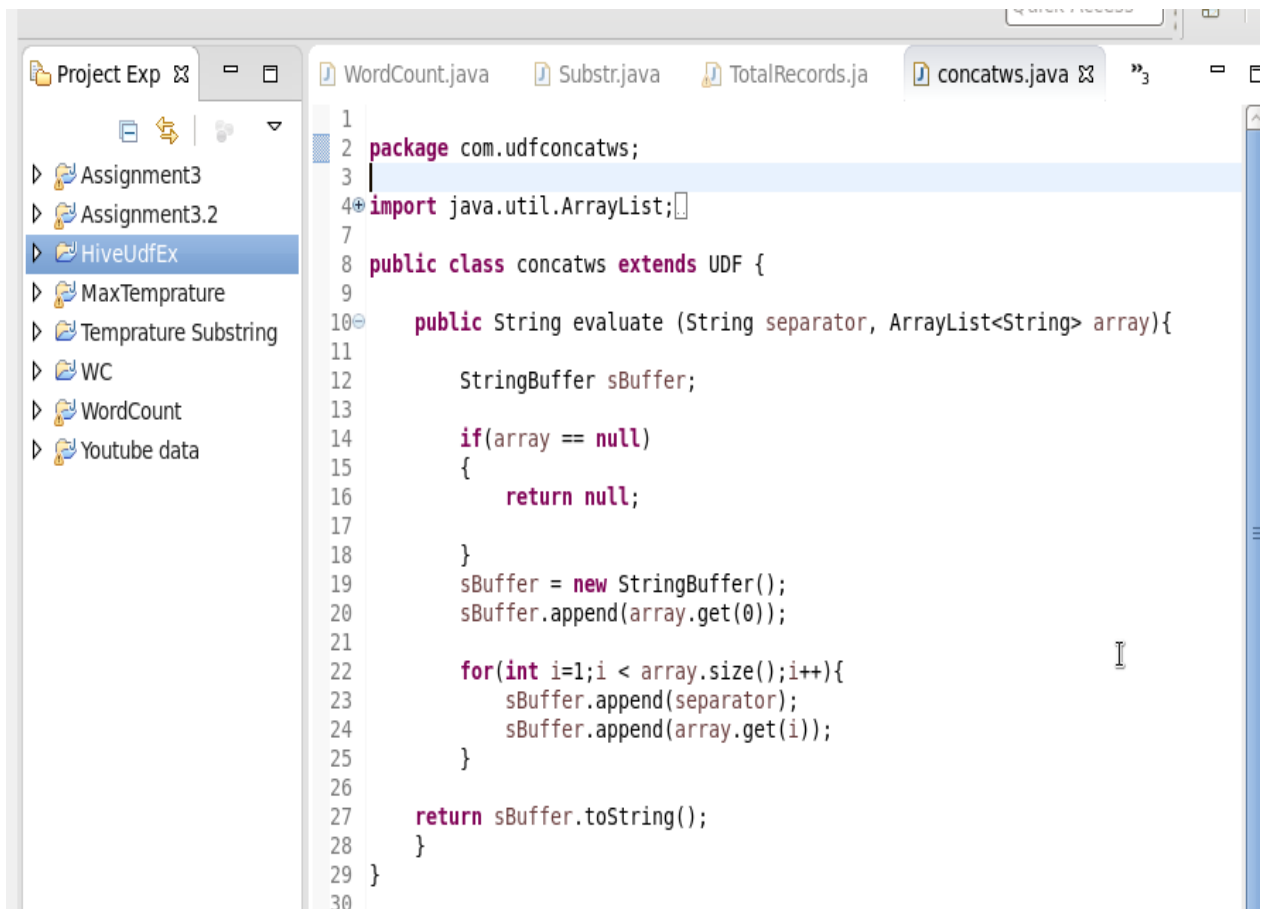


Assignment 8.2

Write hive UDF that implements functionality of string concat_ws(string SEP, array<string>), this UDF will accepts two arguments, one string and array of string. It will return a single string where all the elements of the array are separated by the SEP.

The following is the Hive UDF to accepts strings ,array string . The array string is separated by first string.



The screenshot shows an IDE with a project explorer on the left and a code editor on the right. The project explorer shows a project named 'Project Exp' with several sub-projects, including 'HiveUdfEx' which is selected. The code editor shows the implementation of a Hive UDF named 'concatws' in the file 'concatws.java'. The code is as follows:

```
1 package com.udfconcatws;
2
3
4 import java.util.ArrayList;
5
6
7
8 public class concatws extends UDF {
9
10     public String evaluate (String separator, ArrayList<String> array){
11
12         StringBuffer sBuffer;
13
14         if(array == null)
15         {
16             return null;
17         }
18
19         sBuffer = new StringBuffer();
20         sBuffer.append(array.get(0));
21
22         for(int i=1; i < array.size(); i++){
23             sBuffer.append(separator);
24             sBuffer.append(array.get(i));
25         }
26
27         return sBuffer.toString();
28     }
29 }
30
```

1. The following is the Array string is created with the name empArrayDataset

```
empArrayDataset
[acadgild@localhost Desktop]$ cat empArrayDataset
Alex    Analyst,Data Engineer,Data Consultant
Felix   Analyst,Software Engineer,Software Consultant
[acadgild@localhost Desktop]$
```

2. Now a table name empArray is created in hive as specified below

```
hive> create table empArray(empName string,empDesignation array<string>)
> row format delimited
> fields terminated by '\t'
> collection items terminated by ',';
OK
Time taken: 0.609 seconds
```

3. Data is loaded into the table as specified below

```
hive> load data local inpath '/home/acadgild/Desktop/empArrayDataset' overwrite into table empArray;
Loading data to table custom.emparray
Table custom.emparray stats: [numFiles=1, numRows=0, totalSize=95, rawDataSize=0]
OK
Time taken: 2.87 seconds
```

4. Jar file is exported for the above program with the name HiveUdf.jar
5. Jar file is added in hive as below command

Add jar /home/acadgild/Desktop/HiveUdf.jar

6. Now a temporary function is created as below.

```
hive> create temporary function separa as 'com.udfconcatws.concatws';
OK
Time taken: 3.862 seconds
```

7. Finally by using `separa` function array string is separated by first string Alex.

```
Time taken: 2.107 seconds  
hive> select separa(empName, empDesignation) from empArray;  
OK  
AnalystAlexData EngineerAlexData Consultant  
AnalystFelixSoftware EngineerFelixSoftware Consultant  
Time taken: 0.366 seconds, Fetched: 2 row(s)  
hive> █
```