

# Lecture 11.1

# BigData and Hadoop

**School of Computing and Information Technology**

2020-2021

# Outline

Recap of previous Lecture

---

Topic for the Lecture

---

Objective and Outcome of Lecture

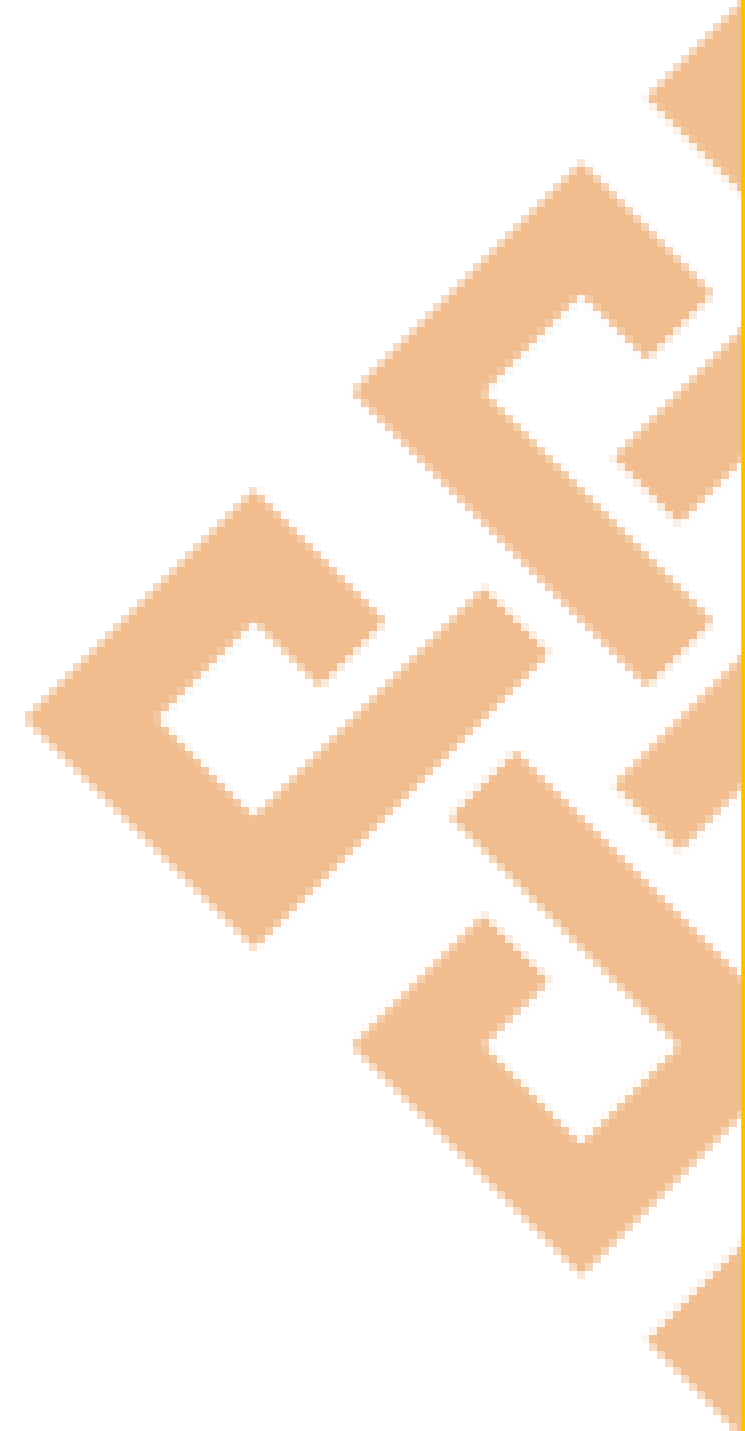
---

Lecture Discussion



# **History of Hadoop**

**Recap of previous Lecture**



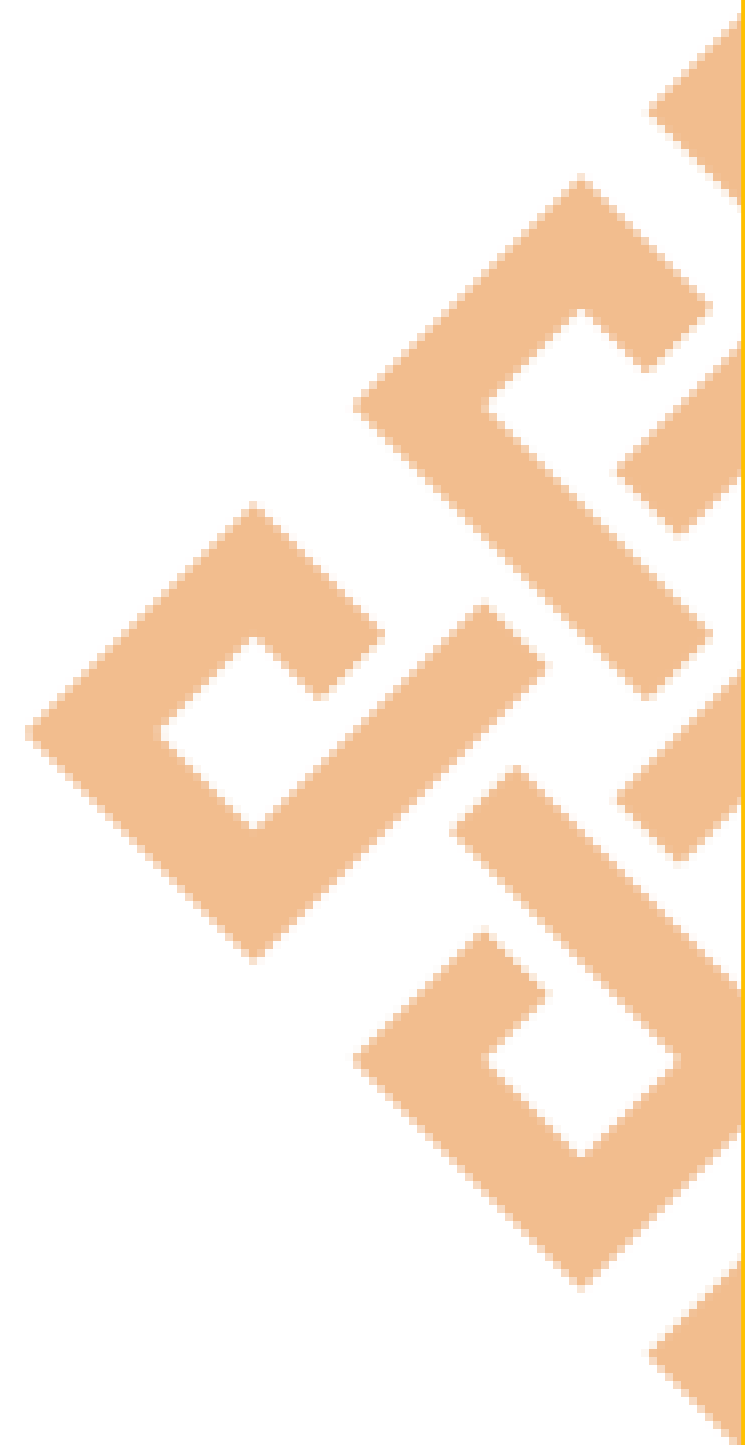
# Recap of Previous Lecture

History of Hadoop



# **Over View of Hadoop**

**Topic of the Lecture**



# Topic of the Lecture

Components of Hadoop

Key aspects of Hadoop

Hadoop ecosystem



# **Hadoop Components**

**Objective and Outcome of Lecture**



# Objective and outcome of lecture

## Lecture Objective

Explain why Hadoop components and Ecosystem.

## Lecture Outcome

Outline the Hadoop components and Ecosystem.





# Hadoop Overview

Basically Hadoop accomplishes 2 Tasks

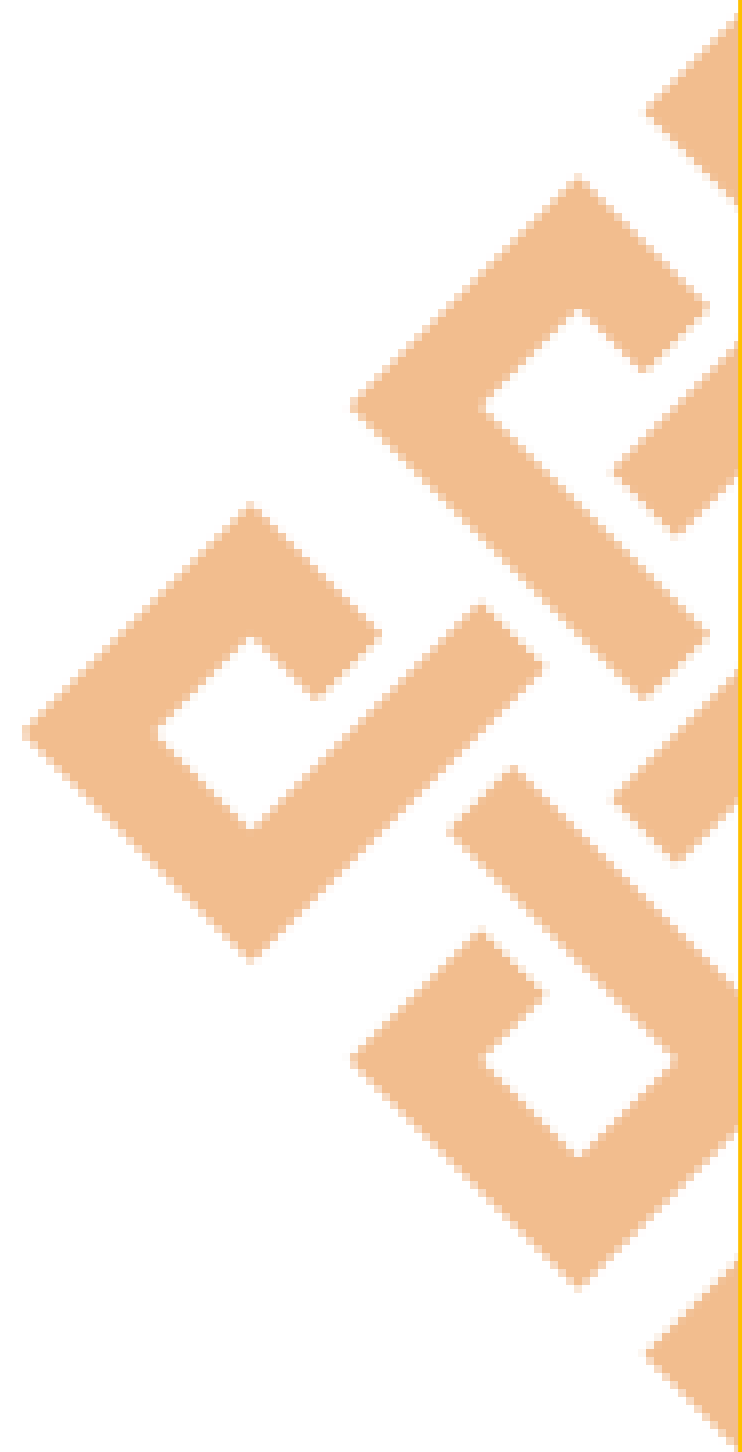


Massive data storage

Faster data processing



# Key Aspects of Hadoop



# Key aspects of Hadoop

Open source software: It is free to download.

Framework: Everything is provided –programs ,tools etc.

Distributed : Divides and stores data across multiple computers.

Massive data storage: stores huge amount of data across nodes of low-cost commodity

Faster data processing: large amount of data is processed in parallel.



# Use cases of Hadoop

clickStream Data:helps to understand the purchasing behaviour of customers. clickStream also helps to online marketers to optimize their product web pages,promotional contents etc.



# Use cases of hadoop



clickStream analysis using Hadoop provides following 3 benefits.

1.Hadoop helps to join clickStream data with other data sources such as customer relationship data.this additional data often provides the much needed information to understand customer behaviour.

2.hadoop scalability property helps you to store years of data without ample increment cost.

