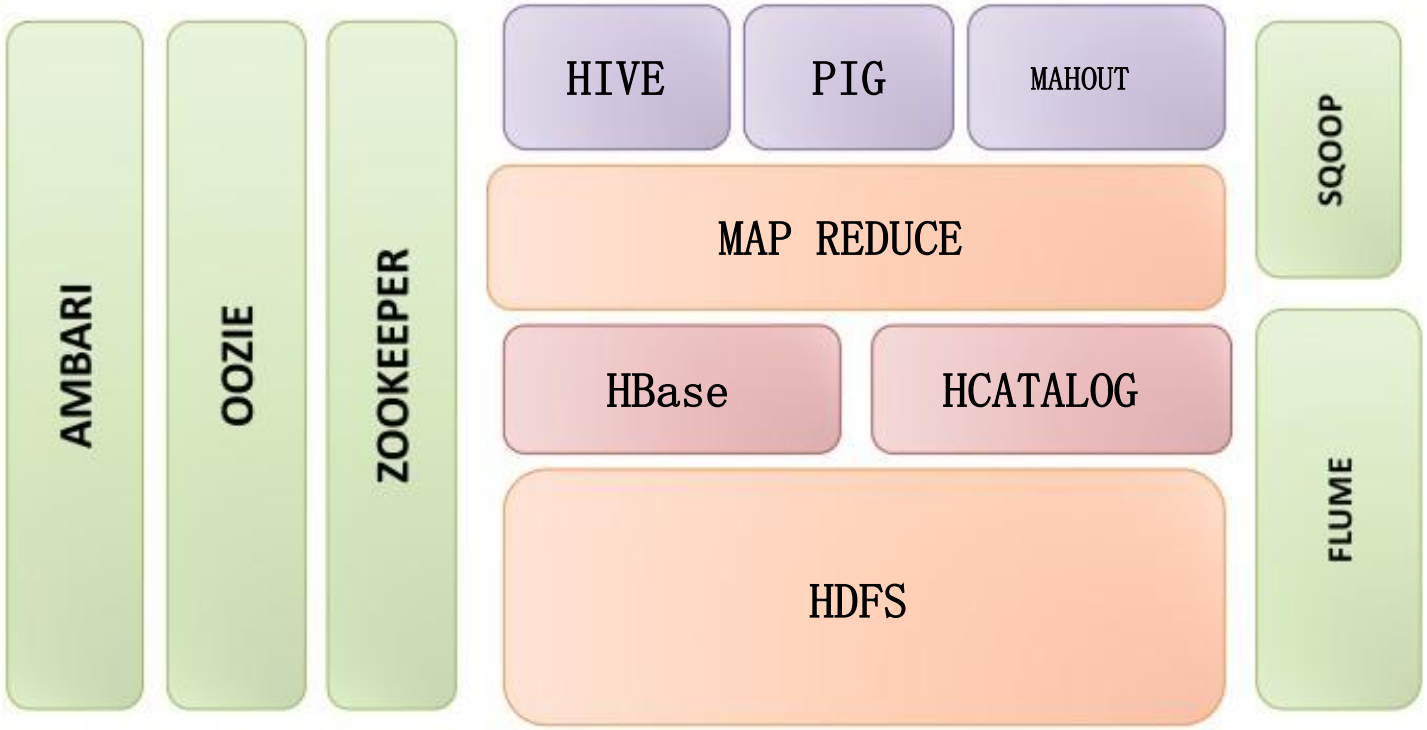
HADOOP 2.0

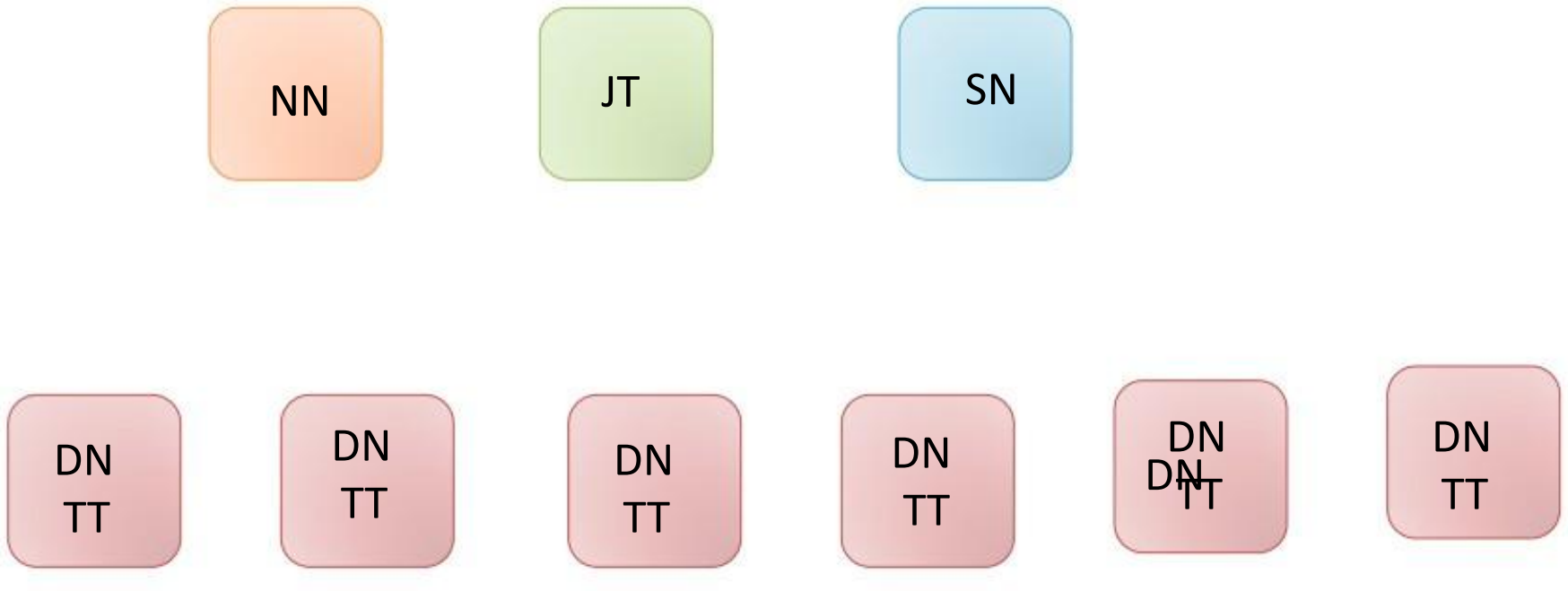
Who Am I ??

- Big Data Engineer in Cognizant High Performance Computing CoE
- Big Data Instructor & Technical Speaker for academic institutions

