Data Analytics and Security in Cloud

ROHIT RAJ[1]
[1] *Department of Computer Engineering, Poornima College of Engineering, Sitapura, Jaipur*

Abstract -- The advent of the digital age has led to a rise in different types of data with every passing day. This data is complex and needs to be stored, processed and analyzed for information that can be used by organizations. Cloud computing provides an apt platform for big data analytics in view of the storage and computing requirements of the latter. We discuss various possible solutions for the issues in cloud computing security and Hadoop. Big data analytics use complex data mining algorithm that require efficient high performance processors. Cloud computing infrastructure is able to provide both computational and data processing applications and also offers elasticity, pay-per- user, low affordable investment. Challenges in data migration on cloud are scalable data management, application security, Map reduce & Hadoop environment.

Index Terms- Big Data, Cloud Computing, Kerberos, Altryx

I. INTRODUCTION

The generation of massive amounts of pervasive and complex data, which needs to be efficiently created, stored, shared and analyzed to extract useful information has huge potential, ever-increasing complexity, insecurity and risk. The requirement of an efficient and effective analytics service, application, programming tool and framework has given birth to the concept of Big Data Processing and Analytics.

Big data analytics examines large amounts of data to uncover hidden patterns, correlations and other insights. With today's technology, it's possible to analyze data and get answers from it almost immediately – an effort that's slower and less efficient with more traditional business intelligence solutions.

Big data analytics has found application in several domains and fields. Some of these applications include medical research, solutions for the transportation and logistics sector, global security and prediction and management of issues concerning the socio- economic and environmental sector.

Enterprises are gathering more data, at a faster pace, than ever before and companies are increasingly looking to tap into the potential of these vast swathes of fast-moving, unstructured and complex streams of data to achieve step-change improvements in growth and performance. Big Data is much more than simply a matter of size – it presents an opportunity to discover key insights and emerging trends in data, make business more agile. The cloud computing environment offers development, installation and implementation of software and data applications 'as a service'. Three multi- layered infrastructures namely, platform as a service (PaaS), software as a service (SaaS), and infrastructure as a service (IaaS), exist. Infrastructure-as-a-service is a model that provides computing and storage resources as a service. On the other hand, in case of PaaS and SaaS, the cloud services provide software platform or software itself as a service to its clients. The cost of storage has considerably reduced with the advent of cloud-based solutions. In addition, the 'pay-as-you-go' model and the concept of commodity hardware allow effective and timely processing of large data, giving rise to the concept of 'big data as a service'. An example of one such platform is Google BigQuery, which provides real time insights from big data in the cloud environment.

With the growth of structured, unstructured, and semi-structured data there has never been a better time to drive improvements in customer engagement, process performance, and strategic decision-making. The challenge is that most solutions for Big Data analytics require either an army of specialists with Ph.D.'s and advanced computing degrees, or focusing on a single source of data such as Hadoop. Alteryx eliminates this challenge by delivering a platform for self-service data analytics that puts the value of Big Data in the hands of all analysts and decision makers. Alteryx gives organizations the power to take the advantage of all data inside Big Data environments and combine it with external datasets to derive the maximum value of all available data sources.

- Prepare and blend data inside and outside Big Data environment in a repeatable workflow that eliminates coding and accelerates the time to insights
- Build sophisticated but accessible predictive, statistical and spatial analytic in a simple and intuitive, GUI-based, workflow design environment
- Simplify sharing of Big Data analytics via outputs to Qlik or Tableau, or interactive analytic apps that can be used by any decision maker

A field of major concern is Big Data Analytics Security. Not only security but also data privacy challenges existing industries and federal organizations. With the increase in the use of big data in business, many companies are wrestling with privacy issues. For marketing and research, many of the businesses uses big data, but may not have the fundamental assets particularly from a security perspective. Data privacy is a liability, thus companies must be on privacy defensive. But unlike security, privacy should be considered as an asset, therefore it becomes a selling point for both customers and other stakeholders. There should be a balance between data privacy and national security. All data security issues are caused by the lack of effective measures provided by antivirus software and firewalls. These systems were developed to protect the limited scope of information stored on the hard disk, but Big Data goes beyond hard disks and isolated systems. Of immediate concern to companies using Big Data is the security of cloud-based systems. Intel Security has recently published the McAfee Labs' Threat Predictions Report that contains their expectations for the near-future of data security. Of particular concern in this report is the supposition that legitimate cloud file hosting services such as Drop box and Stream Nation, are at risk of being used as control servers in upcoming cyber espionage campaigns. If targeted, these popular cloud services could enable the malware to transfer commands without raising suspicion.

