Assignment 3.4

1. HDFS Federation and HDFS High Availability:

Disadvantages in the Hadoop 1.x:

* The name node is not scalable.
* Does not support name node high availability
* Overburdened job Tracker
* Not possible to run non-map reduce jobs

**High availability:**

Name node is a single point of failure, need manual recovery using the secondary name node in case of failure. But this can be overcome in the Hadoop 2.x.

We can demonstrate the high availability from the below example:

Problem in 1.X: IF the name node fails then there is no access to the hadoop cluster.

Solution:

* + Hadoop 2.x supports two Name Nodes at a time one node is active and another is standby node
  + Standby Name Node manages metadata same as Secondary Name Node in Hadoop 1.x

**Federation:**

In the Hadoop 1.x federation contains the one name node and a namespace but in Hadoop 2.x we have the multiple name node and namespaces.

In order to scale the name service horizontally, JDFS federation uses multiple independent namenodes. The name nodes are federated, that is, the name node are independent and do not require coordination with each other.

1. HDFS handle failure while writing the data:

First, the pipeline is closed, and any packets in the acknowledgment queue are added to the front of the data queue so that datanodes that are downstream from the failed node will not miss any packets.

If the data node fails then failed datanode is removed from the pipeline, and the remainder of the block’s data is written to the two working datanodes in the pipeline.

The data nodes which are alive they are communicated with the namenode so that the data node which are not giving the heart break signals can be deleted.