**ASSIGNMENT 29.3**

**Explain Brief of the following in brief**

**● Hive UDF**

**● Hive UDAF**

**● Hive UDTF**

**● Thrift server**

**Hive udf:**

UDF - user-defined function

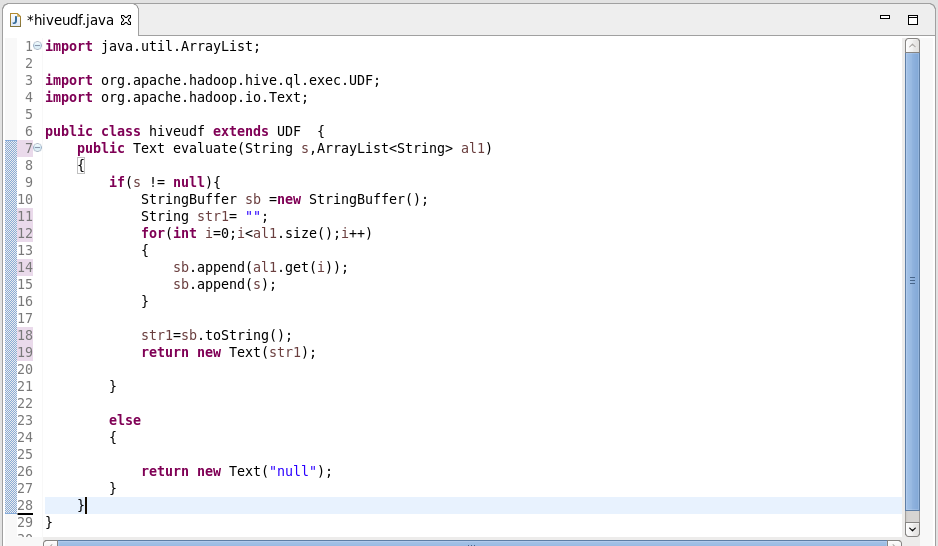
It have a single input value and single output value.

For using those in case of query, one can use it only once for each row in a result set.

*User Defined Functions(UDFs) :*

* 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](https://acadgild.com/blog/beginners-guide-for-hdfs/).
* 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.
* Since the Hadoop framework is written in Java, naturally most of the Hadoop developers prefer Java to write the UDFs.
* However, Apache has also made it easy for non-Java developers to be able to work on Hadoop; this is done using the Hadoop Streaming Interface!

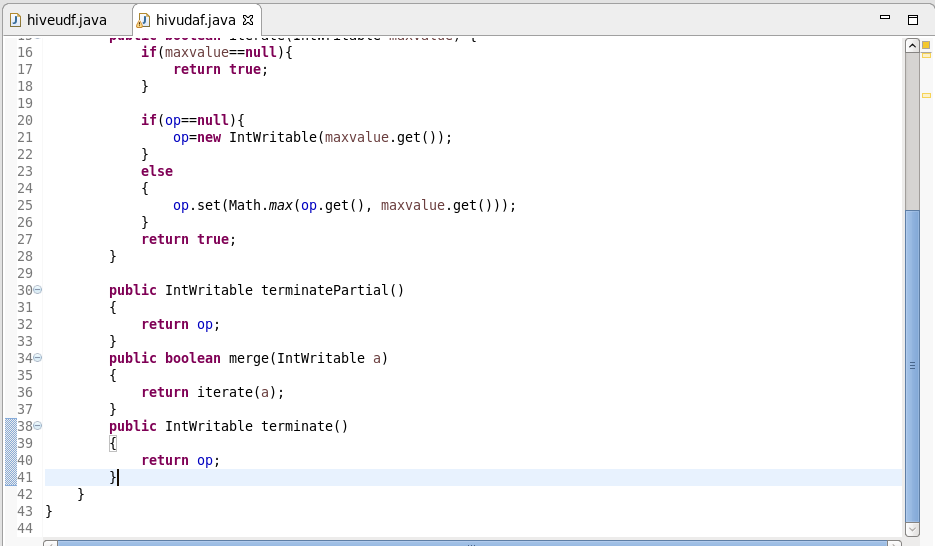
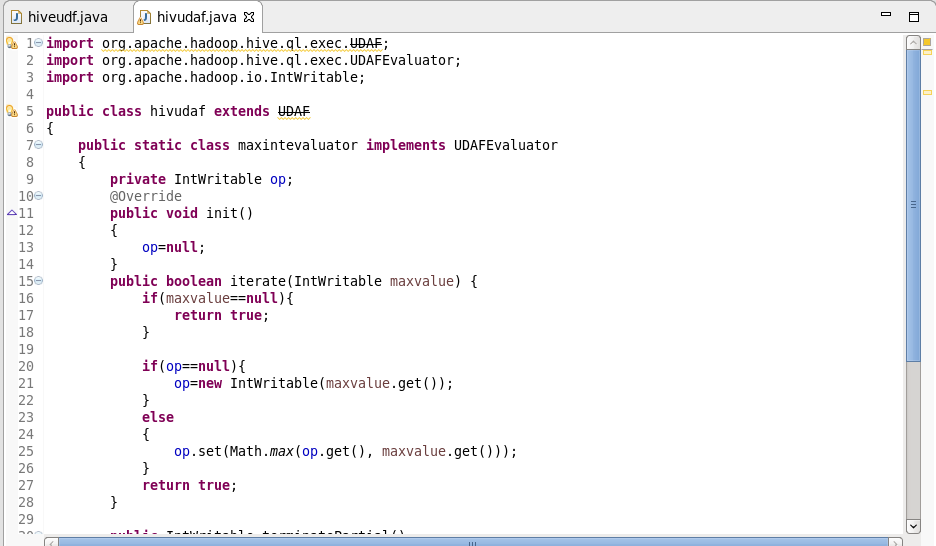
**Example:** I’ve written a java code to concat a string . This UDF will accept two arguments, one string and one array of string. It will return a single string where all the elements of the array are separated by the separator.



