AMC ENGINEERING COLLEGE

18th KM, Bannerghatta Road, Kalkere, Bangalore – 560 083
Phone : 27828655, Telefax: 27828656
E-mail :principal@amcec.edu.in, Website: www.amcec.edu.in

SEMINAR REPORT ON
HADOOP

Name   : ....

USN   : ....

Branch   : ....

Semester   : ....

CONTENTS

✓What is hadoop Technology??

✓Why hadoop?

✓Developers of hadoop Technology

✓Famous hadoop users

✓Hadoop Features

✓Hadoop Architectures

✓Core-Components of Hadoop

✓Hadoop High Level Architechture

✓Hadoop cluster

CONTENTS…

✓What is HDFS

✓HDFS – Name Node features:

✓HDFS-name node architecture

✓HDFS-data node

✓Hadoop MAPREDUCE
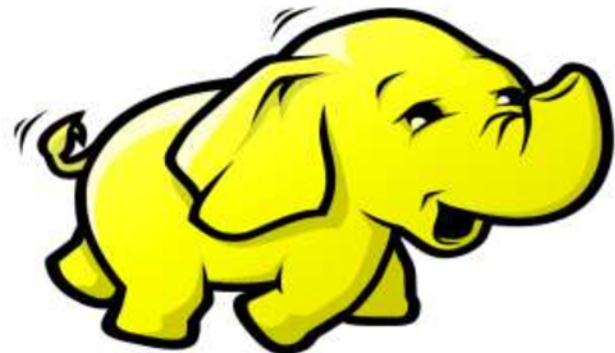
✓Benefits of Hadoop…

✓Conclusion

✓Reference

HADOOP TECHNOLOGY

What is Hadoop Technology??

- The most well known technology used for Big Data is Hadoop.

- It is actually a large scale batch data processing system

Why Hadoop ??

- Distributed cluster system

- Platform for massively scalable applications

- Enables parallel data processing

Hadoop Features

- Hadoop provides <u>access to the file systems</u>

- The Hadoop Common package contains the necessary <u>JAR files</u> and <u>scripts</u>

- The package also provides <u>source code</u>, <u>documentation</u> and a <u>contribution section</u> that includes projects from the Hadoop Community.

Core-Components of Hadoop:

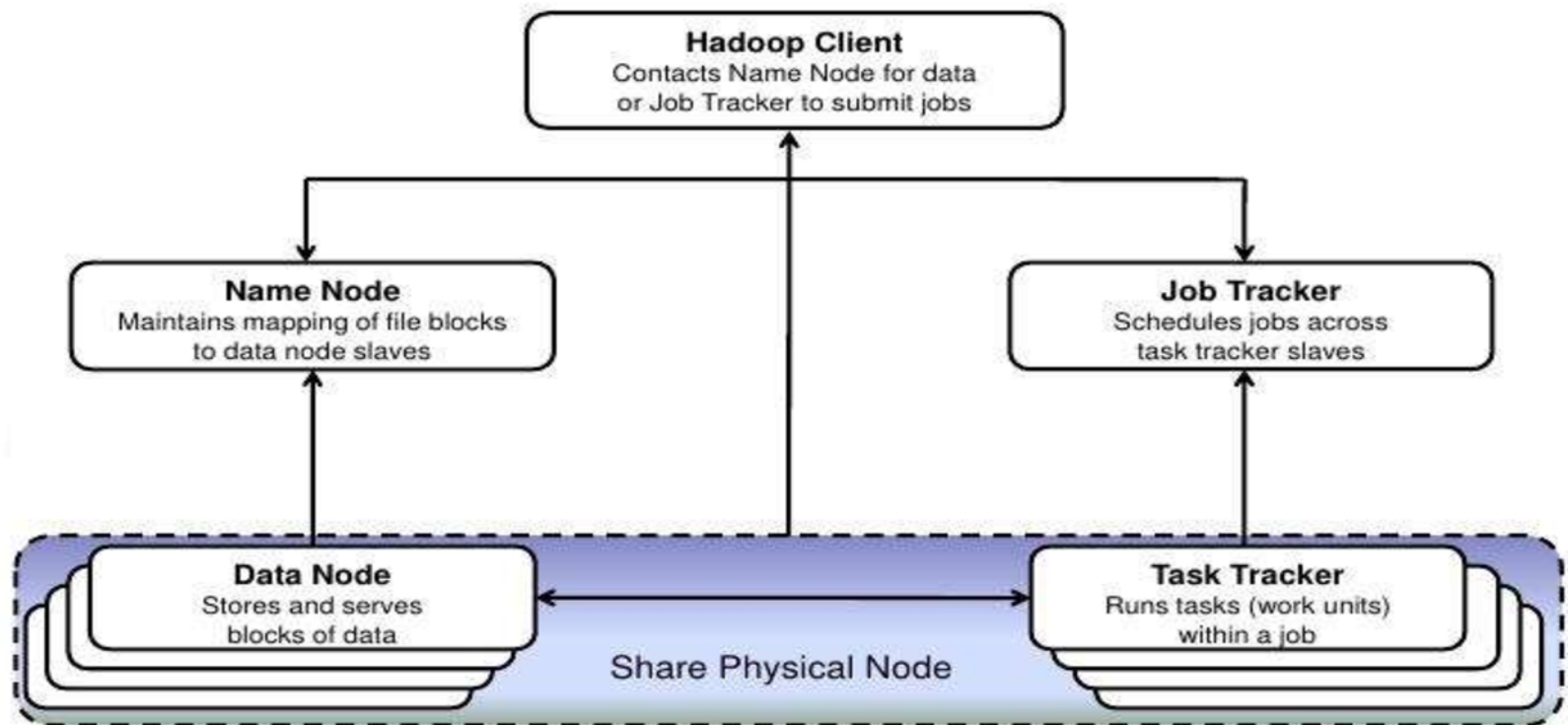Hadoop distributive file system.

