International Journal of Recent Trends in Engineering & Research
ISSN(ONLINE) : 2455-1457
IMPACT FACTOR : 3.344

IMPLICATION OF SECURITY ISSUES ASSOCIATED WITH BIG DATA IN CLOUD COMPUTING

Santosh Kumar Satapathy[1], Santosh Kumar Moharana[2], Avay Kumar Ojha[3]
[1,2,3] cse, Gandhi Engineering College

Abstract— In this paper, we discuss security issues for cloud computing, Big data, Map Reduce and Hadoop environment. The main focus is on security issues in cloud computing that are associated with big data. Big data applications are a great benefit to organizations, business, companies and many large scale and small scale industries.We also discuss various possible solutions for the issues in cloud computing security and Hadoop. Cloud computing security is developing at a rapid pace which includes computer security, network security, information security, and data privacy.Cloud computing plays a very vital role in protecting data, applications and the related infrastructure with the help of policies, technologies, controls, and big data tools. Moreover, cloud computing, big data and its applications, advantages are likely to represent the most promising new frontiers in science.
Keywords— Cloud Computing, Big Data, Hadoop,Map Reduce, HDFS (Hadoop Distributed File System)

I. INTRODUCTION

In order to analyze complex data and to identify patterns it is very important to securely store,manage and share large amounts of complex data.Cloud comes with an explicit security challenge, i.e. the data owner might not have any control of where the data is placed. The reason behind this control issue is that if one wants to get the benefits of cloud computing, he/she must also utilize the allocation of resources and also the scheduling given by the controls. Hence it is required to protect the data in the midst of untrustworthy processes. Since cloud involves extensive complexity, we believe that rather than providing a holistic solution to securing the cloud, it would be ideal to make noteworthy enhancements in securing the cloud that will ultimately provide us with a secure cloud.

MapReduce processes exceedingly large amounts of data without being affected by traditional bottlenecks like network bandwidth by taking advantage of this data proximity. Hadoop, which is an open-source implementation of Google MapReduce, including a distributed file system, provides to the application programmer the abstraction of the map and the reduce. With Hadoop it is easier for organizations to get a grip on the large volumes of data being generated each day, but at the same time can also create problems related to security, data access, monitoring, high availability and business continuity.

1.1 Cloud Computing

Cloud computing is an umbrella term used to refer to Internet based development and services. The cloud is a metaphor for the Internet. A number of characteristics define cloud data, applications services and infrastructure:
Remotely hosted: Services or data are hosted on someone else's infrastructure.
Ubiquitous: Services or data are available from anywhere.
Commodified: The result is a utility computing model similar to traditional that of traditional utilities, like gas and electricity.

Cloud computing is the delivery of computing services over the Internet. Cloud services allow individuals and businesses to use software and hardware that are managed by third parties at remote locations. Examples of cloud services include online file storage, social networking sites, webmail, and online business applications. The cloud computing model allows access to information and computer resources from anywhere that a network connection is available. Cloud computing provides a shared pool of resources, including data storage space, networks, computer processing power, and specialized corporate and user applications.

1.2 Architecture

A basis infromation about the architecture is provided in this chapter, together with the explanations of relevant terms such as virtualization, Frond/Back end or Middleware.

