InternetTechnology and Secured Transactions (ICITST). IEEE, 2015, pp. 202–207.
7. Xinhua Dong, Ruixuan Li_, Heng He, Wanwan Zhou, Zhengyuan Xue, and Hao Wu, “Secure Sensitive Data Sharing on a Big Data Platform”, Tsinghua Science and Technology, ISSN l l1007-0214l l08/11l lpp72-80, Volume 20, Number 1, February 2015

8. I. Hashem, I. Yaqoob, N. Anuar, S. Mokhtar, A. Gani and S. Ullah Khan, "The rise of "big data" on cloud computing: Review and open research issues", Information Systems, vol. 47, 2015, pp. 98-115.
9. Weichang Kong, Qidi Wu, Li Li and Fei Qiao, "Intelligent Data Analysis and its challenges in big data environment", 2014 IEEE International Conference on System Science and Engineering (ICSSE), 2014, pp. 108-113.
10. [Online] Challenges and Opportunities with Big Data", Purdue Univesity, 2016. [https://www.purdue.edu/discoverypark/cyber/assets/pdfs/BigDataWhite Paper.pdf. [Accessed: 12- Jan- 2017].
11. Z. Azmi, "Opportunities and Security Challenges of Big Data", Current and Emerging Trends in Cyber Operations, 2015, pp. 181-197.
12. M. Chen, S. Mao and Y. Liu, "Big Data: A Survey", Mobile Networks and Applications, vol. 19, no. 2, 2014, pp. 171-209.
13. E. Bertino, "Big Data - Security and Privacy", 2015 IEEE International Congress on Big Data, 2015, pp. 757-761.
14. B. Matturdi, X. Zhou, S. Li, F. Lin, "Big Data security and privacy: A review", Big Data, Cloud & Mobile Computing, China Communications vol.11, issue: 14, pp. 135 – 145, 2014.
15. P. Adluru, S.S. Datla, Z. Xiaowen, "Hadoop eco system for big data security and privacy", Systems, Applications and Technology Conference (LISAT), Long Island, Farmingdale, NY, pp. 1 – 6, 2015.
16. S. Marchal, J. Xiuyan, R. State, T. Engel, "A Big Data Architecture for Large Scale Security Monitoring", Big Data (BigData Congress), pp. 56 – 63, Anchorage, AK, 2014.
17. T. Omer, P. Jules, "Big Data for All: Privacy and User Control in the Age of Analytics", Northwestern Journal of Technology and Intellectual Property, article 1, vol. 11, issue 5, 2013.
18. A. Kumar, L. HoonJae, R.P. Singh, "Efficient and secure Cloud storage for handling big data", Information Science and Service Science and Data Mining (ISSDM), pp. 162 – 166, Taipei, 2012.
19.] T. Vijey, A. Aiiad, "Big Data Security Issues Based on Quantum Cryptography and Privacy with Authentication for Mobile Data Center", Procedia Computer Science, vol. 50, pp. 149–156, 2015.
20. H. Chingfang, Z. Bing, Z. Maoyuan, "A novel group key transfer for big data security", Applied Mathematics and Computation, vol. 249, pp. 436–443, 2014.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

A Survey of Intrusion Detection Systems for Cloud Computing

Ms. Savita Singh[1*]

[1]*Assistant Professor, IT Department, Institute of Management Studies, Noida*

Abstract

End customers benefit from scalable, virtualized, on-demand services delivered through the cloud, all while spending less money on infrastructure. Internet-based delivery of such services is made possible by the use of well-established networking protocols as well as formats, which are overseen by a variety of organisations. Deficiencies and flaws in the underlying technology and old protocols may lead to intrusions. Cloud resources as well as services are protected from a variety of threats and assaults by (IDS) that is Intrusion Detection System, which is the most often utilised component of the computer security as well as compliance. Cloud incursions, IDS detection methods, including IDS based over the Cloud Computing are all discussed in this study.

Keywords: Cloud computing, Firewalls, Intrusion detection system, Intrusion prevention system.

1. INTRODUCTION

In today's IT environment, (CC) that is the cloud computingis a rapidly expanding computational paradigm. Internet-based computer resources (example- network, server, storage, apps, and etc.) are made available as the "service" via internet again for benefit of consumers [1]. Platform (the virtualized operating system for the server) as well as application (including web applications) layers are the three primary abstraction levels of this system [2]. CC has a number of distinct features:

- **Virtual**: It's easy for users to know where they are and what's going on below.

* ISBN No. 978-81-955340-6-7

- **Scalable**: Having the ability to decompose large and difficult jobs into smaller, more manageable chunks.
- **Efficient**: Dynamic provisioning of the shared computing resources using the Services Oriented Architectures.
- **Flexible**: It can handle a wide range of workloads, from consumer to business. Platform as a Service also represented as (PaaS), Infrastructure as a Service also represented as (IaaS), and SaaS models are all examples of cloud computing's 3-service models. Using infrastructure as a service approach, consumers may have access to a variety of resources, including hosting servers as well as networks. In tools, PaaS, languages, including APIs are provided for the creation, deployment, including operation of cloud-based applications, whereas in SaaS, even systems provide fully-functional online applications which customers may operate without the need for additional hardware or software installation.

There are two types of NICs in network intrusion detection system also represented as (NIDS) : one for licentious mode and one for management mode. When IDS is installed in network or even at the network's edge, it keeps tabs on all of the traffic that flows through it. (HIDS) that is Intrusion detection systems and software-based programmes are installed on the host to be watched by shareholders. Whenever an assault is detected, the agents would monitor an operating system as well as write data into the log Rles. The Host Intrusion Detection Systems also represented as (HIDS) are able to monitor installed agents on every unique host. It is possible to keep track of any attempted intrusions on crucial servers using Host-based IDS systems.

1.1. INTRUSIONS TO CLOUD SYSTEMS

This section demonstrates a number of common intrusions, that can cause Cloud resources as well as services to be unavailable, compromised, or otherwise compromised.

A. **Insider attack:**Unauthorized rights may be gained by authorised Cloud users. There is a risk that insiders may perpetrate frauds and leak confidential information (or even destroy information with an intent). This raises severe concerns about a person's ability to put their faith in the organisation. As an illustration, an internal denial-of-service attack against an Amazon Elastic Compute Cloud that is represented as (EC2) was presented [4].

B. **User to Root attacks:**their account. In order to get root-level access to the system, someone may take advantage of these vulnerabilities. This may be accomplished by exploiting buffer overflows in a process that is already executing as root. When the application programme code exceeds the static buffer, it results in an error. Weak password recovery procedures, phishing assaults, keyloggers, as well as other forms of security risk can't be prevented by using security methods that are universally accepted as standard practise. In the Cloud, even an attacker may get root access to the VMs or hosts by acquiring access to a legitimate user's instances.

C. **Flooding attack:**In this scenario, the attacker aims to overwhelm the victim through the sending a large number of packets from innocent host (zombies) in network. Packets can be of type TCP, UDP, ICMP or a mix of them. This kind of attack may be possible due to illegitimate network connections. In case of Cloud, the requests for VMs are accessible by anyone through Internet, which may cause DoS (or DDoS) attack via zombies. Flooding attack affects the service's availability to authorized user. By attacking a single server providing a certain service, attacker can cause a loss of availability on the intended service. Such an attack is called direct DoS attack. If the server's hardware resources are completely exhausted by processing the flood requests, the other service instances on the same hardware machine are no longer able to perform their intended tasks. Such type of distributed attack is called indirect attack. Flooding attack may raise the usage bills drastically as the Cloud would not be able to distinguish between the normal usage and fake usage.

D. **Attacks on Virtual Machine (VM) or hypervisor:** On compromise the lower layer hypervisor, attacker can gain control over installed VMs. E.g. BLUEPILL [5], SubVir [6] and DKSM [7] are some well-known attacks on virtual layer. Through these attacks, hackers can be able to compromise installed-hypervisor to gain control over the host. New vulnerabilities, such as zero-day vulnerability, are found in Virtual Machines (VMs) [8] that attract an attacker to gain access to hypervisor or other installed VMs. A zero-day vulnerability is a threat that tries to exploit application vulnerabilities that are unknown to others or the software developer. Zero-day exploits are used by attackers before the developer of the target software knows about the vulnerability. A zero-day vulnerability was exploited in the HyperVM virtualization application which resulted in destruction of many virtual server based websites [9]

E. **User to Root attacks :** Here, a criminal gains access to the user's account via posing as the user. sniffing password. This makes him able to exploit vulnerabilities for gaining root level access to system. For example, Buffer overflows are used to generate root shells from a process running as root. It occurs when application program code overfills static buffer. The mechanisms used to secure the authentication process are a frequent target since there are no universal standard security mechanisms that can be used to prevent security risks like weak password recovery workflows, phishing attacks, keyloggers etc. In the Cloud, an attacker may get root access to the VMs or hosts by acquiring access to the legitimate user's account.

F. **Port Scanning:** It gives a list of the ports that are open, closed, or even filtered. Attackers may locate open ports by scanning for them, and then launch attacks against the services that use those ports. This attack may provide information about the network's IP address, router, MAC address gateway filtering, the firewall rules, and more. TCP scanning,, the SYN scanning, the UDP scanning, even the FIN scanning, the Window scanning, the ACK scanning, (similar as ACK scanning but it examines any alterations in window field of packets) and so on. Are some of several port scanning methods available. Via port scanning, the attacker inside a Cloud environment may find open ports on that these services are offered and exploit them.

2. IDS ARCHITECTURES IN CLOUD

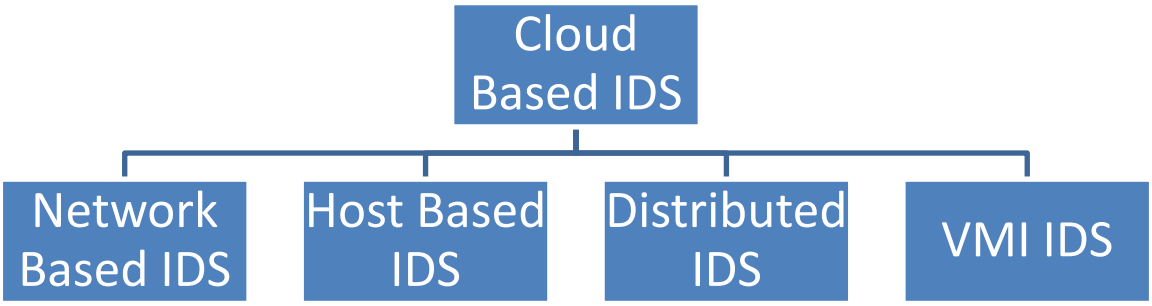

Figure: 1 Types of Cloud-Based IDS

2.1. Network intrusion detection systems

An example of a common network intrusion detection system (NIDS) for virtualized settings is presented by [10]. A virtual switch has an NIDS placed on it (through that the traffic of all the VMs together passes). In a typical computer context, an NIDS would be installed on the border server. There is a strong resemblance between this method and the standard NIDS. A virtualized environment has been designed to fit its needs and tested as a result of the work done by authors All the VM traffic is collected and logged by NIDS just on vSwitch. To detect DoS or even DDoS attacks, it employs SNORT [11] tools. The starting IP address is being used to analyse traffic. Unusual traffic coming from an IP address is immediately stopped, and the affected application is relocated to the different data centre. DDoS assaults and the botnets may be detected and blocked using this method. Only the known attacks may be detected by using SNORT, and also no performance data are provided in the publication [12]. This research does not address the issue of supporting large virtual networks with high traffic volumes [13]. Large virtual networks might provide problems for the detection of attacks by NIDS, since it may not be able to analyse all of the packets in the network at the same time. To recognize DoS attacks on the virtual SIP-based hosts, [14] having simulated IDS at several cloud locations. Detection is done using a signature-based approach. SNORT has indeed been chosen as that of network IDS for testing in Eucalyptus cloud computing environment. [15] Point out the major drawbacks of past strategies and provide a solution to remedy them. The added complexity that comes with checking all virtual machines for possible threats. In their study, they argue that the introduction of profile-based IDS may alleviate this issue. For the cloud, they propose an NIDS-based VMI that creates an individual profile for every virtual machine by comparing known attacks signatures as well as divergence from typical threshold values.

2.2. Host intrusion detection systems

There are three basic deployment-based categories for HIDS approaches with regard to the cloud in general. HIDS may be installed inside the host OS (in which it could monitor either host OS or even guest OS by communication by the VMM [16]) or even in the separate guest OS for the monitoring purposes. The poor attack resilience of an IDS that is totally controlled by the client is a downside of such first scenario. Because it's been widely panned in the literature, it's been ruled out for use in the cloud [17]. They [16] refer to this as a type I or type II situation where VMM seems to be the sole host process running as well as many VMs are operating over it. An alternative type II scenario assumes the virtual machine manager operates as software over host computer. In addition to host's normal processes, VM runs on host's virtual machine manager (VMM). [18] By using concepts of automated computing, we present an intrusion prevention system based on autonomous agents. The employment of the autonomous sensors to keep tabs on system activity including network traffic is indeed an anomaly-based detection strategy for spotting hostile activity. Detection of intrusions is based on abnormal levels of resource use by a user, according to [19]. (AAA) which is Authentication, Authorization, and Accounting is indeed the primary component of approach. The user's current use history is used to calculate the anomalous level. More guest OS may be introduced with no worry about the detection speed thanks to IDS of the medium as well as low-level security using less resources. The administrator has access to log files and may do audits on them.

2.3. Distributed intrusion detection systems

[20] has developed as well as simulated the intrusion detection system to fight DoS as well as DDoS assaults in the cloud. The IDS consists of four parts, each of which serves a distinct purpose. This prevents the single point of the failure inside the system. As a result, it does not identify unknown assaults since it relies on signature-based detections. The authors [21] propose a three-dimensional IDS. It is a cloud-based IDS for the IaaS users. A server as well as several agents make up 3-D IDS. The architecture is offered as a theoretical concept with no accompanying experimental proof. It also needs that the server be installed at the user end that is not always the case for all users. The Multithreaded network intrusion detection systems (MITIDS) are being proposed by [22] to combat Cross Site Scripting (XXS) as well as DDoS assaults. A capture module, an analysis as well as processing module, and just a report generation module make up the approach. In theory, it's a new technique, but no proof has been presented by the researcher. Another intrusion detection system relying on VMs is Siren [23]. Sniffer detects malicious software operating in the VM that tries to communicate with the network it is connected to. Virtual machines (VMs) should not produce traffic on network in absence of the human intervention, according to Siren's design. Siren labels traffic as dangerous if it detects such a situation. Siren's capability to insert designed human input to explore for ad-on malware is among the strongest capabilities. But producing traffic which closely mimics human input is the main issue with this method.

2.4. Virtual Machine Introspection (VMI) based techniques

Out-of-the-box intrusion detection is based on the virtual machine introspection also represented as (VMI). In VMI, inspection module is moved outside of virtual machine. Detection of any intrusions in guest system's software is carried out outside. One benefit of this method would be such malware detection is undisturbed even if an incursion occurs. HIDS as well as NIDS do not have this feature. Confidence is lost when HIDS reports inaccurately when NIDS has restricted sight. VMI-based intrusion detection systems have recently been proven in works by [24], [25], [26], as well as [27]. We'll go into these strategies in a minute. [28] Livewire, an intrusion detection prototype relying on VMI, is proposed. The VMM must be basic and well implemented for this method to work. VMM is protected by such a feature, making it harder for an attacker to get access. VMM's isolating, inspections, as well as interposition features are key to the technique's success. Livewire's OS library interface must be written in the safe programming language for multiple operating systems in order to avoid an attack upon itself. Revolutionary accomplishment in field of the virtual machine IDS, Livewire is regarded to be.

[28] Propose VMwatcher, that is a pre-configured intrusion detection system that is more accurate and resistant to tampering. The underlying VMM is assumed to be safe by VMwatcher. Type -II VMs may make use of this strategy. VMwatcher's best-known feature is its capability to close a semantic gap2 which is always present when obtaining an external perspective of guest OS. VMwatcher employs two distinct strategies: VM introspection that isn't obtrusive, as well as guest viewpoint casting. VMI is a kind of non-intrusive introspection. View casting is a technique used to create VM’s semantics view (— for example, file systems, processes, as well as directories). Memory states are captured in VM raw picture as well as a semantic representation is reconstructed. For both Windows as well as Linux, the prototype has been thoroughly tested. The semantic view is carefully crafted by VMwatcher to retain VM view. Furthermore, authors claim their VMwatcher has indeed the capacity to identify stealthy malware through comparing the internal and exterior views.

CONCLUSION

Intrusion detection would play an increasingly significant role in research landscape now that most of our current computer infrastructure migrates to cloud. Spending large sums of money on information security as well as privacy is a must to secure infrastructures throughout the globe. With the help of an IDS, a computer system may be kept safe and secure. IDSs for the cloud computing are in high demand since the number of people using the cloud increases at such a rapid rate. There are a number of cloud-based intrusion detection technologies now in use. In the cloud-based IDS world, there are four main subtypes to choose from: networks, hosts, distributions, and also introspection-based virtual machines (VMs). According to this poll, a number of potential threats to the security of Cloud services have been identified in future. Cloud security might not have been adequately addressed by one of the conventional options, such as a firewall.

REFERENCES

1. B. Zarpelao, "A survey of intrusion detection in internet of things;' Journal of Network and Computer Applications, 2017.
2. S. Raza, Wallgren , "Svelte: Real-time intrusion detection in the internet of things," Adhocnetworks, vol. 11, no. 8, 2013.
3. M. Nobakht, "A host-based intrusion detection and mitigation framework for smart home IoT using open flow," in Availability, Reliability and Security, 2016 11thInternational Conference on IEEE, 2016.
4. M. Slaviero, "BlackHat presentation demo vids: Amazon." [Online]. Available: http://www.sensepost.com/blog/3797.html
5. J. Rutkowska, "Subverting VistaTM Kernel for Fun and Profit," Black Hat Conference, 2006.
6. S. King, P. Chen, and Y-M. Wang, "SubVirt: Implementing malware with virtual machines," 2006 IEEE Symposium on Security and Privacy, 2006, pp.314-327.
7. S. Bahram, X. Jiang, Z. Wang, and M. Grace, "DKSM: Subverting Virtual Machine Introspection for Fun and Profit," Proceedings of the 29th IEEE International Symposium on Reliable Distributed Systems, 2010.
8. NIST: National vulnerability database. [Online]. Available: http://web.nvd.nist.gov/view/ vuln/detail?vulnId=CVE-2009-3733
9. D. Goodin, "Webhost Hack Wipes Out Data for 100,000 Sites." [Online]. Available: http://www.theregister.co.uk/2009/06/08/webhost_attack/.
10. A. Bakshi, and Y. B. Dujodwala, "Securing Cloud from DDOS Attacks Using Intrusion Detection System in Virtual Machine," Communication Software and Networks, 2010. ICCSN'10. Second International Conference on, pp. 260-264, 2010.
11. "Home - Snort.Org," https://www.snort.org/.
12. C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "A survey of intrusion detection techniques in Cloud," Journal of Network and Computer Applications, vol. 36, no. 1, pp. 42-57, 2013.
13. N. A. Premathilaka, A. C. Aponso, and N. Krishnarajah, "Review on state of art intrusion detection systems designed for the cloud computing paradigm," 2013 47th International Carnahan Conference on Security Technology (ICCST), pp. 1 - 6, 2013.
14. C. Mazzariello, R. Bifulco, and R. Canonico, "Integrating a network IDS into an open source Cloud Computing environment," 2010 Sixth International Conference on Information Assurance and Security, pp. 265 - 270, 2010.