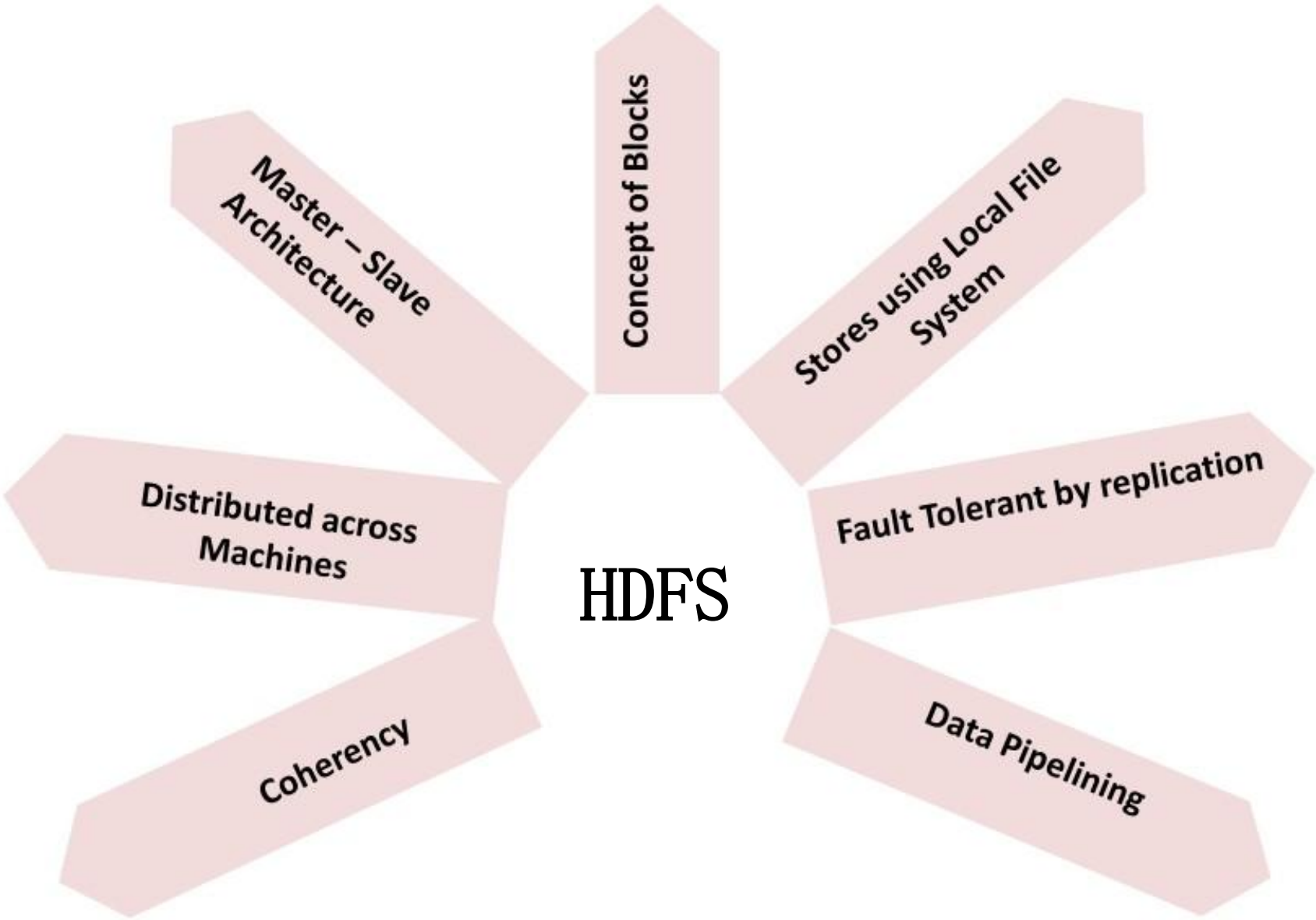
Hadoop 1.X EcoSystem



Hadoop 1.X Cluster

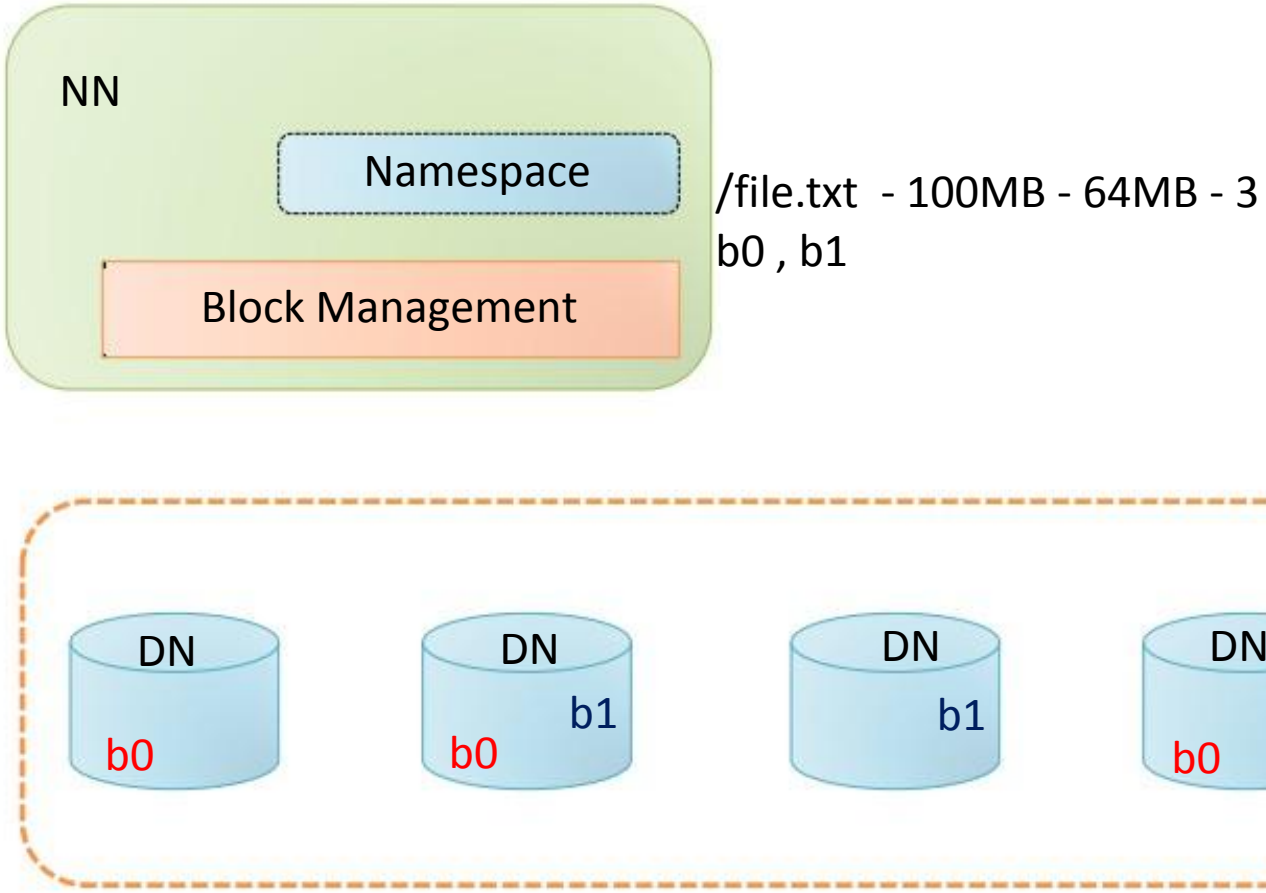


HDFS - Distributed Storage

