

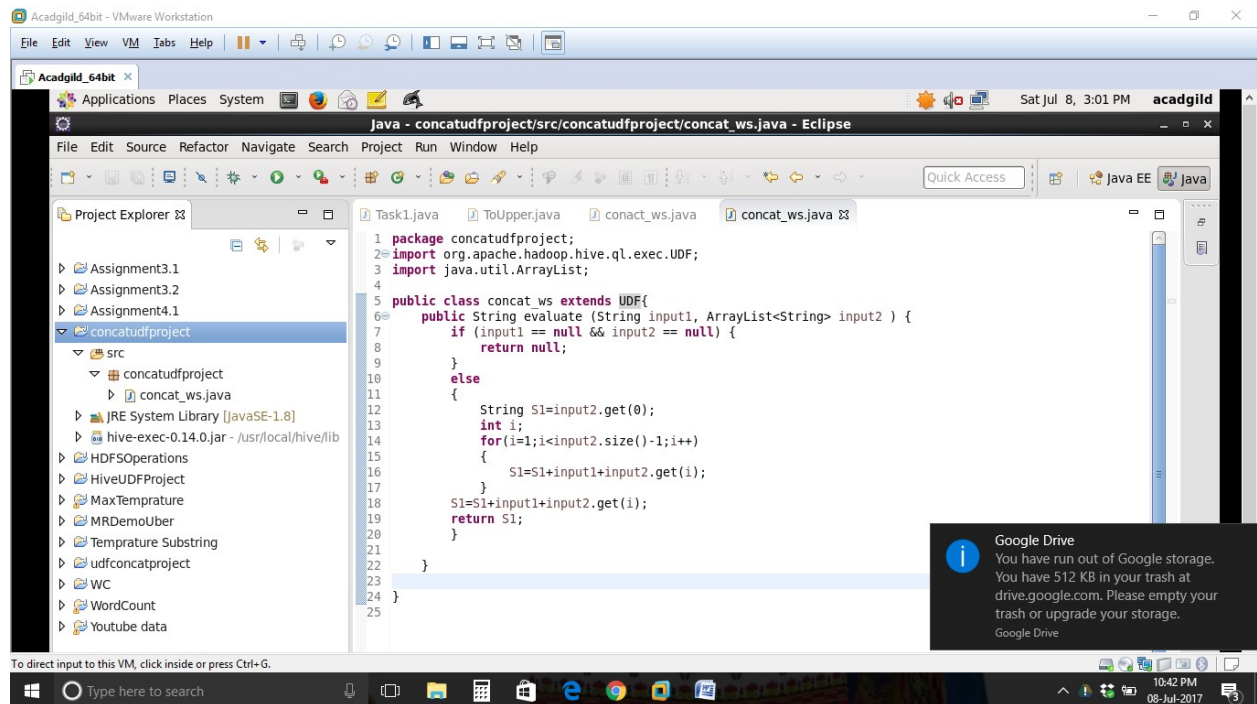
Assignment 16.2

Problem Statement

Write a hive UDF that implements functionality of string concat_ws(string SEP, array). 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.

