Explain Brief of the following in brief

● Hive UDF

● Hive UDAF

● Hive UDTF

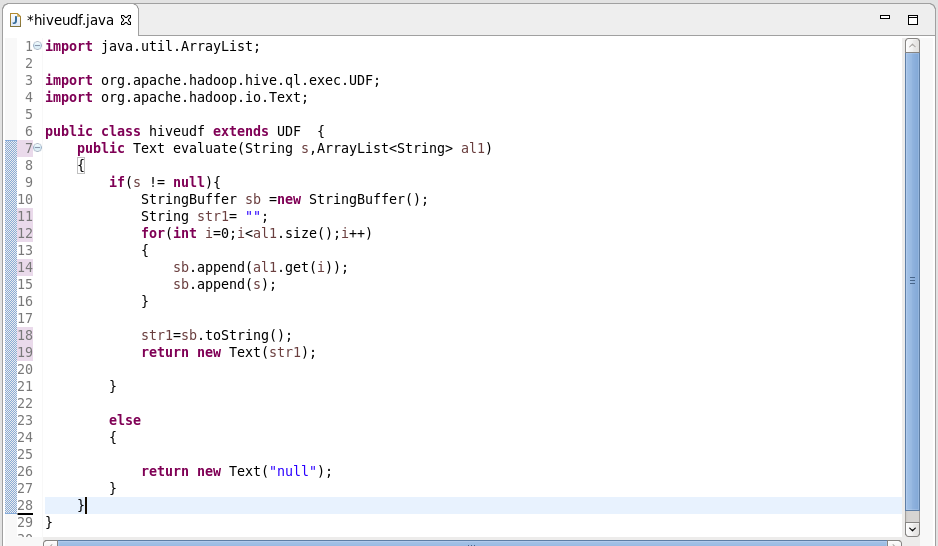
● Thrift server

**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:

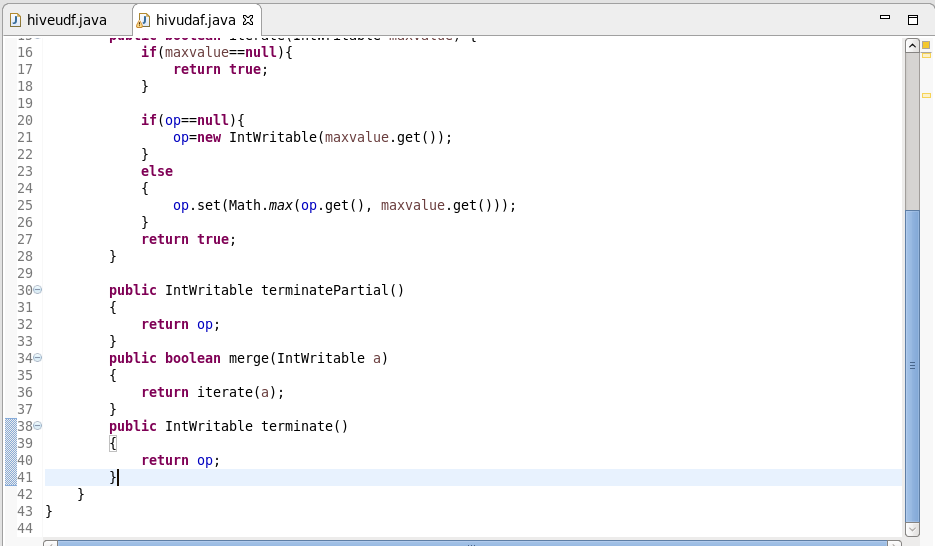
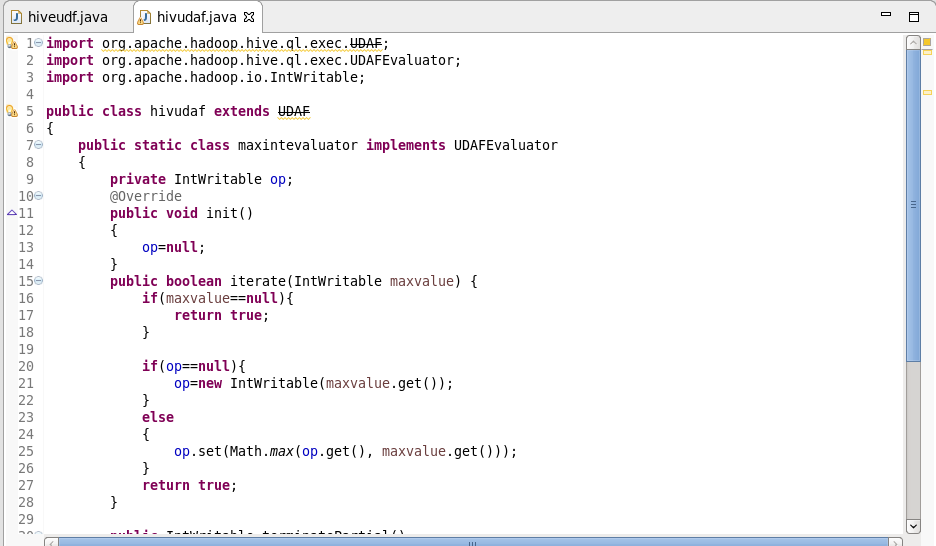
* Extend the functionality of Hive by writing functions that can be evaluated in Hive QL.
* Custom serializers and/or deserializer (“serdes”), which provide a way of either deserializing a custom file format stored on [HDFS](https://acadgild.com/blog/beginners-guide-for-hdfs/).
* Custom mappers/reducers, 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:

**Hive udaf:**

* User defined aggregate functions works on more than one row and gives single row as output. **e.g Hive built in MAX() or COUNT() functions**.
* here the relation is many to one. Lets say you have a table with students name, id and total marks, so here if I have 10 rows in the table .
* if I have to find student who got maximum number then our query need to check each 10 row to find the maximum but ultimately we get only one output which is the maximum.
* Hope this justifies the many to one relationship.
* Hive allows us to define our own UDAFs.

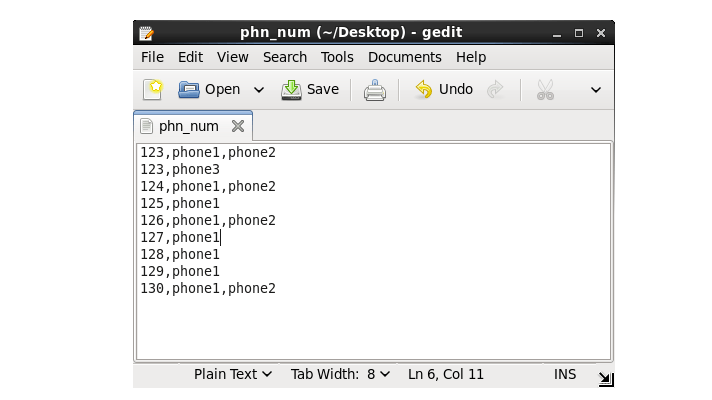
**Sample udaf:**

****

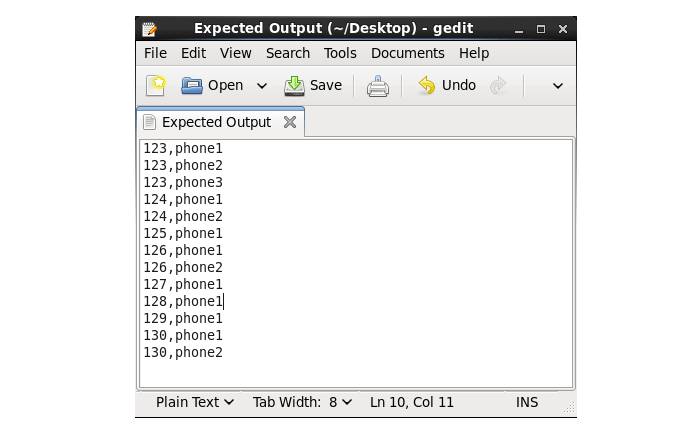
**UDTF:**

* User defined tabular function works on one row as input and returns multiple rows as output. So here the relation in one to many. **e.g Hive built in EXPLODE() function.**
* Now lets take an array column USER\_IDS as ARRAY10,12,5,45> then SELECT EXPLODE(USER\_IDS) as ID FROM T\_USER. will give 10,12,5,45 as four different rows in output.
* UDTF can be used to split a column into multiple column as well which we will look in below example. Here alias "AS" clause is mandatory .

**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