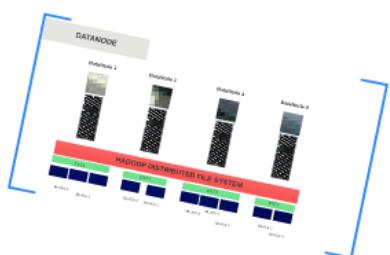




HDFS ARCHITECTURE



HDFS ARCHITECTURE

DATANODE

DataNode 1



DataNode 2



DataNode 3



DataNode 4

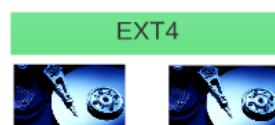


HADOOP DISTRIBUTED FILE SYSTEM



BLOCK 1

BLOCK 3



BLOCK 2

BLOCK 3



BLOCK 3

BLOCK 1

BLOCK 2



BLOCK 1

BLOCK 2

NAMENODE



aka Master

File Name - MyDatasetInHDFS

File Size - 350 MB

Replication Factor - 3

Blocks - BLK_0045732, BLK_9610590, BLK_8851209

Block Locations

Permissions

Created By, Created On

Lasted Modified By, Last Modified On

MyDatasetInHDFS

BLK_0045732	DN20	DN2	DN10
BLK_9610590	DN20	DN4	DN13
BLK_8851209	DN7	DN2	DN10

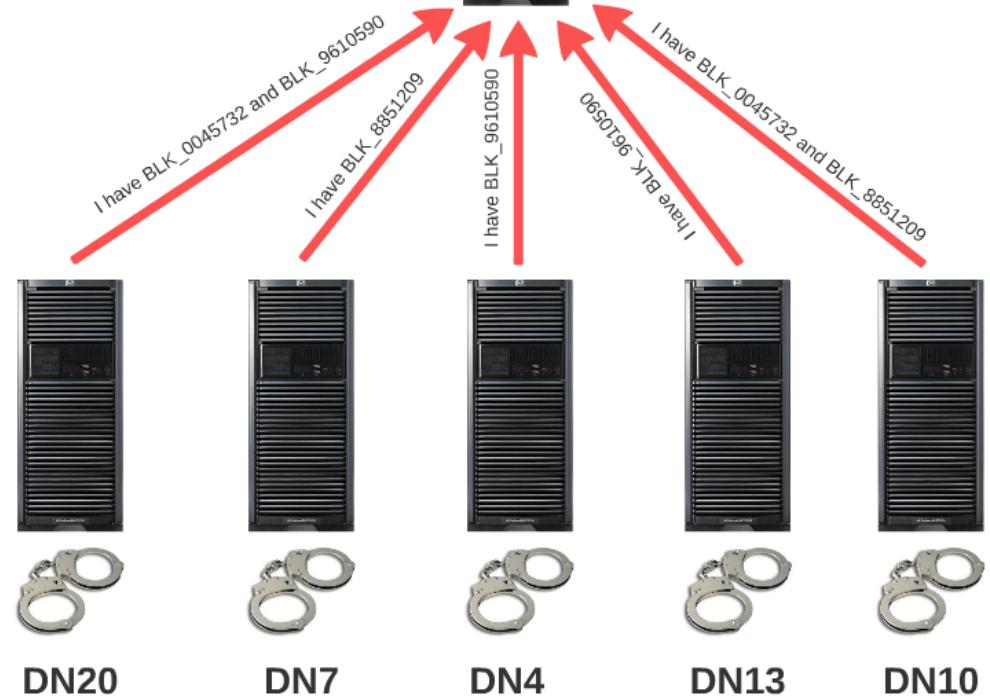


To All Data Nodes:
Send me block locations

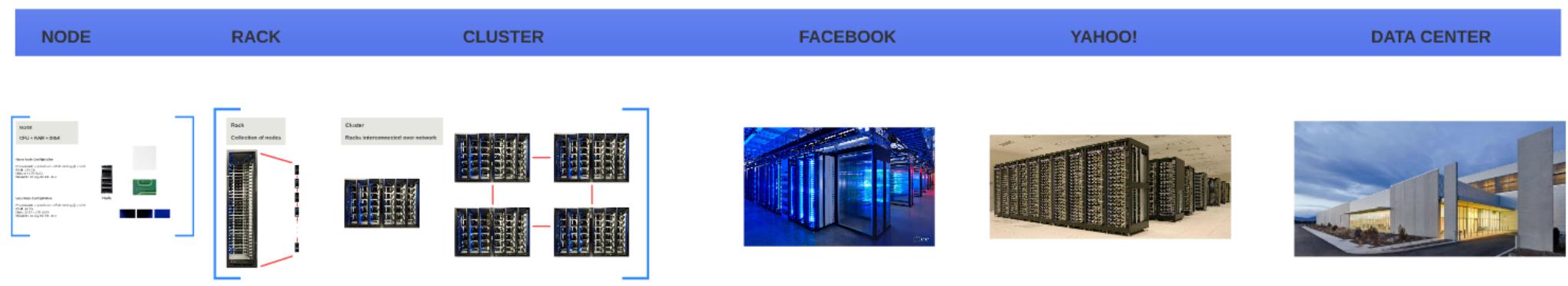