15. S. Gupta, P. Kumar, and A. Abraham, "A Profile Based Network Intrusion Detection and Prevention System for Securing Cloud Environment," International Journal of Distributed Sensor Networks, vol. 2013, pp. 1-12, 2013.

16. M. Laureano, C. Maziero, and E. Jamhour, "Intrusion Detection in Virtual Machine Environments," Proceedings. 30th Euromicro Conference, 2004., pp. 520 - 525, 2004.

17. S. Alarifi, and S. Wolthusen, "Anomaly detection for ephemeral cloud IaaS virtual machines," Network and System Security, pp. 321-335, 2013.

18. A. Patel, Q. Qassim, Z. Shukor, J. Nogueira, J. Júnior, C. Wills, and P. Federal, "Autonomic agent-based self-managed intrusion detection and prevention system," In Proceedings of the South African Information Security Multi-Conference pp. 223-234, 2011.

19. J.-H. Lee, M.-W. Park, J.-H. Eom, and T.-M. Chung, "Multi-level intrusion detection system and log management in cloud computing," Advanced Communication Technology (ICACT), 2011 13th International Conference pp. 552-555, 2011.

20. C.-C. Lo, C.-C. Huang, and J. Ku, "A Cooperative Intrusion Detection System Framework for Cloud Computing Networks," 2010 39th International Conference on Parallel Processing Workshops, pp. 280-284, 2010.

21. J. He, C. Tang, Y. Yang, Y. Qiao, and C. Liu, "3D-IDS: IaaS User-oriented Intrusion Detection System," Information Science and Engineering (ISISE), 2012 International Symposium on, pp. 12-15, 2012.

22. M. P. K. Shelke, M. S. Sontakke, and A. D. Gawande., "Intrusion detection system for cloud computing," International Journal of Scientific & Technology Research, pp. 67-71, 2012.

23. X. Zhao, B. Kevin, and P. Atul, "Virtual Machine Security Systems," Advances in Computer Science and Engineering, pp. 339-365, 2009.

24. T. Garfinkel, and M. Rosenblum, "A Virtual Machine Introspection Based Architecture for Intrusion Detection," In NDSS vol. 3, pp. 191-206, 2003. 35] X. Jiang, X. Wang, and D. Xu, "Stealthy malware detection through vmm-based out-of-the-box semantic view reconstruction," Proceedings of the 14th ACM conference on Computer and communications security - CCS '07, pp. 128-138 2007.

25. M. Laureano, C. Maziero, and E. Jamhour, "Intrusion Detection in Virtual Machine Environments," Proceedings. 30th Euromicro Conference, 2004., pp. 520 - 525, 2004.

26. B. D. Payne, M. Carbone, M. Sharif, and W. Lee, "Lares: An Architecture for Secure Active Monitoring Using Virtualization," 2008 IEEE Symposium on Security and Privacy (sp 2008), pp. 233-247, 2008.

27. T. Garfinkel, and M. Rosenblum, “A Virtual Machine Introspection Based Architecture for Intrusion Detection,” In NDSS vol. 3, pp. 191-206, 2003.

28. X. Jiang, X. Wang, and D. Xu, “Stealthy malware detection through vmm-based out-of-the-box semantic view reconstruction,” Proceedings of the 14th ACM conference on Computer and communications security - CCS '07, pp. 128-138 2007.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

Cloud Architectures Encountering Data Security and Privacy Concerns- A Survey

Ms. Shweta Singh[1*]

[1]*Cloud Computing, architecture, PaaS, SaaS, IaaS, Cloud Computing, Cloud Attacks, Cloud Security.*

Abstract

Emerging as one of the most relevant IT paradigms of recent times is cloud computing. The IT environment is increasingly being transformed by cloud computing. Cloud users may access resources, apps, including infrastructure from cloud providers on the pay-as-you-go basis. As an example, cloud providers may already have apps in place for their customers to utilise. As an example, the cloud service provider may provide the capacity to design and the deploy user apps. The Massive storage infrastructure is also accessible for such database and any user-provided data. There are a slew of different cloud designs, and more are on the way. SaaS, PaaS, as well as IaaS are by far the most common, and they may be set up in private, public, communal, or a mix of these environments. It explores current cloud computing architectural advances and gives a review of numerous investigations undertaken inside cloud computing industry to address various dangers inside its design, with specific reference towards multi-cloud architectures.

Keywords: Research, Methodology, Research Methodology, Research Techniques, Qualitative research, Quantitative Research.

1. INTRODUCTION

There is a growing trend toward the cloud computing, that minimises the administration burden on businesses and enables them to concentrate on their core functions. Cloud computing is one of top 10 computing advancements, according to the Gartner survey [1]. The computer resources, data, as well

* ISBN No. 978-81-955340-6-7

as memory space provided by cloud computing are all extremely reasonably priced. Computing innovation provides numerous benefits over conventional privately held data centres, including commercial innovation, economies of scale, cheap administrative overhead, inexpensive operating as well as maintenance costs, with high quality services. In this regard, cloud computing seems to be an excellent alternative for many IT companies.

When it comes to storing data on cloud, cloud computing is employed by both small businesses as well as major corporations.

Service-oriented as well as event-driven architectures are combined in the cloud computing architecture.

The architecture of cloud computing is broken down into followings two sections: -

- Front End
- Back End

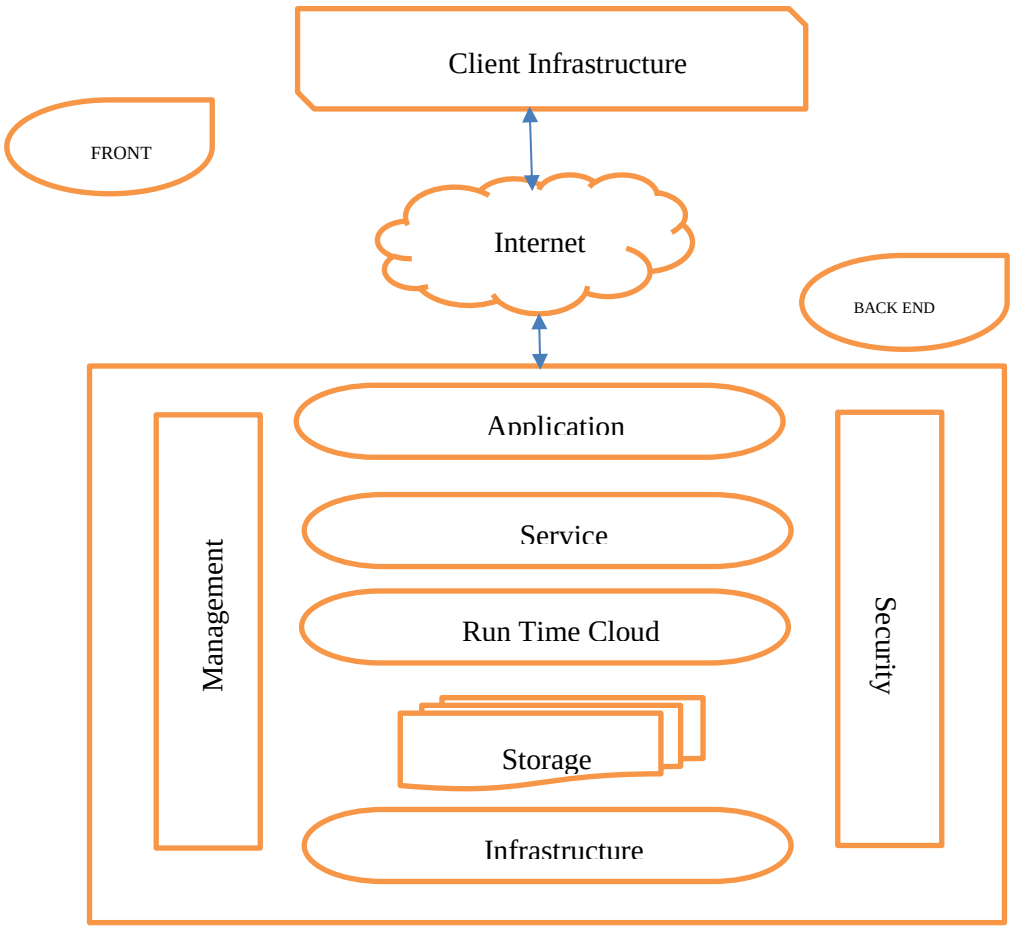

Figure 1: Cloud computing architecture

Front End

The client interacts with front end. It has client-side APIs and applications needed to connect to the cloud computing infrastructures that are included inside. Web servers (such as Chrome, Firefox, and the Internet Explorer, among others) are part of front end, as are thin and the fat clients, the tablets, as well as mobile devices.

Back End

Service provider uses back end. To deliver cloud computing services, it oversees the management of all resources necessary for this to take place. Everything from data storage to the virtual computers to servers to the traffic control systems is included.

Small and medium-sized businesses may save money and time by using the cloud to run their businesses more effectively [2]. Cloud computing has had a slew of different definitions put out over the years. The online environment provides computer resources on-demand as well as that could be remotely controlled by a large number of people may be defined as a generic definition. Visual interfaces allow users to manage as well as control such assets, including storages, infrastructures (like servers and networks), licenced software, as well as services, as well as pay for them as they use them [3][4].

2. CLOUD SERVICES

IAAS stands for the Infrastructure as a Service, SAAS stands for the Software as a Service, while PAAS stands for Platform as a Service, overall of which are offered via the cloud. In general, the cloud provides three types of services to its users: Infrastructure as a Service represented as (IAAS), Software as a Service represented as (SAAS), and Platform as a Service represented as (PAAS).

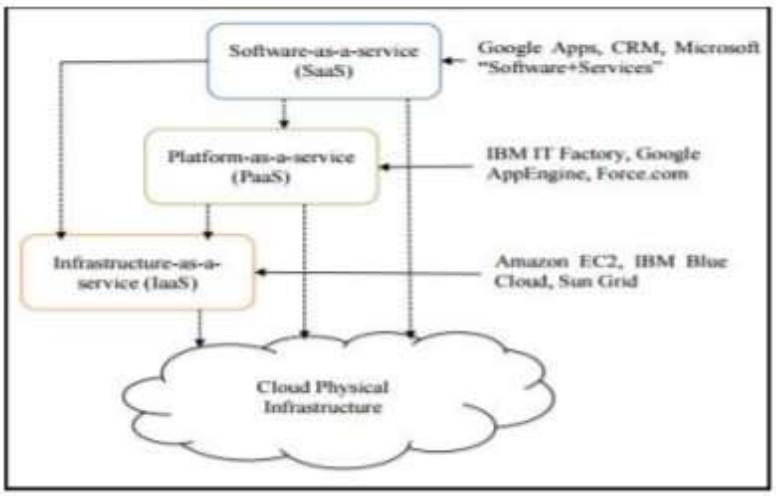

Figure 2: Cloud Services

Google Docs as well as Zoho Suite are two instances of IaaS.

- Rather without having to worry about acquiring, licencing, or maintaining software, the SaaS providers install but also distribute it as needed for their customers [5]. Some of the applications in this list may be rented from service provider for a fee based on how much time or even resources the user uses. Users may access this programme using a web browser since it is hosted on the cloud. Instances of SaaS include, Salesforce.com, GoogleDocs and others.
- If you're looking for a way to build, deploy, run, and manage your apps in cloud with not having to worry about complexity of maintaining your own infrastructure, then PaaS is for you. In this paradigm, the user has access to the actual infrastructure of cloud as well as may choose the settings that are required to execute or deploy his application. Google App Engine as well as Amazon Web Service are two instances of PaaS.

3. CLOUD DEPLOYMENT MODELS

Cloud service providers may provide a variety of deployment methods, including the public, the private, the hybrid, as well as the community cloud [8].

- Because a large number of people may access and use a public cloud, it is less secure and more exposed to various dangers.
- A private cloud is one that is exclusively used by a single organisation and is the only accessible by that business. As a result, private clouds are more secure than the public ones.
- There are two types of hybrid cloud deployments: those that use both public as well as private clouds, and those that use either one or the other. In comparison to public cloud, this solution provides greater security and lower operating costs.
- As being such, community cloud is indeed a private cloud utilized by numerous businesses and isn't really a different deployment architecture.

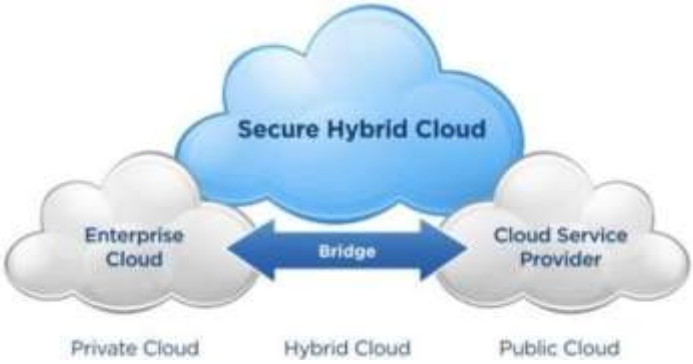

Figure:3 Cloud Models

4. ATTACK AND CORRECTIVE RECTIFICATION WITHIN A CLOUD

4.1. SaaS layer attack:

Among SaaS customers, data security concerns such as data backup, the data access, and the data availability are indeed the most common complaints.

1. **DoS attacks**

Denial-of-service assault (DoS) was among the Cloud's most notable attacks. The hacker's main goal was to exhaust all of the user's personal information by sending out a large number of the request packets across the network..,[9,10]

2. **SQL injection attack:**

A malicious cryptogram or code was injected into network in the shape of legitimate input, with the goal of stealing all data about the victim's usage of the internet, such as their registered user, pin, credit card number, etc. That's when the hacker gets access towards the user's private information illegally. [11].

3. **Authentication attack:**

Weak usernames and passwords were to blame for the assaults on authentication. In just this authentication assault, which is a little out of control, the hacker pretends to become a user to mislead the system as well as get access they shouldn't have. [12].

4.2. PaaS layer attack:

The side channel and cross-site attacks were other names for this assault.

1. **Port Scanning Attack**

This was a well-known exploit wherein a hacker gains access to a portal URL, extracts data, and either destroys or misuses the data. [13].

2. **Metadata spoofing**

attack Here in which the attacker access the file and make some modifications or else delete some of the important operations [14].

3. **Man-in-the-browser**

Attack Here in which the attacker was stand between the sender and the receiver and can access the information [15].

4. **Phishing / Spoofing attacks**

Phishing or the spoofing attacks will have an effect on both the server and users. Here the user can redirected to the spoofed web link and the attacker can access and get the personal information about the user [16].

4.3. IaaS layer attack

Attacks will occur often on such layer because virtualization administrator lacks a defence opening [17].

1) **Cross-virtual-machine attacks** Another name for this technique is side channel attack. While extracting as well as destroying some secondary data like power, volt, but also minutes, here is where user-confidential details may be found. [18].
2) **Virtual machine (VM)** rollback attack: Here, attacker has access to the passcode of virtual machines, thus he or she may take a picture of one and execute it without user's knowledge. The brute-force assault is used in this case. A component for controlling permissions, rollback, may be used by the attacker to alter user's ease of access or authorisation code [19].
3) **VM escape attack:** Attackers target downed guest operating systems or memory information throughout such sort of attack. Attackers have full control of guest operating system beyond this point. [20].

5. SECURITY ARCHITECTURES AND PROTOCOLS

In current history, a number of academics have come up with solutions for cloud security issues. A few of the security designs including models provided by professionals in the fields of availability of services, secrecy, including data integrity, such as proof of the data ownership, recoverability, dynamic audits, as well as data deduplication within single as well as multi-cloud systems, are included below. However, even though multi-cloud storage offers a high degree of security, certain means for verifying data integrity must be in place in the event that unauthorised modifications are made. When it comes to data integrity verification in the cloud computing, public auditing is perhaps the most popular solution. A user's data and also their requester's identification are kept hidden by such auditing procedures in order to preserve complete confidentiality. This may be done by implementing Evidence of Retrievability also represented as (PoR) techniques, which enable verifiers to figure out if the data block or even file is owned or not owned by a prover. [21][22][23].

[23] The PoR technique was created in 2003, and it attempts to guarantee the accessibility of files exchanged across several peer servers. To ensure the integrity of file, they advocated the use of the error-coding to file, as well as peer-to-peer file verification. Their technique assumes that each instance has its own MAC and therefore does not need a solitaire server's error correcting. The study's definition of verification method is also vague.

6. CONCLUSION

Despite the fact that cloud computing security has been extensively explored, the goal of this article, for example, the review of literature within the field of the cloud computing security, has been met (Cloud fundamentals, issues as well as various security infrastructures). We discussed a variety of the cloud deployment techniques and cloud-based services. The issues which prevent most businesses from using cloud services, including as integrity, accessibility, and protection, are examined in depth. Research shows that multi-cloud designs are better equipped to deal with cloud computing challenges, and they've been a popular trend in recent years because of their capabilities. To sum up, we can assume that the multi-cloud systems are always in their early stages; they still have a long way to go before they can be considered secure enough to protect data as well as user privacy, as well as efficient enough to handle complex computations, communications, including dynamic operations inside an efficient manner.