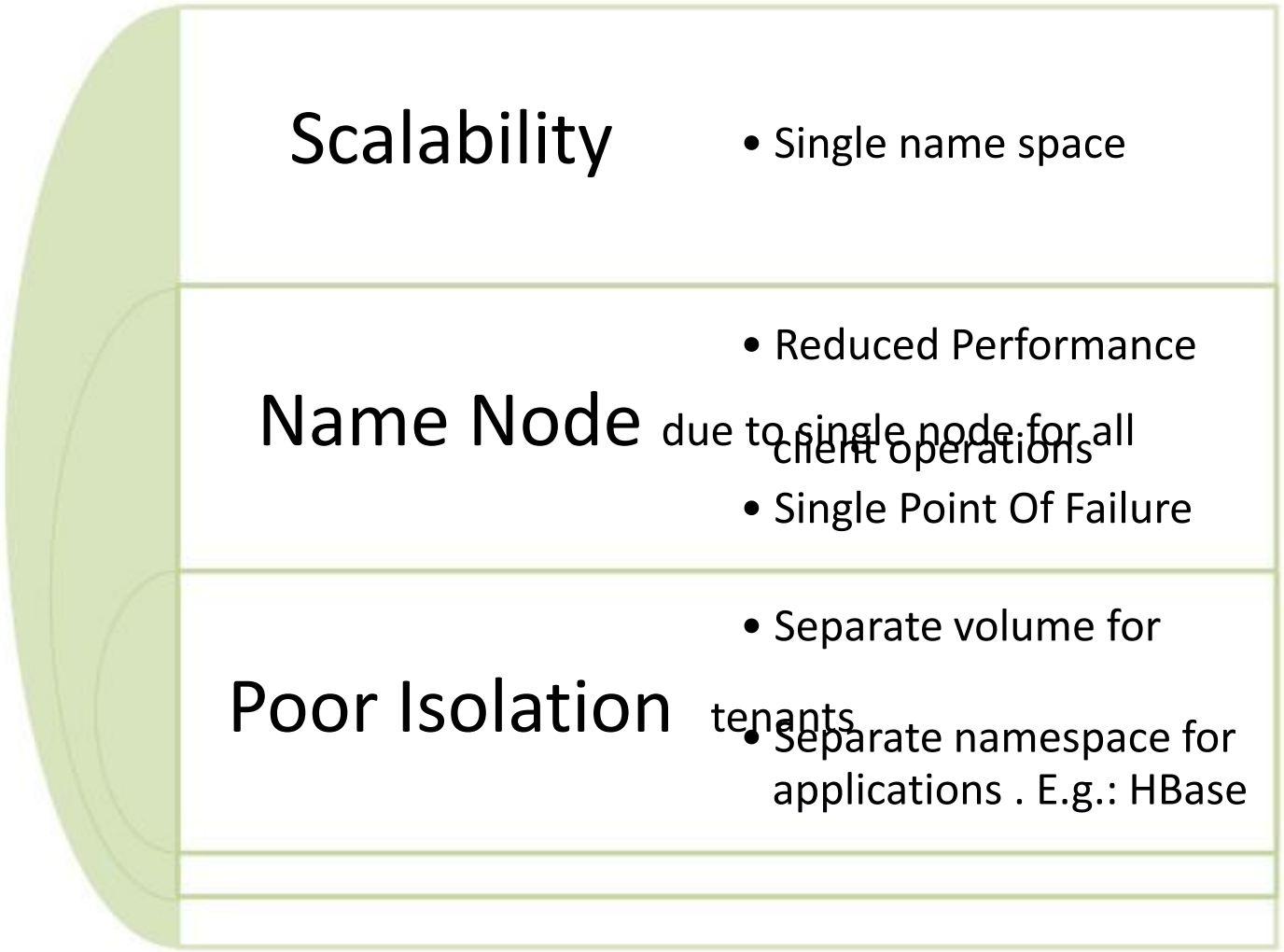


HDFS

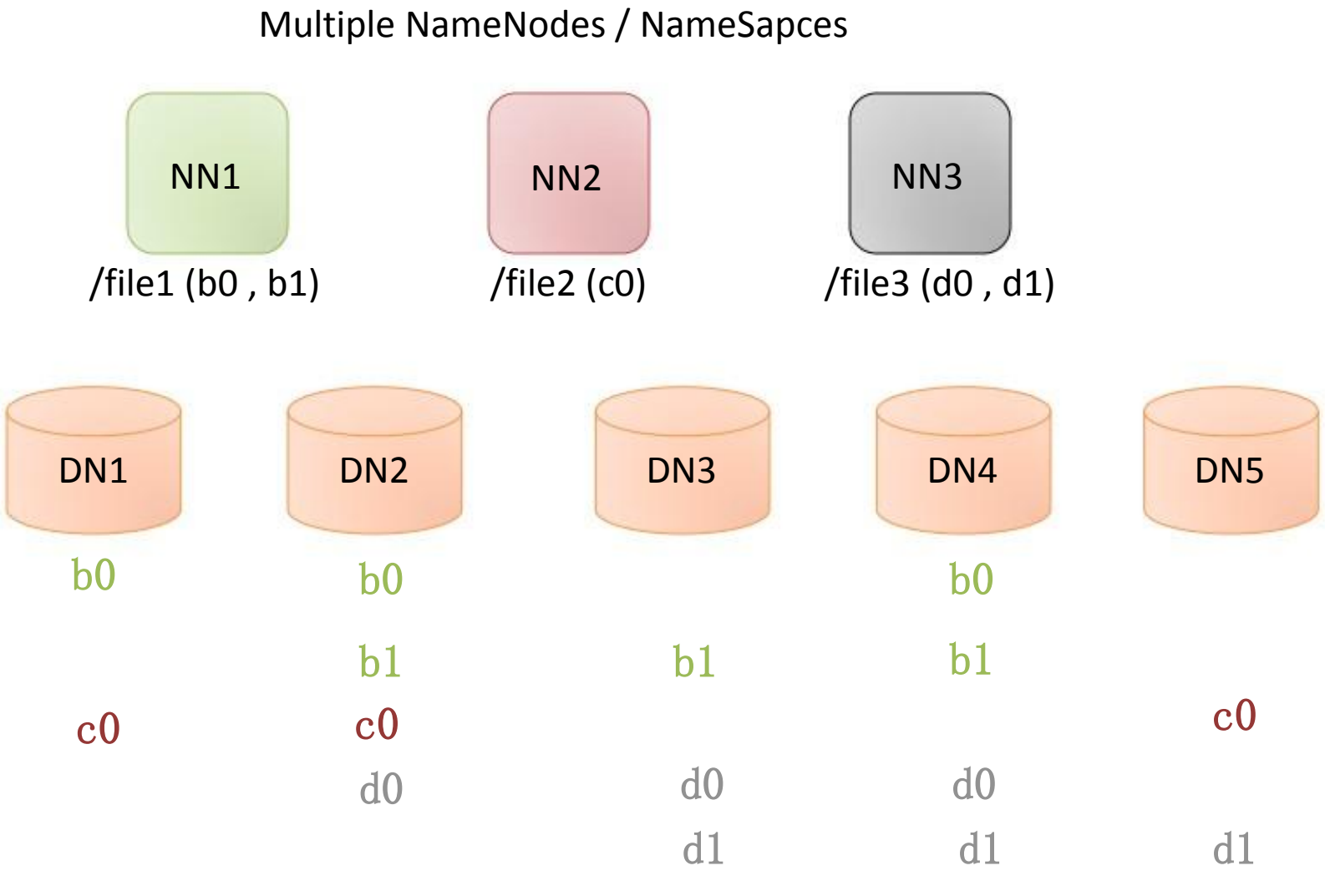
NameSpace Volume = NameSpace + BlockPool



