* Deep Dive into HDFS

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

* grades.csv - : Screenshot proving the data has been loaded.

A screenshot of a computer program

Description automatically generated

* Screenshots of the three chosen HDFS command outputs

1. TAIL

A computer screen shot of a computer code

Description automatically generated

1. MKDIR – Create a new directory “test-dir”

A black screen with white text

Description automatically generated

1. RM – Remove the created directory – “test-dir”

A screenshot of a computer screen

Description automatically generated

1. COUNT – Count the number of directories, files, and bytes

A black screen with white text

Description automatically generated

1. CAT - reads the file in HDFS and displays the content of the file

A computer screen shot of a person

Description automatically generated

* YARN - Screenshot of the results

A black screen with white text

Description automatically generated

* YARN - Screenshot from the YARN UI showing the updated maximum memory (2048 MB)

A screenshot of a computer

Description automatically generated

* A summary of the result and its significance.

A black and white screen

Description automatically generated

A black screen with white text

Description automatically generated

A screen shot of a computer

Description automatically generated

## Summary:

Map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). Secondly, reduce task, which takes the output from a map as an input and combines those data tuples into a smaller set of tuples. The Reducer’s job is to process the data that comes from the mapper.

During a MapReduce job, Hadoop sends the Map and Reduce tasks to the appropriate servers in the cluster. The goal is to obtain an approximation of π that is as accurate as possible based on the number of random points generated and processed within the MapReduce framework. The more points generated and processed, the more accurate the estimation of π will be.

Q: Could the estimation method, in this case, the Monte Carlo method, influence the outcome?

A: Yes, The accuracy of the estimation (the Monte Carlo method) depends on factors such as the number of samples, the quality of random generation, and the algorithm used. Increasing the number of samples generally leads to a more accurate estimation of π.

Q: Ponder on the significance of the parameters 2 and 10.

A: We can see, there are 2 maps with 10 samples per map in the given example for which the pi value is 3.8. Increasing the number of maps and samples to 10 and 70 respectively, we see a value closer to the actual pi value.

A black screen with white text

Description automatically generated

A screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## References:

Hadoop - MapReduce. Online Courses and eBooks Library. (n.d.). [https://www.tutorialspoint.com/hadoop/hadoop\_mapreduce.htm#:~:text=MapReduce%20is%20a%20processing%20technique,(key%2Fvalue%20pairs)](https://www.tutorialspoint.com/hadoop/hadoop_mapreduce.htm" \l ":~:text=MapReduce%20is%20a%20processing%20technique,(key%2Fvalue%20pairs))