Ch1-Introduction to Big 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 data with so large size and complexity that none of traditional data management tools can store it or process it efficiently.
- Big data is also a data but with huge size.

What is an Example of Big Data?

- The New York Stock Exchange is an example of Big Data that generates about *one terabyte* of new trade data per day.
- Social Media
- The statistic shows that *500+terabytes* of new data get ingested into the databases of social media site Facebook, every day..
- This data is mainly generated in terms of photo and video uploads, message exchanges, putting comments etc.

Types Of Big Data

Following are the types of Big Data:

- **Structured**
- **Unstructured**
- **Semi-structured**

1)**Structured Data**

- Any data that can be stored, accessed and processed in the form of fixed format is termed as a ‘structured’ data.
- Over the period of time, talent in computer science has achieved greater success in developing techniques for working with such kind of data (where the format is well known in advance) and also deriving value out of it.
- However, nowadays, we are foreseeing issues when a size of such data grows to a huge extent, typical sizes are being in the rage of multiple zettabytes.
- An ‘Employee’ table in a database is an example of Structure

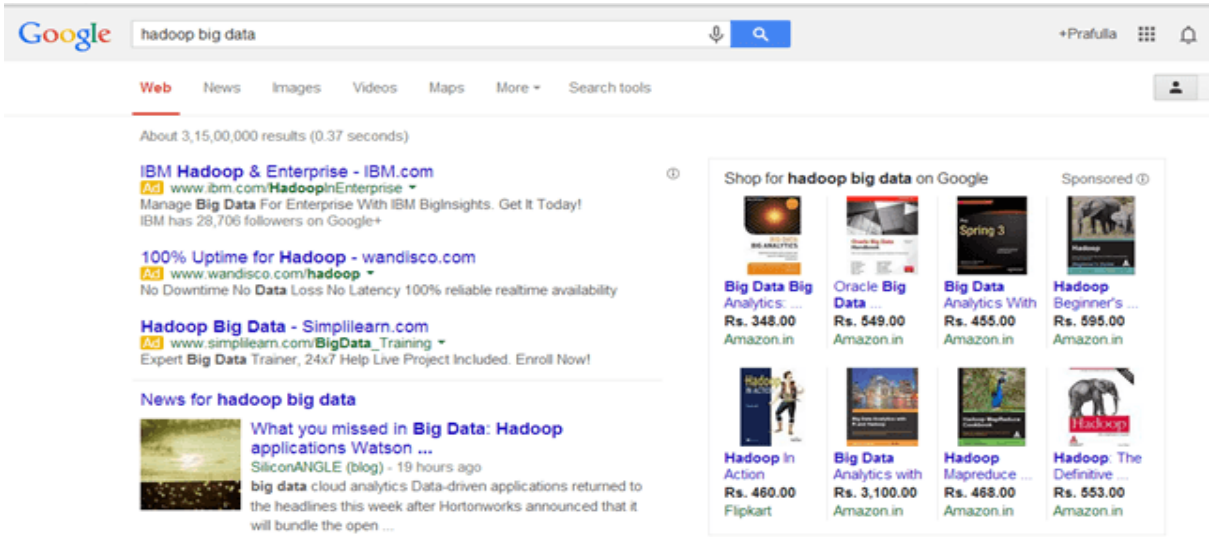
Employee_ID	Employee_Name	Gender	Department	Salary_In_lacs
2365	Rajesh Kulkarni	Male	Finance	650000
3398	Pratibha Joshi	Female	Admin	650000
7465	Shushil Roy	Male	Admin	500000

2)Unstructured

- Any data with unknown form or the structure is classified as unstructured data.
- In addition to the size being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it.
- A typical example of unstructured data is a heterogeneous data source containing a combination of simple text files, images, videos etc.

Examples Of Un-structured Data

The output returned by 'Google Search'



3)**Semi-structured**

- Semi-structured data can contain both the forms of data.
- We can see semi-structured data as a structured in form but it is actually not defined with e.g. a table definition in relational DBMS.
- Example of semi-structured data is a data represented in an XML file.
- Examples Of Semi-structured Data
- Personal data stored in an XML file-

<rec><name>Prashant Rao</name><sex>Male</sex><age>35</age></rec>
<rec><name>Seema R.</name><sex>Female</sex><age>41</age></rec>
<rec><name>Satish Mane</name><sex>Male</sex><age>29</age></rec>
<rec><name>Subrato Roy</name><sex>Male</sex><age>26</age></rec>
<rec><name>Jeremiah J.</name><sex>Male</sex><age>35</age></rec>

Characteristics Of Big Data

1) Volume

- As its name suggests, the most common characteristic associated with big data is its high volume.
- This describes the enormous amount of data that is available for collection and produced from a variety of sources and devices on a continuous basis.

