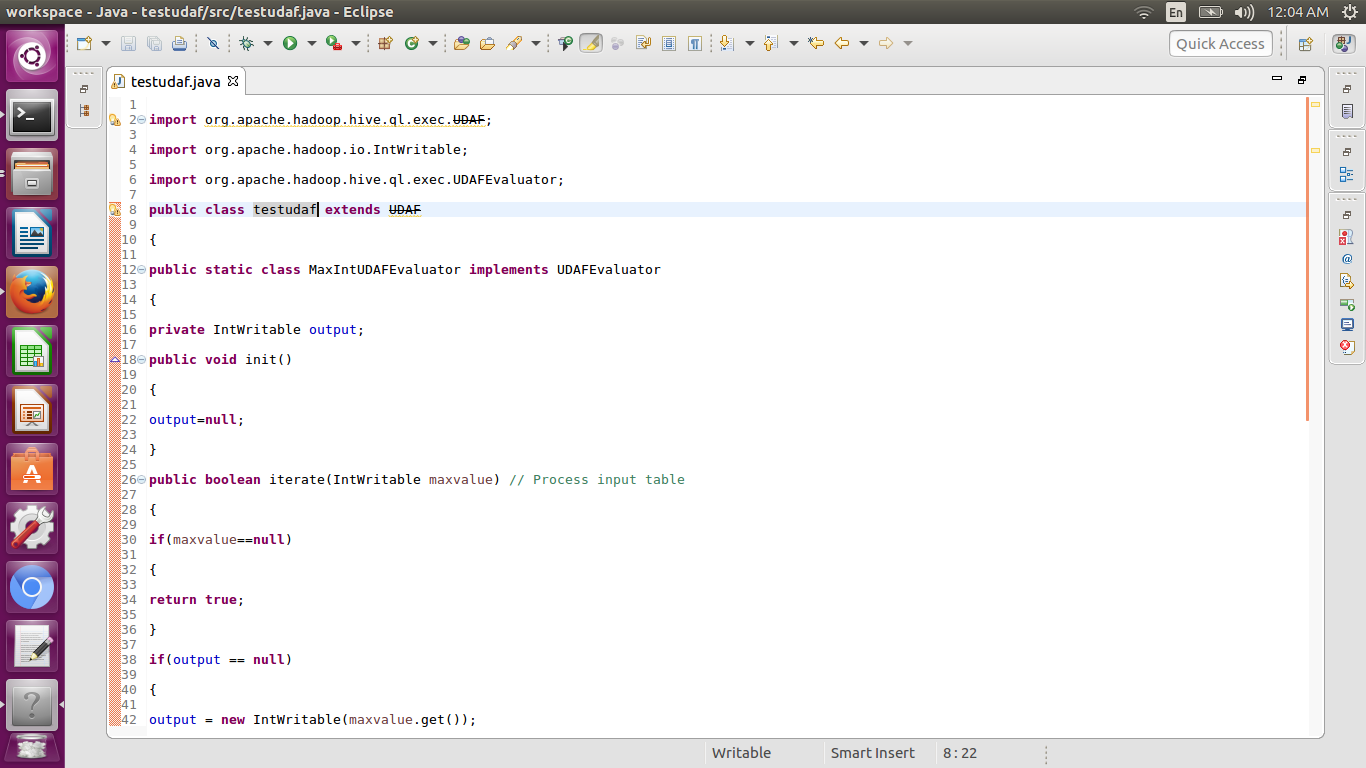
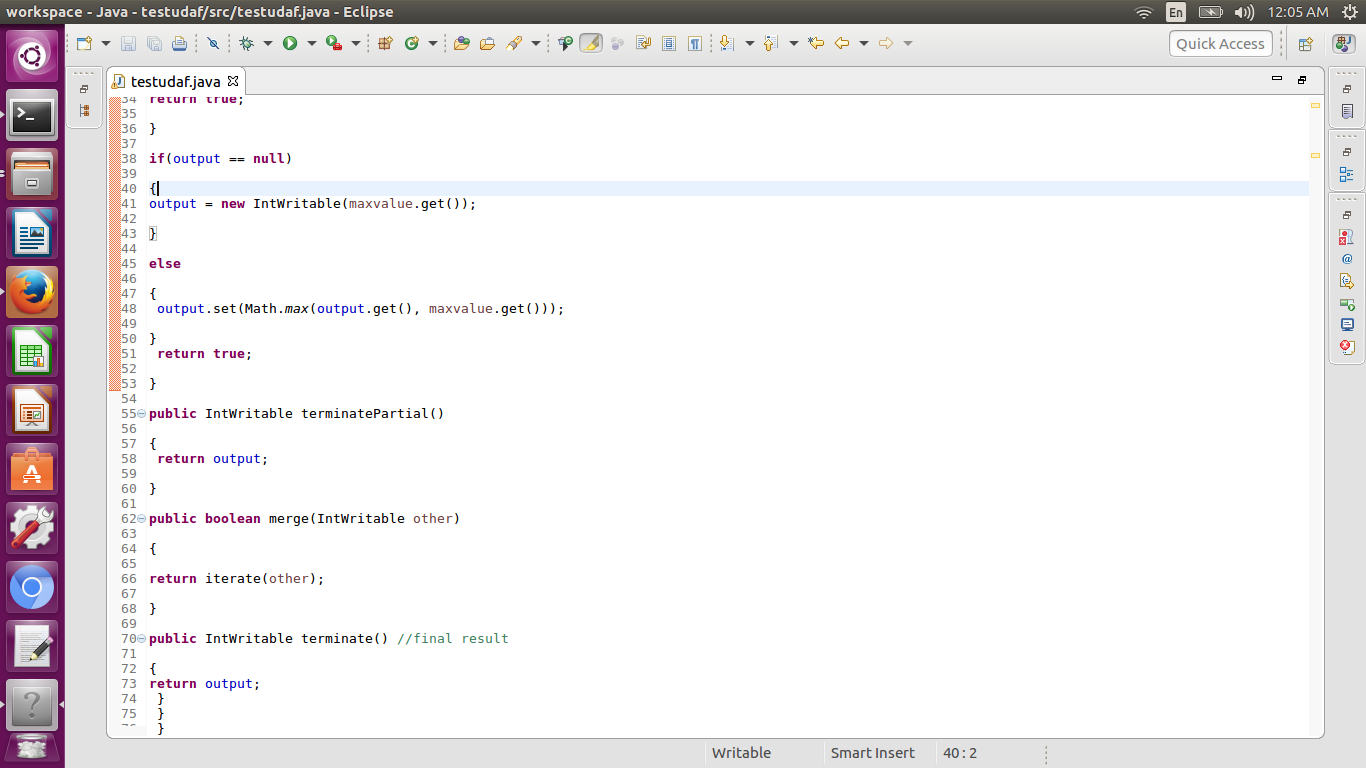
Problem Statement:

