_____

Big- Data Security

Akriti Yadav
Sardar Patel Institute Of Technology,Mumbai

Abstract:-Big data is a term that describes the large volume of data – both structured and   unstructured – that inundates a business on a day-to-day basis or in other words big data is a collection of data sets which is very large in size as well as complex.

As the internet & technology is growing, amount of big data continue to grow. Big Data has been regularly collected and utilized by many organizations for several decades.This era of "big data" has ushered different opportunities to advance science, improve health care, promote economic growth, reform our educational system, and create new forms of social interaction and entertainment but at the same time it lead to availability of data in cloud which may be misused by anyone. The Challenges in security and privacy concerns are growing as big data becomes more and more accessible. Due to all this the sensitive data may get leaked which may be harmful for any organisation or institution. This paper mainly contains and deals with how to work and use big data in a cloud with the security layers to protect the data.

Key Words: - *ELP, Data Authentication, Data Mining*
_____ *********** _____.

I. INTRODUCTION

Big-Data can be defined as the large amounts of digital information companies and governments collect about human beings and the environment we are living in. The amount of data that is getting generated is expected to double every two years, which will surely lead to huge collection of data with obvious risks of its security.Big data not only provides an enormous competitive advantage for corporations, helping businesses tailor their products to consumer needs, but also identify and minimize corporate inefficiencies, and share data with user groups across the enterprise.

So, the big data of organisations which is of great importance to them is at risk, when not properly handled. Cyber Attackers try to breach organization's big data repository to get bigger payoffs and more recognition from a single attack. The data can be personal or can be of great importance for an organisation. Hence, it is required to secure big-data from unauthorized persons & attacks by applying different security techniques.

The security management challenges that were addressed in traditional security management are the variety, velocity & volume of data. Since the big data repositories include information deposited by various sources across the organisations. These variety of data gathered makes the secure access management a challenge. The term velocity means here the speed of generation of data or how fast the data is generated and processed to meet the requirements and the challenges for development & growth. Last one, volume refers here as the quantity of data that is generated which is actually very big in size and leading to the increase in big-data day by day. Thus security is an important concern to be dealt with.
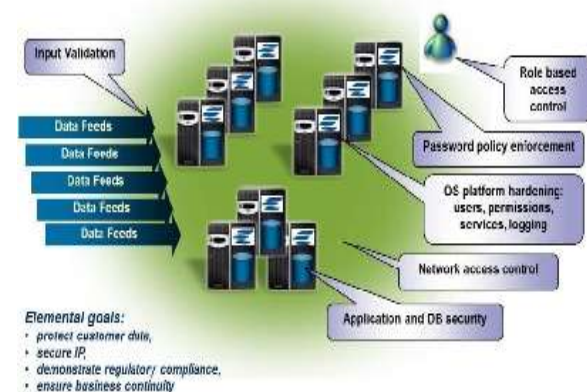
II. BASIC WAYS TO PROVIDE

SECURITY (Elemental Security Platform):-

At initial level, when the data we are dealing with is confidential & related to many aspects should be protected using the following protection schemes. As for our use, we keep our data on cloud which is available to all withminimum efforts to get.

1) Password protection:- Since big data don't have any passwords to protect, so we can use "Elemental security platform", to protect data with multiple layers.

2) Role-based access control:- By this we can define that the data at repository can be accessed by whom providing some security wall.

3) OS Hardening:- OS hardening based on the operating system used, will lock down the data & allow protection focus on the four areas : users, permissions, services, logging.

Elemental provides multilevel protection and deep visibility

III. SECURITY ISSUES RELATED WITH BIG-DATA:-

Since different technologies are being used to store, retrieve and process the data & on the basis of that, we can categorize four main security issues of big-data.
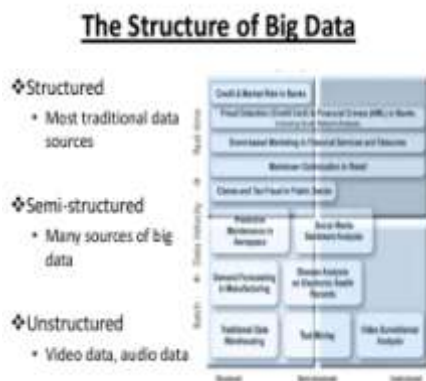
_____

_____

Authentication level:-
There are number of clusters and nodes present in an organisation's tree. Each node with different priority can access data at their level & at the same time administrative rights can access any data. Logging is an important feature for any data under susceptance. Without login facility we can't figure out any user's interference with the data.

