

Influence of Structured, Semi-Structured, Unstructured data on various data models

Shagufta Praveen, Research Scholar, Umesh Chandra, Assistant Professor
Glocal University,

Abstract: Enormous growth of data from diversified sources changed the complete scenario of database world. Most of the surveys say that data is very important for all the organizations and its proper handling will seek attention in future. Various forms of data available in the digital world need different data models for their storage, processing and analysis. This paper discusses various kinds of data with their characteristics with examples, and also represents that the growing data is responsible for the numerous emerging data models and database evolution.

Keywords: Structured, Unstructured, Semi structured, Data Models

1. Introduction:

Big Data is a term that catches attention of everyone today. This attention can be justified through some surveys and facts. These surveys and facts says that each and every second we all users are creating a new data which gives a addition to the rate of data growth. Most of the web applications like Facebook, Twitter, Instagram, Youtube are the ones which connects with 1 billion people every day and these people not only survey but share and create new data every single second [1]. Survey says that the amount of digital universe will double in every two years [2]. Most of the organizations are working on data driven projects [3]. Most of the organization doesn't consider web data as dead data where as different research center using this data for analysis purpose and trying to utilize it for business intelligence and pattern prediction. Data mining and data extraction deals with various algorithms to extract data so that it could help us for betterment in IT industries.

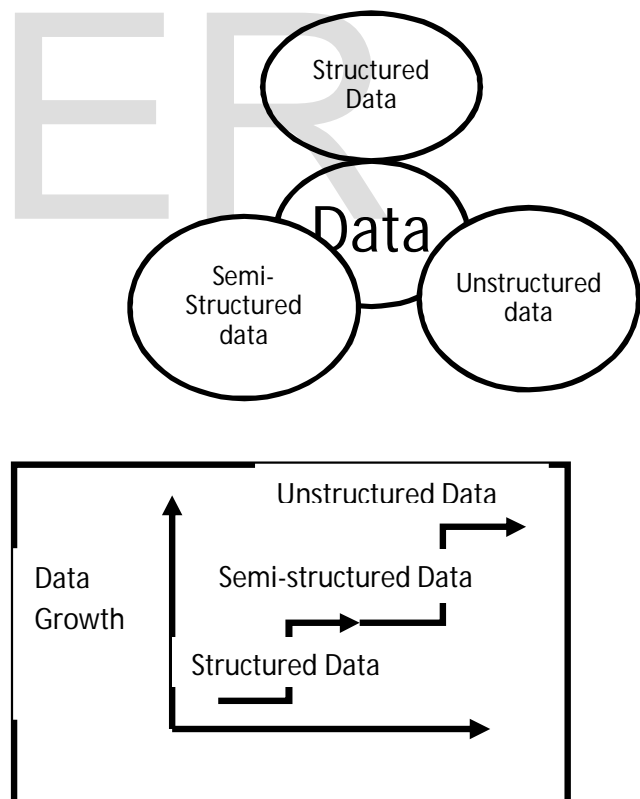


Fig 1. Kinds of Data

2. Various Kind of Data:

2.1. Structured data:

Structured data includes mainly text, these data are easily processed. These data are easily entered, stored and analyzed. Structured data are stored in the form of rows and columns which is easily managed with the a language called "structured query language"(SQL)[4]. Relational model[5] is a data model that supports structured data and manage it in the form of row and table and process the content of the table easily. XML also

Support structured data. Most of the content of the web pages are in the XML forms. These content are included in structured data, companies like Google uses structured data to find on the web to understand the content of the page [6]. This way most of the Google search is done with the help of structured data. Since starting of the revolution of database[7] network[8], hierarchical[9], relational, object relational[10] data model deals with structured data.

2.2. Characteristics of Structured Data

1. Structured data has various data type: date, name, number, characters, address
2. These data are arranged in a defined way
3. Structured data are handle through SQL
4. Structured data are dependent on schema, it is a schema based
5. These data can easily interact with computer

2.3. Semi-Structured Data

Semi-structured data includes e-mails, XML and JSON. Semi structured data is not fit for relational database where it is expressed with the help of edges, labels and tree structures. These are represented with the help of trees and graphs and they have attributes, labels. These are schema-less data. Data models which are graph based can store semi-structured data. MongoDB is a NOSQL model that support JSON (semi-structured data).

Data consist of tags and which are self-describing are generally semi-structured data. They are different from structured and unstructured data. Data object Model [11], Objects Exchange Model [11], Data Guide[11] are famous data model that express semi-structured data. Concepts for semi-structured data model: document instance, document schema, elements attributes, elements relationship sets[11].

