IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

Paradigm Shift of Big-Data Application in Cloud Computing

Ms. Preeti Narooka[1], Dr. Sunita Chaodhary[2]

Ph. D Student, Banasthali University, Jaipur, India[1]

Associate Professor, Banasthati University, Jaipur, India [2]

Abstract: Big Data and Cloud Computing came before few years ago but their capture all kind of businesses, technologies and each and every people. After few years world can't think without Cloud Computing and Big Data Applications. Today is the time of combinations of number of technologies and the combined effects of emerging Internet technologies, increased computing power, and fast, pervasive digital communications were spawning new ways to manage talent and assets as well as new thinking about organizational structures. This paper is about Parallel Processing and Distributed Computing concept in Big Data and Cloud Computing Application. And also about how Big Data help Cloud Computing for Processing and vice-verse. The best examples of all above technologies are Hadoop-HDFS for File System or Storing Unstructured Data and Map Reduce is for Processing of the stored Data.

Keywords: Big Data, Cloud Computing, Distributed Computing, distributed Storage, Hadoop, Map Reduce, Parallel Processing.

I. INTRODUCTION

The term Cloud Computing and Big Data concept is used almost anywhere these days, from news articles to professional magazines, from Social Media to YouTube, blogs discussion, single hand companies to Multi-National companies. Both terms almost introduce nearby in 2005, where Big Data introduce because large data sets almost impossible to manage and process using traditional data management concept-due to size, but also their complexity. Big Data can be seen in all finance and business where enormous amount of stock exchange, banking, online and onsite purchasing data flows through computerized system every day and are then captured and stored for inventory monitoring, customer behaviour and market behaviour. And Cloud computing is just came in frame to maximizing the effectiveness of the shared resources [1]. Cloud Computing and Big Data are originate from distributed computing and parallel Computing respectively. Intuitively, it is not possible to overcome from all the issues and challenges inside the Big data and Cloud. Therefore, there is a requirement of combine methodology for Processing and Storing purpose of the data. In this paper discussed basic, issues and challenges of both. Session second is about the origin of Cloud computing and processed Big Data. Challenges, Issues and Literature reviews are discussed in session third & forth. In last this paper focus on how their combination's application could help to overcome more number of issues and challenges.

II. DISTRIBUTED COMPUTING TO CLOUD COMPUTING

Distributed computing is a base methodology for the Cloud Computing. A distributed system [16] is a collection of autonomous computers, connected through a network and distribution software called middleware, which enables computers to coordinate their activities and to share the resources of the system; the users perceive the system as a single, integrated computing facility. A distributed system has several characteristics: its components are autonomous, scheduling and other resource management and security policies are implemented by each system, there are multiple points of control and multiple points of failure, and the resources may not be accessible at all times. Distributed systems can be scaled by adding additional resources and can be designed to maintain availability even at low levels of hardware/software/network reliability.

Similarly, Distributed Computing perform task on different computer which are connected through network and one system behave as a master of all; coordinate the activities and share the resource of the system.

Distributed resources, such as storage devices, data sources, and supercomputers, are interconnected and can be exploited by users around the world as single, unified resource. To a growing extent, repetitive or resource-intensive IT tasks can be outsourced to service providers, which execute the task and often provide the results at a lower cost. Best example of resource sharing is social network. Mostly social network sites are implemented on the concept of large distributed computing systems.
Where, they are handled by centrally controlled data centers. However, the classification of the distributed computing is of peer-to-peer, utility, cluster, and jungle computing.