REFERENCES

1. Gartner, “Gartner identifies the top 10 strategic technologies for 2011”, “web reference”: http:// www.gartner.com/it/page.jsp?id=1454221, “Last access date”: 02 Dec. 2016.
2. S. Subashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing", Journal of Network and Computer Applications, 34(1), 2011, pp 1-11.
3. Hassan, Qusay (2011). "Demystifying Cloud computing" (PDF). The Journal of Defense Software Engineering. CrossTalk. 2011 (Jan/Feb): 16–21. Retrieved 11 December 2014.
4. Peter Mell and Timothy Grance (September 2011). "The NIST definition of Cloud computing" (technical report), National Institute of Standards and Technology: U.S. Department of Commerce. doi:10.6028/NIST.SP.800-145. Special publication 800-145.
5. Hassan, Qusay (2011). "Demystifying Cloud computing" (PDF). The Journal of Defense Software Engineering. CrossTalk. 2011 (Jan/Feb): 16–21. Retrieved 11 December 2014.
6. Harold C. Lin, Shivnath Babu, Jeffrey S. Chase, Sujay S. Parekh, “Automated control in Cloud computing: Opportunities and Challenges”, Proc. of the 1st Workshop on Automated control for data centres and clouds, New York, NY, USA, pp. 13-18, 2009, ISBN: 978- 1-60558-585-7.
7. William Y. Chang, Hosame Abu-Amara, Jessica Feng Sanford, Transforming Enterprise Cloud Services, London: Springer, 2010, pp. 55-56.
8. E. Gorelik, “Cloud computing models,” 2013. [Online]. Available: http://web.mit.edu/smadnick/www/wp/2013-01.pdf. Accessed: Feb. 12, 2016.

9. G. Somani, M. S. Gaur, D. Sanghi, M. Conti, and R. Buyya, "DDoS attacks in cloud computing: Issues, taxonomy, and future directions," Computer Communications, vol. 107, pp. 30-48, 2017.
10. L. Alhenaki, A. Alwatban, B. Alamri, and N. Alarifi, "A Survey on the Security of Cloud Computing," in 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), 2019, pp. 1-7.
11. P. Deshpande, S. Sharma, S. K. Peddoju, and A. Abraham, "Security and service assurance issues in Cloud environment," International Journal of System Assurance Engineering and Management, vol. 9, pp. 194- 207, 2018.
12. A. Singh and K. Chatterjee, "Cloud security issues and challenges: A survey," Journal of Network and Computer Applications, vol. 79, pp. 88-115, 2017.
13. N. Almasalmeh, F. Saidi, and Z. Trabelsi, "A Dendritic Cell Algorithm Based Approach for Malicious TCP Port Scanning Detection," in 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), 2019, pp. 877-882.
14. R. Anitha, P. Pradeepan, P. Yogesh, and S. Mukherjee, "Data storage security in cloud using metadata," in 2nd International Conference on Machine Learning and Computer Science (IMLCS'2013), Kuala Lumpur (Malaysia), 2013, pp. 26-30.
15. A. Mallik, A. Ahsan, M. Shahadat, and J. Tsou, "Man-in-the-middle-attack: Understanding in simple words," International Journal of Data and Network Science, vol. 3, pp. 77-92, 2019.
16. V. S. P. P. C. Kumar and S. P. Rao, "Phishing attack detection," ed: Google Patents, 2019.
17. F. Mohammed and D. Uliyan, "A New Password Authentication Scheme Resistant against Shoulder Surfing Attack," 技術學刊, vol. 34, 2019.
18. S. Anwar, Z. Inayat, M. F. Zolkipli, J. M. Zain, A. Gani, N. B. Anuar, et al., "Cross-VM cache-based side channel attacks and proposed prevention mechanisms: A survey," Journal of Network and Computer Applications, vol. 93, pp. 259-279, 2017.
19. P. Mishra, E. S. Pilli, V. Varadharajan, and U. Tupakula, "Intrusion detection techniques in cloud environment: A survey," Journal of Network and Computer Applications, vol. 77, pp. 18-47, 2017.
20. Y. Xia, Y. Liu, H. Chen, and B. Zang, "Defending against vm rollback attack," in IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN 2012), 2012, pp. 1-5
21. M. Lillibridge, S. Elnikety, A. Birrell, M. Burrows, and M. Isard. “A cooperative Internet backup scheme”, USENIX Annual Technical Conference, General Track 2003, pages 29—41, 2003.

22. H. Shacham and B. Waters, "Compact proofs of retrievability," Advances in Cryptology-ASIACRYPT 2008, pp. 90–107, 2008.

23. Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, "Enabling public verifiability and data dynamics for storage security in Cloud computing," Computer Security–ESORICS 2009, pp. 355–370, 2009.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

A SURVEY ON GREEN CLOUD COMPUTING

Devaraju Hanumanthu[1*]

[1]*Lecturer in computer science, Government College (Autonomous), Rajahmundry. devarajs.h@gcrjy.ac.in*

Abstract

Green cloud computing has been concerned with building cloud computing resources in the way that minimises their environmental effect. Green clouds have the potential to save the significant amount of energy while also lowering operating expenses. Due to the rapid increase in global temperatures, this is now a requirement. A green cloud computing strategy is essential as more enterprises migrate to the Internet of Things (IoT). HPC that is High Performance Computing, corporate, as well as web applications may all be run on the cloud because to its scalability and economics. There has been a rapid increase in the utilisation of big data centres (DC) as even the demand for such data centres grows. This High energy use leads to large operating costs as well as high emissions of greenhouse gases. Cloud computing may have a negative influence on an environment if the right energy solutions are not implemented. Additional heat is released when the use of processing chips increases. As a result of this wasteful heating, the system must be cooled further, which in turn creates additional heat. We must thus find a way to achieve system equilibrium while using less energy. An energy-optimized cloud computing system is possible using a green algorithm. An overview of the green cloud computing is provided, along with information on how it may be implemented. Green cloud applications as well as problems are also discussed in this article.

Keywords: Cloud computing, Green cloud computing, Environment Sustainability, Energy Consumption.

1. INTRODUCTION

Because of the widespread use of computers as well as data centres, a significant quantity of electricity is already being used. All of today's industries, from healthcare to autos to banking sector are heavily

* ISBN No. 978-81-955340-6-7

reliant on IT, which results in a high amount of the energy consumption and a rise in expenses, that's why Green computing was born. It is the primary goal of the Green computing to ensure that all computer equipment are utilised efficiently and environmentally friendly, such as the design as well as manufacture of devices. Also, the gadgets are to be recycled and reused as part of the plan. Designed to promote digital equipment recyclability while also promoting energy efficiency, the Energy star programme was created by the Environmental Protection Agency to promote a more environmentally friendly approach to the IT sector. A few of the approaches to encourage green computing are just as follows. Design, production, and disposal all fall under the category of environmental friendliness.

2. NEED FOR GREEN CLOUD COMPUTING

In today's environment, it's impossible to envision a day with no usage of some kind of technological innovation. We're dealing with a lot of the data since we're using a lot of the technology-based items. Different data centres throughout the globe maintain this massive quantity of information. Since you don't want your server to fall down, one need to have a lot of the energy and electricity to keep such data centres running smoothly as well as effectively at all times. In today's world, information technology permeates almost every industry.Banking, Healthcare, the media, as well as automobiles are just a few of the industries that depend on IT on a daily basis. Without it, operations would grind to a halt. There are numerous additional industries where it Is widely used, resulting in rising energy consumption. As a result, expenses will rise over time. Technology has also seen a lot of progress in the last few years. One of these is Cloud Computing that is in great demand at the moment, among others. When it comes to using computing services, this implies that we don't need any more hardware on our end to do so. Because everything takes place online, we may take use of a distant server's processing capacity on our local workstation. This same way we presently utilise technology has been fundamentally altered by the advent of cloud computing. It's been a huge help in a lot of other areas, too. However, with the rise of Cloud Computing, we're now dealing with a massive amount of data that we weren't previously dealing with. In addition, the quantity emissions of carbon has risen as a result of the widespread usage of Cloud Computing, that might have a negative impact upon the environment. Although we can't even limit Cloud Computing since it has become an essential part of our life, we can find strategies to lessen the negative emissions that arise from it. Finally, Internet of Things (IoT) concerns with actual goods that connect and interact via the Internet using sensors as well as software built into them. Connectivity, data, as well as the internet are all words that may be used in the same phrase because of Cloud Computing and indeed the massive data centres located at the faraway servers. Thus, the Internet of Things relies heavily on the data centres for its infrastructure. Data centres, on the other hand, need a lot of power. New technology and goods have led to an increase in energy consumption, which is a direct result of this trend. There is a need to lower the quantity of energy used wherease still using such resources, thus we must discover a means to do this Thus, Green Cloud Computing has emerged, referring to the environmentally benign method wherein Cloud Computing is used. It is the goal of Green Cloud

Computer to utilise computing resources in an environmentally and economically responsible manner, while yet delivering the very same value to an end user as before. More and more businesses throughout the globe are embracing Green Cloud Computing. In today's society, there are several reasons why Green Cloud Computing is essential.

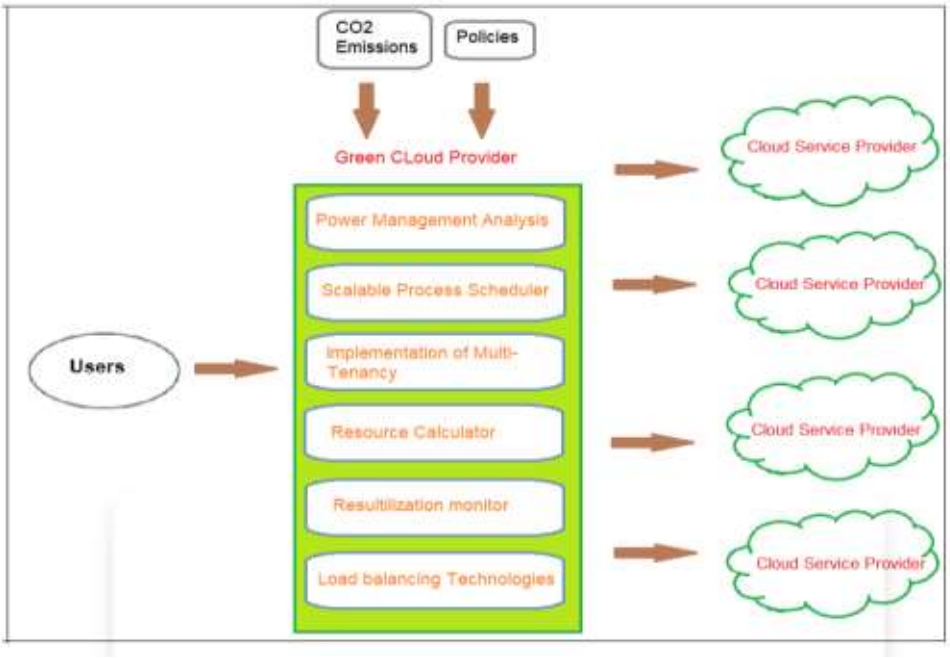

Figure 1: Green Cloud Computing Architecture

Understanding computer's life cycle is essential before implementing the GREEN IT idea. Fig. 1 was used to demonstrate this point. There are a few of these green groups that aren't quite that eco-friendly:

1) Data centres and corporate computing ecosystems throughout the globe may benefit from the Green Grid, a worldwide alliance of IT firms and professionals dedicated to increasing energy productivity in the data centres. Microsoft, HP, APC, as well as Dell are just some of the companies represented just on Green Grid's board of directors, along with other major technology companies like IBM as well as Intel.
2) A federal agency, the U.S. Environmental Protection Agency (EPA), was established to protect human health as well as the natural environment. This group also collaborated with the US Department of Energy as well as the US Environmental Protection Agency to establish the Energy Star programme.

End users in a business have a variety of priorities when it comes to implementing GREEN IT. Such are the ones I'm talking about:

- Energy/Power expenditure – When we use electricity, we must conserve it.

- Reuse - After recycling, it's important to think about whether or not the parts we've collected may be put to good use.
- Hardware/Software- Relevance in terms of computer acquisition.
- Energy utilization- The efficient use of resources and the systematic approach to computing that it requires.
- Minimize misuse- Utilizing computers to reduce the amount of Organic resources that are wasted.

3. PARAMETERS USED FOR MEASURING POWER CONSUMPTION

Processor and data centre power consumption can be assessed using a number of different parameters, including TDP, DCiE, PUE and Performance per Watt, as well as metrics such as Carbon Usage Effectiveness also represented as (CUE), Water Usage Effectiveness also represented as (WUE), and Data Center Productivity (DCP), all of which are used to determine how efficiently a processor or even the data centre is using energy. [1].

Table 1: Parameter used for Power Performance

Parameter	Description
Carbon usage Effectiveness (CUE)	The data center's contribution to global warming pollution is quantified using this metric. CUE=ECO2/EI , and there Eco2 = Entire co2 emitted from the total energy consumed by data centre facility. EIT = IT equipment's total power consumption.
Water Usage Effectiveness (EUE)	Each year, it's a way to calculate how much water a data centre needs. Here's how it's calculated: WUE = Water used yearly /EIT
	In other words, it's an indicator of how much productive labour a datacenter has produced. DCP=Useful Work-done/Tresource, wherein Tresource is the total amount of resources needed to do such useful work.
Data Centre Productivity (DCP)	The greatest amount of electricity that can be dissipated by a computer system's during cooling is measured by it. In the context of an actual application, this is the max amount of power that a computer chip may consume.
Thermal Design Power (TDP)	The greatest amount of electricity that can be dissipated by a computer system's during cooling is measured by it. In the context of an actual application, this is the max amount of power that a computer chip may consume.
Power Usage Effectiveness (PUE)	Comparison of computer application as well as infrastructure equipment energy consumption as well as overhead energy waste is one of its primary uses. PUE = Total Facility Power / IT Equipment Power.
Data Center Infrastructure Efficiency (DCiE)	In other words, it is PUE's opposite. For comparing efficiency of the data centres, PUE as well as DCiE are two measures that are widely used in the industry. 1/PUE is the formula for DCiE. DCiE = IT Equipment Power/Facility Power.
Performance per Wat	The amount of data a processor could process for every watt of electricity it consumes is known as its throughput. This has to be a record. In other words, it assesses how much work a computer can do for each watt of electricity it uses.

Green Energy Coefficient (GEC)	Measurement of green energy (electricity from renewable sources) being utilised by a datacenter's facility. There are 1,000 kilowatt-hours in one kWh. As Green Energy Consumed/Total Energy Consumed, GEC is described as:
Compute Power Efficiency (CPE)	A datacenter's computational efficiency may be measured using this metric. Each watt of electricity spent by a server or cluster was not productive all time, thus some facilities used power even when they were idle, while others used power to do computation. To calculate CPE, divide IT Equipment Utilization by PUE, which is equal to the product of IT Equipment Utilization and IT Equipment Power.
Energy reuse factor (ERF)	Energy that can be reused except in a datacenter and is consumed by the facility is referred to as reusable energy. Total Energy Consumed/Re-sued Energy Used = ERF.

4. GREEN COMPUTING APPROACHES

Businesses are using the given green computing techniques.

Virtualization & Use of Terminal Servers: It is possible to run numerous operating systems on a single machine using virtualization, a technique known as "terminal server virtualization." Applications seem to execute on separate machines. Up to 80percent of total of energy may be saved by using shared servers as well as terminals. [2]

Power Supply & Power Management: Only 60percent of total of power available was utilised for transmission and 40percent of total was squandered. Green computing technology is expected to reduce the amount of energy required. Utilizing a green method for the cloud computing power management lowers overall power usag. [2]

I. APPLICATIONS IN GREEN CLOUD COMPUTING

There is a lot of interest in green computing there in areas as follows:

- Management of data centres' energy use
- a green wireless network
- Using a Big Data Network and Green Parallel Computing together is a powerful combination
- Green computing via the use of algorithms.

5. RENEWABLE ENERGY AND GREEN COMPUTING

The public cloud offers better scalability as well as flexibility to enterprises than on-premises technology. Using the cloud may reduce an organization's carbon footprint, something big cloud providers have been known to highlight on occasion. A good illustration being "Data Center Alley" in the Northern Virginia, which is home to over 100 data centres as well as 10 million sq ft of the data centre space, as well as is not too far from the ParkMyCloud's headquarters. [3] D ata centre business

was welcomed in the Northern Virginia since the good economic effects. More data centres will be required as the demand for the cloud services develops. It is estimated that the installed energy capacity of nine big (500-MW) coal power plants in the Northern Virginia was 4.5 gigatonnes of the commissioned energy previous year. Major cloud providers like (AWS) that is Amazon Web Services have been criticised by Greenpeace as well as the other environmental groups for not doing more to safeguard the environment while running data centres. As per them, the problem would be that the cloud providers depend on the commissioned energy from the energy firms who only concentrate on the dirty energy (coal but also natural gas) as well as very little over renewable energy efforts. Although the claims have drawn attention to energy companies, we demanded to discover what (incase anything) the main cloud providers do to lessen their dependency on such sources of energy as well as supply data centres with the cleaner fuel to make the green computing a reality.

6. FUTURE OF GREEN COMPUTING

Instead than focusing on reducing energy usage, Green Computing's future will be built on improving performance. According to Green IT's primary goal of reducing the organization's own carbon footprint, energy costs are reduced both in Data Centers and on individual computers. Secondary to Data Center's energy consumption, Green IT has to concentrate on an innovation and better integration with broader corporate social responsibility activities. As a result of this supplementary emphasis, Green Computing techniques must be developed [5]. In the context of sustainability, corporate value is created although long-term environmental resources really aren't adversely affected by the activities of the company.

7. GREEN INITIATIVE IN GREEN APPROACHES

Virtualization

Because of the requirement to save separate systems from being overly used, virtualisation was born. However one technology that allows users to connect to servers from a distance is virtualization. As a consequence of this, original hardware as well as system may be disconnected from integrated system, which reduces power usage and need for additional cooling. Instead of putting up a server as well as cooling system, a large system server may be accessed through virtualization rather than a physical installation (Green Computing, in the year 2013). By severing the ties that bind applications, system services, components, as well as storage systems, virtualization is ideally suited for usage in environmentally friendly computing. The existence of virtualization is supporting green computing in many forms. Green initiative nowadays moved into the concept of the virtualization where cloud computing have a major role in that. The statistics of cloud users based upon the benefits provided by them has been increased sufficiently for the last two years.

Power Management

The longer battery life, lower cooling needs, and less noise all emphasise the importance of the power management in almost any computer system. The costs needed for operation of the system is also considered to be one of the main reason for the individuals to concentrate more on the power management support in the system resulting in the stability of the system leading to probable maintenance of the impact that it can create on the environment. As an example of the power management approach that has been widely and efficiently supported, a system's hibernation option turns off the RAM as well as CPU automatically, minimising the amount of the background system activity. There are certain programs available nowadays that can actually alter even the voltages of the system probably resulting in the reduction of the heat produced and electricity consumed in the system which is generally called as under volting.

