

Big Data And Hadoop

Session 19 – Assignment 1

Problem Statement:

1. Create a customer_hive table on the top of 'customer' table created in the last session. Calculate the maximum and minimum age of customer from the table.
2. Access the customer hbase table from pig and compute the maximum and minimum age among all the customers along with their corresponding name and id.

Solution:

A. Start HBase and view existing table

Step 1: We first start HBase using the command **start-hbase.sh** as follows:

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ start-hbase.sh  
starting master, logging to /usr/local/hbase/logs/hbase-acadgild-master-localhost.localdomain.out  
[acadgild@localhost ~]$
```

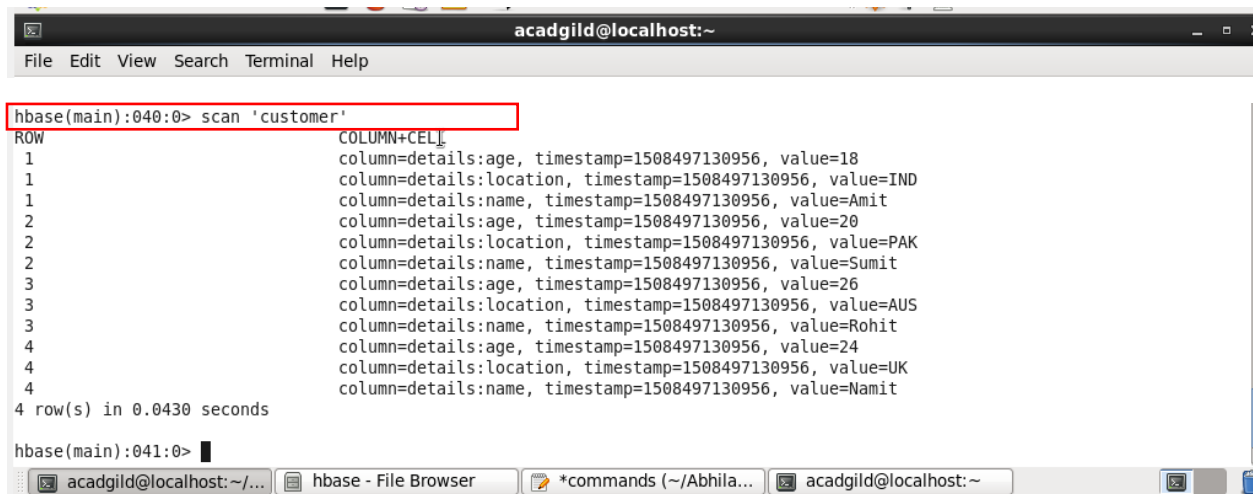
Step 2: Next we start the hbase shell using the command **hbase shell** as follows:

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ hbase shell  
2017-09-24 17:49:57,986 INFO [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.a  
vailable  
HBase Shell; enter 'help<RETURN>' for list of supported commands.  
Type "exit<RETURN>" to leave the HBase Shell  
Version 0.98.14-hadoop2, r4e4aabb93b52f1b0fef6b66edd06ec8923014dec, Tue Aug 25 22:35:44 PDT 2015  
hbase(main):001:0>
```

Step 3: Using **list** command, we can see the table **customer** appeared in the list.

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
hbase(main):002:0* list  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.  
2017-10-20 15:47:33,760 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using b  
uilt-in java classes where applicable  
0 row(s) in 2.4930 seconds  
  
TABLE  
clicks  
customer  
htest  
3 row(s) in 0.0190 seconds  
  
=> ["clicks", "customer", "htest"]  
hbase(main):003:0>
```

Step 4: Data in this table can be viewed using the **scan** command as follows:

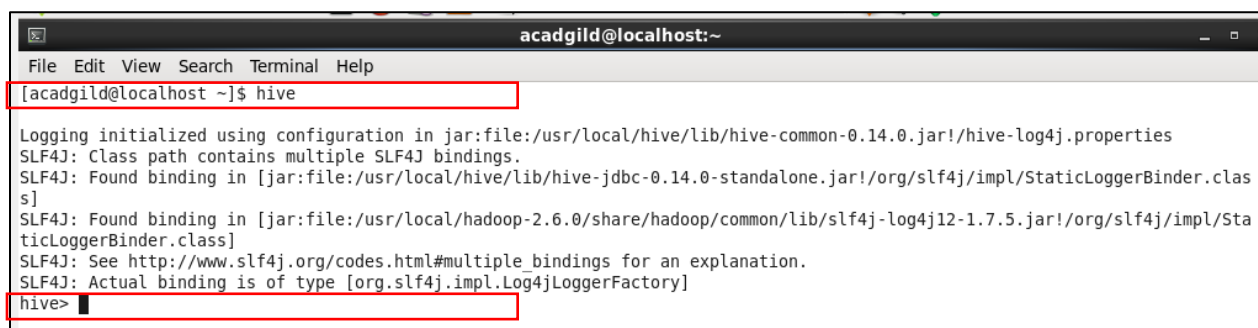


