**Spark Online Auction Case Study**

The wide usage of internet among people has clearly given a greater advantage to the e-auction portals. Right from reaching-out to the most number of buyers to selling different assets for the best prices in the shortest time, the online auction platforms have made everything possible and also easier for people today. XYZ runs online auctions company where they provide us the bidder details along with bid time, bid item and bid price to analysis on the bids happened over a period of time.

**Dataset:**

* auction.csv

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| auctionid | Unique identifier of an auction |
| bid | The proxy bid placed by a bidder |
| bidtime | The time (in days) that the bid was placed, from the start of the auction |
| bidder | eBay username of the bidder |
| bidderrate | eBay feedback rating of the bidder |
| openbid | The opening bid set by the seller |
| price | The closing price that the item sold for (equivalent to the second highest bid + an increment) |
| item | Item for auction |
| daystolive | days for auction |

**Objective**:

To do analysis on online auction dataset using Spark API (RDD/DataFrame or DataSet /Spark Streaming) with Hive

**Problem Statement:**

Load and review data for Ebay online auction data in csv format

Use the below components for the case study

1. Load the source file using RDD API and add schema
2. Use any API of your choice to do the below analysis
3. Total number of items that are auctioned
4. Get the auctions with closing price greater than 100
5. What's the average number of proxy bids per item?
6. How many bids per auction?
7. What is the maximum price of each auction?
8. Write the data to Hive table where bidder feedback rating is not 0 or negative
9. Get the hive table contents and store it as csv file with pipe delimiter and header
10. Create a spark session. Use Structured Streaming API to stream the csv data file and process item and price value for price lesser than 200 and write it as json format in a location. Mention checkpoint directory as well.