



Date:

iii) Now to check the location is directory or not :-

→ HDFS -test -d

MapReduce

⇒ HV<sub>1</sub> → Map-Reduce → Job Tracker  
Task Tracker ]  
daemon

⇒ HV<sub>1</sub>, HV<sub>2</sub>, HV<sub>3</sub>

1) This is the initial version of Hadoop.  
It includes the ~~HDFS~~ HDFS for storage  
and MapReduce for processing data in  
parallel across across a distributed  
cluster.

2) This is the second version with  
new features compared to HV<sub>1</sub>.  
It includes YARN (yet another Resour  
Negotiator),

HYUNDAI

which serve as a resource management, and job scheduling layer, allowing hadoop to support multiple processing engines beyond MapReduce, such as Apache spark.

3) The next major release after hadoop v2, offering further enhancements, optimization, and new features. Some of the key improvements include erasure coding support in HDFS for more efficient data storage, enhancements to YARN for better resource utilization and scalability.

Q2

Hadoop  $\Rightarrow$  primarily designed for distributed storage and processing of large volume of data across across clusters of commodity hardware.

Data warehouse  $\Rightarrow$  Designed for storing and managing structured data from various sources in a centralized repository, optimized for querying and analysis.

YOU ARE ALWAYS A STUDENT, NEVER A MASTER. YOU HAVE TO  
KEEP MOVING FORWARD. - CONRAD HALL

i) Hadoop  $\Rightarrow$  Capable of handling diverse data types, including structured, semi-structured, and unstructured data.

Data warehouse  $\Rightarrow$  Typically optimize for structured data, such as transactional data from relation database.

ii) Hadoop  $\Rightarrow$  utilizes a distributed processing framework, such as MapReduce or Apache Stark, for parallel data processing.

Data warehouse  $\Rightarrow$  Often employs SQL-based querying for performing ad-hoc queries, analytics, and reporting.

iv) Hadoop  $\Rightarrow$  Designed to scale horizontally by adding more commodity hardware to the cluster, allowing it to efficiently process and store petabytes of data.

HAVING PERSEVERANCE IS ALL ABOUT BEING ABLE TO BOUNCE BACK AND KEEP MOVING FORWARD. - JASON SCOTT

▲ HYUNDAI

Data warehouse ↪

Typically scaled vertically  
by upgrading hardware resources, which may  
have limitations in handling extremely  
large datasets compared to hadesb.