Session 8

Assignment 2

|  |  |
| --- | --- |
| **Prepared For:** | AcadGild |
|  |  |
| **Document Approval:** | **AcadGild** |
|  |  |
|  |  |
|  |  |
|  |  |
| **Project Title:** | Session 8 – Assignment 2 |
|  |  |
| **Prepared By:** | Duncan Burgess |
|  |  |
|  | dburgess@duncb.com |
|  |  |
| **Primary Engineer:** | Duncan Burgess |
|  |  |
| **Document Reference:** | **Session 8 – Assignment 2** |
|  |  |
| **Start Date:** | 24/09/2017 |
|  |  |
|  |  |

# 

# Contents

[Contents 2](#_Toc494014693)

[Change History 3](#_Toc494014694)

[1. Problem Statement 4](#_Toc494014695)

[2. Solution 5](#_Toc494014696)

[2.1. Creating Jar 5](#_Toc494014697)

[2.2. Datasets 6](#_Toc494014698)

[2.3. Creating tables and importing Data 6](#_Toc494014699)

[2.4. Adding Jar and Creating temporary function 7](#_Toc494014700)

[2.5. Execution with Table 7](#_Toc494014701)

[2.6. Execution passing parameters 7](#_Toc494014702)

# Change History

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Document Revision** | **Date** | **Authored By** | **Authorised By** | **Sections Affected** | **Reason for Change** |
| Rev 01 | 24/09/2017 | Duncan Burgess |  | All | Initial release. |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

# 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.

# Solution

## Creating Jar

The jar was created in Eclipse to implement the concat\_ws (String SEP,array<String>).

2 jars where imported hadoop-common-2.6.0.jar and hive-exec-0.14.0.jar

Java Code

*//UDF for hive to concatenate a file or sting entry with delimiter //by Duncan Burgess*

*import java.util.List;*

*import org.apache.hadoop.hive.ql.exec.UDF;*

*import org.apache.hadoop.io.Text;*

*public class HiveConcat extends UDF*

*{*

*private String ConcatResults;*

*public String evaluate (String SEP, List<Text> mystring) { if (SEP == null && mystring == null)*

*{return null;*

*}*

*ConcatResults = null;*

*ConcatResults = mystring.get(0).toString();*

*for (int n=1;n<mystring.size();n++)*

*ConcatResults += SEP + mystring.get(n); return ConcatResults;*

*}*

*}*

Once created the jar is exported.



## Datasets

This is created to test the various delimiters.

*hiveconcat.txt*

*; Bing,Bang,Bong*

*' Ford,Vauxhall,Bently,Ferrari*

*- one,twwo,three,four,five*

*. Fish,Meat,Vegtables,Nuts*

*, 2017,2016,2015,2014,2013*

*good,bad,ugly*

## Creating tables and importing Data

**Create table** ***concatdata*** using database ***hiveconcat***

*hive> create table concatdata*

*> (seperator STRING,data array<STRING>)*

*> row format delimited*

*> fields terminated by '\t'*

*> collection items terminated by ',';*

*OK*

*Time taken: 0.092 seconds*

**Load Data**

*hive> load data local inpath '/home/acadgild/Datasets/hiveconcat.txt' into table concatdata;*

*Loading data to table hiveconcat.concatdata*

*Table hiveconcat.concatdata stats: [numFiles=1, totalSize=146]*

*OK*

*Time taken: 0.29 seconds*

**Test data imported correctly**

*hive> select \* from concatdata;*

*OK*

*; ["Bing","Bang","Bong"]*

*' ["Ford","Vauxhall","Bently","Ferrari"]*

*- ["one","twwo","three","four","five"]*

*. ["Fish","Meat","Vegtables","Nuts"]*

*, ["2017","2016","2015","2014","2013"]*

*["good","bad","ugly"]*

*Time taken: 0.053 seconds, Fetched: 6 row(s)*

## Adding Jar and Creating temporary function

*hive> add jar /home/acadgild/Documn ents/HiveConcat.jar;*

*Added [/home/acadgild/Documents/HiveConcat.jar] to class path*

*Added resources: [/home/acadgild/Documents/HiveConcat.jar]*

*hive> create temporary function concater as 'HiveConcat';*

## Execution with Table

Using function concater to concatenate data from concatdata table.

*hive> select concater(seperator,data) as results from concatdata;*

*OK*

*Bing;Bang;Bong*

*Ford'Vauxhall'Bently'Ferrari*

*one-twwo-three-four-five*

*Fish.Meat.Vegtables.Nuts*

*2017,2016,2015,2014,2013*

*good bad ugly*

*Time taken: 0.071 seconds, Fetched: 6 row(s)*

## Execution passing parameters

Using command line various tests with different delimiters

select concater('***delimeter***',array (***array of entries***));

*hive> select concater(' ',array ('Testing','test','Tested'));*

*OK*

*Testing test Tested*

*Time taken: 0.068 seconds, Fetched: 1 row(s)*

*hive> select concater('-space-',array ('Testing','Test','Tested'));*

*OK*

*Testing-space-Test-space-Tested*

*Time taken: 0.051 seconds, Fetched: 1 row(s)*

*hive> select concater('',array ('Testing','Test','Tested'));*

*OK*

*TestingTestTested*

*Time taken: 0.066 seconds, Fetched: 1 row(s)*

*hive> select concater(':',array ('Testing','Test','Tested'));*

*OK*

*Testing:Test:Tested*

*Time taken: 0.066 seconds, Fetched: 1 row(s)*