Power Supply

The use of green technologies for power supply will also aid in the realisation of green computing idea. M The ore power the system uses, more effective it may be designed. Individuals may save the most electricity in system by acquiring and utilising power providers that have earned “80 plus" accreditation "State Legislation on E-Waste, 2008". Energy consumption as well as heat generation may be reduced by using this kind of energy-efficient and usable power supply.

Displays

The heat generated by the screens directly affects the system's power consumption. As a result, it's generally agreed that switching to LED displays instead of LCD monitors is the best option. It's since the fluorescent bulb being used consumes more power and generates more heat than is healthy for the bulb. LCDs have been shown to be 66percentage points more energy effective and 80percent more qualified in reducing the size as well as weight of such system. In comparison it is found that the CRT is actually consuming around 120W power which is double the power that is been used by 22” LCD. As a result, it is critical for anybody acquiring a computer system to thoroughly inspect all of its components, such as displays, before making a final decision.

Video Cards

As video cards can't utilise shared terminals, think clients, or even the desktop sharing capabilities that are very beneficial in conserving energy in system, reducing their utilisation is regarded a sensible option. Older video cards may be reused since they utilise less power, lowering the need for heat sinks or even fans. Choosing a GPU that has an average wattage or even the performance per watt is often regarded to become a much better option when it comes to implementing a green system.

CONCLUSION

The quantity of energy used by the cloud data centres is expanding rapidly until enough companies migrate to the cloud, which has a huge impact on the planet's carbon footprint. This issue can be solved

with green cloud computing, which reduces energy usage and optimises the distribution of resources. The development of the green cloud computing is being aided by cutting-edge AI approaches. It is possible to use green computing in a variety of sectors, including IoT as well as big data analytics. The general public must be made aware of the significance of eco-friendly computing. When it comes to the ecology, adopting green computing would be incredibly advantageous in the future.

REFERENCES

1) Jain, A.; Mishra, M.; Peddoju, S.K.; Jain, N., "Energy efficient computing- Green cloud computing," Energy Efficient Technologies for Sustainability (ICEETS), 2013 International Conference on , vol., no., pp.978,982, 10-12 April 2013

2) https://www.jigsawacademy.com/blogs/cloudcomputing/green-cloud-computing

3) https://mytechdecisions.com/facility/cloud-providers-greencomputing/

4) http:/www.mosqueterofas.blogspot.in/2012/09/benefits-of-green-computing.html.„Science And Technology".

5) http://www.greencompute.com/green_computing.html.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

An Analysis of Cloud Computing's Resource Allocation Methods

A. Shenbaga Bharatha Priya[1*]

[1]*Teaching Fellow, Ramanujan Computing Centre, Anna University*

Abstract

In order for a cloud subscriber to run their applications on many platforms, anywhere, at any time, utilising just the resources given by the cloud service provider, the cloud service providers play a critical role. As a result, cloud customers must still deal with the challenge of quickly accessing the computing resources they need. For many users in several applications, this will have an effect on the service time and service level agreements. As a consequence, earlier research on cloud resource allocation need to be completed. This study examines the distribution and management of cloud computing resources. There is a lot of discussion here on alternative resource allocation systems and the difficulties they pose.

Keywords: Cloud computing; Resource Allocation; Infrastructure; service of Clouds.

1. INTRODUCTION

Cloud computing is nothing more than the Internet's mutation. On the cloud, it's possible that people to have all of the resources they need. When it comes to on-demand IT services and products, cloud computing is the natural next step. It's just a matter of time until cloud computing becomes a critical component in the development and deployment of distributed applications. The availability of a wide range of resources in the cloud is helping to make cloud computing more popular among the general public. Many cloud service providers including Microsoft, Amazon, Google and IBM provide platform as a service (PaaS). Interoperability capabilities enabled developers to distribute apps among machines hosted by a single enterprise. A cloud computing provider manages and deploys a broad network of

* ISBN No. 978-81-955340-6-7

computer resources for these applications. While developers benefit from a managed computing platform, they do not have to spend time and money setting up and maintaining their own network infrastructures. As a result, the cloud must address how to manage quality of service (QoS) for cloud users who share resources and establish service level agreements (SLAs). Clients may access all of their resources via a single point of access provided by cloud computing. Cloud computing provides several advantages, including a defined and abstracted infrastructure, a completely virtualized environment, dynamic infrastructure, pay-as-you-use, and no software or hardware installations. As a result, the most pressing concern is the sequence in which the requests are granted. As a consequence, the scheduling of resources is affected. Effective usage of system resources is crucial to getting the best potential performance out of the system. As a result, cloud computing services are mostly marketed on a per-minute or per-hour basis. Because of this, it is important to plan the use of resources effectively. When it comes to cloud platforms, there are two degrees of resource allocation (or load balancing). At the time of upload, a load balancer distributes the requested instances among physical machines in an effort to balance the computational demand of various applications. Multiple requests for the same application should each be assigned to a distinct instance in order to spread out the computational load as evenly as possible. Elastic load balancing (ELB), for example, is used by Amazon EC2 to manage how incoming requests are handled. Requests to specified availability zones, individual instances, or instances with the fastest response times may be directed by application designers. Here, we explore the importance of allocating resources in the following paragraphs.

2. SIGNIFICANCE OF RESOURCE ALLOCATION

RA in cloud computing is an internet-based technique for distributing resources to online cloud applications. It is not necessary for suppliers of service to manage the resources assigned to each individual module as a consequence of the resource provisioning process. The Resource Allocation Method (RAS) is an integrated strategy for utilising and allocating limited resources in the cloud environment to meet cloud application needs. All of the resources required to execute a task requested by a user must be gathered in this manner. The best RAS takes into account the timing and sequence of resource distribution. The following characteristics should be avoided in an ideal RAS:

- There is a problem known as resource disputation when two applications try to use the same resource at the same time.
- There is a high demand for resources when there are a limited number of resources.
- When resources are separated, a scenario known as "resource fragmentation" emerges. [Resources will be available there, however they can't be allocated to the required application.]
- A situation known as "over-provisioning" occurs when an application is provided with more resources than it requested

- Resources are not allocated to an application in sufficient quantity to meet the expected demand.

3. AT A GLANCE: THE MANAGEMENT OF RESOURCES

3.1. Resource Management

In our view, research management encompasses the finding, allocation, and monitoring of resources, as seen in Fig. 1 below. These processes are responsible for managing physical resources such as CPU cores, disc storage, and network bandwidth. Slices of these resources must be allocated to virtual machines executing a wide variety of tasks, all of which might potentially compete for the same resources. Here is a breakdown of the taxonomy for resource management components:

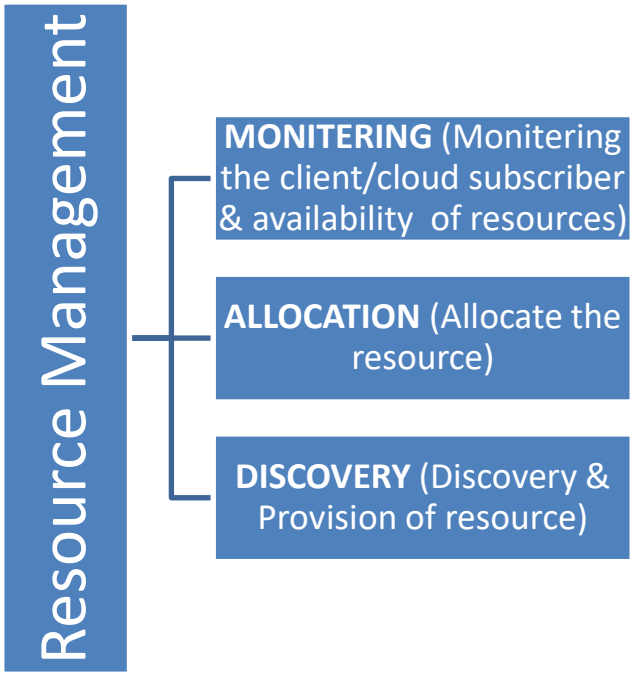

Figure: 1 Elements of resource management.

The discovery process is the core of resource management. Searching for the right resources to meet the application's needs is the first step. [1] In this case, the cloud provider is in charge of overseeing the whole thing. The resource broker or user broker is doing this step in order to find out what resources are accessible. The discovery process includes detailed explanations of the resources accessible. A resource management system (RMS) and other RMSs that interact with it may learn the status of the resources that they manage via resource discovery, according to [2]. The discovery of resources works in

conjunction with the distribution of resources to give the information server with data on the current condition of resources.

Using the Internet, a cloud application may be assigned to a source that is accessible and ready to go. Using a pay-as-you-go model, these resources are distributed to users depending on their requests. Scheduling and dispatching are used to assign resources in this process. The client's given resources will be scheduled by the scheduler. As a result, the dispatcher will assign the client's resources. There are a number of ways to monitor and manage hardware and software infrastructures, including resource monitoring as specified in paper [4]. It also gives data and Key Performance Indicators (KPIs) to assist in the allocation of resources for both platforms and applications in the cloud. When there is a problem with the physical layer or the services layer, this is a key component for monitoring the state of the resources. In our perspective, allocation and discovery should be included in provisioning, but tracking should be managed as a distinct process. All three processes, however, are interrelated in order to offer customers with resources. In the next part, we'll go through how the resource provisioning and resource management processes work.

4. RESOURCE ALLOCATION STRATEGIES & ALGORITHMS

a) Topology Aware Resource Allocation (TARA)

Allocating resources in the cloud may be done in a number of ways. In the view of [5,] it proposes an architecture for optimising the distribution of resources in cloud systems based on infrastructure as a service. The performance of distributed data-intensive applications hosted on existing IAAS systems may be affected by a lack of understanding of the hosted application's needs. This resource allocation problem may be solved by implementing an architecture that employs a "what if" approach to guide the IaaS's allocation decisions. The concept utilises a prediction engine and a genetic algorithm to evaluate the performance of resource allocations and a large search space with a lightweight simulator. TARA reduced job completion times by as much as 59 % to application-independent allocation solutions. TARA's architecture is shown in Figure 2, which depicts the inputs used in the scoring process.

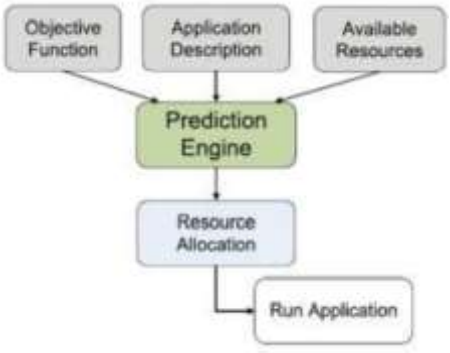

Figure: 2 Basic Architecture of TARA

b) Resource Allocation Resource Allocation Using a Linear Scheduling Strategy

It is feasible to provide the highest level of service to all customers [3] by employing the cloud environment's service node to handle all client requests and use all CPU, memory, and throughput resources. Waiting and response times increase when scheduling resources and tasks independently. Scheduling jobs and resources is handled by the LSTR scheduler, which uses a linear approach to the problem of resource allocation. KVM/Xen virtualization and LSTR scheduling are used in conjunction with one server node to create an IaaS cloud environment that maximises system performance and resource usage. There must be an integrated approach to resource consumption and allocation to maximise resource usage. The scheduling algorithms are primarily concerned with distributing the resources among the requestors in a way that maximises the given Quality of Service (QoS) characteristics. The cost function is the QoS parameter that we used in our assessment. The LSTR scheduling approach takes into account both the jobs and the number of virtual machines available. Because of this, resources will be used more efficiently.

c) Resources for Parallel Data Processing Allocated Dynamically

Efficient Use of Resources via Dynamic Allocation [6,7] presents a new cloud-optimized framework called Nephele for parallel data processing. Nephele is the first data processing framework which enable the dynamic allocation and de-allocation of various cloud computing resources during the execution of jobs and task scheduling. Virtual machines may be allocated to certain tasks in a processing job and automatically created and destroyed as the work progresses.

I. Benefits and Drawbacks of Resource Allocation Methods

Cloud computing offers various benefits in terms of resource allocation, regardless of the size of the firm or the business marketplaces in which it works. Due to the fact that this is a new technology, there will always be certain drawbacks. The benefits and drawbacks of allocating cloud-based resources will be discussed in the next paragraph.

Benefits:

- There is no need for a user to install any software or hardware in order to access, build, or host an application using a service like resource allocation.
- Another important advantage is that there are no restrictions on location or media. Our apps and data are accessible from any system, anywhere in the globe.
- There is no need for the user to invest in expensive hardware and software.
- A cloud provider's resources may be shared across the internet when there is a shortage of such resources.

Drawbacks:

- It is impossible for users to monitor the amount of resources they utilise since they rent them from remote servers.
- When a user wishes to migrate to a different service provider for better data storage, a migration issue arises. It's difficult to move large amounts of data from one service provider to another.
- Hacking and phishing assaults are possible in public cloud environments. It's simple for malware to proliferate since cloud servers are linked.
- Peripheral devices like printers may not be able to communicate with the cloud. Many of them need the installation of local software. Networked peripherals are less susceptible to errors and malfunctions.
- To properly allocate and manage cloud resources, a more comprehensive understanding of how the cloud works is necessary, and this expertise is mostly dependent on the cloud service provider.[8]

5. CONCLUSION

In this article, we've examined resource management as a whole, as well as contemporary research-based methodologies for allocating and monitoring resources. This paper summarises various theories and methods (algorithms) in order to have a better structure and model for allocation of resources and monitoring in order to enhance effectiveness, competitive nature, and efficiency in order to achieve the desired SLA, improved resource performance, and decreased power consumption. This study's main objective is to develop a new method for allocating and monitoring cloud computing resources.

REFERENCES

1. R. Buyya and R. Ranjan, “Special section: Federated resource management in grid and cloud computing systems,” Future Generation Computer Systems, vol. 26, no. 8, pp. 1189–1191, June, 2010.
2. K. Krauter, R. Buyya, and M. Maheswaran, “A taxonomy and survey of grid resource management systems for distributed computing,” Software: Practice and Experience, vol. 32, no. 2, pp. 135–164. 2002
3. V. Vinothina, R. Sridaran, and P. Ganapathi, “A survey on resource allocation strategies in cloud computing,” International Journal of Advanced Computer Science and Applications, vol. 3, no. 6, pp. 97– 104, 2012.

4. G. Aceto, A. Botta, W. De. Donato, and A. Pescapè, "Cloud monitoring: A survey," Computer Networks, vol. 57, no. 9, pp. 2093– 2115, 2013.
5. Gunho Lee, Niraj Tolia, Parthasarathy Ranganathan, and Randy H. Katz, Topology aware resorce allocation for data-intensive workloads, ACM SIGCOMM Computer Communication Review, 41(1):120--124, 2011.
6. Daniel Warneke and Odej Kao, Exploiting dynamic resource allocation for efficient parallel data processing in the cloud, IEEE Transactions On Parallel And Distributed Systems, 2011.
7. Shenbaga Bharatha Priya, A., Ganesh, J. and Devi, M.M., 2014. Dynamic Load Rebalancing Algorithm for Private Cloud. In Applied Mechanics and Materials (Vol. 573, pp. 556-559). Trans Tech Publications Ltd.
8. Priya.A, Bhuvaneswaran R.S., 2021. Cloud service recommendation system based on clustering trust measures in multi-cloud environment. Journal of Ambient Intelligence and Humanized Computing, 12(7), pp.7029-7038.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

A Survey on Intelligence Data Analysis: Issues and Challenges

Dr. Pankaj Saxena[1*]

[1]*Professor, R.B.S. Management Technical Campus, Agra (India) , E-mail : pankajrbsmtc@gmail.com*

Abstract

Many real-world applications now need automated or semi-automatic analysis of the data, which has given rise to the new area of (IDA) "intelligent data analysis," which combines several disciplines, including artificial intelligence and statistics in particular. As a whole, they work well together: Computing power is not a replacement for statistical understanding when it comes to statistical procedures for huge data sets. As a result, data analysis systems are becoming more sophisticated. When analysing data, there are a broad variety of issues that might arise. This article discusses some of these issues and provides possible solutions. Using data from a real-world level crossing risk assessment scenario, we investigate some of these issues and ideas.

Keywords: analysis of data, mining of data, level crossing risk evaluation, extraction of rules, artificial neural networks, induction of rules.

1. INTRODUCTION

Data intelligence refers to the process of using AI and machine learning to analyse and transform massive information into intelligent data insights which can be used to improve services and investments. The use of data intelligence tools and methods may assist in the development of improved business processes via the development of a better understanding of acquired information.

The data-driven intelligence process consists of five main components: descriptive data, prescriptive information, diagnostic information, and predictive information. In these fields, you'll learn how to

* ISBN No. 978-81-955340-6-7

analyse data to comprehend it, acquire new knowledge, and come up with solutions to problems. Cybersecurity, finance, health, and insurance, as well as law enforcement, are some of the most pressing fields in need of data intelligence. An intelligent data capture system may be used to turn print documents or images into usable data in these situations.

Big data and business intelligence are strongly dependent on the utilisation of intelligent data. An intelligent data processing technique that restructured and enhanced the enormous datasets utilised by AI is capable of helping humans discover patterns, develop well-informed judgments, and adapt to changing circumstances. In order to improve the visualisation of prescriptive and predictive analytics, advanced analytics approaches are also used.

AI, high-performance computing (HPC), pattern recognition (PCR), and statistics (STAT) are all employed in IDA, which is an interdisciplinary study focused on extracting useable data. Strategic Data Intelligence, Global Data Intelligence, and the like all offer platforms and solutions for the analysis of large amounts of raw data.

There has been an increase in the need for more advanced IDA approaches as a consequence of the expansion in online and multimedia activities, as well as electronic commerce and others [1]. Analyzing vast volumes of data with detailed descriptions seems tempting on the surface, but in practise it is quite tough. There must be a plan in place to properly use the vast amounts of data that are generated by such massive and complicated datasets.

IDA derives value from data by uncovering patterns and rules in the data. It is impossible to count the number of IDA algorithms in the world, but the trend of their development may be characterised in three ways: (a) algorithm principle, (b) magnitude of dataset, and (c) kind of dataset.