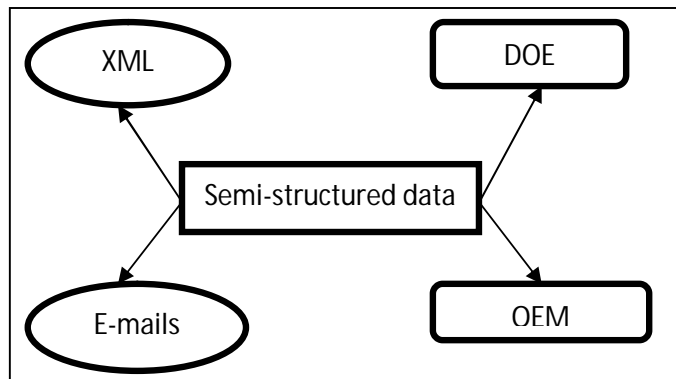


Fig.3 Attributes of Semi-Structured Data

2.4. Example of Semi-structured Data

```
{  
  Row:{Emp_id:" 12345",Emp_name:"Ram"},  
  Row:{Emp_id:" 56786",Emp_name:"Hari"},  
  Row:{Emp_id:" 67858",Emp_name:"Shyam"},  
  Row:{Emp_id:" 90890",Emp_name:"John"},  
}
```

2.5. Characteristics of Semi-structured Data

1. It is not based on Schema
2. It is represented through label and edges
3. It is generated from various web pages
4. It has multiple attributes

3. Unstructured Data

Unstructured data includes videos, images, and audios. Today, in our digital universe 90% of data which is increasing is unstructured data. This data is not fit for relational database and in order to make them store, scenario came up with NoSQL database. Today there are four family of NoSQL database: key-value, column-oriented, graph-oriented, and document-oriented. Most of the famous organization today (Amazon, LinkedIn, Facebook, Google, Youtube) is dealing with NoSQL data [12] and they are replaced their convention database to NoSQL database.

3.1. Characteristics of Unstructured Data

1. It is not based on Schema
2. It is not suitable for relational database
3. 90% of unstructured data is growing today
4. It includes digital media files, Word doc, .pdf files,
5. It is stored in NoSQL database

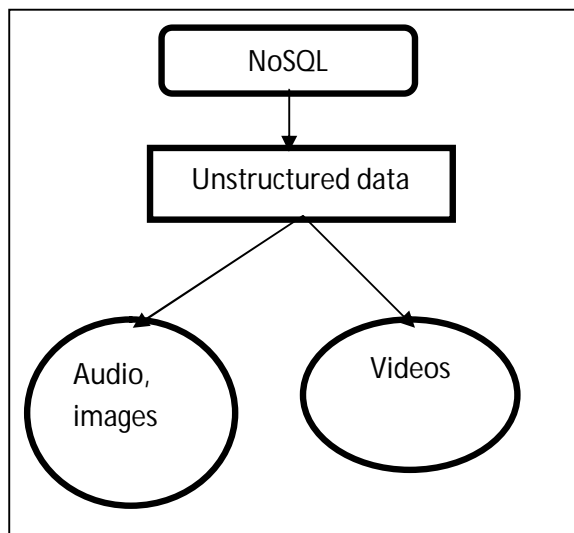


Fig.4. Attributes of Unstructured Data

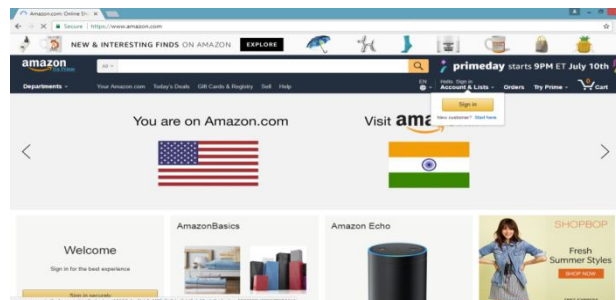


Fig.5. Example of Unstructured Data

4. Conclusion: This paper emphasize on the concept that growing data directly influence its related data models and database technologies, it represents that big data concept not only deals with huge and vast data but it gives a new gate to database analyst and researchers to work on various data and data models for survival of new kinds of data in upcoming and present scenario.

References:

1. <https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#7e621bc417b1>
2. <https://insidebigdata.com/2017/02/16/the-exponential-growth-of-data/>
3. <https://www.idgenterprise.com/resource/research/2015-big-data-and-analytics-survey/>
4. J. R. Groff, P. N. Weinberg SQL: The complete reference second addition, 2002, Mc-Graw Hills Companies
5. E.F. CODD, 1970. A Relational Model of Data for Large Shared Data Banks.
6. <https://developers.google.com/search/docs/guides/intro-structured-data>
7. S. Praveen, Dr. U. Chandra, Arif ali wani, a literature review on evolving database, IJCA, March 2017.
8. https://en.wikipedia.org/wiki/Network_model
9. <http://www.edugrabs.com/hierarchical-model/>
10. <http://www.learn.geekinterview.com/it/data-modeling/object-relational-model.html>
11. T.W Ling., G. Dobbie, Semi-structured database design., 2005, Springer, 178,978-0-387-23567-7
12. S. Praveen, Dr. U. Chandra, NoSQL: IT Giant Perspectives, 2017, IJCIR