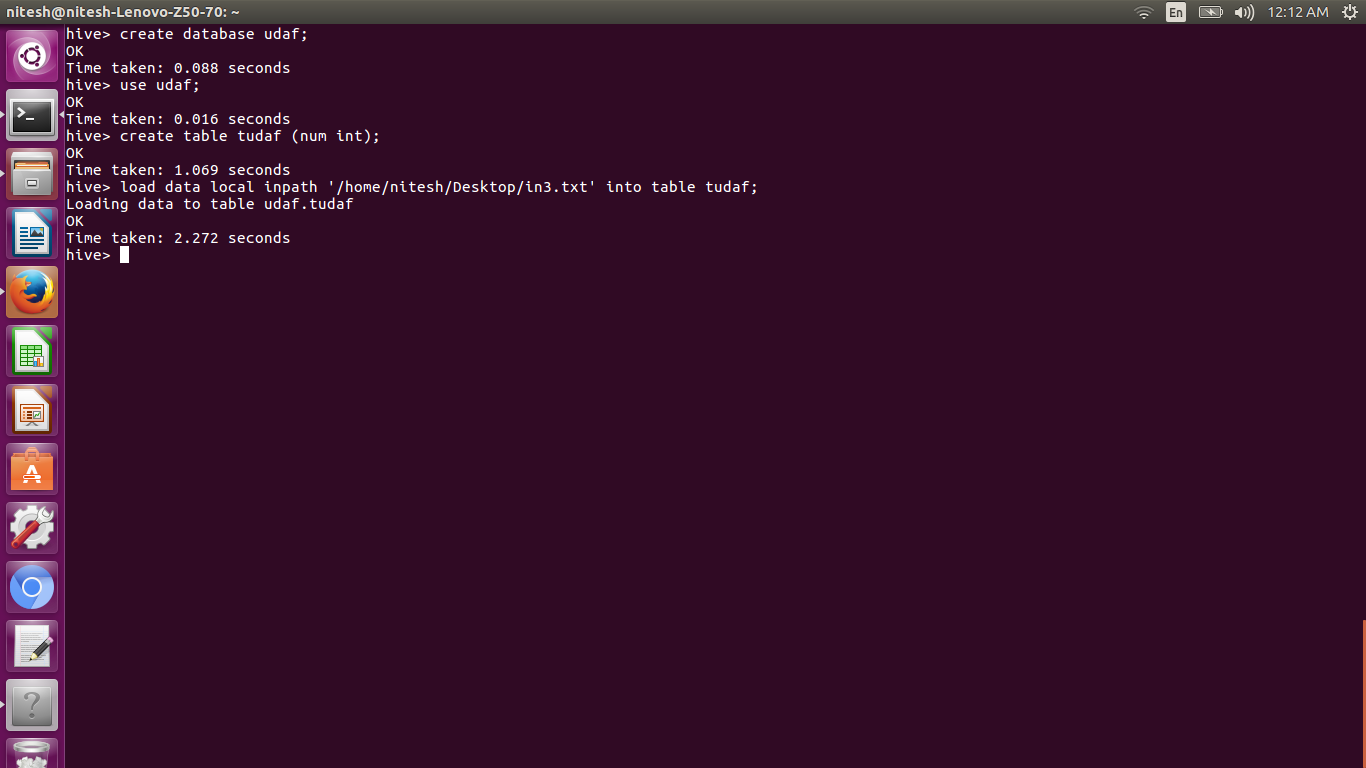
Write a hive UDAF to find the largest integer in a column. Use your own dataset for the above problem statement and attach the steps performed, codes used with the result.