3.business analysis can use Apache pig or Apache Hive for website analysis.with these tools we can organize clickStream data by user session ,refine it ,and feed it to visualization or analytics tools



# Lecture 11.2

# BigData and Hadoop

School of Computing and Information Technology

2020-2021

# **Outline**

---

**Recap of previous Lecture**

---

---

**Lecture Discussion**

---

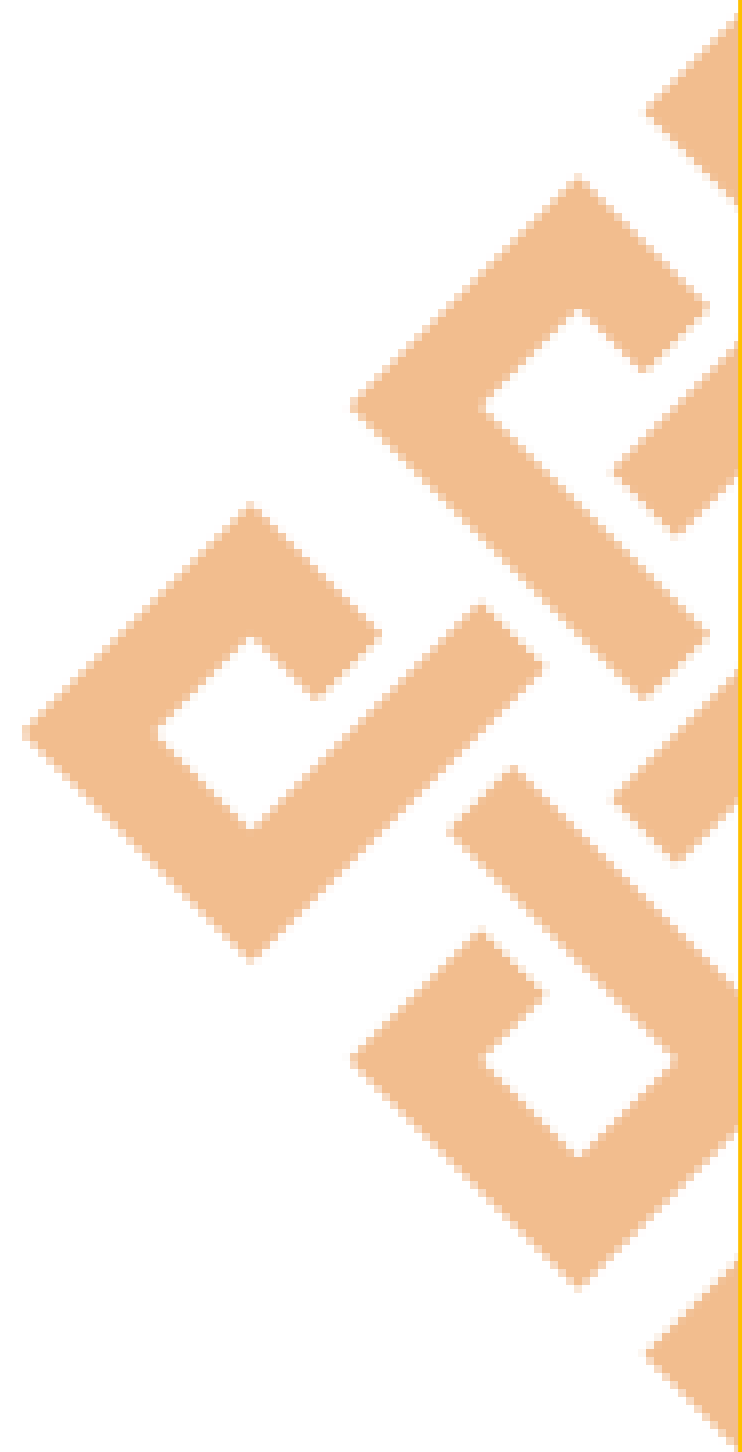
**Overview of The Hadoop**

---

**Summary**



# Overview of the Hadoop





# Recap of Previous Lecture

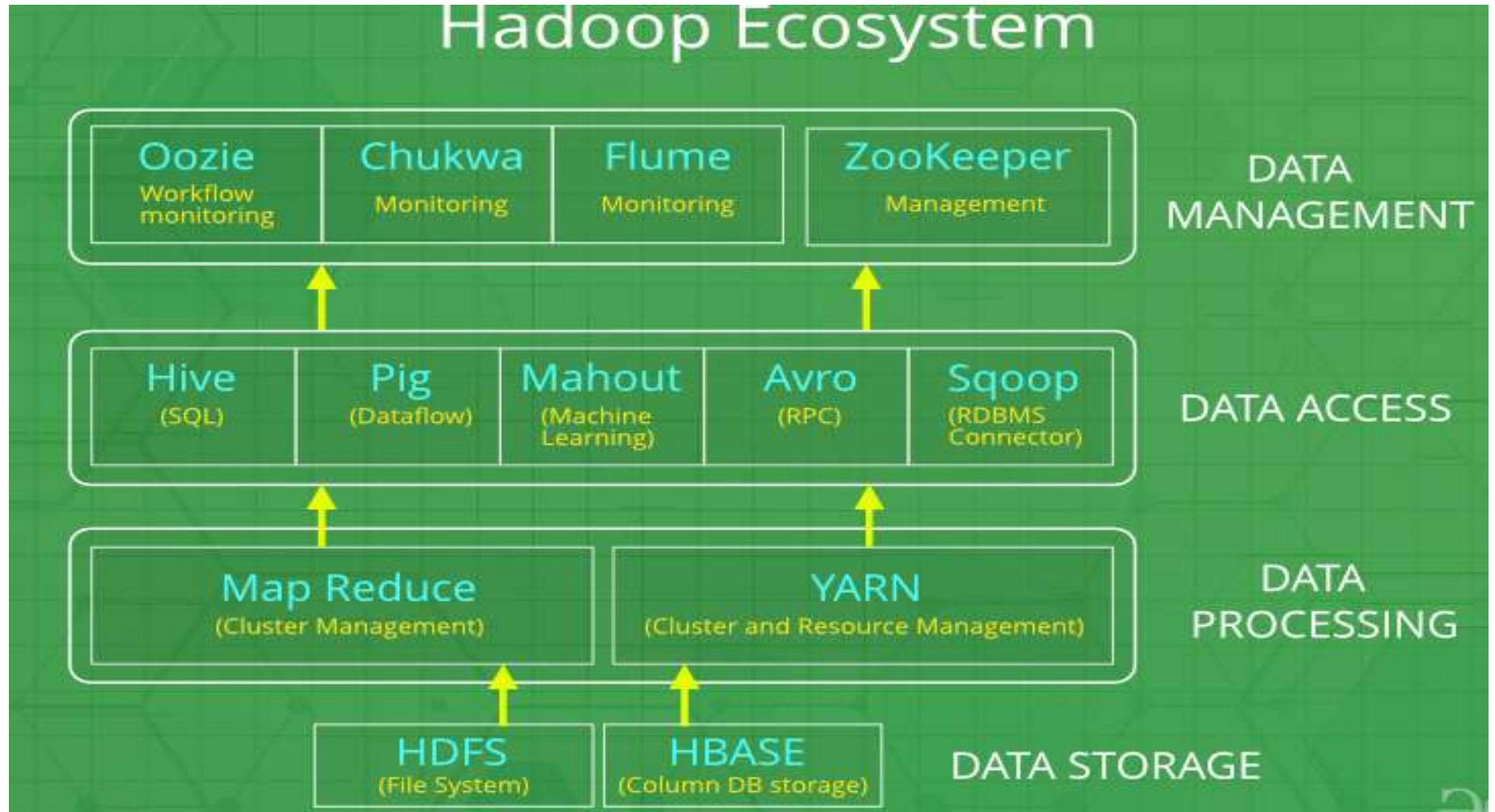
Components of Hadoop

Key aspects of Hadoop

Use Cases of Hadoop

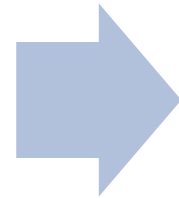


# Hadoop Ecosystem



## Contd.....

*Hadoop Ecosystem* is a platform or a suite which provides various services to solve the big data problems.



