**Problem Statement**:

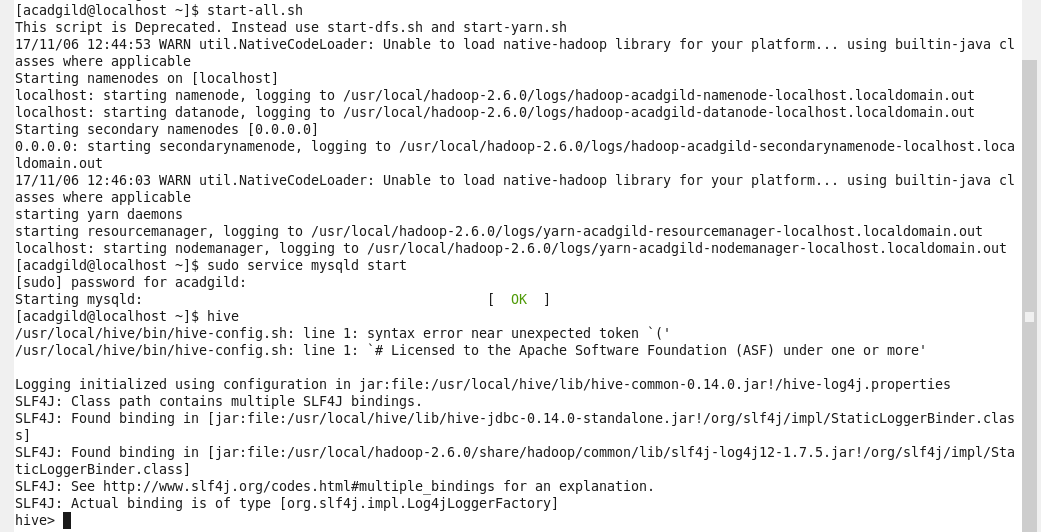
Write a hive UDF that implements functionality of string concat\_ws(string SEP, array<string>).

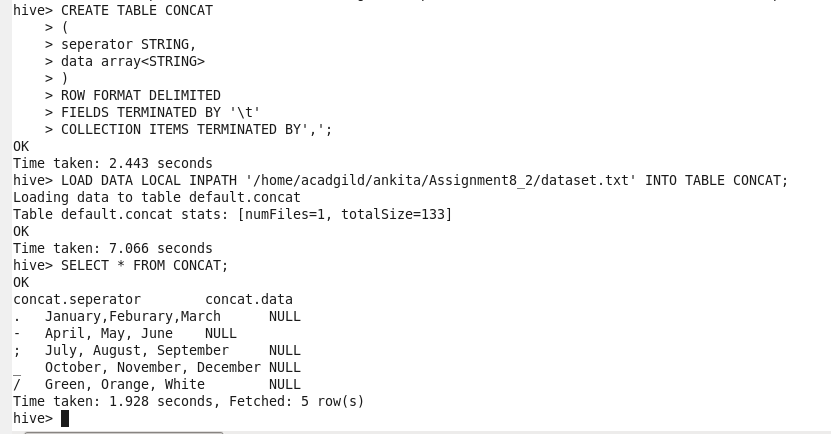
This UDF will accept two arguments, one string and one array of string.

It will return a single string where all the elements of the array are separated by the SEP.