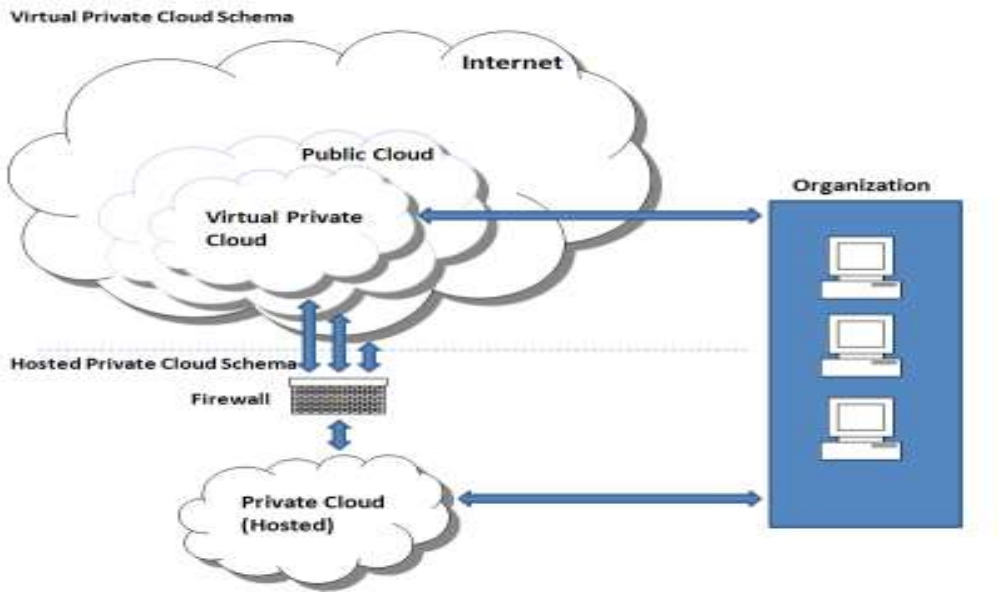
Virtualization is best described as essentially designating one computer to do the job of multiple computers by sharing the resources of that single computer across multiple environments. Virtual servers and virtual desktops allow you to host multiple operating systems and multiple applications locally and in remote locations, freeing your business from physical and geographical limitations. [5]

The Cloud Computing architecture can be divided into two sections, the front end and the back end, connected together through a network, usually Internet. The Front End includes the client's computer and the application required to access the cloud computing system. Not all cloud computing systems have the same user interface. Services like Web-based e-mail programs leverage existing Web browsers like Internet Explorer or Firefox. Other systems have unique applications that provide network access to clients.

The Back End of the system is represented by various computers, servers and data storage systems that create the "cloud" of computing services. Practically, Cloud Computing system could include any program, from data processing to video games and each application will have its own server.

A central server administers the system, monitoring traffic and client demands to ensure everything runs smoothly. It follows a set of rules called protocols and uses a special kind of software called Middleware. Middleware allows networked computers to communicate with each other. [6]

Public Cloud (external cloud) is a model where services are available from a provider over the Internet, such as applications and storage. There are free Public Cloud Services available, as well as pay per usage or other monetized models. Private Cloud (Internal Cloud/Corporate Cloud) is computing architecture providing hosted services to a limited number of people behind a company’s protective firewall and it sometimes attracts criticism as firms still have to buy, build, and manage some resources and thus do not benefit from lower up-front capital costs and less hands-on management, the core concept of Cloud Computing. [7]



Private/Public cloud

1.3 What Is Big Data

Big data refers to the collection and subsequent analysis of any significantly large collection of data that may contain (user data, sensor data, and machine data). When analyzed properly, big data can deliver new business insights, open new markets, and create competitive advantages. Compared to the structured data in business applications, big data consists of the following three major attributes:

- Variety—Extends beyond structured data and includes semi-structure or unstructured data of all varieties, such as text, audio, video, click streams, log files, and more.
- Volume—Comes in one size: large. Organizations are awash with data, easily amassing hundreds of terabytes and petabytes of information.
- Velocity—Sometimes must be analyzed in real time as it is streamed to an organization to maximize the data's business value.

1.4 Big Data Use Cases

There are many examples of big data use cases in virtually every industry imaginable. Some businesses have been more receptive of the technologies and faster to integrate big data analytics into their everyday business than others. It is evident that organizations embracing this technology not only will see significant first-mover advantages but will be considerably more agile and cutting edge in the solutions and adaptability of their offerings.

