Section A-

1. Data Architecture in Big Data refers of the design and structure of systems used to collect, store, manage, and analyze large volumes of data. It includes data sources, storage formats, processing frameworks, and access methods, ensuring scalability, speed, and integration.

2. ~~What is~~
Networking is crucial in Big Data for enabling high-speed data transfer between distributed nodes in a cluster. It supports data replication, communication between Processing units, and access to cloud or on-premis storage, ensuring low latency and high throughput.

3. Hp Vertica is a columnar storage-based analytical database designed for high-speed querying of large datasets. It supports real-time analytics and Parallel Processing, making it ideal for Big data applications like business intelligence and data warehousing.

4. Redis is an in-memory key-value data store used for fast caching and real-time analytics. Its high speed and low latency make it suitable for Big Data use cases such as session storage, leader boards, and real-time recommendation engines.

Section - B

1. What is Impala, and how does it enhance big Data querying? Compare it with SparksQL

Ans Impala is a distributed SQL query engine for Apache Hadoop that allows low-latency and high throughput SQL queries. Unlike SparksQL, which is optimized for batch Processing using RDDs and DataFrames, Impala provides real-time querying capabilities directly over HDFS and HBase without moving data, making it faster for interactive analytics.

2. Explain Big Data Integration & its importance with Examples.

Ans: Big Data Integration combines data from various sources like relational databases, cloud storage, and streaming Platforms into a unified view. It ensures data consistency and supports analytics.

for example: Integrating customer data from CRM, web logs, and social media can help businesses Personalize Services.

Section - C

1. Describe mongoDB and its role in Big Data Processing. How does it integrate with Spark?

ans: MongoDB is a NoSQL document-Oriented database Suitable for handling unstructured BigData. It Stores data in flexible, JSON - like documents.

It integrates with Spark using the mongoDB Spark Connector, enabling distributed data processing and real-time analytics on mongoDB datasets.

2. Explain Neo4j and its significance in handling graph - based Big Data applications.

Neo4j is a graph database designed to efficiently store and query graph structures like social networks, recommendation engines and fraud detection systems.

It enables fast traversal of connected data, which is challenging for traditional relational databases.