**Session 8: ADVANCED HIVE**

Assignment 8.2

Student Name: Abarajithan SA

Course: Big Data Hadoop & Spark Training

Start Date:  2017-09-09

End Date:  2017-11-26

**Assignment 8.2**– 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.

Contents

[Introduction 2](#_Toc499156694)

[Prerequisites 2](#_Toc499156695)

[Create Database and Table 2](#_Toc499156696)

[HIVE QL 2](#_Toc499156697)

[HIVE QL 2](#_Toc499156698)

[HIVE UDF Java code 3](#_Toc499156699)

[HIVE UDF CONCAT\_WS function 4](#_Toc499156700)

[HIVE QL 4](#_Toc499156701)

[Required Output 5](#_Toc499156702)

# Introduction

In this assignment we are going to write HIVE UDF using Java in order to achieve the CONCAT\_WS function.

# Prerequisites

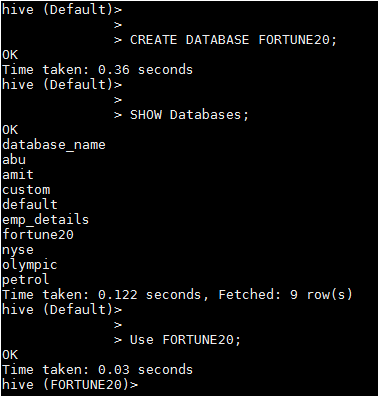
## Create Database and Table

**Create Database FORTUNE20**

### HIVE QL

***CREATE DATABASE FORTUNE20***

***Use FORTUNE20;***



**Create Table Fortune\_company**

### HIVE QL

***CREATE TABLE fortune\_company(rank int, company\_name string,website string, protocal string)***

***ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t';***

***LOAD DATA LOCAL INPATH '/home/acadgild/hadoop/fortune20.txt’***

***INTO TABLE fortune20.fortune\_company;***



# HIVE UDF Java code

**package** concatws;

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

import org.apache.hadoop.hive.ql.exec.Description;

@Description(name = "concatws", value = "\_FUNC\_(string SEP, array<string>) - RETURN\_TYPE(STRING)\n" + "Description: Concatenate two strings, separated by the seperator",

extended = "Example:\n"

+ " > SELECT CONCAT\_WS (website,’.’,protocal) FROM src;\n"

+ "www.walmart.com")

**public** **class** concatws **extends** UDF

{

**public** String evaluate(String param1, String[] param2)

{

String Output = "";

**if**(param1==**null** && param2==**null**)

{

**return** **null**;

}

**for**(**int** i = 0; i < param2.length; i++)

{

Output+= param2[i];

}

**return**(param1.concat(Output));

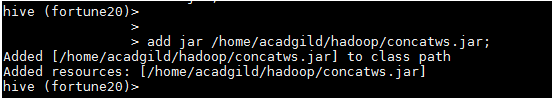
}

}

After that we are adding JAR created from the JAVA class which is defining the UDF using below syntax-

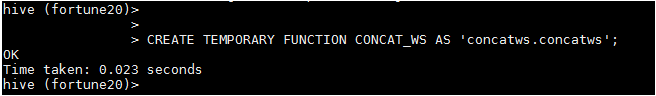
# HIVE UDF CONCAT\_WS function

***add jar /home/acadgild/hadoop/concatws.jar;***



After that we are creating a temporary function “CONCAT\_WS” using below syntax-

***CREATE TEMPORARY FUNCTION CONCAT\_WS AS 'concatws.concatws';***



After that we run below query to take one column (company\_name) input as String and another array(website,’.’,protocal) as Array of Strings and concatenate them,

## HIVE QL

***SELECT company\_name, CONCAT\_WS(website,'.',protocal) From fortune\_company;***



## Required Output