2. ALGORITHM PRINCIPLE

IDA's algorithm has progressed from a basic to a complicated state during the course of its development. Early IDA algorithms were built on a basis of classical probability theory and a distance-based similarity theory based on Euclidean geometry. Computational intelligence was included into the IDA later, making its principles more sophisticated.

2.1. Probability Based Algorithm

The IDA methods based on probability probability theory are often used for the classification and grouping because of the property of probability theory itself. Prior probability and posteriori probability are used by the Naive Bayes Classifier (NBC) to classify sample data. Classification is performed by C4.S using the sample data entropy gain, while clustering is performed by Expectation Maximization (EM) using the maximum likelihood estimate of the parameters. They are extensively utilised because of their ease of implementation and high performance.

Second feature selection may increase the accuracy of NBC in huge text classification using the auxiliary feature strategy. Rework over-fitting and boost classification accuracy by using Decision Trees and neural networks. The Randomly Selected Naive Bayes (RSNB) technique overcomes the local optimum problems that afflict standard NBC through using stochastic processes in the NBC's feature stage of the selection process.

EM may be used to identify change points in multivariate data in plant monitoring. The authors [6] recommend adopting an EM hybrid approach based on forward-backward Kalman filtering for data-driven fault identification. [7] do high-dimensional Boolean factor analysis using two novel EM approaches.

2.2. Euclidean Distance Based Algorithm

It's possible to visualise the similarity between distinct components in an n-dimensional dataset by comparing the Euclidean distances between the dataset's n-dimensional vectors. Euclidean distance IDA techniques that focus on finding the cluster centres by minimising the total number of mean-square errors, such as the k-Means and the k-nearest-neighbor (k-NN) algorithms, are widely used. Example points in space are represented in a higher-dimensional context by use of the Support Vector Machine (SVM) model. These points are then translated to a higher-dimensional space in order to create two hyperplanes that are as broad as feasible.

[8] identify benign and malignant breast cancer tumours using a combination k-Means and SVM algorithm.

By incorporating an evolutionary approach into the k-Means method, we may lessen our dependence on the original cluster centres while simultaneously improving our capacity to handle scattered data. The repetitious training for continuous input condition may be eliminated by using a quick k-Means algorithm to graphic processing [10].

3. DATA ANALYSIS TASKS AND TECHNIQUES

Predictive modelling, clustering, and link analysis may all be part of a data analysis process, depending on the end user's goals and interests [11]. Making predictions based on fundamental aspects of the data is the purpose of predictive modelling. Using a mathematical model, data must be mapped to one of several established classes or to a real-valued forecasting variable. Predictive modelling may be done using any supervised machine learning approach that trains a model based on previous or current data. In order to train the model, we give it a set of previously known information and ask it to predict the future with the correct answers. Neural networks, decision trees, Bayesian classifiers, K-nearest neighbour classifiers, case-based reasoning, genetic algorithms, and rough and fuzzy sets are some of the approaches used to map discrete-valued target variables. Variables with continuous values may be mapped in many ways, including regression, induction trees, neural networks, and radial basis

functions.Clustering is a technique used to create hierarchies of events by grouping together those having similar features. Clustering may be accomplished using any unsupervised machine learning approach for which the incoming data set does not include a preset set of data categories. There are certain pre-existing facts that the model is given, from which it produces categories of data with comparable features. These include partitioning, hierarchical strategies based on density and model-based methods [12]. "

Using link analysis, one may discover the intrinsic connections between data points. Achieving this objective is made possible by tasks such as discovering associations, identifying sequential patterns, and performing other sequence-discovery activities [11]. They reveal samples and patterns by forecasting correlations between elements that are not evident. When it comes to link analysis, it's all about counting all conceivable combinations. Apriori and its variants [13] are among the most often utilised algorithms.

4. DIFFICULTIES THE IDA FACE IN A BIG DATA ENVIRONMENT

There are significant roadblocks for IDA in the age of big data, when people's desire to get the most out of their data has never been greater. In a big data world, the IDA is confronted with four perspectives: (a) Large data management, (b) data gathering, (c) data analysis, and (d) application pattern are all aspects of the data lifecycle.

4.1. Management of Big Data

Hadoop Distributed File Solution (HDFS), for example, is a rather established system for managing enormous amounts of data. A good large data management system, on the other hand, does more than just store information correctly. Managing data lifecycles, data security, and costs in the context of big data management are all important considerations for getting the most out of data.

- **Data Life Cycle Management**

The most difficult part of managing the life cycle of data is determining how long a piece of data should be kept in storage. The conventional wisdom is that the data life cycle ends when the analysis of the data is done. Data lifecycle management is no longer a straightforward subject in the big data era. Even for an identical dataset, the value extracted from data by various users is varied. Data lifecycles are also varied because of this. When a patient has fully healed, their medical record's life cycle comes to an end from the patient's perspective. However, if a clinician is interested in learning more about a patient's family history of allergies, the medical record might provide valuable information. Medical records integration is critical for epidemiologists looking to learn more about an outbreak. If you maintain a dataset differently, you could get different information out of it, as shown in this scenario.

4.2. Data Security Management

The IDA has another challenge: managing data security throughout its lifecycle. People in a big data environment are always worried about their data's privacy and security. Personal, corporate, and national secrets may be compromised at any point in the data lifecycle in such an open environment. Data encryption may help protect data, but it can also limit the speed at which it can be processed if the encryption is very complicated. Data fuzzification, in addition to encryption, is another option for data security, albeit it has the potential to distort the data. Data security management's primary issue is to optimise the performance of IDA in the context of data security.

4.3. Cost Management

The IDA is also dealing with the difficulty of keeping costs down. The expense of improving IDA performance must be kept under control in order to ensure long-term sustainability. It is possible to increase the value derived from data by using diverse life cycle management techniques, such as distributed data analysis, complicated data encryption, and a range of other life cycle management alternatives, all while raising the costs of hardware and network transmission. The major purpose of cost management is to lower both explicit and implicit costs as long as the value offered is adequate for a specific application and a balance between value and cost can be established.

4.4. Data Collection

Multi-source and heterogeneous big data features have emerged in addition to their rising size. Integrating and pre-processing huge multi-source heterogeneous data sets is a challenge in data collection.

- Multi-source Heterogeneous Fusion

The great variety of data sources accessible in a big data context results in data heterogeneity.. The data may also include additional elements like as text, images, audio, and other electrical impulses. A new big data framework is essential to integrating all of the heterogeneous multi-mode data into a format that IDA can process.

- Pre-processing of Messy Data

Noise and redundancy in raw data may have a significant impact on IDA's performance, accuracy, and resiliency [15]. We can't avoid preparing the data in some way. There is a problem, however, with typical data preparation solutions in a large data context. When dealing with noisy and redundant data, the primary goal of pre-processing is to rapidly execute feature selection to identify the most important characteristics so that real-time changing analysis requirements may be followed in a timely manner.

4.5. Analysis of Data

Many IDA approaches, including as feature selection for imbalanced datasets, distributed analysis of data, and big data modelling, have made great progress in data analysis, but there are still many challenges to be faced.

- Imbalanced Feature Selection

Dataset Data preparation may result in an unbalanced dataset, even if the most significant characteristics are picked from the multi-source heterogeneous dataset. Because traditional IDA algorithms focus on broad generalisation, the unbalanced dataset's minority are ignored and treated as random noise. However, in other situations, such as fault diagnosis, the knowledge contained in these outliers may be quite significant. The unbalanced dataset necessitates the development of feature selection methods that are tailored to the problem at hand.

4.6. Application Pattern

Traditional IDA use cases tend to follow a set pattern. Patterns for cross-platform applications.

Data sharing standards and complicated data visualisations face additional problems in the big data age.

- Data Exchange Standard

As IDA's application pattern evolves, so does the amount of data that may move across platforms. For cross-platform IDA applications, it's difficult to unify the many storage formats and data structures which are employed in the data interchange. Standardizing data storage formats and systems will need further funding. Data interchange standards exist in certain industries, such as RosettaNet, but they haven't gained much traction because of issues with universality and usability. Therefore, in the big data era, developing an exchange standard with high pertinency, universality, and usability will become a major challenge.

- Visualization

Complicated Data In a decision support system, data visualisation may provide a straightforward and user-friendly man-machine interface (DSS). In a big data context, correlations between data grow increasingly complicated because of the rising size, dimensionality, data sources, and heterogeneity of the data. The decision maker may be better able to understand the IDA findings and make informed judgments if the data is shown. However, the ability to easily understand complicated data via the use of IDA tools may help them become more well known.

5. CONCLUSION

Data analysis is an iterative and interactive process that includes problem conceptualization, data quality assurance, model construction, and interpreting and post-processing of the results..Using intelligent data analysis, this article looked at the underlying concerns and challenges. Issues in real-world applications provide the most difficult difficulties, and the best solutions are those that focus on solving those problems. Data models and lDAs should be tailored to specific applications in order to get the most value and actionable insights from them. IDA researchers should interact more with industry and mix actual applications and theoretical research in order to solve the issues that may arise in the next years.

REFERENCES

1. R. Nayak, Data Mining for Web-Enabled Electronic Business Applications, to be published in Architectural Issues of Web-Enabled Electronic Business, Shi Nansi Ed., Idea Publishing Group, April 2002.
2. Zhang, Wei, and Feng Gao, "An Improvement to Naive Bayes for Text Classification," Procedia Engineering, vol. 15, pp. 2160-2164, 2011.
3. Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A., and Strachan, R., "Hybrid decision tree and naive Bayes classifiers for multi-class classification tasks," Expert Systems with Applications, vol. 41(4), pp. 1937-1946,2014.
4. Liangxiao Jiang, Zhihua Cai, Harry Zhang, and Dianhong Wang, "Not so greedy: Randomly Selected Naive Bayes," Expert Systems with Applications, vol. 39(12), pp. 11022-11028,2012.
5. Keshavarz, M., and Huang, B., "Bayesian and Expectation Maximization methods for multivariate change point detection," Computers & Chemical Engineering, vol. 60, pp. 339-353,2014.
6. Mahmoud, M. S., and Khalid, H. M., "Expectation maximization approach to data-based fault diagnostics," Information Sciences, vol. 235, pp. 80-96,2013.
7. Frolov, A. A., Husek, D., and Polyakov, P. Y., "Two Expectation-Maximization algorithms for Boolean Factor Analysis," Neurocomputing, vol. 130, pp. 83-97,2013
8. Zheng, B., Yoon, S. W., and Lam, S. S., "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," Expert Systems with Applications, vol. 41(4), pp. 1476-1482,2014.
9. M.e.Naldi, and RJ.G.B. Campello, "Evolutionary k-means for distributed data sets," Neurocomputing, vol. 127, pp. 30-42,2014.

10. Lin, e. H., Chen, e. e., Lee, H. L., and Liao, 1. R., "Fast K-means algorithm based on a level histogram for image retrieval," Expert Systems with Applications, vol. 41(7), pp. 3276-3283,2014.
11. P. Cabena, P. Hadjinian, R. Stadler, J. Verhees& A. Zanasi, Discovering Data Mining from Concept to Implementation, Prentice Hall PTR, 1997.
12. J. Han & M. Kamber, Mastering Data Mining, San Francisco: Morgan Kaufmann, 2001.
13. R. Agrawal & R. Srikant, Fast Algorithms for Mining Association Rules, IBM Research Report RJ9839, IBM Almaden Research Center, 1994.
14. Kambatla, K., Kollias, G., Kumar, V., and Grama, A., "Trends in big data analytics," Journal of Parallel and Distributed Computing, in press.
15. Kwon, 0., and Sim, 1. M., "Effects of data set features on the performances of classification algorithms," Expert Systems with Applications, vol. 40(5), pp. 1847-1857,2013.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

A Review on Big Data: Security Challenges

A Sangeerani Devi[1*]

[1]*Assistant Professor, Department of Computer Science and Engineering , Sri Sairam Engineering College*

Abstract

Big data's widespread use has led to rapid growth of data resources, and new data analysis methods like conventional data mining and statistical analysis are helping to fuel this growth. In the world of big data, data from a variety of sources may be analysed, combined, and used in a variety of ways, resulting in new insights. Although each step of the life cycle offers data security and dependability concerns, the protection of personally identifiable information is a vital goal. Big data analytics, in particular, may be used to analyse user preferences, and this information can lead to the violation of personal privacy. The scope of big data is examined, as is the state of the art in big data security research. The concerns and causes influencing security are laid forth. The authors also touch on and expound on methods that protect individual privacy.

Keywords: large data; cycle of life; security of big data; privacy.

1. INTRODUCTION

There is a lot of interest in big data these days in business, science/technology/media and various government agencies. Healthcare, medicine, government agencies, distribution, marketing, and manufacturing are just a few of the industries used by many countries to turning big data in order to improve their services. It uses information-based technology to analyse vast amounts of data and predict future changes based on the information gained. Economic progress and technical advancement are both aided by this new source of energy. Big data is driven by a wide range of commercial and political objectives, including data integration, analysis, and mining. This is especially true when it comes to structured large data that comes from a wide range of sources, such as social media platforms, websites,

[*] ISBN No. 978-81-955340-6-7

and global positioning systems. Data mining and statistical analysis approaches, such as standardised data mining, are propelling the growth of the market for big data because of their ability to mine large amounts of data quickly and efficiently. As a result of the data life cycle, which includes collection, analysis, fusion, and application, new information may be gleaned from big data.

2. ISSUES IN BIG DATA RELATED TO SECURITY AND PRIVACY

In light of Section II, discussion of the security and privacy implications of big data, this section takes a closer look at some of the most prevalent security and privacy concerns. There are a few conceptual and operational taxonomies of security and privacy provided by [1] and [2] to create vulnerabilities in big data systems. However, in order to establish a causal link between the properties of big data and vulnerabilities, we suggest and develop the following category based on the subjects and challenges in study disciplines, as shown in Fig.1:

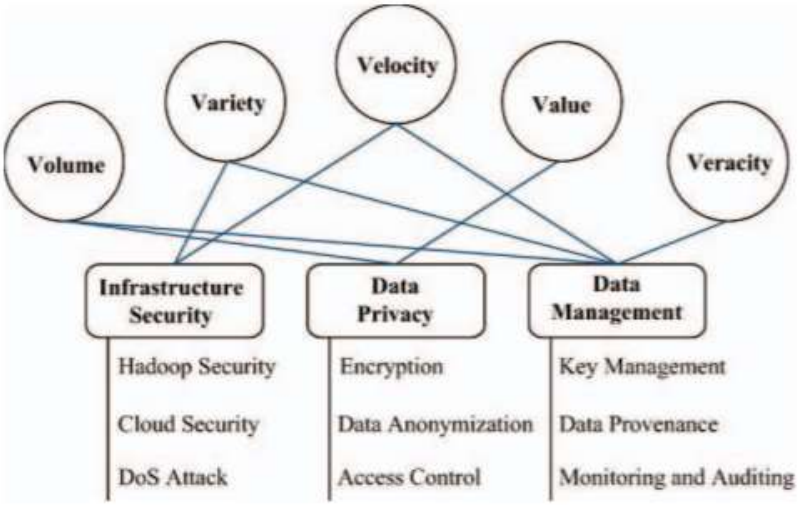

Figure 1. Category of Security Challenges in Big Data.

- Infrastructure Security
- Privacy of Data
- Data Management

There have been a lot of prior studies on the big data's security and privacy . The security and privacy issue may be better understood from several angles, as shown in Fig. 2.

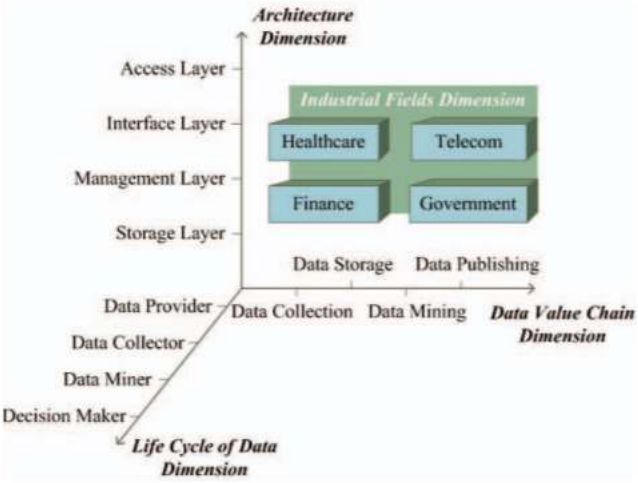

Figure. 2. Perspectives of Big Data Security and Privacy Analytics

3. ISSUES IN BIG DATA

Existing security for large data security and privacy are explored in this section. The researchers have devised a set of theoretical and operational privacy and security categories to better understand the threats presented by big data. At each of the four levels of a big data system, security and privacy are necessary, such as at the storage layer for the secure storage and control of monitoring devices. Hadoop Distributed File System (HDFS), data encryption, and so on [3] are included in the second management layer. There are three layers in the interface layer: an application programming interface (API), identity verification, and access control. Lastly, the access layer contains the user's cybersecurity. Big data presents a new set of challenges, as well as new opportunities, for organisations. This poll examines these challenges and opportunities. Security and privacy issues are examined and explored in detail. Furthermore, a study found that data sources, storage, and output all need some kind of security. The protection of the data types listed above will go a long way toward ensuring the safety of large amounts of data.

A variety of data encryption, access rights, transport layer security, and firewalls may be easily hacked. As a result of these factors, new technologies and strategies are being created to protect large data. The following portion of the study goes into depth into the Fig. 3 shown elements.