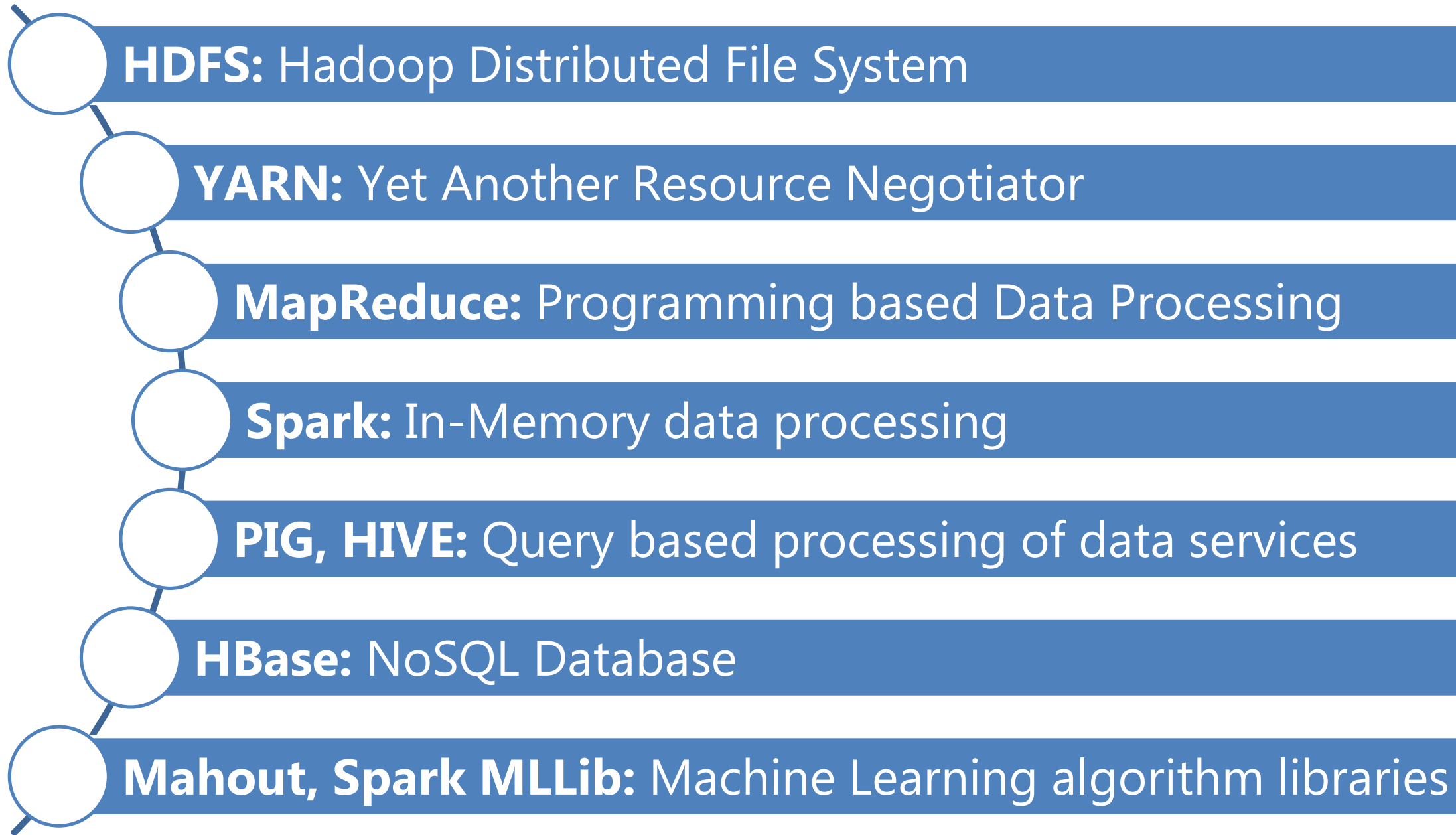
It includes Apache projects and various commercial tools and solutions. There are *four major elements of Hadoop* i.e. HDFS, MapReduce, YARN, and Hadoop Common.



Most of the tools or solutions are used to supplement or support these major elements. All these tools work collectively to provide services such as absorption, analysis, storage and maintenance of data etc.