Map reduce.

What is HDFS ?

- Distributed file system

- Traditional hierarchical file organization

- Single namespace for the entire cluster

- Write-once-read-many access model

- Aware of the network topology

Hadoop High Level Architechture

Hadoop Client
Contacts Name Node for data
or Job Tracker to submit jobs

Name Node
Maintains mapping of file blocks
to data node slaves

Job Tracker
Schedules jobs across
task tracker slaves

Data Node
Stores and serves
blocks of data

Task Tracker
Runs tasks (work units)
within a job

Share Physical Node

Hadoop cluster

• A Small Hadoop Cluster Include a single master &

multiple worker nodes

Master node:
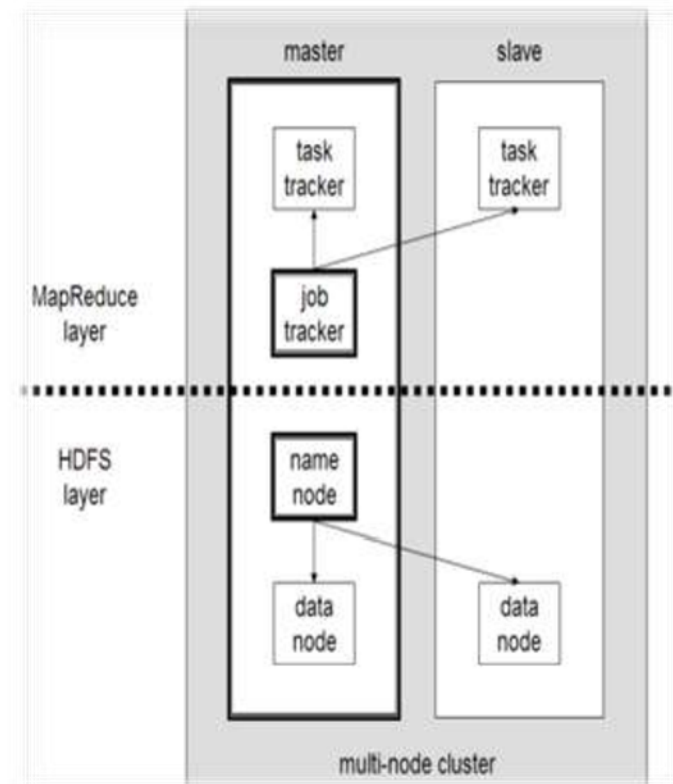Data Node
Job Tracker
 Task Tracker
 Name Node

Slave node:
   Data Node
   Task Tracke

HDFS – Name Node Features

Metadata in main memory:

- List of files

- List of blocks for each file

- List of Data Nodes for each block

- File attributes

- Creation time

- Records every change in the

metadata

HDFS-Data node

• Block Server Stores data in the local file system

• Periodic validation of checksums

• Periodically sends a report of all existing blocks
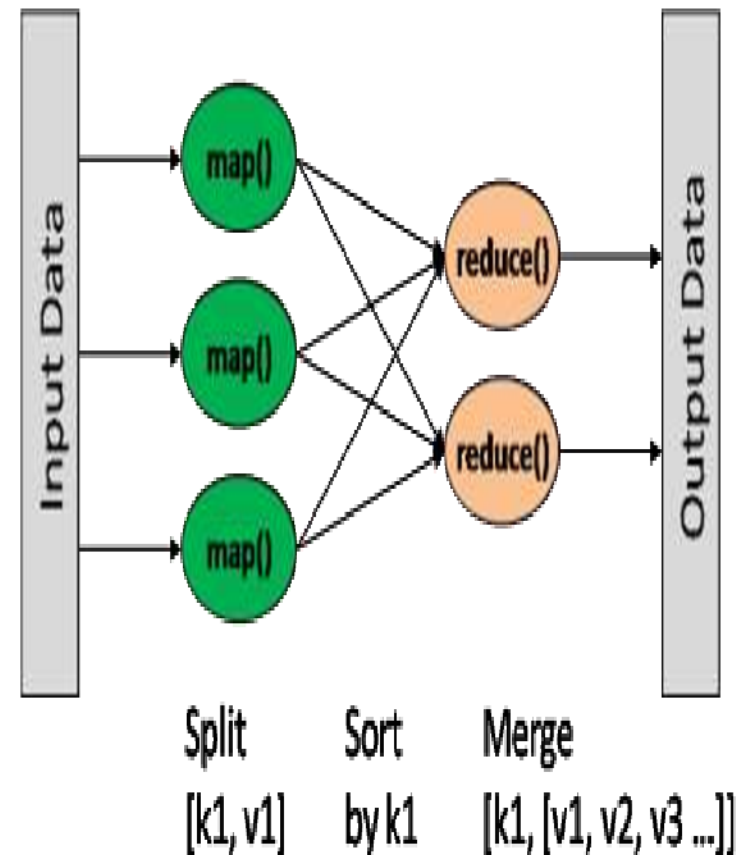
to the Name Node

Hadoop MAPREDUCE

Map reduce implementation:

Job Tracker:
Splitting into map and reduce tasks
Scheduling tasks on a cluster node
Task Tracker:
Runs Map Reduce tasks periodically

Benefits of Hadoop…

- Cost Saving and efficient and reliable data processing

- Provides an economically scalable solution

- Storing and processing of large amount of data

- Data grid operating system

- It is deployed on industry standard servers rather than expensive specialized data storage systems.

- Parallel processing of huge amounts of data across inexpensive, industry-standard servers.