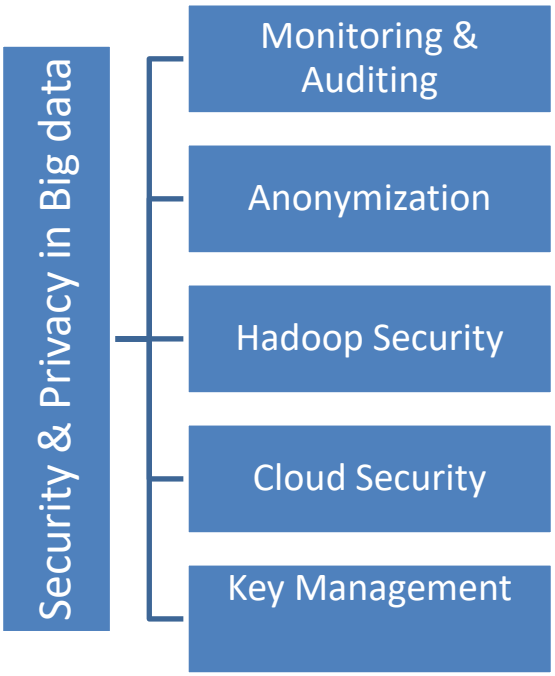

Fig. 3 Security and Privacy Categorization

A. Security in Hadoop

As one of the distributed process frameworks that isn't built for security, Hadoop is often used as an example. Big data analytics with Hadoop necessitates the use of a secure platform. In order to prevent hackers from stealing data from the cloud, two methods were recommended. HDFS has a trust mechanism that establishes a connection between the user and the name node. Authentication to the name node is required as part of the procedure. Users and name nodes create the hash function jointly. A comparison of the hash functions is made. If the two are identical, the user is granted access to the massive database. SHA-256, a hashing algorithm, is used for authentication in this method. The security of HDFS is of great importance. As a result, three strategies for enhancing security have been devised. The Kerberos system, which relies on a Service Ticket for security, is the first option. The algorithm known as the Bull Eye algorithm [4] is used to monitor sensor data and sensitive information in the second way. Replicated data and original data may be managed by this algorithm's primary benefit. HDFS security is further enhanced by using the Master Slave technique. There are two nodes in HDFS that act as a name node: the master and slave nodes. Using the Name Node Security Enhancement (NNSE) authorization, a slave node may respond and respond to an issue in the master node.

B. Monitoring as well as Auditing

For network security, detection of intrusions is an essential aim [5]. In order to identify intrusions on the whole network, DNS, HTTP, and other monitoring systems were created. Utilizing correlation

techniques, the scattered data is collected and analysed. To assess whether a node, flow, or packet has been malicious, the matrix has been set up. The detection system receives an alert message if any of these are detected or if the procedure is terminated by the prevention system. [5] Data availability, integrity, consistency, aggregation, and confidentiality are all factors that contribute to a security hole in huge data. There is a demand for security solutions to cover this gap in the market.

C. Security on Cloud Platform

Some of the reasons of cloud computing is extensively used because it provides on-demand services and resources that may be pooled. Despite the possibility of an attack, the cloud architecture's hosts are impervious to it. As a result, the cloud architecture service provider must take preventive measures. The cloud platform uses a variety of security measures, including authentication, compression, encryption, and decryption, to protect massive data. The cloud platform uses a security approach called Cryptographic Virtual Mapping to create a data route. Protection of vital, sensible and valuable pieces of information is the goal of this method. In order to protect sensitive, valuable, and vital data, the encryption is applied just to the storage channel that leads to it. Data parts and their accessing indexes must be available at all times in order to attain a factor of availability. This means that even if some data is lost, the overall availability of the system may be considered a success.

D. Anonymization

The volume characteristics of big data make it impossible for any of the traditional approaches to guarantee anonymity, despite the best efforts of researchers. In order to maximise scalability and privacy, a hybrid technique to anonymization has been developed that combines the traditional approaches of Bottom-Up and Top-Down. The t-ancestor clustering approach and a proximity-aware agglomerative algorithm are used to partition the dataset and record data, respectively, to overcome the issue of scalability [5]. Big data may be protected by using a differentiated privacy method. The model is constructed in such a manner that each piece of input data has an equal chance of being released. One of the two processes for guaranteeing differential privacy in large data is known as the Exponential mechanism and the Laplace mechanism. The Laplace mechanism is used to generate noise based on Laplace distribution for current and accurate results. The Exponential mechanism rewards outputs with higher scores with exponentially larger chances of success when the results are fictitious. This makes it more likely that you'll be referred to as [5]. Discriminating between private and public data is an important consideration in any research, even though it has been proved to be beneficial for trajectory data. As a result, it has been shown that all data sources, stored data, and output data need to be safeguarded [5]. The protection of the data types listed above will go a long way toward ensuring the safety of large amounts of data. In order to keep up with the ever-changing security landscape, methods like machine learning and statistical analysis, generally referred to as "data science," are often used. Analyses are performed depending on kind of the data present in the ecosystem, and a variety of machinery learning methods are used to spot any kind of alterations [6]. In order to safeguard big data environment, firms already utilise a variety of security techniques. Network device configuration may be managed using the IBM risk manager tool, which can be used to report and manage risks [7].

E. Key Management

A problem with large data security arises when keys are exchanged between the server and the user. A group key transfer mechanism is needed to distribute a key across numerous groups. As a result, a new protocol based on the Diffie-Hellman key agreement and a linear secret key method is implemented to guard against online attackers. As part of a collection of complicated networks, the Outsourcing Conditional Proxy Re-Encryption (CPRE) is used. Unstructured data, such as email, text, and XML, is also challenging to secure in large data systems [8]. Unstructured data may be protected by the use of data analytics techniques such as data filtering, grouping, and classification based on the degree of data sensitivity. In the second level, database's node of data is arranged, and the relevant and significant service (identity, integrity confidentiality, nonrepudiation, and authentication) is picked by a scheduling algorithm from the security suite to offer security.

4. CONCLUSION

Predicting future trends has become one of the most promising and dominant technologies because to the rise of big data. The privacy and the security should be taken into account when developing apps in these situations. When it comes to sensitive applications in big data, privacy and security must take precedence over anything else. This article examines the effects of big data features on infrastructure security and privacy, as well as cloud security and data management, on the privacy and the security of the big data systems. Data collection, storage, use, transport, and analysis all face unique challenges in the modern world. Security and privacy of Big data are our primary concerns. It's imperative that we create a framework for safe data sharing on a semi-trusted platform for big data that includes secure data transmission, storage, use, and destruction.

REFERENCES

1. Cloud Security Alliance Big Data Working Group, Expanded Top Ten Big Data Security and Privacy Chanllenges[R], Apr. 2013
2. NIST, NIST Big Data Interoperability Framework: Volume 4, Security and Privacy[R], National Institute for Standards and Technology, 2015, http://dx.doi.org/10.6028/NIST.SP.1500-4.
3. F. McSherry and K. Talwar, “Mechanism design via differential privacy,”in 48th Annual IEEE Symposium on Four.
4. D. S. Terzi, R. Terzi, and S. Sagiroglu, “A survey on security and privacy issues in big data,” in2015 10th International Conference for InternetTechnology and Secured Transactions (ICITST). IEEE, 2015, pp. 202–207.

5. F. McSherry and K. Talwar, “Mechanism design via differential privacy,”in 48th Annual IEEE Symposium on Foundations of Computer Science(FOCS’07). IEEE, 2007, pp. 94–103.
6. C. Thota, G. Manogaran, D. Lopez, and R. Sundarasekar, “Architecture for big data storage in different cloud deployment models,” inResearch Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing. IGI Global, 2021, pp. 178–208.
7. R. Bhatia and M. Sood, “Security of big data: A review,” in 2018 Fifth International Conference on Parallel, Distributed and Grid Computing(PDGC). IEEE, 2018, pp. 182–186.
8. D. S. Terzi, R. Terzi, and S. Sagiroglu, “A survey on security and privacy issues in big data,” in2015 10th International Conference for InternetTechnology and Secured Transactions (ICITST). IEEE, 2015, pp. 202–207.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

Review On Data Security Technology Based on Cloud Storage

Dr. Surender Kumar[1*]

[1]*Head/Assistant Professor, P.G. Department of Computer Science, Sri Guru Teg Bahadur Khalsa College, Sri Anandpur Sahib, (An Autonomous College) Punjab (India), drsurender.sgtb@gmail.com*

Abstract

Data security has become a major concern as cloud storage systems have been more widely used in a variety of complicated environments. It is possible that data may be incomplete due to node failures or other external intrusions, but it is also possible that cloud service providers actively hide or other circumstances make it difficult for the user to be aware of the change.

Approaching the issue of security with more prudence is necessary There have been several techniques created to safeguard files and other information as computer and communication technologies have evolved. The term "computer network security" refers to a combination of tools, processes, rules, and solutions that are used to prevent and respond to assaults on a network. All of these ideas must be defined and learned in order to properly assess an organization's security posture. Some of the methods and dangers to the network and computers are discussed in this document, along with possible programmes.

Keywords: Big Data, Cloud Computing, Data Security, Technical Analysis.

1. INTRODUCTION

Many firms consider data security to be a top priority. For cloud users, this means first identifying the data objects that need to be safeguarded, classifying the data according to its security implications, and defining the data protection policy and enforcement procedures. When it comes to most cloud-based applications, data objects would include not just bulky data stored in the cloud (e.g., a user database

* ISBN No. 978-81-955340-6-7

and/or a filesystem), but also data in transit between the cloud and the user(s) (In many circumstances, it would be more cost-effective and convenient to move large volumes of data to the cloud by mobile media like archive tapes than transmitting over the Internet.). It's possible that additional types of application data, such as user IDs or service audit data generated by the auditing model or service profiles or transient runtime data generated by the service instances, may also be included in data objects. The security implications for cloud users will vary according to the value of the data and the kind of data that is being stored in the cloud. Such as a user database at rest on the cloud servers, cloud users need robust security to ensure the confidentiality of their data as well as the integrity and availability of their information. The privacy of a user's identification information might be jeopardised if it contains any personally identifiable information (PII). Therefore, only authorised users should have access to the user's identification data. Data from service audits give proof of compliance and SLA fulfilment and should not be intentionally distorted. In order to prevent attackers from finding and identifying service instances, attackers need to keep their hands off the service profile information. During runtime, temporary runtime data should be separated from vital user business data and safely removed.

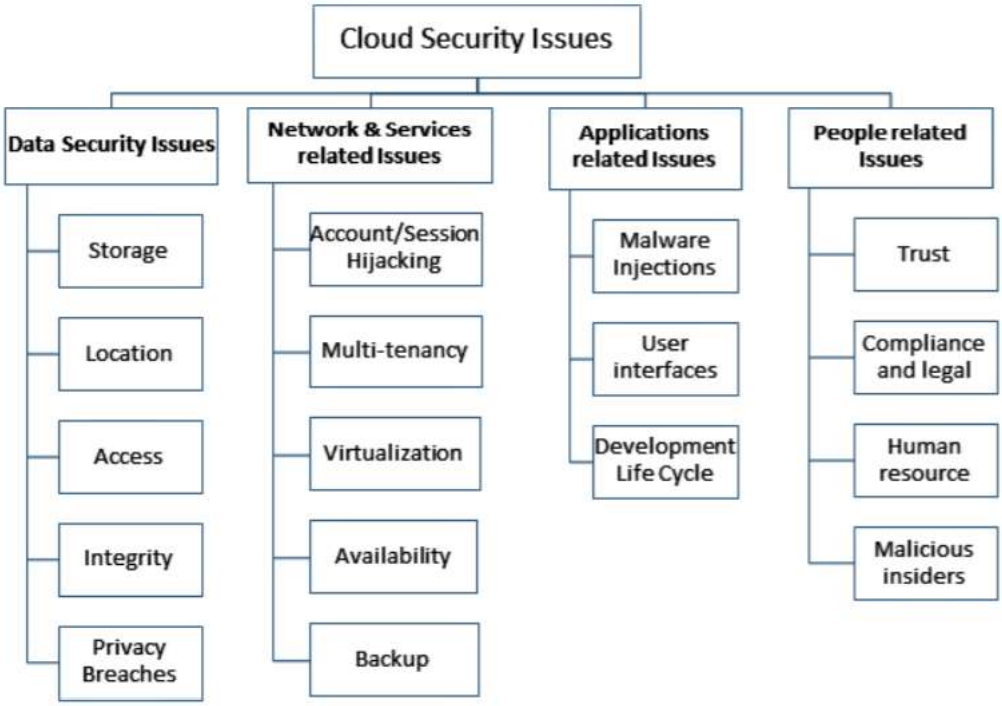

Figure 1. Summary of security issues in each category of Cloud Computing

2. Data Security Technology in Big Data Cloud Computing Environment

Improved data security in a cloud computing environment is dependent on thorough study and analysis of different data security technologies, which must be based on an understanding of cloud computing's huge data security challenges in order to increase data security protection technology's efficiency. It is now possible to significantly increase data security in the cloud computing environment by using the following data security protection technologies:

a) Data Encryption Security Technology

China's economic growth may be considerably accelerated if the full potential of cloud computing is used. However, there are a number of security issues that will impair the cloud platform's application effect in the big data cloud computing environment. Data leakage and theft have become an issue with the implementation of cloud storage and transmission. Big data security privacy must be examined more thoroughly in the research of data security technology, and attention must be paid to the successful implementation of data encryption security technology. The following elements of data encryption technology may be used to verify its efficacy as a security measure. In order to protect personal information, some users may encrypt and process it using the cloud platform's storage service. To a certain degree, this may increase the level of security of information. Cloud computing platforms, on the other hand, may encounter mistakes in data analysis as a result of the framework's inherent irrationality, reducing the effectiveness of encryption processing. Data encryption is a major concern for cloud computing workers, therefore they must be aware of it. The integrity of data must also be taken into consideration while transmitting data via the internet. Because it is necessary to adhere to the applicable criteria while transmitting data. When uploading data, however, the integrity of the data may be compromised. Users' ability to make effective use of the data will be severely hampered if it is not uploaded in its entirety. In addition, it will have an influence on the performance of cloud service platform applications. Third, while calculating data, we should be mindful of safeguarding user privacy. Data computing information and outcomes must be monitored by relevant departments, and the cloud computing platform's security must be regularly improved. Data leakage may be reduced by doing this. Data encryption technologies must be bolstered in order to secure the safety and privacy of information.

One of the large data cloud computing environment's data security methods is data encryption. This data protection technology's primary goal is to keep personal information private and secure. In order to avoid different data security difficulties, the system platform must properly optimise the data encryption technology. Traditional information encryption technology and innovative cloud server configurations are often used in data encryption processing. There must be a download of data to the local before encryption and combination work can be done using typical encryption techniques. To complete the encryption process, the cloud server setting technology may employ cloud server operations to establish relevant keywords. The old technique of data encryption is difficult to use and has a poor fault tolerance rate. It is possible that the new data encryption approach will significantly enhance the effectiveness of data search operations. However, we must be aware that data security cannot be completely ensured

during the use of the new encryption technology. Because of this, we must constantly optimise and develop data encryption technology in order to limit the frequency of sensitive information and increase data security [1].

b) Data Access Security Technology

Big data access security privacy protection technology and data destruction technology are two of the most important components of data access security protection technology. To begin, there is the issue of data security and privacy protection for large datasets. Cloud computing is required for the application process in order to store large amounts of huge data, which may be classified into public and private clouds. As one of them, the public cloud has a big storage capacity and is accessible to everyone. There is a large number of internal data resources that may be accessed via public cloud services because of the data's great openness. In most cases, private clouds are built on top of existing business infrastructure. There is a certain amount of privacy with private clouds. During the course of company growth and practical implementations, private data will be generated. This data must be stored and protected in order to guarantee that the company's regular and steady growth continues. Because of its uniqueness, a private cloud may make use of more sophisticated technology to safeguard its data resources. However, the expense of using private cloud data access security protection technologies must also be taken into account. If you want to assure data security and decrease costs, most firms will employ a combination of public/private cloud solutions. Second, the use of data erasure tools. Data collection, data administration, and data storage all play a role in the big data operation. Big data's destruction connection is also one of its most important linkages. The application value of the data will be directly impacted by the data deletion procedure. The research on data destruction technology must be strengthened to guarantee the soundness and comprehensiveness of data screening techniques in order to assure prompt and effective data destruction and avoid destroyed data surviving or being leaked. Furthermore, we must guarantee that data can be erased swiftly and properly to avoid harmful use of data and harm to the interests of businesses [2].

c) Data Sharing Security Technology

Data sharing security protection technology is an essential technology type for enhancing data transmission security. The following three components make up the bulk of shared encryption technology. First and foremost, cloud server encryption. The cloud server encryption technique may utilise public key encryption to retransmit the data to the cloud once it has been downloaded. The efficiency of this method of operation is poor, and it's a lot more difficult to use. Secondly, proxy re-encryption. To complete the data transmission procedure, this encryption technology must be sent from the authorised person to the agent, and finally to the accepting talent. In order to transmit data, there must be a number of steps. As a result, data exchange has a poor effectiveness. Third, conditional re-encryption through a proxy proxy is a possibility. Today's data sharing security protection technology is mostly used

in this manner. The proxy re-encryption approach necessitates the addition of encryption execution conditions throughout the application process. Using this method, not only can you better organise your data and make it easier to share, but it may also increase the safety of your data in transit. However, in the process of implementing data encryption techniques, it is also important to enable consumers to make a suitable choice of encryption methods based on their specific demands. This is the most effective method of ensuring the safety of sensitive data.

3. Security Techniques For Securing Cloud

Cloud data encryption is not a solution for data that has trust in the cloud's ability to protect it. Authentication and identity management, encryption, integrity checks and data masking may all be applied to cloud data using currently available security mechanisms. Here, security approaches are explained in Figure 1.

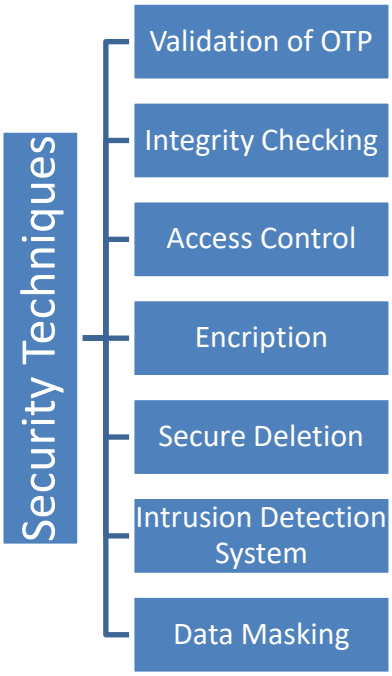

Figure 1 Security Techniques for Securing Cloud

- **Validation of OTP**

To authenticate a cloud user's identity, many banks now employ the One Time Password (OTP) approach, which generates an OTP using a random number generator and is often referred to as system factor authentication (see Figure 2). When used in conjunction with a second authentication factor, it is known as a Multiple Authentication Factor (MAF).

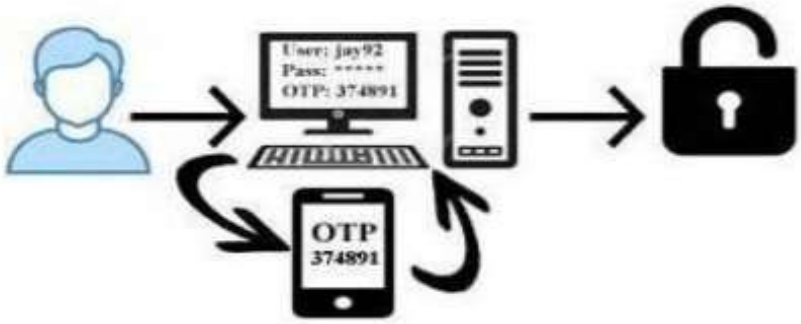

Figure 2 OTP Authentication

- **Access Control**

An authorised cloud data user may see and modify their own data, but an unauthorised cloud data user cannot do so because of access control measures in place at both the cloud service provider and the cloud data owner.

