

# CSCE 5300 Introduction to Big data and Data Science

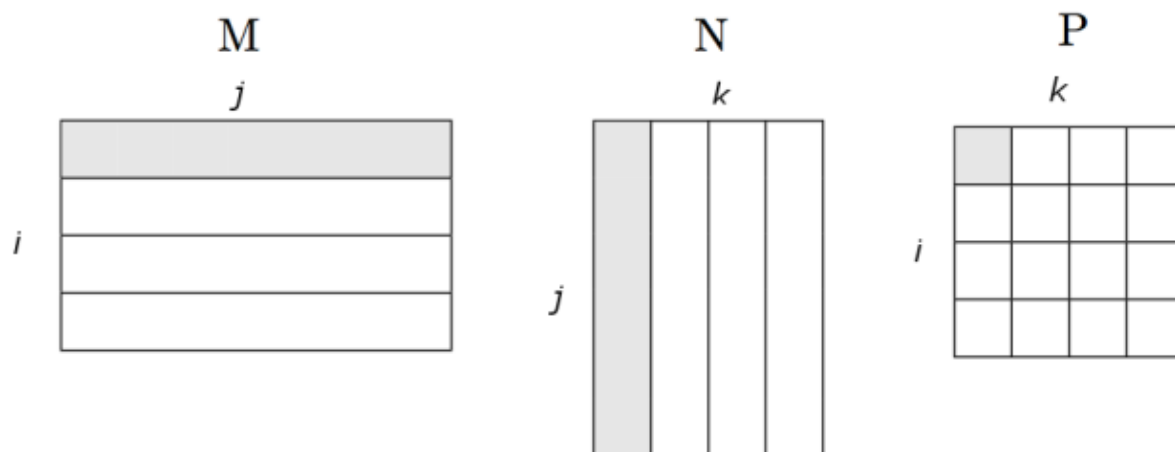
## ICE-3

**Lesson Title:** Hadoop MapReduce and Hadoop Distributed File System (HDFS)

**Lesson Description:** Overview of Hadoop and Map Reduce Paradigm. The Lesson focuses on map reduce applications with coding exercises by actual implementation

### In class exercise

#### 1. Matrix Multiplication in Map Reduce



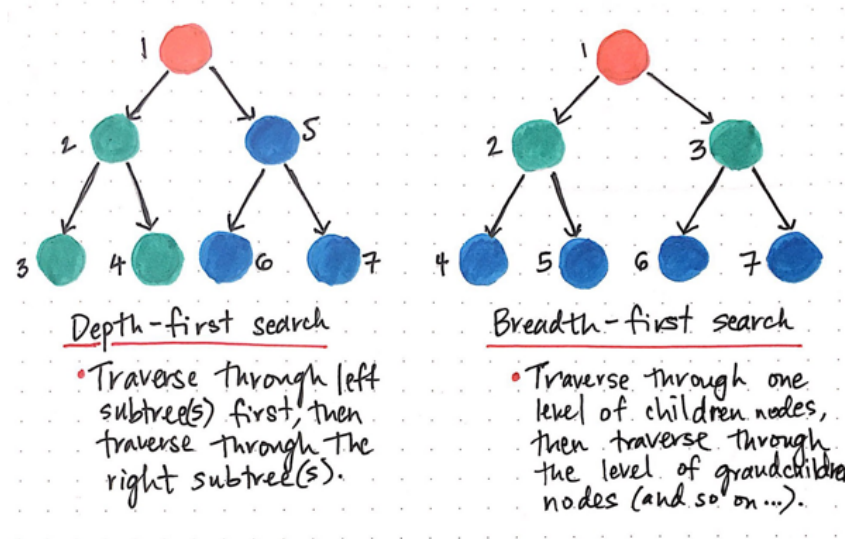
Suppose we have a  $i \times j$  matrix  $M$ , whose element in row  $i$  and column  $j$  will be denoted  $m_{ij}$  and a  $j \times k$  matrix  $N$  whose element in row  $j$  and column  $k$  is denoted by  $n_{jk}$  then the product  $P = MN$  will be  $i \times k$  matrix  $P$  whose element in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

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

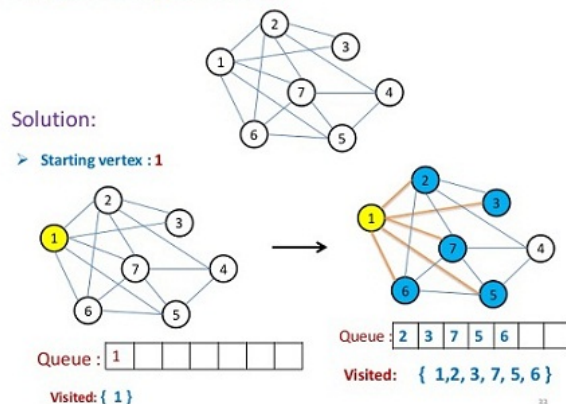
#### **Reference:**

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

2. Breadth First Search using Map Reduce
3. Depth First Search using Map Reduce



Example: BFS Algorithm Tracing



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

4. 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.

Marks will be distributed between logic, implementation and UI

### Programming elements:

Hadoop MapReduce and HDFS

### Source Code:

Given in canvas.

## **ICE Submission Guidelines**

1. ICE Submission is individual.
2. ICE code has to be properly commented.
3. The documentation should include the screenshots of your code/results with explanation.
4. Provide the explanation of the dataset/exercise as per your understanding.
5. The similarity score for your document should be less than 15%.
6. All you need to do is submit the source code (properly commented) and documentation (.pdf/.doc) with explanation and screenshot of source code having input logic and output results.
7. Submission after the deadline is considered as late submission.