# Components that Collectively form a Hadoop Ecosystem



# Summary of the Lecture

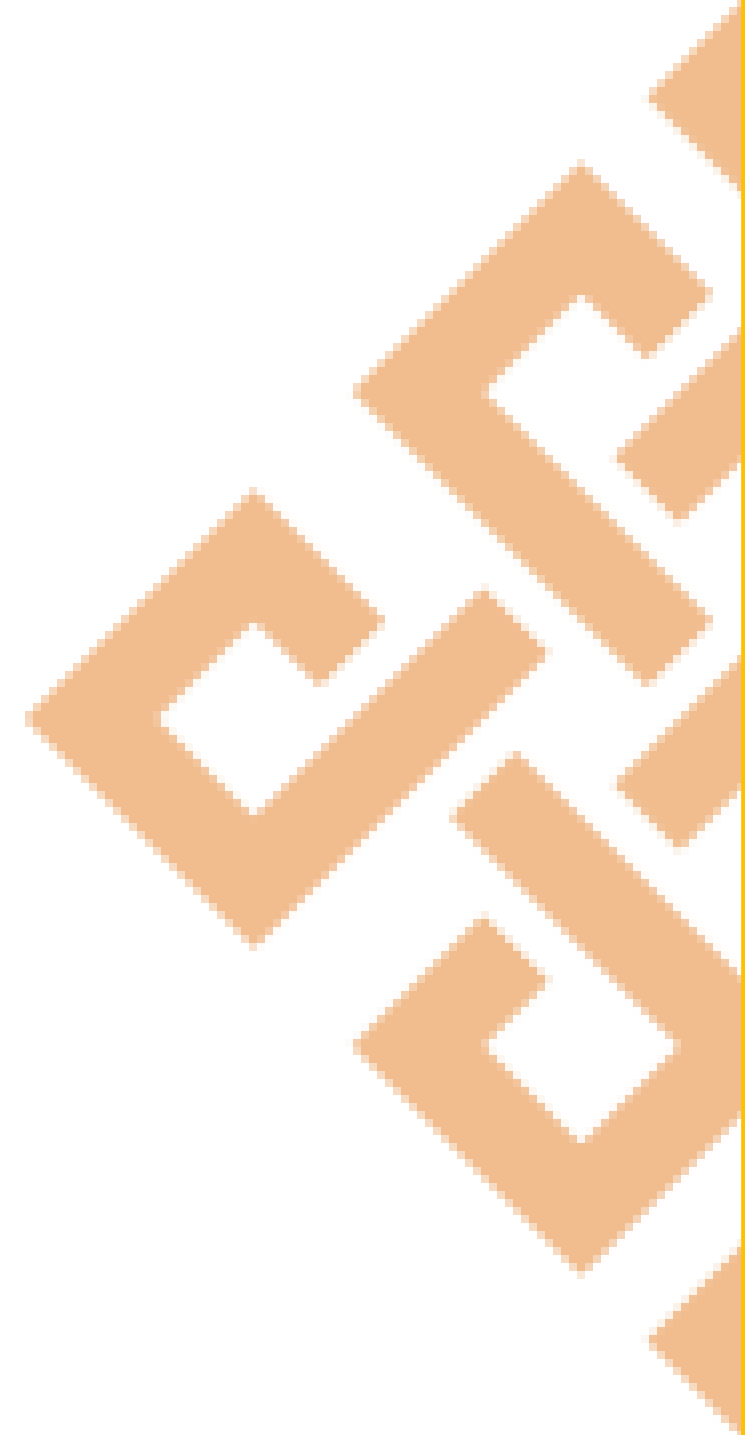
Over view  
of Hadoop

Hadoop  
Ecosystem



# Overview of Hadoop

**Resources and Tasks to be completed**



# Lecture 12.1

## BigData and Hadoop

**School of Computing and Information Technology**

2020-2021

# Outline

Recap of previous Lecture

---

Topic for the Lecture

---

Objective and Outcome of Lecture

---