- **Integrity Checking**

Cloud data integrity is an assurance that only authorised users may update or access cloud data. Simple cloud-based data verification assures the integrity of the data and that it has not been altered in any way. As a result of PDP and POR, it is possible to secure the integrity of cloud data on a distant server while verifying that the proof that cloud data is saved by the user on the server has not been altered [3].

- **Encryption**

Using cloud storage security encrypts all of your data before it travels from your local computer to the cloud, making it virtually impossible for anyone else to read your private information. Only an authorised user with access to your decryption key can decrypt your encrypted files, making it essential to keep encrypted data separate from the encryption key.

The method of encrypting and decrypting is shown in Figure 3.

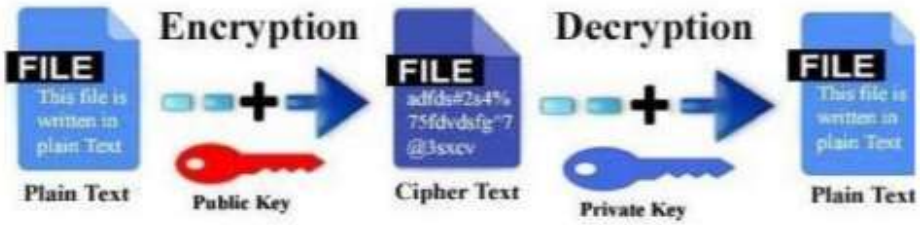

Figure 3 Data Encryption and Decryption

- **Secure Deletion**

It's critical to know how the data on the server gets purged. In this strategy, we destroy the media before it is reused and at the same time give protection for accepting the data that was on the media before to deletion. This sort of data is often transmitted at a lower level of categorization since the security for accepting earlier data is not supplied [4].

- **Intrusion Detection System**

To put it simply, an intrusion detection system (IDS) monitors system activity and network traffic in order to look for any indications of illicit activity. In the modern day, most hackers use a variety of various methods to get access to private information. Any unauthorised or harmful use of IT resources is considered an incursion. It is the goal of the invaders, who want to get access to sensitive information, to wreak damage. An Intrusion Detection System (IDS) may be divided into two categories: a network-based IDS (NIDS) that monitors network traffic and keeps an eye on ongoing assaults, as well as a Host-based IDS (HIDS), which is placed on a single system or server and monitors unlawful activity on that system.

- **Data Masking**

Data masking is a method for protecting cloud data from unauthorised access and theft while also ensuring that the data is replaced with fictitious but plausible data. De-recognition, cleansing, and comprehending the phrase are all terms that are used interchangeably to describe the same muddled process. Not only is data masking an algorithm, but it is also a collection of publicly available data. Static Data Masking (SDM) is utilised by most businesses when developing tests, and this is the only form of masking that is available when utilising outsourced developers in a separate firm or location to create tests. Duplicating the database is the only option in these situations. Based on the user's function in the organisation, Dynamic Data Masking (DDM) allows access to certain information.

4. CONCLUSION:

In the recent decade, businesses, companies, and hackers have all benefited from the widespread use of cloud computing technologies. Cloud computing security has been threatened by the rise of current cloud architectures and high-speed internet with new developments. One benefit of moving to a cloud-based system was that it allowed a company's capacity to adapt and grow with the ever-changing industrial landscape. For a variety of reasons, this rendered their data less secure and more prone to attack. There is a need to focus more on research and use of different data security protection solutions in the cloud computing environment because of its unique properties. Only by improving data transmission efficiency can the security and reliability of data transmission be enhanced.

REFERENCES

1. Zhang Qian, Yang Huibi. Exploration of big data security and privacy protection under cloud computing[J]. Science Popular (Science Education), 2017, 000(010):192-192.
2. Yuan Huihua. Research on Data Security in Big Data Cloud Computing Environment[J]. Information Technology and Informatization, 2019.
3. S. Sharma, "Data Integrity Challenges in Cloud Computing", 4 th international conference on recent innovations in science engineering and management, pp. 736-7436, 2016.
4. DIGITAL GUARDIAN [online] https://digitalguardian.com/blog/what-cloud-encryption (Accessed 25 December 2019).
5. G.K. Ravikumar "Design of Data Masking Architecture and Analysis of Data Masking Techniques for Testing", International journal of engineering science and Technology, vol. 3, no. 6, pp. 5150-5159, 2011.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

Mobile Cloud Computing Applications survey: Security and Privacy

Kumar Rahul[1*], Rohitash Kumar Banyal[2]

[1]*Assistant Professor, Department of Basic and Applied Science, NIFTEM, Sonipat 131028*

[2]*Associate Professor, Department of Computer Science and Engineering, Rajasthan Technical University, Kota 324010*

Abstract

Mobile cloud computing, a novel approach for mobile computing that has been in development for decades, results from the combination of powerful mobile devices and cloud computing services. The MCC enables cloud computing to be integrated into the mobile environment and solves performance issues. As a result, The Cloud Technology is able to transcend the limits of mobile computing. For mobile apps, cloud technology offers adequate computational capacity to operate on a cloud platform. By offloading applications to a resource-rich Remote server, the novel technology known as mobile cloud computing (MCC) is introduced to overcome the limitations of mobile devices (such as battery life, storage capacity, processing capacity). MCC integrates two technologies (Mobile Computing, Cloud Computing). MCC is described in this study, as well as security and privacy issues, threats, and obstacles that are associated with MCC.

Keywords: Computing of Mobile, Cloud Computing, Mobile Cloud Computing, security.

1. INTRODUCTION

Mobile gadgets have played a significant influence in our contemporary and virtual lifestyles during the last couple decades. For example, according to an International Data Corporation (IDC) report in 2016, the use of tablets and mobile devices grew by 1.6 billion units [1]. For a variety of reasons, mobile apps have become more popular over the last several years across a wide range of industries. Apps may

* ISBN No. 978-81-955340-6-7

be downloaded from Android play store and Apple iTunes, regardless of location, thanks to mobile computing. Despite the high-end capabilities of mobile cloud computing for executing a variety of real-time apps, consumers continue to ask for greater processing power. The battery life, storage capacity, processing power, and connectivity capabilities of mobile computing devices are all constrained.

In its most basic form, mobile cloud computing (MCC) refers to a system in which data storage and processing take place outside to mobile device. As an alternative, MCC may be described as a mix of the mobile web and cloud computing, which is the most common method to access apps for mobile users and the Internet services.

2. MOBILE CLOUD COMPUTING

Computers and other devices are given with shared resources and information (such the electrical grid) across a network as a service rather than a product, and this is known as Cloud Computing (typically the Internet). An end-awareness users of the physical location and system setup is not required in order to utilise this service. Computing, communication, and storage resources which are shared in a virtualized and isolated environment by many users that are the primary focus of most of the cloud research.

- There are several common obstacles for mobile devices (battery life, storage, bandwidth etc.)
- Using cloud computing infrastructure, platforms, and applications is a low-cost and scalable way for consumers to benefit from cloud computing's benefits.
- Since all resource-intensive calculation may be performed in the cloud, a strongest device configuration (e.g., CPU performance, RAM capacity, etc.) is not required for mobile cloud computing.

For example, base transceiver stations (BTS, access points or satellites) create and regulate the connections (air links and functional interfaces) between the mobile networks and mobile devices shown in Figure 1. Central processors linked to servers delivering mobile network services receive mobile user requests and information (e.g., ID and location).

Using the home agent (HA) and subscriber data stored in the database, mobile network operators may deliver mobile users services like AAA (authentication, authorization, and accounting). Subscriber requests are then sent to a cloud computing facility via the internet. Requests for cloud services are processed by cloud controllers in the cloud, which then supply the services to mobile users. The utility computing, virtualization, and service-oriented architecture principles are used to create these services (e.g., web, application, and database servers)

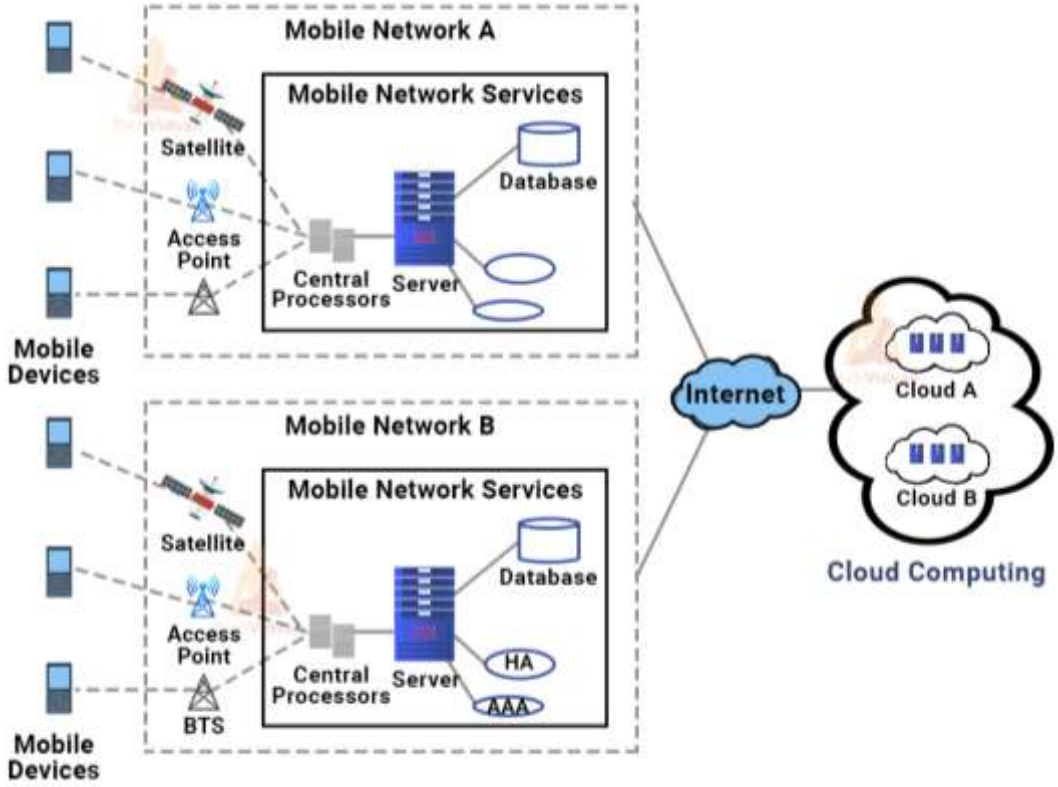

Figure 1: Mobile Cloud Architecture

3. MOBILE CLOUD COMPUTING SERVICE MODELS

Similar to CC architecture, these services are utilised to develop the MCC's infrastructure. [2]:used to build CC architecture [2]:

- *Mobile Cloud Infrastructure-as-a-Service (MIaaS)*—In a pay-as-you-use model, computation, storage, network components, and devices are supplied, managed, and returned in response to a customer's request in a pay-as-you-use model. There are two ways to set up the mobile cloud's infrastructure:
- In the pay-as-you-go model, clients are offered mobile cloud infrastructure and resources in accordance with value of their payment, where compute, storage resource and network components are provisioned, maintained & returned to clients according to their request.
- *Mobile Network-as-a-Service (MNaaS)*—As a result, the vendor may react to on-demand requests from customers and dynamically create, implement, and design a wireless network infrastructure for mobile connection, to cloud infrastructure. Because of its scalability and flexibility, the startup costs are inexpensive. Consider OpenStack, Google App Engine, and

Microsoft Azure, just to name a few examples.

- *Mobile Platform-as-a-Service (MPaaS)*—App hosting, development, validation, and deploymenttools are all supported by this service. App Mobi, for example, is a tool that aids in the creation, distribution, and validation of mobile applications (as easy as possible).
- *Mobile Software-as-a-Service (MSaaS)*—This kind of software is delivered to customers in form of Mobile Software-as-a-Service (MSasS). Thin mobile client-based Internet connection is used in this paradigm to enable to access mobile application services (deployed and operated in the cloud) to the mobile users.

The following are the main categories into which Mobile Cloud Computing applications may be broken down:

- Sensing capability: It is possible to use a smartphone as a sensor. Some of the properties that sensors are capable of monitoring are humidity,temperature andblood pressure. Sensor data may be uploaded to cloud at a further time. Information stored in the cloud is available to a global audience.
- With mobile cloud, users may pick what information they want to keep private and what information they want to make available to the public.
- Reliability and Data storage: Cloud storage preserves and backs up users' data if there is a problem with their device.
- Personal information security: When a storage device fails, users' data is safely stored in the mobile cloud, where it is automatically restored.
- In the cloud, virtual computers and a secure search engine work together to increase the security of user data.
- Health monitoring: Sensors may be used to store and send personal health information to the cloud through mobile devices. To help people maintain their health, health care providers may access such information in the cloud and provide them advice. For health monitoring, mobile devices may be utilised as sensors, too.
- Sensing as a service: The mobile cloud offers platform,infrastructure, and software as a service for sensing purposes. When it comes to using several programmes, compatibility isn't a problem.

4. MOBILE CLOUD COMPUTING APPLICATION

The worldwide market for mobile apps is growing at an accelerating rate. MCC has been included into a variety of mobile apps. Mobile commerce, mobile learning,mobile healthcare,mobile social networking, mobile sensing, multimedia sharing, mobile gaming, location-based mobile service, and augmented reality are some of the current examples of MCC applications. As a result of simultaneous user access and transaction processing, mobile commerce such as e-banking, e-advertising, and e-

shopping makes advantage of scalable processing capacity and security mechanisms. Security and administrative controls are provided to guarantee that smartphone users have the privileges and access permissions essential for secure viewing and sharing of multimedia material. Learning resources on the cloud may be accessed at any time and from any location thanks to mobile learning. Many MCC applications, including healthcare, social networking, and environmental/health monitoring, will be revolutionised by mobile sensing using sensor-equipped cellphones. Massive amounts of patient data may now be saved in cloud in real time thanks to mobile healthcare. Scalability in mobile gaming is achieved by the use of scalable compute in the cloud and real-time data updates on the mobile device. It is possible to upload audio/video/multimedia material for real-time sharing through mobile social networking, with cloud computing offering not only storage but also security to safeguard the confidentiality and data integrity.

5. ADVANTAGES AND DISADVANTAGES OF MCC

Some benefits of Mobile Cloud Computing are [3]:

1) Extending Battery Lifetime—One of the most significant considerations to every mobile device maker is the battery's lifespan. Many approaches to reducing power usage have been put out so far, such as optimising CPU performance or adjusting screen brightness. Improvements to the mobile infrastructure (hardware) are required to make these modifications viable, however these changes may have an impact on manufacturing costs and may not be achievable on all devices. Computing offloading, a method for moving resource-constrained devices (such as mobile devices) to the cloud, has been presented as a way to address all of these issues. This reduces the amount of energy needed to run apps that take a long time to complete.
2) Improving Storage Capacity—In today's mobile devices, the most common issue is a lack of available storage space. MCC have solved an issue that has plagued the mobile industry for years. Users of mobile devices may now store and access almost unlimited amounts of data in the cloud. Amazon Simple Storage Service [Amazon.com/s3], Flicker, Facebook, and many more apps employ this cloud approach.
3) Reliability—When data backup was put into mobile phones, it improved even more since data that had been lost (even accidentally) could no longer be recovered. It was MCC that revolutionised the storing of data (by users). Cloud storage means that data may now be accessible from any (different) device.
4) Dynamic Provisioning—It has long been an issue in operating systems to have to wait for a resource while a process completes its job. There are numerous scheduling algorithms that may be used to distribute resources to process based on its wait time. Introducing dynamic on-demand resource provisioning as a flexible option to execute programmes without reservations or wait times for resources was MCC's solution to this challenge.

5) Scalability—In order for a service provider to satisfy the needs of its customers on a daily basis, they must be able to adapt quickly and easily. With MCC's flexible resource provisioning, the applications may be extended (add/modify) with little or no limitation on the resources used.

6) Multitenancy—There are several ways to save production costs in software, and multitenancy is one of the most effective (e.g.: network operator, data centre owners, etc.).

Some drawbacks of Mobile Cloud Computing are [4]:

1) Security—For years, the security of user data has been a top priority. Users of mobile devices save critical data on the cloud, which must be safeguarded to prevent severe losses. As a result, one of MCC's drawbacks is the inability to adequately safeguard user data.

2) Performance—Two things that have always been a concern for mobile devices are performance and user experience. When comparing the mobile app to the original, there would be significant reductions and modifications (that we use in desktop). For instance, Chase offers its customers a mobile banking app (to make transactions easy and fast). Things like "Things you can do" are missing from the mobile version of the programme compared to the PC/desktop version, which has more features (that cuts down performance).

3) Connectivity and Latency—In data transport, latency and connectivity play a significant impact. Data transmission may be slowed down or even halted entirely by a single weak signal or weak link. Because weather, signal traffic, and other factors such as these may easily impact bandwidth and signal strength, the service provider should do frequent checks on these parameters.

4) Compatibility—In order for a mobile phone to switch operating systems, it must be able to work with a wide range of devices and apps. It is difficult for a service provider to function in a consistent environment due to the usage of multiple applications for different technologies, which reduces compatibility.

5) Limited Resources— Memory storage, battery, and other variables have weakened mobile devices' ability to receive and analyse information. This is due to the fact that mobile devices have less processing power than a desktop/PC. Because of their resource limitations, mobile devices have become a roadblock in MCC.

6. MOBILE CLOUD COMPUTING ISSUES

Mobile communication issues:

- ✓ Low bandwidth: Low bandwidth is a major problem in mobile communication since radio resources for wireless networks are substantially more limited than those for wired networks.
- ✓ Service availability: Because of traffic jams, network outages, and poor signal strength, customers' mobile devices may have trouble accessing the cloud to utilise a service.
- ✓ Heterogeneity: In order to meet MCC criteria (always-on connection, on-demand scalability, and efficiency of energy), it’s challenging to handle wireless communication with extremely diverse networks.

- Problems with computing: Offloading of computing:
- MCC has one of the most important characteristics.
- Energy savings may not always be achieved by offloading.
- There are a number of considerations to keep in mind when deciding whether and how much of a service code should be offloaded.