Data level:-
Data is nothing but some important and personal information about someone or something. Data level issues deals with data integrity and availability such as data protection to improve efficiency of big data environments like Hadoop, store the data as it is without encryption. In distributeddata store, information is stored in many nodes with replicas.

Network level:-
This level mainly deals with processing of data at different nodes. The problem arises when; it is not not known which data has to be protected. Two or more nodes can communicate with each other or share their data/resources through network, making it tough to understand the situation.

General level:-
Big data environment uses many new technologies for processing the data along some traditional security tools for security purposes.There may occur complexities with these technologies for data storing, data processing and data retrieval.

IV.APPROACHES TO SOLVSECURITY

Data encryption:-
The approach is for dealing with data level issues. Data encryption is converting data into secret message using encryption algorithms. Different algorithms used for encryption are RSA, DES algorithm. These algorithms use private keys to encrypt data and receivers use their keys to decrypt

the data which protects the important and sensitive information.

Network encryption:-
This deals with network level issues. As data or message is sent over the network for communication, network must be encrypted. If network is encrypted with appropriate industry standards then hacker will not able to crack the network. So to make that secure, RPC and SSL are used with their specifications.

Logging:-
This approach works towards for authentication level issues. Logging keeps account of recording the logs & for maintaining the changes in data. The details of modifying the data get capture.

V.Algorithm:-

Data Acquisition:-
Firstly the data is gathered from different sources. These enormous amounts of data are saved in smart devices which are made to store large amount of data, taking minimum spaces.

Fig-2:Data Acquisition

Data Extraction:-
All the data that is gathered is of not use because of redundancy and unimportant data for that different techniques are used to discard the useless data.

Fig-3: Data Extraction

Data Structuring:-
Once the data is extracted, there is a need to arrange data in a proper format so as to use it with ease. Because the unstructured data is more complex to understand and we can't fire any queries on that.

_____

_____

Fig-4:Data Structuring

*Data Visualization***:-**

Now on the structured data we fire queries, and get the visual representation of data. These are the actual insights of data in the form of patterns.

Fig-5: Data Visualization

*Data Interpretation***:-**

The important step in all is interpreting and gaining the most important data from the cluster of data. At this stage we try to get the useful data and discarding the un-useful data.

Securing Data:-

This is an important step which in itself is a big-process. Here we protect the data that is of sensitive nature to anyone or any organisation with different ways discussed above. It is one of the important concerns we are dealing with in present times with the increase of data day by day.

Fig-6: Data Visualization

RESULTS:-

By doing the above research regarding big-data and it's security, the results that can be extracted are –

1. The whole process follows a specific sequence of collecting data from various sources to get the valuable output. So, the security steps needs to be followed from initial level.
2. If security approaches are not followed, it may lead to unwanted results which may cause significant effects for any organization.
3. The increase in data continuously is leading to increase in securityproblems which may arise more in future.   But if proper methods of discovering data is applied, will reduce the problem.

CONCLUSION:-

Big-Data is continuously changing the way, we perceive the world. So, our focus is mainly on Protecting the data, thus making them secure at different levels of use. Elemental Security Platform is one of the important ways to understand how the data flow should be there to secure it. Thus, it's our data we are dealing with, so it's ourresponsibility to use proper methodologies and techniques to keep it safe.

REFERENCES

[1] www.wikipedia.com
[2] "Big Data for Development: Challenges and Opportunities", Global Pulse, May 2012
[3] "Big-data security and privacy" by Elisa Bertino.
[4] Xingdong Wu, Wing-Ding, "Data mining with Big-Data ".
[5] "Big Data security"byKalyaniShirodkar.

_____