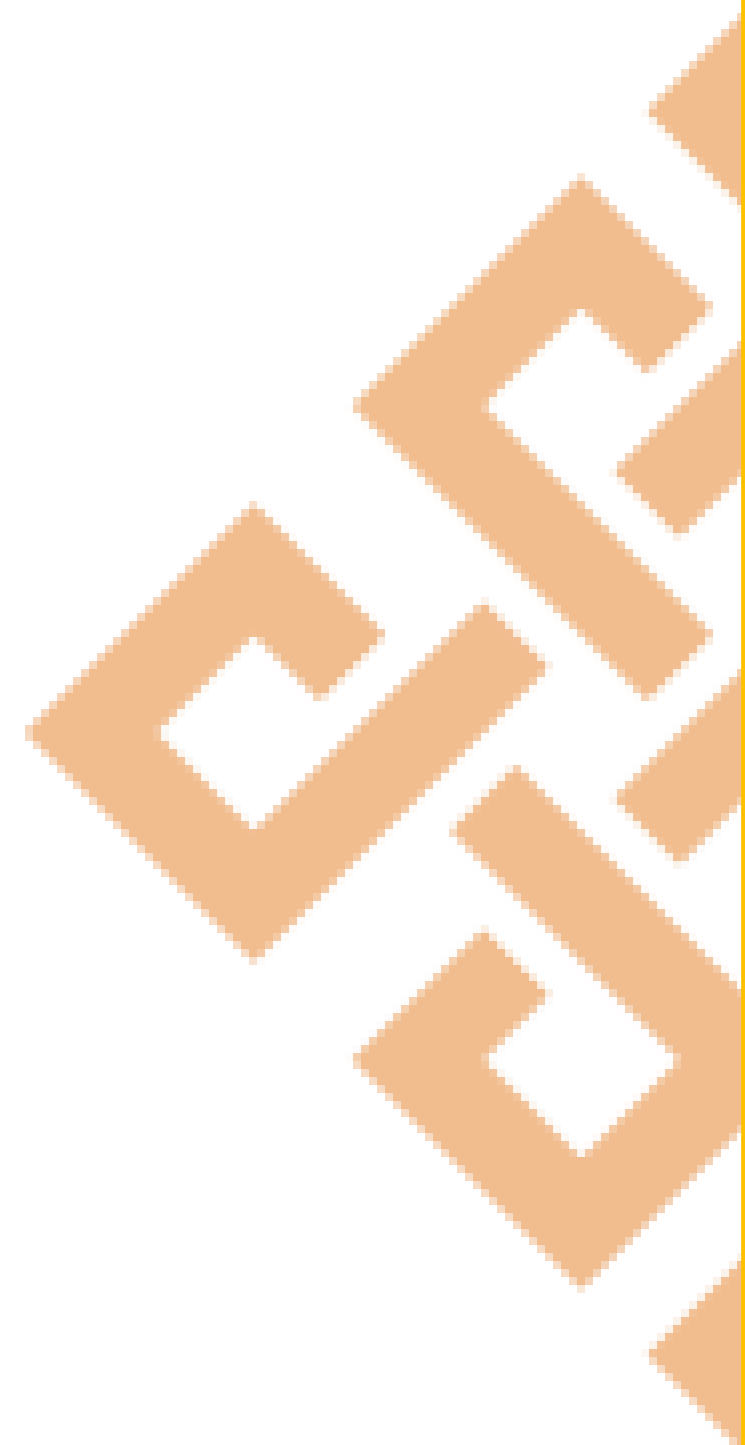
Lecture Discussion





# **Hadoop Ecosystem**

**Recap of previous Lecture**



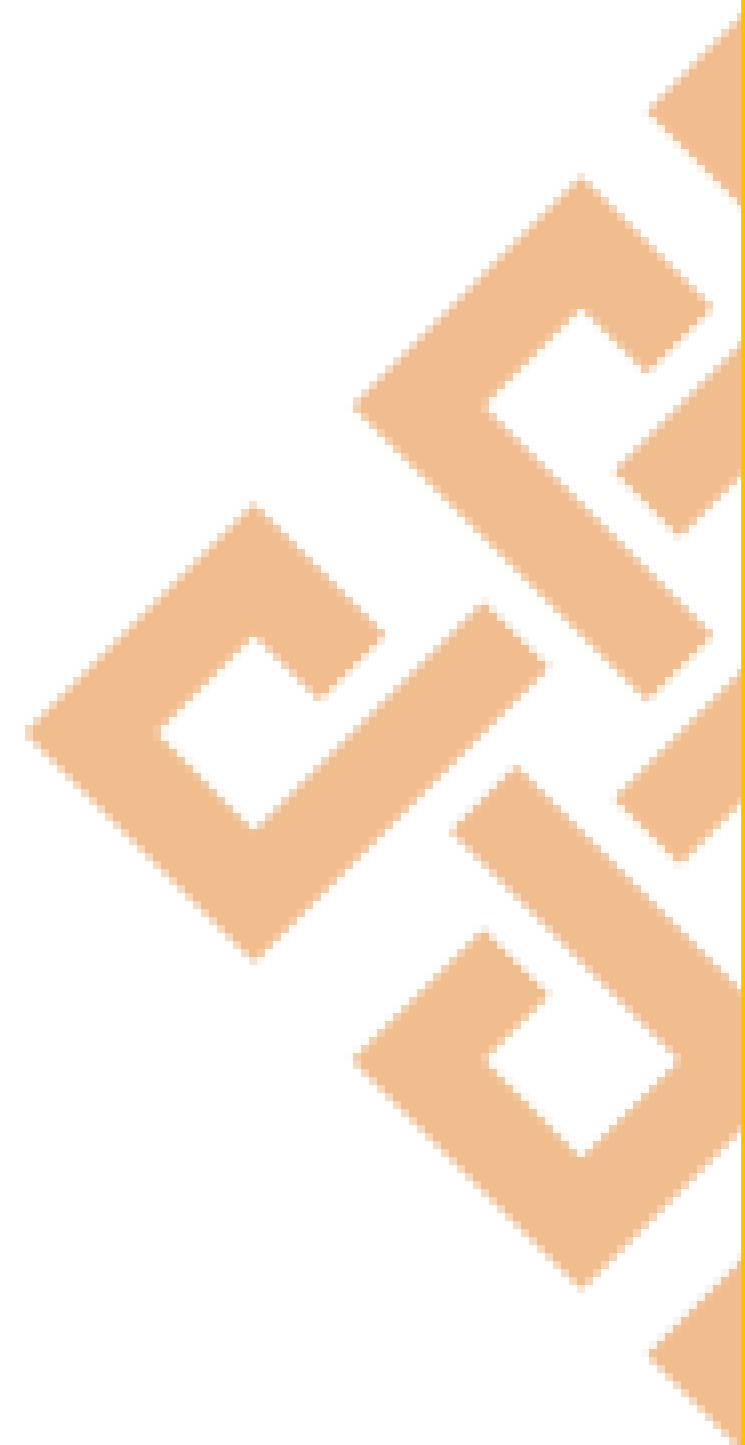
# Recap of previous lecture

## Hadoop Ecosystem



# **High Level Architecture of Hadoop**

**Topic of the Lecture**



# Topic of the Lecture

Hadoop Architecture/Master slave Architecture

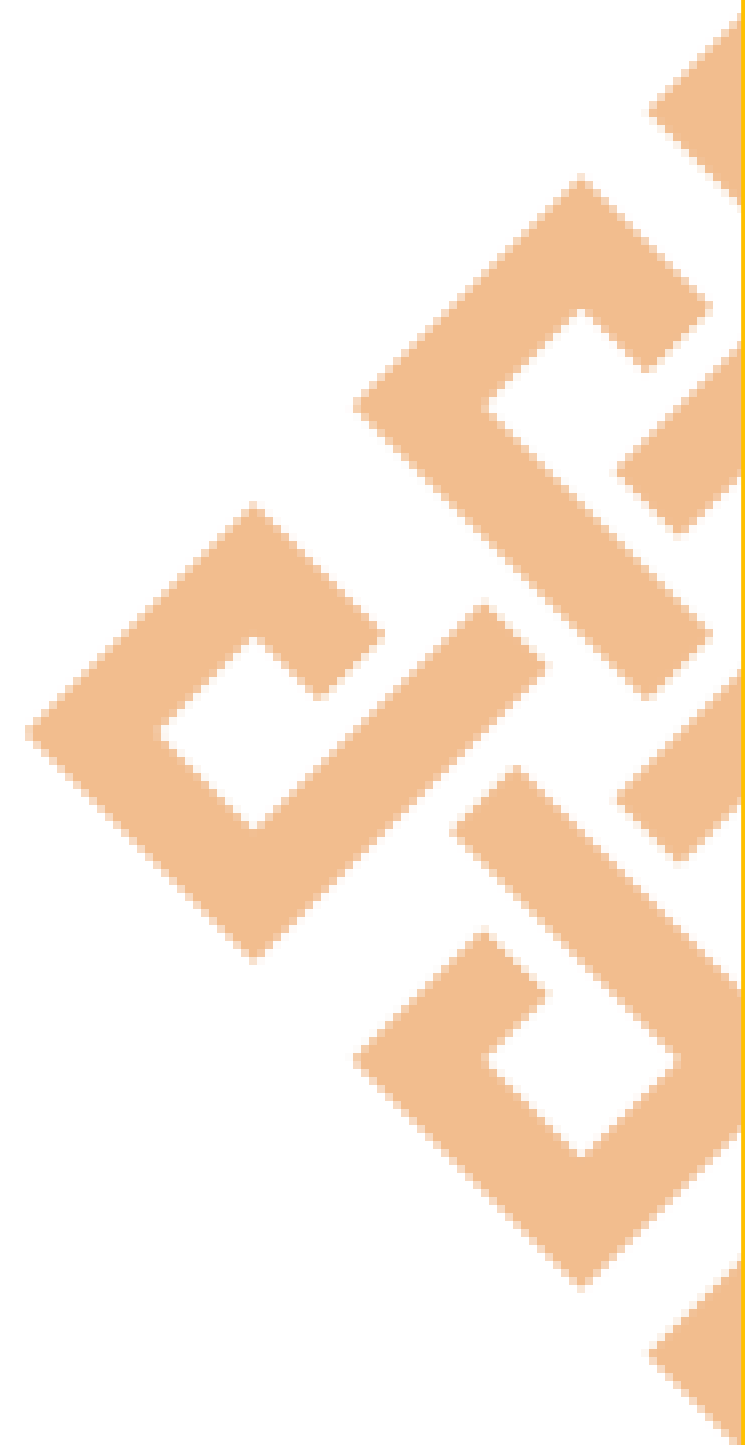
Introduction to Hadoop Distributed File System

