CS426 Mining Massive Data Sets
Project 1: Hadoop Setup

Through this project, you will understand the architecture of Hadoop and know the terms like master, slave, job, task, name node, data node and so on.

You need to install Hadoop cluster environment and run WordCount example on the platform. You can follow the steps and get detailed information from course slides and the official website of Hadoop.

A report is required to prove your progress. In the report, you should detail your setup process by including the logs and screenshots of the web pages in running status. Any mention of more than just-the-basic will be a plus.

In your report, you must list ALL members' names and student IDs.

Some notes:

1. You may find useful information from: The official website: http://hadoop.apache.org, Michael G. Noll's Blog: http://www.michael-noll.com/.
2. Reports should be printed out. Each team only needs to submit one copy by the team leader.
3. Only cluster mode will be accepted. It is strongly recommended to use real Ethernet instead of virtual machines. If you do so, please mention it in your report.