2) Velocity

- Big data velocity refers to the speed at which data is generated.
- Today, data is often produced in real time or near real time, and therefore, it must also be processed, accessed, and analyzed at the same rate to have any meaningful impact.

3) Variety

- Data is heterogeneous, meaning it can come from many different sources and can be structured, unstructured, or semi-structured.
- More traditional structured data (such as data in spreadsheets or relational databases) is

now supplemented by unstructured text, images, audio, video files, or semi-structured formats like sensor data that can't be organized in a fixed data schema.

4)Veracity:

- Big data can be messy, noisy, and error-prone, which makes it difficult to control the quality and accuracy of the data.
- Large datasets can be unwieldy and confusing, while smaller datasets could present an incomplete picture.
- The higher the veracity of the data, the more trustworthy it is.

5)Variability:

- The meaning of collected data is constantly changing, which can lead to inconsistency over time.
- These shifts include not only changes in context and interpretation but also data collection methods based on the information that companies want to capture and analyze.

6)Value:

- It's essential to determine the business value of the data you collect.
- Big data must contain the right data and then be effectively analyzed in order to yield insights that can help drive decision-making.

How does big data work?

The central concept of big data is that the more visibility you have into anything, the more effectively you can gain insights to make better decisions, uncover growth opportunities, and improve your business model.
Making big data work requires three main actions:

Integration: Big data collects terabytes, and sometimes even petabytes, of raw data from many sources that must be received, processed, and transformed into the format that business users and analysts need to start analyzing it.

Management: Big data needs big storage, whether in the cloud, on-premises, or both. Data must also be stored in whatever form required. It also needs to be processed and made available in real time. Increasingly, companies are turning to cloud solutions to take advantage of the unlimited compute and scalability.

Analysis: The final step is analyzing and acting on big data—otherwise, the investment won't be worth it. Beyond exploring the data itself, it's also critical to communicate and share insights across the business in a way that everyone

can understand. This includes using tools to create data visualizations like charts, graphs, and dashboards.

Big Data Advantages

-Businesses can utilize outside intelligence while taking decisions

Access to social data from search engines and sites like facebook, twitter are enabling organizations to fine tune their business strategies.

-Improved customer service
Traditional customer feedback systems are getting replaced by new systems designed with Big Data technologies. In these new systems, Big Data and natural language processing technologies are being used to read and evaluate consumer responses.
-Early identification of risk to the product/services, if any
-Better operational efficiency

What is big data analytics

Big data analytics is the often complex process of examining big data to uncover information -- such as hidden patterns, correlations, market trends and customer preferences -- that can help organizations make informed business decisions.

Why is big data analytics important?

Organizations can use big data analytics systems and software to make data-driven decisions that can improve business-related outcomes.
The benefits may include more effective marketing, new revenue opportunities, customer personalization and improved operational efficiency.
With an effective strategy, these benefits can provide competitive advantages over rivals.

Applications of Big Data

1)Travel and Tourism

- Travel and tourism are the users of Big Data.
- It enables us to forecast travel facilities requirements at multiple locations, improve business through dynamic pricing, and many more.

2)Financial and banking sector

- **The financial and** banking sectors use big data technology extensively.
- **Big data** analytics help **banks** and customer behavior on the basis of **investment patterns, shopping trends, motivation to invest**, and inputs that are obtained from **personal** or **financial** backgrounds.

3)Healthcare

- Big data has started making a massive difference in the **healthcare** sector, with the help of **predictive analytics, medical professionals,** and health care personnel.
- It can produce **personalized healthcare** and **solo patients also**.

4)Telecommunication and media

- **Telecommunications and the multimedia** sector are the main users of **Big Data**.
- There are **zettabytes** to be generated every day and handling large-scale data that require big data technologies.

5)Government and Military

- **The government and military** also used **technology** at high rates.
- We see the figures that the **government** makes on the record. In the **military**, a fighter plane requires to process **petabytes** of data.
- Government agencies use Big Data and run many agencies, managing utilities, dealing with traffic jams, and the effect of crime like **hacking** and **online fraud**.

Aadhar Card: The government has a record of **1.21 billion** citizens. This vast data is analyzed and store to find things like the number of youth in the country. Some schemes are built to target the maximum population. Big data cannot store in a traditional database, so it stores and analyze data by using the Big Data Analytics tools.

6)E-commerce

- E-commerce is also an application of Big data.
- It maintains relationships with customers that is essential for the e-commerce industry.
- E-commerce websites have many marketing ideas to retail merchandise customers, manage transactions, and implement better strategies of innovative ideas to improve businesses with Big data.

Amazon: Amazon is a tremendous e-commerce website dealing with lots of traffic daily.

But, when there is a pre-announced sale on Amazon, traffic increase rapidly that may crash the website.

So, to handle this type of traffic and data, it uses Big Data. Big Data help in organizing and analyzing the data for far use.

****