

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

## Analysis of Security in Big Data

P.Tamil Arasan<sup>1</sup>, Dr.A.Valarmathi<sup>2</sup>

PG Scholar, Department of Computer Applications, Anna University, BIT-Campus, Trichy, India<sup>1</sup>

Head of the Department, Department of Computer Applications, Anna University, BIT-Campus, Trichy, India<sup>2</sup>

**ABSTRACT:** All the source of data collection and analysis of data provided in various formats (raw data from sensors, imagery, text-based from human reports, etc.) and distributed across heterogeneous data stores. The advance incensing technologies, the acquisition of new sensors that use of mobile devices result in the production of an overwhelming amount of sensed data that augment the challenges to transform these raw data into useful, actionable intelligence in a timely manner. Leveraging recent advances in data integration and Big Data technologies, we are adapting key concepts of unified data spaces for the design and implementation of intelligence data integration platform Data Integration Services. The development of this scalable data integration platform rests on the layered data space approach, makes use of recent Big Data technologies and leverages. On to logical models, and semantic-based analysis services developed for various purposes as part of the semantic layer.

**KEYWORDS:** big data, data integration

### I. INTRODUCTION

#### 1.1. Big Data:

Big data, a general term for the massive amount of digital data being collected from all sorts of sources, is too large, raw, or unstructured for analysis through conventional relational database techniques. Almost 90% of the world's data today was generated during the past two years, with 2.5 quintillion bytes of data added each day. Moreover, approximately 90% of it is unstructured. Still, the overwhelming amount of big data from the Web and the cloud offers new opportunities for discovery, value creation, and rich business intelligence for decision support in any organization. Big data also means new challenges involving complexity, security, and risks to privacy, as well as a need for new technology and human skills. Big data is redefining the landscape of data management, from extract, transform, and load, or ETL, processes to new technologies (such as Hadoop) for cleansing and organizing unstructured data in big-data applications.

The purpose of the big data panel was to provide a forum for exchange of ideas on curricula content in the area of data science and big data. Challenges include analysis, capture, curation, search, sharing, storage, transfer, visualization, and information privacy.

#### 1.2. Why big data?

Key enablers for the growth of “Big Data” are:

- Increase of storage capacities
- Increase of processing power
- Availability of data

#### 1.3.HADOOP

Hadoop is an open-source software framework for storing and processing big data in a distributed fashion on large clusters of commodity hardware. Essentially, it accomplishes two tasks: massive data storage and faster processing.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

## II.MODULES

- Data collection
- Data cleaning & exploration
- Data Analysis & Processing
- Data Visualization

### 2.1 DATA COLLECTION

Data Collection is an important aspect of any type of research study. Inaccurate data collection can impact the results of a study and ultimately lead to invalid results.

Data collection methods for impact evaluation vary along a continuum. At the one end of this continuum are quantitative methods and at the other end of the continuum are Qualitative methods for data collection .

### 2.2 DATA CLEANING

Data cleaning, also called data cleansing or scrubbing, deals with detecting and removing errors and inconsistencies from data in order to improve the quality of data. Data quality problems are present in single data collections, such as files and databases, e.g., due to misspellings during data entry, missing information or other invalid data. When multiple data sources need to be integrated, e.g., in data warehouses, federated database systems or global web-based information systems, the need for data cleaning increases significantly. This is because the sources often contain redundant data in different representations. In order to provide access to accurate and consistent data, consolidation of different data representations and elimination of duplicate information become necessary.

Big data require and provide extensive support for data cleaning. They load and continuously refresh huge amounts of data from a variety of sources so the probability that some of the sources contain “dirty data” is high. Furthermore, big data are used for decision making, so that the correctness of their data is vital to avoid wrong conclusions.

### DATA EXPLORATION

Big data exploration addresses the challenge faced by every large organization: business information is spread across multiple systems and silos and people need access to that data to meet their job requirements and make important decisions. That Big Data Exploration enables you to explore and mine big data to find, visualize, and understand all your data to improve decision making. By creating a unified view of information across all data sources - both inside and outside of your organization - you gain enhanced value and new insights.

### 2.3 DATA ANALYSIS

Big data analytics enables organizations to analyze a mix of structured, semi-structured and unstructured data in search of valuable security information. Big data analytics is the process of examining large data sets containing a variety of data types -- i.e., big data-- to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information. The analytical findings can lead to more effective marketing, new revenue opportunities, better customer service, improved operational efficiency, competitive advantages over rival organizations and other business benefits.

