BIG DATA HADOOP

1: Introduction to Big Data and Hadoop

1.1.What is Big Data?

- 1.1.1. Definition and Characteristics
- 1.1.2. Sources of Big Data

1.2.Challenges of Big Data

- 1.2.1. Volume, Velocity, Variety, and Veracity
- 1.2.2. Privacy and Security Concerns

1.3.Introduction to Hadoop

- 1.3.1. Origins of Hadoop
- 1.3.2. Hadoop's Role in Big Data

2: Hadoop Fundamentals

2.1.Hadoop Distributed File System (HDFS)

- 2.1.1. Architecture and Concepts
- 2.1.2. HDFS Commands and Operations

2.2.MapReduce

- 2.2.1. MapReduce Basics
- 2.2.2. MapReduce Workflow

2.3.YARN (Yet Another Resource Negotiator)

- 2.3.1. Introduction and Role in Hadoop

3: Hadoop Ecosystem

3.1.Hive

- 3.1.1. Introduction to Hive
- 3.1.2. HiveQL (Hive Query Language)

3.2.Pig

- 3.2.1. Introduction to Pig

- 3.2.2. Writing Pig Latin Scripts

3.3.HBase

- 3.3.1. Introduction to HBase
- 3.3.2. HBase Data Model

4: Hadoop Installation and Configuration

4.1.Hadoop Installation

- 4.1.1. Single-Node Cluster Setup
- 4.1.2. Multi-Node Cluster Setup

4.2.Configuration Files

- 4.2.1. Core-site.xml and HDFS-site.xml
- 4.2.2. MapReduce Configuration

4.3.Cluster Management

- 4.3.1. Namenode and Datanode Management
- 4.3.2. ResourceManager and NodeManager

5: Data Ingestion and Processing with Hadoop

5.1.Importing Data into HDFS

- 5.1.1. Hadoop Copy Commands
- 5.1.2. Data Loading Techniques

5.2.Data Processing with MapReduce

- 5.2.1. Writing MapReduce Programs
- 5.2.2. Running MapReduce Jobs

6: Hadoop Data Storage and Management

6.1.Hive for Data Warehousing

- 6.1.1. Creating Hive Tables
- 6.1.2. Running Hive Queries

6.2.Pig for Data Transformation

- 6.2.1. Data Transformation with Pig

- 6.2.2. Complex Data Processing

7: Hadoop Administration and Monitoring

7.1.Cluster Monitoring and Maintenance

- 7.1.1. Monitoring Cluster Health
- 7.1.2. Cluster Upgrades and Patches

7.2.Backup and Recovery

- 7.2.1. Backup Strategies
- 7.2.2. Data Recovery Procedures

8: Advanced Hadoop Topics

8.1.Hadoop Security

- 8.1.1. Authentication and Authorization
- 8.1.2. Kerberos Integration

8.2.Hadoop Ecosystem Extensions

- 8.2.1. Introduction to Spark
- 8.2.2. Integrating Spark with Hadoop

9: Hadoop Use Cases and Best Practices

9.1.Real-World Applications

- 9.1.1. Big Data in Industry
- 9.1.2. Case Studies

9.2.Best Practices

- 9.2.1. Performance Tuning
- 9.2.2. Scalability Strategies

10: Final Project and Course Review

10.1.Project Proposal and Planning

- 10.1.1. Identifying a Real-World Problem
- 10.1.2. Designing a Hadoop Solution

10.2.Implementation and Presentation

- 10.2.1. Building and Deploying the Solution
- 10.2.2. Final Project Presentation