● **Hive udf:**

UDF - user-defined function

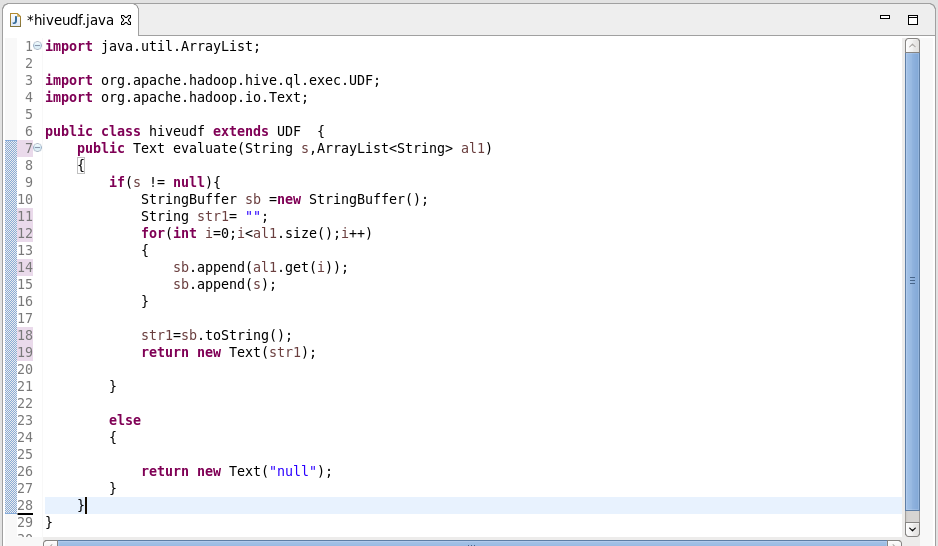
It have a single input value and single output value.

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

User Defined Functions(UDFs) :

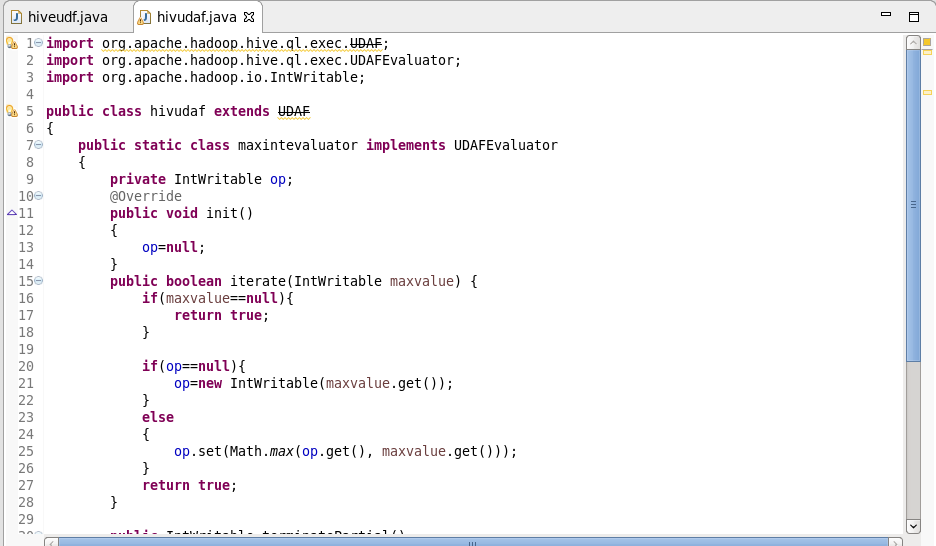
* Extend the functionality of Hive by writing functions that can be evaluated in Hive QL.
* Custom serializers and/or deserializer (“serdes”), which provide a way of either deserializing a custom file format stored on [HDFS](https://acadgild.com/blog/beginners-guide-for-hdfs/).
* Custom mappers/reducers,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

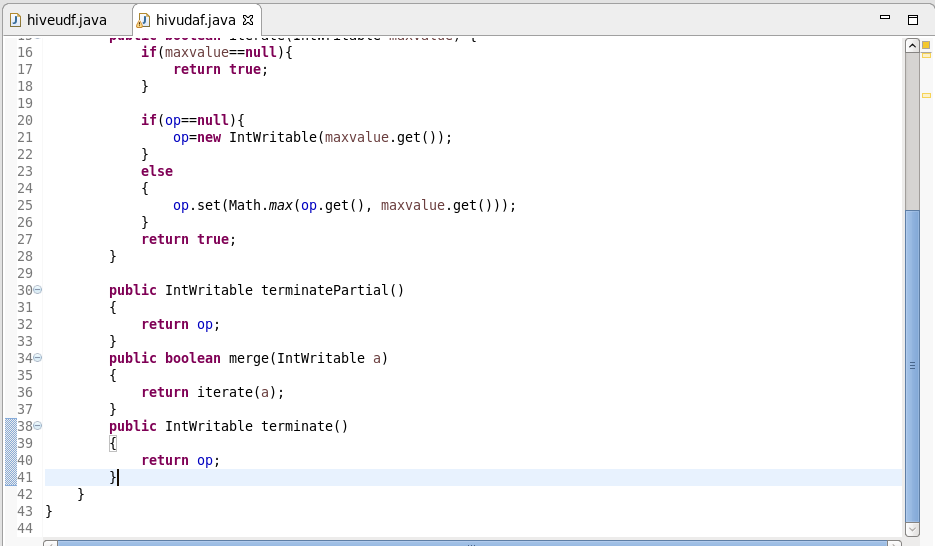
**Example:** I’ve written a java code to concat a string . This UDF will accept two arguments, one string and one array of string. It will return a single string where all the elements of the array are separated by the separator.



● **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.





**● 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.

**● Thrift server**

[Apache Hive](https://acadgild.com/big-data/apache-spark-training-certification) is a data warehouse software that facilitates reading, writing and managing large data sets residing in distributed storage using SQL.

Let’s consider a scenario, where the user is looking forward to performing an operation on Hive server, and the [Hadoop cluster](https://acadgild.com/big-data/big-data-development-training-certification) or Hive software setup is not installed in his/her system. The solution for the above scenario is that the user can write codes in other languages and access Hive server using Apache Thrift interface.

Apache Thrift is a software framework for scalable cross-language services development, which combines a software stack with a code generation engine to build services that work efficiently and seamlessly between C++, Java, Python, PHP, Ruby, Perl, C#, JavaScript, Node.js and other languages.

Thrift can be used when developing a web service that uses a service developed in one language access that is in another language.

HiveServer is a service that allows a remote client to submit requests to Hive, using a variety of programming languages, and retrieve results. It is built on Apache Thrift, therefore it is sometimes called as the Thrift server.

In the context of Hive, Java language can be used to access Hive server. The Thrift interface acts as a bridge, allowing other languages to access Hive, using a Thrift server that interacts with the Java client.