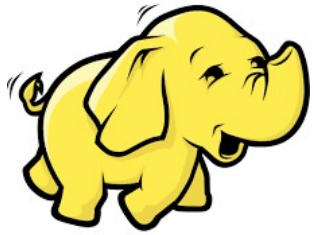
HDFS Daemons

Tushar B. Kute,
http://tusharkute.com



HDFS Daemons

- Daemons mean Process. Hadoop Daemons are a set of processes that run on Hadoop. Hadoop is a framework written in Java, so all these processes are Java Processes.
- Apache Hadoop 2 consists of the following Daemons:
  - NameNode
  - DataNode
  - Secondary Name Node
  - Resource Manager
  - Node Manager

HDFS Daemons

- Namenode, Secondary NameNode, and Resource Manager work on a Master System while the Node Manager and DataNode work on the Slave machine.

Namenode

- NameNode works on the Master System. The primary purpose of Namenode is to manage all the MetaData. Metadata is the list of files stored in HDFS(Hadoop Distributed File System).
- As we know the data is stored in the form of blocks in a Hadoop cluster. So the DataNode on which or the location at which that block of the file is stored is mentioned in MetaData.
- All information regarding the logs of the transactions happening in a Hadoop cluster (when or who read/wrote the data) will be stored in MetaData. MetaData is stored in the memory.

Namenode

- Features:
  - It never stores the data that is present in the file.
  - As Namenode works on the Master System, the Master system should have good processing power and more RAM than Slaves.
  - It stores the information of DataNode such as their Block id’s and Number of Blocks

Namenode

- How to start Name Node?

  hadoop-daemon.sh start namenode

- How to stop Name Node?

- hadoop-daemon.sh stop namenode

Datanode

- DataNode works on the Slave system. The NameNode always instructs DataNode for storing the Data.
- DataNode is a program that runs on the slave system that serves the read/write request from the client.
- As the data is stored in this DataNode, they should possess high memory to store more Data.

Datanode

- How to start Data Node?

  hadoop-daemon.sh start datanode

- How to stop Data Node?

  hadoop-daemon.sh stop datanode

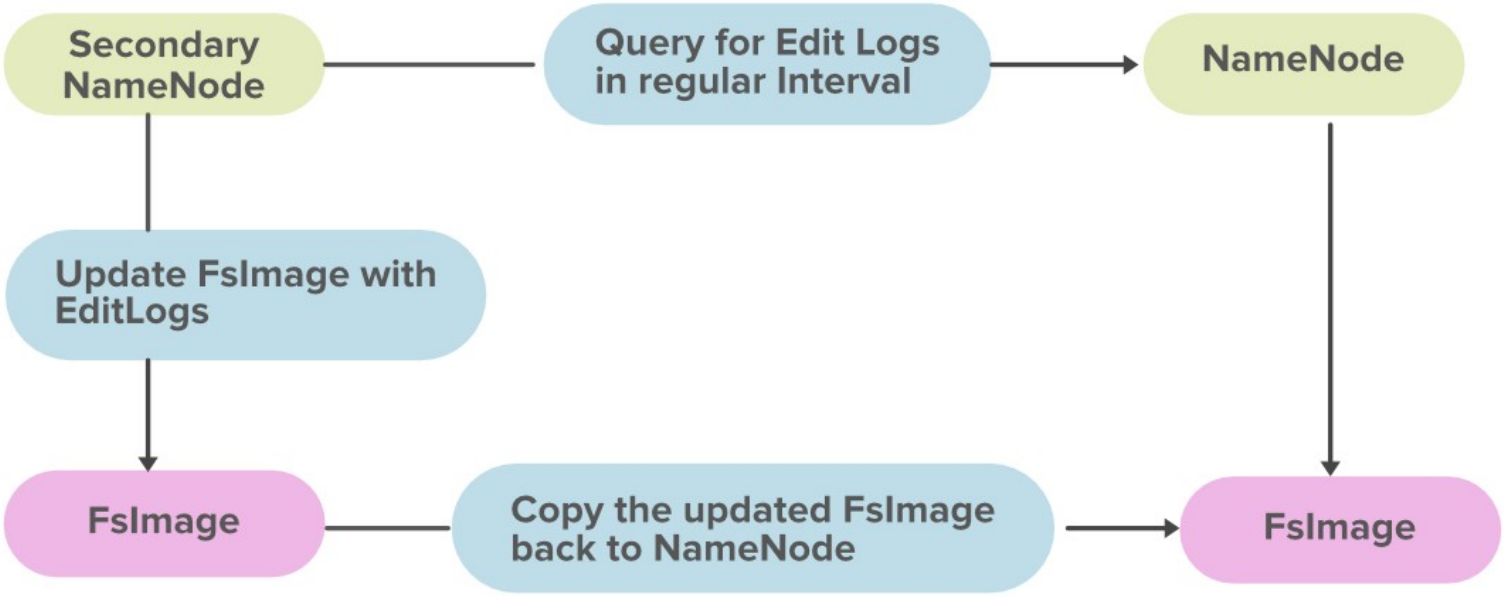
Secondary Namenode

- Secondary NameNode is used for taking the hourly backup of the data. In case the Hadoop cluster fails, or crashes, the secondary Namenode will take the hourly backup or checkpoints of that data and store this data into a file name fsimage.
- This file then gets transferred to a new system. A new MetaData is assigned to that new system and a new Master is created with this MetaData, and the cluster is made to run again correctly.
- Now in Hadoop2, we have High-Availability and Federation features that minimize the importance of this Secondary Name Node in Hadoop2.

