

Technical Seminar

on

HADOOP TECHNOLOGY

Under the Guidance of

P.V.R.K.MURTHY, M.Tech

Assistant Professor

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 Architecture
- ✓Hadoop cluster

CONTENTS...

- ✓What is HDFS
- ✓HDFS – Name Node features:
- ✓HDFS-name node architecture
- ✓HDFS-data node
- ✓Hadoop MAPREDUCE
- ✓Benefits of Hadoop...
- ✓Conclusion
- ✓Reference

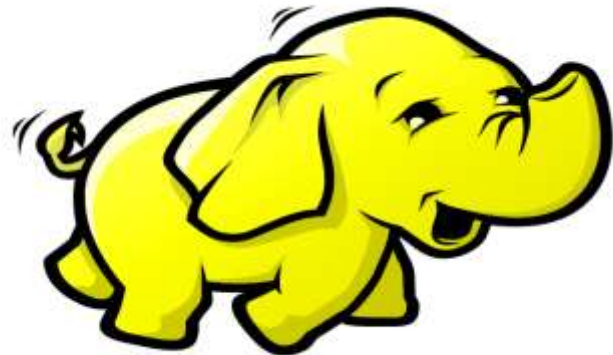
HADOOP TECHNOLOGY

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



Developers of Hadoop Technology:



Michael j. cafarella



Doug cutting

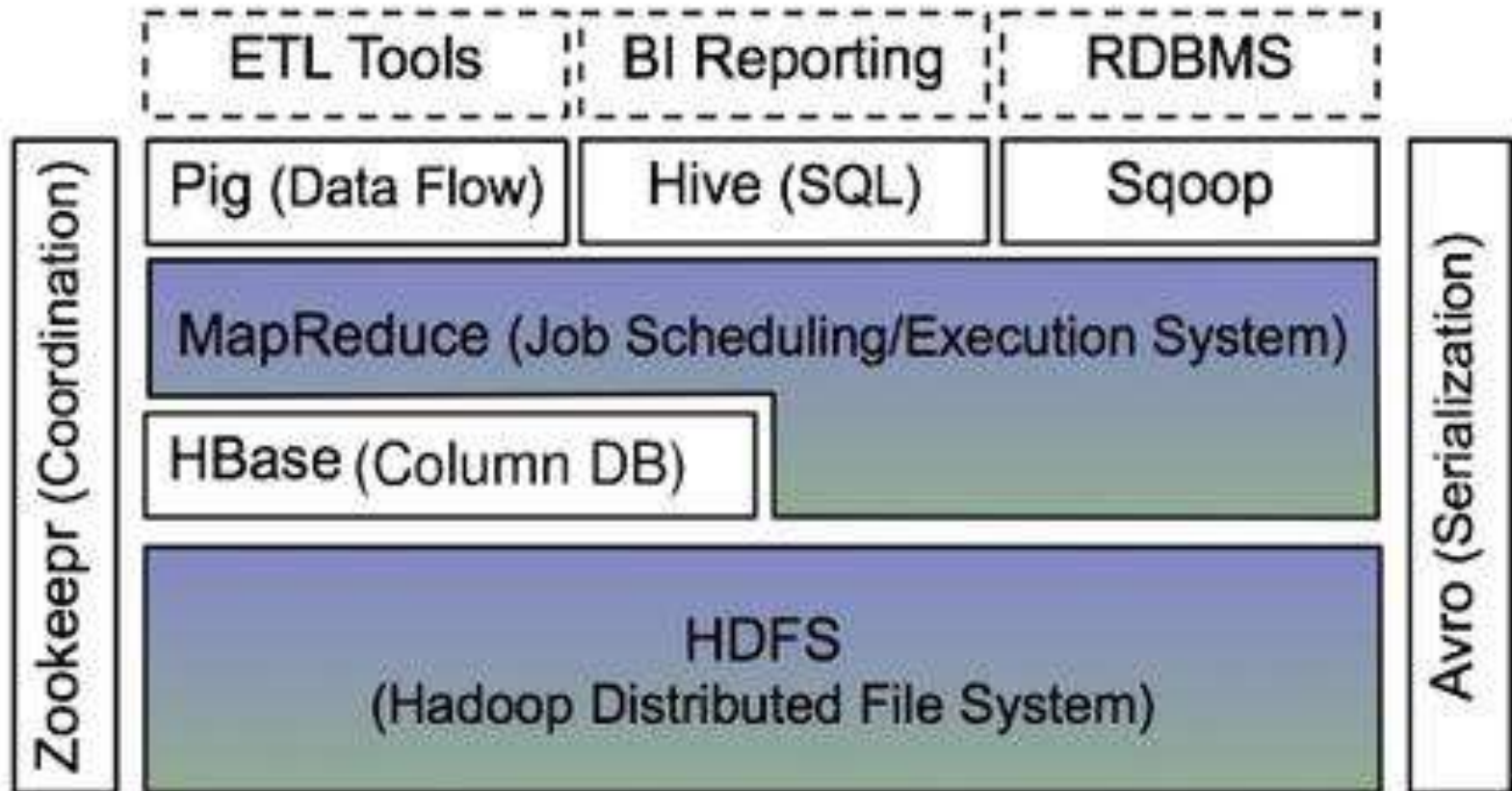
Famous Hadoop users



Hadoop Features

- Hadoop provides access to the file systems
- The Hadoop Common package contains the necessary JAR files and scripts
- The package also provides source code, documentation and a contribution section that includes projects from the Hadoop Community.

HADOOP ARCHITECTURE



Core-Components of Hadoop:



Hadoop distributive file system.