Malicious attacks on IT systems are becoming more complex and new malware is constantly being developed. Unfortunately, companies that work with Big Data face these issues on a daily basis. The challenge of detecting and preventing advanced threats and malicious intruders, must be solved using big data style analysis. These techniques help in detecting the threats in the early stages using more sophisticated pattern analysis and analyzing multiple data sources. In many organizations, the deployment of big data for fraud detection is very attractive and useful.

II. CHALLENGES IN DATA ANALYTICS SECURITY

Some Data Analytics Security Challenges are:

1. Mostd is tribute systems computations have only a single level of protection, which is not recommended.
2. Non-relational databases (NoSQL) are actively evolving, making it difficult for security solutions to keep up with demand.
3. Automated data transfer requires additional security measures, which are often not available.
4. When a system receives a large amount of information, it should be validated to remain trustworthy and accurate; this practice doesn't always occur, however.
5. Unethical IT specialists practicing information mining can gather personal data without asking users for permission or notifying them.
6. Access control encryption and connections security can become dated and inaccessible to the IT specialists who rely on it.
7. Some organizations cannot – or do not institute access controls to divide the level of confidentiality within the company.
8. Recommended detailed audits are not routinely performed on Big Data due to the huge amount of information involved.

III. SOLUTIONS TO IMPROVE SECURITY IN DATA ANALYTICS

Cloud computing experts believe that the most reasonable way to improve the security of Big Data is through the continual expansion of the antivirus industry. A multitude of antivirus vendors, offering a variety of solutions, provides a better defense against Big Data security threats.

Refreshingly, the antivirus industry is often touted for its openness. Antivirus software providers freely exchange information about current Big Data security threats, and industry leaders often work together to cope with new malicious software attacks, providing maximum gains in Big Data security.

Here are some additional recommendations to strengthen Big Data security:

- Focus on application security, rather than device security.
- Isolate devices and servers containing critical data.
- Introducereal – time security information and event management.
- Provide reactive and proactive protection.

IV. DATA SECURITY

Modern computer systems provide service to multiple users and require the ability to accurately identify the user making a request. The process of verifying the user's identity is called authentication. Today, most common computer network architecture is a distributed architecture consisting of dedicated user workstations (clients) and distributed or centralized servers. In this environment, the network connection to other machines is provided. Thus, we need to protect user information and resources kept at the server. The authentication service in such environment can be achieved by using Kerberos Security Protocol.

I t is one of the most widely used authentication protocols. It addresses an open distributed environment in which users at workstations wish to access the services on the server. Kerberos employs one or more

Kerberos servers (the KDC: Kerberos Distribution Center) to provide an authentication service. Kerberos requires the user to prove his or her identity for every service invoked. It also requires that server prove it's identity to the clients. The overall scheme of Kerberos is a trusted third party that uses a protocol proposed by Needham and Schroeder. It is trusted in the sense that clients and servers trust Kerberos to mediate their mutual authentication. Assuming the Kerberos protocol is well designed, then the authentication service is secure if the Kerberos server itself is secure. Kerberos provides a centralized authentication server whose function is to authenticate users to servers and servers to users.

Kerberos relies exclusively on symmetric encryption, making no use of public-key encryption. Most of the secure routing protocols rely on public key infrastructures (PKI) to authenticate communicating nodes. Although PKI is secure, it is based on asymmetric cryptography and hence requires excessive processing and communication resources. This resource hungry feature makes PKI based systems more susceptible to Denial of Service attacks. In contrast, Kerberos is a symmetric key based authentication mechanism making a request.