### 2.4 DATA VISUALIZATION

Data visualization is a general term used to describe any technology that lets corporate executives and other end users “see” data in order to help them better understand the information and put it in a business context. Process by which numerical data are converted into meaningful images.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

## III.LITERATURE REVIEW

- ❖ Big Data is the new experience curve in the new economy driven by data with high volume, velocity, variety, and veracity. They come from various sources that include the Internet, mobile devices, social media, geospatial devices, sensors, and other machine-generated data
- ❖ Joseph O. Chan (2013) An Architecture for Big Data Analytics Communications of the IIMA ©2013 Volume 13 Issue 2
- ❖ No SQL databases provide distributed, highly scalable data storage for Big Data. Yen (2009) presented a No SQL taxonomy that includes key- value- cache, key- value- store, eventually- consistent key- value- store, ordered- key- value- store, data- structures server, tuple- store, object database, document store, and wide columnar store. A popular example of No SQL database is Apache Hbase.
- ❖ Apache Hadoop uses the processing framework known as MapReduce. MapReduce helps programming solve data parallel problems for which data set can be sub-divided into small parts and processed independently.
- ❖ Shreyas Kudale, Advait Kulkarni, Asst. Prof. Leena A. Deshpande Predictive Analysis Using Hadoop: A Survey International Journal of Innovative Research in Computer and Communication Engineering Vol. 1, Issue 8, October 2013.

## IV.ARCHITECTURE

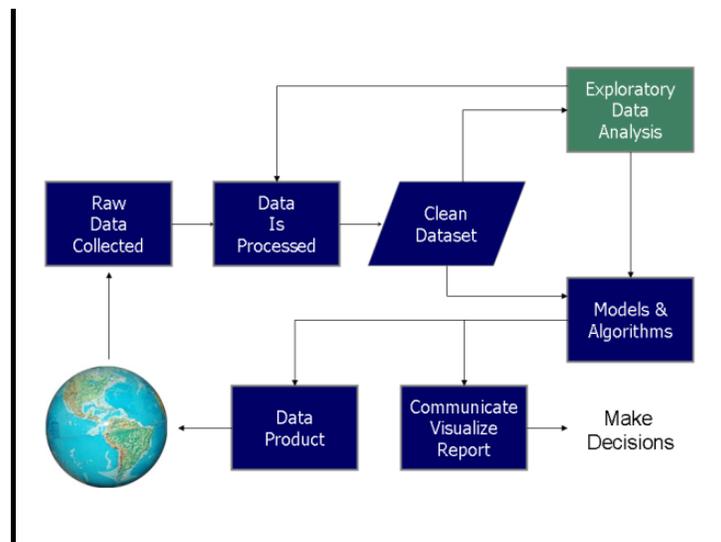


Fig.1.System Architecture

## V.CONCLUSION

Big Data helps to create new growth opportunities and entirely new categories of companies, such as those that aggregate and analysis industry data. Big Data can unlock significant value by making information transparent. All the organizations store more transactional data in digital form. Big data allows ever-narrower segmentation of customers and therefore much more precisely tailored product or service. Sophisticated analytics can substantially improve decision

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

making, minimize risks, and unearth valuable insights that would otherwise remain hidden. Hence the security intimateness in military service will enhance the hidden treats in the process of valuable during the emergency services.

## REFERENCES

- [1] Joseph O. Chan (2013) An Architecture for Big Data Analytics Communications of the IIMA ©2013 Volume 13 Issue 2
- [2] -YoonDeuk Seo and Jinho Ahn Enhancing User-friendliness of the User Taste Prediction Service Using MapReduce Framework International Journal of Multimedia and Ubiquitous Engineering Vol.9, No.5 (2014), pp.263-272 <http://dx.doi.org/10.14257/ijmue.2014.9.5.26> (2)
- [3] Bharti Thakur et al. (2014) Data Mining for Big Data: A Review International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 5, May 2014
- [4] Hassanin M. Al-Barhamtoshy and Fathy E. Eassa. A Data Analytic Framework for Unstructured Text. Life Sci. J 2014; 11(10):339-350] (ISSN: 1097-8135). <http://www.lifesciencesite.com>.
- [5] -T.K.Das , P.Mohan Kumar BIG Data Analytics: A Framework for Unstructured Data Analysis International Journal of Engineering and Technology (IJET) ISSN : 0975-4024 Vol 5 No 1 Feb-Mar 2013 (5)
- [6] G Geethakumari et al. Big Data Analysis for Implementation of Enterprise Data Security IRACST - International Journal of Computer Science and Information Technology & Security (IJSITS), ISSN: 2249-9555 Vol. 2, No.4, August 2012