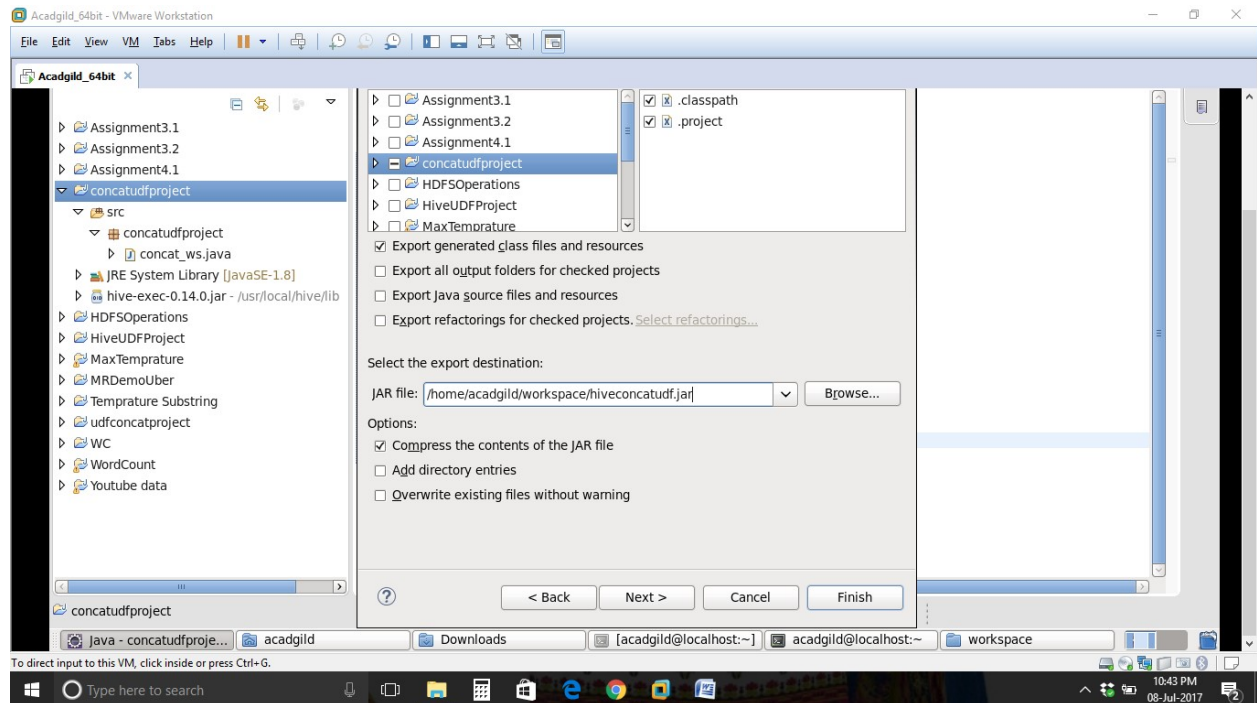
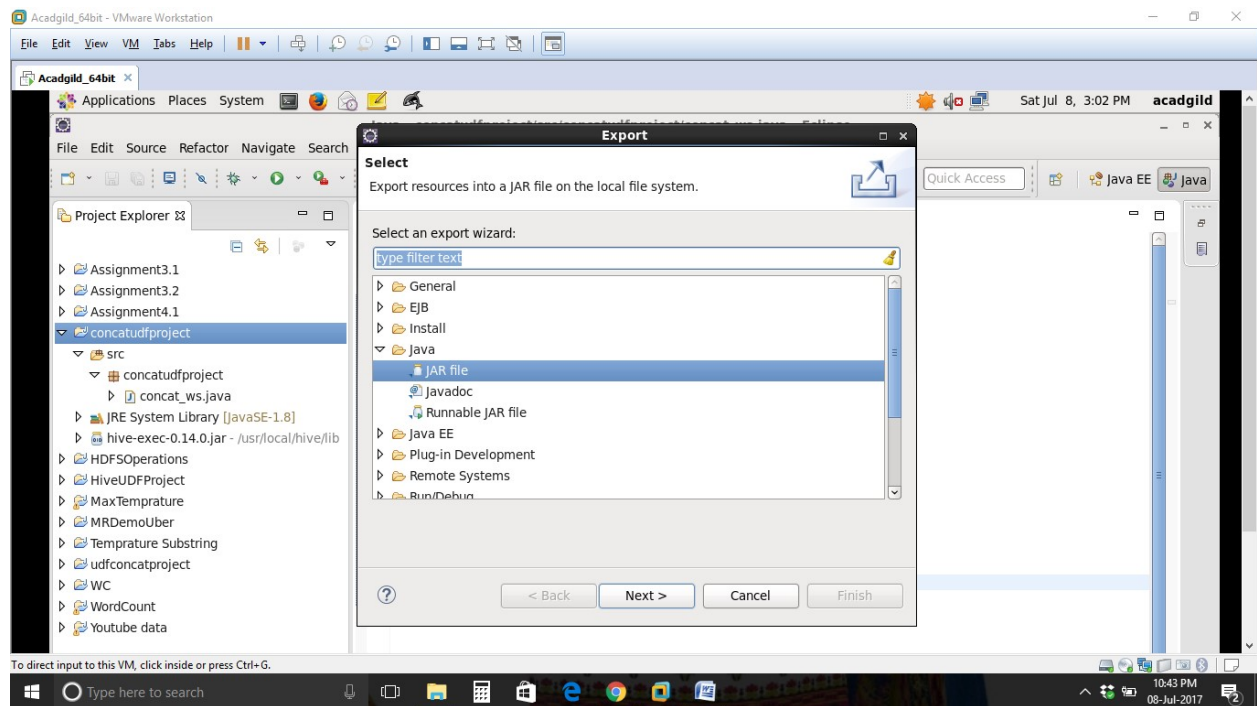
I have uploaded the project zip file in git hub