Use case examples of big data solutions include

- Financial services providers are adopting big data analytics infrastructure to improve their analysis of customers to help determine eligibility for equity capital, insurance, mortgage, or credit.
- Airlines and trucking companies are using big data to track fuel consumption and traffic patterns across their fleets in real-time to improve efficiencies and save costs.
- Healthcare providers are managing and sharing patient electronic health records from multiple sources—imagery, treatments, and demographics—and across multiple practitioners. In addition, pharmaceutical companies and regulatory agencies are creating big data solutions to track drug efficacy and provide more efficient and shorter drug development processes.
- Telecommunications and utilities are using big data solutions to analyze user behaviors and demand patterns for a better and more efficient power grid. They are also storing and analyzing

environmental sensor data to provide insight into infrastructure weaknesses and provide better risk management intelligence.

Big Data versus Traditional Data Types

Components Traditional Data Big Data components	Traditional data	Bigdata
Architecture	Centralized	Distributed
Data volume	Terabytes	Petabytes to exabytes
Data type	Structured or transactional	Unstructured or semi-structured
Data relationships	Known relationship	Complex/unknown relationships
Data model	Fixed schema	Schema-less

1.5 Big Data Technologies (Hadoop)

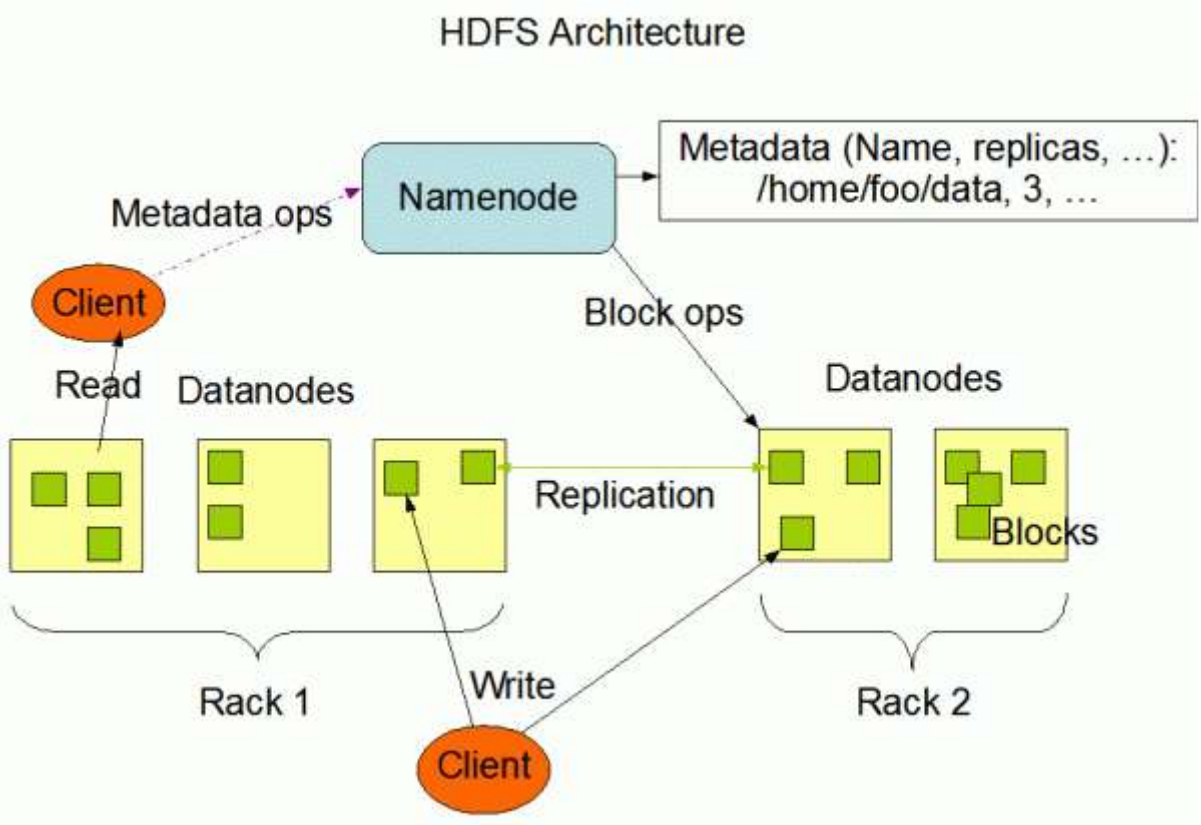
The driving force behind an implementation of big data is the software—both infrastructure and analytics. Primary in the infrastructure is Hadoop. Hadoop is the big data management software infrastructure used to distribute, catalog, manage, and query data across multiple, horizontally scaled server nodes. Yahoo! created it based on an open source implementation of the data query infrastructure (originated at Google) called MapReduce. It has a number of commercially supported distributions from companies such as MapR Technologies and Cloudera. Hadoop is a framework for processing, storing, and analyzing massive amounts of distributed unstructured data. As a distributed file storage subsystem, Hadoop Distributed File System (HDFS) was designed to handle petabytes and exabytes of data distributed over multiple nodes in parallel.