Where, the utility computing is parent of the grid computing and the cloud computing. Classification of Distributed computing is shown in the below figure 1

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

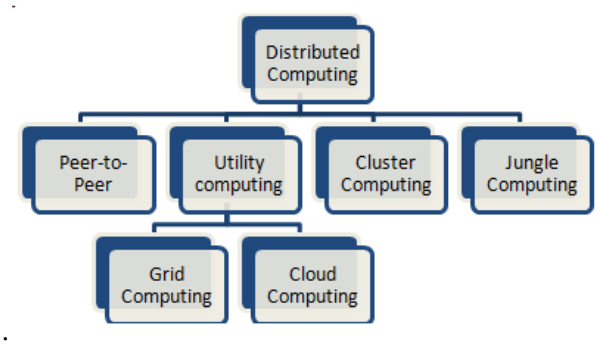

Figure:1

Utility Computing: After few years Distributed computing moving/upgraded toward Utility Computing. The concept of Utility computing is simple: rather than operating servers in-house, organizations subscribe to an external utility computing service provider and pay only for those hardware and software resources which they want to use. Utility computing is based on the consolidation principle, where organization's resources are shared by a number of applications and users. The principal resources offered include, but are not limited to, virtual computing environments (paid per hour and data transfer), and storage capacity (paid per GB or TB used).

Grid Computing: Are very large-scale virtualized, distributed computing systems. They cover multiple administrative domains and enable virtual organizations. Their shared large grids for instance, for example 80,000 CPU cores are shared within Enabling Grids for E-science, which is one of the largest multi-disciplinary grid infrastructure in the world.

This brings together more than 10,000 users in 140 institutions (300 sites in 50 countries) to produce a reliable and scalable computing resource available to the European and global research community [16].

Cloud computing [2], [3]: is a type of internet-based computing that relies on sharing computing resources rather than having personal devices or local servers to handle applications. Cloud computing [4], [5] allows people to set up their virtual office and gives the flexibility to connecting with any business anywhere and anytime.

Definition and Characteristics:

The **Cloud Computing** paradigm originates mainly from research on distributed computing, utility computing and virtualization, as it is based on principles, techniques and technologies developed in these areas [7].[Table1]. The informal definition of Cloud Computing proposed in Mell and Grance (2011) as below: "Cloud Computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

TABLE1: COMPUTING AND THEIR SERVICES & CHARACTERISTIC

Computing Paradigms	Services	Enhanced Features
A. Cloud computing B. Edge computing C. Grid computing D. Utility computing	a. Software as a Service (SaaS) b. Infrastructure as a Service (IaaS) c. Platform as a Service (PaaS) d. Service-Oriented Architecture (SOA)	a. Ubiquitous access b. Reliability c. Scalability d. Virtualization e. Exchangeability / Location independence f. Cost-effectiveness

There are multiple characteristics associated with Cloud Computing. They are mentioned below:

1) Elasticity /Scalability
2) Outsourcing Security Multi-Tenancy
3) Virtualisation
4) Pay per use
5) Provenance
6) Ease of Use
7) Agility
8) Adaptability
9) Resource utilisation
10) Privacy
11) Availability
12) Energy efficiency
13) Data Management Cost efficiency
14) Reliability Metering
15) Programmability

III. PARALLEL COMPUTING AND BIG DATA TECHNIQUES

In company's employees' performing task in a group, where task is divides into small job and given to each and every person of that group. This type of process is known as a parallel performance of the task.

Similarly, Parallel Computing [figure 2] solves large problems by splitting them into smaller ones and solving them concurrently. Parallel computing was considered for many years for solving data-intensive problems encountered in many areas of science, engineering, and enterprise computing; it required major advances in several areas including, algorithms, programming languages and environments, performance monitoring, computer architecture, interconnection networks, and, last but not least, solid-state technologies.

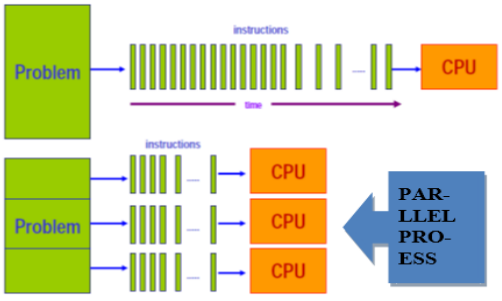

Figure 2:

DOI 10.17148/IJARCCE.2016.55130

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

Parallel hardware and software systems to solve problems demanding more resources than those provided by a single system and, at the same time, to reduce the time required to obtain a solution. It also speed-up the process and increase the efficiency.

