Industry 4.0
BIG DATA &ANALYTICS

Prepared by: Prof. Rasmita Lenka
rasmitafet@kiit.ac.in

TOPICS TO BE COVERED

 Introduction Data Types Charaterstics of Big Data Data Sources Data Acquistion Data Storage Big Data Analytics for Industry.4.0

Introduction

What is Data?

The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.

What is Big Data?

Big Data is also data but with a huge size. Big Data is a term used to describe a collection of data that is huge in volume and yet growing exponentially with time. In short such data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently.

"Extremely large data sets that may be analyzed computationally to reveal patterns , trends and association, especially relating to human behavior and interaction are known as Big Data."

It refers to a massive amount of data that keeps on growing exponentially with time. It is so voluminous that it cannot be processed or analyzed using conventional data processing techniques. It includes data mining, data storage, data analysis, data sharing, and data visualization. The term is an all-comprehensive one including data, data frameworks, along with the tools and techniques used to process and analyze the data.

Traditional Data Vs Big Data

- Traditional data is generated in enterprise level.
- Its volume ranges from Gigabytes to Terabytes.
- Traditional data source is centralized and it is managed in centralized form
- Traditional data base tools are required to perform any data base operation
- Its data sources includes ERP ,SQL,transaction data, CRM transaction data, financial data, organizational data, web transaction data etc..

- Big data is generated outside the enterprise level.
- Its volume ranges from Petabytes to Zettabytes or Exabytes
- Big data source is distributed and it is managed in distributed form.
- Special kind of data base tools are required to perform any databaseschema-based operation
- Its data sources includes social media, device data, sensor data, video, images, audio etc.

Data Types

Structured

Structured data is created using a fixed schema and is maintained in tabular format. The elements in structured data are addressable for effective analysis. It contains all the data which can be stored in the SQL database in a tabular format. Today, most of the data is developed and processed in the simplest way to manage information.

Relational data, Geo-location, credit card numbers, addresses, etc.

Unstructured

It is defined as the data in which is not follow a pre-defined standard or you can say that any does not follow any organized format. This kind of data is also not fit for the relational database because in the relational database you will see a pre-defined manner or you can say organized way of data. Unstructured data is also very important for the big data domain and To manage and store Unstructured data there are many platforms to handle it like No-SQL Database.

Word, PDF, text, media logs,

Semi-structured

Semi-structured data is information that does not reside in a relational database but that have some organizational properties that make it easier to analyze. With some process, you can store them in a relational database but is very hard for some kind of semi-structured data, but semi-structured exist to ease space.

XML data.

TYPES OF BIG DATA

Structured

- Main Frame
- SQL Server
- Oracle
- DB2
- Sybase
- Access, Excel, txt, etc
- Teradata
- Neteeza, Other mpp
- SAP, JDE, JDA, Other ERP.

Un-Structured

- Social Media
  - Chatter, Text Analytics, Blogs, Tweets, Comments, Likes, Followers, Social Authority, Clicks, Tags, etc.
- Digital, Video, QR
- Audio
- Geo-Spatial

Multi-Structured /Hybrid

- Emerging Market Data
- Loyalty
- E-Commerce
- Other Third Party Data
  - Weather
  - Currency Conversion
  - Demographic
  - Panel
- POS, POL, IR, EDI, RFID, NFC, QR, IRI, Rsi, Nielsen, Other Syndicated, IMS, MSA, etc.