HDFS - Limitations

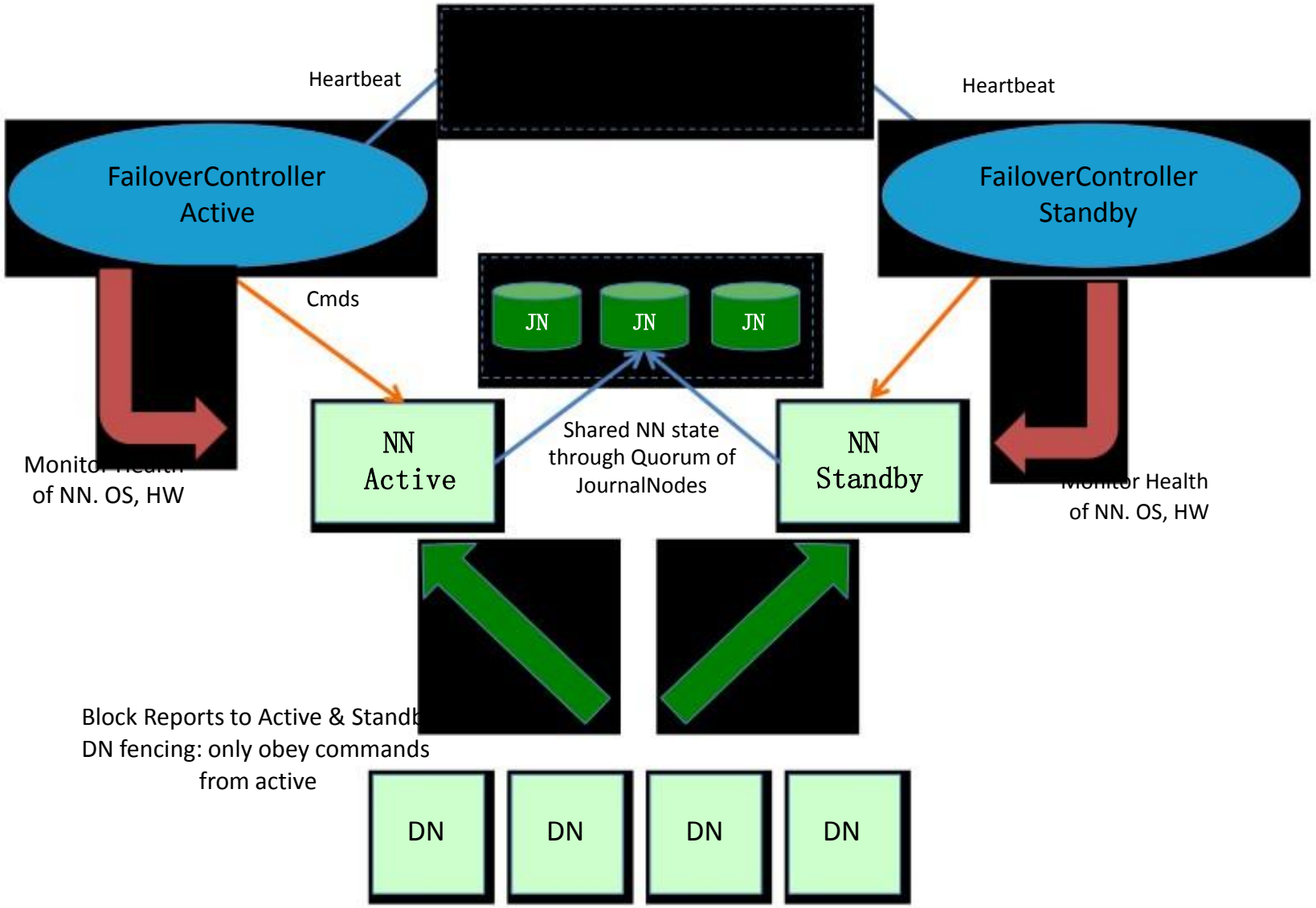


HDFS Federation



High Availability

- To avoid SPOF
- A pair of redundant NameNodes
  - One Active and One Hot Standby
  - Block Reports to both NNs from DNs
- Standby Node does checkpointing
  - Disappearance of Secondary NameNode
- Supports both Manual and Automatic Failover
- Backward Compatibile


Heartbeat
Heartbeat
FailoverController
Active
FailoverController
Standby
JN
JN
JN
Cmds
NN
Active
NN
Standby
Shared NN state
through Quorum of
JournalNodes
of NN. OS, HW
Monitor Health
of NN. OS, HW
Block Reports to Active & Standb
DN fencing: only obey commands
from active
DN
DN
DN
DN

Automatic FailOver

- Two Things
  - Failure Detection
  - Election of a New NN
- Two Components
  - ZooKeeper Quorum
  - Zookeeper Client (ZKFC - ZooKeeperFailoverController)

Split Brain Scenario

- Occurs when both NN acts as active NNs
  - Both make changes to the namespace
- Fencing
  - To Make Sure only one NN is active
  - Approach
    - Quorum based Storage

NFS Gateway - NFS Access to HDFS

- browse the HDFS file system through their local file system
- download files from the HDFS file system
- can upload files
- Supports NFSv3
- Can not write directly into HDFS
- HDFS- 4750

HDFS SnapShots

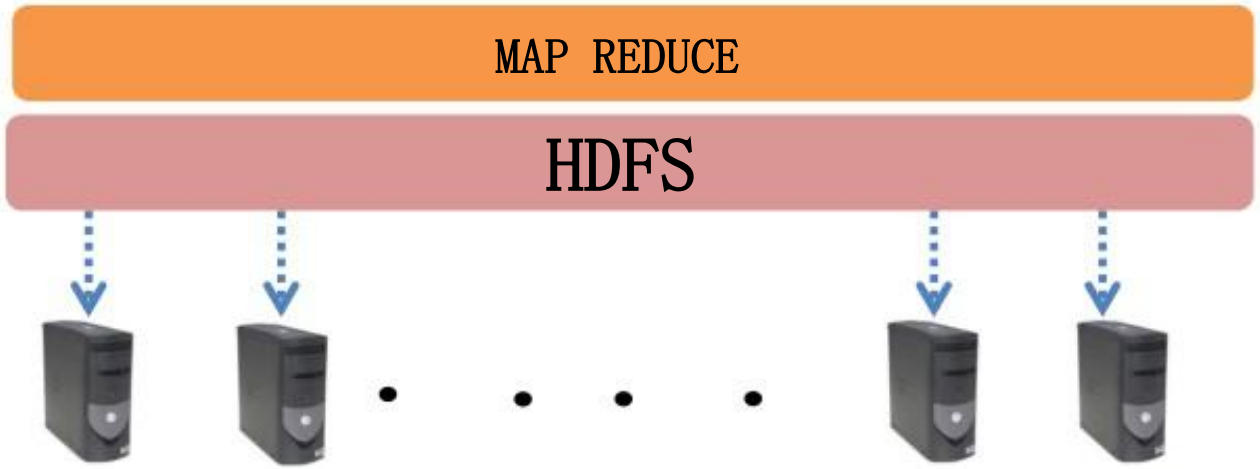
- Read-Only Copy-on-Write (COW) snapshots
- Snapshots for entire namespace or sub directories
- use cases
  - data backup - protection against user errors - disaster recovery.
- JIRA - HDFS-2802

Other Features

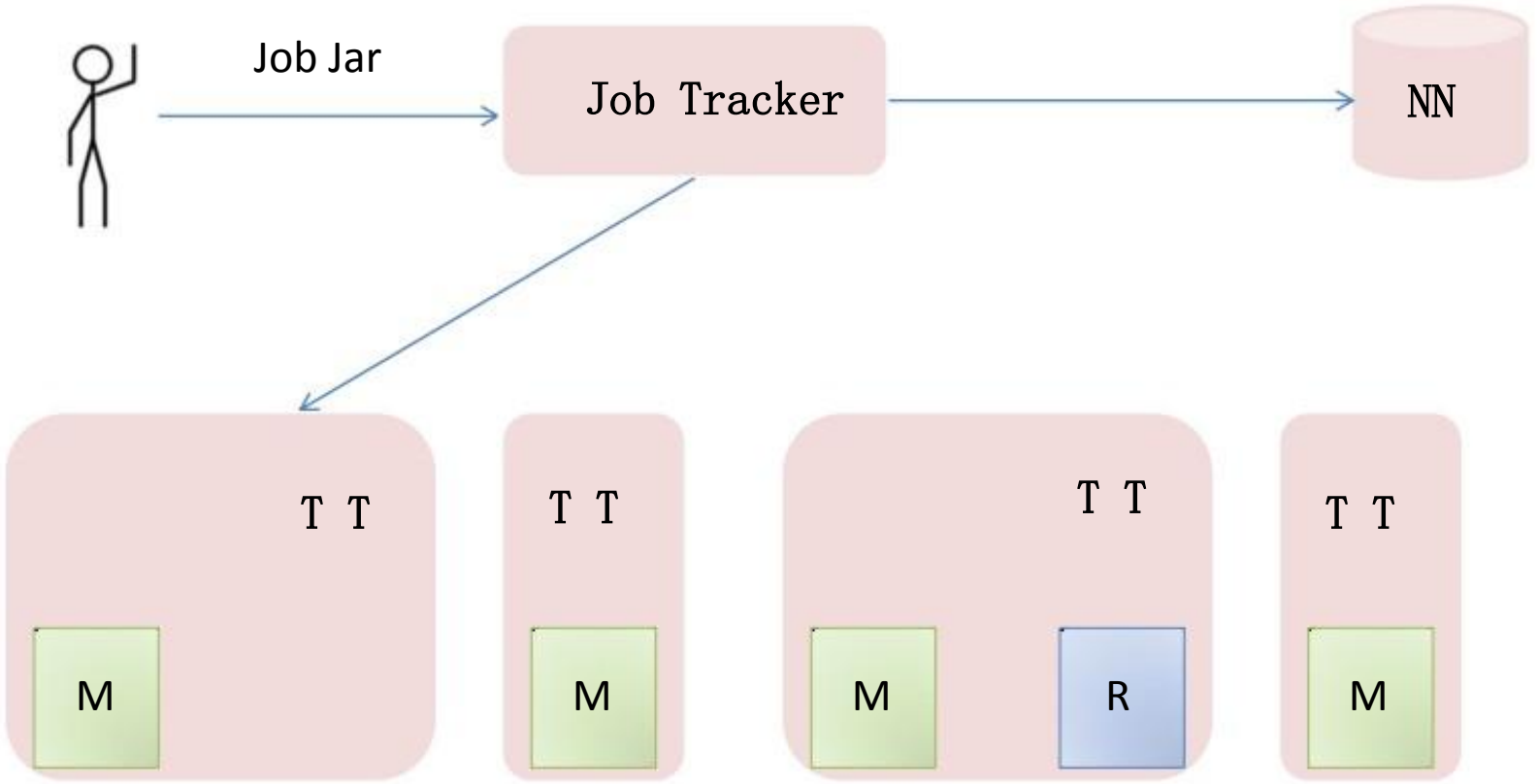
- Short-Circuit Local Reads - (Impala)
- Protobuf - Wire compatibility
  - Rolling Upgrades
- On the wire encryption
- Client Side Mount table
- More Performance Improvements

