2023-CSE5BDC-T5-W - Big Data Management on the Cloud

Big Data Management on the Cloud Week 2. Quizzes

Topic 1: Introduction to Hadoop

##### **Quiz: Check your understanding 1**

##### **Which of the following statements is FALSE when describing the Hadoop Distributed File System (HDFS)?**

Answer: Exclusively designed to store relational data.

**Which of the following are advantages that Hadoop has over traditional RDBMS?**

-Node recovery

-High horizontal scalability

-Fault-tolerance

NameNode (NN) is a

**centralised**

 service/daemon that acts as a cluster

**storage**

 manager. Its primary responsibilities are:  
  
• maintaining

**meta-data**

 of files and directories.  
• controlling and coordinating

**Data Nodes**

 for file system operations such as create, read, write, etc.

Clients communicate with NN in order to perform everyday file system operations. In turn, NN gives clients the

**location**

 of DNs in the cluster (where the data block is available) to carry out operations.

**Quiz: Check your understanding 2**

**Which of the following is NOT a property of Hadoop?**

Fast data updates

**Which of the following are the responsibilities of the Job Tracker (there is more than one correct answer)?**

* Preparing execution plan and job coordination.
* Scheduling map/reduce tasks to run on TaskTrackers

**TaskTracker/taskTracker** is a daemon that runs in every slave node in the **Hadoop/hadoop** cluster. It is responsible for starting and managing **map/Map/reduce tasks/Reduce tasks** in its dedicated slots. It periodically sends a report to **JobTracker/jobTracker** about resource availability, **task progress/Task progress**, and its **status/Status**.

TaskTrackers send HeartBeat to JobTracker to indicate ready for service, similar to

**NameNode**

 sending the same type of indication to

**DataNodes**

 in HDFS. If a DataNode does not

**receive a heartbeat message**

 every 10 minutes, it is dead and will be removed from HDFS cluster.

## Topic 2: Introduction to Hive

**Quiz: Check your understanding 3**

**Which of the following is a characteristic of Hive?**

Hive has no control over underlying storage, so it can only enforce queries on read, i.e., schema on read.

**Which of the following command would create a table in Hive, if the table does not already exist? (You can select as many answers as you want.)**

CREATE TABLE IF NOT EXISTS Employees

CREATE TABLE Employees

Which of the following command would append data to the table Employees?

INSERT INTO TABLE Employees

## Topic 3: Querying data in Hive

**Quiz: Check your understanding 4**

**Which of the following is NOT a type of User-Defined Functions (UDFs)?**

User-Defined Macro Functions (UDMFs)

Which one of the following is NOT correct about ORDER BY and SORT BY?

ORDER BY runs faster than SORT BY