Tech Mahindra

Hadoop 1.x Architecture

TLS

5 Daemons of Hadoop Architecture

Name Node	– Master Node
Secondary Name Node	– Master Node
Job Tracker	– Master Node
Data Node	– Slave Node
Task Tracker	– Slave Node

Tech Mahindra

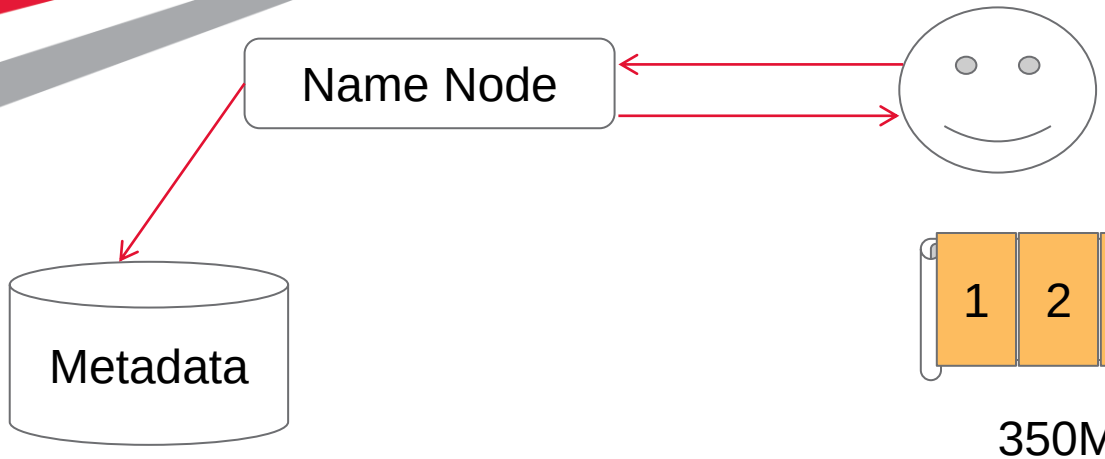
5 Daemons of Hadoop Architecture cont…

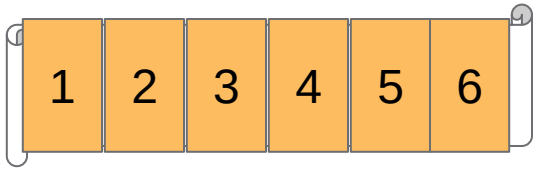
- ✓ Let us assume, a user wants to store 350MB data file in to cluster.
- ✓ In our cluster, every single node can accommodate 64MB block size of file split.
- ✓ There fore, user need to split the 350MB size file into 64MB file splits.
- ✓ After splitting user will get 6 splits, 5 - 64MB splits and 1 – 30MB split