There are different levels of parallelism likewise:
1) Bit level parallelism.
2) Instruction-level parallelism.
3) Data parallelism or loop parallelism.
4) Task parallelism.

Where parallel computing concept is useful in big data for dividing all tasks into small job and pursue further. Basically its concept helps to solve large problems concurrently. And there many other techniques which are also part of big data likewise Data Mining, Data Analysis, Data Clustering and managing and many more. Before discussing big data techniques and applications everyone should know the basic information about big data.

Big Data [14]: is a large amount of data which requires advance technologies and architectures to extract data by capturing and analysis process. Data is growing at a huge speed making it difficult to handle such a large amount. The large amount is mainly difficult to handle because of rapid increment of data in compare to performance and computing of resources. Big data is an advance technology for bring huge benefits to the business organizations in place of traditional techniques. Big data due to its various properties like volume, velocity, variety, variability, value and complexity solve many issues and challenges [15].

Definition and Characteristics:
Big Data is a data (Exabyte) which is growing at a huge speed and handling of it's also too difficult. The main difficulty in handling such large amount of data is because that the volume is increasing rapidly [figure 3].
The following definition is also good way to define Big Data [14]:

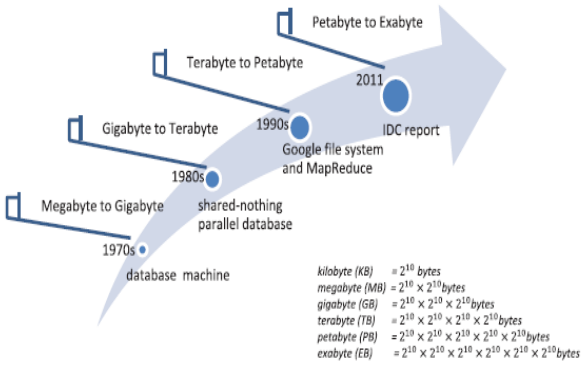

Figure 3

"Big Data" is a term encompassing the use of techniques to capture, process, analyse and visualize potentially large datasets in a reasonable timeframe not accessible to standard IT technologies. By extension, the platform, tools and software used for this purpose are collectively called "Big Data technologies".

Big Data have three most important characteristics as VOLUME, VELOCITY and VARIETY which are called as "3 Vs of Big Data [7]".

a) Volume: Volume is the most important feature of Big Data which impose additional and specific requirements to all traditional technologies and tools currently used. In Big Data Volume includes size, scale, amount, and dimension for Terabyte- and Exabyte data from different processes which stored in databases. In world, social services providers such as Google, Facebook, and Twitter are producing, analysing and storing data in huge volume.

b) Velocity: Big Data are generated at high speed, including also data generated by multiple events, and need to be processed in real-time, near real-time or in batch. Meaning of above statement is petabyte size data generated in less amount of time.

c) Variety: Its deals with the collected complex big data. Variety of collected are structured, unstructured, semi- structured, and a mixed data. Data variety imposes new requirements to data storage and database design which should dynamic adaptation to the data format.
Above session is about background and characteristic of Cloud computing and Big data. And in next session is about Literature review, Challenges and Issues.

IV. CHALLENGES AND ISSUES

Cloud Computing provide advance technologies but with advancement they have issues and challenges too. In this paper some challenges and issues of Cloud Computing are described.

a) Cost Models: Cloud computing is a large scope for business models, but to make this models viable they have face lot of problems. It should be of low cost or to be free. Also the growing competition between cloud service providers makes its more and more difficult for user.

b) Data Management and Handling: Plenty of work has been invested into distributed data bases etc. And in cloud computing data production and consumption is constantly increasing. Data management is just not for maintaining the data, but it also about efficiently computing, communicating and distributing to fulfilment of quality requirement.

c) Resource Awareness: The relationship between resources and their impact on execution of different applications/ use cases is still widely unknown.

d) Moving to the Cloud Computing: Only switching cost does not arise but also most organization or applications are not ready for cloud. They are not programmed in a fashion of the Cloud Computing Characteristics.