Classic MapReduce

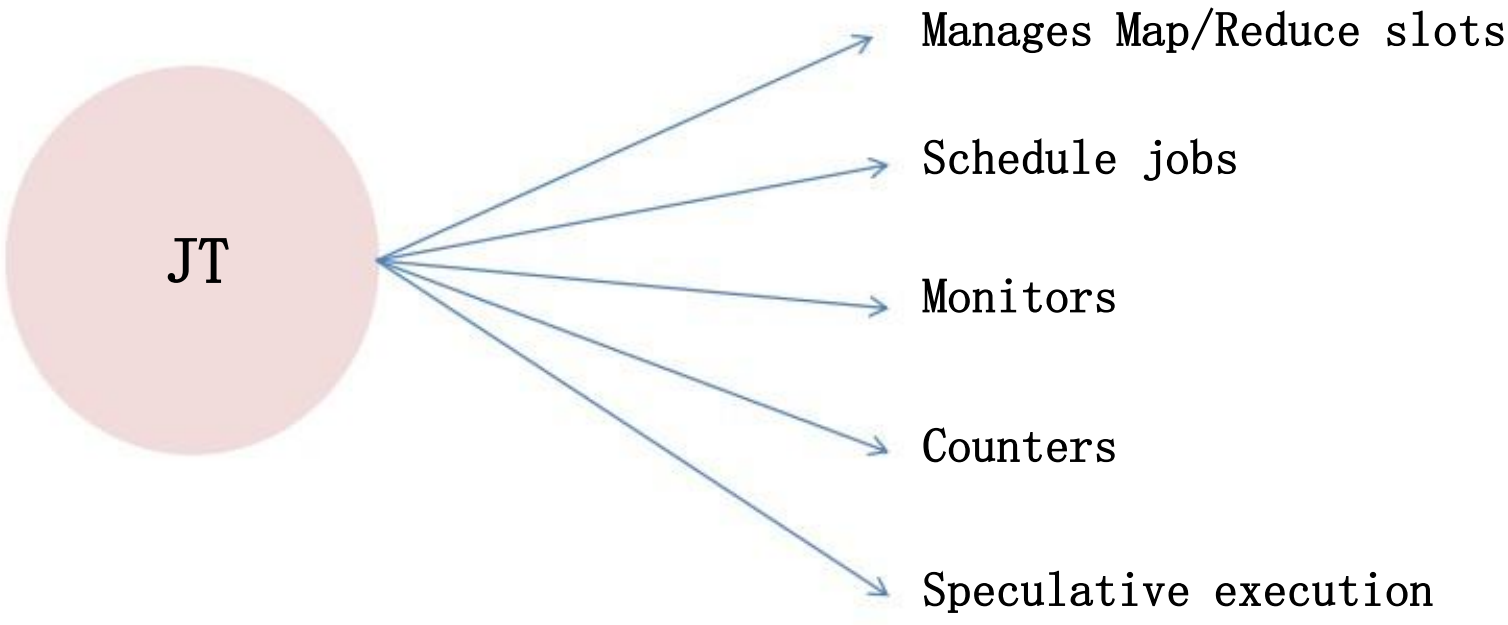
- Distributed Parallel Processing
- Codesigned , colocated , codeployed with HDFS
- Complete Abstraction - Programming APIs exposed
- Component Failure Recovery
- Consistency



