**BDSP CA1 Screenshots**

1. Taking the **nba\_logreg.csv** file from downloads and placing into the **/home/hduser/Desktop/Lab04** folder.

A computer screen shot of a program

Description automatically generated

1. Starting up **dfs** and **yarn** for **hadoop**

A computer screen with white text

Description automatically generated

1. Moving the **nba\_logreg.csv** file into the folder on **Hadoop file system**

A computer screen shot of a computer program

Description automatically generated

1. Moving empty file **Sem\_2\_Integrated\_CA1.ipynb** to **Lab04** directory, then to **user1** folder in **Hadoop file system**

A screenshot of a computer

Description automatically generated

1. Launching **pyspark**

A screenshot of a computer program

Description automatically generated

1. Starting the **spark context**, bringing in the **nba\_logreg.csv** file as an **RDD** and **counting rows** in the dataset

A screenshot of a computer

Description automatically generated