Name node and Data node, MapReduce



# Hadoop Architecture

Objective and Outcome of Lecture



# Objective and Outcome of Lecture

## Lecture Objective

**Explain the Hadoop High Level Architecture**

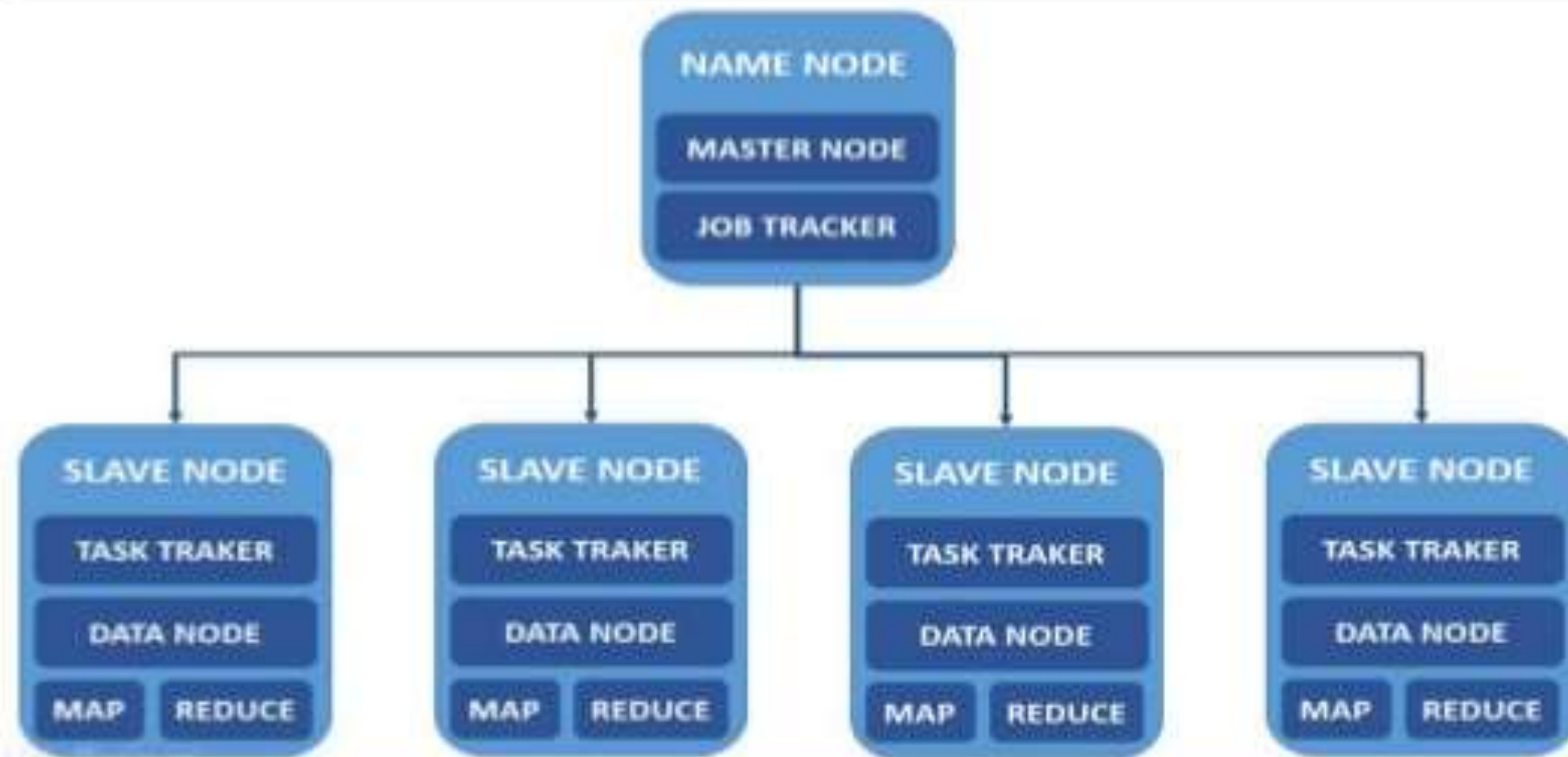
## Lecture Outcome

**Outline the Hadoop High Level Architecture.**

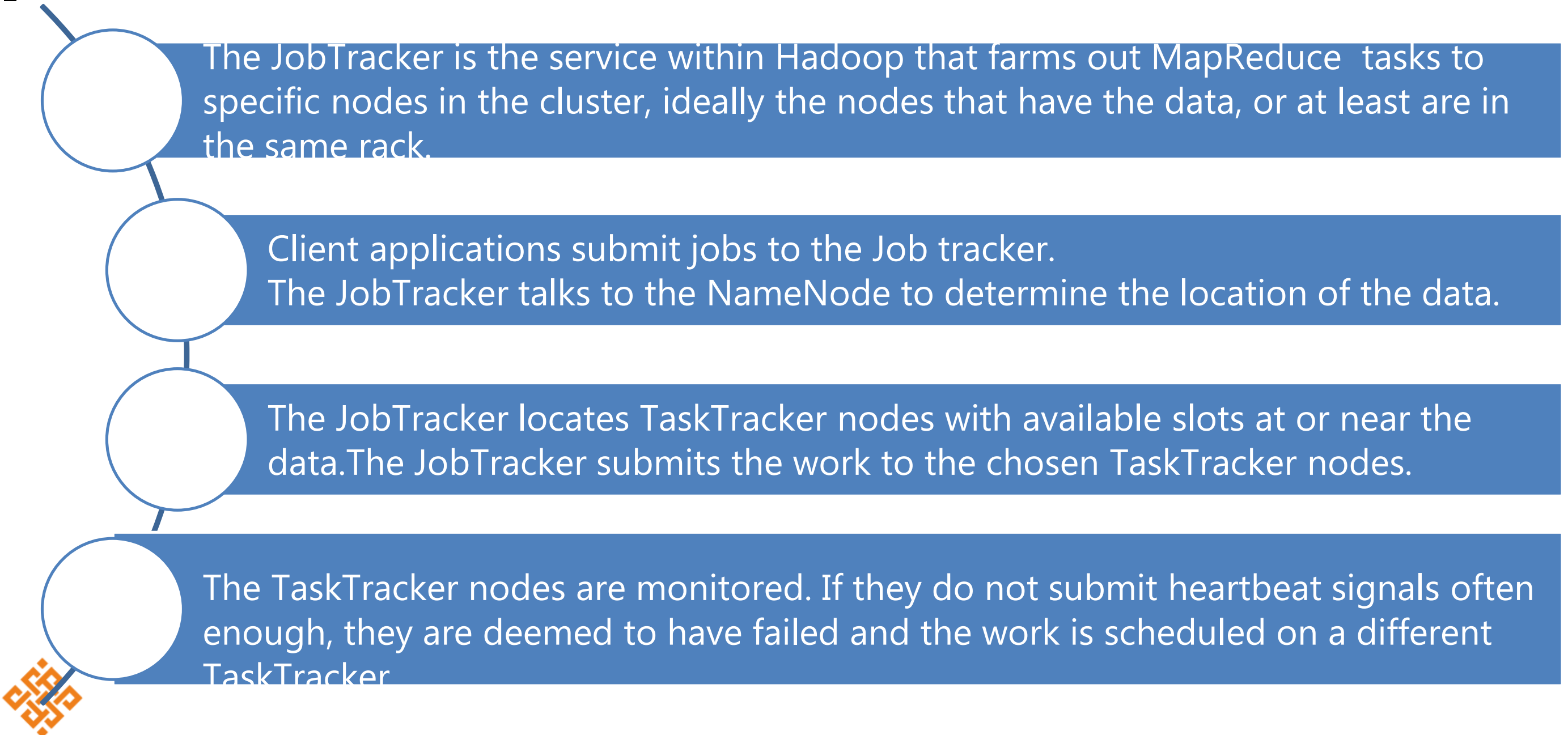


# HADOOP MASTER/SLAVE ARCHITECTURE

---



# Job Tracker and Task Trackers



The JobTracker is the service within Hadoop that farms out MapReduce tasks to specific nodes in the cluster, ideally the nodes that have the data, or at least are in the same rack.

