CSEE5590-0001/490-0003: Big Data Programming

Lesson Plan # 9

ICP Feedback and Submission Link :

https://forms.gle/xMAmr3zATrtMG5cX7

Lesson Title: ***Apache Spark II***

Lesson Description: ***Apache Spark Practice***

Lesson Overview:
Apache Spark is a unified analytics engine for big data processing, with built-in modules for streaming, SQL, machine learning and graph processing.

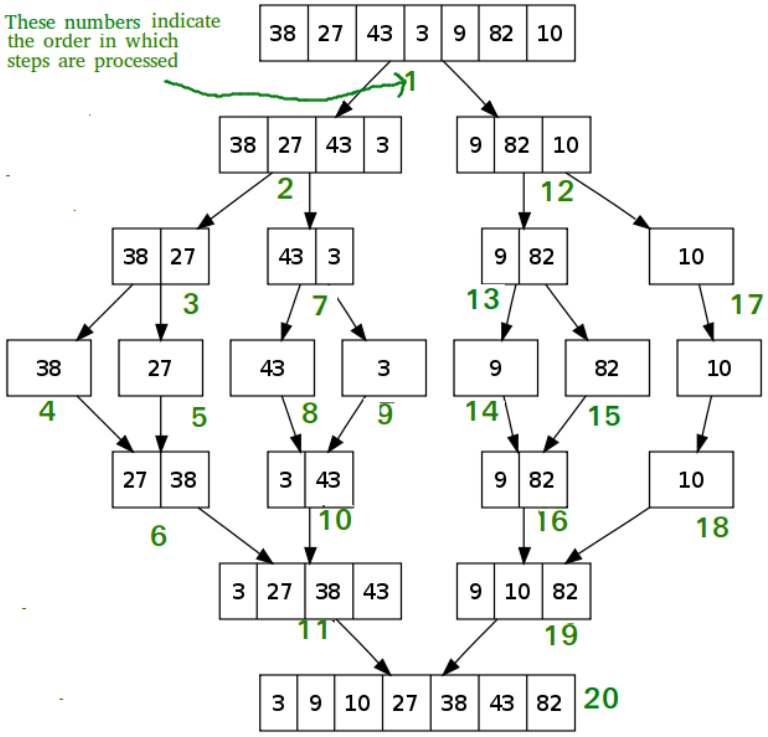
In class exercise:

1. ***K-Means Clustering Algorithm***

Review of the appraoch:
https://umkc.box.com/s/74rda02dy0uy1qpbasel5pl06gksf4wz
Review of Source Code:
https://github.com/apache/spark/blob/master/examples/src/main/scala/org/apache/spark/examples/LocalKMeans.scala

2. ***Merge Sort Algorithm***

Merge Sort is a Divide and Conquer algorithm. It divides input array in two halves, calls itself for the two halves and then merges the two sorted halves. The merge () function is used for merging two halves. The merge (arr, l, m, r) is key process that assumes that arr[l..m] and arr[m+1..r] are sorted and merges the two sorted sub-arrays into one.



Create a Map-Reduce Program to perform Merge-Sort Algorithm in Spark.

3. ***DepthFirst Search***

Implement Depth First Search in Graph in Apache Spark

https://medium.com/basecs/deep-dive-through-a-graph-dfs-traversal-8177df5d0f13

ICP Submission Guidelines:

1. ICP Submission is individual however, it can be completed as a Team during session.
2. If completed, should 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 GitHub and 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/*