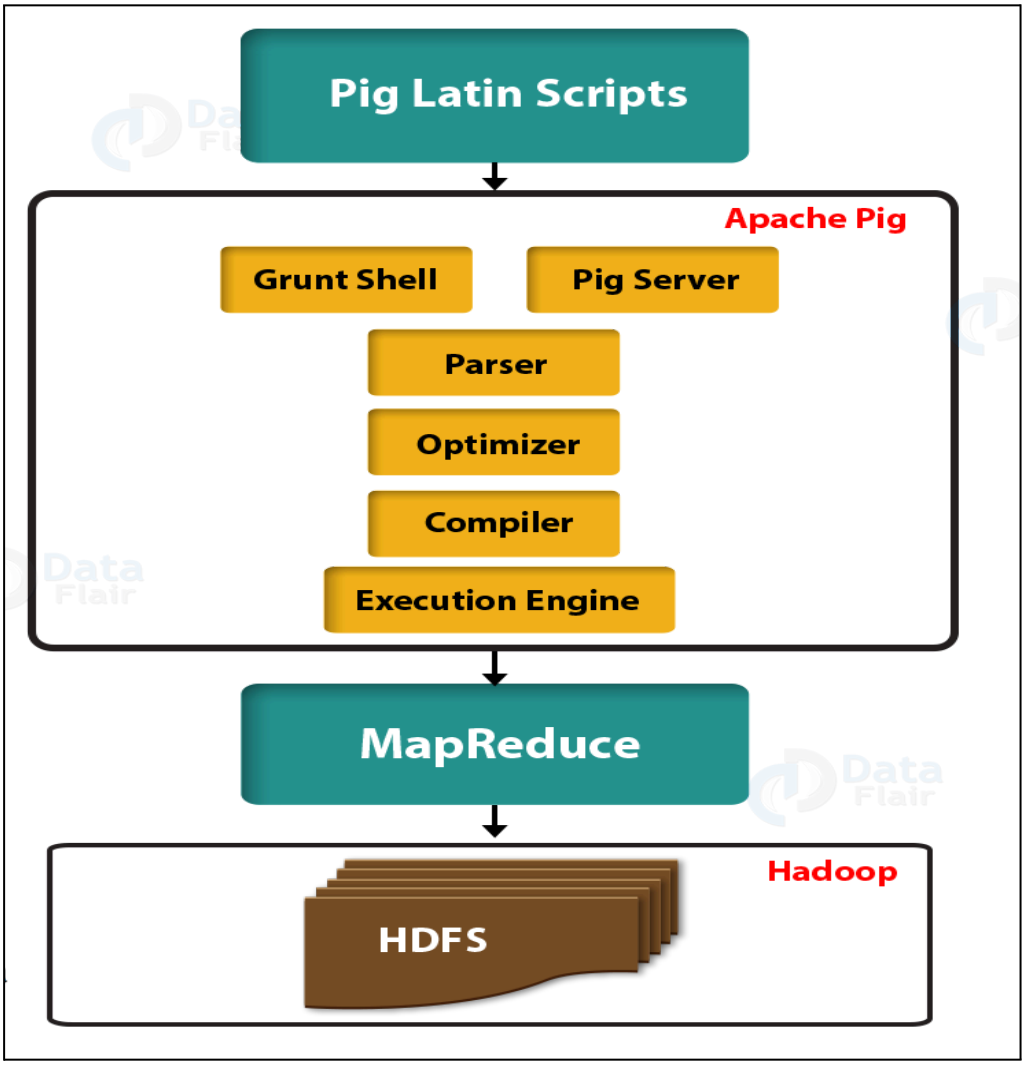
EXPERIMENT NO. 6

Aim: Execute PIG built-in commands and run pig scripts on HDFS

Theory:

1. Explain the Working of a Pig with Architecture.



Practical Execution

1. Enter the pig command localhost.

```
[training@localhost ~]$ pig
```

2. Fs: This will list all the files in the HDFS

```
grunt> fs  -ls
```

3. We can invoke any shell commands using sh

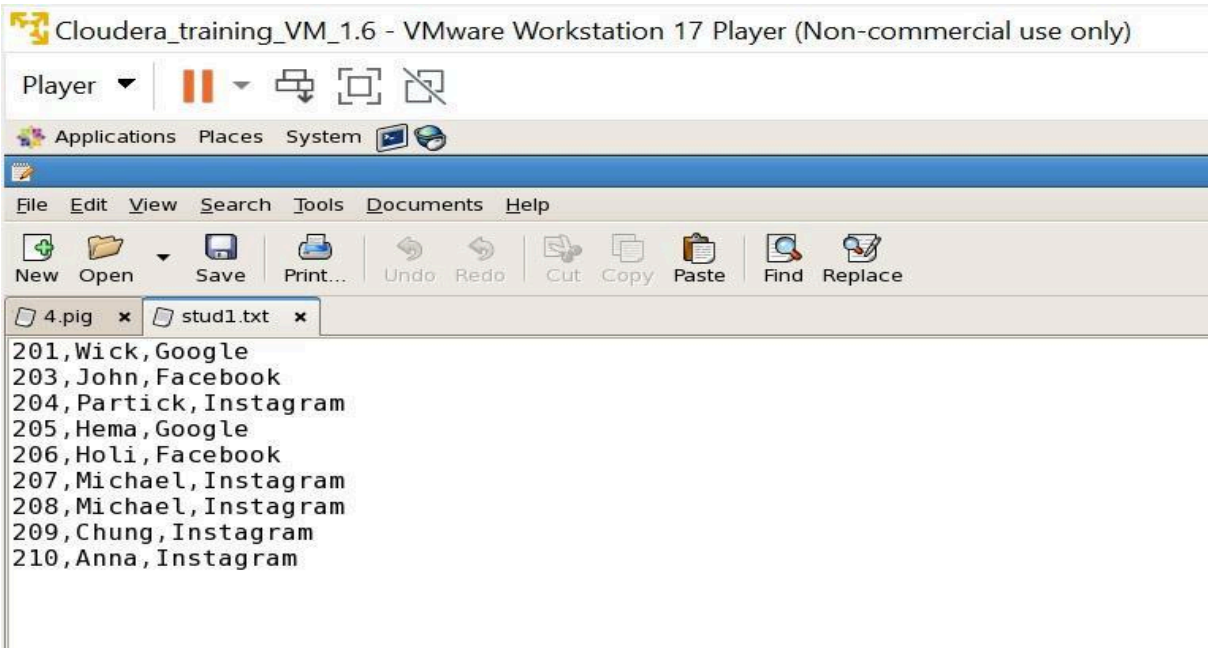
```
grunt> fs –ls
grunt> fs –ls   /
```

4. Create a folder on Hdfs

```
grunt> fs -mkdir   /aids
```

5. create text/csv file on local user account and enter employee's details

```
grunt> sh gedit stud1.txt
```



6. Copy stud1.txt file on HDFS

```
grunt>  fs -put   /home/training/stud1.txt  /aids/
```

7. Check that the file has been copied properly.

```
grunt> fs -cat   /aids/stud1.txt
```

8. load the stud1.txt to pig using the load command and store the result in student Variable

```
grunt> student = LOAD   '/aids/stud1.txt'   USING PigStorage(',')  as
(id: int ,name:chararray ,Company: chararray);
```

9. Display the result using the dump operator.

```
dump student;
```

10. You may create one pig Script file and name it yukt.pig

```
sh gedit yukt.pig
```

In that file Write code as given below

```
 student = LOAD   '/aids/stud1.txt'   USING PigStorage(',')  as
(id:int,name:chararray,Company:chararray);
```

```
dump student;
```

11. Save that file.

12.go on pig prompt and execute that file 2.pig

```
grunt> exec   /home/training/2.pig
        OR
grunt> run   /home/training/2.pig
```

13. Display separate columns from a given file.

```
grunt> ds1 = foreach student generate id, city;
grunt> dump ds1;
```

14. Use order operator

```
grunt> ds2= order student by id desc;
grunt> dump ds2;
```

15. Use group

```
grunt> gr = group ds2 by name ;
grunt> dump gr;
```

16. Use filter operator.

```
grunt> ds3 = filter student by company == 'Instagram';
grunt> dump ds3;
```

17 . Union operator

```
grunt> Result = UNION ds1, student;  (Note: make sure both tables have the same attribute)
grunt> dump Result
```

17. Store files on HDFS using the store command.

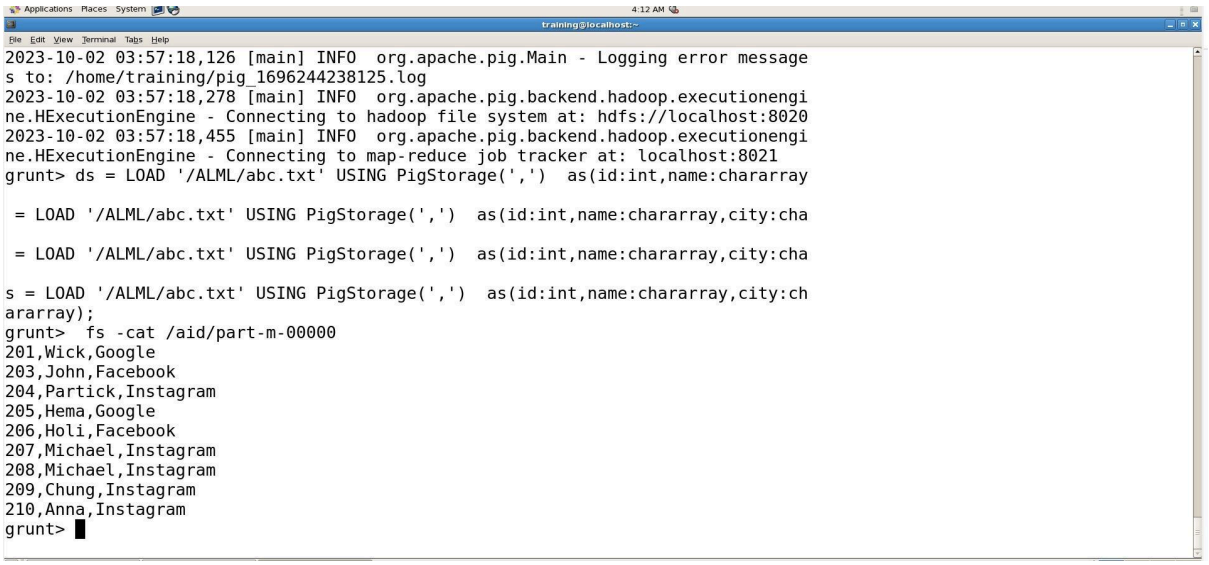
```
grunt> store result into  '/a6/'  USING PigStorage(',');
```

18. Check the file has been copied.

```
grunt> fs -ls   /a6/
```

19.Display the result.

```
grunt> fs -cat   /a6/part-m-00000
```



Conclusion: hence we study how to run the Pig command and Pig script using Pig Hadoop ecosystem.