UNIT-II

1 Mark Questions

1. What is Data Science?
2. Define Hadoop?
3. What are the two main parts of Hadoop 1.0?
4. Define HDFS?
5. Define Big Data Analytics?
6. What are the core aspects of Hadoop?

Storage and processing

1. What are the Distributed Computing Challenges of Hadoop?

Hetrogenity and resource sharing concurrency

1. What are the Key Aspects of Hadoop?

Storage scalability utilzation

1. Give example for Use Case for Hadoop?

On click mouse

1. What are the HDFS Daemons?
2. Define YARN?

Yet another resource negotiatior

1. Name any two key points of HDFS?

Block structed file ,default file replication factor,default block size

1. What is NameNode?
2. What is meant by Default Replica Strategy in Hadoop?
3. Name any four HDFS Commands?
4. What is the HDFS command used to get the list of directories and Files?

Hdfs dfs -ls /path

1. What is the Command for copying a file from HDFS to local File System?

Hdfs dfs copyToLocal path

1. How to display the contents of an HDFS file on console?

Hdfs dfs -cat <filename>

1. What are the special features of HDFS?
2. What is Data-Pipeline?

By contrast, "**data pipeline**" is a general term that includes ETL and ELT. It refers to a system for moving **data** from one system to another. The **data** may be processed either in batches or in real time. ELT/ ETL **pipelines**, **data** lakes, and **data**

1. What is meant by Secondary Name Node?
2. What is Anatomy of File Write?
3. What is Data Node?
4. Name any two Hadoop Distributors?

Cdh,hdp

1. Define Hadoop’s Scalability Property?
2. Define Pig?
3. Define NYSE?

New York Stock Exchange

1. What is DFS?

Disrtibuted file system

1. How to process Gigantic Store of Data?

Cluster of computers

1. What is the use of MAHOUT?

Machine learning

1. Write a short note on HDFS. 5M
2. Explain about different modes in Hadoop Cluster? 5M
3. Distinguish Hadoop versus RDBMS. 5M
4. Explain Hadoop Eco System with neat sketch. 10M
5. Write about Data serialization formats used in Hadoop MapReduce. 10M