**Hive udaf:**

* It is a User-Defined Aggregation Functions (UDAFs) which gives a exceptional way to integrate an advanced data-processing into Hive.
* Aggregate functions helps to do the for a set of calculation to some values and in return will provide with single value.
* Aggregate function is difficult to write than UDF.
* Aggregated Values in chunks ,so the implementation should be capable of combining partial aggregations into final result.

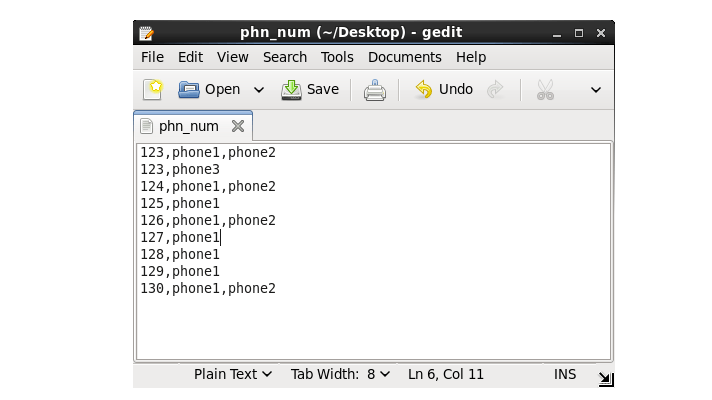
**Sample udaf:**

****

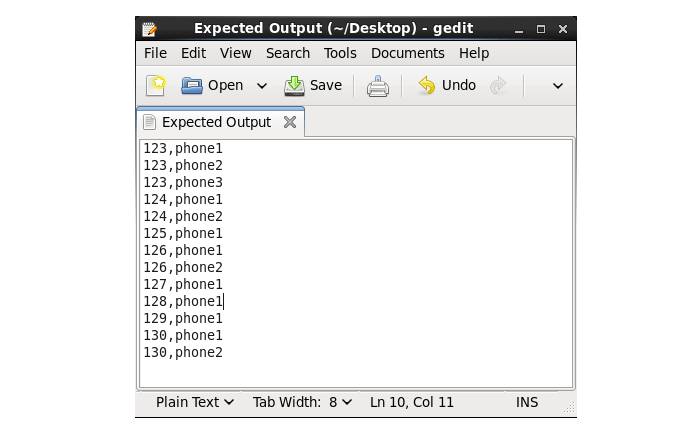
**UDTF:**

* UDTF - User Defined Table Generating Function
* Its operates will be on a single row
* And the output will be of multiple rows in a table.
* It will be opposite to that of udaf.
* And will be much similar to the explode function.

**Before using udtf:**

****

**After using udtf:**

****

**Thrift server:**

* Apache Thrift software framework, for scalable cross-language services development, will combine software stack with code generation engine
* In order to build services which work efficiently between C++, Java, Python, PHP, Ruby, Erlang, Perl, Haskell, C#, Cocoa, JavaScript, Node.js, Smalltalk, OCaml and Delphi and other languages.
* RPC framework is for building the cross-platform services.
* Stack will contain 4 layers:

1. Server
2. Transport
3. Protocol
4. Processor.

* When a hive tables or database is queried , in background automatically your requests is transferred between hive service and hive server
* when you want to create your own service or project you can use thrift protocols which will help you in creating layers, think this as you are creating your user defined functions using libraries, so in that case libraries will be thrift.