7. CONCLUSIONS

Because of the rapid rise in data computing, data processing capacity is increasingly seen as a critical resource in many nations. Mobile cloud computing, which combines the benefits of both cloud computing and mobile, has opened up a number of new research avenues in recent years. The combination of cloud computing with mobile devices in mobile cloud computing is an important method to improving capabilities of real-time applications. Mobile devices are utilised to gather data from any location, while the cloud is used for data processing and delivery. There are several uses for mobile cloud computing. Different MCC applications, such as mobile health monitoring, MMS, MCC in the military sector were explored in this article. Mobile cloud computing, on the other hand, has certain drawbacks, such as the risk of data loss when mobile devices save and retrieve data from the cloud. As a result, we must enhance the level of protection for users who need to authenticate in order to access data.

REFERNCES

1. Song, Weiguang, and XiaolongSu. "Review of mobile cloud computing." 2011 IEEE 3rd International Conference on Communication Software and Networks. IEEE, 2011.
2. Ruay-Shiung-Chang, Jerry Gao, Volker Gruhn, Jingsha He, George Roussos, Wei-Tek Tsai, “Mobile Cloud Computing Research- Issues, Challenges and Needs,” 2013 IEEE Seventh International Symposium on Service-Oriented System Engineering, pp. 442- 453.
3. Hoang T. Dinh, Chonho Lee, Dusit Niyato, Ping Wang, “A survey of mobile cloud computing: Architecture, applications, and approaches,” Wireless Communications and Mobile Computing, vol. 13(18), pp. 1587-1611. DOI: 10.1002/wcm.1203.
4. Pragati Redekar, Rakesh Rajini, “Towards Mobile Cloud Computing,” 2014 International Journal of Advanced Research in Computer Engineering and Technology (IJARCET), vol. 3(10), pp. 3335-3338.

Academic Guru Publication House

Advances in Cloud Computing Security
Techniques and Applications
Volume 1
Year: 2021

AG PH Books

Hybrid cloud computing: Security Aspects and Challenges

Dr. Umakant Bhaskar Gohatre[1*]

[1]*Assistant Professor, Smt. Indira Gandhi College of Engineering, Navi Mumbai*

Abstract

Innovative business results may be achieved via the growth of cloud infrastructures to hybrid cloud models, which are driven by the need for increased Information technology (IT) agility and overall cost-control constraints Hybrid cloud solutions integrate advantages of both public and private cloud infrastructures. Different security risks have been proven to be addressed in order to get the most out of the hybrid cloud approach. The security of hybrid clouds for major corporations and governments is the subject of this article. It explores numerous security types inside the IaaS and SaaS concepts, different authentication and security principles, and security difficulties in this field. In this case, a comparative assessment of several current solutions and the common issue areas and security threads are the focus of this work.

Keywords: Migration; Hybrid cloud; security issues; security techniques.

1. INTRODUCTION

The phrase "Cloud Computing" has gained a lot of traction recently in the IT sector. People on the internet use it a lot, and various writers have given it many diverse interpretations. When it comes to software development, cloud computing is reshaping the industry. Personal and professional elements of life and work are trending toward the idea that everything can be found on the internet. Big online-based corporations like Google and Amazon have come up with a concept called "Cloud Computing," which

* ISBN No. 978-81-955340-6-7

is the sharing of web infrastructure to cope with the storage, scalability, and calculation of Internet data. Using this trend Customer-specific hardware and software services are provided through the internet via the cloud computing concept. Cost and maintenance are reduced by using cloud computing. When it comes to cost-saving IT cloud solutions, many companies are turning to hybrid clouds, which combine advantages of constructing private and public clouds and also using the scalability inherent in their current Information technology (IT) infrastructure. Numerous companies are now quickly implementing a multi-cloud strategy that makes use of a variety of cloud service providers for supporting their IT infrastructure [1]. 58 % said they use Microsoft Azure as their cloud platform provider, while 52% said they use Amazon Web Services. A further 19 % goes to Google Cloud; a further 9 % goes to Oracle Cloud; and a last 7.3 % goes to RackSpace.

2. OVERVIEW OF HYBRID CLOUD

Hybrid cloud computing discuss about aggregating and incorporating computer, connectivity, applications and storage into a unified management framework that allows enterprise IT and developers to leverage the scale, flexibility and cost savings of existing in-house IT investment tools and systems to manage in data of enterprise centre with their newly adopted cloud services. More than 80% of IT organisations are expected to use hybrid architectures, according to an IDC survey. Figure 1 shows Hybird prototypes.

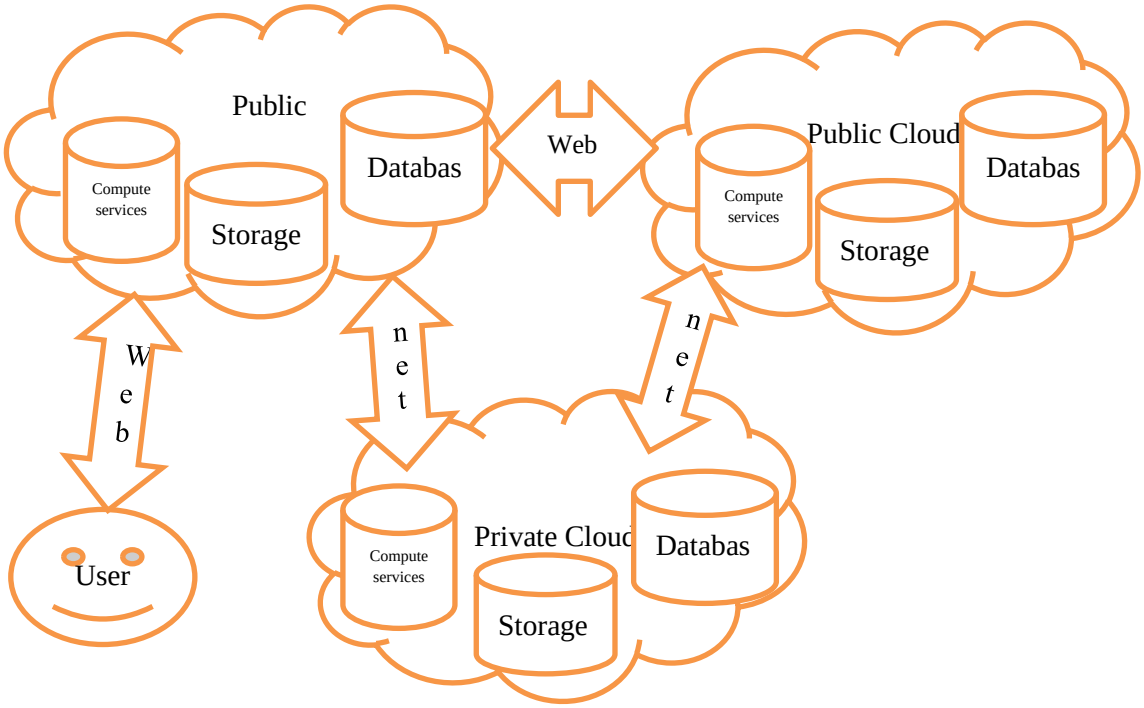

Figure:1 Hybrid cloud Services Model

Some features that a Hybrid cloud includes are discussed below.

- *Infrastructure Integration and Application Environment:* Workloads or virtual machines may be created in both private and public clouds with the use of a hybrid cloud platform.
- *Interconnectivity*: When two separate environments can communicate and interact with one other, the data, virtual machines, and applications they contain are more likely to be interconnected.
- *Applications Portability:* System components may be reused across several cloud environments using cloud-aware development.
- *Monitoring and Management:* The ability to keep track of and manage several cloud environments is essential nowadays. System health monitoring and management are critical in a hybrid cloud environment since it allows for cross-cloud insight into the overall health of the hybrid cloud system.

Some cloud computing resources are supplied and managed inside the company while others are outsourced to a hybrid cloud. It's possible that a firm uses Amazon Simple Storage Service (Amazon S3) for archived data but keeps its operating customer data on-premises. Ideally, a hybrid strategy enables a company to take benefit of the scalability and cost-effectiveness of public cloud computing environment without exposing mission-critical programmes and data to third-party risks. The term "hybrid IT" is also used to describe this form of hybrid cloud.

This specific topic is being concentrated on by a couple of the primary unique and cloud suppliers and providers. Several hybrid cloud storage providers are shown in Figure 2.

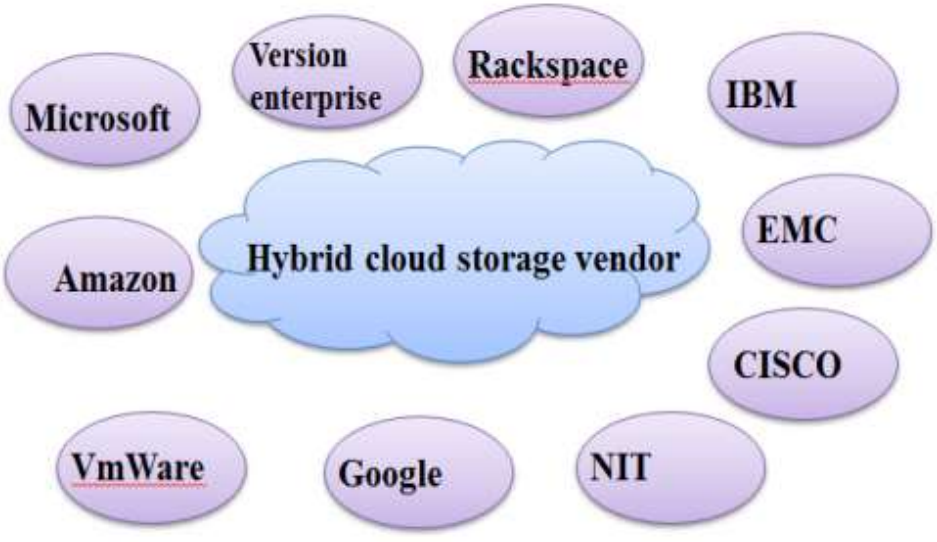

Figure 2. Hybrid Cloud Storage Vendors

3. HYBRID CLOUD COMPUTING CHALLANGES

Some challenges to consider when setting up hybrid clouds are:

i. ***Startup and Shutdown of On Demand*:** Cloud nodes must be able to be started and shut off on demand by your infrastructure. Cloud nodes must be started or stopped based on some kind of policy that takes into account the features of your application. With this method, you may launch new nodes if the main cloud becomes overcrowded and shut them down if it becomes underloaded by reacting to CPU consumption in either direction.

ii. ***Cloud-based Node Discovery:*** It is difficult to set up normal discovery protocols in the cloud since many cloud providers do not support IP Multicast (including Amazon and Go Grid). TCP would be required for your node discovery mechanism to function. However, you are also unaware of IP addresses of newly created nodes in cloud. If you want to avoid this, you should keep the IP addresses of new nodes in a cloud storage service like Amazon S3 or Simple DB.

iii. ***One-Directional Communication:*** It's a difficulty for large organisations to open up additional ports in firewalls to connect to cloud services. You may only be able to connect to a cloud in an outgoing fashion rather often. Such scenarios should be supported by your middleware. On top of that, you may come into a situation where A cloud can speak to B cloud, and the b cloud can talk to the other C cloud, but the A cloud cannot talk to the C cloud. Ideally, cloud A should be able to communicate with cloud C through cloud B.

iv. ***Latency Communication:*** Latency Cloud communication may take longer time than communication between nodes inside same cloud. Cloud-to-cloud connectivity, on the other hand, is sometimes substantially slower than cloud-to-local data centre connection. In order to avoid breaking down the cluster, your middleware layer should be able to respond and manage delays.

v. ***Reliability and Atomicity:*** On the cloud, many operations are unstable and non-transactional. F or example, when you store anyrthing on Amazon S3 storage, it doesn't guarantee that another programme will be able to access the data you've put there immediately. Data can't be protected from being overwritten, and file locking isn't possible. It is only possible to provide this functionality at the application or middleware levels.

4. ADVANTAGE OF HYBRID CLOUD

(1) In terms of wording, it’s more diverse since it includes both private and public cloud.

(2) As a result of this, the organization's demands may be met quickly with a hybrid cloud, as described in [5][6]. When an organization's in-house server can't manage a project that requires

a lot of computing power, a cloud-based solution is ideal. It would also save the company money by avoiding the need to purchase high-end server hardware, which is essential.

(3) From anywhere in the globe, a hybrid cloud may be used to work on a project at any time. Having a global presence enables them to serve enterprises who need to extend their impact beyond local boundaries. It's a safe haven for private information as well as a useful public resource.

(4) Due to the obvious allocated private cloud, it consistently offers the highest degree of security. Depending on the situation, it may be able to reduce and manage costs.

(5) Hybrid cloud may be quite expensive for a company if it choose to invest resources into a hosting provider or outsource the same. However, this innovation may be obtained for very low costs, making it a far better investment for the charitable organisation in the long run.

5. TRAITS OF HYBRID CLOUD

- *Security*: Keeping one's personal information safe is a constant concern. In addition to access control measures while data is stored, safeguarding efforts are put in place when it is transferred between storage locations and on-premises places [6]. The storing of documents should also be safe.
- *Reliability*: When it comes to data integrity, the hybrid cloud is also a factor. There must be no tampering with t data sent from person A to person B In the cloud, the data would be indexed by the cloud service. Likewise, its integrity should endure even if it isn't around anymore. For example, When indexes are damaged, the data is lost.
- *Business Coherence*: Scheduled and unscheduled downtime may have a negative impact on businesses' ability to function effectively. Capacity providers must offer backup mechanisms like snapshots, replication, and reinforcements, as well as quick recoveries in the event that their own infrastructure goes down.
- *Reporting And Charge-Back:* A cloud storage may be a compensatory pay-you-pay model, in which the bill is due at the conclusion of the charge cycle. This may serve as an example of any value-based fees that a provider can impose in addition to the capacity expenses.
- *Management*: The customer should be ready to cope with the circumstances in a hybrid cloud environment if they want to keep part of their data on-premises and some in the cloud.

6. CONCLUSION

Security and privacy are of the utmost importance in the cloud computing data storage. Despite fact that cloud storage & administration provides more flexibility and convenience, there are still risks of intruders and criminal behaviour. Cloud servers can provide more protection and privacy for data kept there.

Hybrid Cloud computing is an inevitable paradigm in which private and public clouds may be used simultaneously. Any new technology should take into account the many security risks that might be associated with it. In light of the different security concerns, cloud users and hybrid cloud providers alike will be better equipped to deal with these risks. In addition, a study of hybrid models has exploited and targeted with issue concerns a framework for security and a necessity for cloud security. It lessens the load on users of the majority of cost savings and complexity. Organizations are confident in the safety of their data in the face of security threats and system failures. For service delivery needs, it recommends a modernised IT operational agility.

Cloud computing and the security challenges that arise as a result of its fertilised, shared, public, private, and hybrid nature have been the subject of this study. There are a numerous ways to deal with cloud computing security concerns, and this article outlines some of them.

REFERENCES

1. Nitin Kumar, Shrawan Kumar Kushwaha and Asim Kumar, “Cloud Computing Services and its Application,” Advance in Electronic and Electric Engineering, Volume 4, Number 1 (2014), pp. 107-112.
2. Caifeng Zou, Huifang Deng and Qunye Qiu, “Design and Implementation of Hybrid Cloud Computing Architecture Based on Cloud Bus,” IEEE 9th International Conference on Mobile Ad-hoc and Sensor Networks, 2013.
3. M. Posey, “Journey to the Hybrid Cloud,” White Paper Sponsored by: VMware, IDC #242798R2, September 2015.
4. “Hybrid Cloud 101: Hybrid Cloud Computing - Intel” Intel IT Center Solution Brief | The Path to Hybrid Cloud, September 2013.
5. Rahul Khurana1, Himanshu Gupta, A Hybrid Model on Cloud Security 5th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) ,Volume.16,pp:347-352, ,2016.

6. Saurabh Singh, Young-Sik Jeong, Jong Hyuk park, A Survey on Cloud Computing Security: Issues, Threats, and Solutions ,Journal of Network and Computer Applications, SI1084-8045, ,2016.

EDITION 2021

Advances in Cloud Computing Security

Techniques and Applications

EDITORS

Dr. Stuti Asthana is currently working as Independent Researcher and Analyst. In 2019, she received doctoral degree, Doctor of Philosophy with specialization Artificial Intelligence from Suresh Gyan Vihar University, NAAC "A" Grade, Jaipur, India. In 2012, she received Master's degree, Master of Technology with specialization Computer Science and Engineering from Rajiv Gandhi Prodhyogiki Vishwavidhyalaya , Bhopal, M.P., India.. In 2006, she received Bachelor's degree, Bachelor of Engineering from Oriental Institute of Science & Technology, Bhopal, M.P., India, in Computer Science and Engineering. Within the period of her doctoral-level studies at the Doctor of Philosophy of computer science and engineering, she has demonstrated himself extraordinary as disciplined and dedicated research scholar. As an unwavering dedicated researcher, he has published 12 research papers in various reputed Web of Science, SCOPUS, DBLP indexed international journals and flagship international Scopus Indexed conferences. Her doctoral research investigates the design and development of Multiscript Handwriting Recognition Using Neural Network. She is an active Editorial Board member for various reputed international journals. She has also published National and International Patents in the areas of Data Mining, Machine Learning, IoT and Neural Network.

Dr. Rakesh Kumar Bhujade completed Ph.D. in Artificial Intelligence in 2016, cracked UPSC and currently working as Head of the Information Technology Department in Government Polytechnic, Daman. He has published 01 Philippines Patent, 02 Australian Government Patents (Grants), 05 Indian Patents, 03 Books, 02 AWS Global Certifications, 01 Research Grant and more than 30 papers in Scopus Indexed International Journal/Conferences. He also acted as Advisory/Reviewer Board member in many International Journals and Conferences. Dr. Rakesh is expertise in Neural Network, Soft Computing, Artificial Intelligence and Machine Learning.

Prof. Ghanshyam Prasad Dubey is a post graduate in Engineering with specialization in the stream of Information Technology. He is having a vast experience of over 14 years in academics. His subjects of interest include Database Management System, Computer Networking, Operating Systems, Basic Computer Engineering and JAVA Programming. His recent research interests include Cloud Computing, Big Data, Machine Learning, Information Security and Algorithms. He is having 10 publications in reputed international conferences and 6 research papers in Scopus indexed journals.

Price : 1200 INR

AGPH **Books**

www.agphbooks.com

9 788195 534067