ISSN: 0474-9030

Our Heritage

Vol-68-Issue-15

January-2020

UGC Care Listed Journal

Securities to Big Data Analytics

Ms.Swati Jadhav[1],Dr.Manisha Kumbhar[2]

Assistant Professor,Institute of Business Management & Research ,Chinchwad.

Professor,Sinhgad Institute of Management,Pune

Abstract:

Big Data analysis and Big Data analytics are very useful concepts in current computing trends or environments. As the increasing volume of massive data of various social sites or web sites it is very difficult to maintain analyse data as well as to collect knowledge from data. Securities and privacy to maintain the Big Data is very essential. Basically Big Data generally available on clouds. It is necessary to take care while uploading data on clouds.

Keywords: *Security, Approaches, Engineering, Cloud environment, NoSQLDatabase.*

Introduction

Big data is a word used for detailed information of massive amounts of data that are structured, semi-structured or unstructured. Big Data generally is not handled by traditional database software technologies.Users of Big Data can keep away to intruders by applying or providing Firewall services as well as Intrusion Detection and Prevention systems.All the enterprises handled Big Data by their web sites and are not structured since it contains figures, numerical data, Images comments, Email attachment etc. and are in unstructured data.Five**V**'s are plays very important role in Big Data Analytics.

Volume: It includes storage of data, blogs, emails, You tube audio video streaming etc. .
Variety: It consist of types of data which is supported.
Velocity:In how much time the files are created and running processes are carried out.
Veracity: Reliability of data testing is done
Value : Big Data is not in fix quantity is fast moving and fast growing.

Security Technology in Big Data Environment

It is essential to protect Big Data from intruders. A single ransomware attack may destroy a Big Data. To protect the privacy of massive data though the info with privacy leaks, the attacker can't obtain the effective value of knowledge. We will use encoding and Data anonymity technology

(1) Encoding Technology Data encryption technology is a crucial means to guard data confidentiality, it safeguards the confidentiality of the info, but it hampers the performance of the system at an equivalent time. the info processing ability of a enormous information system is fast and efficient, which may satisfy the wants of the hardware and software required for encryption. Therefore the homomorphic encryption has become a search hotspot in data privacy protection.

The homomorphic encryption may be a model for the calculation of the ciphertext, avoiding the encryption and decryption within the unreliable environment, and directly operate on the ciphertext. . Homomorphic encryption remains within the exploratory stage, the algorithm is immature, low efficiency, and there's a particular distance faraway from practical application.

(2) Data Anonymity Technology Data anonymity is another important technology for privacy protection, though the attacker gets the info that contains the privacy, he can't get the first exact data, because the worth of the key field is hidden. However, within the background of massive data, the attacker can obtain data from multiple sources, then associate the info from one source with another source, they're going to find the first meaning of the hidden data.

Approaches to privacy preservation storage on cloud

Mainly have three dimensions, confidentiality, integrity and availability. The first two are directly related to the privacy of the data i.e. if data confidentiality or integrity is related it will have a direct effect on user's privacy. The availability of information refers to ensuring that authorized parties can access the information when needed. Big Data will stored in cloud by providing security like encryption. In this approach a sender can send data using pubic key encryption (PKE) and a authorised user can receive it.The following are approaches to protect user privacy when data is being stored on the cloud.Attribute-based encryption -Access control is based on the identity of a user complete access to overall resources.Storage path -Encryption secure data of cloud.

Big Data Engineering

Big Data Engineering collects all data from horizontal scalable servers.New engineering techniques in the data layer have been driven by the increasing importance of data types that cannot be handled efficiently in traditional relational models.

1. Non-Relational refers to logical data models such as document, graph, key-value and others that are used to provide more efficient storage and access to nontabular data sets.
2. NoSQL (alternately called “no SQL” or “not only SQL”) refers to data stores and interfaces that are not tied to strict relational approaches.
3.Data Modelling is used for sorting and storing data. In big enterprises it need to store data in huge form we may say this is Big Data. If we select appropriate model to maintain Big
Data it offers following benefits.

- **Performance:** Good Data model are useful to extract complex data more easily as well as to maintain their performance more accurately.
- **Cost:** Such models reduce speed, storage and computing cost and reuse results for Big Data system.
- **Efficiency:** Good data models increase the efficiency
- **Quality:** Good data models maintain the consistency and reduce the possibility of any errors.

Many enterprises are gathering their data through different sources and also work in parallel. Due to models availability it results in parallel across distributed data from one or more data sources.

Precaution of Analysis and Analytics of Big Data:

This looks like any network security strategy. Yet, big data environments add another level of security because security tools must operate during three data stages that aren't all present within the network. These are 1) data ingress (what’s coming in), 2) stored data (what’s stored), and 3) data output (what’s going bent applications and reports).

Stage 1: Data Sources. There are various sources of Big Data. User generated data can include email messages, transactions of user, CRM data all this data is unstructured data. Security is essential while transit to source to platform.

Stage 2: Stored Data. Secure stored data by various security toolset like encryption, user authentication, Intrusion Detection etc.

Stage 3:Output Data. The whole reason for the complexity and cost of big data platforms is to drive massive data volumes and semantic analytics across different types of data. This analysis gives results on applications, reports, and dashboards. This very valuable intelligence makes it a rich target for intrusion, and the output is difficult to integrate as well. Also, secure compliance at this point: Make sure that the results to end-users do not contain regulated data.

