Assignment 8.2 ( Advanced Hive)

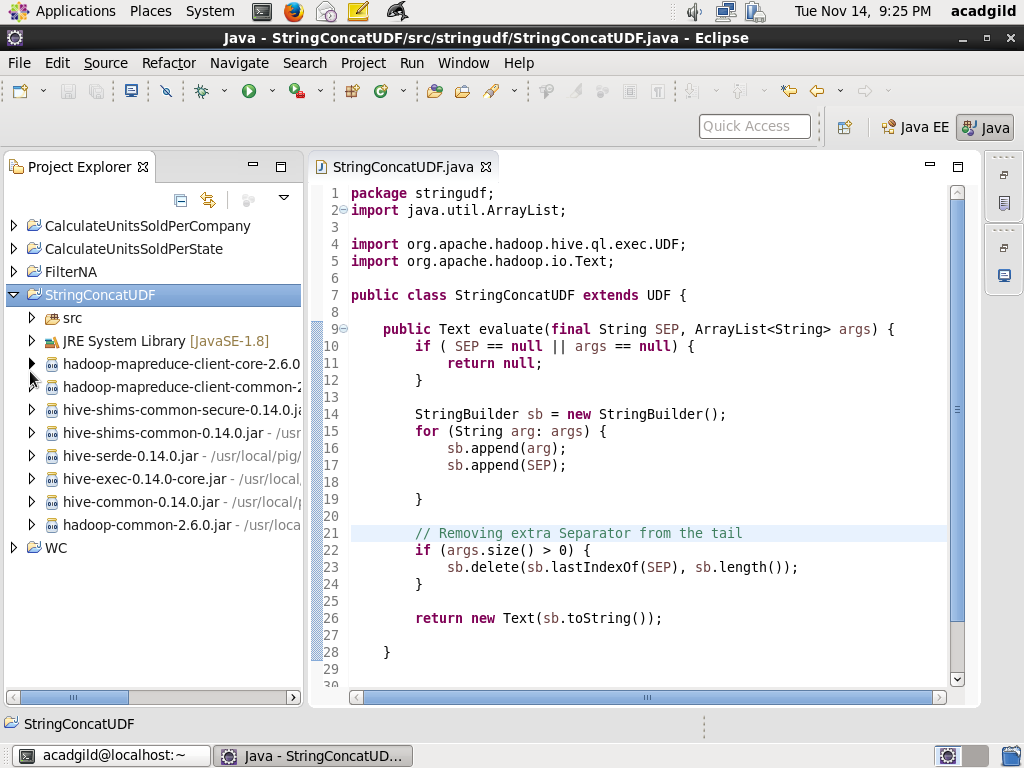
Task Create a Hive UDF concat\_ws which can concatenate array separated by a separator

Step1: Create a Java Project StringConcatUDF and add a class StringConcatUDF in package stringudf

Create a Java Project StringConcatUDF and add a class StringConcatUDF in package stringudf.

The class StringConcatUDF is extended from UDF and override the evaluate method which takes arguments, a SEP which is used for separator and ArrayList of string. In the method iterate over the list and for each element concatenate with SEP

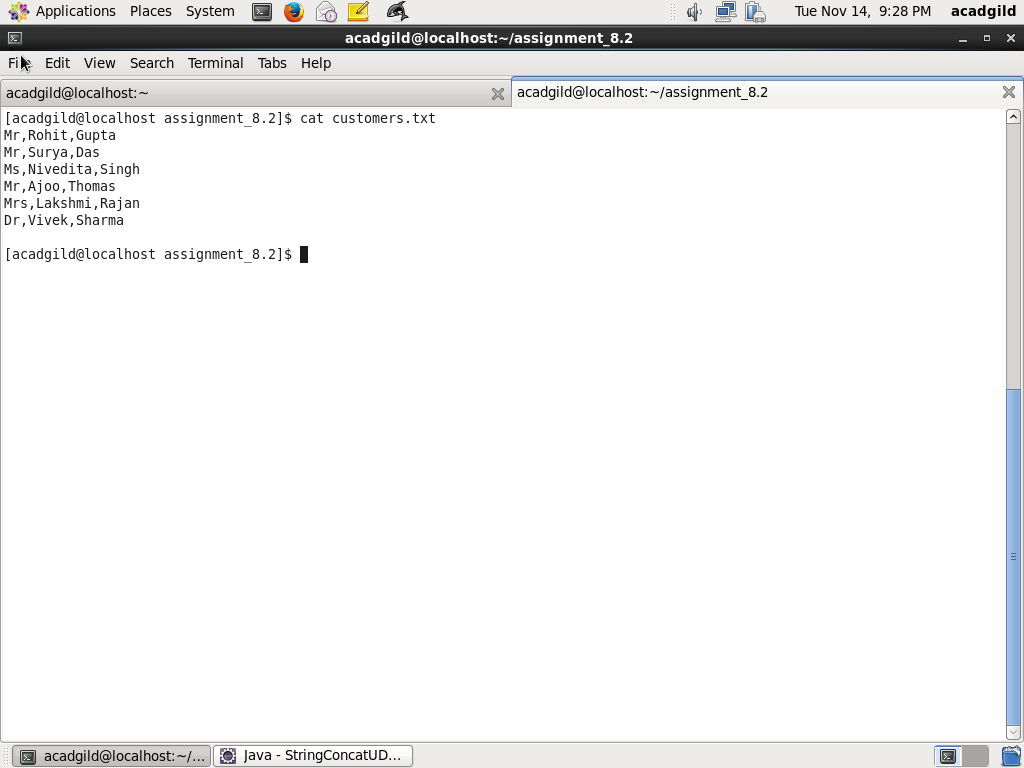
The screenshot of java class in eclipse is as below (it is also uploaded)



Step2: Export the project as StringUDFConcat.jar and store into /home/acadgild/assignment\_8.2 directory

Step3: Create a sample dataset for customers.txt which has tile,first name, last name

Screenshot is as below:



Step3: Registering the UDF function

This is done using the command below in hive shell

|  |
| --- |
| ADD JAR /home/acadgild/assignment\_8.2/StringConcatUDF.jar; |
|  |

|  |
| --- |
| CREATE TEMPORARY FUNCTION concat\_ws AS 'stringudf.StringConcatUDF'; |
| Screenshot is as below: |
|  |

Step4: Create the table customers, load the dataset and use select command with UDF concat\_ws

Table customer is created using the following query:

|  |
| --- |
| CREATE TABLE IF NOT EXISTS customers |
|  |

|  |
| --- |
| ( |
|  |

|  |
| --- |
| title string, |
|  |

|  |
| --- |
| first\_name string, |
|  |

|  |
| --- |
| last\_name string |
|  |

)

|  |
| --- |
| ROW FORMAT DELIMITED FIELDS TERMINATED BY ','; |
|  |

Load dataset using the following query:

|  |
| --- |
| LOAD DATA LOCAL INPATH '/home/acadgild/assignment\_8.2/customers.txt' |
|  |

OVERWRITE INTO TABLE customers;

Next, use select clause with concat\_ws as below which will concatenate title, first name, last name separated by –

SELECT concat\_ws('-', array(title, first\_name, last\_name)) FROM customers;

The screenshot is as below:

