Big Data and NoSQL

Contents

- What is Data?
- What is Big Data?
- Characteristics of Big Data
- Example of Big Data: Digital Data
- Types of Digital Data: Unstructured, Semi-structured, Structured data
- Big Data Analytics
- Managing Big Data: NoSQL
- Types of NoSQLs
- SQL vs. NoSQL

What is Data?

Any abstract value (“numeric”, “alphabetic” and/or, “alpha-numeric”) stored/recorded about a person, an object or an event can be called “Data”.

Once the “Data” is collected and stored formally, it can be utilized for various computational operations like data retrieving, analyzing and report generation.

Those computational operations generate meaningful “Information” from the stored data.

What is Big Data?

Big Data is a collection of data that is huge in volume yet growing exponentially with time. It is a kind of data that has such large size and complexity that none of traditional data management tools/methods can store it or process it efficiently. Big data is also data but with several additional characteristics.

****Big Data can be used to analyze and reveal patterns, understanding the trends and finding associations.*

Characteristics of Big Data: V's of Bigdata

- **Volume:** *Amount/Size of Data*
- **Velocity:** *Rapid growing and real time*
- **Variety:** *Various Datatypes and various forms*
- **Variability:** *Variable insights generated from the same data*
- **Veracity:** *Accuracy and truthfulness of the Data*
- **Visualization:** *Effectively representable Data*
- **Value:** *Using the aforementioned characteristics data brings value to the business*
- …

Big Data Application

Data/Big Data Analytics

Big Data Analytics is the process of collecting, organizing and using sets of data (or Big Data) to discover patterns and other useful information.

Data Analytics Continued...

Data Analytics is typically performed using specialized software tools and applications for:

- **Predictive Analytics**
- **Data Mining**
- **Text Mining**
- **Forecasting**
- **Data Optimization**

Which enable organizations to take better and effective business decisions in future!

Types of Data/Big Data

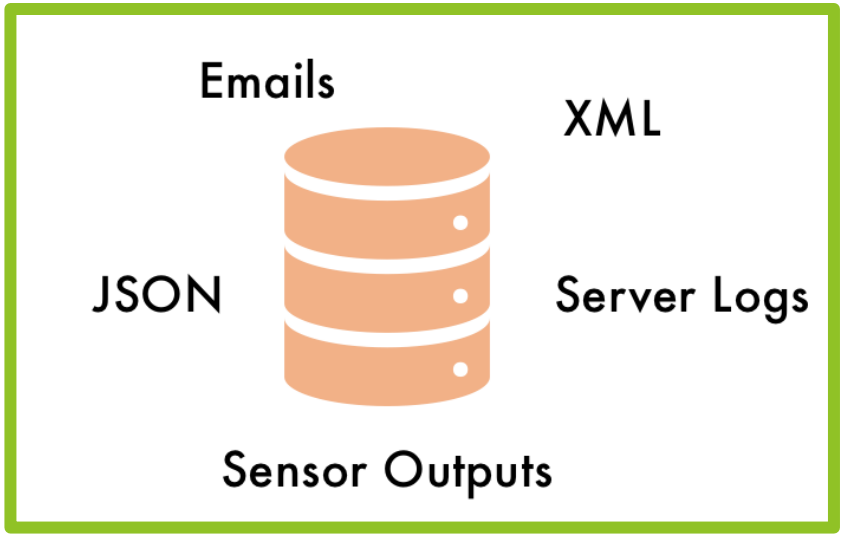
- Unstructured
- Semi-Structured
- Structured

Unstructured Data

Unstructured data is information, in many different forms, that doesn't follow conventional data models, making it difficult to store and manage in a mainstream relational database structure.