e) Data Challenges - Big Data: Cloud Computing following the Web Service development, they have very classical servers and clients likewise organization. Specific technology requires interactive communication with multiple users which includes e.g. voice, video chat, live virtualization etc. All this increases the complexity of data. Since application or organizations needed to handle or deal with a different-different type of data. When relevant data dispersed over the network to different place which become complicate programming. All this are strongly related to the Big Data problem.

f) Performance: The Cloud computing performance highly depends on the Scalability, Utilization of resources and Communication strength. It's all about how Cloud computing is maintain balance with all above characteristics. And the necessity for high data throughput constantly increases due to the amount of users and satisfying the need of data. This is also effects the performance of Clouds.

g) Lack of Better Specialization Support: Though even generic platforms imply specialisation to specific needs, they mostly try to address as many use cases as possible thus not offering the performance and capabilities needed for specific users. Given the growing interest in services comprising multiple domains, it is thereby also not clear to which degree CLOUD offerings should be specialised, respectively generic.

h) Trust (legislation, policy, security, and privacy etc.): A Cloud provides "global" data hosting, possibly across multiple legal jurisdictions, raising compliance issues for both users and providers. The additional lack of proper data encryption modes that support remote computing without decryption makes many users distrust the CLOUD – without additional means, such as homomorphism encryption, data will always be accessible to the provider himself, at least. But it is not only the lack of security support that is cause for distrust, but also the lack understanding of CLOUD behaviour, quality of service to be realistically expected, performance rating etc. Where the Big data techniques, increases the performance and computation of the resources but parallel the complexity of the process is rapidly increased. Since Big data has so many challenges and issues. In this paper some of the common challenges [9], [15] and issues [10], [11] are discussed.

i) Privacy and Security: Privacy and Security are big concern, and one that increases in the context of Big data. When in organization personal in information in combined with external large data it's become difficult to handle with care. Another important consequence arising would be Social stratification where a literate person would be taking advantages of the Big data predictive analysis and on the other hand underprivileged will be easily identified and treated worse.

j) Storage and Processing Issues: The storage available is not enough for storing the large amount of data which is being produced by almost everything: Social Media sites are themselves a great contributor along with the sensor devices etc. Uploading the large amount of data in cloud doesn't solve the problem.
Processing of such large amount of data also takes large amount of time. To find suitable elements whole of data Set needs to be Scanned which is somewhat not possible.

k) Data Access and Sharing of Information: To make accurate decisions for the data processing, system needs accurate, complete and timely mannered data. This entire thing makes data management more complex for decision making, business intelligence and productivity improvements. Expecting sharing of data between companies is awkward because of the need to get an edge in business. Sharing data about their clients and operations create difficulty to built trust and haziness in existence of the organizations.

l) Analytical challenges: Big data brings some huge analytical challenges. The huge amount of data which can be unstructured, semi structured or structured requires a large number of advance analytic skills. Moreover the type of analysis which is needed to be done on the data depends highly on the results to be obtained i.e. decision making. This can be done by using one using two techniques: either incorporate massive data volumes in analysis or determine upfront which Big data is relevant.

m) Skill Requirement: Since Big Data is at its youth and emerging technology so it needs to attract organizations and youth with diverse new skill sets. Youth should take interest to extend research, analytical and creative idea for huge data. For all this university need to introduce Big data to produce skilled youth toward this emerging technology.

n) Technical Challenges:
a) Quality of Data: Collection of huge amount of data and its storage comes at a cost. More data if used for decision making or for predictive analysis in business will definitely lead to better results. This further leads to various questions like how it can be ensured that which data is relevant, how much data would be enough for decision making and whether the stored data is accurate or not to draw conclusions from it etc.
b) Heterogeneous Data: Unstructured data represents almost every kind of data like social media interactions, to recorded meetings, to handling of PDF documents, fax transfers, to emails and more. It's a completely raw and unorganized and conversion of unstructured to structured data is not feasible. To process unstructured data is cumbersome and very costly.

