Zeyobron Bigdata Analytics

Project phase-1

Name   : Navaneetha Krishnan

Mobile no   : 9600730899

Project Reporting : Mr - Sai Adithya

Contents

- ➢ **Verify you have customer_total table**

- ➢ **Create customer_src table like customer_total**

- ➢ **Insert into customer_src from customer_total where id>300 and id<350**

- ➢ **Create a password file your edge Node == (echo -n cloudera>passfile)**

- ➢ **Create a directory in Edge Node**

- ➢ **Go inside that directory**

- ➢ **Create sqoop job for customer_src table with id last value 0 along with psw_file**

- ➢ **Execute the Job**

- ➢ **Verify whether data got imported**

- ➢ **AVSC file should have generated locally-- Copy that to HDFS**

- ➢ **Go inside Hive shell and create database prodb**

- ➢ **Now create Hive table on top of raw data with AVSC FILE as reference**

Project phase-1

1) Verify you have customer_total table

Login to MySQL [mysql -uroot -pcloudera]

View all databases [show databases;]

```
mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| add1               |
| cm                 |
| dataa              |
| datap              |
| datap1             |
| db1                |
| firehose           |
| hue                |
| inc                |
| incjob             |
| metastore          |
| mysql              |
| nav                |
| navms              |
| oozie              |
| prac1              |
| prac2              |
| prodb              |
| retail_db          |
```

Select prodb [use prodb;]

View all tables [show tables;]

```
mysql> use prodb;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables;
+----------------+
| Tables_in_prodb |
+----------------+
| customer_total |
+----------------+
1 row in set (0.00 sec)
```

2) Create customer_src table like customer_total

Command to create customer_src table

[create table customer_src(id int(10),username varchar(100),sub_port varchar(100),host varchar(100),date_time varchar(100),hit_count_var_1 varchar(100),hit_count_var_2 varchar(100),hit_count_var_3 varchar(100),timezone varchar(100),method varchar(100),procedure varchar(100),value varchar(100),sub_product varchar(100),web_info varchar(100),status_code varchar(100));   ]

```
mysql> desc customer_total;
+-----------------+--------------+------+-----+---------+-------+
| Field           | Type         | Null | Key | Default | Extra |
+-----------------+--------------+------+-----+---------+-------+
| id              | int(10)      | YES  |     | NULL    |       |
| username        | varchar(100) | YES  |     | NULL    |       |
| sub_port        | varchar(100) | YES  |     | NULL    |       |
| host            | varchar(100) | YES  |     | NULL    |       |
| date_time       | varchar(100) | YES  |     | NULL    |       |
| hit_count_val_1 | varchar(100) | YES  |     | NULL    |       |
| hit_count_val_2 | varchar(100) | YES  |     | NULL    |       |
| hit_count_val_3 | varchar(100) | YES  |     | NULL    |       |
| timezone        | varchar(100) | YES  |     | NULL    |       |
| method          | varchar(100) | YES  |     | NULL    |       |
| procedure       | varchar(100) | YES  |     | NULL    |       |
| value           | varchar(100) | YES  |     | NULL    |       |
| sub_product     | varchar(100) | YES  |     | NULL    |       |
| web_info        | varchar(100) | YES  |     | NULL    |       |
| status_code     | varchar(100) | YES  |     | NULL    |       |
+-----------------+--------------+------+-----+---------+-------+
```

View all tables [show tables;]

```
mysql> show tables;
+------------------+
| Tables_in_prodb  |
+------------------+
| customer_src     |
| customer_total   |
+------------------+
2 rows in set (0.00 sec)
```

3) Insert into customer_src from customer_total where id>300 and id<350

*View all tables [insert into customer_src select * from customer_total where id>300 and id<350;]*

4) Create a password file your edge Node == (echo -n cloudera>passfile)

Create edge node pass file [echo -n Aditya908>/home/itv004084/apassfile]

5) Create a directory in Edge Node

Directory in edge node [mkdir pharsedir]

6) Go inside that directory

Directory inside edge node [cd pharsedir]

