CS455 - Lab 08

Thilina Buddhika

March 23, 2015
Agenda

- Course logistics
- Quiz 6 Review
Course Logistics

- HW2-PC grades are released. To check the feedback, 
  `~cs455/bin/peek HW2-PC feedback.txt`
- HW2-WC grades are also released. Feedback was provided only if there are deductions.
The number of map instances in a MapReduce stage typically exceeds the number of reduce instances. [True/False]

The number of partitions for intermediate outputs generated in a MapReduce operation at each mapper is equal to the number of mappers. [True/False]

Relational databases are ideal for ad hoc analysis where the data access patterns are write-once and read-many-times. [True/False]
4. Intermediate output files generated by a mapper are stored at the Master. [True/False]

5. A single reducer task will handle only a single intermediate key. [True/False]

6. In the case of processing large datasets, pushing computations to the data is far more efficient than pulling data to the computation. [True/False]
The partitioning function in MapReduce is used for splitting the input and assigning data splits among the Mappers. [True/False]

In contrast to MPI, MapReduce dataflow is implicit. [True/False]

Which of the following statements is incorrect about MapReduce:

a) Data locality for processing is assured at the mappers.

b) Data locality for processing is assured at the reducers.

c) Implements the shared nothing architecture.

d) None of the above.
When updating small portions of a dataset, the traditional B-Tree approach used in databases works well. [True/False]
Next Week..

- Assignment 3
- Setting up Hadoop
Questions ?