V. LITERATURE REVIEW

A. Bo Li, (2013) [9]
This paper reveals most recent progress on big data networking and big data. We have categorized reported efforts into four general categories. First, efforts related to classic big data technology such as storage, Software-

Defined Network, data transportation and analytics are reported. Second, important aspects of big data in cloud computing such as recourse management and performances optimization are introduced. Lastly, we introduce interesting benchmarks and progress in both search engines and mobile networking.

B. Zibin Zheng, Jieming Zhu, and Michael R. Lyu(2013) [8]
The overwhelming service-generated data become too large and complex to be effectively processed by traditional approaches. How to store, manage, and create values from the service-oriented big data become an important research problem. On the other hand, with the increasingly large amount of data, a single infrastructure which provides common functionality for managing and analyzing different types of service-generated big data is urgently required. To address this challenge, this paper provides an overview of service-generated big data and Big Data-as-a-Service.

C. Yuri Demchenko, Cees de Laat, Peter Membrey (2014) [12]
This paper discusses a nature of Big Data that may originate from different scientific, industry and social activity domains and proposes improved Big Data definition that includes the following parts: Big Data properties (also called Big Data 5V: Volume, Velocity, Variety, Value and Veracity), data models and structures, data analytics, infrastructure and security. The paper discusses paradigm change from traditional host or service based to data centric architecture and operational models in Big Data. The Big Data Architecture Framework (BDAF) is proposed to address all aspects of the Big Data Ecosystem and includes the following components: Big Data Infrastructure, Big Data Analytics, Data structures and models, Big Data Lifecycle Management, Big Data Security. The paper analyses requirements to and provides suggestions how the mentioned above components can address the main Big Data challenges. The presented work intends to provide a consolidated view of the Big Data phenomena and related challenges to modern technologies, and initiate wide discussion.

D. Weidong Bao, Wenhua Xiao, Haoran Ji, Chao Chen, Xiaomin Zhu and Jianhong Wu (2016) [13]
As the data are dynamically generated and the resource pricing varies over time, moving the data from differently geographic locations to different data centers while provisioning adequate computation resource to process them is an essential task to achieve cost effectiveness for DSP. In this paper, a joint online approach is proposed to address this task. We formulate the problem into a joint stochastic optimization problem, which is then decoupled into two independent sub-problems via the Lyapunov framework.

VI. MIGRATION AND PARADIGM SHIFT

Several new technologies have emerged with the development of network technology (Social Networks) and distributed computing. Because of IT technology and the Internet, tremendous amount of digital information is produced and distributed every day. As a result, efforts to extract information from large-scale information are being accelerated. It is generally agreed today that big data and cloud are major trend of modern computer technology [7]. Danah Boyd et. al. [9] presented that modern society is the era of big data.

In today's time everything related to computer are revolving around cloud and big data. Most of the people are on social network, in organization they having their own network; number of e-commerce sites, all these are reasons of huge data and processing become complex which requires cloud as base. Basically cloud is a perfect fit for Big Data analytics. The hyperscale industry leaders -- Google, Facebook and so on -- have been doing it for many years. But for small enterprises could be more important and useful. It would help for:

- Cost reduction
- Reduced overhead
- Rapid provisioning/time to market
- Flexibility/scalability

"Cloud computing provides enterprises cost-effective, flexible access to Big Data's enormous magnitudes of information. Big Data on the cloud generates vast amounts of on-demand computing resources that comprehend best practice analytics. Both technologies will continue to evolve and congregate in the future." -- Oracle said.

The data which generated by companies and social networks are in huge size which also known as data management issues. To solve this problem, technology is distributed database because traditional management systems are failed to handle huge size of data. Traditional data warehouses are structured database where they should be cleansed, documented and even trusted. But all this constraint is not with big data. They can handle any form of data because it's an unstructured form of data.

Organizations continue to store more and more data in cloud environments, and clouds offer business users scalable resources on demand. This networking resources are processed with big data tools like Apache Hadoop (Software provides the high-performance computing power needed to analyze vast amounts of data efficiently and cost effectively).

Running Hadoop in virtualized environments continues to evolve and mature with initiatives like VMware's open-source project Serengeti*, among others [13].

