Date:

HYUNDAI

Integrate with different data storage systems and processing frameworks, providing flexibility in building data processing pipelines.

7) Extensibility- The hadoop ecosystem is highly extensible, with a wide range of open-source projects and tools that extend its functionality. These include apache hive for data warehousing, Apache pig for data flow scripting, Apache Hbase for real-time read/write access, and Apache Spark for in-memory processing, among others.

Date

HYUNDA

Q// hadoop characteristic and features.

1) Distributed Computing - Hadoop is designed for distributed computing, allowing it to process large dataset by distributing the workload

2) Scalability - Hadoop is highly scalable, meaning it can easily accommodate growing amounts of data by adding more nodes to the cluster.

3) Fault tolerance - Hadoop is fault-tolerance, meaning it can run on function even if some of the nodes in the cluster fail. data replication

4) Cost - Effectiveness - Hadoop is built on commodity hardware, meaning it can run on inexpensive, off-the-shelf hardware components. This makes it a cost-effective solution for storing.

5) Parallel Processing - Hadoop leverages the map-Reduce Programming model for parallel processing of data. This allows it to process data in parallel across multiple nodes in the cluster

6) flexibility - Hadoop is flexible and can handle various types of data, including structured, semi-structured, and unstructured data. It can also

EVERYTHING HAS TO KEEP MOVING FORWARD IN 'ENDEAVOUR.'
OTHERWISE, IT WILL STAGNATE. - SHAUN EVANS

Date:

HYUNDAI

Name node → This is like the central manager of the hadoop filling filling system (HDFS). It keeps track of where all the files are stored in e cluster and manages access to them.

Data node → These are the workers of hadoop. They actual store the ctual data in the hadoop filing system. ach datanode manages the storage of data on the machine it runs on.

) Resource manager → In a hadoop ~~clust~~ cluster, Resource Manager is esponsible for managing the computing esources available in the cluster.

) Node manager - Node managers are ~~respon~~ responsible for managing the computing resource (CPU and memory) o individual machines in the cluster.

They report back to the the resou Manager about available resource and execute tasks assigned ~~to~~ to them.

MOVE FAST. SPEED IS ONE OF YOUR MAIN ADVANTAGES OVER LARGE COMPETITORS. - SAM ALTMAN

Date

HYUNDAI

Q what is hadoop?

⇒ Hadoop is ~~like~~ a popular open-source framework used for distributed storage and processing of large datasets across cluster of computers.

It is design to handle massive amounts of data in a cost-effective and scalable manner.

⇒ Distributed storage →

⇒ MapReduce → 1) Map phase 2) Reduce phase

⇒ Fault Tolerance →

⇒ Scalability →

Q/ what are hadoop "daemons"?

⇒ In Hadoop, "daemons" are special ~~progra~~ programs or process that run in the background and handle various tasks within the hadoop ecosystem.

PERSISTENCE IS THE ABILITY TO KEEP MOVING FORWARD IN SPITE OF DIFFICULTIES. - BYRON PULSIFER

Date

HYUNDAI

Q. History of hadoop.

⇒ The history of hadoop to begins in the early ~~2000~~ 2000s at yahoo!, where engineers Doug cutting and mike cafarella were working on a project to index the web.

In 2006, Doug cutting created an open-source implementation of ~~google~~ google's Mapreduce programming model and ~~distributed~~ distributed file system called 'Hai Hadoop'.

⇒ Named after cutting's son's toy elephant.

⇒ It become an apache software foundation project in 2008.