Network Characterization of Hadoop

Abdul Navaz and Deniz Gurkan

UNIVERSITY of
HOUSTON
YOU ARE THE PRIDE

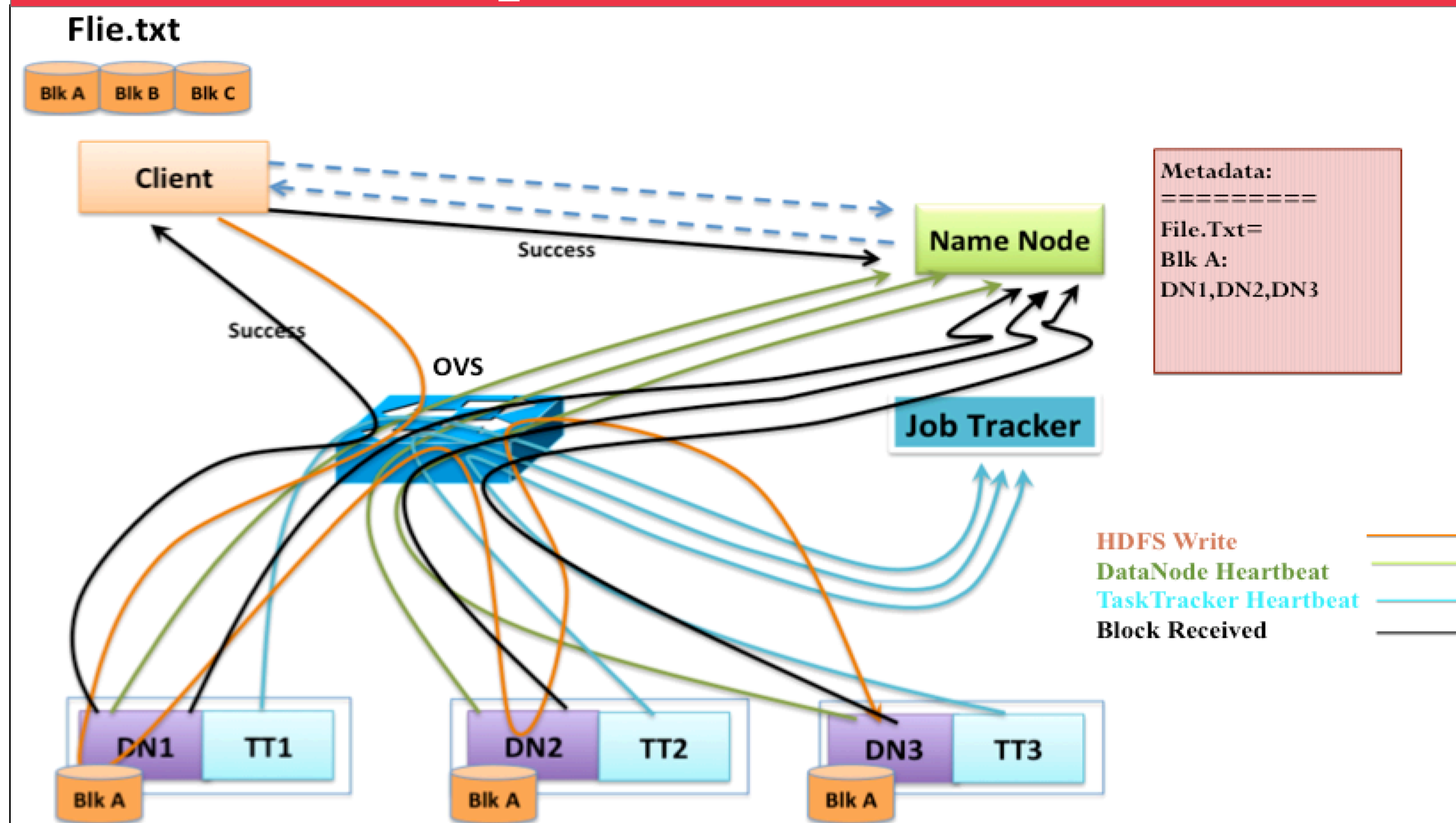COLLEGE of TECHNOLOGY

Goal

To demonstrate the networking messages in Hadoop cluster, separating