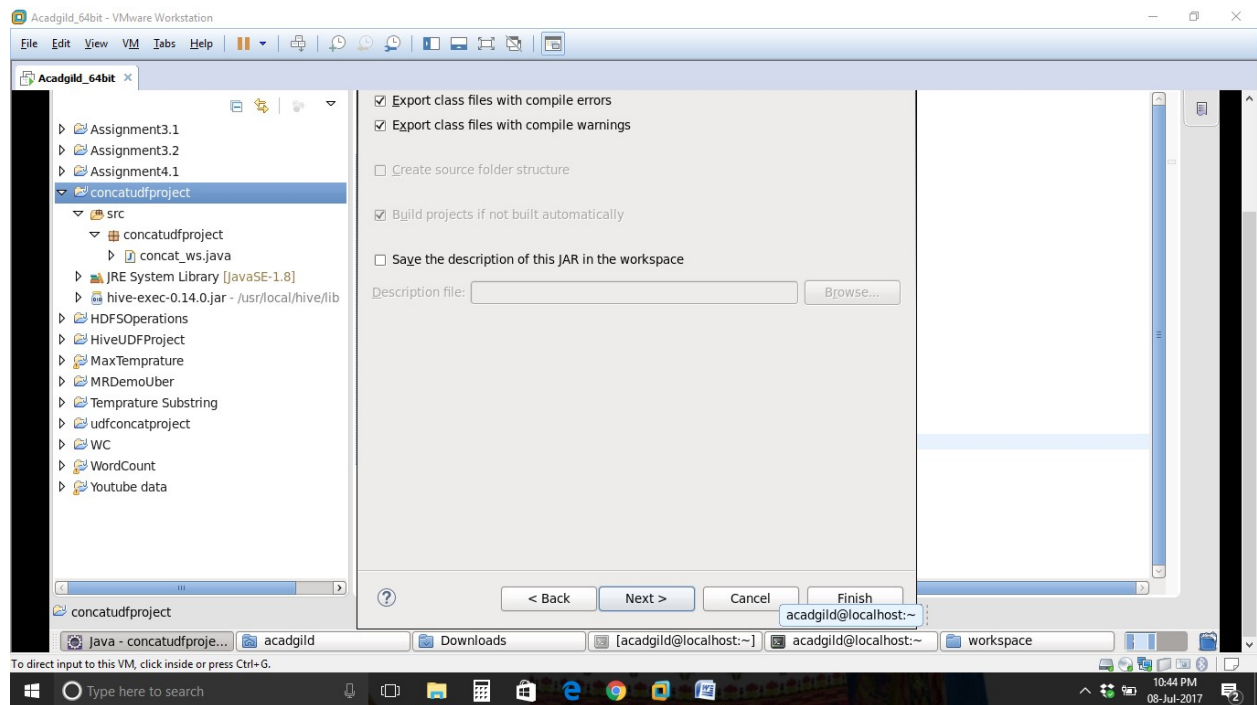
The java class :



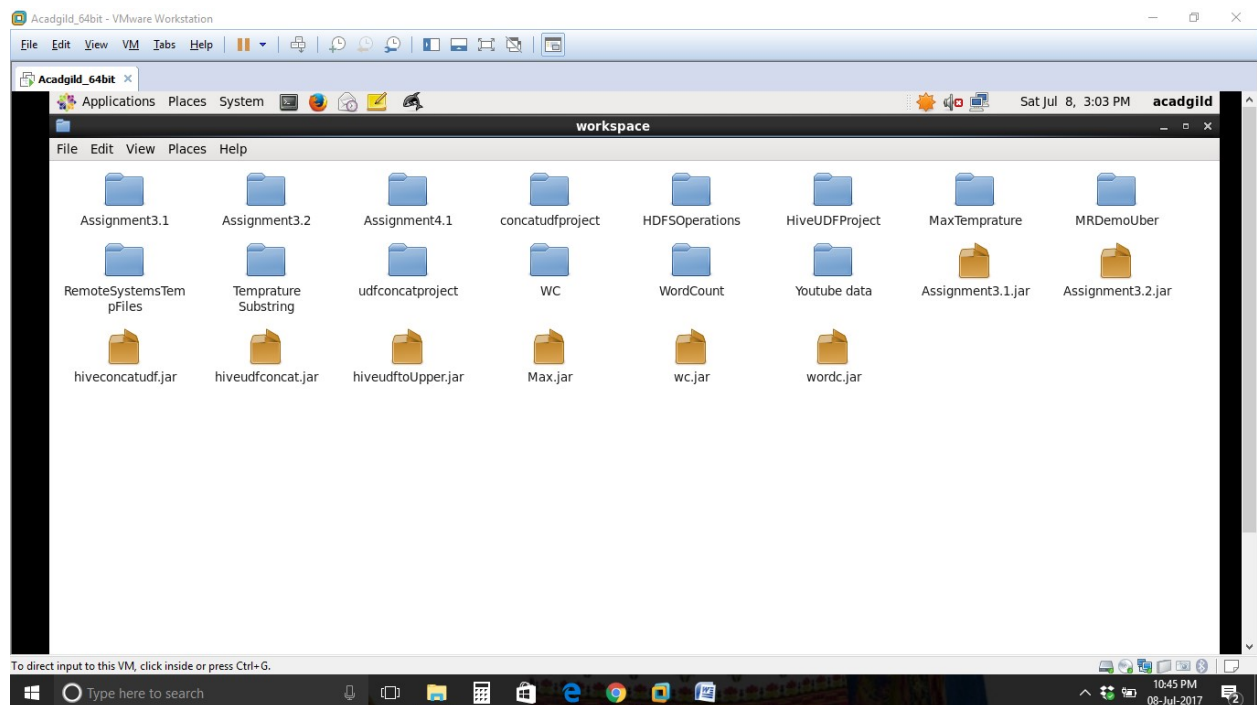
```
1 package concatudfproject;
2 import org.apache.hadoop.hive.ql.exec.UDF;
3 import java.util.ArrayList;
4
5 public class concat_ws extends UDF{
6     public String evaluate (String input1, ArrayList<String> input2 ) {
7         if (input1 == null && input2 == null) {
8             return null;
9         }
10        else
11        {
12            String S1=input2.get(0);
13            int i;
14            for(i=1;i<input2.size()-1;i++)
15            {
16                S1=S1+input1+input2.get(i);
17            }
18            S1=S1+input1+input2.get(i);
19            return S1;
20        }
21    }
22 }
23
24 }
25 }
```

