**Create UDF (User Defined Functions) in Apache Pig**  **and execute it in MapReduce / HDFS mode**

**Aim:**

To create UDF in Apache Pig and execute it in MapReduce/HDFS mode.

**Procedure:**

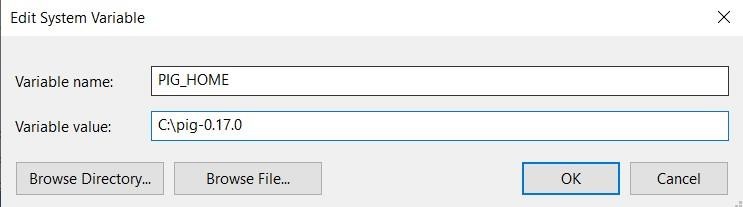
**Pig Download and installation:**

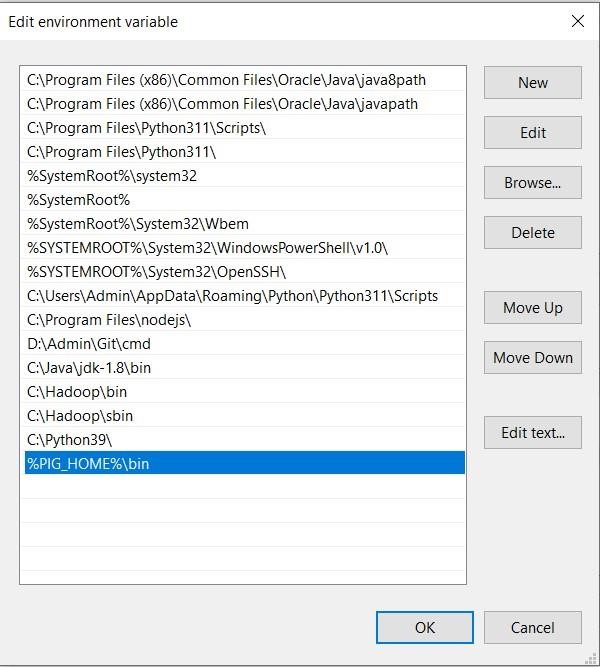
1. Download Pig:

Download Pig from “https://downloads.apache.org/pig/pig-0.17.0/”

