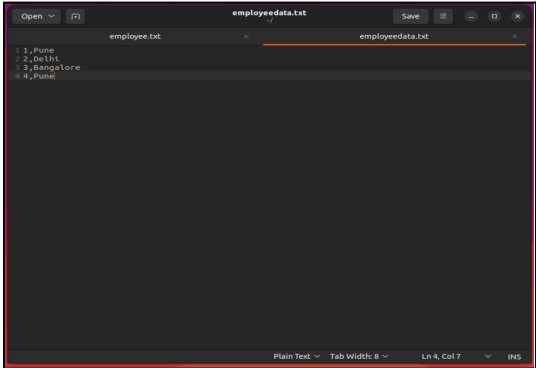
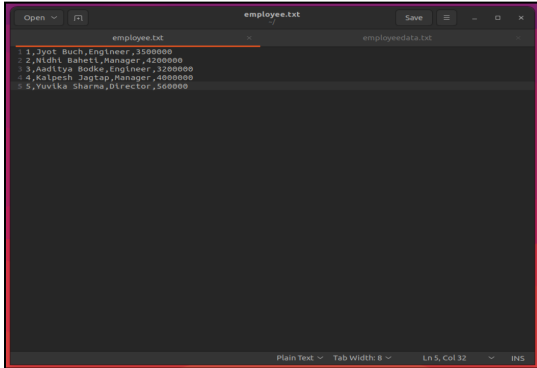


Vishwakarma University, Pune
University Grants Commission (UGC) Approved State Private University

Assignment 5	
Name - Jyot Buch	Roll No. - 09
Date of Submission - 01/05/2023	Class - A (2023 Batch)

1.	<p>Explain the steps required for installation of Apache Pig and then take any Pig relation with schema of: (id: int, name: chararray, address:chararray, age:int, salary:int) You are required to perform Pig operations of Join, Grouping, Filtering and Sorting. Support your answer with screenshots.</p>
	<ol style="list-style-type: none">1. Download the latest version of Apache Pig from the official website (https://pig.apache.org/).2. Extract the downloaded file using the following command: <code>tar -xzf pig-0.16.0.tar.gz</code>3. Move the extracted directory to the desired location using the following command: <code>sudo mv pig-0.16.0 /usr/local/pig</code>4. Set the PIG_HOME environment variable to the Pig installation directory using the following command: <code>export PIG_HOME=/usr/local/pig</code>5. Add the Pig binary directory to the PATH environment variable using the following command: <code>export PATH=\$PATH:\$PIG_HOME/bin</code>6. Verify the installation by running the Pig version command: <code>pig -version</code> <div></div> <p style="text-align: center;">Data to be manipulated</p>

```
jyotbuch@jyotbuch-VirtualBox: ~  
jyotbuch@jyotbuch-VirtualBox: ~  
jyotbuch@jyotbuch-VirtualBox: ~  
Counters:  
Total records written : 4  
Total bytes written : 193620558  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
job_local1583864248_0012  
  
2023-05-01 14:41:22,426 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:41:22,428 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:41:22,429 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:41:22,461 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2023-05-01 14:41:22,475 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address  
2023-05-01 14:41:22,475 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2023-05-01 14:41:22,475 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2023-05-01 14:41:22,499 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1  
2023-05-01 14:41:22,499 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1  
(1,Jyot Buch,Engineer,3500000,1,Pune)  
(2,Nidhi Baheti,Manager,4200000,2,Delhi)  
(3,Aaditya Bodke,Engineer,3200000,3,Bangalore)  
(4,Kalpesh Jagtap,Manager,4000000,4,Pune)  
grunt>
```

Output for JOIN clause

```
jyotbuch@jyotbuch-VirtualBox: ~  
jyotbuch@jyotbuch-VirtualBox: ~  
jyotbuch@jyotbuch-VirtualBox: ~  
35361"  
Counters:  
Total records written : 3  
Total bytes written : 118323616  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
job_local679725602_0011  
  
2023-05-01 14:39:38,284 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:39:38,288 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:39:38,291 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:39:38,325 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2023-05-01 14:39:38,325 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address  
2023-05-01 14:39:38,326 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2023-05-01 14:39:38,326 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2023-05-01 14:39:38,344 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1  
2023-05-01 14:39:38,345 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1  
(Manager,{{(4,Kalpesh Jagtap,Manager,4000000),(2,Nidhi Baheti,Manager,4200000)}})  
(Director,{{(5,Yuvika Sharma,Director,5600000)}})  
(Engineer,{{(3,Aaditya Bodke,Engineer,3200000),(1,Jyot Buch,Engineer,3500000)}})  
grunt>
```

Output using GROUP BY clause

```
jyotbuch@jyotbuch-VirtualBox: ~  
jyotbuch@jyotbuch-VirtualBox: ~  
Successfully stored 2 records (161350300 bytes) in: "hdfs://localhost:54310/tmp/temp-1560251009/tmp2022932080"  
Counters:  
Total records written : 2  
Total bytes written : 161350300  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
Job DAG:  
job_local1441328994_0010  
2023-05-01 14:37:38,183 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:37:38,205 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:37:38,221 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:37:38,224 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2023-05-01 14:37:38,254 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.jobtracker.address is deprecated. Instead, use mapreduce.jobtracker.address  
2023-05-01 14:37:38,263 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2023-05-01 14:37:38,264 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2023-05-01 14:37:38,283 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1  
2023-05-01 14:37:38,283 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1  
(1,Jyot Buch,Engineer,3500000,1,Pune)  
(4,Kalpesh Jagtap,Manager,4000000,4,Pune)  
grunt>
```

Output using FILTER BY clause

```
jyotbuch@jyotbuch-VirtualBox: ~  
jyotbuch@jyotbuch-VirtualBox: ~  
ics system already initialized!  
2023-05-01 14:31:08,585 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:31:08,587 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:31:08,616 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:31:08,621 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:31:08,633 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:31:08,645 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:31:08,655 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:31:08,669 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!  
2023-05-01 14:31:08,688 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2023-05-01 14:31:08,705 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.jobtracker.address is deprecated. Instead, use mapreduce.jobtracker.address  
2023-05-01 14:31:08,705 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2023-05-01 14:31:08,706 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2023-05-01 14:31:08,740 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1  
2023-05-01 14:31:08,740 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1  
(5,Yuvika Sharma,Director,5600000)  
(3,Aaditya Bodke,Engineer,3200000)  
(1,Jyot Buch,Engineer,3500000)  
(4,Kalpesh Jagtap,Manager,4000000)  
(2,Nidhi Baheti,Manager,4200000)  
grunt>
```

Output using ORDER BY clause