Exporting the java class as jar file

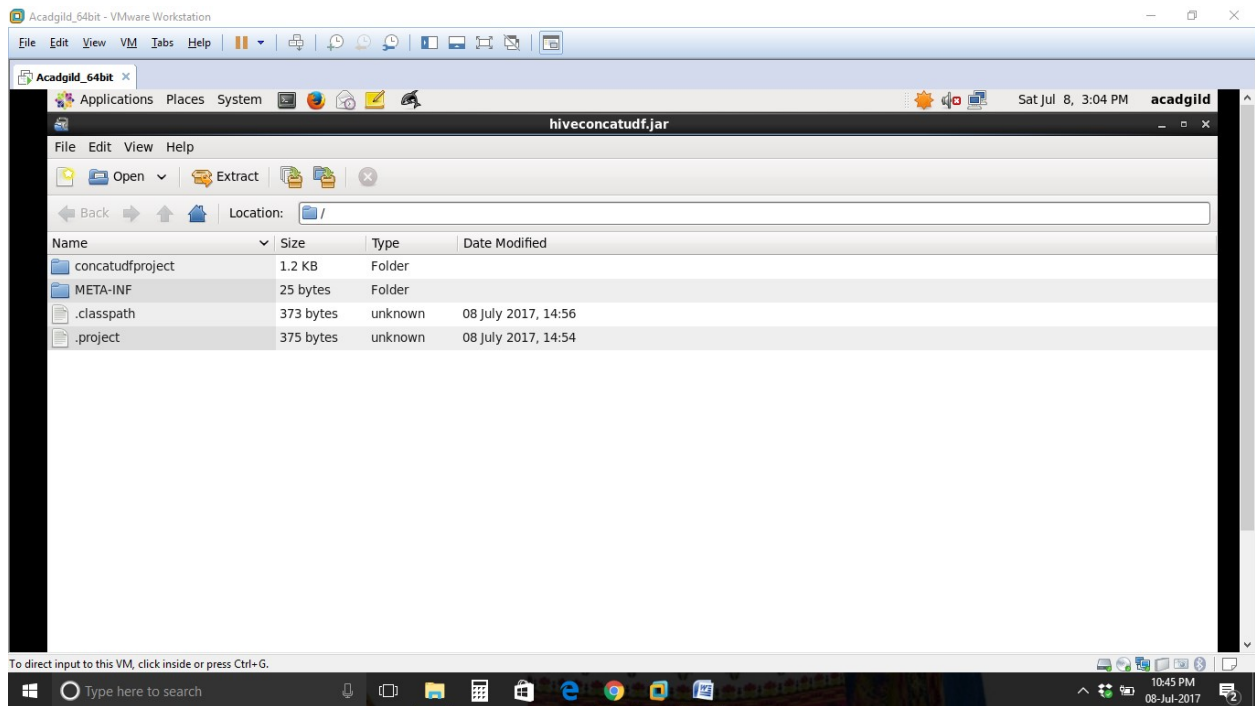




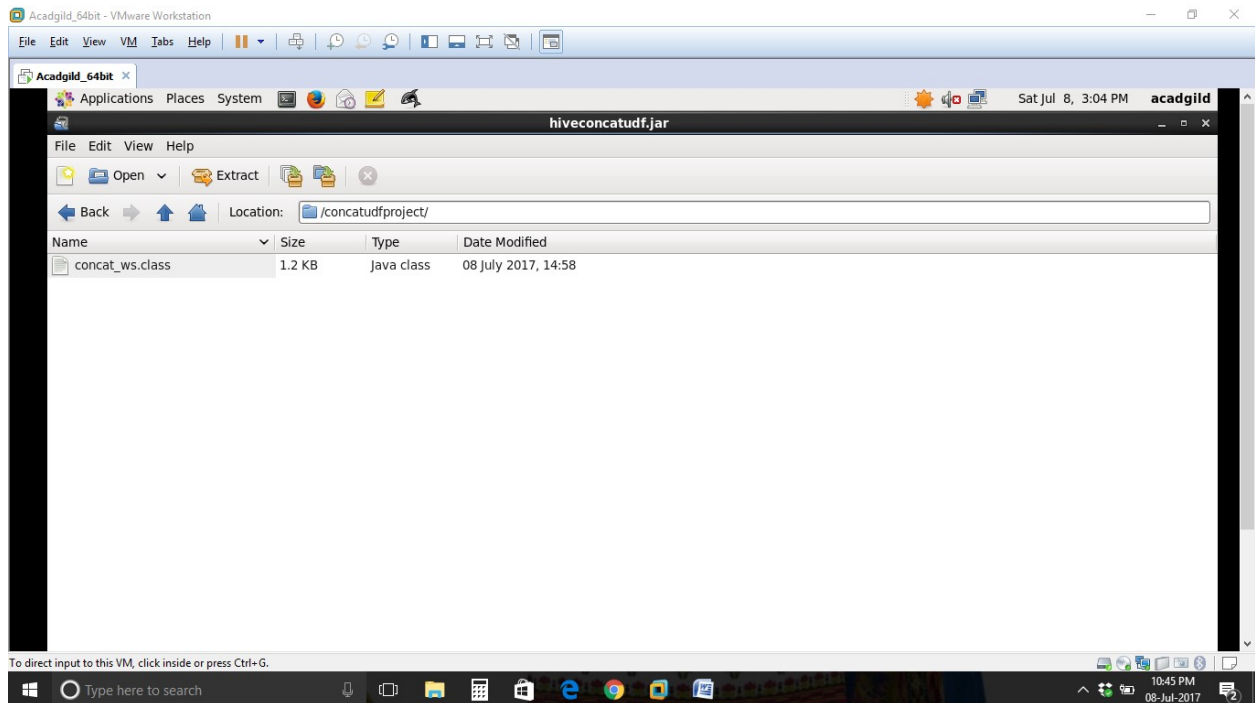
Hiveconcatudf.jar file is there under /home/acadgild/workspace