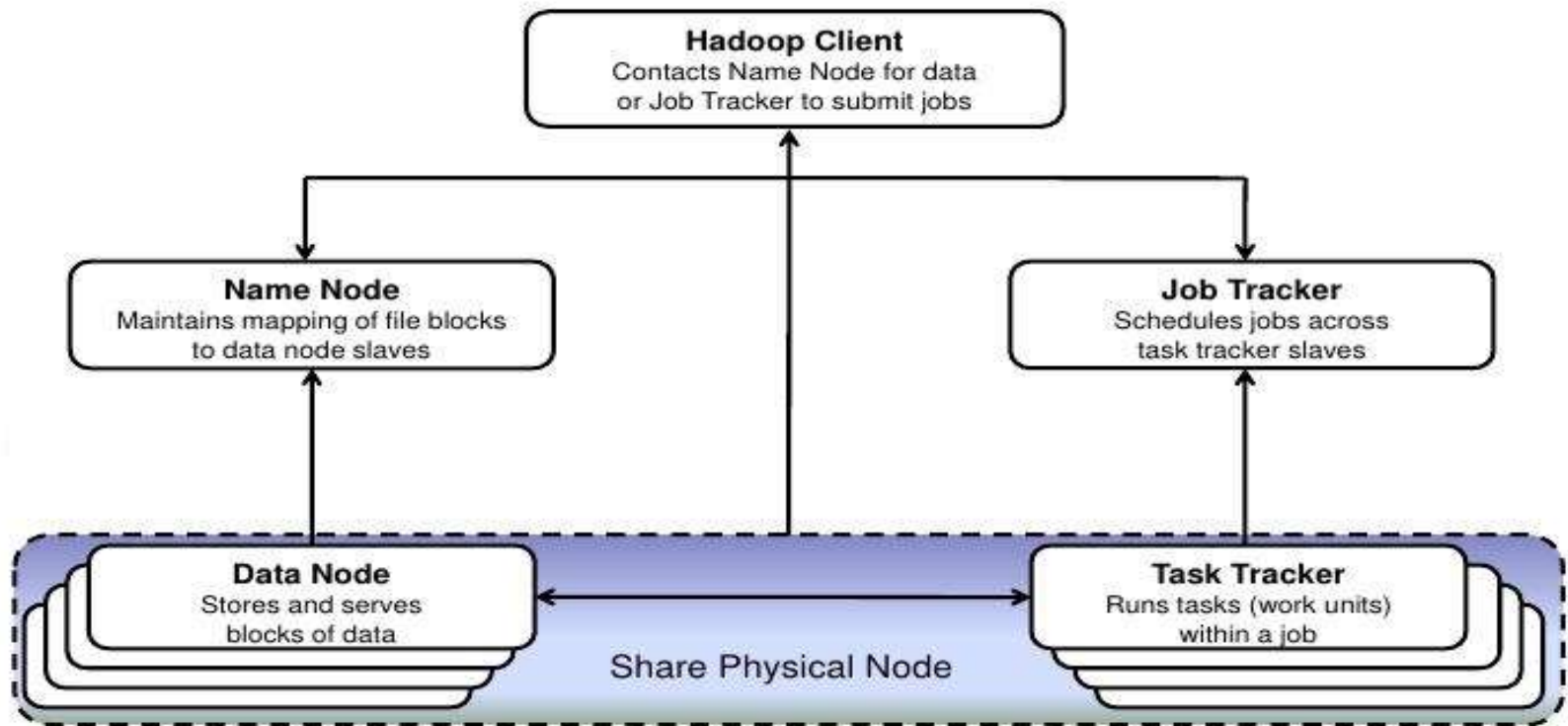


Map reduce.

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 Architecture



Hadoop cluster

- A Small Hadoop Cluster Include a single master & multiple worker nodes

Master node:

Data Node

Job Tracker

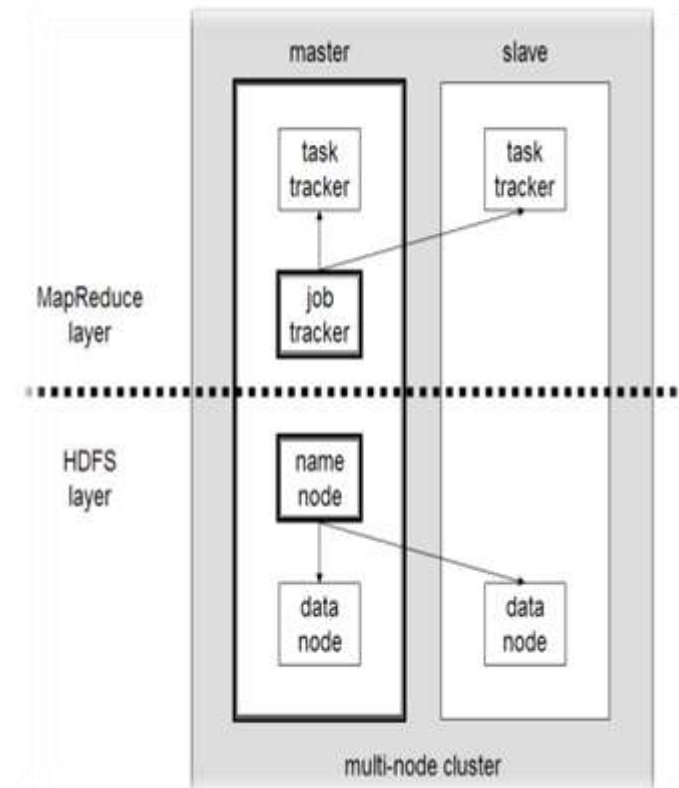
Task Tracker

Name Node

Slave node:

Data Node

Task Tracker

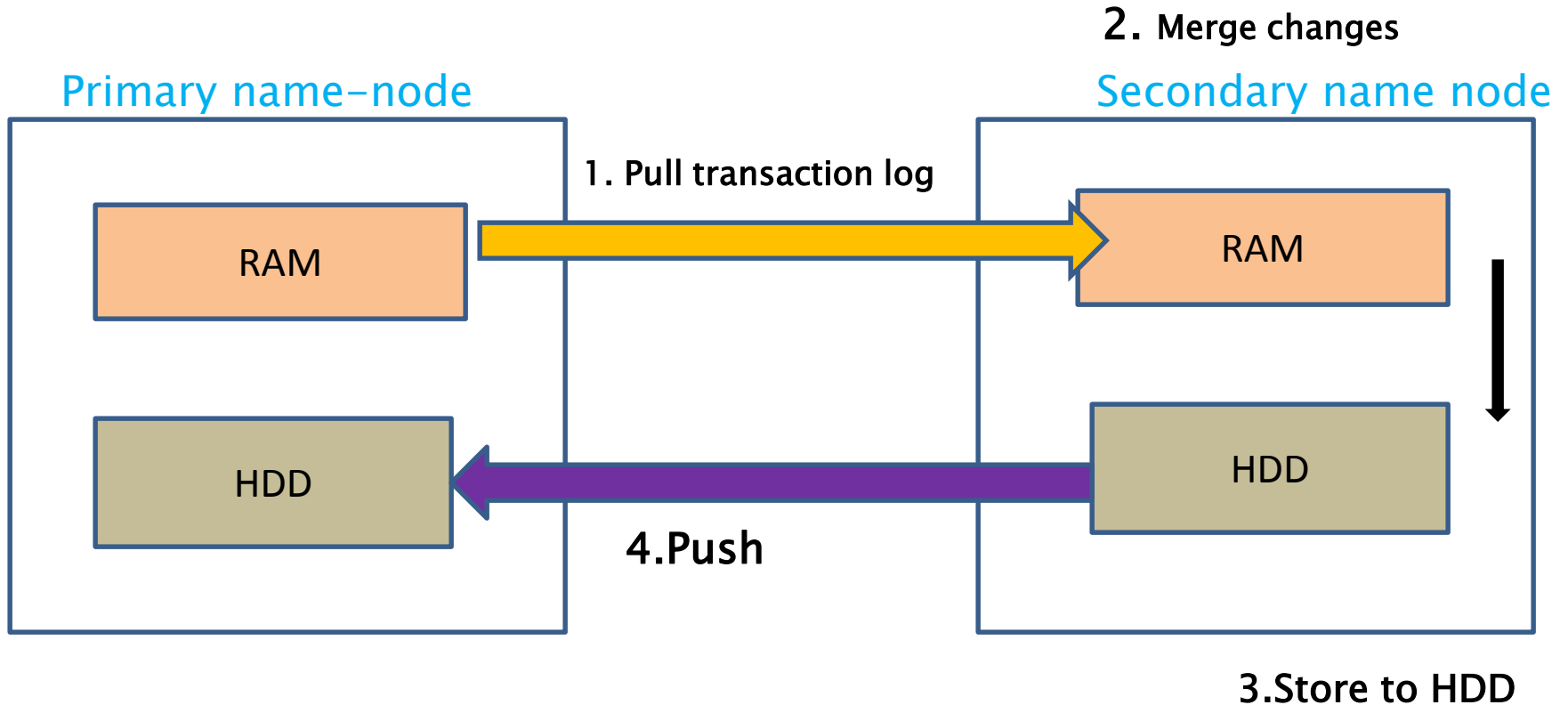


HDFS – Name Node Features

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-name node architecture



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

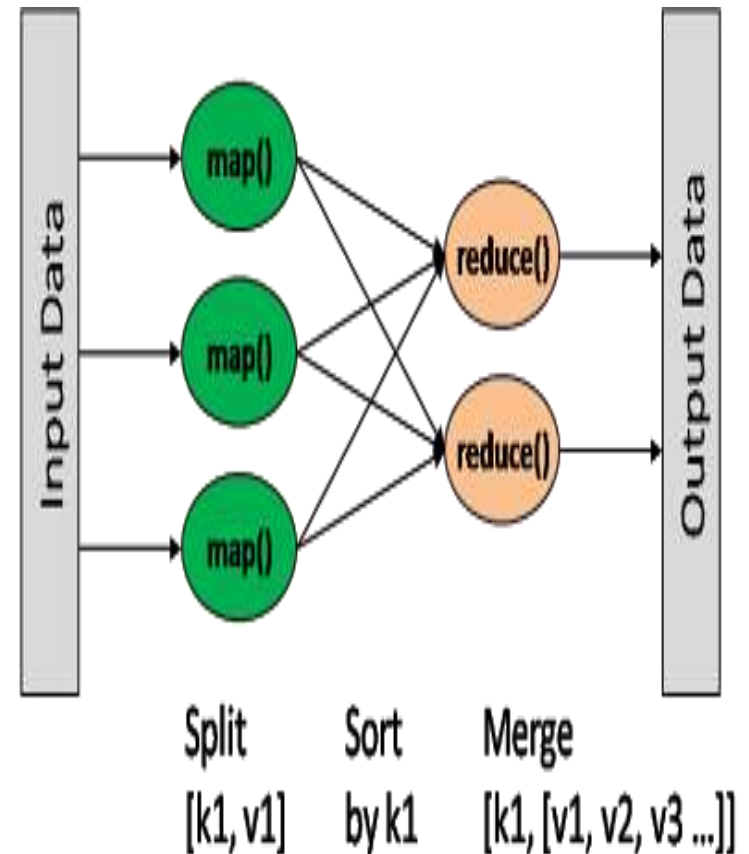
Map reduce implementation:

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.