the data and control traffic and to show the need of better utilization of network.
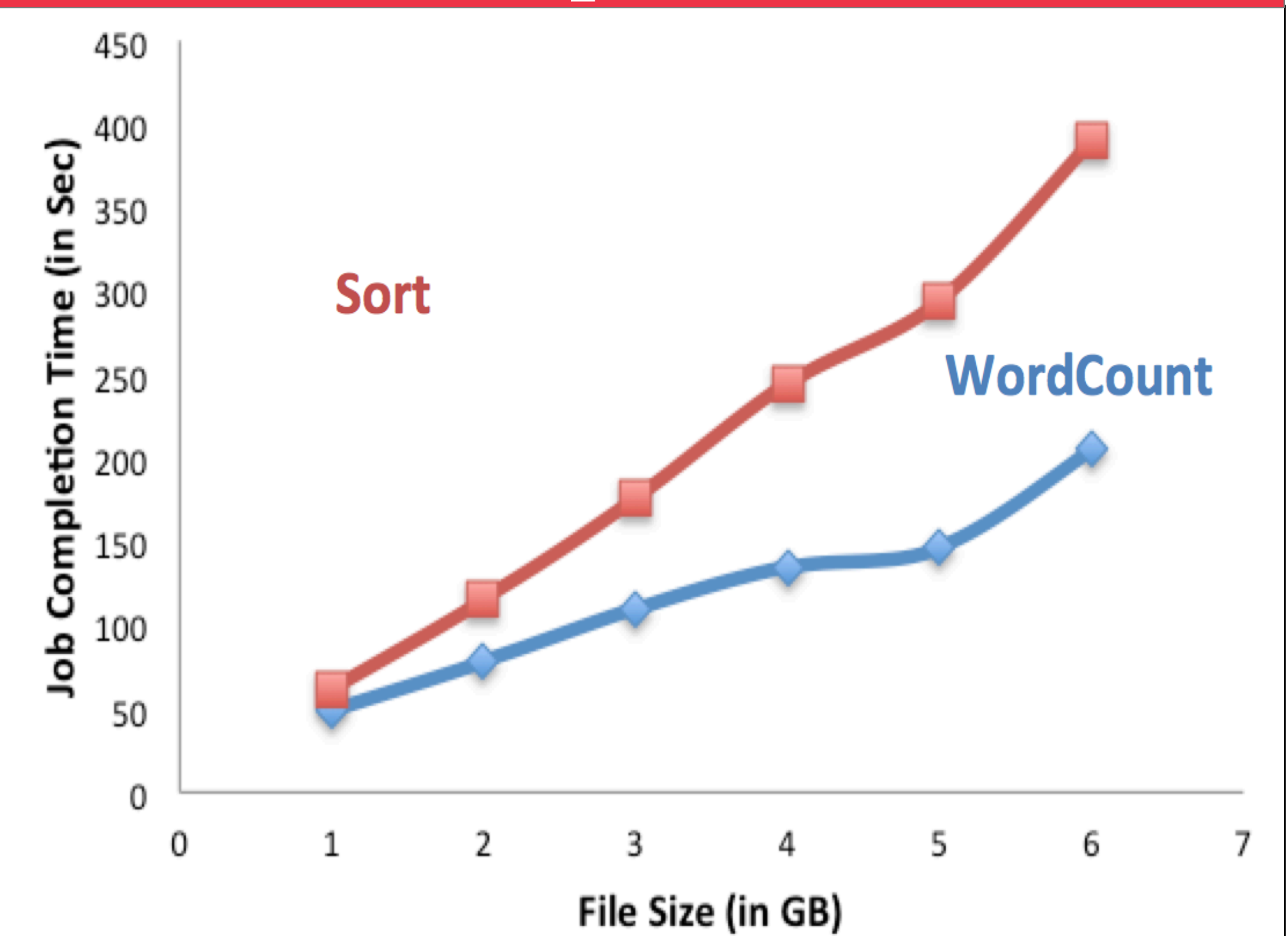
Hadoop Running on VTS
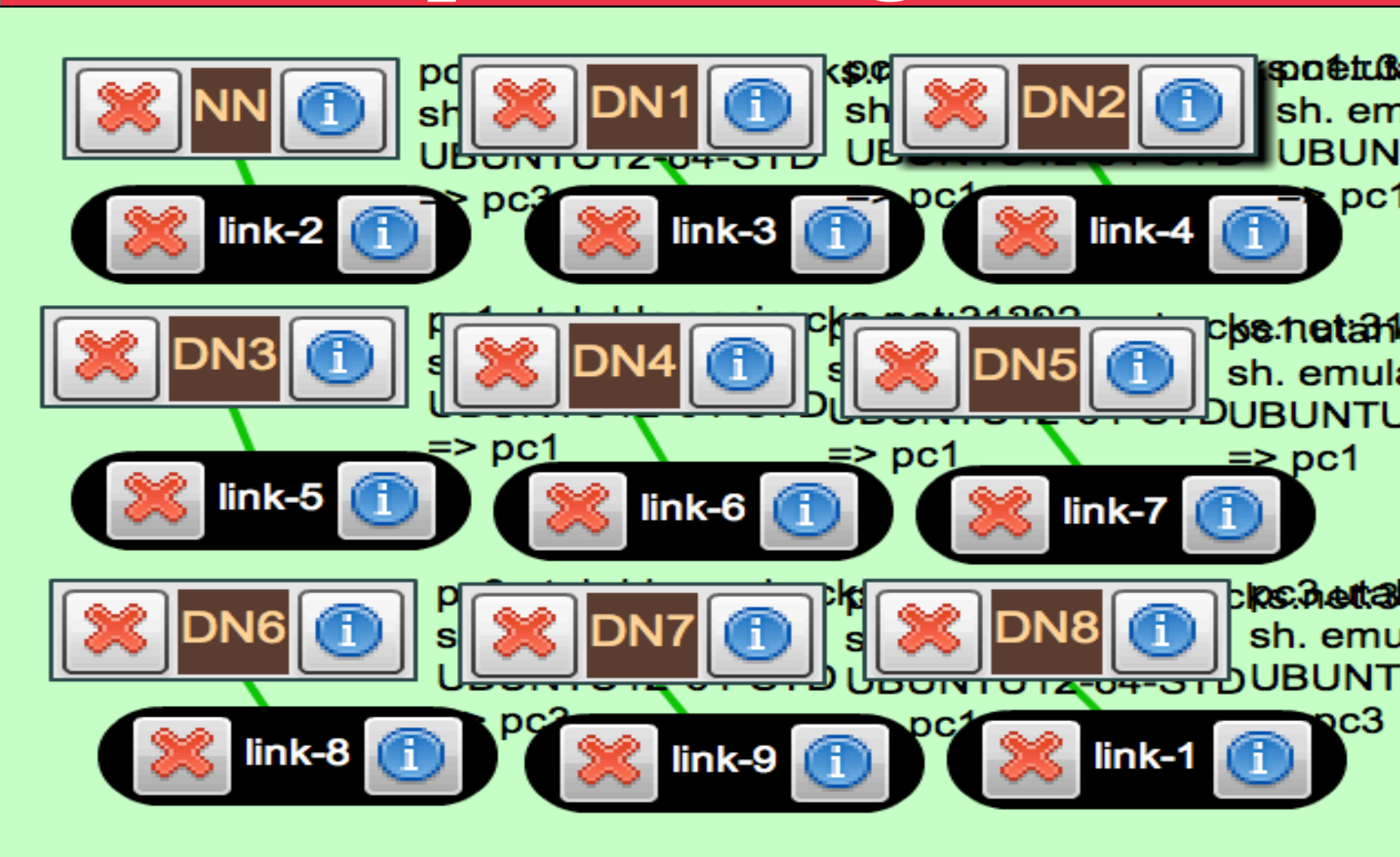
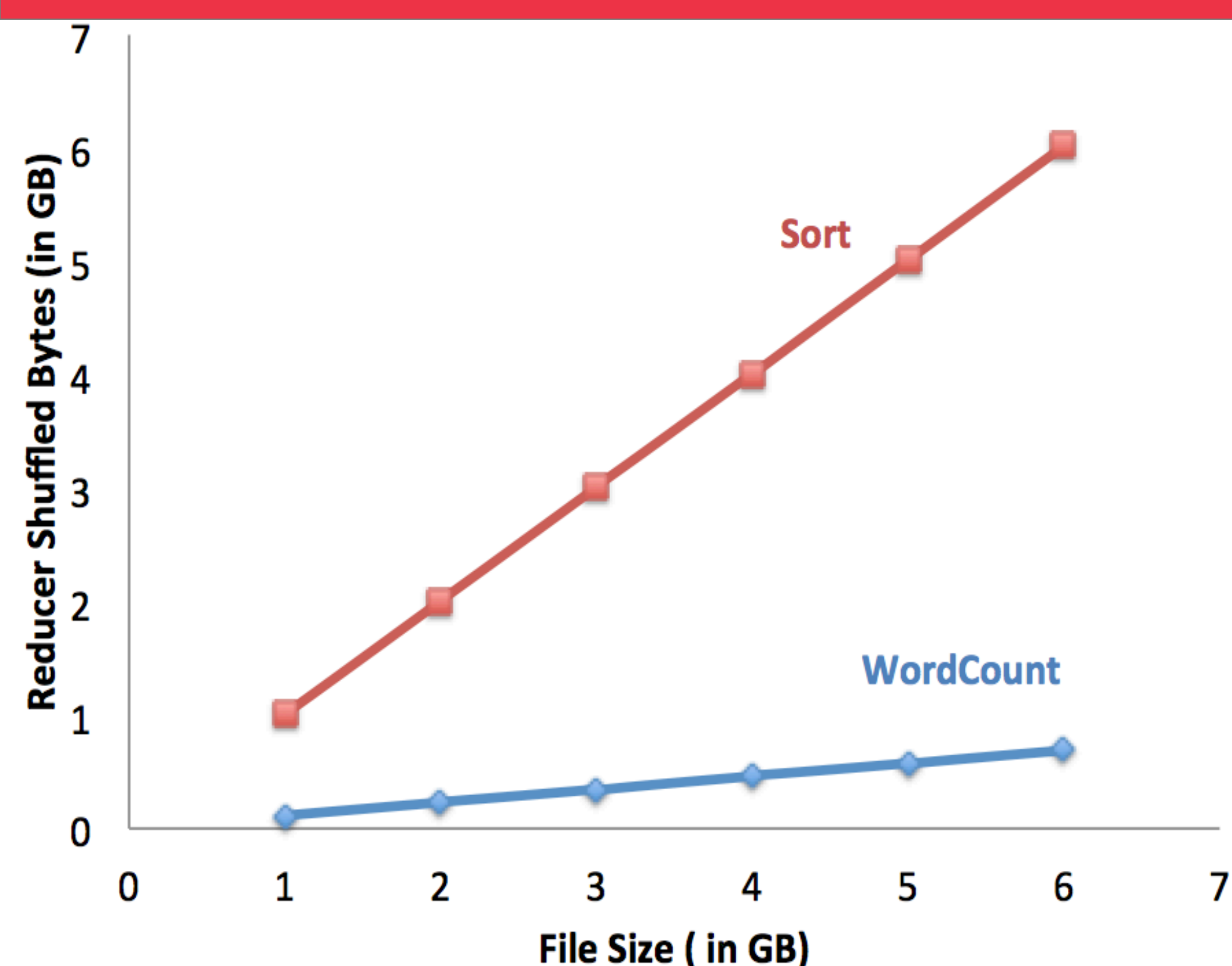1 Master Node , 8 Slave Node
176 Static Flows

Hadoop HDFS Architecture

Job Completion Time

Shuffled Traffic

Analysis

➢ Control traffic increases with the job completion Time
➢ Amount of Shuffling Traffic Depend on type of the job.
➢ Task Tracker Sends Heartbeats every 0.3secs
➢ Datanodes sending Heartbeats every 3secs.
➢ TCP 3 way handshake and pipelining.
➢ Protocols: Namenode protocol
   Datanode protocol
   Interdatanode protocol
   Client Protocol
   Data Transfer Protocol

Control Vs Data Traffic