Tech Mahindra

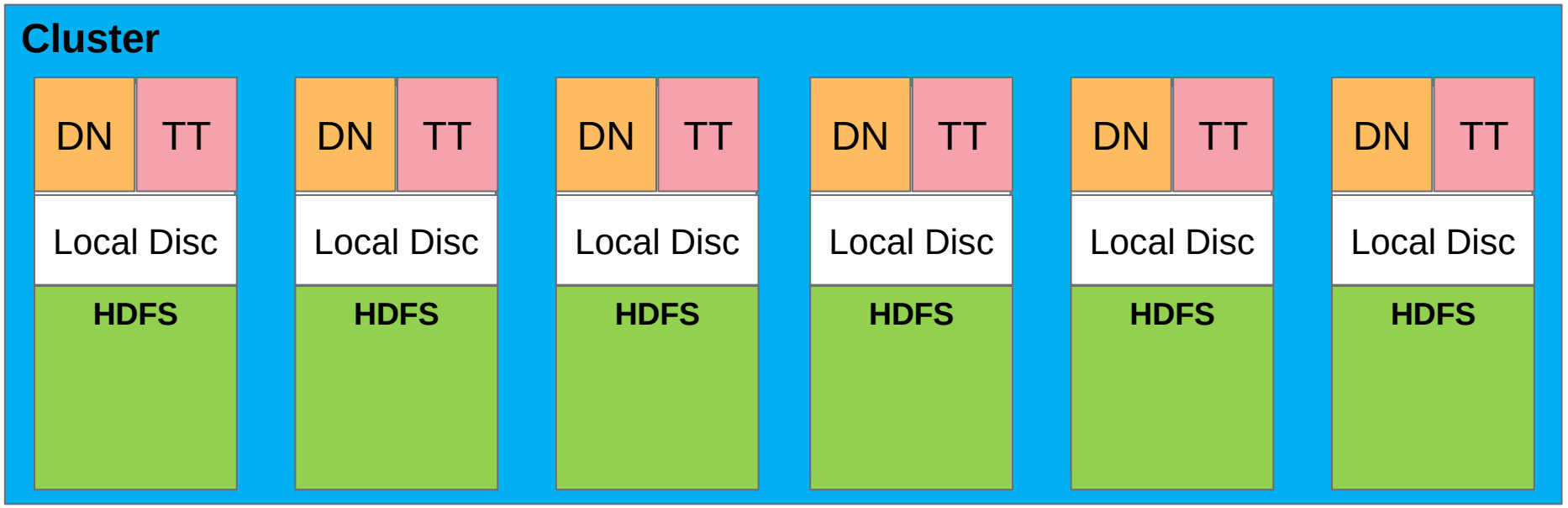
HDFS - Data Storage Process

TLS

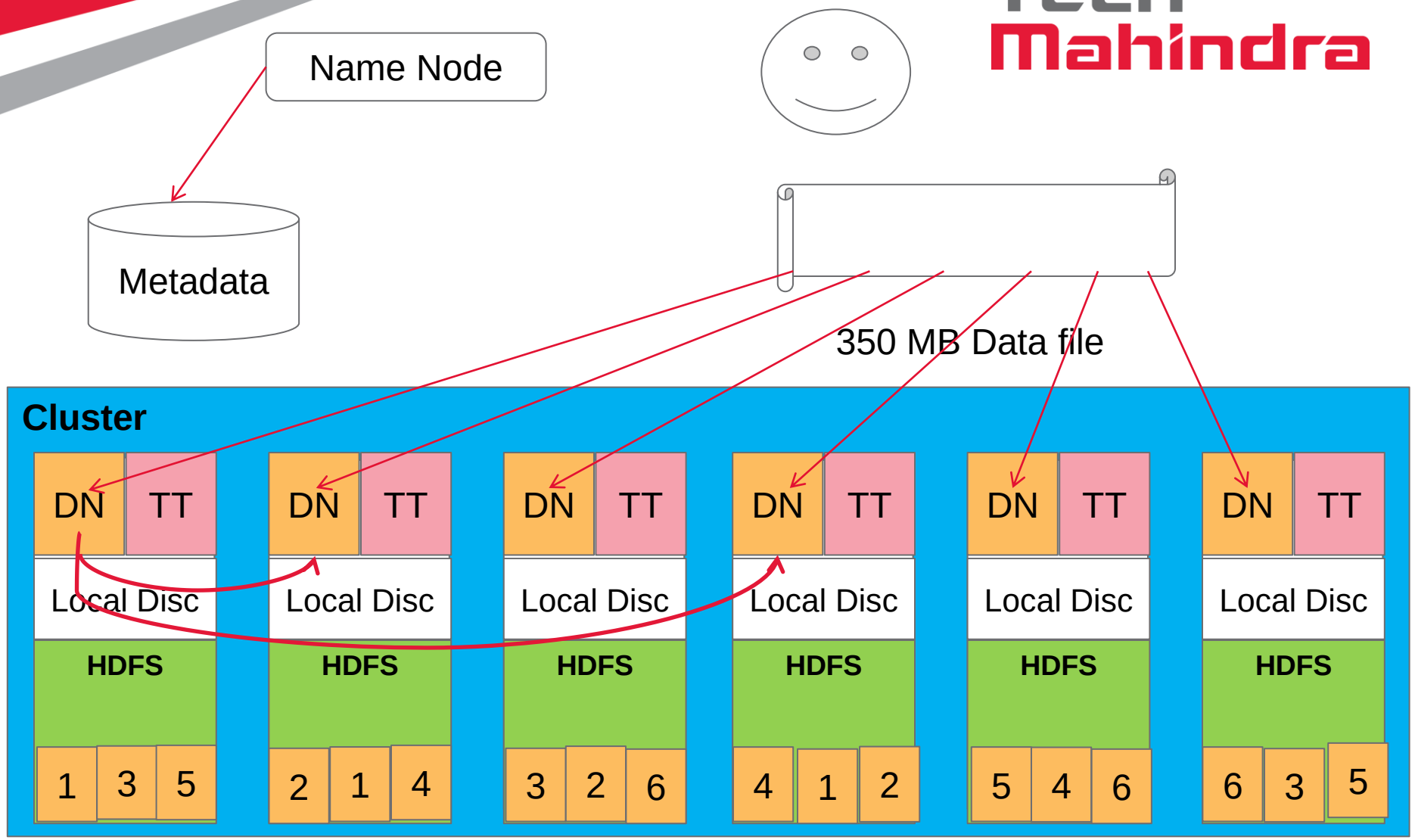




350MB Data file



Tech Mahindra



Data Storage on HDFS with Replication Factor=3

TLS

Tech Mahindra

For Data Storage on HDFS:

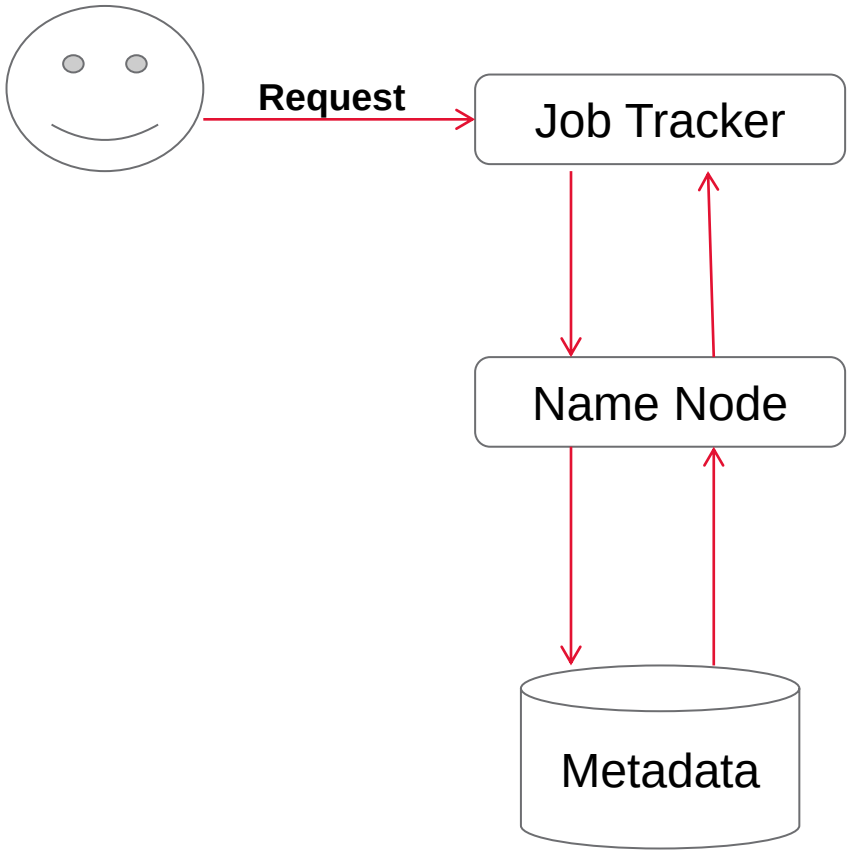
Client/User sends a request to Name Node.
Name Node will send the Data node (DN) references where Data blocks can be stored.
Client/User will split the file in to 64MB data blocks.
Client/User will send the blocks to Data nodes.
Data nodes will internally communicate with other Data nodes for replication