Hadoop Ecosystem

The Hadoop platform consists of two key services: a reliable, distributed file system called Hadoop Distributed File System (HDFS) and the high-performance parallel data processing engine called Hadoop MapReduce. The combination of HDFS and MapReduce provides a software framework for processing vast amounts of data in parallel on large clusters of commodity hardware in a reliable, fault-tolerant manner. Hadoop is a generic processing framework designed to execute queries and other batch read operations against massive datasets that can scale from tens of terabytes to petabytes in size.

The Hadoop Cluster

Hadoop, which includes a distributed file system known as Hadoop Distributed File System (HDFS) and MapReduce, is a critical big data technology that provides a scalable file system infrastructure and allows for the horizontal scale of data for quick query, access, and data management.



1.6 MapReduce

MapReduce is a massively scalable, parallel processing framework that works in tandem with HDFS. With MapReduce and Hadoop, compute is executed at the location of the data, rather than moving data to the compute location; data storage and computation coexist on the same physical nodes in the cluster. MapReduce processes exceedingly large amounts of data without being affected by traditional bottlenecks like network bandwidth by taking advantage of this data proximity.

Key MapReduce Features:

Scale-out Architecture - Add servers to increase processing power
Security & Authentication - Works with HDFS and HBase security to make sure that only approved users can operate against the data in the system
Resource Manager - Employs data locality and server resources to determine optimal computing operations
Optimized Scheduling - Completes jobs according to prioritization
Flexibility – Procedures can be written in virtually any programming language
Resiliency & High Availability - Multiple job and task trackers ensure that jobs fail independently and restart automatically

Big Data Analytics for Security This section explains how Big Data is changing the analytics landscape. In particular, Big Data analytics can be leveraged to improve information security and situational awareness. For example, Big Data analytics can be employed to analyze financial transactions, log files, and network traffic to identify anomalies and suspicious activities, and to correlate multiple sources of information into a coherent view. Data-driven information security dates back to bank fraud detection and anomaly-based intrusion detection systems. Fraud detection is one of the most visible uses for Big Data analytics. Credit card companies have conducted fraud detection for decades. However, the custom-built infrastructure to mine Big Data for fraud detection was not economical to adapt for other fraud detection uses. Off-the-shelf Big Data tools and techniques are now bringing attention to analytics for fraud detection in healthcare, insurance, and

other fields. In the context of data analytics for intrusion detection, the following evolution is anticipated:

● 1 st generation: Intrusion detection systems – Security architects realized the need for layered security (e.g., reactive security and breach response) because a system with 100% protective security is impossible.

● 2 nd generation: Security information and event management (SIEM) – Managing alerts from different intrusion detection sensors and rules was a big challenge in enterprise settings. SIEM systems aggregate and filter alarms from many sources and present actionable information to security analysts.

● 3 rd generation: Big Data analytics in security (2nd generation SIEM) – Big Data tools have the potential to provide a significant advance in actionable security intelligence by reducing the time for correlating, consolidating, and contextualizing diverse security event information, and also for correlating long-term historical data for forensic purposes.

