ASSIGNMENT-1

1.Compare Big Data and Data Mining

Big Data: It is huge, large or voluminous data, information or the relevant statistics acquired by the large organizations and ventures. Many software and data storage created and prepared as it is difficult to compute the big data manually.
It is used to discover patterns and trends and make decisions related to human behavior and interaction technology.

Data Mining: Data Mining is a technique to extract important and vital information and knowledge from a huge set/libraries of data. It derives insight by carefully extracting, reviewing, and processing the huge data to find out pattern and co-relations which can be important for the business. It is analogous to the gold mining where golds are extracted from rocks and sands.

Data Mining	Big Data
It is one of the method in the pipeline of Big Data.	Big Data is a technique to collect, maintain and process the huge information. It explains the data relationship.
Data mining is a part of Knowledge Discovery of the Data. It is close view of the data.	It is about extracting the vital and valuable information from huge amount of the data.It is a technique of tracking and discovering of trends of complex data sets. It is a large or overall view of the data.
The goal is same as Big Data as it is one of the tool of Big Data.	The goal is to make data more vital and usable i.e. by extracting only important information from the huge data within existing traditional aspects.
It is manual as well as automated in nature.	It is only automated as computing huge data is difficult.

Data Mining	Big Data
It only focuses on only one form of data. i.e. structured.	It focuses and works with all form of data i.e. structured, unstructured or semi-structured.
It is used to create certain business insights. Data mining is a manager of the mine.	It is mainly used for business purposes and customer satisfaction. Big Data is a mine.
It is a sub set of Big Data. i.e. one of the tools.	It is a super set of Data Mining.
It is a tool to dig up the vital information from the large data. Data can be large as well as small.	It is more involved with the processes of handling voluminous data. Data can only be large.

2.Discuss the role of traditional on-disk storage devices (HDDs, SSDs) in Big Data environments. Evaluate the advantages and limitations of using on-disk storage for managing large volumes of data. How do modern storage technologies, such as SSDs and hybrid storage solutions, impact the performance and scalability of Big Data platforms?

In Big Data environments, traditional on-disk storage devices like HDDs and SSDs play key roles in managing vast data volumes.

- **HDDs** are cost-effective, providing large storage capacities, but suffer from slower performance due to mechanical parts, making them ideal for cold storage or archival purposes.

- **SSDs** are faster, with lower latency and higher throughput, improving performance for data-intensive tasks such as real-time analytics and machine learning. However, they are more expensive and have limited write cycles, making them less suitable for certain use cases without wear leveling.

Advantages:

- **HDDs**: Low cost per GB, suitable for large storage.

- **SSDs**: Faster read/write speeds, improved performance for real-time data access.

- Both are scalable and reliable when paired with redundancy techniques.

Limitations:

- **HDDs**: Slow speeds and higher latency.

- **SSDs**: Higher cost and limited write endurance.

Modern Storage Technologies Impact:

- **Hybrid Storage** (HDD + SSD): Balances performance and cost by using SSDs for high-priority, frequently accessed data and HDDs for cold storage.

- **Cloud Storage**: Modern Big Data platforms leverage scalable cloud solutions that combine SSDs and HDDs, improving both performance and cost-efficiency.

Overall, a mix of SSDs, HDDs, and hybrid storage systems is often used to optimize for performance, cost, and scalability in Big Data platforms.