All the technologies discussed in above session having their owe specification and process to perform task, but today's time required more robust task performing technologies. Since single technology is not satisfactory for good output.

Need to combine/mix number of technologies to have best outcome likewise Hadoop and MapReduce are most

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

suitable example of all above combined techniques. Refer figure 4.

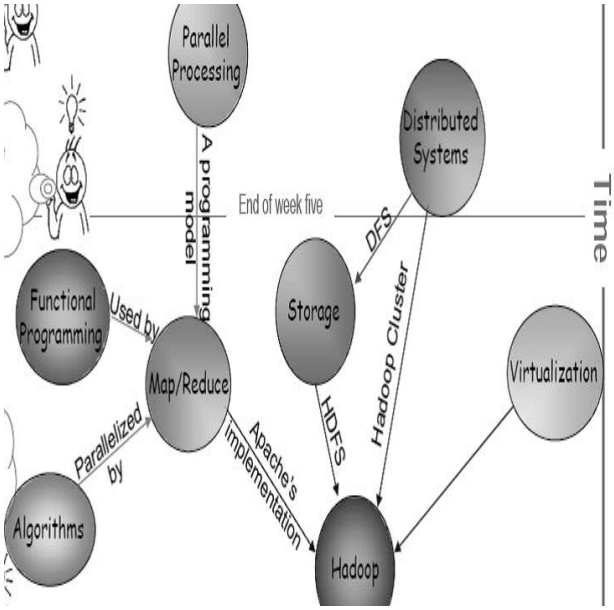

Figure 4

Paradigm shift of Big Data Application in Cloud Computing

Cloud computing is a very powerful technology to perform complex computing and also maintain expensive computing hardware, space and software. Where big data need large on-demand compute power and distributed storage to crunch the 3V data problem and cloud seamlessly provides this elastic on-demand.

Data storage [12] plays an important or critical role in the performance of the data applications; where the mostly problems faced as a storage location, as well as, the decision making application performance. Clouds support's storage system for further process.

Big data application performance in cloud is connected through number of network resources to complete the task efficiently and fast. But it become very complex structure and reduces the processing performance.

Cloud Computing requires to distribute computations and data at a multiple system which is very hyper critical function to perform by organizations. The coordination's of the distributed environment also known as coordination of the data storage and data analysis. Here is the solution for the both problems known as Big Data applications. As earlier discussed in this paper Big data applications help to store and perform task on huge unstructured size of database.

Combinations of Hadoop and Map reduce application can reduce the complexity of the data coordination and analytical performance of the data in cloud.

ZooKeeper performs as a distributed coordinator which can coordinate huge number of nodes in the cloud. ZooKeeper is an open-source software which written in Java. It works as a front end of the networks which help clients to connect any servers. A client can watch on znode which is similar to the inode of the hadoop distributed file system. The organization allows updating, retrieval, and read-write of data to the clients.

The main advantage of cloud computing is optimally respond to the cost and the reduce time constraints of the application. All this process can be done by MapReduce programming. MapReduce is based on the Parallel processing of data-intensive applications supporting arbitrarily divisible load sharing. First, split all file into blocks and assign each block to task trackers (task-nodes) and they process it parallel known as Mapper. Second phase, merge all the partial results of task trackers into one known as Reducer (merging process perform by single task tracker). And we get proper result from unstructured data in the cloud without wasting a long time.

Since all this application of big data in Cloud computing can perform more efficiently and effectively for data process. Application of Big data in cloud can solve the problem of Data storage and Data analysis but still one issue couldn't completely solved by all above techniques which known as Processing of Data. Still it required to optimize more efficiently.

