



JAYOTI VIDYAPEETH WOMEN'S UNIVERSITY, JAIPUR
Government of Rajasthan established
Through ACT No. 17 of 2008 as per UGC ACT 1956
NAAC Accredited University

Faculty of Education and methodology

Department of Computer Science and Engineering

Faculty Name- Jv'n Narendra Kumar Chahar (Assistant Professor)

Program- B.Tech 6thSemester

Course Name – Web Intelligence, HADOOP and Big Data Analysis

Session no.: 5

Session Name- HADOOP Operation Modes

Academic Day starts with –

- Greeting with saying '**Namaste**' by joining Hands together following by 2-3 Minutes Happy session, Celebrating birthday of any student of respective class and **National Anthem**.

Lecture starts with- quotations' answer writing

- Review of previous Session- **Introduction**

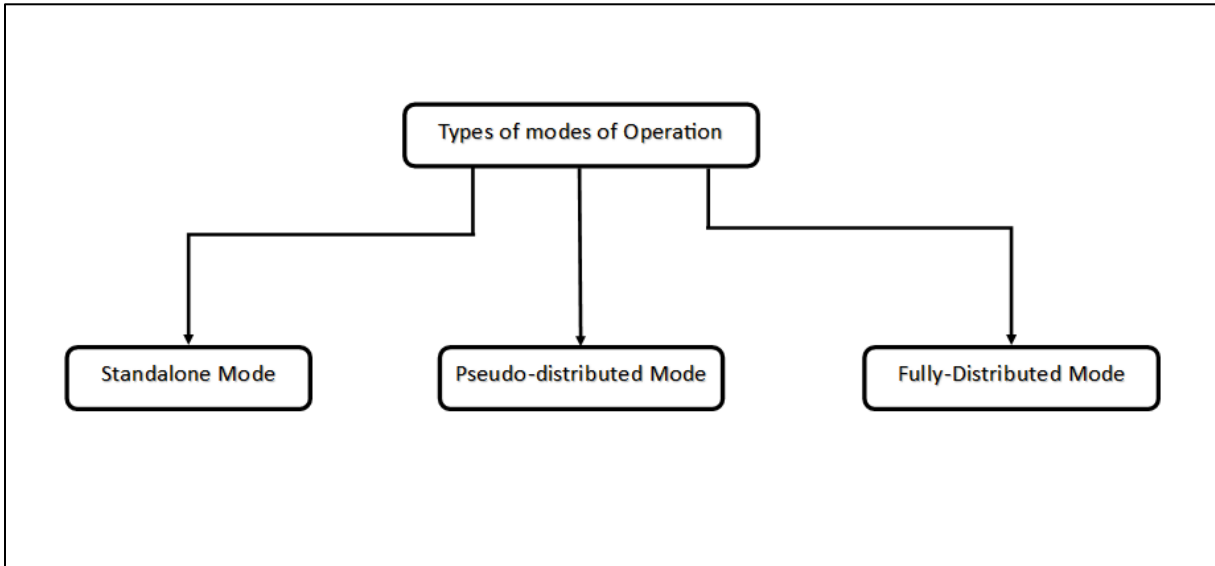
Topic to be discussed today- Today We will discuss about – **HADOOP Operation Modes**

- Lesson deliverance (ICT, Diagrams & Live Example)-
 - Diagrams

Introduction & Brief Discussion about the Topic – **An overview Hadoop Environment Setup**

HADOOP – Operation Modes

As we all know Hadoop is an open-source framework which is mainly used for storage purpose and maintaining and analyzing a large amount of data or datasets on the clusters of commodity hardware, which means it is actually a data management tool. Hadoop also possess a scale-out storage property, which means that we can scale up or scale down the number of nodes as per are a requirement in the future which is really a cool feature.



Hadoop Mainly works on 3 different Modes:

1. Standalone Mode
2. Pseudo-distributed Mode
3. Fully-Distributed Mode

1. Standalone Mode

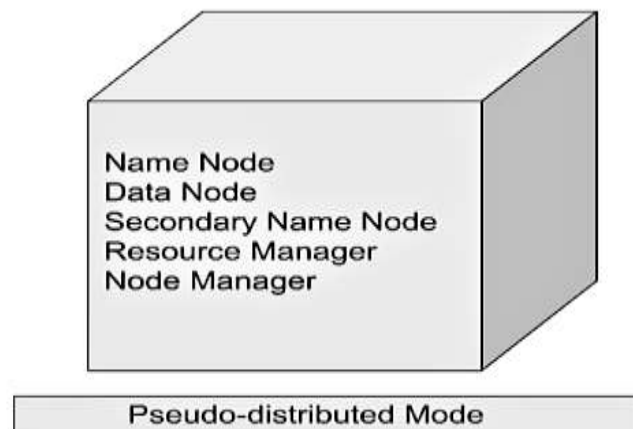
In Standalone Mode none of the Daemon will run i.e., Namenode, Datanode, Secondary Name node, Job Tracker, and Task Tracker. We use job-tracker and task-tracker for processing purposes in Hadoop1. For Hadoop2 we use Resource Manager and Node Manager. Standalone Mode also means that we are installing Hadoop only in a single system. By default, Hadoop is made to run in this Standalone Mode or we can also call it as the Local mode. We mainly use Hadoop in this Mode for the Purpose of Learning, testing, and debugging.