Classic MapReduce



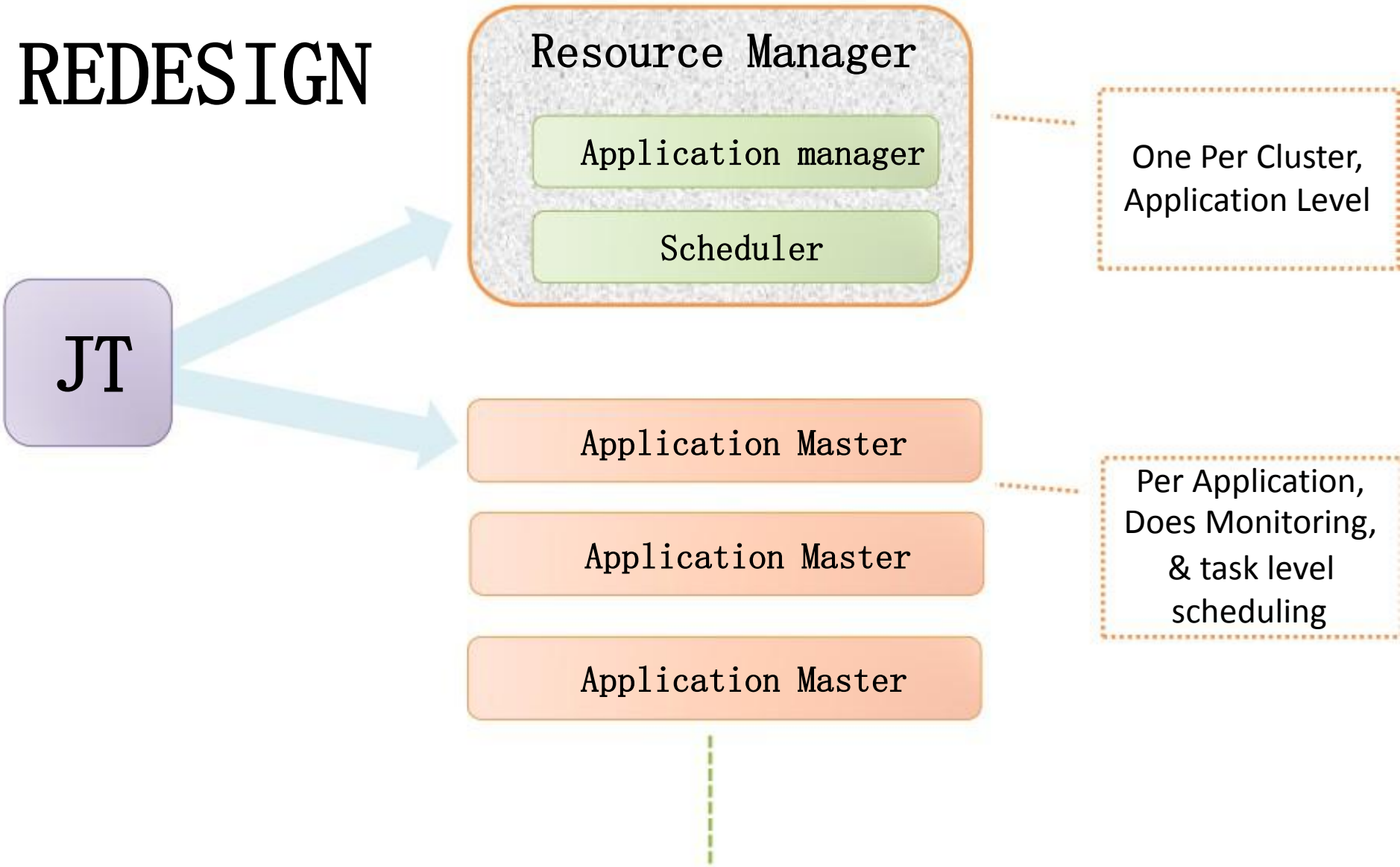
Job Tracker Responsibilities

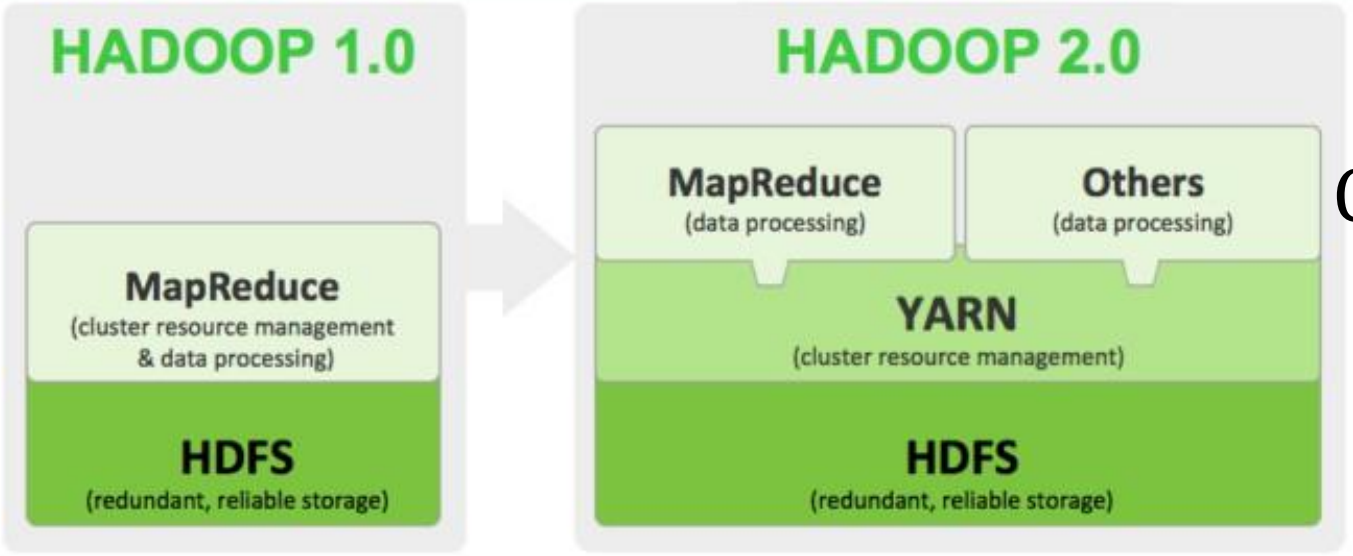


MR1 Limitations

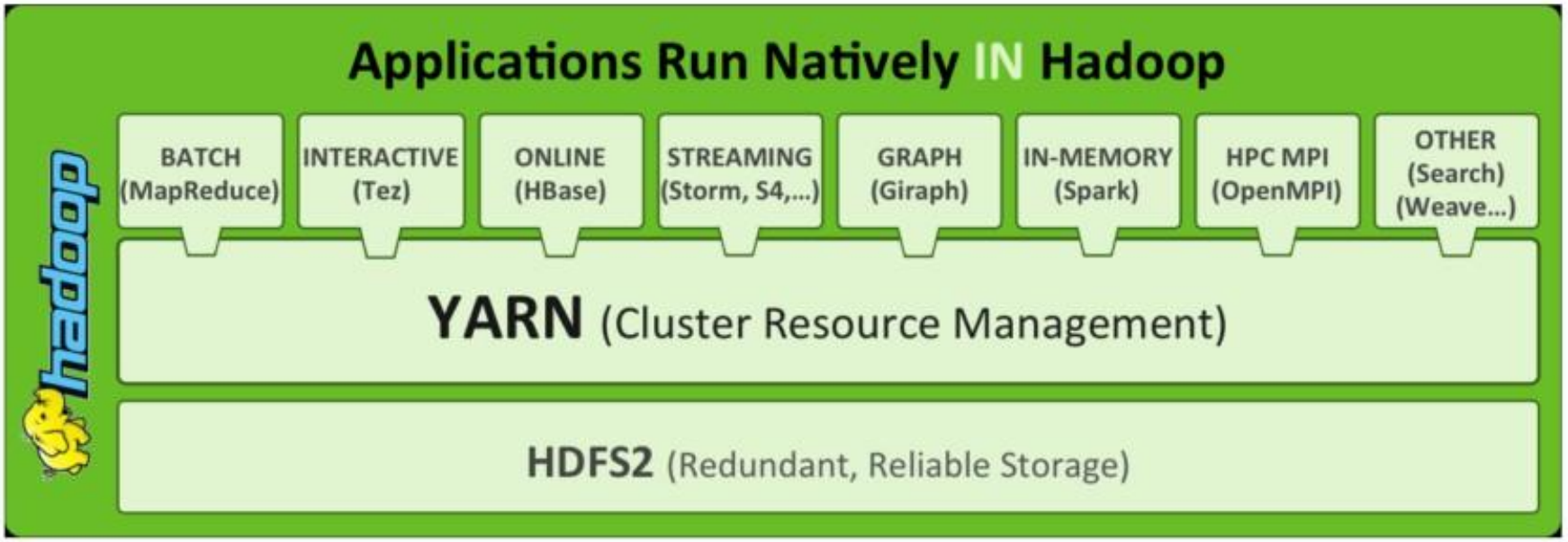
- Scalability
- Availability
- Predicatable Latency
- Cluster utilization
- Lack of Support for alternate paradigms
- iterative Applications
- Wire Compatible Protocols

REDESIGN


Resource Manager
Application manager
Scheduler
One Per Cluster, Application Level
JT
Application Master
Application Master
Application Master
Per Application, Does Monitoring, & task level scheduling



