



# **SNS COLLEGE OF ENGINEERING**



**Kurumbapalayam(Po), Coimbatore – 641 107**

**Accredited by NAAC-UGC with 'A' Grade**

**Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai**

## **Department of Artificial Intelligence and Data Science**

**Course Name – Big Data Analytics  
III Year / V Semester**

**Unit 3 – DATA ANALYTICAL FRAMEWORKS**

**Topic - MapReduce**





# MapReduce



## Working

- During a MapReduce job, Hadoop sends the Map and Reduce tasks to the appropriate servers in the cluster.
- The framework manages all the details of data-passing such as issuing tasks, verifying task completion, and copying data around the cluster between the nodes.
- Most of the computing takes place on nodes with data on local disks that reduces the network traffic.
- After completion of the given tasks, the cluster collects and reduces the data to form an appropriate result, and sends it back to the Hadoop server.
- Typically both the input and the output are stored in a file-system. The framework takes care of scheduling tasks, monitoring them and re-executes the failed tasks.



# MapReduce



## The MapReduce Framework

- The MapReduce framework consists of a single **master JobTracker** and one **slave TaskTracker per cluster-node**.

## Job Tracker

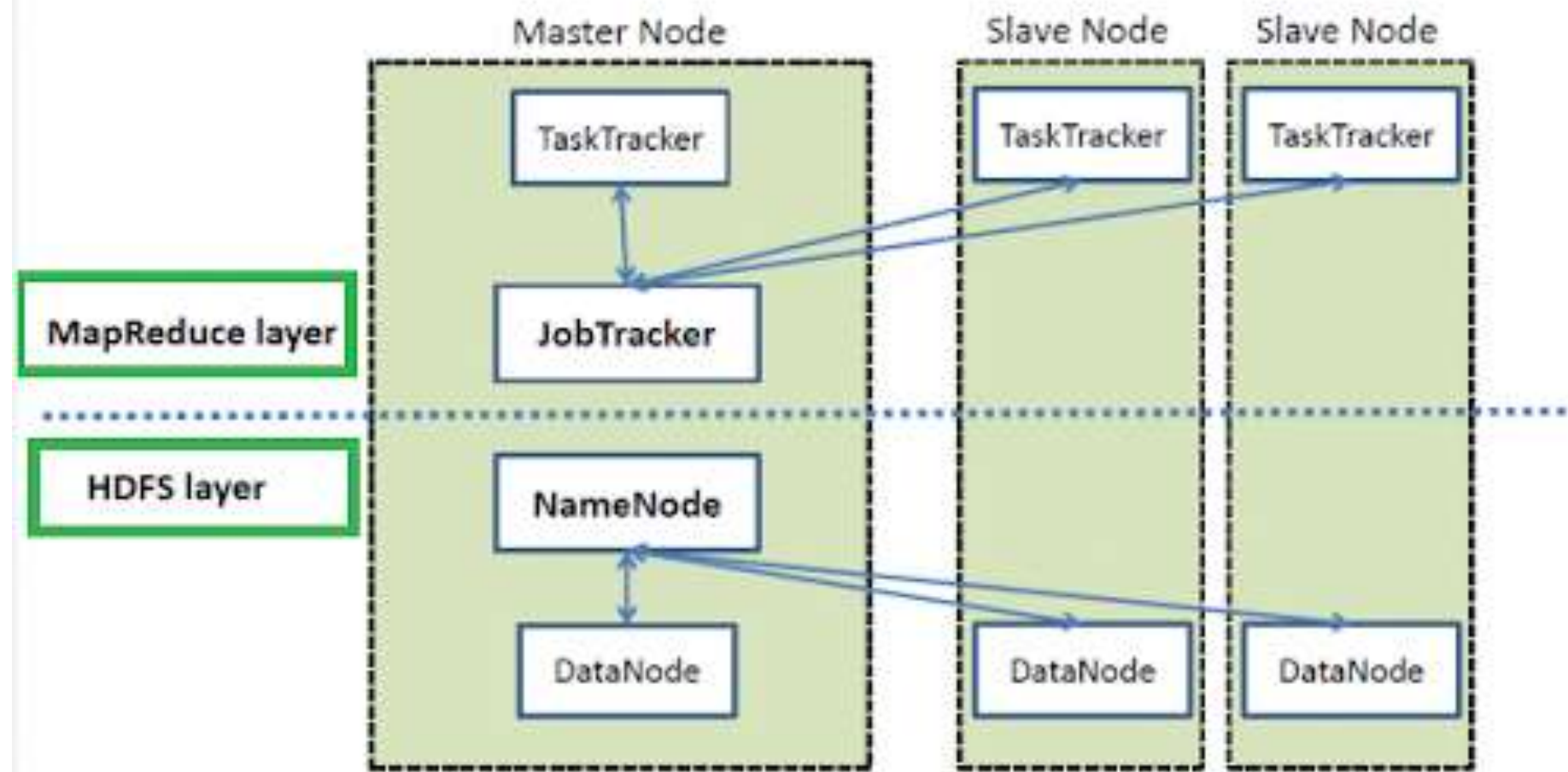
- The master Job Tracker is responsible for resource management, tracking resource consumption/availability and scheduling the jobs component tasks on the slaves, monitoring them and re-executing the failed tasks.

## Task Tracker

- The slaves TaskTracker execute the tasks as directed by the master and provide task-status information to the master periodically.
- The JobTracker is a single point of failure for the Hadoop MapReduce service which means if JobTracker goes down, all running jobs are halted.

# MapReduce

## High Level Architecture of Hadoop





**THANK YOU**