Hadoop works very much Fastest in this mode among all of these 3 modes. As we all know HDFS (Hadoop distributed file system) is one of the major components for Hadoop which utilized for storage Permission is not utilized in this mode. You can think of HDFS as similar to the file system's available for windows i.e. NTFS (New Technology File System) and FAT32(File Allocation Table which stores the data in the blocks of 32 bits). when your Hadoop works in this mode there is no need to configure the files – hdfs-site.xml, mapred-site.xml, core-site.xml for Hadoop environment. In this Mode, all of your Processes will run on a single JVM(Java Virtual Machine) and this mode can only be used for small development purposes.

2. Pseudo Distributed Mode (Single Node Cluster)

In Pseudo-distributed Mode we also use only a single node, but the main thing is that the cluster is simulated, which means that all the processes inside the cluster will run independently to each other. All the daemons that are Namenode, Datanode, Secondary Name node, Resource Manager, Node Manager, etc. will be running as a separate process on separate JVM(Java Virtual Machine) or we can say run on different java processes that is why it is called a Pseudo-distributed.

One thing we should remember that as we are using only the single node set up so all the Master and Slave processes are handled by the single system. Namenode and Resource Manager are used as Master and Datanode and Node Manager is used as a slave.

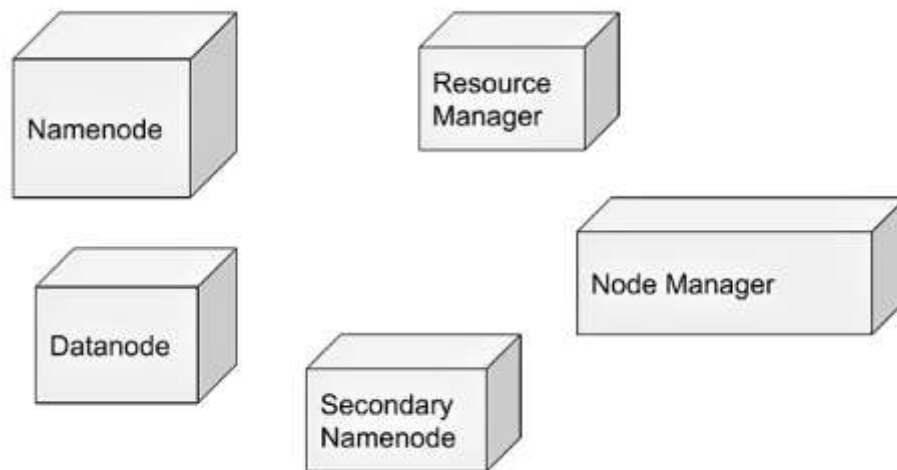


A secondary name node is also used as a Master. The purpose of the Secondary Name node is to just keep the hourly based backup of the Name node. In this Mode,

- Hadoop is used for development and for debugging purposes both.
- Our HDFS(Hadoop Distributed File System) is utilized for managing the Input and Output processes.
- We need to change the configuration files mapred-site.xml, core-site.xml, hdfs-site.xml for setting up the environment.

3. Fully Distributed Mode (Multi-Node Cluster)

This is the most important one in which multiple nodes are used few of them run the Master Daemon's that are Namenode and Resource Manager and the rest of them run the Slave Daemon's that are DataNode and Node Manager. Here Hadoop will run on the clusters of Machine or nodes. Here the data that is used is distributed across different nodes. This is actually the Production Mode of Hadoop let's clarify or understand this Mode in a better way in Physical Terminology.



Once you download the Hadoop in a tar file format or zip file format then you install it in your system and you run all the processes in a single system but here in the fully distributed mode, we are extracting this tar or zip file to each of the nodes in the Hadoop cluster and then we are using a particular node for a particular process. Once you distribute the process among the nodes then you'll define which nodes are working as a master or which one of them is working as a slave.

References-

1. **Book:** Hadoop: The Definitive Guide by Tom White, 3rd Edition, O'reilly Hadoop in Action by Chuck Lam, MANNING Publications
2. **Online:** <https://www.geeksforgeeks.org/>
3. **Online:** <http://www.oracle.com>

QUESTIONS: -

Q1. What is pre-requisite of installing HADOOP?

Q2. What are some different modes in Hadoop?

Next, we will discuss about Installing Hadoop in Standalone Mode.

- Academic Day ends with-
National song 'Vande Mataram'