Master node

jobtracker | namenode | Ceph MDS

metadata

②

①

Slave node

tasktracker | datanode | Ceph

GIF

replication

...

Computation layer (MapReduce)

Storage layer (HDFS)

Compute and Storage Cluster

Data blocks

Ceph objects

① Scan Ceph to retrieve metadata for existing data

② Write Ceph metadata to HDFS in Global Information File