# 15-440: Recitation 9

## School of Computer Science
## Carnegie Mellon University, Qatar
## Fall 2012
## Date: Nov 29, 2012
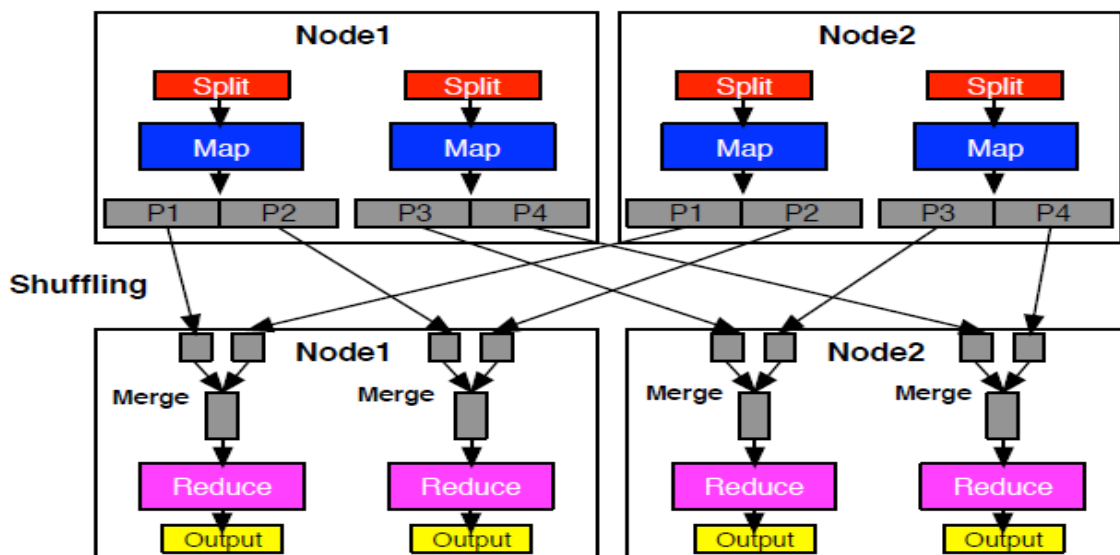
### I-    Intended Learning Outcome (ILO):

The ILO of this recitation is:

- Apply MapReduce to a real problem.

### II-    Objectives:

- Discuss with students the high-level MapReduce data flow.
- Help students develop and run MapReduce programs.

### III-    High-Level MapReduce Data Flow:

**IV-    The WordCount Problem:**

Write a MapReduce application, referred to as WordCount, that computes the occurrence frequency of each word in a text file. After you write your application, follow the following steps:

1) Create a folder for the .class files of your application using the following command:

   **$ mkdir WordCount_Classes**

2) Compile your WordCount program using the following command:

   **$ javac –classpath $HADOOP_HOME/ hadoop-0.20.2-core.jar –d WordCount_Classes   WordCount.java**

   Assuming that the name of your WordCount application is WordCount.java and that HADOOP_HOME is the root of the Hadoop installation.

3) Create the jar file required by Hadoop to run your application using the following command:

   **$ jar –cvf WordCount.jar –C WordCount_Classes/ .**

   The -C WordCount_Classes part of this command directs the Jar tool to go to the WordCount_Classes directory, and the **.** following

   -C WordCount_Classes/ directs the Jar tool to archive all the contents of that directory

4) Create two simple sample text files, file01 and file02. For instance you can have them as follows:

   **file01:**

   Welcome to MapReduce

**file02:**

Welcome to MapReduce in 15440

5) Create an input directory in HDFS using the following command:

**$ hadoop dfs -mkdir /user/hadoop/wordcount/input**

6) Copy file01 and file02 to your HDFS input directory using the following commands:

**$ hadoop dfs -copyFromLocal file01 /user/hadoop/wordcount/input**
**$ hadoop dfs -copyFromLocal file02 /user/hadoop/wordcount/input**

7) Check that file01 and file02 now exist at user/hadoop/wordcount/input using the following command:

**$ hadoop dfs -ls /user/hadoop/wordcount/input**

8) Run your WordCount application using the following command:

**hadoop jar WordCount.jar WordCount /user/hadoop/wordcount/input /user/hadoop/wordcount/output**

9) Use the MapReduce Web UI to view information about your job.

http://your_user_name-hadoop-node1.qatar.cmu.local:50030/jobtracker.jsp