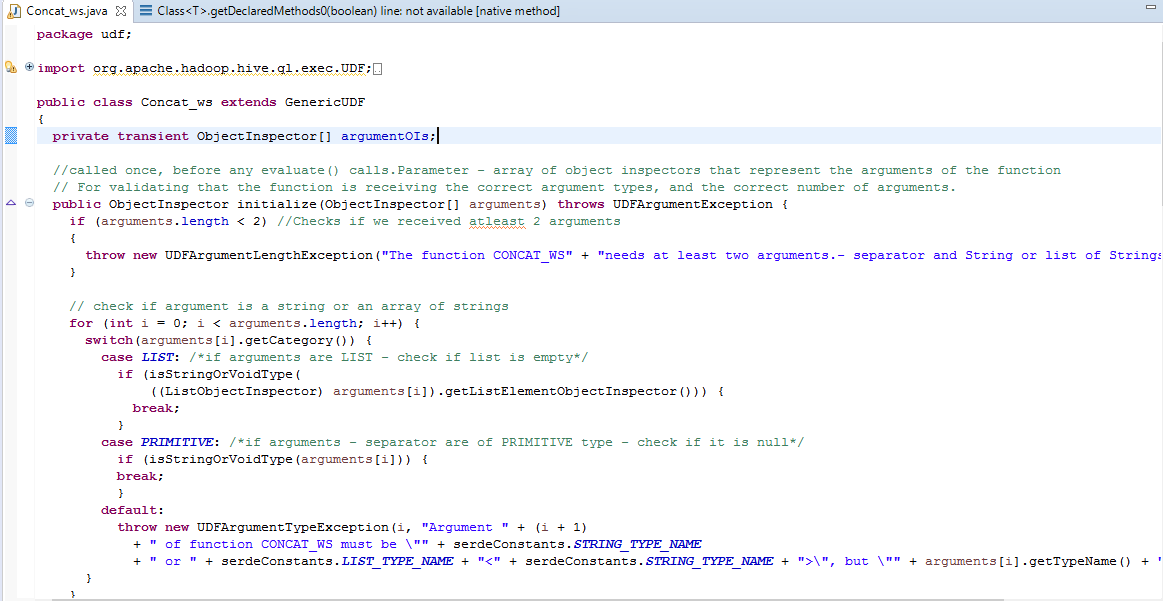
**Assignment3.5**

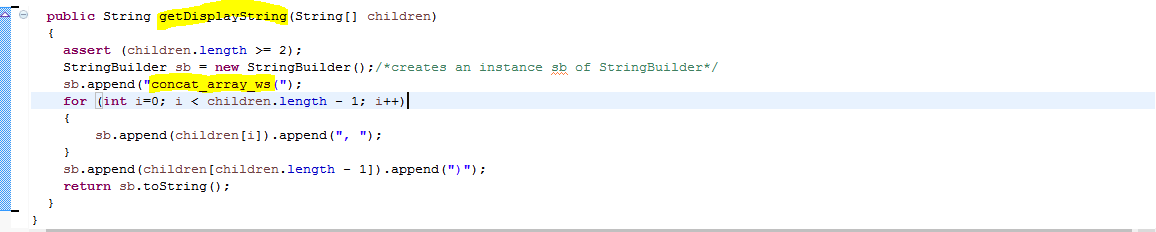
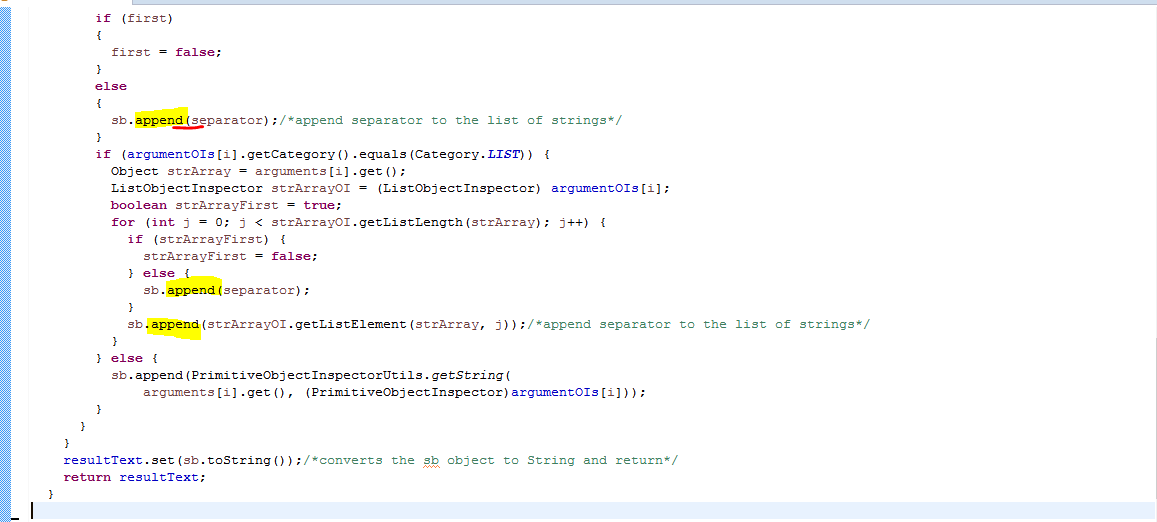
**Problem Statement:**