Cut Copy Past



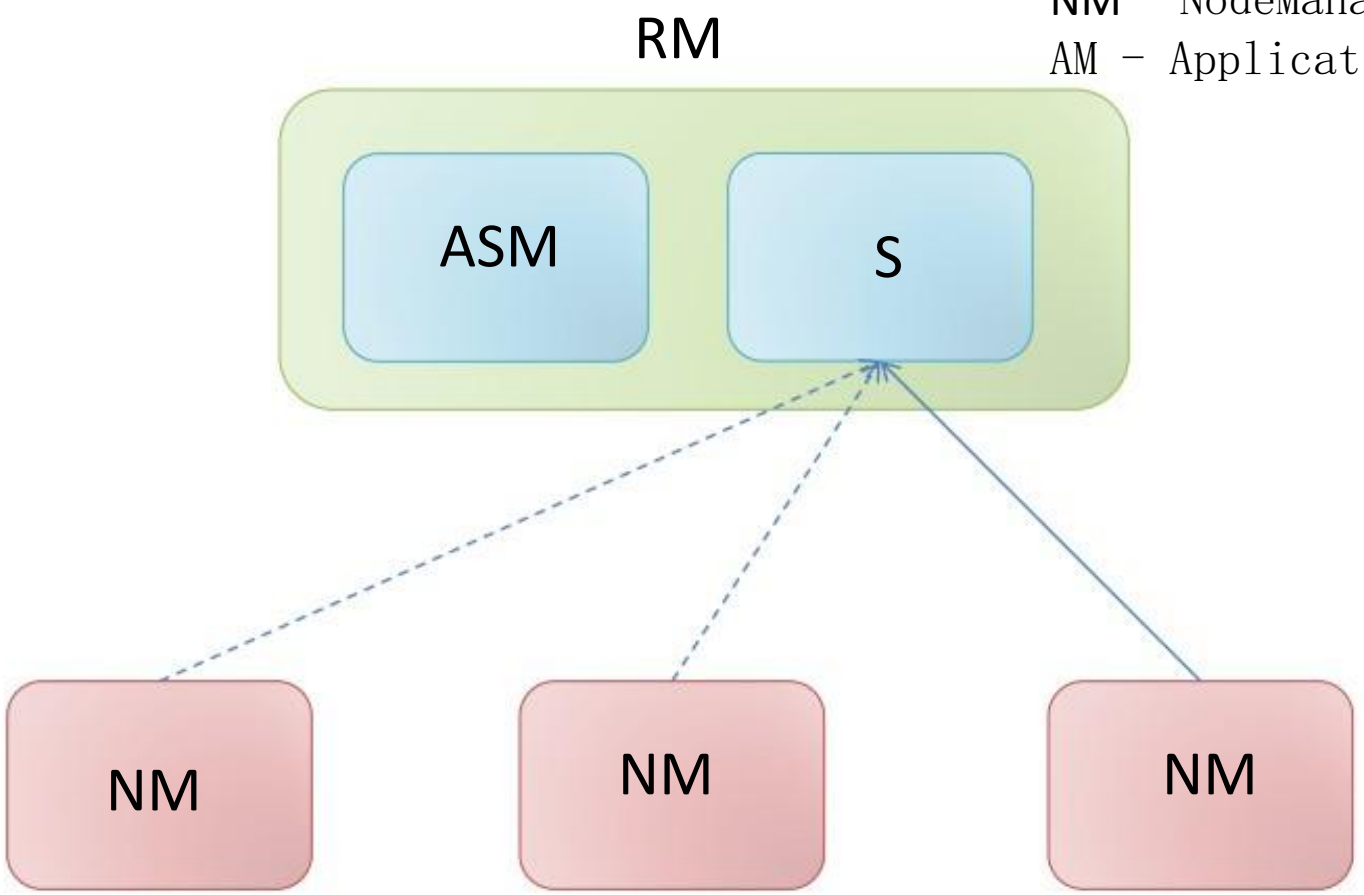
Resource: hortonworks.com

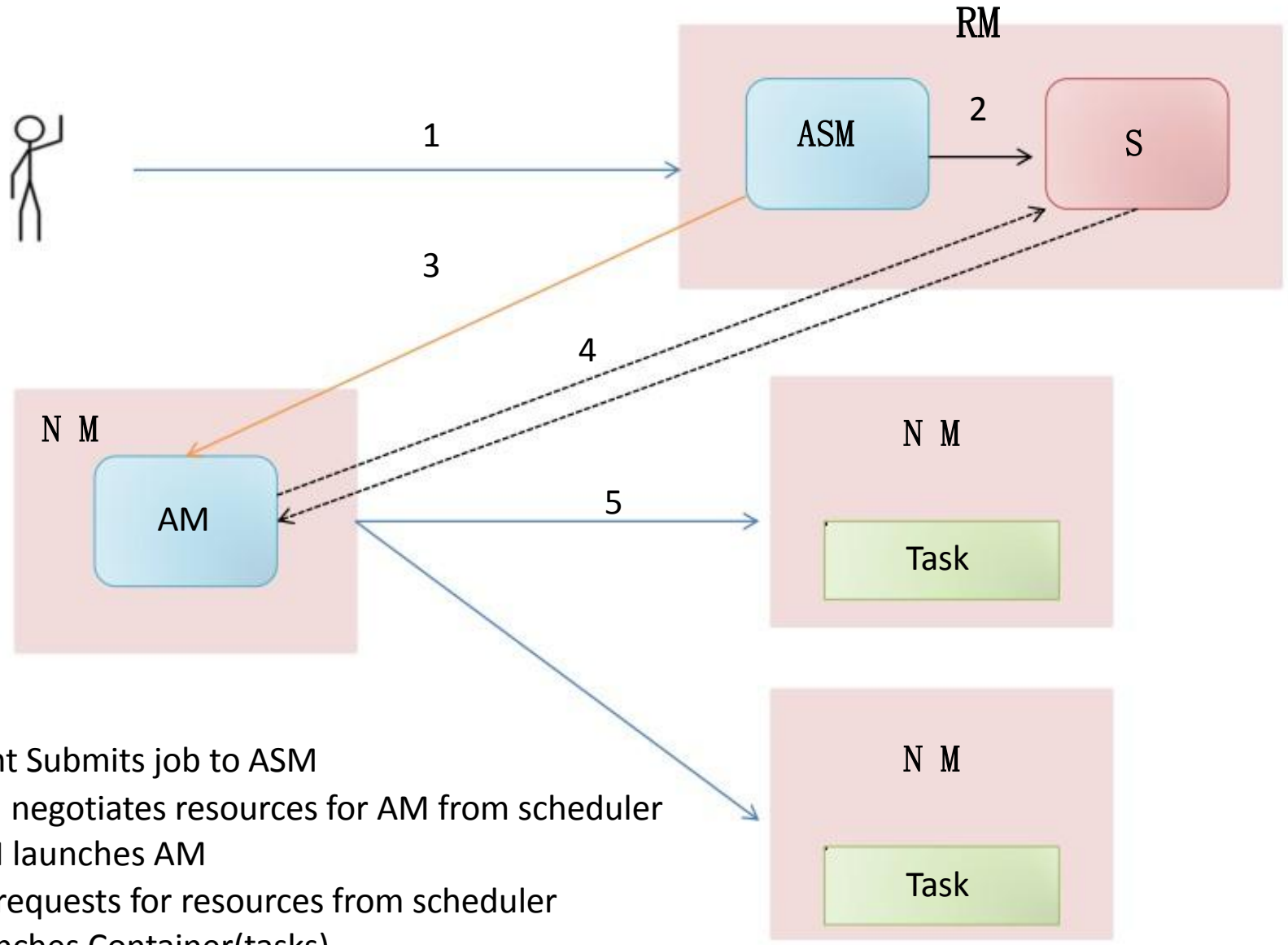
Concept

- Application
  - A Job submitted to the framework. Eg . MR Job
- Container
  - Unit of allocation
  - Replaces the fixed map/reduce slots

YARN

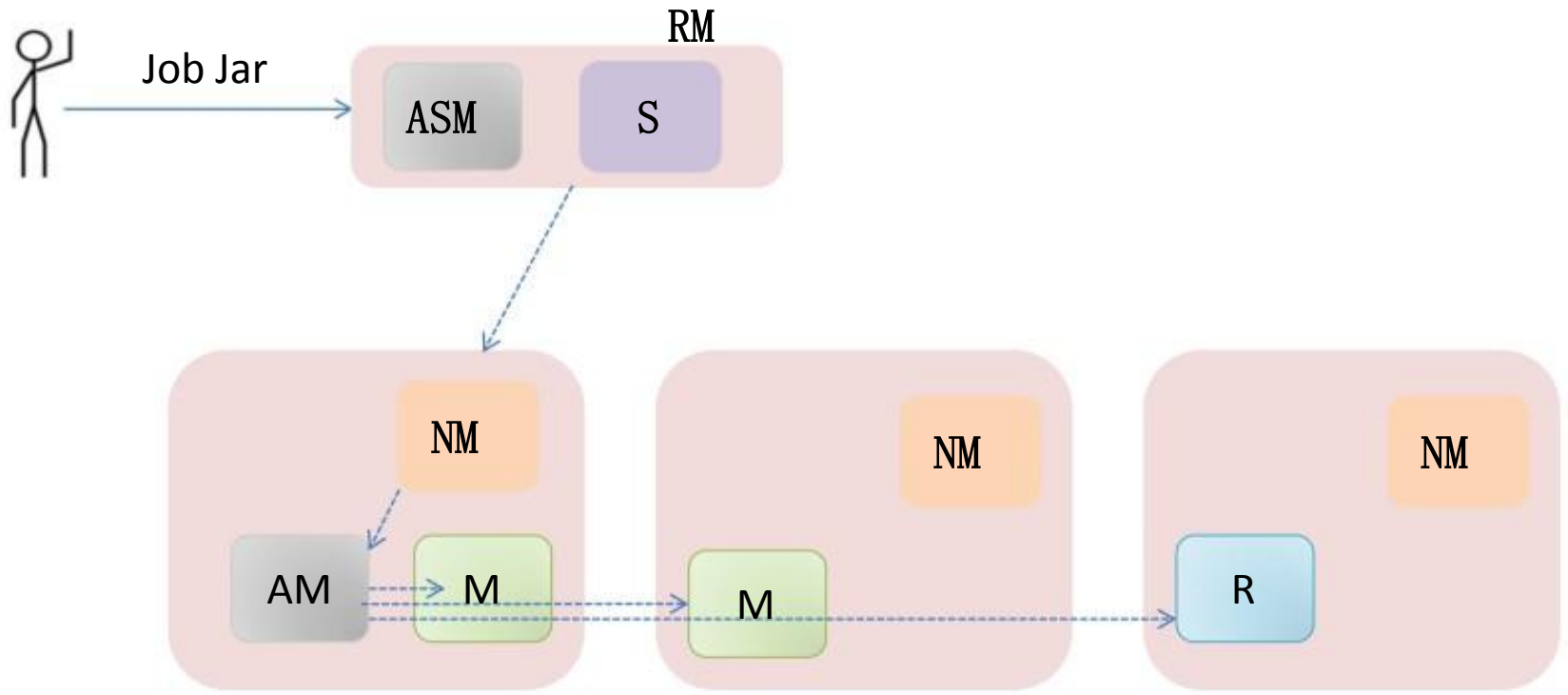
RM - Resource Manager
ASM - ApplicationManager
S - Scheduler
NM - NodeManager
AM - ApplicationMaster**





1.Client Submits job to ASM
2.ASM negotiates resources for AM from scheduler
3.ASM launches AM
4.AM requests for resources from scheduler
5. Launches Container(tasks)

Classic MR in Yarn



BackWard Compatibility

- Mapred APIs
  - No Change
- MapReduce APIs
  - Source Compatibility
  - Recompilation required
- Command line Scripts
  - MRAdmin -> RMAdmin
- Pig Scripts
    - Pig-0.10.1 ++ work on YARN
- Hive Queries
  - Hive-0.10.0 ++ work on YARN
- Oozie
  - 3.2.0+ version work