```
acadgild@localhost:~  
File Edit View Search Terminal Help  
hbase(main):040:0> scan 'customer'  
ROW COLUMN+CELL  
1 column=details:age, timestamp=1508497130956, value=18  
1 column=details:location, timestamp=1508497130956, value=IND  
1 column=details:name, timestamp=1508497130956, value=Amit  
2 column=details:age, timestamp=1508497130956, value=20  
2 column=details:location, timestamp=1508497130956, value=PAK  
2 column=details:name, timestamp=1508497130956, value=Sumit  
3 column=details:age, timestamp=1508497130956, value=26  
3 column=details:location, timestamp=1508497130956, value=AUS  
3 column=details:name, timestamp=1508497130956, value=Rohit  
4 column=details:age, timestamp=1508497130956, value=24  
4 column=details:location, timestamp=1508497130956, value=UK  
4 column=details:name, timestamp=1508497130956, value=Namit  
4 row(s) in 0.0430 seconds  
hbase(main):041:0>
```

B. Create a customer_hive table on the top of 'customer' table created in the last session.

Calculate the maximum and minimum age of customer from the table.

Step1: We start hive command line interface as follows:



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ hive  
Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-0.14.0.jar!/hive-log4j.properties  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/hive-jdbc-0.14.0-standalone.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.  
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]  
hive>
```

Step 2: Create an external table in hive using the command below

```
create external table  
customer  
(  
    id int,name String,age int,location String  
)  
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'  
with serdeproperties  
("hbase.columns.mapping"=":key, details:name, details:age, details:location")  
tblproperties("hbase.table.name"="customer");
```

Here we have used HiveStorageHandler. HiveStorageHandler is the primary interface Hive uses to connect with NoSQL databases

such as HBase, Cassandra, and others. If we use HiveStorageHandler we can create hive table which will point to Hbase table and we can query the data in sql style. For every single column family, one Hive table has to be created.

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
hive> create external table customer(id int, name String, age int, location String) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' with serdeproperties("hbase.columns.mapping"=":key,details:name,details:age,details:location")  
> TBLPROPERTIES("hbase.table.name"="customer");  
OK  
Time taken: 2.406 seconds  
hive>
```

Step 3: After execution the above command, table in hive is created. Its schema can be seen using the **describe** command as follows:

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
hive> describe customer;  
OK  
id                int                from deserializer  
name              string             from deserializer  
age               int                from deserializer  
location          string             from deserializer  
Time taken: 0.22 seconds, Fetched: 4 row(s)  
hive>
```

Step 4: Data that got inserted can be seen using **select** query as follows:

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
hive> select * from customer;  
OK  
1      Amit      18      IND  
2      Sumit     20      PAK  
3      Rohit     26      AUS  
4      Namit     24      UK  
Time taken: 0.666 seconds, Fetched: 4 row(s)  
hive>
```

Step 5: Now to find the customer with minimum age, we use the following query:

*Select * from customer order by age limit 1;*

Here, we order the records by age using **ORDER BY**. Then we fetch only one record, hence used **LIMIT** clause.

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
hive> select * from customer order by age limit 1;  
Query ID = acadgild_20171108221212_24f54aed-341c-41e4-9179-6c36cfd39fa  
Total jobs = 1  
Launching Job 1 out of 1  
Number of reduce tasks determined at compile time: 1  
In order to change the average load for a reducer (in bytes):  
  set hive.exec.reducers.bytes.per.reducer=<number>  
In order to limit the maximum number of reducers:  
  set hive.exec.reducers.max=<number>  
In order to set a constant number of reducers:  
  set mapreduce.job.reduces=<number>  
Starting Job = job_1510157240052_0001, Tracking URL = http://localhost:8088/proxy/application_1510157240052_0001/  
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1510157240052_0001  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2017-11-08 22:13:15,361 Stage-1 map = 0%, reduce = 0%  
2017-11-08 22:13:24,458 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.25 sec  
2017-11-08 22:13:34,206 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.73 sec  
MapReduce Total cumulative CPU time: 6 seconds 730 msec  
Ended Job = job_1510157240052_0001  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.73 sec HDFS Read: 249 HDFS Write: 14 SUCCESS  
Total MapReduce CPU Time Spent: 6 seconds 730 msec  
OK  
1      Amit      18      IND  
Time taken: 40.14 seconds, Fetched: 1 row(s)  
hive>
```