In 2007, MIT formed the Kerberos Consortium along with some of the major vendors and users of Kerberos such as Sun Microsystems, Apple, Google, Microsoft, etc., to foster continued development. The MIT Kerberos Consortium was created to establish Kerberos as the universal authentication platform for the world's computer networks. Kerberos has grown to become the most widely deployed system for authentication and authorization in modern computer networks. Kerberos is currently shipped with all major computer operating systems and is uniquely positioned to become a universal solution to t h e d i s t r i b u t e d a u t h e n t i c a t i o n a n d authorization problem of communicating parties.

V. KERBEROS MESSAGES EXCHANGE

Kerberos Message Actions is shown in Fig. 1. The exchange between the client and the Kerberos AS (Authentication Server) in messages 1 and 2 are used only when the user first logs in to the system. Exchange between the client and the Kerberos TGS (Ticket Granting Server) in messages 3 and 4 are used whenever a user authenticates to a new server. Message 5 is used each time the user authenticates itself to a server. And finally, message 6 is the mutual-authentication response by the server. The ticket plus the secret session key are the user credentials to be authenticated to a specific server.

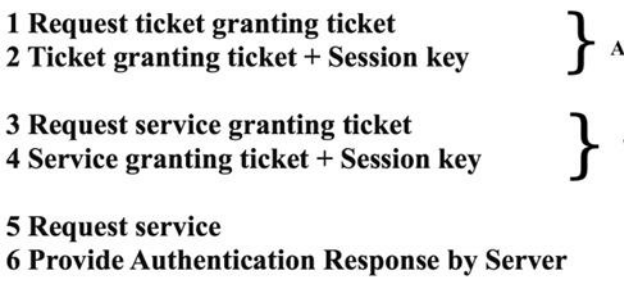

Fig. 1. Kerberos Message Actions

VI. FUTURE ASPECTS

The increase in the size of data is one the major concern of today's world. The storage of this large amount of data is provided in Hadoop's HDFS (Distributed File System) and computation on this data is done by either Map Reduce (batch processing) or Apache Spark (real time processing). The scope in the field of business analytics is ever expanding and is helping it become mainstream as companies of all sizes and analytics skill levels get into the big data game Exploring business analytics needs the right focus, right technology, right people, right culture and top management commitment. Companies like IBM, Accenture, and Deloitte are using business analytics tools and coming up with decisions that are useful and profitable. Also, the researchers are still researching ways to implement scalability in Kerberos. Scalability will allow Kerberos to be implemented in large networks with more secure environment.

VII. CONCLUSION

Big Data is an issue that requires huge storage and complex computing environment which is achieved by Hadoop and Spark. Cloud is a suitable platform for dealing with this data.The generation of massive amount of data gave rise to the concept of Data Analytics. 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.

If Enterprises follow these security privacy policy then they can safely keep their data on cloud. Kerberos is a symmetric key based authentication mechanism that helps in protecting the data to a greater extent. Kerberos' many strengths are attested to by it's wide adoption as an industry standard. It seems certain to continue to play a role in small networks with strict authentication requirements. Modifications to Kerberos and research on extending Kerberos are still active. Even with technical improvements in its scalability, Kerberos may remain restricted to the small network context.

REFERENCES

[1] Venkata Narasimha Inukollu, Sailaja Arsi And Srinivasa Rao Ravuri. "Security Issues Associated With Big Data In Cloud Computing, International Journal Of Network Security & Its Applications (Ijnsa), Vol.6, No. 3, May 2014.

[2] Amitkumar Manekar And Dr. G. Pradeepini. "A Review On Cloud Based Big Data Analytics.".Icses Journal On Computer Networks And Communications(Ucnc),May 2015,Vol. 1, No.1

[3] S. Mahdi Shariati,Abouzarjomehri, M.Hossein Ahmadzadegan. "Challenges And Security Issues In Cloud Computing From Two Perspectives: Data Security And Privacy Protection" 2015 2Nd International Conference On Knowledge-Based Engineering And Innovation(Kbei) November 5-6, 2015.

[4] http://www.kerberos.org

[5] Boldyreva And V. Kumar, "Provable-Security Analysis Of Authenticated Encryption In Kerberos". Ieee Symposium On Security And Privacy (Sp'07). May 2007.