7) Create sqoop job for customer_src table with id last value 0 along with password file

sqoop job delete command [sqoop job --delete pjob]

sqoop job create   [sqoop job --create pjob -- import --connect jdbc:mysql://localhost/datap --username root --password-file file:///home/cloudera/pfile -- m 1 --table intab --target-dir /user/cloudera/datap --incremental append --check-column id --last-value 0 --as-avrodatafile]

```
[itv004084@g02 ~]$ sqoop job --create mobilejob -- import --connect jdbc:mysql://zeyod2.czgbk94pezaa.ap-south-1.rds.amazonaws.com/itv004084 --username root --password-f
ile file:///home/itv004084/pfile --m 1 --table itv004084 --target-dir /user/itv004084/datap --incremental append --check-column id --last-value 0
Warning: /opt/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /opt/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/hadoop-3.3.0/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2022-11-03 02:53:34,215 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
[itv004084@g02 ~]$
```

8) Execute the Job

sqoop job execute command [sqoop job --exec pjob]

```
[itv004084@g02 ~]$ sqoop job --list
Warning: /opt/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /opt/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/hadoop-3.3.0/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2022-11-03 02:56:12,229 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
Available jobs:
  iincjob
  incjob
  injob
  injobs
  itv_job
  itvjob
  mobilejob
  naveenjob
  ncjob
  newj
  pjob
  psjob
[itv004084@g02 ~]$
```

9) Verify whether data got imported

Data importing command [hadoop fs -ls /user/itv004084/customer_data]

```
drwx------   - itv004084 supergroup          0 2022-10-16 02:17 /user/itv004084/.Trash
drwxr-xr-x   - itv004084 supergroup          0 2022-11-03 02:59 /user/itv004084/_sqoop
drwxr-xr-x   - itv004084 supergroup          0 2022-10-17 13:41 /user/itv004084/aavdirr
drwxr-xr-x   - itv004084 supergroup          0 2022-10-03 02:09 /user/itv004084/adata
drwxr-xr-x   - itv004084 supergroup          0 2022-10-30 03:38 /user/itv004084/adir
-rw-r--r--   3 itv004084 supergroup        508 2022-10-30 03:28 /user/itv004084/atab.avsc
drwxr-xr-x   - itv004084 supergroup          0 2022-10-17 11:39 /user/itv004084/avdir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-17 13:37 /user/itv004084/avdirr
drwxr-xr-x   - itv004084 supergroup          0 2022-09-24 20:39 /user/itv004084/avrodata
drwxr-xr-x   - itv004084 supergroup          0 2022-10-30 03:49 /user/itv004084/avscdirpro
drwxr-xr-x   - itv004084 supergroup          0 2022-09-22 00:58 /user/itv004084/chdata
drwxr-xr-x   - itv004084 supergroup          0 2022-09-25 11:46 /user/itv004084/chennaidir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-16 02:43 /user/itv004084/confirm
drwxr-xr-x   - itv004084 supergroup          0 2022-10-30 03:47 /user/itv004084/customer_stage_loc
-rw-r--r--   3 itv004084 supergroup         13 2022-10-16 02:20 /user/itv004084/data.csv
drwxr-xr-x   - itv004084 supergroup          0 2022-09-22 00:51 /user/itv004084/datadir
drwxr-xr-x   - itv004084 supergroup          0 2022-11-03 02:59 /user/itv004084/datap
drwxr-xr-x   - itv004084 supergroup          0 2022-10-16 00:54 /user/itv004084/ddir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-06 12:51 /user/itv004084/diinimport
drwxr-xr-x   - itv004084 supergroup          0 2022-10-04 01:38 /user/itv004084/dinimport
drwxr-xr-x   - itv004084 supergroup          0 2022-10-08 03:11 /user/itv004084/djinr
drwxr-xr-x   - itv004084 supergroup          0 2022-10-31 10:51 /user/itv004084/dyndir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-28 02:53 /user/itv004084/edir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-16 02:38 /user/itv004084/exdir
drwxr-xr-x   - itv004084 supergroup          0 2022-09-19 00:15 /user/itv004084/h_zzdir
drwxr-xr-x   - itv004084 supergroup          0 2022-09-26 12:59 /user/itv004084/indir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-06 14:03 /user/itv004084/jobdir
drwxr-xr-x   - itv004084 supergroup          0 2022-09-26 12:18 /user/itv004084/joindir
drwxr-xr-x   - itv004084 supergroup          0 2022-09-26 16:09 /user/itv004084/naveen
drwxr-xr-x   - itv004084 supergroup          0 2022-09-22 00:26 /user/itv004084/newimp
drwxr-xr-x   - itv004084 supergroup          0 2022-09-26 15:38 /user/itv004084/pardir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-02 00:52 /user/itv004084/pdata
drwxr-xr-x   - itv004084 supergroup          0 2022-10-22 01:03 /user/itv004084/pdir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-16 01:05 /user/itv004084/rdir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-22 23:41 /user/itv004084/sdir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-22 23:41 /user/itv004084/sidir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-27 12:38 /user/itv004084/srcd
drwxr-xr-x   - itv004084 supergroup          0 2022-10-30 03:41 /user/itv004084/stdir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-31 10:59 /user/itv004084/subdyndir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-31 10:54 /user/itv004084/subsrcdir
drwxr-xr-x   - itv004084 supergroup          0 2022-10-27 12:38 /user/itv004084/tard
drwxr-xr-x   - itv004084 supergroup          0 2022-10-31 10:31 /user/itv004084/warehouse
-rw-r--r--   3 itv004084 supergroup          9 2022-09-19 01:29 /user/itv004084/zfile
[itv004084@g02 ~]$
```