Big Data and Cloud environment

Different services are run under different clouds in clustered form. There are three models of cloud services Software as a Service (SaaS), Infrastructure as a Service (IaaS), Platform as a Service (PaaS),

Figure 1: Cloud Services

NoSQL Database Encryption and Security

Rapid NoSQL Database Adoption

Business industry and Enterprises rapidly adopting NoSQL databases due to growing amount of data (Big Data). It is very beneficial to organization to extract important as well as intelligent data.

Cybercriminals Target NoSQL Databases

As many popular enterprises like Amazon, Google ,eBay, Facebook, , LinkedIn, Mozilla, Netflix and Twitter are maintaining Big Data in their Databases therefore criminals are targeting these databases .

Security provision for Organizations

Organization which is maintained or stored their data using NoSQL must provide any other geographical security.

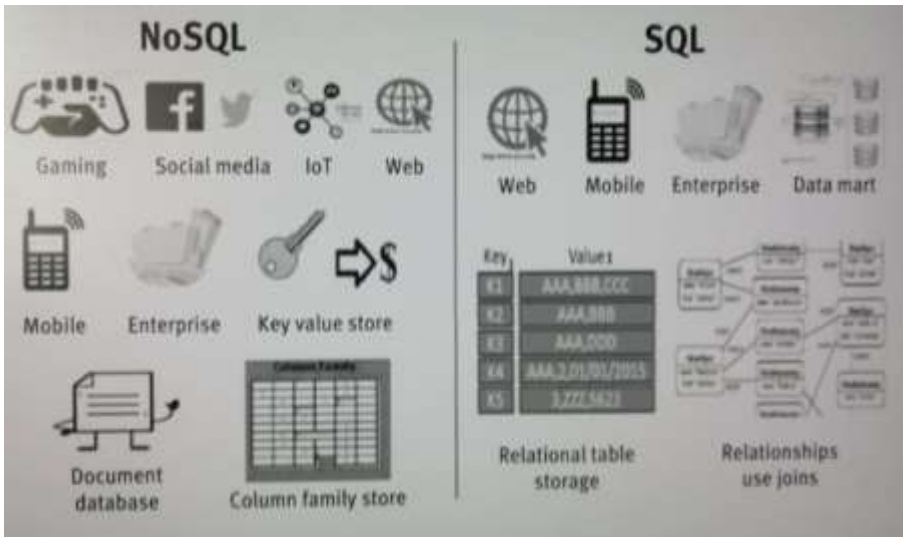

Figure 2: Difference between SQL and NoSQL

Following is the difference between SQL and NoSQL which works with structured and non-structured data simultaneously.

Characteristic	SQL	NoSQL
Data Storage	Information is stored in Table format. Each row contains data items.	Data is not stored in Table format. It is stored in different formats of databases as Text documents, graphs, etc.

Schemas	To alter Schema definition is very complex once table is created	Information can be changed easily as compared with relational databases
Scalability	Due to vertical scaling it is possible to scale A RDBMS across multiple servers and is time consuming.	Due to horizontal scaling More servers can be added to increase the performance.
Integrity Compliance	Existence of ACID properties	Nonexistence of ACID properties

Table 1. Differences between SQL and NoSQL

NoSQL query language:This query language is inspired by MongoDB.A query consists of these parts:
fields to be extractedtable to extract the records fromexpression for filtering the table rows
group by - fields to group the data underaggregate functions to be applied to columns in fields
order by - fields to order the return data bylimit - an integer number of records to return.

Document oriented Databases:
A document-oriented database or a NoSQL document store is a new way to store data in JSON format rather than simple rows and columns. It allows you to express data in its original form the way it's meant to be.With such problems faced by data-intensive and fast-moving organizations, new technology solutions were demanded and the answer is NoSQL Document Databases. In contrast to rows and columns, NoSQL databases keep data in documents. These documents follow a minimum of standard format rules. The format used could be JSON, XML, YAML, etc. The JSON format is the format of choice for NoSQL databases and good reason. A JSON document is simply more compressed, simple and readable

Conclusion:
Big Data processing can be done using NoSQL , massive amount of data available in the form of paragraph or in text documents . It is very difficult to sort and search useful data for accurate and concrete information. To store data is in encryption form is very essential.

References:
[1] SanehLata, Yadav , Asha Sohal," Big Data Analytics in Cloud Computing ",International Journal of Computer Trends and Technology (IJCTT) – Volume 49 Number 3 July 2017 ISSN: 2231-2803 http://www.ijcttjournal.org Page 156 .
[2] Samir A. El-Seoud, Hosam F. El-Sofany, Mohamed Abdelfattah, RehamMohamed,"Big Data and Cloud Computing: Trends and Challenges", https://doi.org/10.3991/ijim.v11i2.6561, iJIM – Vol. 11, No. 2, 2017.
[3] Hossain Shahriar1 ,Hisham M. Haddad2," Security Vulnerabilities of NoSQL and SQL Databases for MOOC Applications, International Journal of Digital Society (IJDS), Volume 8, Issue 1, March 2017.
[4] Gang Zeng," Big Data and Information Security",ISSN (e): 2250 – 3005 , Volume, 05 , Issue, 06 , June – 2015 ,International Journal of Computational Engineering Research (IJCER).
[5] Priyank Jain, Manasi Gyanchandani & NilayKhare," Big data privacy: a technological perspective and review",25 (2016) doi:10.1186/s40537-016-0059-y

[6] Christine Taylor, " Big Data Security", June 2017,https://www.datamation.com/big-data/big-data-security.html

Web-sites

1. https://www.thalesesecurity.com/solutions/use-case/data-security-and-encryption/database-security/nosql-encryption
2. https://wikibon.com/wikibons-2018-big-data-analytics-trends-forecast/