

BigData Assignment 3.5

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 -

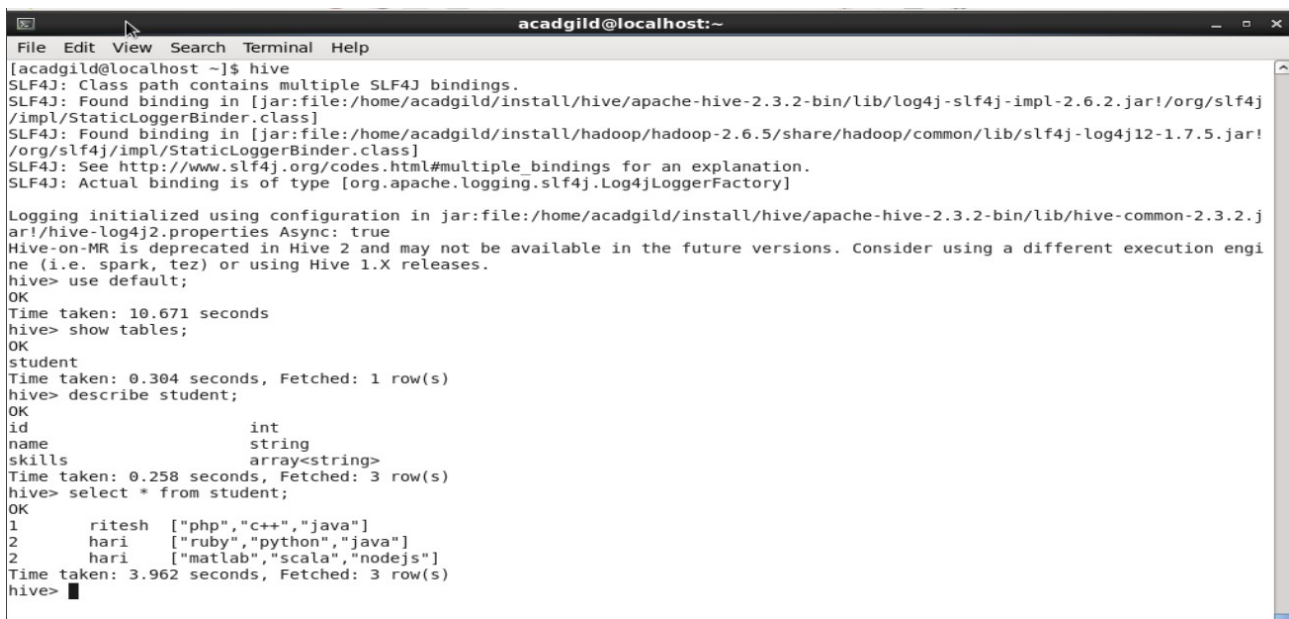
StringUDF.java

```
package hive;

import java.util.ArrayList;
import org.apache.hadoop.hive.ql.exec.UDF;

public class StringUDF extends UDF{
    public String evaluate(String sep, ArrayList<String> arr) {
        String res = "";
        if(sep==null || arr==null) {
//To check input is null or not
            return null;
        }
        for(int i=0;i<arr.size()-1;i++) {
            res = res+arr.get(i)+sep;
//To add separator
        }
        res = res+arr.get(arr.size()-1);
        return res;
    }
}
```

- First started the all the required daemons
start-all.sh
- Started the hive shell
hive
- Checked the database and the content of the table
use default;
select * from student;



```

[acadgild@localhost ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/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.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> use default;
OK
Time taken: 10.671 seconds
hive> show tables;
OK
student
Time taken: 0.304 seconds, Fetched: 1 row(s)
hive> describe student;
OK
id                int
name              string
skills            array<string>
Time taken: 0.258 seconds, Fetched: 3 row(s)
hive> select * from student;
OK
1      ritesh  ["php","c++","java"]
2      hari   ["ruby","python","java"]
2      hari   ["matlab","scala","nodejs"]
Time taken: 3.962 seconds, Fetched: 3 row(s)
hive>

```

- Added the hive jar for string separation
ADD jar /home/acadgild/eclipse/hive_stringConcat.jar;
- Created a hive function to use hive custom UDF
create temporary function arraySep as 'hive.StringUDF';
- Separated the skills array from the table using separator '|' and using hive custom UDF
select id,name,arraySep("|",skills) from student;

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
hive> ADD JAR /home/acadgild/eclipse-workspace/hive_stringConcat.jar ;  
Added [/home/acadgild/eclipse-workspace/hive_stringConcat.jar] to class path  
Added resources: [/home/acadgild/eclipse-workspace/hive_stringConcat.jar]  
hive> create temporary function arraySep as 'hive.StringUDF' ;  
OK  
Time taken: 0.531 seconds  
hive> select id,name,arraySep("|",skills) from student;  
OK  
1      ritesh  php|c++|java  
2      hari    ruby|python|java  
2      hari    matlab|scala|nodejs  
Time taken: 3.779 seconds, Fetched: 3 row(s)  
hive> █
```

Output -

In the above screenshot , it is evident that we using hive custom UDF function , we added separator '|' in the skills array using that function.