10) AVSC file should have generated locally-- Copy that to HDFS

Create one directory command [hadoop fs -mkdir /user/cloudera/projectavsc]

Move file edge node command [hadoop fs -put /home/cloudera/phase1/customer_src.avsc /user/cloudera/projectavsc]

```
[cloudera@quickstart phase1]$ ls -l
total 44
-rw-rw-r-- 1 cloudera cloudera  2280 Oct 29 13:24 customer_src.avsc
-rw-rw-r-- 1 cloudera cloudera 38088 Oct 29 13:24 customer_src.java
[cloudera@quickstart phase1]$
```

11) Go inside Hive shell and create database prodb;

Hive shell command [hive *], [* create database prodb; *], [* use prodb; *]*

```
[cloudera@quickstart ~]$ hadoop fs -ls
Found 25 items
drwxr-xr-x   - cloudera cloudera          0 2022-10-29 13:26 _sqoop
drwxr-xr-x   - cloudera cloudera          0 2022-10-06 12:14 adata
drwxr-xr-x   - cloudera cloudera          0 2022-10-29 10:39 adir
-rw-r--r--   1 cloudera cloudera        508 2022-10-29 10:19 atab.avsc
drwxr-xr-x   - cloudera cloudera          0 2022-09-17 23:09 chdata
drwxr-xr-x   - cloudera cloudera          0 2022-09-23 21:20 chennaidir
drwxr-xr-x   - cloudera cloudera          0 2022-10-29 13:26 customer_data
-rw-r--r--   1 cloudera cloudera      29156 2022-09-16 22:59 data.avro
drwxr-xr-x   - cloudera cloudera          0 2022-09-17 21:59 datadir
```