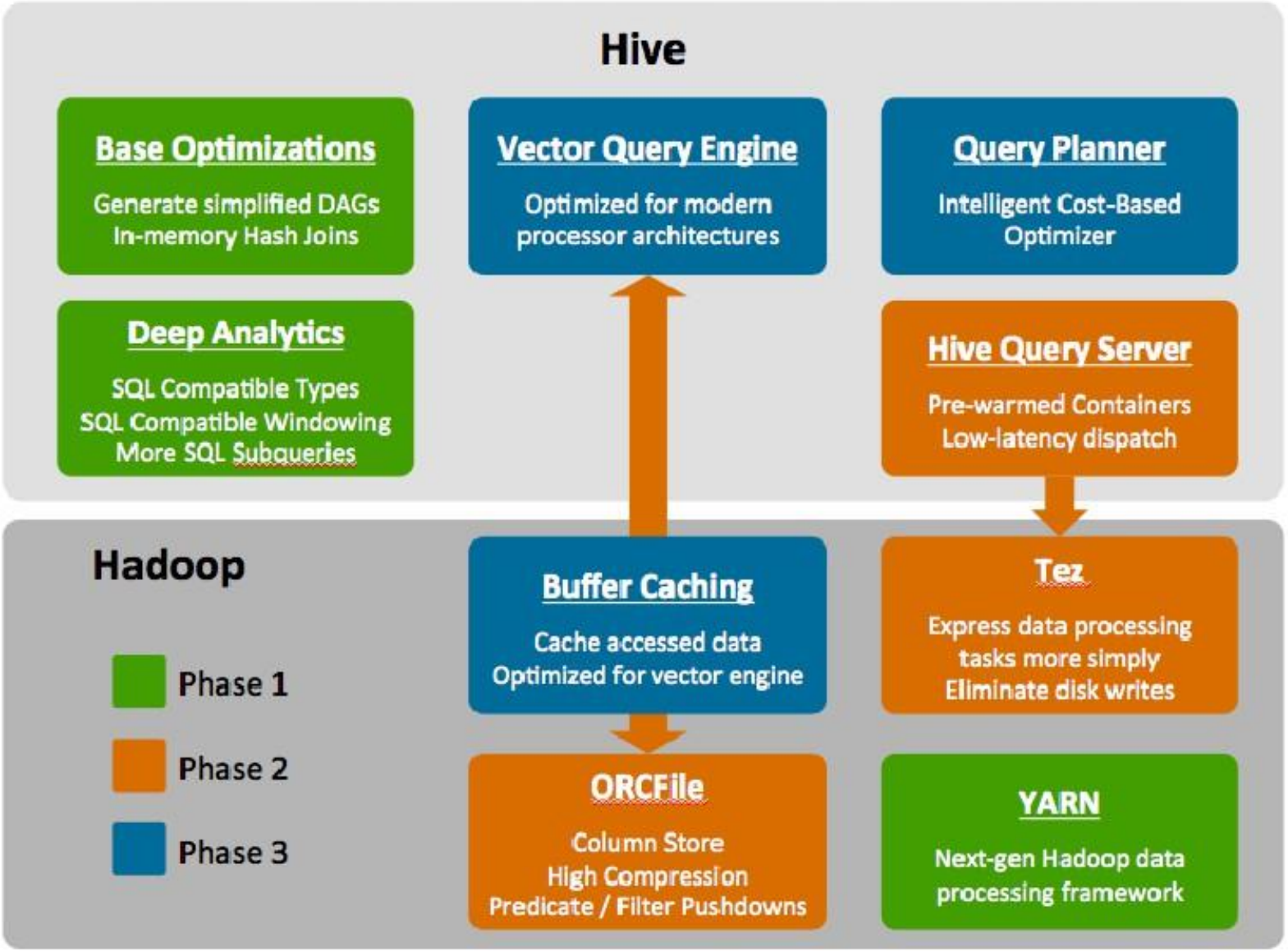
High Availability RM

- YARN -149 JIRA in Progress

Wire Compatible Protocol

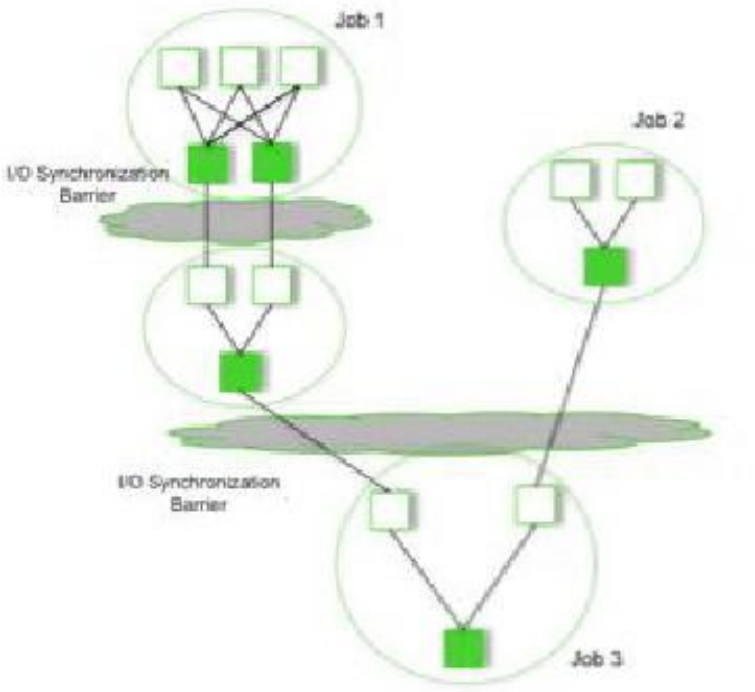
- Used Protocol buffers instead of writables
- HDFS also uses the same
- Client and YARN cluster can use different versions

The Stinger Initiative: Making Apache Hive 100x Faster

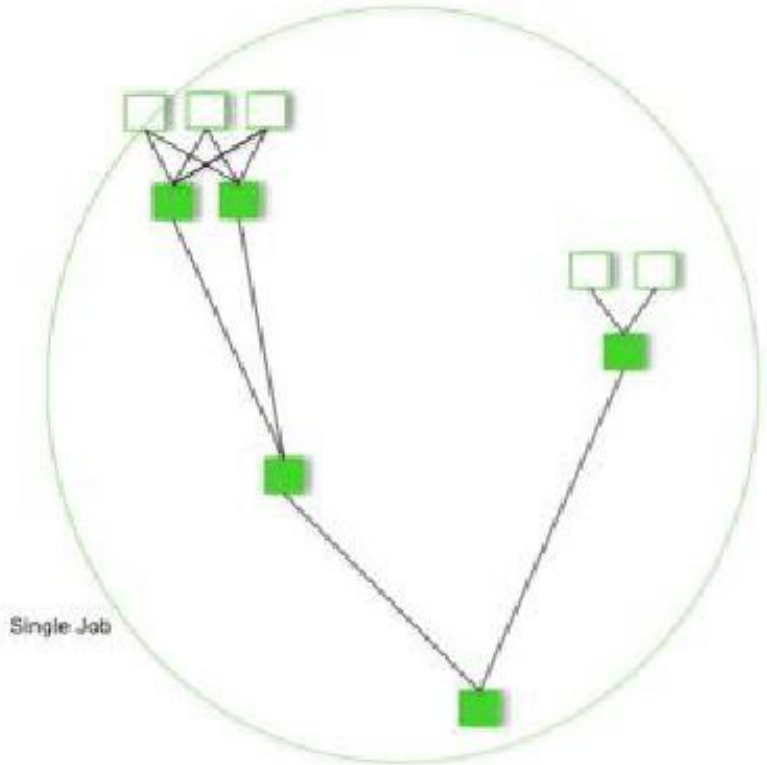


Resource : hortonworks.com

Tez - DAG of Tasks



Pig/Hive - MR



Pig/Hive - Tez

Resource : hortonworks.com

Apache Tez

- Data Processing expressed as DAG (Direct Acyclic Graph)
- Built on YARN
- Adopted by Hive as well as Pig
- Performance gains over MR
- Expressive dataflow definition APIs
- Flexible Input-Processor-Output runtime model

Happy Hadooping!!!