CONCLUSION

Why commodity hw ?

- ✓ because cheaper
- ✓ designed to tolerate faults

Why HDFS ?

- ✓ network bandwidth vs seek latency

Why Map reduce programming model?

- ✓ parallel programming
- ✓ large data sets
- ✓ moving computation to data
- ✓ single compute + data cluster

REFERENCES

- **Apache Hadoop!**

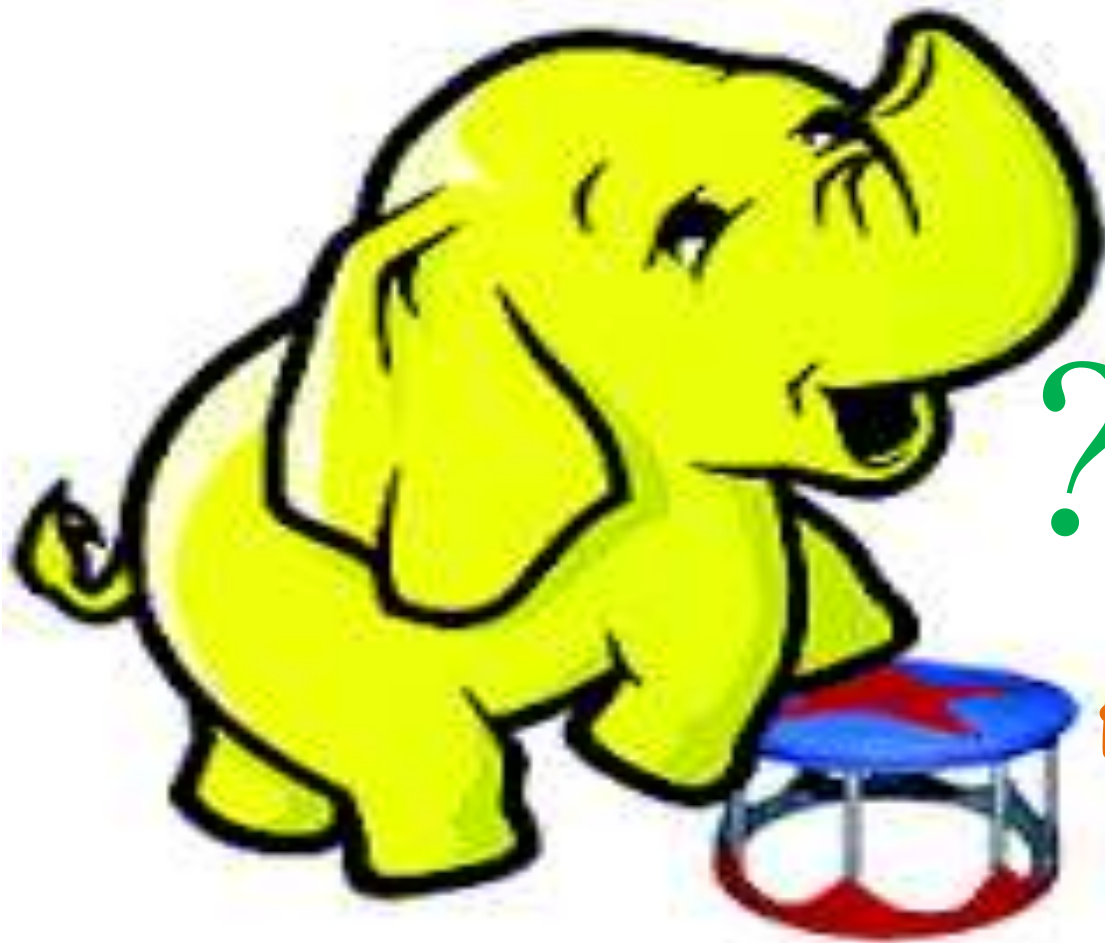
(<http://hadoop.apache.org>)

- **Hadoop on Wikipedia**

(<http://en.wikipedia.org/wiki/Hadoop>)

- **Cloudera - Apache Hadoop for the Enterprise**

(<http://www.cloudera.com>)



? ? ?
Any
Queries



Thank
you