12) Now create Hive table on top of raw data with AVSC FILE as reference

*Hive raw data command [*create table customer_src row format serde 'org.apache.hadoop.hive.serde2.avro.AvroSerDe' STORED AS AVRO LOCATION '/user/cloudera/customer_data' TBLPROPERTIES ('avro.schema.url'='/user/cloudera/projectavro/customer_src.avsc');*]*

```
307    beautifulbear   06865   664.687.377.602 12/Sep/2012:18:09:29 -0500      18      09      29      -0500   GET   /download/download3.zip 0      /download      M
ozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1; .NET CLR 2.0.50727; .NET CLR 3.0.04506.30; InfoPa      100
308    beautifulbear   01759   668.686.42.61   28/Dec/2011:00:32:33 -0500      00      32      33      -0500   GET   /partners        0      /partners      Mozilla/
4.0 (compatible; MSIE 8.0; Windows NT 6.1; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR      100
309    beautifulbear   01701   14.323.74.653   13/Sep/2012:02:29:30 -0500      02      29      30      -0500   GET   /demo   0        -      Jakarta Commons-HttpClie
nt/3.0-rc4      100
310    beautifulbear   43137   52.43.672.362   04/Jan/2012:12:12:06 -0500      12      12      06      -0500   GET   /docs/doc2.pdf 0      /docs/doc2.pdf Mozilla/
5.0 (Windows; U; Windows NT 5.1; de; rv:1.9.2.3) Gecko/20100401 Firefox/3.6.3 ( .NET CLR 3.5      100
311    beautifulbear   04126   53.378.683.602  14/Sep/2012:13:11:32 -0500      13      11      32      -0500   GET   /feeds/press     0      /feeds/press   Mozilla/
5.0 (Windows; U; Windows NT 6.0; fr; rv:1.9.2.3) Gecko/20100401 Firefox/3.6.3 ( .NET CLR 3.5      100
312    beautifulbear   32800   75.622.36.380   05/Jan/2012:09:12:28 -0500      09      12      28      -0500   GET   /products        0      /products      Mozilla/
5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.1.5) Gecko/20091102 Firefox/3.5.5 GTB5      100
313    beautifulbear   02621   14.610.30.67    18/Sep/2012:08:54:30 -0500      08      54      30      -0500   GET   /product/product3        0      -      Mozilla/
4.0 (compatible; MSIE 8.0; Windows NT 5.2; Trident/4.0; .NET CLR 1.1.4322; .NET CLR 2.0.5072      100
314    beautifulbear   80516   680.330.387.7   05/Jan/2012:12:48:38 -0500      12      48      38      -0500   GET   /product/product3        0      /product/product
3      Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; .NET CLR 1.1.4322)      100
315    beautifulbear   94600   361.631.17.30   20/Sep/2012:23:24:59 -0500      23      24      59      -0500   GET   /demo   0      /demo  Jakarta Commons-HttpClie
nt/3.0-rc4      100
316    beautifulbear   08102   322.6.648.325   05/Jan/2012:14:05:22 -0500      14      05      22      -0500   GET   /download/download3.zip 0      /download/downlo
ad3.zip Mozilla/4.0 (compatible; MSIE 5.00; Windows 98) 100
317    beautifulbear   04731   87.86.81.40     21/Sep/2012:05:54:30 -0500      05      54      30      -0500   GET   /feeds/press     0      -      Mozilla/5.0 (Win
dows; U; Windows NT 6.1; it; rv:1.9.2.3) Gecko/20100401 Firefox/3.6.3 (.NET CLR 3.5.      100
318    beautifulbear   06435   30.672.8.336    06/Jan/2012:20:01:36 -0500      20      01      36      -0500   GET   /home   0      /home  Jakarta Commons-HttpClie
nt/3.0.1       100
319    beautifulbear   06597   367.34.686.631  22/Sep/2012:23:09:30 -0500      23      09      30      -0500   GET   /demo   0        -      Jakarta Commons-HttpClie
nt/3.0-rc4      100
320    beautifulbear   90110   322.6.648.325   06/Jan/2012:20:30:54 -0500      20      30      54      -0500   GET   /download/download3.zip 0      /download/downlo
ad3.zip Mozilla/4.0 (compatible; MSIE 5.00; Windows 98) 100
321    beautifulbear   61561   13.626.82.338   24/Sep/2012:19:11:38 -0500      19      11      38      -0500   GET   /product/product3        0      /product/product
3      Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; Mozilla/4.0 (compatible; MSIE 6.0; W      100
322    beautifulbear   01794   650.47.72.628   06/Jan/2012:21:22:30 -0500      21      22      30      -0500   GET   /misc   0        -      Python-urllib/2.5      1
00
323    beautifulbear   60546   41.631.382.601  25/Sep/2012:22:40:00 -0500      22      40      00      -0500   GET   /feeds/press     0      -      Apple-PubSub/65.
12.1   100
324    beautifulbear   04825   322.6.648.325   07/Jan/2012:06:32:49 -0500      06      32      49      -0500   GET   /download/download3.zip 0      /download/downlo
ad3.zip Mozilla/4.0 (compatible; MSIE 5.00; Windows 98) 100
325    beautifulbear   61449   361.54.46.628   27/Sep/2012:12:07:24 -0500      12      07      24      -0500   GET   /partners        0      /partners      Mozilla/
4.0 (compatible; MSIE 7.0; Windows NT 6.0; Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1      100
326    beautifulbear   40 53   322.6.648.325   07/Jan/2012:09:04:53 -0500      09      04      53      -0500   GET   /download/download3.zip 0      /download/downlo
ad3.zip Mozilla/4.0 (compatible; MSIE 5.00; Windows 98) 100
327    beautifulbear   09264   628.44.58.626   30/Oct/2011:07:26:30 -0500      07      26      30      -0500   GET   /feeds/press     0      -      Apple-PubSub/65.
11     100
328    beautifulbear   21605   322.6.648.325   08/Jan/2012:05:18:12 -0500      05      18      12      -0500   GET   /download/download3.zip 0      /download/downlo
```