**DATA SCIENCE + CLOUD**

**\*\*Big Data – Hadoop :**

BigData – huge amount of data

-problems are 1. How to store 2. Where to store

## \*its hard disk that play role in data transfer not CPU or RAM

Problems are -

1. Store [speed , size]
2. Process
3. Business
4. **Store:**

Done by Hadoop – {Technology , framework}

**Hadoop :-**

* Open Source Apache – free and best for learning
* Cloudera – not good for learning
* HortonWorks - not free

**APACHE HADOOP :-**

->It has three versions – v1, v2, v3

\*hadoop 2.7 + redhat 7.5 -> best combination

->for storage HDFS is used –(Hadoop Distributed File System)

3gb

3gb

C:\Program Files (x86)\Microsoft Office\MEDIA\CAGCAT10\j0285750.wmf C:\Program Files (x86)\Microsoft Office\MEDIA\CAGCAT10\j0285750.wmf

C:\Program Files (x86)\Microsoft Office\MEDIA\CAGCAT10\j0285750.wmf C:\Program Files (x86)\Microsoft Office\MEDIA\CAGCAT10\j0285750.wmf

C:\Program Files (x86)\Microsoft Office\MEDIA\CAGCAT10\j0285750.wmf C:\Program Files (x86)\Microsoft Office\MEDIA\CAGCAT10\j0285750.wmf

 C:\Program Files (x86)\Microsoft Office\MEDIA\CAGCAT10\j0285750.wmf

100mb/s

200mb/s

user

Data node

Name Node

Name node – hdd(less) , RAM/CPU(high)

Data node – hdd(more) , RAM/CPU(less)

**PATH-SETUP FOR HADOOP AND JAVA JDK:**

JAVA\_HOME=/usr/java/jdk1.8.0

HADOOP\_HOME=/home/ec2-user/hadoop2

PATH=$JAVA\_HOME/bin:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin

~

|  |  |  |
| --- | --- | --- |
| AWSAccessKeyId=AKIAJRSS22I5KOBMNW6A |  |  |
| AWSSecretKey=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 | | |