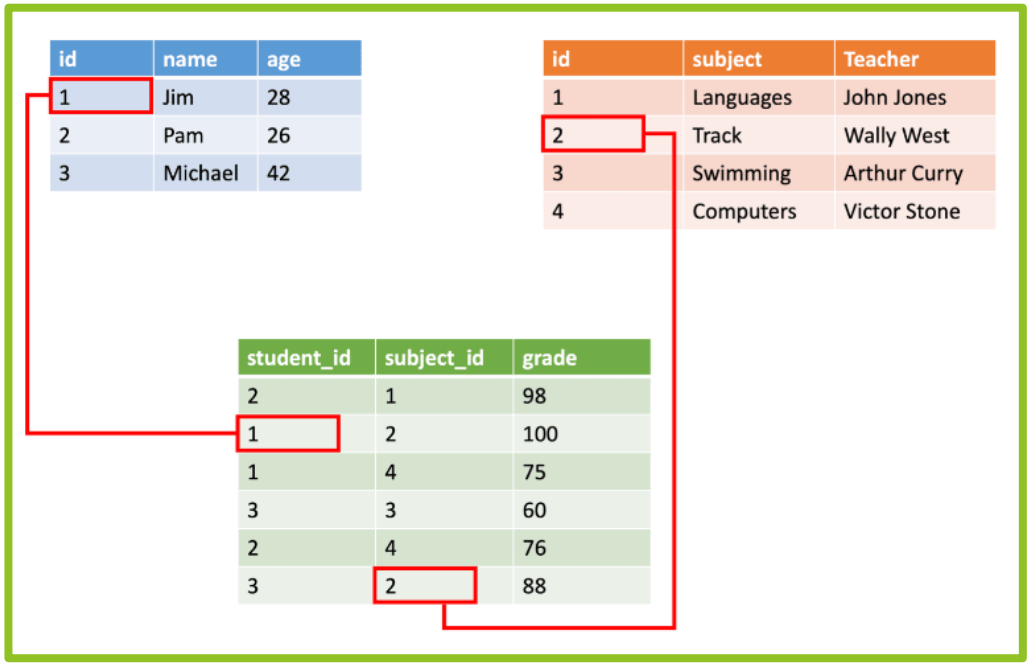
Semi-structured Data

Semi-structured data is a form of structured data that does not obey the tabular structure of data models associated with relational databases, however, contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data.



Structured Data

Structured data is when data is in a standardized format, has a well-defined structure, complies to a data model, follows a persistent order, and is easily accessed by humans and programs. This data type is generally stored in a database.



id	name	age
1	Jim	28
2	Pam	26
3	Michael	42

id	subject	Teacher
1	Languages	John Jones
2	Track	Wally West
3	Swimming	Arthur Curry
4	Computers	Victor Stone

student_id	subject_id	grade
2	1	98
1	2	100
1	4	75
3	3	60
2	4	76
3	2	88

Types of Data/Big Data Continued…

The university has 5600 students.
John's ID is number 1, he is 18 years old and already holds a B.Sc. degree.
David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

Unstructured Data

```
<University>
  <Student ID="1">
    <Name>John</Name>
    <Age>18</Age>
    <Degree>B.Sc.</Degree>
  </Student>
  <Student ID="2">
    <Name>David</Name>
    <Age>31</Age>
    <Degree>Ph.D. </Degree>
 </Student>
....
</University>
```

Semi-structured Data

ID	Name	Age	Degree
1	John	18	B.Sc.
2	David	31	Ph.D.
3	Robert	51	Ph.D.
4	Rick	26	M.Sc.
5	Michael	19	B.Sc.

Structured Data

NoSQL

NoSQL databases (also known as "not only SQL") are non-tabular databases and store data differently than relational tables. NoSQL databases come in a variety of types based on their data model.

The main types are: *document*, *key-value*, *wide-column*, *and graph*.

Types of NoSQL

Advantages of NoSQL Database

- Handle Large Volumes of Data that may change frequently with a Scale-Out Architecture
- Store Unstructured, Semi-Structured, or even Structured Data
- Enables Easy Updates to Schema and Fields
- Developer-Friendly
- Take Full Advantage of the Cloud to Deliver Zero Downtime

When to use NoSQL vs. SQL?

NoSQL databases are usually “highly specialized systems” and have their special usage and limitations. NoSQL is more suitable for those who handle huge volumes of data that usually come in from several, often different sources.

So, other than these scenarios the SQL is good enough for processing data and generating reports.


Do you have any
QUESTIONS?