

# CSEE5590-0001/490-0003: Big Data Programming

## Lesson Plan # 3

ICP Feedback and Submission Link :

<https://forms.gle/xMAmr3zATrtMG5cX7>

### Importing Libraries

```
import org.apache.commons.logging.log;  
import org.apache.commons.logging.logFactory;
```

### Required Libraries (POM.XML)

<https://umkc.box.com/s/at5sr2ms080gt7lh53msce3gtoxml1gt>

<https://umkc.box.com/s/fi9zvl413ts3um4zx1xbgkw3ooj0z8x0>

**Lesson Title:** *Hadoop 2*

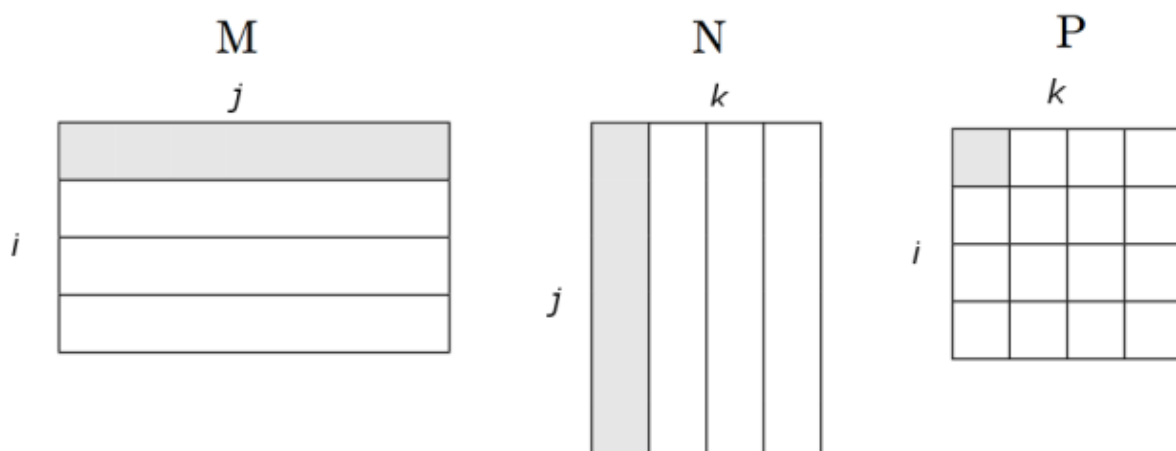
**Lesson Description:** *Hadoop Distributed File System (HDFS)/ MapReduce and Big Data Applications*

### Lesson Overview:

Overview of Hadoop and Map Reduce Paradigm. The Lesson focuses on map reduce coding exercises by actual implementation

### In class exercise:

#### 1. Matrix Multiplication in Map Reduce



Suppose we have a  $i \times j$  matrix  $M$ , whose element in row  $i$  and column  $j$  will be denoted  $m_{ij}$  and a  $j \times k$  matrix  $N$  whose element in row  $j$  and column  $k$  is denoted by  $n_{jk}$  then the product  $P = MN$  will be  $i \times k$  matrix  $P$  whose element in row  $i$  and column  $k$  will be denoted by  $p_{ik}$ , where  $P(i, k) = m_{ij} * n_{jk}$ .

Create a Map-Reduce Program to perform the task of matrix multiplication

Marks will be distributed between logic, implementation and UI (presentation in GITHUB Wiki)

**Programming elements:**

Hadoop MapReduce and HDFS

**Reference:**

<https://lendap.wordpress.com/2015/02/16/matrix-multiplication-with-mapreduce/>

**Source Code:**

<https://umkc.box.com/s/x9exae99zk1dk5317gqjegolt92vba0q>

**Prerequisites:**

Ensure that Hadoop is installed, configured and is running. More details:

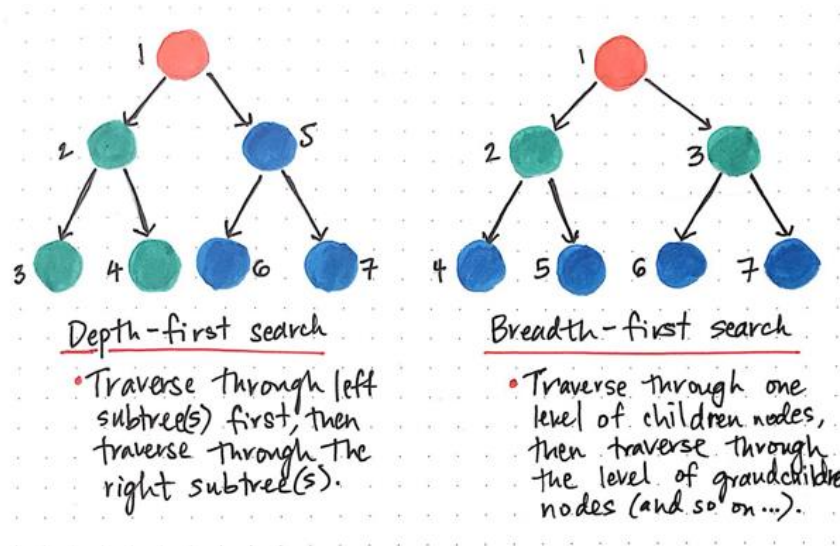
[Single Node Setup](#) for first-time users.