Secondary Namenode

- Major Function Of Secondary NameNode:
  - It groups the Edit logs and Fsimage from NameNode together.
  - It continuously reads the MetaData from the RAM of NameNode and writes into the Hard Disk.
- As secondary NameNode keeps track of checkpoints in a Hadoop Distributed File System, it is also known as the checkpoint Node.

Secondary Namenode



Hadoop Daemon's Port

- These ports can be configured manually in hdfs-site.xml and mapred-site.xml files.

The Hadoop Daemon's	Port
Name Node	50070
Data Node	50075
Secondary Name Node	50090

Resource Manager

- Resource Manager is also known as the Global Master Daemon that works on the Master System.
- The Resource Manager Manages the resources for the applications that are running in a Hadoop Cluster.
- The Resource Manager Mainly consists of 2 things.
  - 1. ApplicationsManager
  - 2. Scheduler

ApplicationsManager

- An Application Manager is responsible for accepting the request for a client and also makes a memory resource on the Slaves in a Hadoop cluster to host the Application Master. The scheduler is utilized for providing resources for applications in a Hadoop cluster and for monitoring this application.
- How to start ResourceManager?

  yarn-daemon.sh start resourcemanager
- How to stop ResourceManager?

  yarn-daemon.sh stop resoucemnager

Node Manager

- The Node Manager works on the Slaves System that manages the memory resource within the Node and Memory Disk.
- Each Slave Node in a Hadoop cluster has a single NodeManager Daemon running in it. It also sends this monitoring information to the Resource Manager.
- How to start Node Manager?

  yarn-daemon.sh start nodemanager
- How to stop Node Manager?

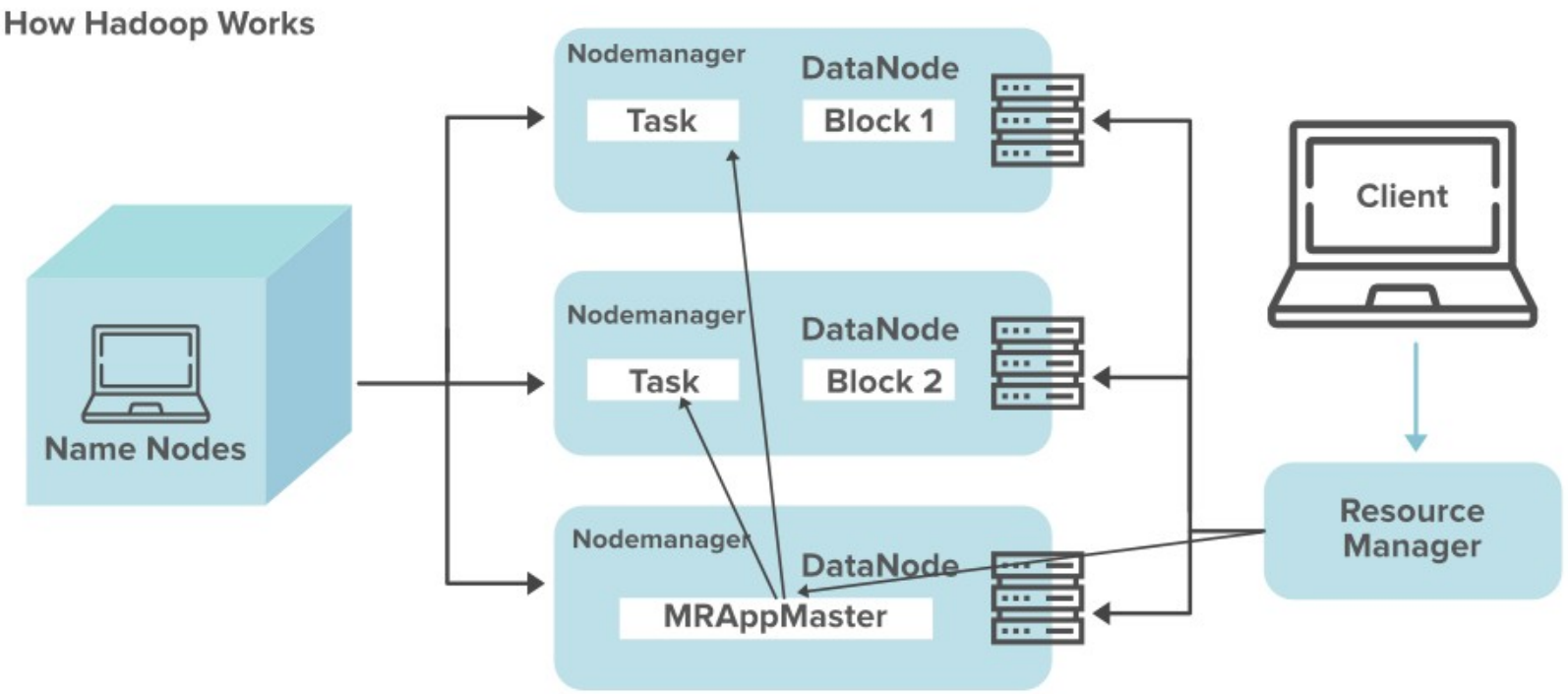
  yarn-daemon.sh stop nodemanager

Node Manager

- In a Hadoop cluster, Resource Manager and Node Manager can be tracked with the specific URLs, of type http://:port_number

The Hadoop Daemon's	Port
ResourceManager	8088
NodeManager	8042

Detailed Working



Thank you

This presentation is created using LibreOffice Impress 5.1.6.2, can be used freely as per GNU General Public License

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com