II. Motivation and Related Work

2.1. Motivation

Along with the increasing popularity of the Could Computing environments, the security issues introduced through adaptation of this technology are also increasing. Though Cloud Computing offers many benefits, it is vulnerable to attacks. Attackers are consistently trying to find loopholesto attack the cloud computing environment. The traditional security mechanisms which are used are reconsidered because of these cloud computing deployments. Ability to visualize, control and inspect the network links and ports is required to ensure security. Hence there is a need to invest in understanding the challenges, loop holes and components prone to attacks with respect to cloud computing, and come up with a platform and infrastructure which is less vulnerable to attacks.

2.2. Related Work

Hadoop (a cloud computing framework), a Java based distributed system, is a new framework in the market. Since Hadoop is new and still being developed to add more features, there are many security issues which need to be addressed. Researchers have identified some of the issues and started working on this. Some of the notable outcomes, which is related to our domain and helped us to explore, are presented below. The World Wide Web consortium has identified the importance of SPARQL which can be used in diverse data sources. Later on, the idea of secured query was proposed in order to increase privacy in privacy/utility tradeoff. Here, Jelena, of the USC Information Science Institute, has explained that the queries can be processed according to the policy of the provider, rather than all query processing. Bertino et al published a paper on access control for XML Documents [12]. In the paper, cryptography and digital signature technique are explained, and techniques of access control to XML data document is stressed for secured environment. Later on, he published another paper on authentic third party XML document distribution [13] which imposed another trusted layer of security to the paradigm.Kevin Hamlen and et al proposed that data can be stored in a database encrypted rather than plain text. The advantage of storing data encrypted is that even though intruder can get into the database, he or she can't get the actual data. But, the disadvantage is that encryption requires a lot of overhead. Instead of processing the plain text, most of the operation will take place in cryptographic form. Hence the approach of processing in cryptographic form added extra to security layer.

IBM researchers also explained that the query processing should take place in a secured environment. Then, the use of Kerberos has been highly effective. Kerberos is nothing but a system of authentication that has been developed at MIT. Kerberos uses an encryption technology along with a trusted third party, an arbitrator, to be able to perform a secure authentication on an open network. To be more specific, Kerberos uses cryptographic tickets to avoid transmitting plain text passwords over the wire. Kerberos is based upon Needham-Schroeder protocol. Airavat [14] has

shown us some significant advancement security in the Map Reduce environment. In the paper, Roy and et al have used the access control mechanism along with differential privacy. They have worked upon mathematical bound potential privacy violation which prevents information leak beyond data provider's policy. The above works have influenced us, and we are analyzing various approaches to make the cloud environment more secure for data transfer and computation.

III.ISSUES AND CHALLENGES

Cloud computing comes with numerous security issues because it encompasses many technologies including networks, databases, operating systems, virtualization, resource scheduling, transaction management, load balancing, concurrency control and memory management. Hence, security issues of these systems and technologies are applicable to cloud computing. For example,it is very important for the network which interconnects the systems in a cloud to be secure. Also, virtualization paradigm in cloud computing results in several security concerns

There are a number of security risks associated with cloud computing that must be adequately addressed

