**ASSIGNMENT 29.3**

**Explain Brief of the following in brief**

**● Hive UDF**

**● Hive UDAF**

**● Hive UDTF**

**● Thrift server**

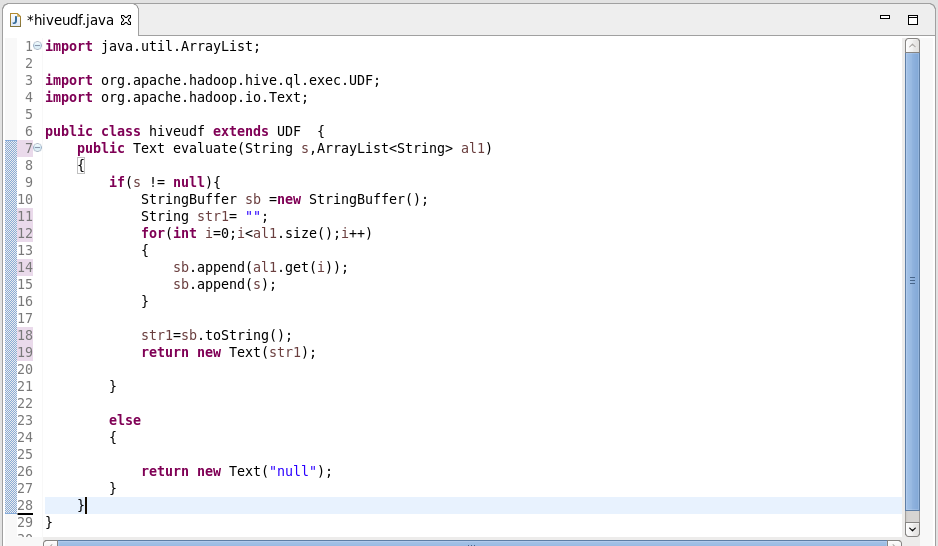
**Hive udf:**

UDF is a user-defined function that takes a single input value and produces a single output value. When used in a query, we can call it once for each row in the result set.

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](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!

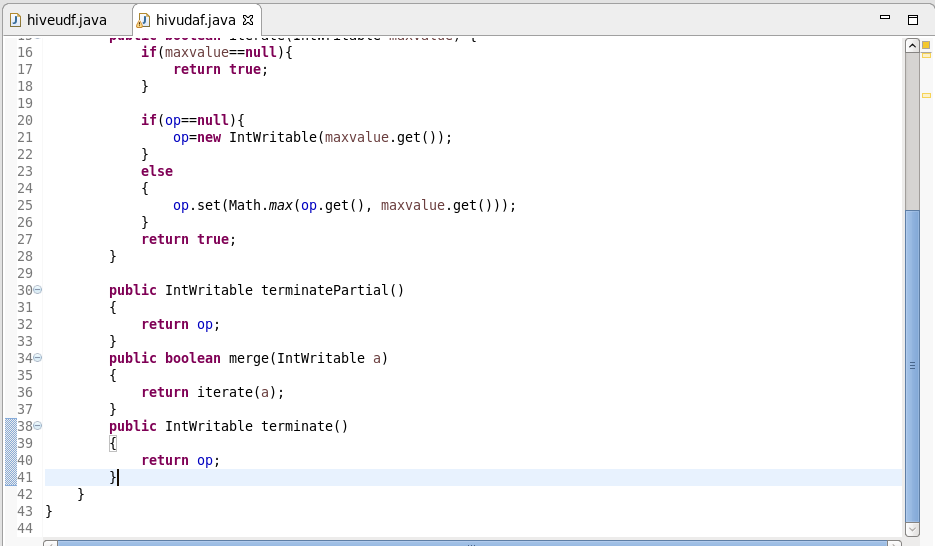
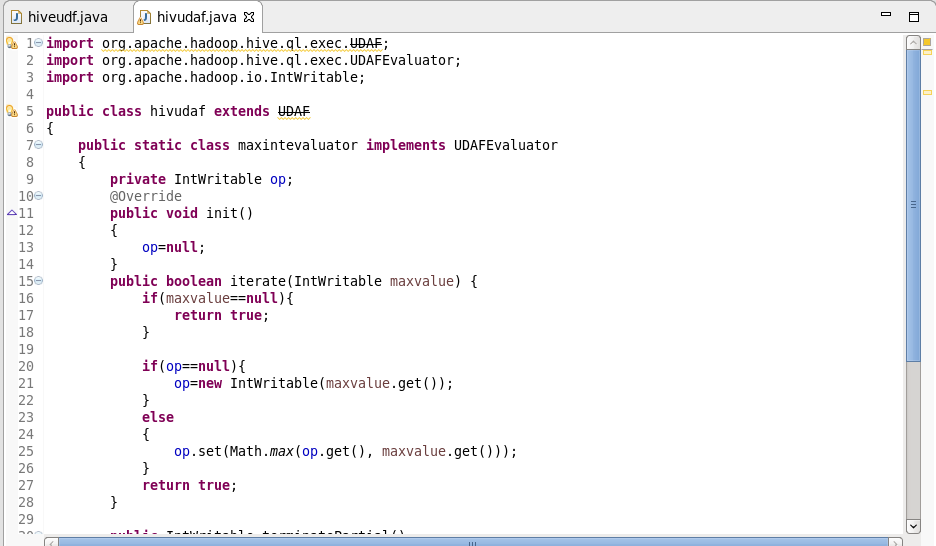
**Example:** I’ve written a java code to concat a string . This UDF will accept two arguments, one string and one array of string. It will return a single string where all the elements of the array are separated by the separator.