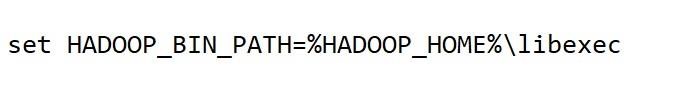


1. Add the environment variable for Pig:





1. Go to C:\pig-0.16.0\bin and open pig (Windows Command Script)



1. Open Windows Powershell and type “pig –x local” and check whether pig grunt appears.

**Pig is successfully installed.**

**Create UDF:**

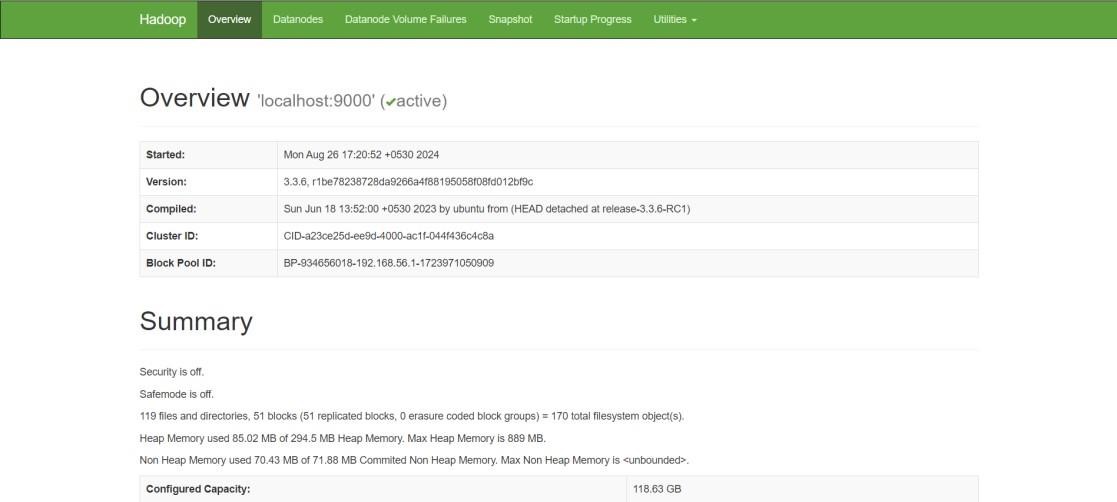
1. **Start Hadoop services:**

Open command prompt as an administrator

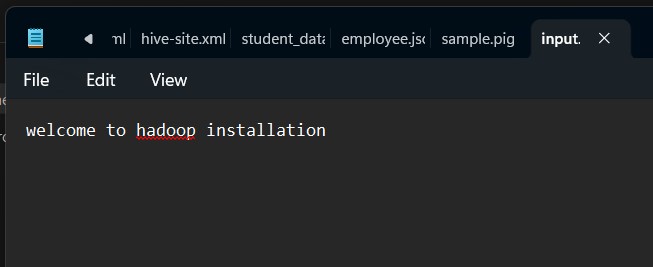
start-dfs.cmd start-

yarn.cmd

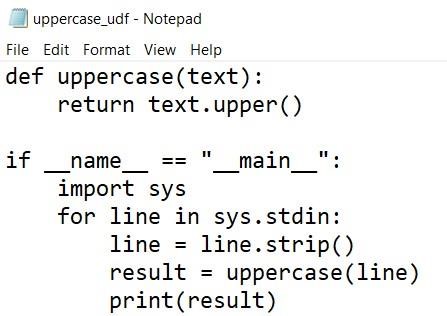
1. Open the browser and go to the URL “localhost:9870”



1. Create a text file “input.txt”:

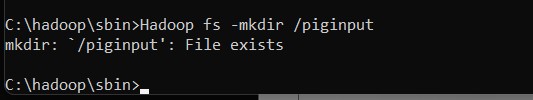


1. Create a Python file “uppercase\_udf.py”:

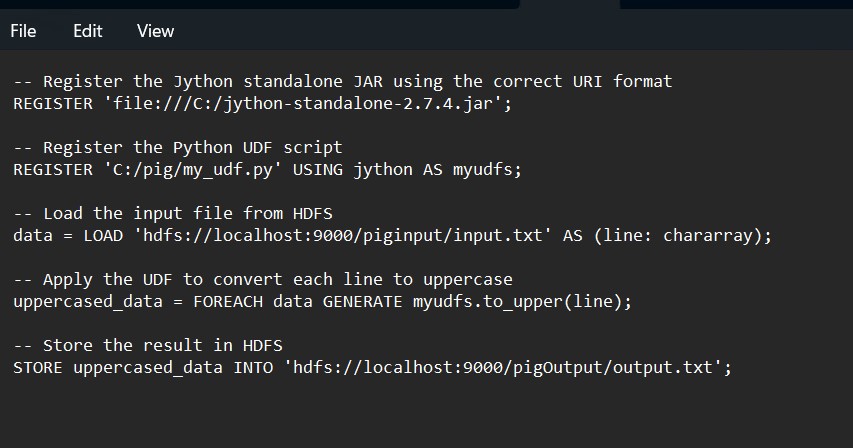


6.Create a Directory in HDFS and copy the Input File to HDFS Hadoop fs -mkdir /piginput