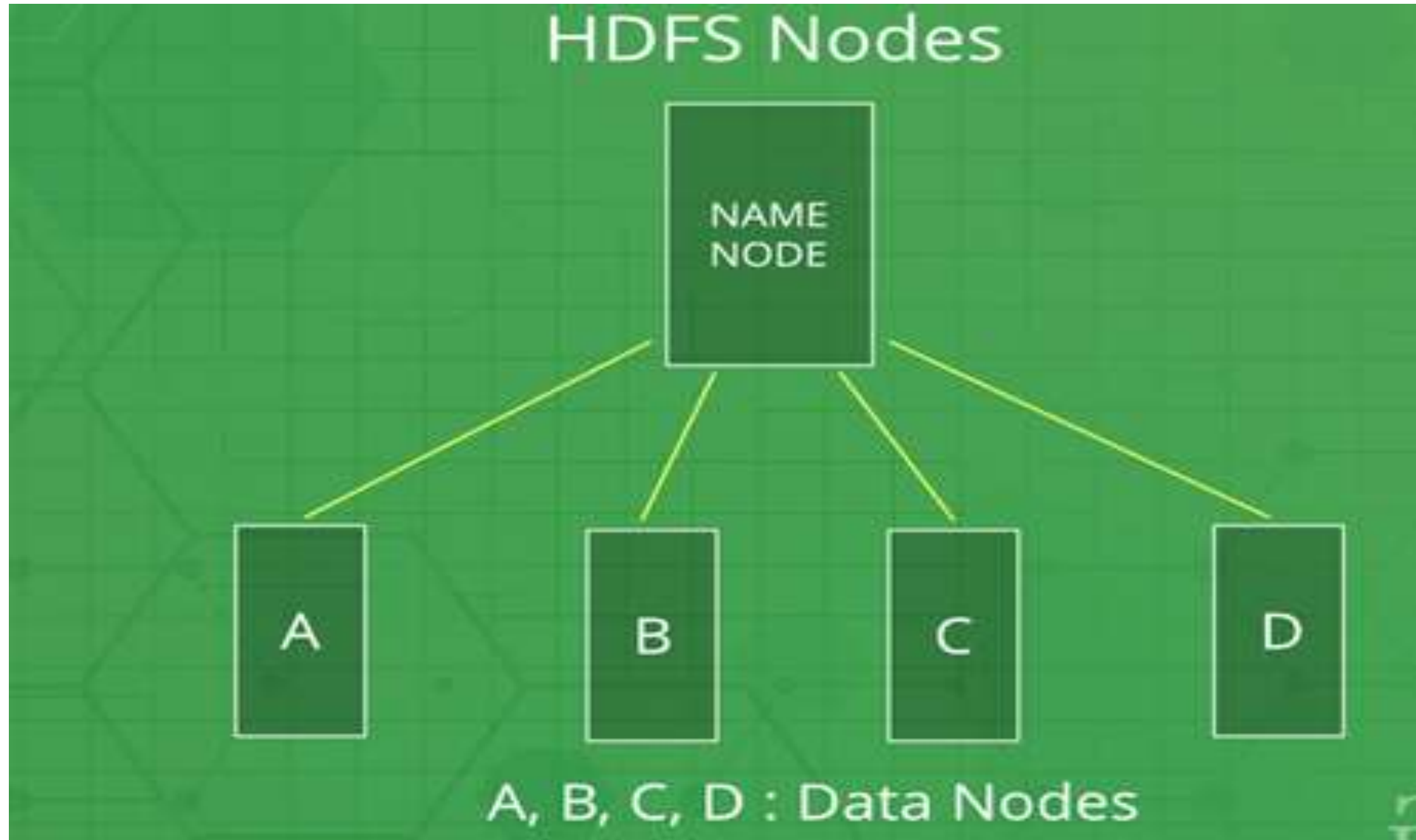
Client applications submit jobs to the Job tracker.  
The JobTracker talks to the NameNode to determine the location of the data.

The JobTracker locates TaskTracker nodes with available slots at or near the data. The JobTracker submits the work to the chosen TaskTracker nodes.

The TaskTracker nodes are monitored. If they do not submit heartbeat signals often enough, they are deemed to have failed and the work is scheduled on a different TaskTracker



# | Contd.....



**Thank You**

