**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](#_Toc499197320)

[Dataset 2](#_Toc499197321)

[Prerequisites 2](#_Toc499197322)

[Create Database and Table 2](#_Toc499197323)

[HIVE QL 2](#_Toc499197324)

[HIVE QL 3](#_Toc499197325)

[HIVE UDF Java code 5](#_Toc499197326)

[HIVE UDF CONCAT\_WS function 5](#_Toc499197327)

[HIVE QL 6](#_Toc499197328)

[Required Output 6](#_Toc499197329)

# Introduction

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

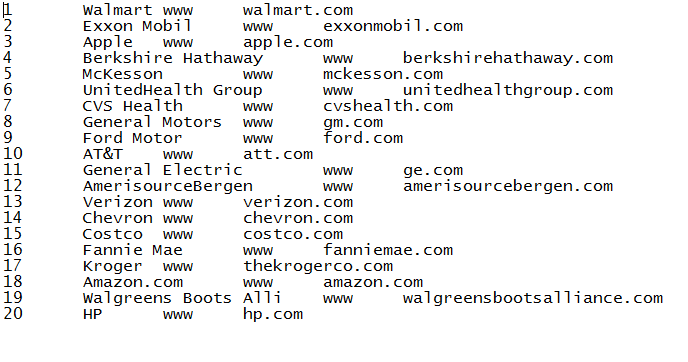
We have fortune 20 companies list and its company website URL, but the ‘www’ and the remaining domain are separated. In our output we try to achieve the output as below,

**1 walmart www.walmart.com**

# Dataset

The below data contains, the column name as,

Rank, company\_name, website, protocal.



# Prerequisites

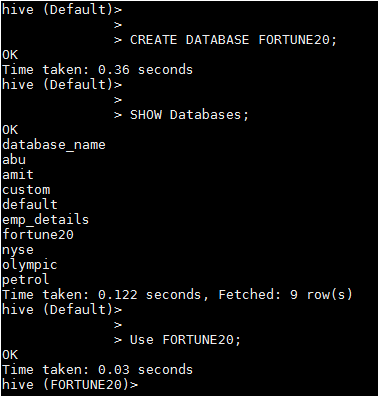
## Create Database and Table

**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';***

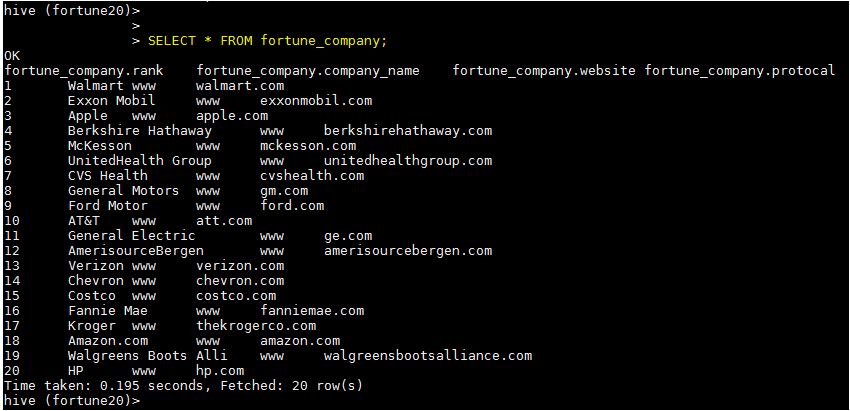
***LOAD DATA LOCAL INPATH '/home/acadgild/hadoop/fortune20.txt’***

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



Viewing the data in the table fortune\_company,

***SELECT \* FROM 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));

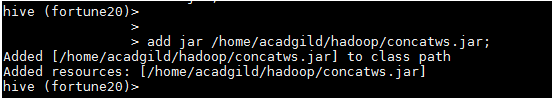
}

}

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

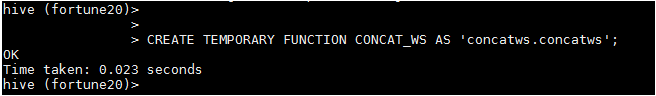
# HIVE UDF CONCAT\_WS function

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

## Required Output

