HDFS Mini Project

1. Create a rdd for the file /home/hduser/sparkdata/home/hduser/sparkdata
2. Create a case class as per the file structure

auctionid: String, bid: Float, bidtime: Float, bidder: String, bidderrate:

Integer, openbid: Float, price: Float, item: String, daystolive: Integer

1. Create highprice\_rdd for the records which has the price value 100 and above
2. Create lessprice\_rdd for the records which has the price value less than 100
3. Write the rdd date into a directory names (/home/hduser/highprice and /home/hduser/lessprice)
4. Convert highprice\_rdd to df\_hp
5. Convert highprice\_rdd to df\_lp
6. Create a view for df\_hp dataframe as high\_price
7. Create a view for df\_lp dataframe as low\_price
8. Print the schema of high\_price
9. Print the schema of low\_price
10. Show 10 lines from high\_price
11. Show 10 lins from low\_price
12. Write a query to get maximum bid amount and minimum big amount from high\_price view
13. Write a query to get maximum bid amount and minimum big amount from low\_price view