We get the output as shown in the above screenshot.

Step 6: Now to find the customer with maximum age, we use the following query:

*Select * from customer order by age DESC limit 1;*

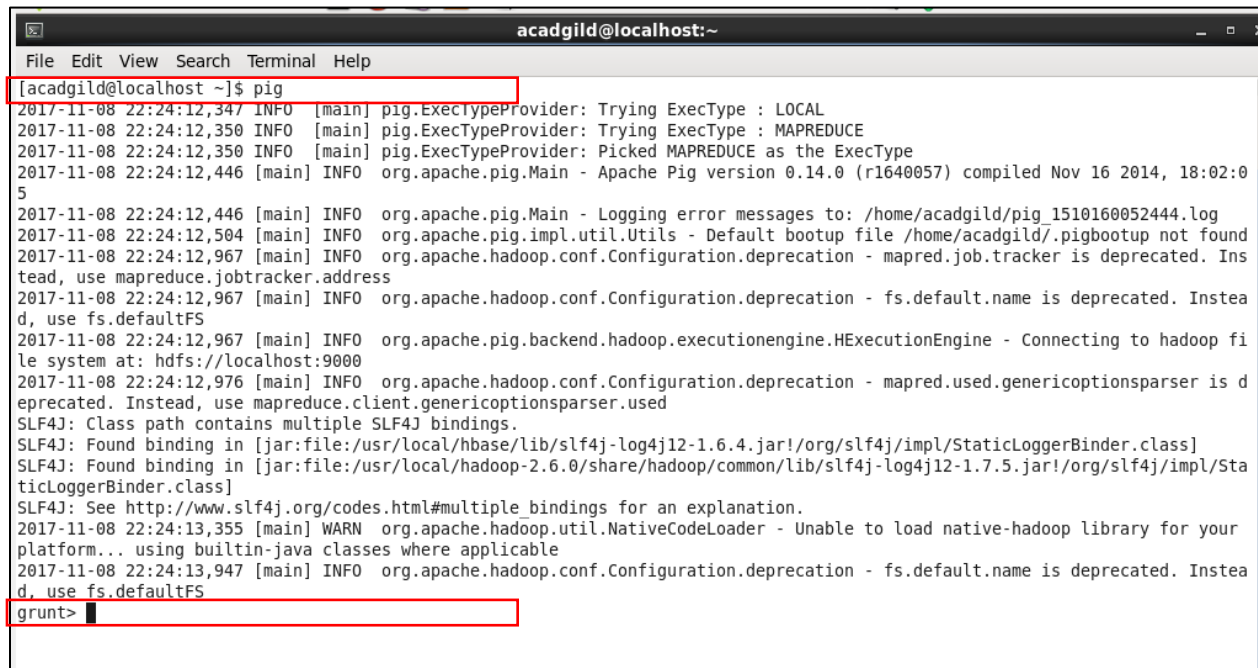
Here, we order the records by age using **ORDER BY**. Also, to get records in descending order, we use the key word **DESC**. Then we fetch only one record, hence used **LIMIT** clause.