hadoop fs -put udfs C:\pig\sample.pig /piginput

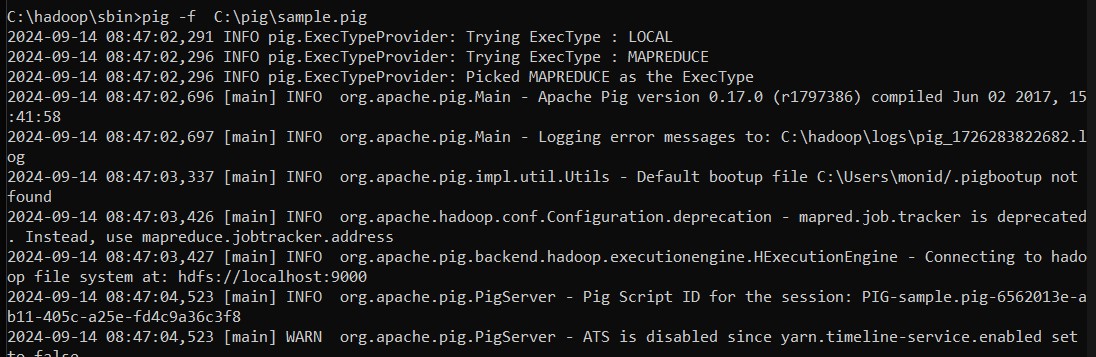


1. Create pig file “sample.pig”:

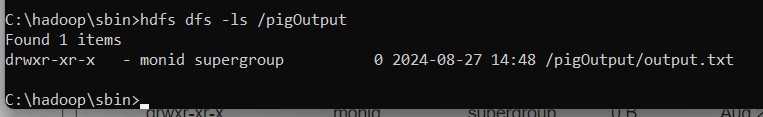


1. Execute Pig file:

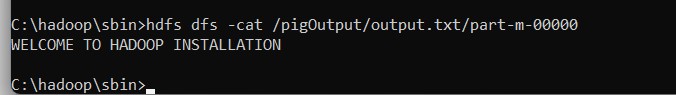
pig -f C:\pig\sample.pig



1. View the Output hdfs dfs -ls /pigOutput

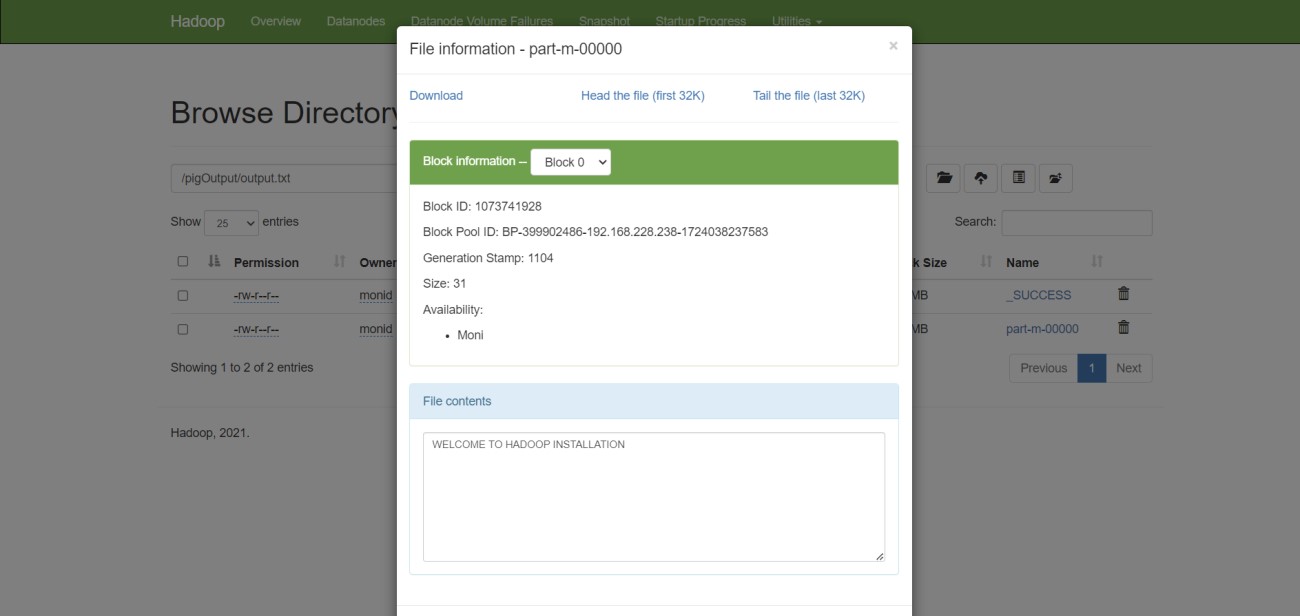


hdfs dfs -cat /pigOutput/output.txt/part-m-00000



1. Once the map reduce operations are performed successfully, the output will be present in the specified directory.

“/pigOutput/output.data/part-m-00000”



1. Stop Hadoop Services stop-dfs.cmd stopyarn.cmd

**Result:**

Thus, UDF in Apache Pig has been created and executed in MapReduce/HDFS mode successfully.