Charaterstics of Big Data

Big data is a collection of data from many different sources and is often describe by five characteristics: volume, value, variety, velocity, and veracity.
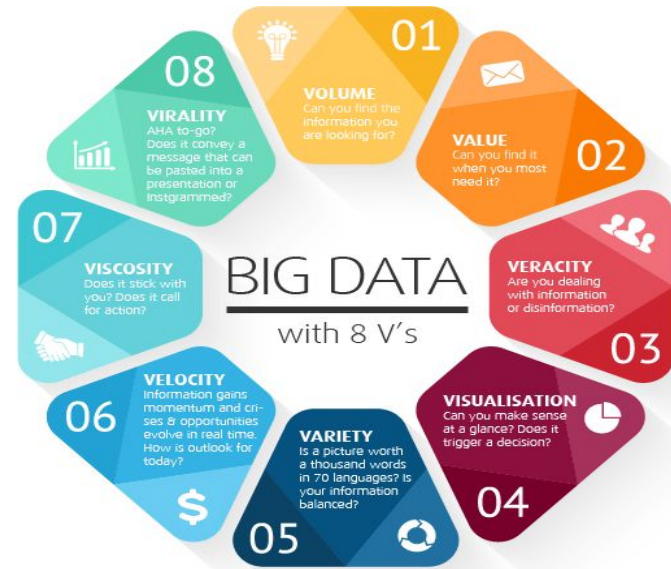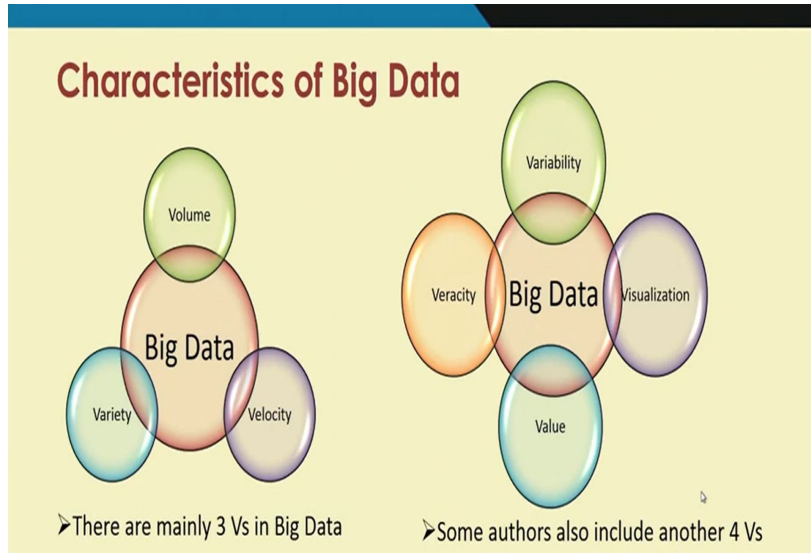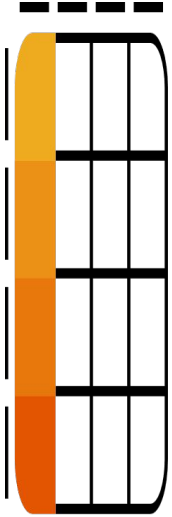
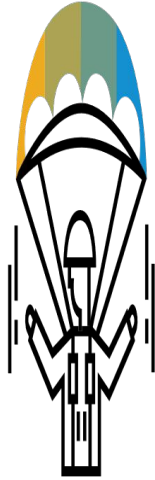Volume Velocity Variety Veracity Value

Big data is a collection of data from many different sources and is often describe by five characteristics: volume, value, variety, velocity, and veracity.
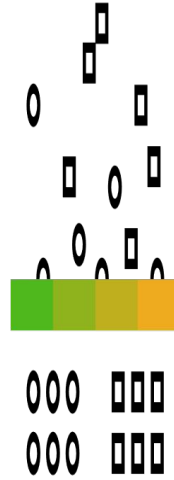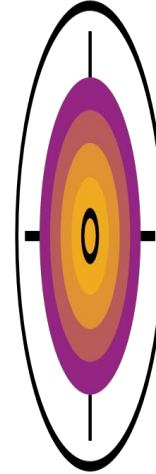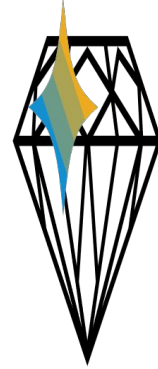
- *Volume: the size and amounts of big data that companies manage and analyze*
- *Value: the most important "V" from the perspective of the business, the value of big data usually comes from insight discovery and pattern recognition that lead to more effective operations, stronger customer relationships and other clear and quantifiable business benefits*
- *Variety: the diversity and range of different data types, including unstructured data, semi-structured data and raw data*
- *Velocity: the speed at which companies receive, store and manage data – e.g., the specific number of social media posts or search queries received within a day, hour or other unit of time*
- *Veracity: the "truth" or accuracy of data and information assets, which often determines executive-level confidence*
- *The additional characteristic of variability can also be considered:*
- *Variability: the changing nature of the data companies seek to capture, manage and analyze – e.g., in sentiment or text analytics, changes in the meaning of key words or phrases*

Data Sources

- Big data is used by organizations for the sole purpose of analytics.Main sources of big data can be grouped under the headings of social (human), machine (sensor) and transactional.

