Problem Statement:

Write a hive UDF that implements functionality of string concat\_ws(string SEP, array). This UDF will accept two arguments, one string and one array of string. It will return a single string where all the elements of the array are separated by the SEP.

User Defined Functions(UDFs) provides us a way to:

Extend the functionality of Hive by writing functions that can be evaluated in Hive QL.

Custom serializers and/or deserializer (“serdes”), which provide a way of either deserializing a custom file format stored on HDFS to a POJO (plain old Java object), or serializing a POJO to a custom file format (or both).

Custom mappers/reducers, which allow you to add a custom map or reduce steps into your Hive query. These map/reduce steps can be written in any programming language, and not just in Java.

There are two different interfaces you can use for writing UDFs for Apache Hive. One is really simple, the other… not so much.

The simple API (org.apache.hadoop.hive.ql.exec.UDF) can be used so long as your function reads and returns primitive types. By this I mean basic Hadoop & Hive writable types - Text, IntWritable, LongWritable, DoubleWritable, etc.

However, if you plan on writing a UDF that can manipulate embedded data structures, such as Map, List, and Set, then you’re stuck using org.apache.hadoop.hive.ql.udf.generic.GenericUDF, which is a little more involved.

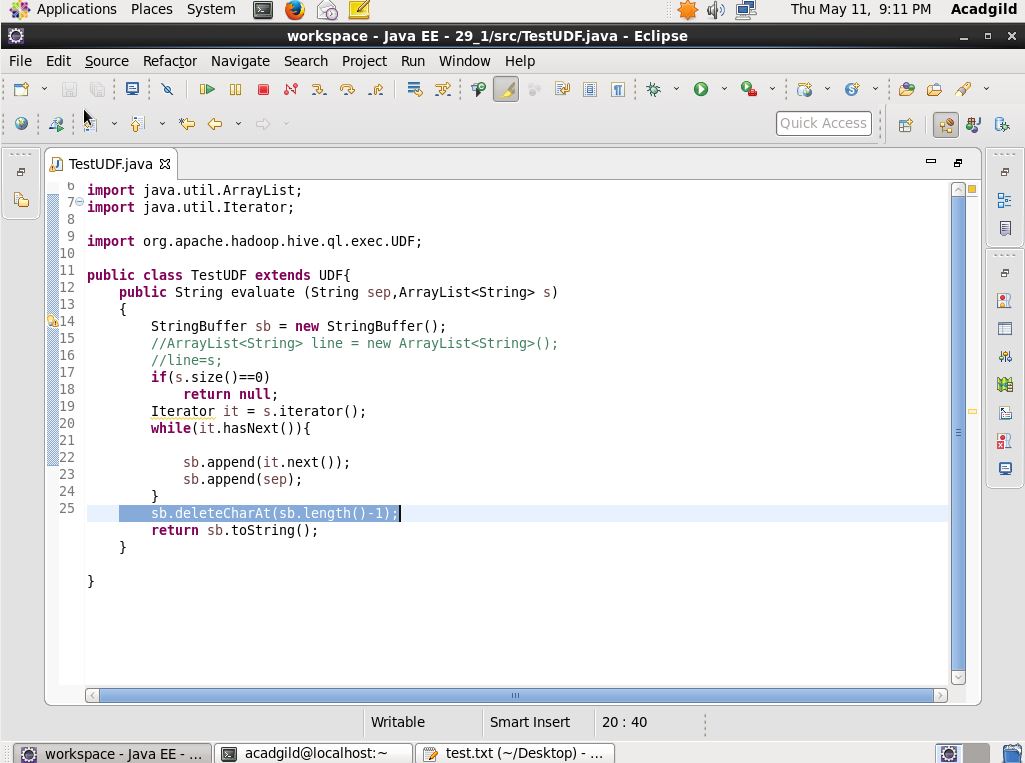
* Simple API - org.apache.hadoop.hive.ql.exec.UDF
* Complex API - org.apache.hadoop.hive.ql.udf.generic.GenericUDF

The Simple API Building a UDF with the simpler UDF API involves little more than writing a class with one function (evaluate).

UDF:

Here we have written a code for udf in java

We have override an evaluate method as it is mandatory. Here we are passing a separator as well as the collection of strings to be concatenated.



Terminal:

Here we have created a table and the data is loaded from the local system. Row fields are terminated by \t and the collections are terminated by ‘,’

We have used a created function in the select clause and the output is shown in the below screen shot.



We have added this jar in the hive using add command. The jar is created by the eclipse where we have written the code. Also we have created a temporary function named func and it is used in the select statement to get the concat function to be implemented with separator as shown above.