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.

**Java program:**



**Code explanation:**

The org.apache.hadoop.hive.ql.udf.generic.GenericUDF - API provides a way to write code for objects that are not writable types, for example - struct, map and array types.

An object inspector provides a consistent interface for underlying object types so that different object implementations can all be accessed in a consistent way from within hive. For example, we can implement a struct as a Map so long as we provided a corresponding object inspector.

**For implementing complex udf - From API we have to implement three methods :**

// overridden method - takes the actual arguments and returns the result

abstract Object **evaluate**(GenericUDF.DeferredObject[] arguments);

// To specify String representation of the function.

abstract String **getDisplayString**(String[] children);

// called once, before calling evaluate().This function receives an array of object inspectors that represent the arguments of the function for validating that the function is receiving the correct argument types, and the correct number of arguments.

abstractObjectInspector**initialize**(ObjectInspector[] arguments);

**initialize:**

* The UDF is initialized using a default constructor.udf.initialize() is called with the array of object instructors for the udf arguments .
* In this method we can check that we have the right number of arguments and that they are the

right types(String,array<strings>) if not throws error.

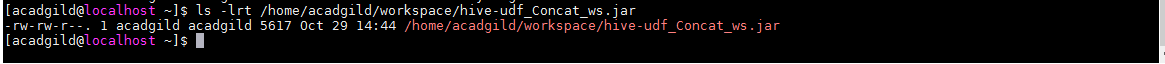
* If the arguments were proper returnsancorresponding object inspector so that Hive can read the result of the function.

**getDisplayString():**

* If there are more that 2 arguments specify String representation of the function which need to be created in hive and returns String.
* sb.append("concat\_array\_ws(") - representation of new function

**evaluate():**

* Overridden method which is called for each row in your query with the arguments provided.This function takes the command line argument and returns the result.
* We store the object instructors for use in evaluate() and return an object inspector.
* In this program in evaluate() assign the 1st argument as separator and rest as Stringbulilder object as it is array of strings. And appends the separator to the stringbuilder instance.



**Registering the exported JAR:**

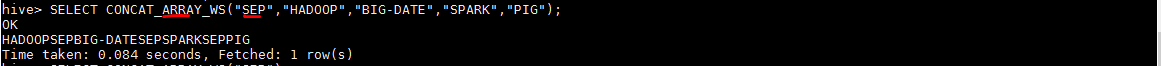
Registered the exported jar in hive by adding the jar.

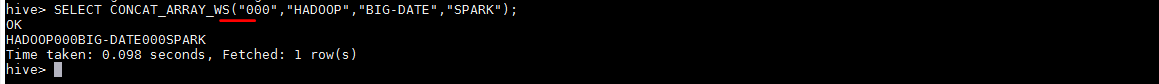


**Creating the temporary function – concat\_array\_ws:**

Created function **“concat\_array\_ws”** with the newly developed class **“udf.Concat\_ws”.**

**Testing the new function - concat\_array\_ws:**

Tested the newly developed function “CONCAT\_ARRAY\_WS” with two arguments “SEP” – separator and the array of strings. And function returned a string – combination of strings and separator.



Tested the newly developed function “CONCAT\_ARRAY\_WS” with 1 argument “SEP” – separator and function failed with error that there should be 2 arguments.(As per design)