- **Loss of governance**. In a public cloud deployment, customers cede control to the cloud provider over a number of issues that may affect security. Yet cloud service agreements may not offer a commitment to resolve such issues on the part of the cloud provider, thus leaving gaps in security defenses.
- **Responsibility ambiguity**. Responsibility over aspects of security may be split between the provider and the customer, with the potential for vital parts of the defenses to be left unguarded if there is a failure to allocate responsibility clearly. This split is likely to vary depending on the cloud computing model used (e.g., IaaS vs. SaaS).
- **Authentication and Authorization**. The fact that sensitive cloud resources are accessed from anywhere on the Internet heightens the need to establish with certainty the identity of a user -- especially if users now include employees, contractors, partners and customers. Strong authentication and authorization becomes a critical concern.
- **Isolation failure**. Multi-tenancy and shared resources are defining characteristics of public cloud computing. This risk category covers the failure of mechanisms separating the usage of storage, memory, routing and even reputation between tenants (e.g. so-called guest-hopping attacks).
- **Compliance and legal risks**. The cloud customer's investment in achieving certification (e.g., to demonstrate compliance with industry standards or regulatory requirements) may be lost if the cloud provider cannot provide evidence of their own compliance with the relevant requirements, or does not permit audits by the cloud customer. The customer must check that the cloud provider has appropriate certifications in place.Finally, data mining techniques can be used in the malware detection in clouds. The challengesof security in cloud computing environments can be categorized into network level, user authentication level, data level, and genericissues.
- **Network level**: The challenges that can be categorized under a network level deal with network protocols and network security, such as distributed nodes, distributed data, Internode communication.
- **Authentication level:** The challenges that can be categorized under user authentication level deals with encryption/decryption techniques, authentication methods such asadministrative rights for nodes, authentication of applications and nodes, andlogging.
- **Data level :**The challenges that can be categorized under data level deals with data integrity and

availability such as data protection and distributed data.

- **Generic types:** The challenges that can be categorized under general level are traditional security tools, anduse of different technologies.

3.1 Distributed Nodes

Distributed nodes [15]are an architectural issue. The computation is done in any set of nodes. Basically, data is processed in those nodes which have the necessary resources. Since it can happen anywhere across the clusters, it is very difficult to find the exact location of computation. Because of this it is very difficult to ensure the security of the place where computation is done.

3.2 Distributed Data

In order to alleviate parallel computation,a large data set can be stored in many pieces across many machines. Also, redundant copies of data are made to ensure data reliability. In case a particular chunk is corrupted, the data can be retrieved from its copies. In the cloud environment, it is extremely difficult to find exactly where pieces of a file are stored. Also, these pieces of data are copied to another node/machines based on availability and maintenance operations. In traditional centralized data security system, critical data is wrapped around various security tools. This cannot be applied to cloud environments since all related data are not presentedin one place and it changes.

3.3 Internode Communication

Much Hadoop distributions use RPC over TCP/IP for user data/operational data transfer between nodes. This happens over a network, distributed around globe consisting of wireless and wired networks. Therefore, anyone can tap and modify the inter node communication [15] for breaking into systems.

3.4 Data Protection

Many cloud environments like Hadoop store the data as it is without encryption to improve efficiency. If a hacker can access a set of machines, there is no way to stop him to steal the critical data present in those machines.

3.5 Administrative Rights for Nodes

A node has administrative rights [15] and can access any data. This uncontrolled access to any data is very dangerous as a malicious node can steal or manipulate critical user data.

3.6 Authentication of Applications and Nodes

Nodes can join clusters to increase the parallel operations. In case of no authentication, third part nodes can join clusters to steal user data or disrupt the operations of the cluster.

3.7 Logging

In the absence of logging in a cloud environment, no activity is recorded which modify or delete user data. No information is stored like which nodes have joined cluster, which Map Reduce jobs have run, what changes are made because of these jobs. In the absence of these logs, it is very difficult to find if someone has breached the cluster if any, malicious altering of data is done which needs to be reverted. Also, in the absence of logs, internal users can do malicious data manipulations without getting caught.

3.8 Traditional Security Tools

Traditional security tools are designed for traditional systems where scalability is not huge as cloud environment. Because of this, traditional security tools which are developed over years cannot be directly applied to this distributed form of cloud computing and these tools do not scale as well as the cloud scales.

3.9 Use of Different Technologies

Cloud consists of various technologies which has many interacting complex components. Components include database, computing power, network, and many other stuff. Because of the wide use of technologies, a small security weakness in one component can bring down the whole system. Because of this diversity, maintaining security in the cloud is very challenging.