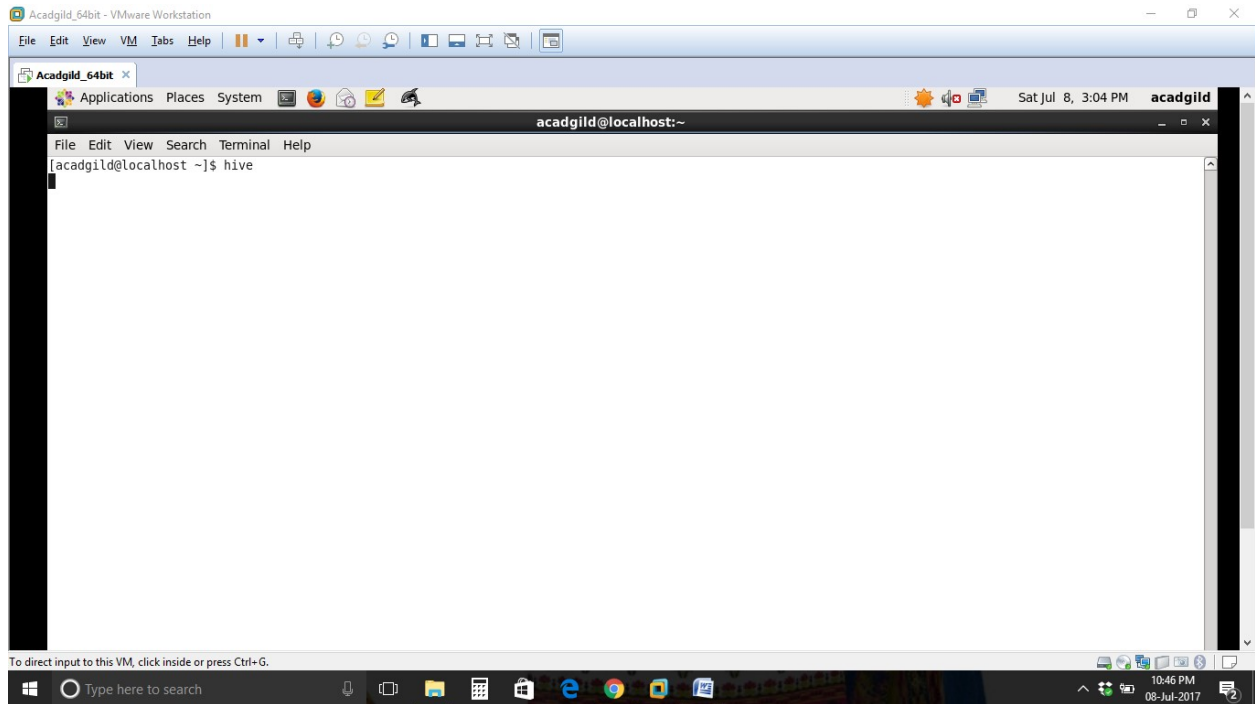
Folder found in the name of the package



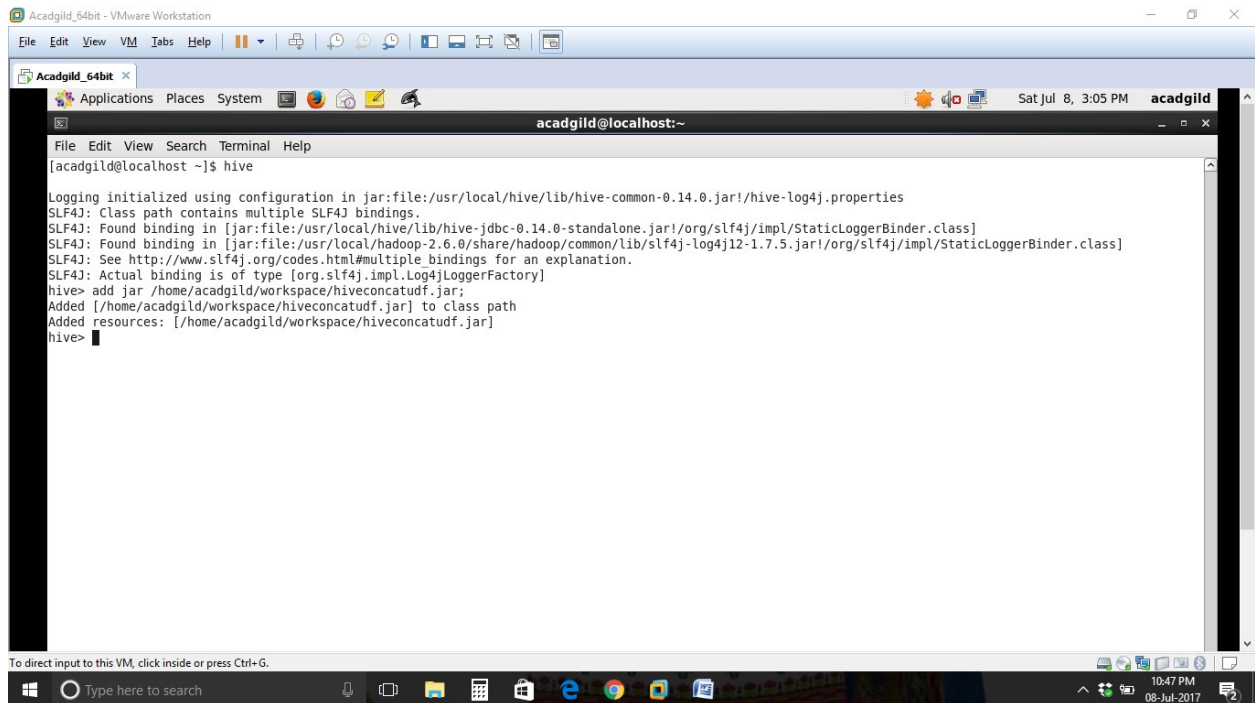
Class file inside the package name folder