i.e.. Client -> Name Node -> Client -> Data nodes ->another Data node

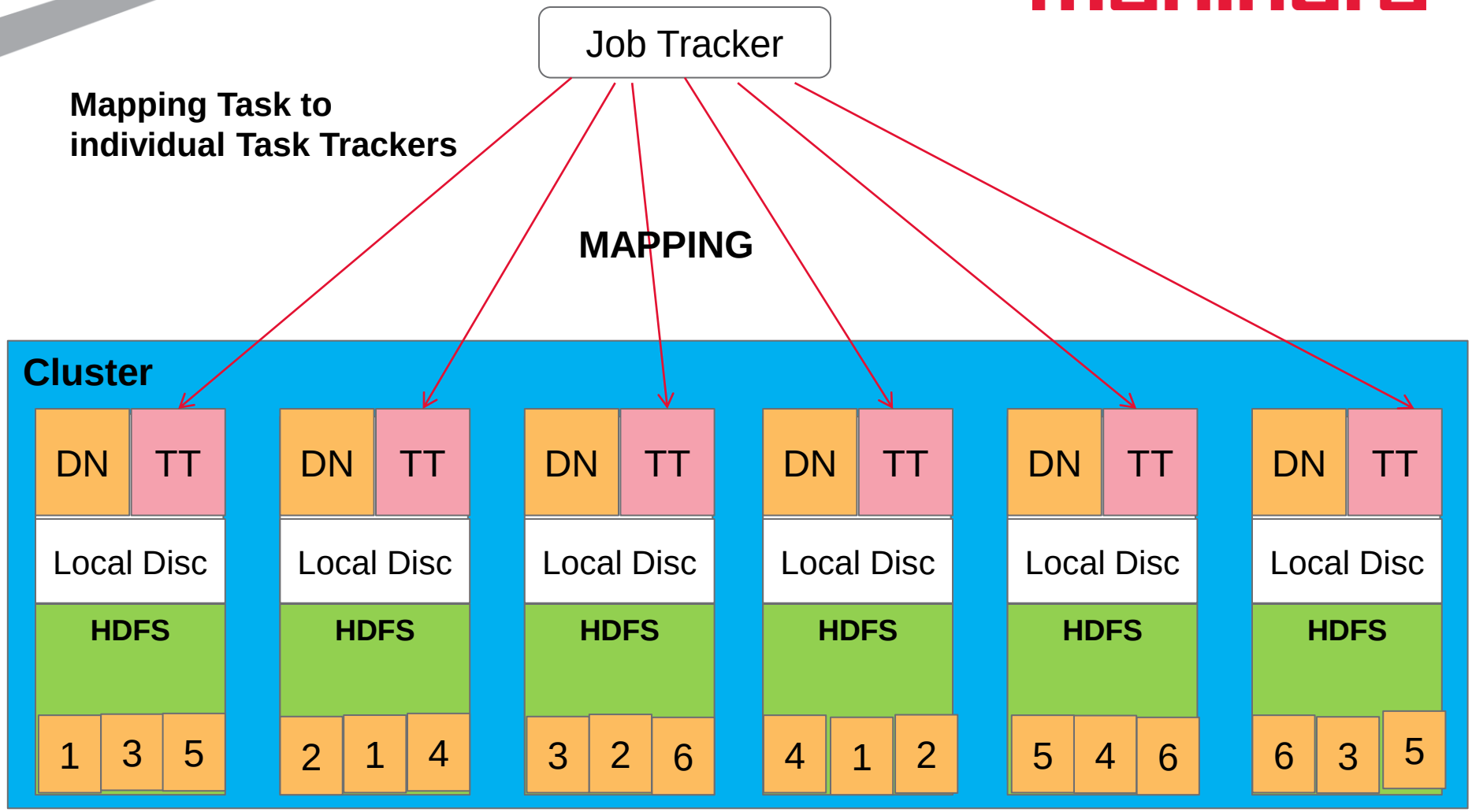
Tech Mahindra

MapReduce –Data Processing

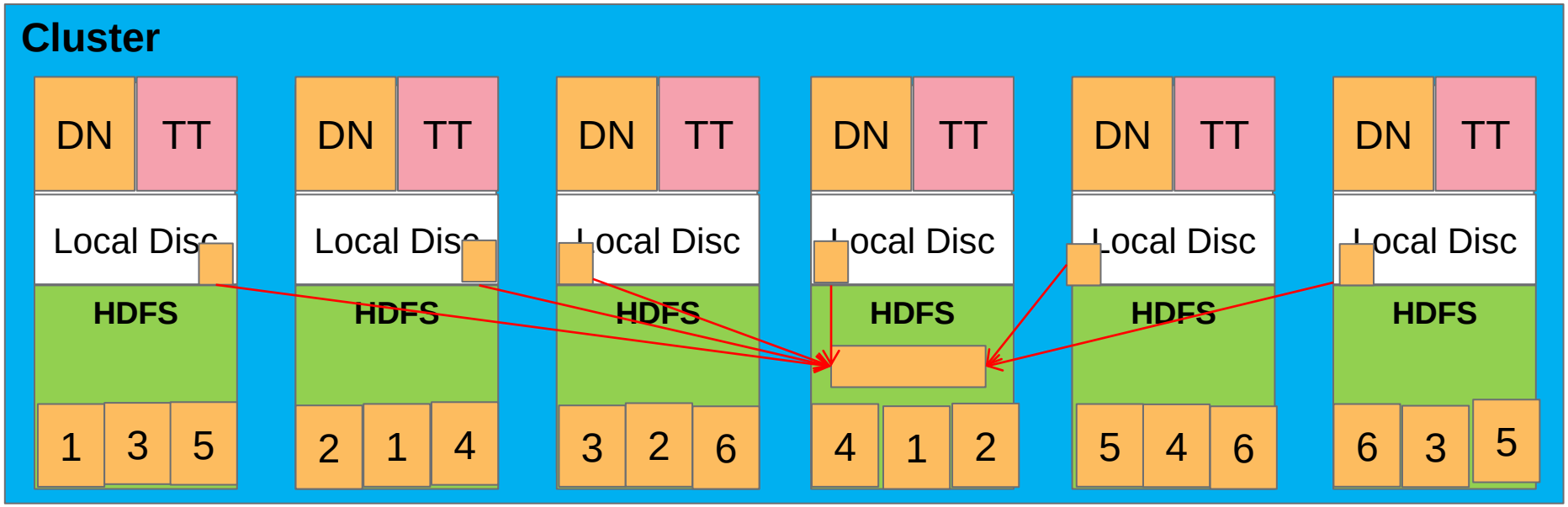
TLS



Data Processing on HDFS using MapReduce



Data Processing on HDFS using MapReduce



After processing individual splits, result will be stored as intermediate results.
These intermediate results will be merged and stored on HDFS itself.
This process is called as **REDUCTION** process.

Data Processing on HDFS using MapReduce

Tech Mahindra

For Data Processing Using MAP REDUCE

Client/User sends a request to Job Tracker
Job Tracker will check with Name node
Name node will give all the details of Date nodes, file storages, replication references, etc..
Job tracker will create job to perform
That Job will be divided in to tasks and assigned to individual Task trackers (TT), it is called as MAPPING/MAP
Task tracker will process the data and creates intermediate results which store in local disk
Finally all the intermediate results will be combined and will create a output file. This process is called REDUCTION/REDUCER.

TLS