IV.THE PROPOSED APPROACHES

We present various security measures which would improve the security of cloud computing environment. Since the cloud environment is a mixture of many different technologies, we propose various solutions which collectively will make the environment secure. The proposed solutions encourage the use of multiple technologies/ tools to mitigate the security problem. specified in previous sections. Security recommendations are designed such that they do not decrease the efficiency and scaling of cloud systems.

Following security measures should be taken to ensure the security in a cloud environment.

4.1 File Encryption

Since the data is present in the machines in a cluster, a hacker can steal all the critical information. Therefore, all the data stored should be encrypted. Different encryption keys should be used on different machines and the key information should be stored centrally behind strong firewalls. This way, even if a hacker is able to get the data, he cannot extract meaningful information from it and misuse it. User data will be stored securely in an encrypted manner.

4.2 Network Encryption

All the network communication should be encrypted as per industry standards. The RPC procedure calls which take place should happen over SSL so that even if a hacker can tap into network communication packets, he cannot extract useful information or manipulate packets.

4.3 Logging

All the map reduce jobs which modify the data should be logged. Also, the information of users, which are responsible for those jobs should be logged. These logs should be audited regularly to find if any, malicious operations are performed or any malicious user is manipulating the data in the nodes.

4.4 Software Format and Node Maintenance

Nodes which run the software should be formatted regularly to eliminate any virus present. All the application softwares and Hadoop software should be updated to make the system more secure.

4.5 Nodes Authentication

Whenever a node joins a cluster, it should be authenticated. In case of a malicious node, it should not be allowed to join the cluster. Authentication techniques like Kerberos can be used to validate the authorized nodes from malicious ones.

4.6 Rigorous System Testing of Map Reduce Jobs

After a developer writes a map reduce job, it should be thoroughly tested in a distributed environment instead of a single machine to ensure the robustness and stability of the job.

4.7 Honeypot Nodes

Honey pot nodes should be present in the cluster, which appear like a regular node but is a trap. These honeypots trap the hackers and necessary actions would be taken to eliminate hackers.

4.8 Layered Framework for Assuring Cloud

A layered framework for assuring cloud computing [16] as shown in Figure (1) consists of the secure virtual machine layer, secure cloud storage layer, secure cloud data layer, and the secure virtual network monitor layer. Cross cutting services are rendered by the policy layer, the cloud monitoring layer, the reliability layer and the risk analysis layer.

4.9 Cloud Security Guidance

As customers transition their applications and data to the cloud, it is critical for them to maintain, or preferably surpass, the level of security they had in their traditional IT environment.
This section provides a prescriptive series of steps for cloud customers to evaluate and manage the security of their use of cloud services, with the goal of mitigating risk and delivering an appropriate level of support. The following steps will be discussed in detail below:

1) Ensure effective governance, risk and compliance processes exist
2) Audit operational and business processes
3) Manage people, roles and identities
4) Ensure proper protection of data and information
5) Enforce privacy policies
6) Assess the security provisions for cloud applications
7) Ensure cloud networks and connections are secure
8) Evaluate security controls on physical infrastructure and facilities
9) Manage security terms in the cloud service agreement
10) Understand the security requirements of the exit process.

V. CONCLUSION

Cloud environment is widely used in industry and research aspects; therefore security is an important aspect for organizations running on these cloud environments. Using proposed approaches, cloud environments can be secured for complex business operations. Cloud computing is clearly one of the most enticing technology areas of the current times due, at least in part to its cost-efficiency and flexibility. However, despite the surge in activity and interest, there are significant, persistent concerns about cloud computing that are impeding the momentum and will eventually compromise the vision of cloud computing as a new IT procurement model. Despite the trumpeted business and technical advantages of cloud computing, many potential cloud users have yet to join the cloud, and those major corporations that are cloud users are for the most part putting only their less sensitive data in the cloud. Lack of control is transparency in the cloud implementation – somewhat contrary to the original promise of cloud computing in which cloud implementation is not relevant. Transparency is needed for regulatory reasons and to ease concern over the potential for data breaches. Because of today's perceived lack of control, larger companies are testing the waters with smaller projects and less sensitive data. In short, the potential of the cloud is not yet being realized.