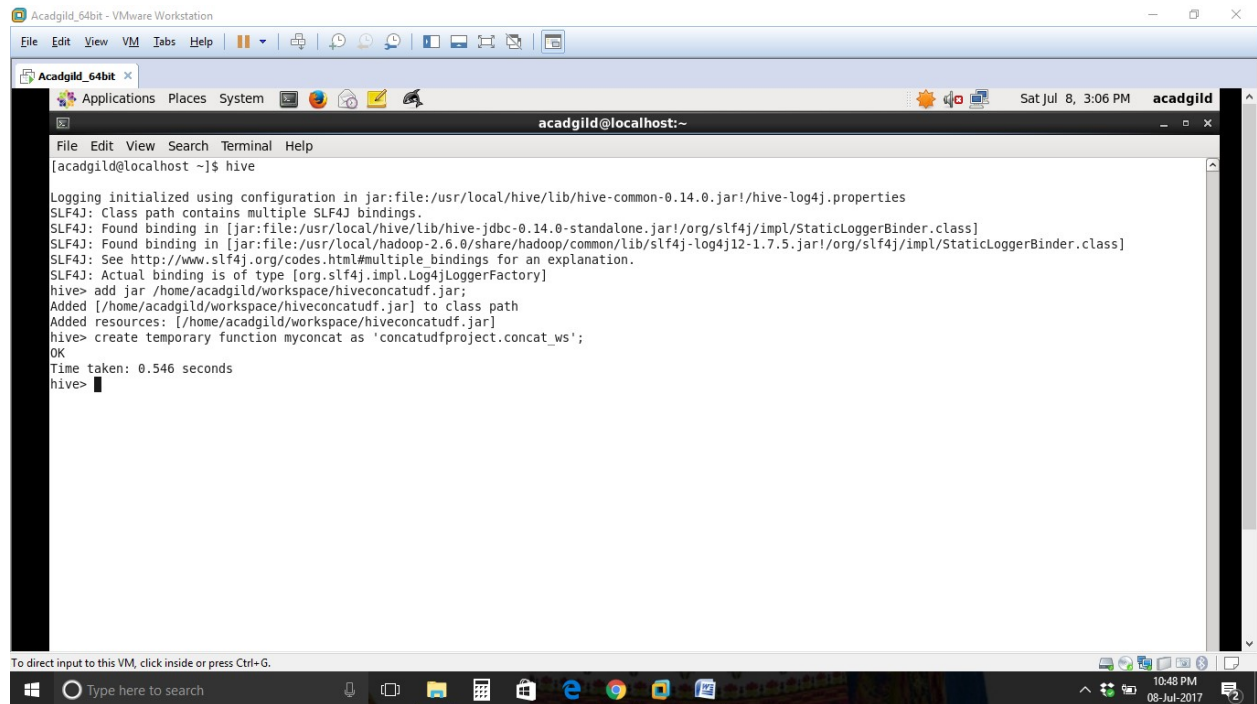
Starting hive



Adding the jar file in hive

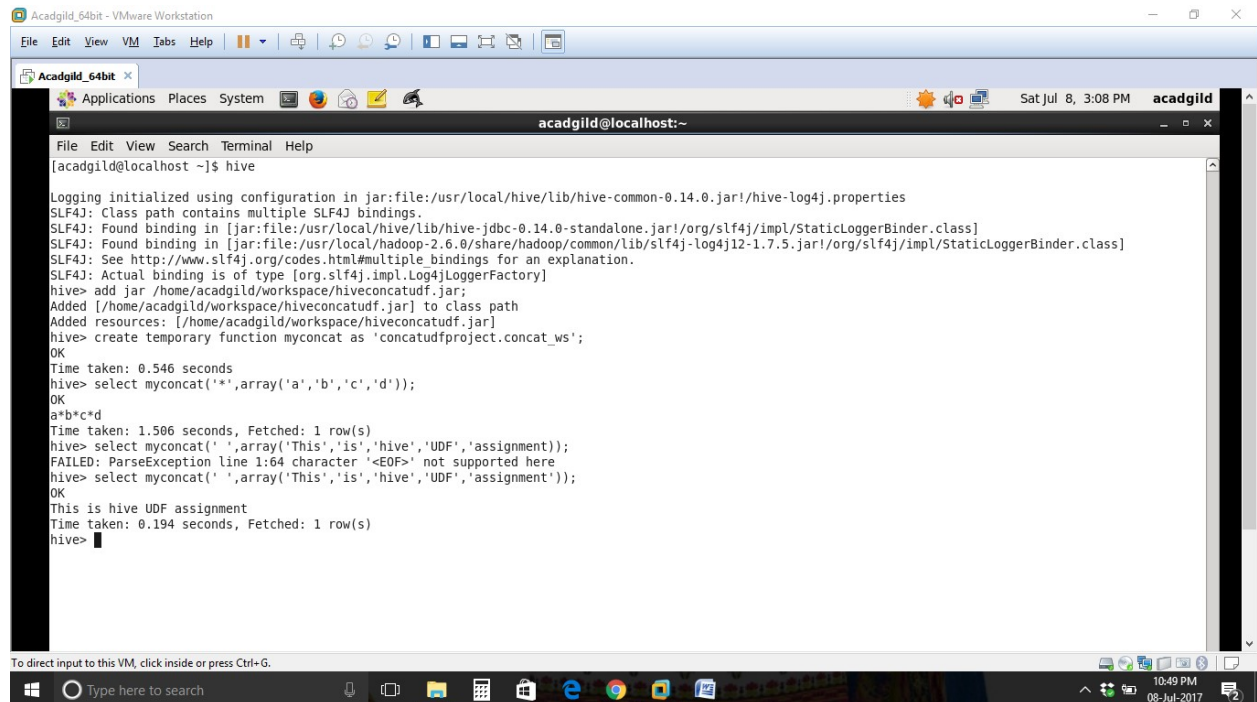


Creating temporary function myconcat from the UDF java class in the Jar file added