**Hive udaf:**

* User-Defined Aggregation Functions (UDAFs) are an exceptional way to integrate advanced data-processing into Hive.
* Aggregate functions perform a calculation on a set of values and return a single value.
* An aggregate function is more difficult to write than a regular UDF.
* Values are aggregated in chunks (potentially across many tasks), so the implementation has to be capable of combining partial aggregations into a final result.

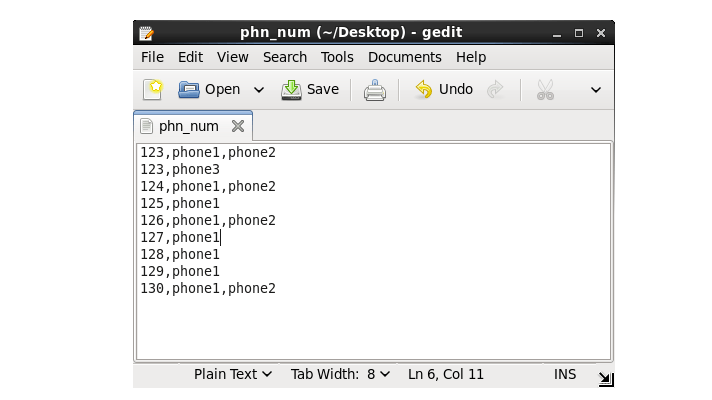
**Sample udaf:**

****

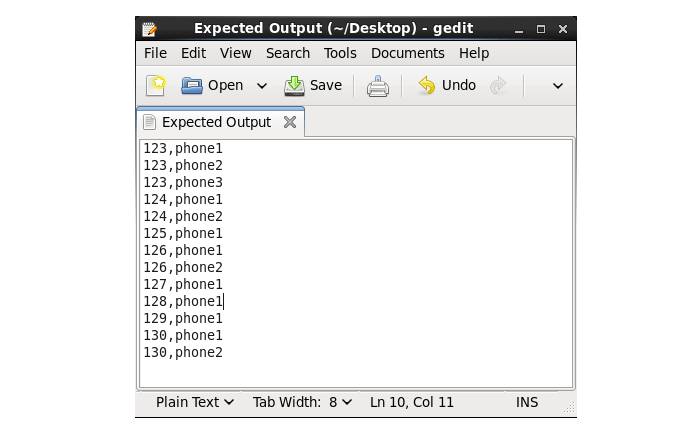
**UDTF:**

* UDTF is a User Defined Table Generating Function that operates on a single row and produces multiple rows a table as output.
* This is exactly opposite to the udaf.
* More or less similar to explode function

**Before using udtf:**

****

**After using udtf:**

****

**Thrift server:**

* The Apache Thrift software framework, for scalable cross-language services development, combines a software stack with a code generation engine to build services that work efficiently and seamlessly between C++, Java, Python, PHP, Ruby, Erlang, Perl, Haskell, C#, Cocoa, JavaScript, Node.js, Smalltalk, OCaml and Delphi and other languages.
* Thrift is an RPC framework for building cross-platform services. Its stack consists of 4 layers: Server, Transport, Protocol, and Processor.
* When you query any hive tables or database, 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.