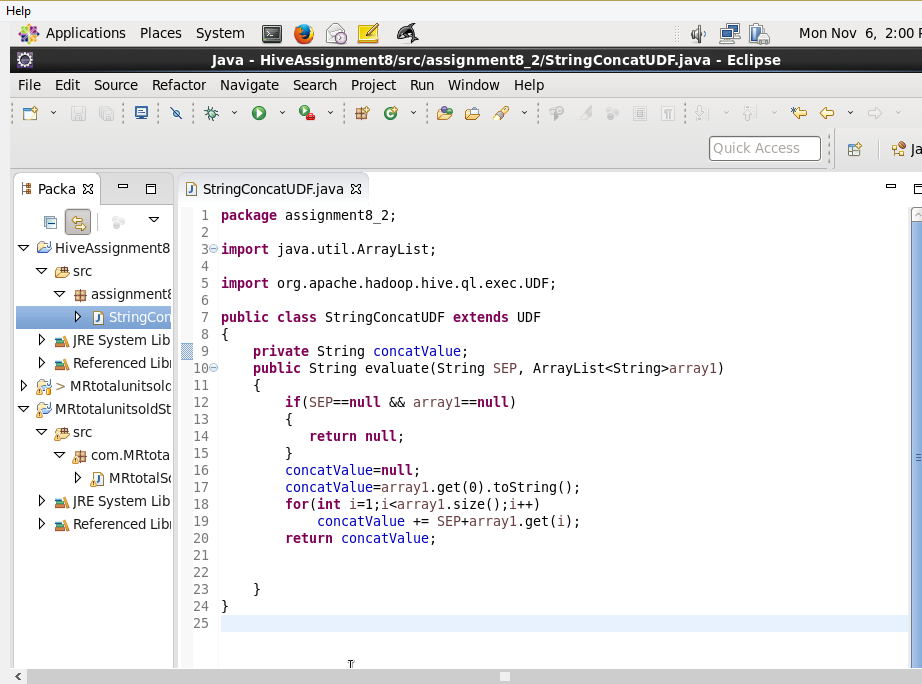
**Solution:**

1. Starting all the daemons and the mysql service (Required for Hive).
2. Starting Hive.
3. Setting print header to true. Prints Column Headers.
4. Creating table ‘CONCAT’ with fields for **separator** (Type: String) and **data** (Type: Array<String>).
5. Loading the dataset into the table CONCAT from the Local FS.
6. Checking to see if the data has been loaded properly by using the **SELECT** command.

****

****

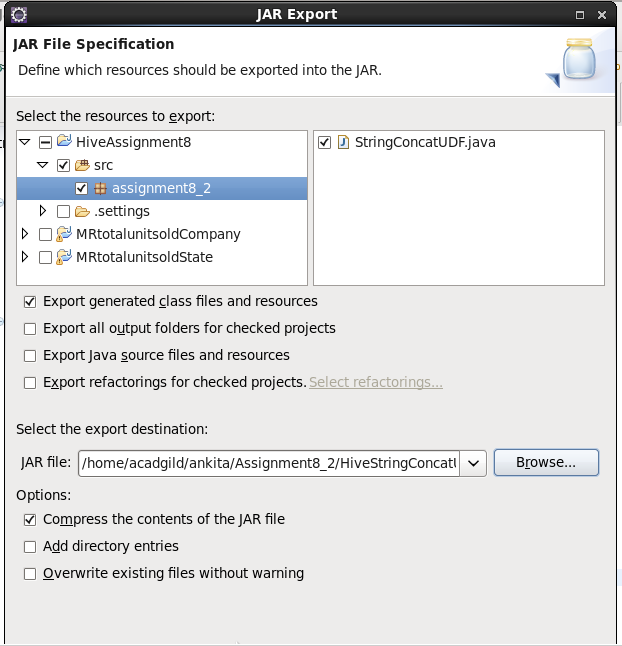
**HIVE UDF:**

****

The Java Program above implements the functionality of string concat function **concat\_ws (string SEP, array<string**>) as a UDF using the following steps:

* Import the jars: **hadoop-common-2.6.0.jar** and the **hive-exec-0.14.0.jar** for UDF and Text support
* Give main class (**StringConcatUDF**) an extension: UDF
* Create a string variable (**concatValue**) for the single-string result
* Override the function **evaluate** and take two elements as parameters: A **String(SEP)** variable to separate the array elements with and the an **ArrayList(array1)** of type:String
* In the function, check if the both strings not null, if not then proceed with the evaluation or return null
* Clear the result string **concatValue** and add the first element of the array to it.
* In a recursive loop, add the Seperator string **SEP** and the next array element **strarr** to the **concatValue**
* Return the result string **concatValue** as the output.

**EXPORTING THE CODE:**

****

In the above image, I have selected the java program named ‘StringConcatUDF’ in the package ‘assignment8\_2’ and exported the code as a .jar file to the a folder ‘Assignment8\_2’ in the path ‘/home/acadgild/ankita/Assignment8\_2/’ and named the jar file as ‘HiveStringConcatUDF.jar’.

**Run Program In Hive -**

To run the program in hive, the below steps are implemeted:

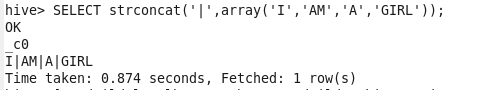
* Adding the Jar file that was imported from the Eclipse IDE to the Hive Jar Database (FileSystem)
* Creating a temporary function(**strconcat**) to run as the Java Program(**StringConcatUDF**) in package ‘**assignment8\_2’**
* Executing the UDF in hive using the statement:

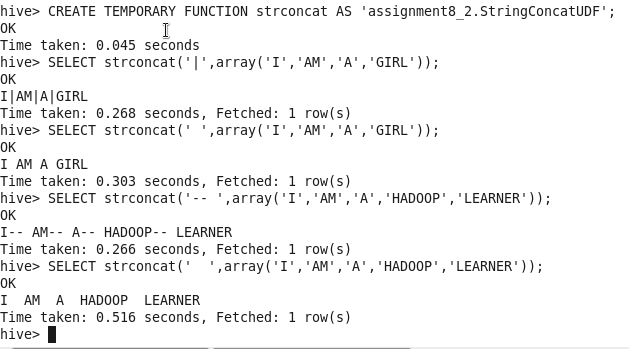
SELECT temp\_func\_name(seperator\_name,array\_name) from table\_name;

The 1st executed statement is with the table CONCAT.

Then execute statement by giving the parameters directly.

****

****

****