VII. CONCLUSION

This paper provides an overview of Big data and Cloud Computing, their origins, applications, challenges, issues and also about the literature review of both. Reviewed the progresses in fundamental Big data technologies such as storage and analysis in Cloud. Important aspects of big data networking in cloud computing such as new challenges and opportunities, resource management and performance, optimizations are also discussed. To sum up, the conclusion is that Big data and Cloud computing are latest technology and together they are more promising for good results. Both progresses have been made in the area of distributed file system and distributed database. Almost all the issues can overcome from Big Data and Cloud computing combination Likewise, Data storing, Sharing, Processing, Analysis, and many more. Still some solutions required improvement. Data Processing and Analysis are those issues could be optimized with better solution because increasing the network size, cost estimation affected organization system. There should be a solution which perform data processing without increasing the cost and can optimize the performance.

REFERENCE

Books

[1]. AlexaHuth and James Cebula; "The Basics of Cloud Computing"; US-CERT; Cloud Computing Huth Cebula.pdf www.btpsonline. com/wysk.html

[2]. Han Qi, Abdullah Gani "Research on Mobile Cloud Computing: Review, Trend and Perspectives".

[3]. Yanpei Chen, Vern Paxson, Randy H. Katz, "What's New About Cloud Computing Security?", Electrical Engineering and Computer

Sciences University of California at Berkeley.

[4]. Issa M. Khalil, AbdallahKhreishah and Muhammad Azeem; "Cloud Computing Security: A Survey"; Computers 2014, 3, 1-35; doi: 10.3390/computers3010001; ISSN 2073-431X; www.mdpi. com/journal / computers

[5]. Keiko Hashizume, David G Rosado, Eduardo Fernández-Medina and Eduardo B Fernandez "An analysis of security issues for cloud computing";Hashizume et al. Journal of Internet Services and Applications 2013, 4:5; http://www.jisajournal.com/content/4/1/5.

Research papers

[6]. Yuri Demchenko, Zhiming Zhao, Paola Grosso, AdiantoWibisono, Cees de Laat (2012), "Addressing Big Data Challenges for Scientific Data Infrastructure", IEEE , 4th International Conference on Cloud Computing Technology and Science

[7]. Dan C. Marinescu (December 6, 2012), "Cloud Computing: Theory and Practice", Computer Science Division, Department of Electrical Engineering & Computer Science, University of Central Florida, Orlando, FL 32816, USA,

[8]. Zibin Zheng, Jieming Zhu, and Michael R. Lyu, "Service-generated Big Data and Big Data-as-a-Service: An Overview", Department of Computer Science & Engineering, The Chinese University of Hong Kong, Hong Kong, China, {zbzheng,jmzhu,lyu}@cse.cuhk.edu.hk

[9]. Bo Li, "Survey of Recent Research Progress and Issues in Big Data", boli@seas.wustl.edu (A paper written under the guidance of Prof. Raj Jain)

[10]. Yuri Demchenko, Paola Grosso, Cees de Laat, "Addressing Big Data Issues in Scientific Data Infrastructure", System and Network Engineering Group, University of Amsterdam, Amsterdam, The Netherlands, e-mail: {y.demchenko, p.grosso, C.T.A.M.deLaat}@uva.nl and Peter Membrey, Hong Kong Polytechnic University, Hong Kong SAR, China.

[11]. Avita Katal, Mohammad Wazidand R H Goudar, Big Data: Issues, Challenges, Tools and Good Practices, Department of CSE, Graphic Era University, Dehradun, India.

[12]. Yuri Demchenko, Cees de Laat , Peter Membrey, "Defining Architecture Components of the Big Data Ecosystem", vol. - 978-1-4799-5158-1/14/$31.00 ©2014 IEEE

[13]. Weidong Bao, Wenhua Xiao, Haoran Ji, Chao Chen, Xiaomin Zhu and Jianhong Wu, "TOWARDS BIG DATA PROCESSING IN CLOUDS: AN ONLINE COST-MINIMIZATION APPROACH", Big Data and Information Analytics, American Institute of Mathematical Sciences, Volume 1, Number 1, January 2016.

White paper

[14]. NESSI – Big Data White Paper; NESSI_WhitePaper_BigData.pdf

[15]. Challenges and Opportunities with Big Data, A community white paper developed by leading researchers across the United States

[16]. International Telecommunication Union – "Distributed Computing: Utilities, Grids & Clouds", ITU-T Technology Watch Report, March 2009