Name Node

In Disk - Metadata of Files & Folders

In Memory - Block locations



HADOOP CLUSTER



NODE

CPU + RAM + DISK

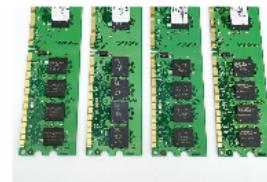
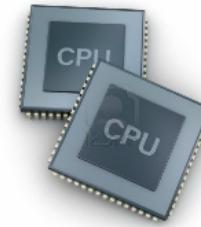
Name Node Configuration

Processors: 2 Quad Core CPUs running @ 2 GHz

RAM: 128 GB

Disk: 6 x 1TB SATA

Network: 10 Gigabit Ethernet



Node

Data Node Configuration

Processors: 2 Quad Core CPUs running @ 2 GHz

RAM: 64 GB

Disk: 12-24 x 1TB SATA

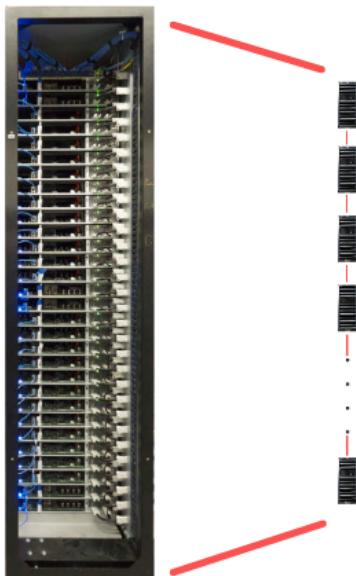
Network: 10 Gigabit Ethernet



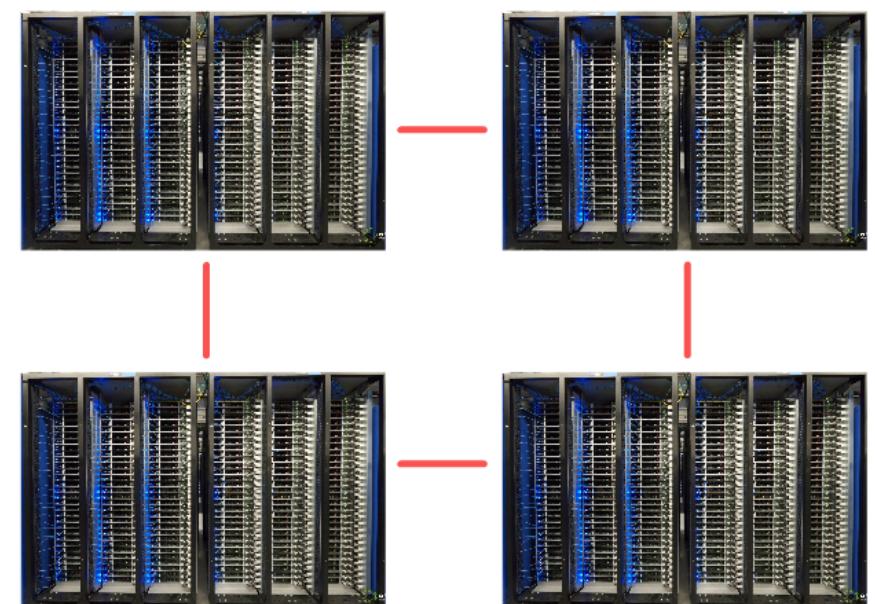
RACK

CLUSTER

Rack
Collection of nodes



Cluster
Racks interconnected over network





www.hadoopinrealworld.com







www.hadoopinrealworld.com

HADOOP CLUSTER