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
hive> select * from customer order by age DESC limit 1;  
Query ID = acadgild_20171108221414_401759cd-4ae8-4e18-b9b1-fc5262db19b8  
Total jobs = 1  
Launching Job 1 out of 1  
Number of reduce tasks determined at compile time: 1  
In order to change the average load for a reducer (in bytes):  
  set hive.exec.reducers.bytes.per.reducer=<number>  
In order to limit the maximum number of reducers:  
  set hive.exec.reducers.max=<number>  
In order to set a constant number of reducers:  
  set mapreduce.job.reduces=<number>  
Starting Job = job_1510157240052_0002, Tracking URL = http://localhost:8088/proxy/application_1510157240052_0002/  
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1510157240052_0002  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2017-11-08 22:14:32,612 Stage-1 map = 0%, reduce = 0%  
2017-11-08 22:14:42,358 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.51 sec  
2017-11-08 22:14:49,915 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.91 sec  
MapReduce Total cumulative CPU time: 6 seconds 910 msec  
Ended Job = job_1510157240052_0002  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.91 sec HDFS Read: 249 HDFS Write: 15 SUCCESS  
Total MapReduce CPU Time Spent: 6 seconds 910 msec  
OK  
3      Rohit     26      AUS  
Time taken: 30.643 seconds, Fetched: 1 row(s)  
hive>
```

We get the output as shown in the above screenshot.

C. Access the customer hbase table from pig and compute the maximum and minimum age among all the customers along with their corresponding name and id.

Step 1: We first start pig command line interface as follows:



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ pig  
2017-11-08 22:24:12,347 INFO [main] pig.ExecTypeProvider: Trying ExecType : LOCAL  
2017-11-08 22:24:12,350 INFO [main] pig.ExecTypeProvider: Trying ExecType : MAPREDUCE  
2017-11-08 22:24:12,350 INFO [main] pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType  
2017-11-08 22:24:12,446 [main] INFO org.apache.pig.Main - Apache Pig version 0.14.0 (r1640057) compiled Nov 16 2014, 18:02:05  
2017-11-08 22:24:12,446 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/pig_1510160052444.log  
2017-11-08 22:24:12,504 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/acadgild/.pigbootup not found  
2017-11-08 22:24:12,967 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address  
2017-11-08 22:24:12,967 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2017-11-08 22:24:12,967 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000  
2017-11-08 22:24:12,976 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.used.genericoptionsparser is deprecated. Instead, use mapreduce.client.genericoptionsparser.used  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.  
2017-11-08 22:24:13,355 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
2017-11-08 22:24:13,947 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt>
```

Step 2: Next, we load data from HBase into pig using the command below:

```
data = LOAD 'hbase://customer' USING  
org.apache.pig.backend.hadoop.hbase.HBaseStorage('details:* ', '-loadKey true') as (id:INT,  
details:MAP[]);
```

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
grunt> data = LOAD 'hbase://customer' USING org.apache.pig.backend.hadoop.hbase.HBaseStorage('details:*','-loadKey true') as  
(id:int,details:MAP[]);  
2017-11-08 22:27:01,544 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.persist.jobstatus.  
hours is deprecated. Instead, use mapreduce.jobtracker.persist.jobstatus.hours  
2017-11-08 22:27:01,544 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.heartbeats.in.second is deprec  
ated. Instead, use mapreduce.jobtracker.heartbeats.in.second  
2017-11-08 22:27:01,544 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - jobclient.completion.poll.interval is  
deprecated. Instead, use mapreduce.client.completion.pollinterval  
2017-11-08 22:27:01,544 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.tasktracker.tasks.sleep-time-be  
fore-sigkill is deprecated. Instead, use mapreduce.tasktracker.tasks.sleep-time-before-sigkill  
2017-11-08 22:27:01,544 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is de  
precated. Instead, use mapreduce.jobtracker.http.address  
2017-11-08 22:27:01,544 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - dfs.replication.considerLoad is depre  
cated. Instead, use dfs.namenode.replication.considerLoad  
2017-11-08 22:27:01,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - dfs.block.size is deprecated. Instead  
, use dfs.blocksize  
2017-11-08 22:27:01,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - dfs.permissions is deprecated. Instea  
d, use dfs.permissions.enabled  
2017-11-08 22:27:01,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - topology.node.switch.mapping.impl is  
deprecated. Instead, use net.topology.node.switch.mapping.impl  
2017-11-08 22:27:01,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - dfs.access.time.precision is deprecate  
d. Instead, use dfs.namenode.access-time.precision  
2017-11-08 22:27:01,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.skip.map.max.skip.records is d  
eprecated. Instead, use mapreduce.map.skip.maxrecords  
2017-11-08 22:27:01,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is deprecate  
d. Instead, use mapreduce.client.submit.file.replication  
2017-11-08 22:27:01,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - dfs.https.client.keystore.resource is  
deprecated. Instead, use dfs.client.https.keystore.resource  
2017-11-08 22:27:01,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - job.end.retry.attempts is deprecated.
```

Step 3: After the execution of the above command, data got loaded into pig. This can be seen using **scan** command. On executing the command **dump**, we get the following output:


```
acadgild@localhost:~  
File Edit View Search Terminal Help  
Total records written : 0  
Total bytes written : 0  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
job_1510157240052_0003  
  
2017-11-08 22:28:14,857 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032  
2017-11-08 22:28:14,873 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp  
plicationStatus=SUCCEEDED. Redirecting to job history server  
2017-11-08 22:28:14,981 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032  
2017-11-08 22:28:14,991 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp  
plicationStatus=SUCCEEDED. Redirecting to job history server  
2017-11-08 22:28:15,097 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032  
2017-11-08 22:28:15,121 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp  
plicationStatus=SUCCEEDED. Redirecting to job history server  
2017-11-08 22:28:15,332 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable  
to retrieve job to compute warning aggregation.  
2017-11-08 22:28:15,333 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success  
!  
2017-11-08 22:28:15,351 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea  
d, use fs.defaultFS  
2017-11-08 22:28:15,351 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is depre  
cated. Instead, use mapreduce.job.counters.max  
2017-11-08 22:28:15,360 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated.  
Instead, use dfs.bytes-per-checksum  
2017-11-08 22:28:15,369 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g  
enerate code.  
2017-11-08 22:28:15,447 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2017-11-08 22:28:15,448 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro  
cess : 1  
(1,[name#Amit,location#IND,age#18])  
(2,[name#Sumit,location#PAK,age#20])  
(3,[name#Rohit,location#AUS,age#26])  
(4,[name#Nimit,location#UK,age#24])  
grunt>
```

Step 4: Now to find the record with minimum age, we have a pig script that can be run as follows:

```
acadgild@localhost:~/Abhilasha  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ cd Abhilasha/  
[acadgild@localhost Abhilasha]$ pig min.pig  
2017-11-08 22:51:18,763 INFO [main] pig.ExecTypeProvider: Trying ExecType : LOCAL  
2017-11-08 22:51:18,765 INFO [main] pig.ExecTypeProvider: Trying ExecType : MAPREDUCE  
2017-11-08 22:51:18,765 INFO [main] pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType  
2017-11-08 22:51:18,845 [main] INFO org.apache.pig.Main - Apache Pig version 0.14.0 (r1640057) compiled Nov 16 2014, 18:02:0  
5  
2017-11-08 22:51:18,845 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Abhilasha/pig_1510161678  
843.log  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.  
2017-11-08 22:51:19,146 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your  
platform... using builtin-java classes where applicable  
2017-11-08 22:51:19,429 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/acadgild/.pigbootup not found  
2017-11-08 22:51:19,619 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Ins  
tead, use mapreduce.jobtracker.address  
2017-11-08 22:51:19,619 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea  
d, use fs.defaultFS  
2017-11-08 22:51:19,620 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop fi  
le system at: hdfs://localhost:9000  
2017-11-08 22:51:19,627 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.used.genericoptionsparser is d  
eprecated. Instead, use mapreduce.client.genericoptionsparser.used
```

The steps used are:

- a. *data = LOAD 'hbase://customer' USING org.apache.pig.backend.hadoop.hbase.HBaseStorage('details:* ', '-loadKey true') as (id:INT, details:MAP[]);*

This is to load data from HBase to pig as mentioned in the previous step.

- b. *record = FOREACH data GENERATE id,details#'name' AS name,details#'age' AS age;*

Here, we have created alias for the columns we need, for our ease.

- c. *orderedData = ORDER record BY age;*
Here, we arrange the records in ascending order of the age.

- d. *limitedData = LIMIT orderedData 1;*
Here, we limit the data to only 1 record.

- e. *dump limitedData;*
This is to dump data on console.

The output of this script execution is as follows:


```
acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help
2017-11-08 22:53:12,447 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-11-08 22:53:12,452 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2017-11-08 22:53:12,524 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-11-08 22:53:12,528 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2017-11-08 22:53:12,595 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-11-08 22:53:12,598 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2017-11-08 22:53:12,665 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable
to retrieve job to compute warning aggregation.
2017-11-08 22:53:12,666 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-11-08 22:53:12,670 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2017-11-08 22:53:12,738 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-11-08 22:53:12,744 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2017-11-08 22:53:12,811 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-11-08 22:53:12,819 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2017-11-08 22:53:12,884 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable
to retrieve job to compute warning aggregation.
2017-11-08 22:53:12,884 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2017-11-08 22:53:12,888 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2017-11-08 22:53:12,888 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is depre
cated. Instead, use mapreduce.job.counters.max
2017-11-08 22:53:12,889 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated.
Instead, use dfs.bytes-per-checksum
2017-11-08 22:53:12,889 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2017-11-08 22:53:12,903 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-08 22:53:12,903 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(1, Amit, 18)
2017-11-08 22:53:13,050 [main] INFO org.apache.pig.Main - Pig script completed in 1 minute, 54 seconds and 354 milliseconds
(114354 ms)
[acadgild@localhost Abhilasha]$
```

Step 5: Now to find the record with maximum age, we have a pig script that can be run as follows:

```
acadgild@localhost:~/Abhilasha
File Edit View Search Terminal Help
[acadgild@localhost Abhilasha]$ pig max.pig
2017-11-08 22:53:50,458 INFO [main] pig.ExecTypeProvider: Trying ExecType : LOCAL
2017-11-08 22:53:50,460 INFO [main] pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2017-11-08 22:53:50,460 INFO [main] pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2017-11-08 22:53:50,524 [main] INFO org.apache.pig.Main - Apache Pig version 0.14.0 (r1640057) compiled Nov 16 2014, 18:02:0
5
2017-11-08 22:53:50,524 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Abhilasha/pig_1510161830
523.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/Sta
ticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2017-11-08 22:53:50,844 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
2017-11-08 22:53:51,103 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/acadgild/.pigbootup not found
2017-11-08 22:53:51,300 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Ins
tead, use mapreduce.jobtracker.address
2017-11-08 22:53:51,300 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2017-11-08 22:53:51,302 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop fi
le system at: hdfs://localhost:9000
2017-11-08 22:53:51,310 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.used.genericoptionsparser is d
eprecated. Instead, use mapreduce.client.genericoptionsparser.used
```

The steps used are:

- a. *data = LOAD 'hbase://customer' USING
org.apache.pig.backend.hadoop.hbase.HBaseStorage('details:* ', '-loadKey true') as
(id:INT, details:MAP[]);*

This is to load data from HBase to pig as mentioned in the previous step.

- b. *record = FOREACH data GENERATE id,details#'name' AS name,details#'age' AS age;*

Here, we have created alias for the columns we need, for our ease.

- c. *orderedData = ORDER record BY age DESC;*

Here, we arrange the records in descending order of the age.

- d. *limitedData = LIMIT orderedData 1;*

Here, we limit the data to only 1 record.

- e. *dump limitedData;*

This is to dump data on console.

The output of this script execution is as follows:

```
acagdild@localhost:~/Abhilasha
File Edit View Search Terminal Help
2017-11-08 22:55:52,265 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-11-08 22:55:52,270 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2017-11-08 22:55:52,342 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-11-08 22:55:52,351 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2017-11-08 22:55:52,424 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-11-08 22:55:52,430 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2017-11-08 22:55:52,506 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable
to retrieve job to compute warning aggregation.
2017-11-08 22:55:52,510 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-11-08 22:55:52,514 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2017-11-08 22:55:52,617 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-11-08 22:55:52,625 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2017-11-08 22:55:52,707 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-11-08 22:55:52,711 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalAp
plicationStatus=SUCCEEDED. Redirecting to job history server
2017-11-08 22:55:52,777 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable
to retrieve job to compute warning aggregation.
2017-11-08 22:55:52,777 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2017-11-08 22:55:52,782 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2017-11-08 22:55:52,782 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is depre
cated. Instead, use mapreduce.job.counters.max
2017-11-08 22:55:52,782 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated.
Instead, use dfs.bytes-per-checksum
2017-11-08 22:55:52,783 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2017-11-08 22:55:52,798 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-08 22:55:52,798 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(3.Rohit,26)
2017-11-08 22:55:52,936 [main] INFO org.apache.pig.Main - Pig script completed in 2 minutes, 2 seconds and 522 milliseconds
(122522 ms)
[acagdild@localhost Abhilasha]$
```

attachments - File Bro... min.pig (~/Abhilasha) ... acagdild@localhost:~ acagdild@localhost:~/...