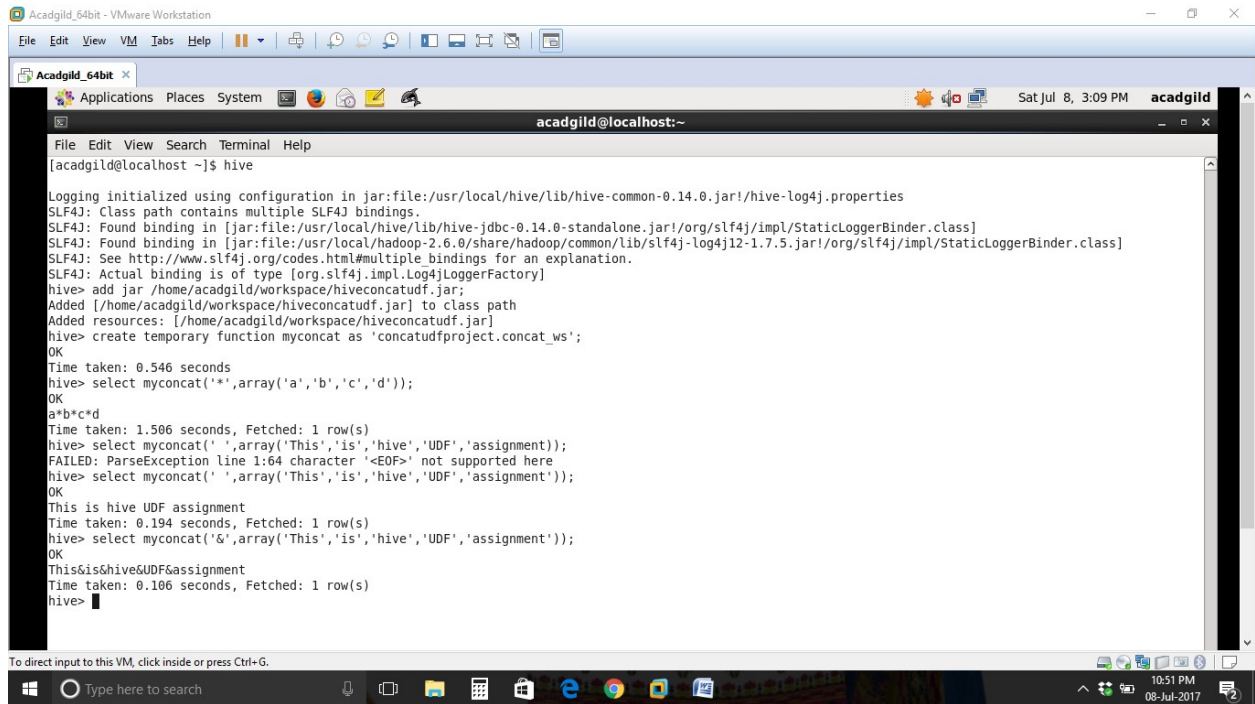
```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hive
Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-0.14.0.jar!/hive-log4j.properties
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/hive-jdbc-0.14.0-standalone.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
hive> add jar /home/acadgild/workspace/hiveconcatudf.jar;
Added [/home/acadgild/workspace/hiveconcatudf.jar] to class path
Added resources: [/home/acadgild/workspace/hiveconcatudf.jar]
hive> create temporary function myconcat as 'concatudfproject.concat_ws';
OK
Time taken: 0.546 seconds
hive>
```

Using the myconcat UDF passing separator as a string and a array of Strings which returns a String with the array elements separated by the separator.



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hive
Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-0.14.0.jar!/hive-log4j.properties
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/hive-jdbc-0.14.0-standalone.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
hive> add jar /home/acadgild/workspace/hiveconcatudf.jar;
Added [/home/acadgild/workspace/hiveconcatudf.jar] to class path
Added resources: [/home/acadgild/workspace/hiveconcatudf.jar]
hive> create temporary function myconcat as 'concatudfproject.concat_ws';
OK
Time taken: 0.546 seconds
hive> select myconcat(' ',array('a','b','c','d'));
OK
a*b*c*d
Time taken: 1.506 seconds, Fetched: 1 row(s)
hive> select myconcat(' ',array('This','is','hive','UDF','assignment'));
FAILED: ParseException line 1:64 character '<EOF>' not supported here
hive> select myconcat(' ',array('This','is','hive','UDF','assignment'));
OK
This is hive UDF assignment
Time taken: 0.194 seconds, Fetched: 1 row(s)
hive>
```

Calling the same myconcat function with different separators and different size string arrays



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hive
Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-0.14.0.jar!/hive-log4j.properties
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/hive-jdbc-0.14.0-standalone.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
hive> add jar /home/acadgild/workspace/hiveconcatudf.jar;
Added [/home/acadgild/workspace/hiveconcatudf.jar] to class path
Added resources: [/home/acadgild/workspace/hiveconcatudf.jar]
hive> create temporary function myconcat as 'concatudfproject.concat_ws';
OK
Time taken: 0.546 seconds
hive> select myconcat(' ',array('a','b','c','d'));
OK
a*b*c*d
Time taken: 1.506 seconds, Fetched: 1 row(s)
hive> select myconcat(' ',array('This','is','hive','UDF','assignment'));
FAILED: ParseException line 1:64 character '<EOF>' not supported here
hive> select myconcat(' ',array('This','is','hive','UDF','assignment'));
OK
This is hive UDF assignment
Time taken: 0.194 seconds, Fetched: 1 row(s)
hive> select myconcat('&',array('This','is','hive','UDF','assignment'));
OK
This&is&hive&UDF&assignment
Time taken: 0.106 seconds, Fetched: 1 row(s)
hive>
```