sqoop import-all-tables \

-m 12 \

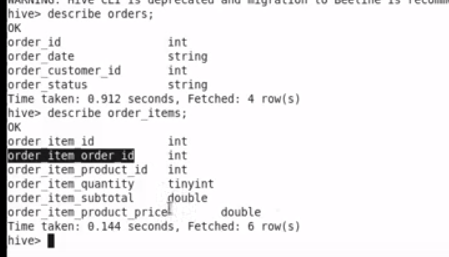
--connect "jdbc:mysql://quickstart.cloudera:3306/retail\_db" \

--username=retail\_dba \

--password=cloudera \

--as-textfile \

--warehouse-dir=/user/hive/warehouse/



* **val** ordersParsedRdd=ordersRdd.map { record => (record.split(",")(0),record) } //\*\*\* You get Mapped RDD Here
* **val** ordersItemsParsedRdd=ordersItemsRdd.map { record => (record.split(",")(1),record) } // You get Mapped RDD here

Getting TotalAmount by Order (Basket) 🡪 We need to use reduceByKey similar to wordcount

To find out maximum value, we have to use reduce and accumulator inside it to check if and else conditions.

Reduceby key also returns tuple, go inside the tuple to get maximum value

/\*(18624,199.99)

(20484,439.95000000000005)

(62544,929.9200000000001)

(20904,649.86)

(57720,579.9200000000001)\*/

/\* In the above case it is tuple , value of first accumlator is 0 by default and value of first value is 199.99 in this case

so you are comparing, 2nd element of tuple with accumulator value i.e. acc.\_2 with val.\_2\*/

in above case acc.\_1, tuple always starts with 1 and accessed by \_. Acc\_1=0 and val\_1=18624

//----------------------- Asmath one more approach to get max value. Use reduce function to get the max value.

**val** totalAmountRdds=ordersItemsRdd.map { record => (record.split(",")(4).toDouble) }

**val** maxTxnAmountOfAllOrders=totalAmountRdds.max();

// If it is more than one value then it is tuple. In above case it is never tuple