Monica Santana – Week 2 Exercise

HDFS Report

```
The Thought of Carlot Services and Carlot Serv
```

Load grades.csv into HDFS

```
| According | Acco
```

SSH into worker1 and worker2 nodes

```
### Announced Control of the Control C
```

3 chosen HDFS commands – 1st one is the -cat command to print file contents

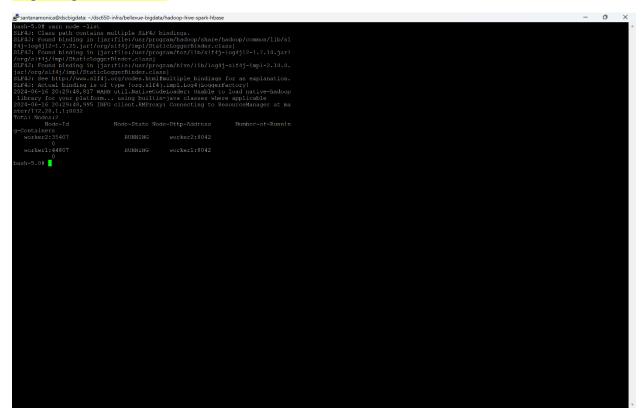
```
### Antonionical distribution - videolic - v
```

 2^{nd} command is -stat using different format sequences: %a = permissions in octal, %A = permissions in symbolic style, %b = file size in bytes

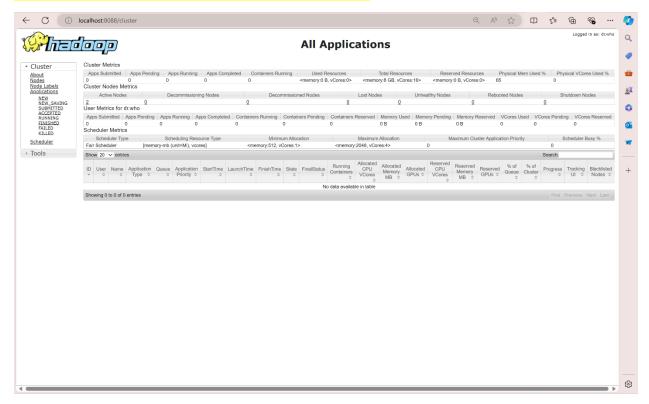
```
Memoriany of the common by the common by the foreign of the common by the common by the common by the common by the common of th
```

3rd command is -touchz to create an empty file and -rm to delete it

Inspecting YARN nodes



YARN UI with updated maximum memory of 2048 mb



Running the MapReduce Pi

The program calculated the pi to be 3.80 in 32.27 seconds. It uses the quasi-monte carlo method using the number of maps which is 2, and the number of sample of maps which is 10. MapReduce uses the mapper to generate points in a unit square to count the points inside and outside the square. The reducer then accumulates the points inside and outside of the mapper.

For pi, it was calculated by this formula of 1 (side of the square) and the area of the circle/square would be pi/4 (sides of the square).

The area of the circle is divided by the area of the square which is equivalent to the number of points inside the circle divided by the number of points of both inside and outside of the circle.

Number of points inside the circle / number of points of both inside and outside of the circle = pi/4 which is value of pi = 4 (number of points inside / total number inside & outside).

I assume larger sample sizes would calculate the pi size to be more accurate to 3.14, this sample was small with only 10.