[Cluster Setup](#) for large, distributed clusters.

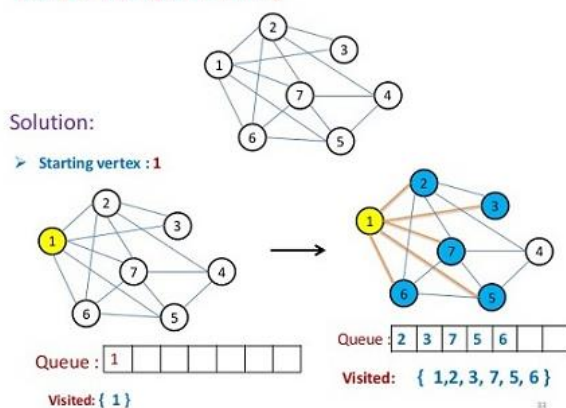
## **2. Bonus**

### **1. BFS and DFS**

**Breadth First Search and Depth First Search in Graph using Map Reduce**



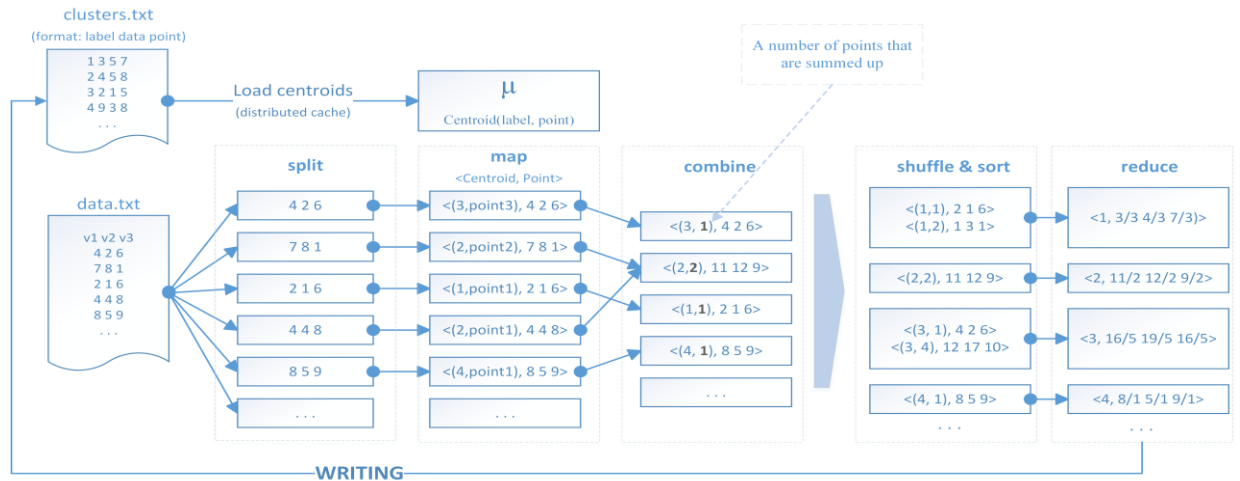
#### Example: BFS Algorithm Tracing



Implement Breadth First Search and Depth First Search in Map Reduce Paradigm

After completion of your ICP fill in the form. Any available TA/instructor will come to you and evaluate ICP

2. Apply Map reduce problem using K-Means Clustering Technique. A view point of the such algorithms are presented in the screenshot. Convert this into code and use right dataset to implement this scenario.



Reference: <https://www.edureka.co/blog/mapreduce-tutorial/#usecase>

### ICP Submission Guidelines:

1. ICP Submission is in pairs of four students.
2. Once completed, must be presented to TA or Instructor before the completion of the class
3. Submission after the deadline is considered as late submission. (Check the late submission policy in the syllabus)
4. ICP Code with brief explanation should be pushed to GitHub.
5. Submit your screenshots as well to GitHub and documentation. The screenshot should have both the code and the output.
6. Submit a demo video 2-3 min showing your assignment with a voice over explaining your work if you are unable to complete ICP within the deadline due to genuine reason.
7. Provide the video submission link through the submission form <https://forms.gle/xMAmr3zATrtMG5cX7>

***Cheating, plagiarism, disruptive behavior and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL: <https://catalog.umkc.edu/special-notices/academic-honesty/>***