* Create a Java class which extends org.apache.hadoop.hive.ql.exec.hive.UDAF;
* Create an inner class which implements UDAFEvaluator;
* Implement five methods ()

1. init() – The init() method initializes the evaluator and resets its internal state. We are using new Column() in the code below to indicate that no values have been aggregated yet.
2. iterate() – this method is called every time there is a new value to be aggregated. The evaulator should update its internal state with the result of performing the aggregation (we are doing sum – see below). We return true to indicate that the input was valid.
3. terminatePartial() – this method is called when Hive wants a result for the partial aggregation. The method must return an object that encapsulates the state of the aggregation.
4. merge() – this method is called when Hive decides to combine one partial aggregation with another.
5. terminate() – this method is called when the final result of the aggregation is needed.

* Compile and package the JAR
* CREATE TEMPORARY FUNCTION in hive CLI 



UDAF receives different input at different MapReduce stages. iterate reads a line from our table (or an input record as per the InputFormat of our table to be more precise), and outputs something for aggregation in some other format. partialAggregation combines a number of these elements into an aggregated form of the same format. And then the final reducer takes this input and outputs a final result a format of which may be different from format in which the data was received. 

1. Create a new table  and load the input dataset

Here we have created a table tudf

1. We have loaded the input dataset
2. We have Add the Jar file in hive with complete path
3. Create temporary function as shown below
4. Use the select statement to find the largest number from the table

