Explain Hadoop architecture and its components with proper diagram ?

Table of Contents

In Previous Years Questions

Hadoop is a distributed processing framework designed to efficiently process large datasets across clusters of computers.

It consists of four core components, each playing a crucial role in data management and processing:

1. Hadoop Distributed File System (HDFS)

- Function: Stores and manages large datasets across multiple nodes in a cluster.
- Components:

Explain Hadoop architecture and its components with proper diagram ?

    - NameNode: Central server managing file system metadata (data block locations, replication factors).
    - DataNode: Storage nodes where actual data blocks reside.
- Benefits:
    - High availability: Data replicated across nodes ensures access even if some fail.
    - Scalability: Easily expands to accommodate larger datasets by adding nodes.
- Diagram:

---

2. Yet Another Resource Negotiator (YARN)

- Function: Allocates and manages resources (CPU, memory) for applications running on the cluster.
- Components:
    - ResourceManager: Oversees all resource management within the cluster.
    - NodeManager: Manages resources on individual nodes.
    - ApplicationMaster: Negotiates resources for specific applications and coordinates their execution.
- Benefits:
    - Efficient resource utilization: Ensures applications receive necessary resources while maximizing overall cluster performance.
    - Multi-application support: Allows multiple applications to run concurrently on the cluster.
- Diagram:

Explain Hadoop architecture and its components with proper diagram ?

3. MapReduce

- Function: Programming model for parallel processing of large datasets.
- Process:
    - Map phase: Processes data in parallel on individual nodes.
    - Reduce phase: Combines and aggregates results from the map phase to produce final output.
- Benefits:
    - Simplified implementation for large-scale data processing tasks.
    - Efficient parallelization for faster execution.

4. Hadoop Common

- Function: Provides utilities and libraries supporting other Hadoop components.
- Includes:

    - File system operations
    - Networking functionalities
    - Security mechanisms
- Benefits:
    - Facilitates development and interoperability among different Hadoop components.