Media as a big data source
- Images, videos, audios, podcasts
- Social media platforms like Facebook, Twitter, YouTube, Instagram

Cloud as a big data source
- Public, private, or third party cloud platforms

Web as a big data source
- Data publically available on the web

IoT as a big data source
- Data generated from the interconnection of IoT devices

Databases as a big data source
- Traditional and modern databases

Data Acquistion

- Data acquisition has been understood as the process of gathering, filtering, and cleaning data before the data is put in a data warehouse or any other storage solution.Big Data architecture has to acquire high speed data from a variety of sources like web, DBMS(OLTP), NoSQL, HDFS and the data is also diverse in nature .
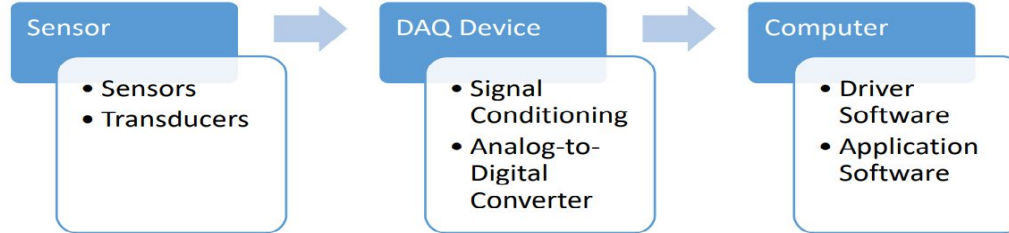
- Frameworks and tech for data gathering
(□Storm,□Simply Scalable Streaming System,Kafka,Flume,Hadoop,)

Data Storage

- Big data storage is a compute-and-storage architecture that collects and manages large data sets and enables real-time data analytics.
- ❖ The Data warehouse
- ❖ Enterprise Data Warehouse:
- ❖ Operational Data Store:
- ❖ Data Mart:
- ❖ The Data Lake
- ❖ Network Attached Storage (NAS)
- ❖ The Cloud
- ❖ Object Storage

An ideal Big Data storage system stores an infinite amount of data.

- Provide fast random read and write access,
- handle different data models flexibly and efficiently,
- support both structured and unstructured data,
- keep data encrypted so that confidentiality can be protected.

Big Data Analytics for Industry.4.0

Big Data Analytics –

❖ Big data analytics describes the process of uncovering trends, patterns, and correlations in large amounts of raw data to help make data-informed decisions. These processes use familiar statistical analysis techniques—like clustering and regression

❖ Hadoop, Spark, and NoSQL databases were created for the storage and processing of big data.

❖ It evolve to integrate the vast amounts of complex information created by sensors, networks, transactions, smart devices, web usage, and more.

❖ big data analytics methods are being used with emerging technologies, like machine learning, to discover and scale more complex insights.

Importance & challenges of big data analytics

Big data analytics is important because it lets organizations use colossal amounts of data in multiple formats from multiple sources to identify opportunities and risks, helping organizations move quickly and improve their bottom lines

Importance

❖ Cost savings. Helping organizations identify ways to do business more efficiently

❖ Product development. Providing a better understanding of customer needs

❖ Market insights. Tracking purchase behavior and market trends

challenges of big data

❖ Making big data accessible

❖ Maintaining quality data.

❖ Keeping data secure

❖ Finding the right tools and platforms

Big data analytics working process and models

Steps for Analytic process

- ❖ Collect Data
- ❖ Process Data
- ❖ Clean Data
- ❖ Analyze Data

Analytic Models: There are basically four models big data analytics

- ❖ Data mining sorts through large datasets to identify patterns and relationships by identifying anomalies and creating data clusters.
- ❖ Predictive analytics uses an organization's historical data to make predictions about the future, identifying upcoming risks and opportunities.
- ❖ Deep learning imitates human learning patterns by using artificial intelligence and machine learning to layer algorithms and find patterns in the most complex and abstract data.

❑Big data analytics used in Industry 4.0

❖ Product and/or machine design data such as threshold specifications

❖ Machine-operation data from control systems

❖ Product- and process-quality data

❖ Records of manual operations carried out by staff

❖ Manufacturing execution systems

❖ Information on manufacturing and operational costs

❖ Fault-detection and other system-monitoring deployments

❖ Logistics information including third-party logistics

❖ Customer information on product usage, feedback,