Page No.
Date

Devanshu Surana
1032210755
PC-23, BDT-1 Batch 2

BDT Lab Assignment 1

Problem statement:
Installation of Big Data tools.

Objectives:
1. To learn concepts of Big Data
2. To learn how to install and use of different big data tools.

Theory:
Big Data:
It is a combination of structured, semistructured, and unstructured data collected by organizations that can be mined for information and used in other advanced analytics applications.

The characteristic that define Big Data are:
1. Volume: Big data involves extremely large quantities of data. This data may be generated rapidly and accumulate quickly, making it challenging to store, process, analyze using traditional databases and tools.
2. Velocity: Speed at which data is generated, collected and processed is a crucial aspect of big data.
3. Variety: Big Data comes in various formats and types including structured data (traditional relational databases), semi-structured (JSON or XML) and unstructered (img, txt, audio).

Page No.
Date

Some popular Big Data tools are :

1. Hadoop; Hive and Pig, MongoDB, Spark, AWS, Snowflake

Use of Big Data tools:

1. Hadoop: 1. Large scale data storage
   2. Batch Processing and Analysis
   3. Data Warehousing
   4. Extraction, Transform, Load
   5. Data Exploration and Analysis.

2. Hive and Pig: 1. Data Warehousing
   2. Data Transformation and ETL.
   3. Querying and Analysis of Large Datasets
   4. Data Processing Pipelines.
   5. Analytics on structured and semi structured Data.

3. MongoDB : 1. Document based storage and retrieval
   2. Flexible Schema Design
   3. Real time Data processing
   4. storage and analysis of semi-structured and unstructured data.
   5. Content management system

4. Spark : 1. Realtime and Batch Data Processing
   2. Machine learning and Advanced Analytics.
   3. Stream Processing
   4. Graph Processing & Analysis.
   5. ETL

Page No.
Date

5. AWS : 1. Cloud based storage
2. Scalable Computing
3. Data Processing and Analytics
4. Serverless computing
5. IoT Data Processing

6. Snowflake : 1. Cloud based Data Warehousing
2. ETL
3. Support for Concurrency and Collaboration

Program Statement : Install Big Data technology tools and learn its various options.

Platform : 64-bit Open source Linux / Windows

Conclusion : Hence, I learned different tools of Big Data technologies.

FAQ's.

Q1) Explain 7 V's of Big Data.

→ 7 V's of Big Data are :

1. Volume : Refers to vast amount of data being generated, collected and stored.
2. Velocity : Describes the speed at which data is generated and how quickly it needs to be processed, often in real-time.
3. Variety : Encompasses the different types and format of data, including structured, semi-structured, and unstructured data from diverse sources.
4. Veracity : Focuses on quality and reliability of data.
5. Value : Represents the goal of extracting meaningful

insights and actionable information.
6. Variability: Accounts for inconsistency and volatility in data flow.
7. Visibility: Refers to need have clear insights into data, its sources, and its flow.

Q) Explain Architecture of Big Data System.
→ Architecture of Big Data System involves several key components.
1. Big Data Sources: Where data originates, such as sensors, devices, and databases.
2. Data Ingestion: Collects and brings data into system from various sources.
3. Data Storage: Large-scale storage systems. store raw and processed data.
4. Data Processing: Technologies like Hadoop, Spark analyze and transform data.
5. Data Analytics: Machine learning and analytics tools uncover insights and patterns.
6. Data Visualisation: Tool represent data insights in a user-friendly way.
7. Data Security: Ensures data confidentiality and Integrity.
8. Scalability: Design allows for handling increasing data and demand.
9. Monitoring: Tools track system health & performance.

Page No.
Date

Q) Explain Big Data applications in any three domains.

→ 1. Healthcare:

- Medical Research: Analyze large-scale genomic and proteomic data to identify disease patterns and potential treatments.
- Predictive Analytics: Health organization use data from wearable and patient records to predict disease.
- Drug Discovery: Big Data assist in screening and analyzing vast chemical and biological datasets to expedite discovery of new drugs and therapies.

2. Retail and Marketing:

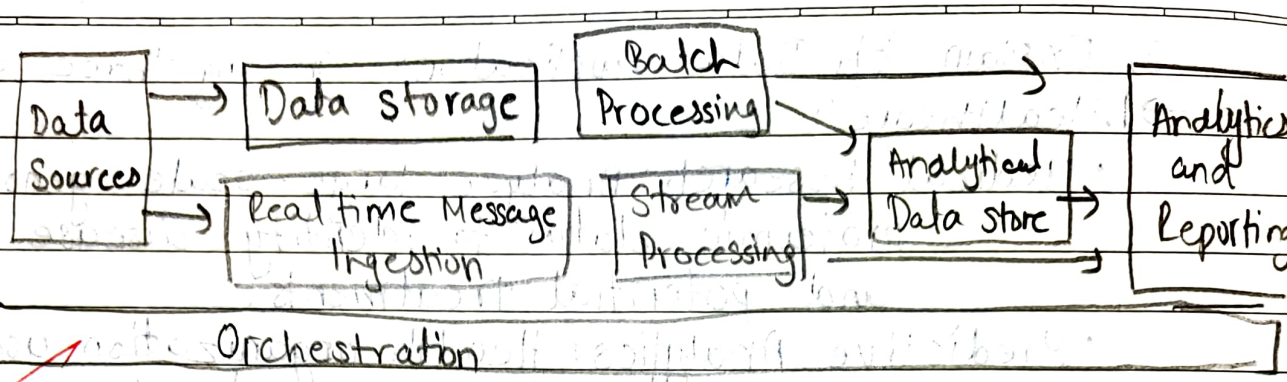
- Customer Insights: Retailers analyze customer purchase history, social media and browsing behaviour to personalize recommendation.
- Inventory Management: Big Data helps optimize inventory levels, reducing waste and ensuring products are available when needed.
- Market trend Analysis: Helps in identifying emerging trends.

3. Transportation and Logistics:

- Route optimization: Big-Data algorithms process real time traffic data to optimize routes, reduce fuel consumption.
- Fleet management: sensor data from vehicles is analyzed to monitor vehicle health.
- Supply Chain Analytics: Helps optimize processes, reduce delays and enhance overall efficiency.

Page No.
Date

Ans 3)



(3) 9/8/23