



Hive UDFs





HIVE USER DEFINED FUNCTIONS



Objectives

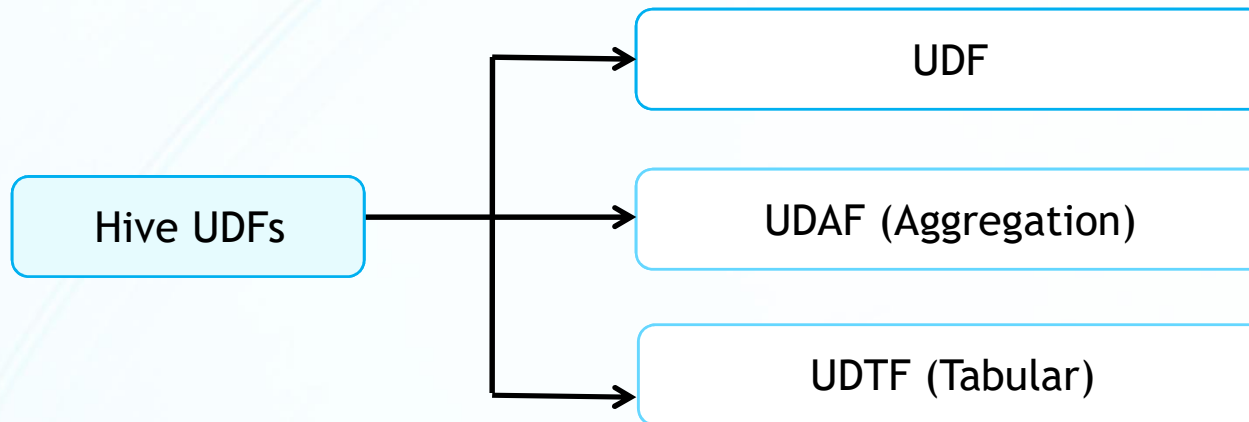
In this module, we are going to look at the following topics:

- ✓ Hive User Defined Functions
- ✓ UDF
- ✓ UDAF
- ✓ UDTF



Hive UDFs

- UDFs in hive are used to plug in our own logic into hive when we are not able to get the desired result from hive's built in functions.
- We can invoke the UDFs from hive query.



Creating Regular UDF

- **Step 1:** Extend a base Class UDF to write our business logic in Java.
- **Step 2:** Overload a method called *evaluate()* inside our class.
- **Step 3:** Export the JAR files to HDFS where hive is running.
- **Step 4:** Add the exported JAR file to hive classpath using ADD JAR EXPORTED_FILE_NAME.jar
 - **Alternate:** Add exported JAR file in bashrc file using command "**nano ~/.bashrc**" as `HIVE_AUX_JAR_PATH = '/hive-path/hive-udf.jar'`.
- **Step 5:** Create a temporary function for the exported jar file.



Regular UDF

- UDFs works on a single row in a table and produces a single row as output.
- Its one to one relationship between input and output of a function.
- e.g. Hive built in TRIM() function.



UDAF

- User defined aggregate functions works on more than one row and gives single row as output. Here the relation is many to one.
- e.g. Hive built in MAX() or COUNT() functions.



Creating UDAF

Step 1: Extend a base Class UDAF to write our business logic in Java.

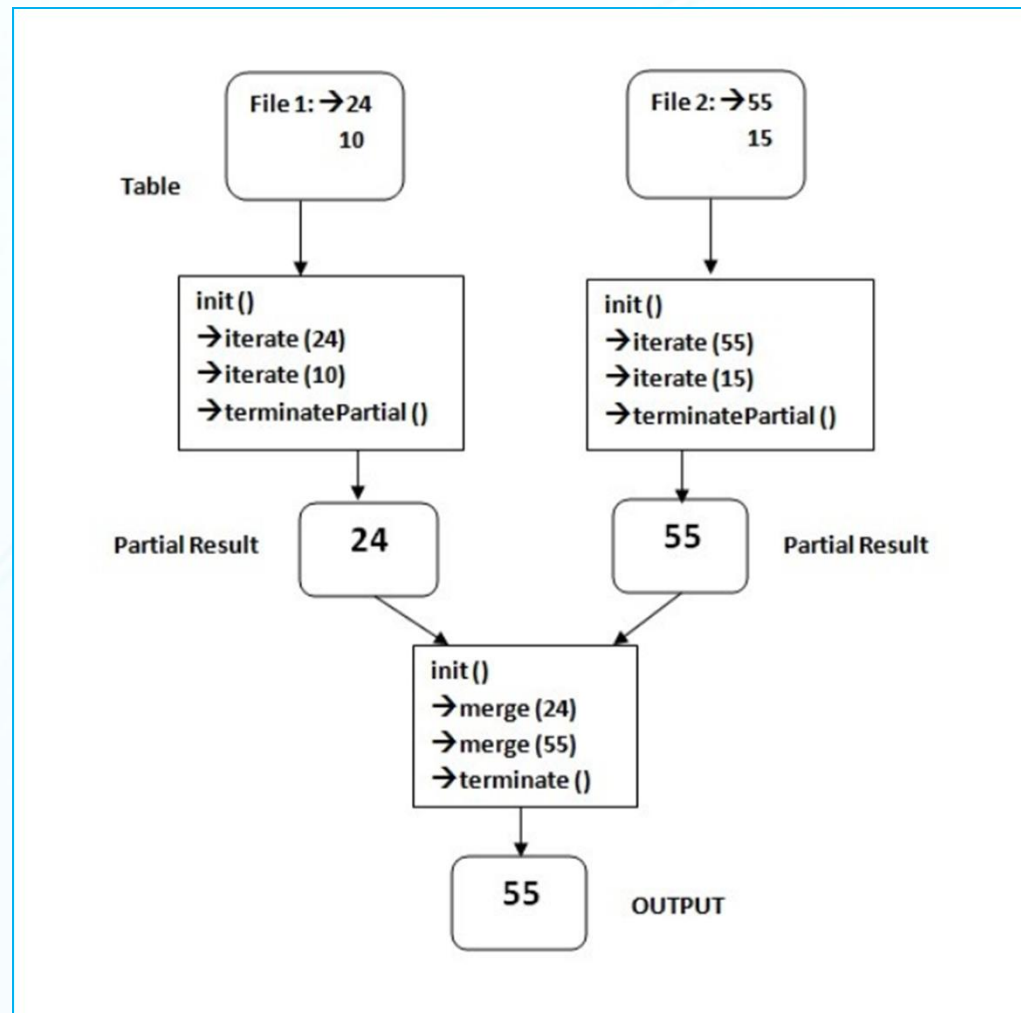
Step 2: We need to overwrite five methods

- *init()*
- *iterate()*
- *terminatePartial()*
- *merge()*
- *terminate()*

Step 3 to Step 5: It is same as explained in UDF example.



UDAF Process Flow



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.



Creating UDTF

Step 1: Extend a base Class GenericUDTF to write our logic in Java.

Step 2: Override 3 methods:

- initialize()
- process()
- close()

Step 3 to Step 5: It is same as explained in UDF example.



THANK YOU