REFERENCES

1. Ren, Yulong, and Wen Tang. "A SERVICE INTEGRITY ASSURANCE FRAMEWORK FOR CLOUDCOMPUTING BASED ON MAPREDUCE."Proceedings of IEEE CCIS2012. Hangzhou: 2012, pp 240 – 244, Oct. 30 2012-Nov. 1 2012
2. N, Gonzalez, Miers C, Redigolo F, Carvalho T, Simplicio M, de Sousa G.T, and Pourzandi M. "AQuantitative Analysis of Current Security Concerns and Solutions for Cloud Computing.". Athens:2011., pp 231 – 238, Nov. 29 2011- Dec. 1 2011
3. Hao, Chen, and Ying Qiao. "Research of Cloud Computing based on the Hadoop platform.". Chengdu, China: 2011, pp. 181 – 184, 21-23 Oct 2011.
4. Y, Amanatullah, Ipung H.P., Juliandri A, and Lim C. "Toward cloud computing reference architecture: Cloud service management perspective.". Jakarta: 2013, pp. 1-4, 13-14 Jun. 2013.
5. A, Katal, Wazid M, and Goudar R.H. "Big data: Issues, challenges, tools and Good practices.". Noida:2013, pp. 404 – 409, 8-10 Aug. 2013.

6. Lu, Huang, Ting-tin Hu, and Hai-shan Chen. "Research on Hadoop Cloud Computing Model and its Applications.". Hangzhou, China: 2012, pp. 59 – 63, 21-24 Oct. 2012.
7. Wie, Jiang , Ravi V.T, and Agrawal G. "A Map-Reduce System with an Alternate API for Multi-core Environments.". Melbourne, VIC: 2010, pp. 84-93, 17-20 May. 2010. International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.3, May 201456
8. K, Chitharanjan, and Kala Karun A. "A review on hadoop — HDFS infrastructure extensions.". JeJu Island: 2013, pp. 132-137, 11-12 Apr. 2013.
9. F.C.P, Muhtaroglu, Demir S, Obali M, and Girgin C. "Business model canvas perspective on big data applications." Big Data, 2013 IEEE International Conference, Silicon Valley, CA, Oct 6-9, 2013, pp.32 - 37.
10. Zhao, Yaxiong , and Jie Wu. "Dache: A data aware caching for big-data applications using the MapReduce framework." INFOCOM, 2013 Proceedings IEEE, Turin, Apr 14-19, 2013, pp. 35 - 39.
11. Xu-bin, LI , JIANG Wen-rui, JIANG Yi, ZOU Quan "Hadoop Applications in Bioinformatics." Open Cirrus Summit (OCS), 2012 Seventh, Beijing, Jun 19-20, 2012, pp. 48 -52.
12. Bertino, Elisa, Silvana Castano, Elena Ferrari, and Marco Mesiti. "Specifying and enforcing access control policies for XML document sources." pp 139-151.
13. E, Bertino, Carminati B, Ferrari E, Gupta A , and Thuraisingham B. "Selective and Authentic Third- Party Distribution of XML Documents."2004, pp. 1263 - 1278.
14. Kilzer, Ann, Emmett Witchel, Indrajit Roy, Vitaly Shmatikov, and Srinath T.V. Setty. "Airavat: Security and Privacy for MapReduce."
15. "Securing Big Data: Security Recommendations for Hadoop and NoSQL Environments."Securosis blog, version 1.0 (2012)
16. P.R , Anisha, Kishor Kumar Reddy C, Srinivasulu Reddy K, and Surender Reddy S. "Third Party Data Protection Applied To Cloud and Xacml Implementation in the Hadoop Environment With Sparql."2012. 39-46, Jul – Aug. 2012.
17. "Security-Enhanced Linux."Security-Enhanced Linux. N.p. Web. 13 Dec 2013.

