# Data Engineer coding task

## Convert level 3 to level 1 data.

You may write code in any language you are comfortable with. There should be clear instructions on how to execute the code.

The purpose of an exchange is to match market participants orders (buy against sell). The state of the market is represented in an order book. This shows for each price level and each side (buy/sell) the orders for that level.

Level 3 data are updates to market participants' orders. Each order has a side, price and quantity. They are uniquely identified by an order id. An order update can be one of ADD, UPDATE, DELETE, or TRADE. ADD inserts a new order into the book. UPDATE modifies an existing order on the book. DELETE removes the order from the book. TRADE matches orders on opposite sides of the book on the same price level; the corresponding update to the book being a reduction of quantity outstanding in the respective orders according to the amount traded.

Note that if there are any orders on both sides of a price level, then they would already be matched and traded. Hence at any given time all the outstanding sell orders are above a certain level and all the outstanding buy orders would be below that level. The lowest sell (ask) level and the highest buy (bid) level constitute what we call the BBO (best bid and ask (offer)). Level 1 data refers to the stream of updates to BBO as a consequence of updates to the book state. It is derived data, with the Level 3 market data being the raw data.

Things to consider:

- How would you deal with data which arrive out of order?
- Bonus points for a streaming implementation.
- How might it be possible for a unified batch and streaming implementation to work?

Use the provided CSV files as L3 market data and expected L1 market data to test your program.

Explanation on L3 market data file format:
(https://drive.google.com/open?id=1Kus049iLkqi6q-g3gkrypQIY65LUU3Ox)

| Column Name | Notes |
|---|---|

| | |
|---|---|
| HOST | Timestamp in UTC |
| seq_num | Sequence number of L3 market data. All the L3 market data entries carrying the same sequence number are received from exchange in a single packet. So you should process all of them before output the latest L1 market data |
| is_image | True or Null for a given seq_num<br>Upon receiving such seq_num, you should clear existing book content you have built. Such image is received upon exchange open session. |
| add_orderid | OrderId is unique for side.<br>You can have same orderId for buy and sell, you need to treat them as two different orders. |
| add_side | BUY or SELL |
| add_price | Price of this order<br>For market orders, the value will be Integer.MAX for BUY and negative Integer.MAX for SELL |
| add_qty | Quantity of this order |
| add_position | Position of this order among all the orders queuing for the same price.<br>You can ignore this column for this exercise |
| update_orderid | |
| update_side | OrderId and Side are used together to identify the original Add order it's trying to update |
| update_price | |
| update_qty | |
| update_position | |
| delete_orderid | OrderId and Side are used together to identify the order it's trying to delete |
| delete_side | |

| | |
|---|---|
| trade_orderid | OrderId and Side are used together to identify the order the trade comes from |
| trade_side | |
| trade_qty | |
| trade_price | |

Expected L1 market data output:
(https://drive.google.com/open?id=1gt2BbVAehVo5_XKidSi2eyAqWMiJMBK2)

| Column Name | Notes |
|---|---|
| time | Timestamp of L3 market data which triggers L1 market data update (in UTC) |
| bid_price | Best bid price |
| ask_price | Best ask price |
| bid_size | Total quantity for the best bid price |
| ask_size | Total quantity for the best ask price |
| seq_num | Sequence number of L3 market data which triggers L1 market data update |

Common Questions:

1) How do I handle 1.7976931348623157e+308 in input L3 file?

This is Integer.MAX, you should treat any order carrying price equals Integer.MAX or Integer.MIN as market order, and exclude such **order** from book calculation.

2) Why my L1 output has more records compared with expected output?

The expected L1 result was generated with consideration of other info, which was not exacted out in L3 market data. So your own output will have more L1 updates. This is expected.
You can use time column as the key to match your output to expected file. For any matched row, the rest of columns should carry the same values as the expected file.