LinkedIn www.linkedin.com/in/naresh-neelam

Apache Spark Architecture

Apache Spark Unified API

Cluster Resource Manager

Node Manager	Node Manager	Node Manager	Node Manager
Operating System	Operating System	Operating System	Operating System
Hardware	Hardware	Hardware	Hardware

Cluster:

is group of connected computers that pools the resources of many machines and give us the ability to use all the resources as a single machine**.**

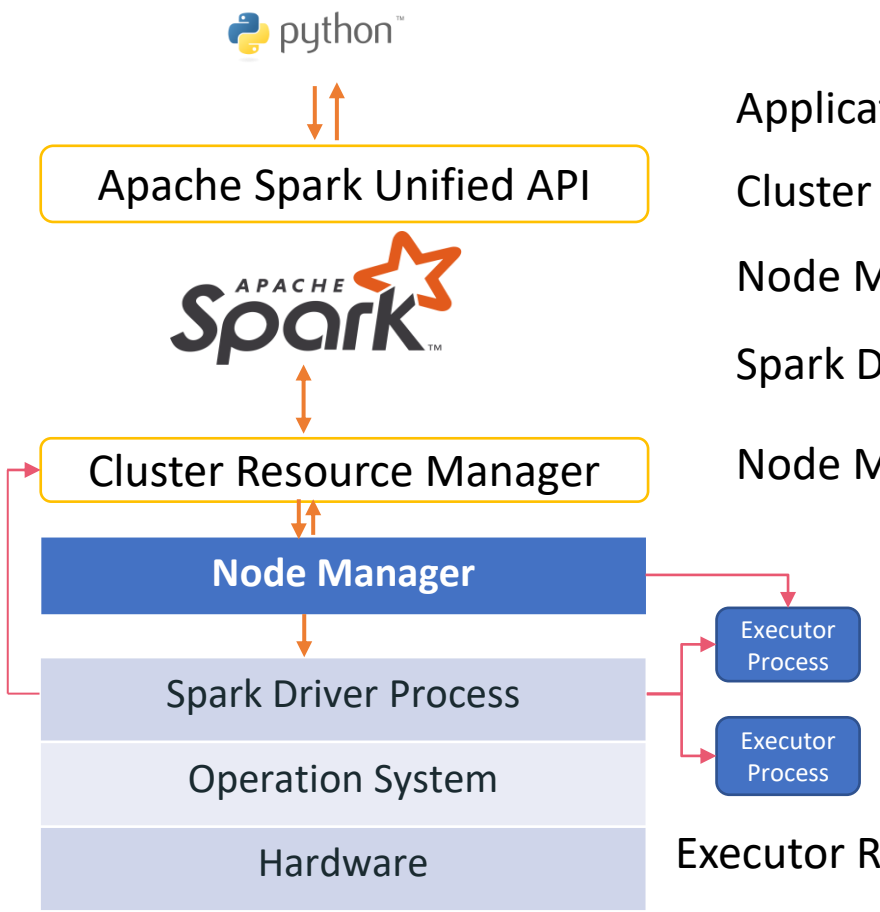
Cluster Resource Manager:

takes resource allocation and applications running on the cluster, it checks and monitor of the health of individual machines on the cluster

Node manager:

manages the resources and inform to the cluster manager about each machine state

Cluster Manager + Node Manager work as an operating system.



Application Submitted

Cluster Manager Requests Spark Driver to be Created

Node Manger Created Spark Driver

Spark Driver requests Spark Executors

Node Manager Creates Spark Executors

Spark Driver sends tasks to Executors

Executor Run the tasks and report status to Spark Driver