Sqoop Export

•••

Transferring data from HDFS to RDBMS

Create table in mysql

In mysql Create a table in banking database.

create database banking;
use banking;

CREATE TABLE card_transactions (transaction_id INT,
card_id BIGINT,
member_id BIGINT,
amount INT,
postcode INT,
pos_id BIGINT,
transaction_dt varchar(255), status varchar(255),
PRIMARY KEY (transaction_id));

Now export the data

Move the card_trans.csv file in /data folder in hdfs & execute the sqoop export command.

```
hadoop fs -mkdir /data
hadoop fs -ls /data
hadoop fs -put Desktop/card_trans.csv /data

sqoop export \
--connect jdbc:mysql://quickstart.cloudera:3306/banking \
--username root \
--password cloudera \
--table card_transactions \
--export-dir /data/card_trans.csv \
--fields-terminated-by ','
```

Sqoop export command Screenshot

```
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hadoop fs -ls /data
Found 2 items
-rw-r--r--    1 cloudera supergroup         19 2020-04-10 15:35 /data/file1.txt
drwxr-xr-x    - cloudera supergroup          0 2020-04-10 15:38 /data/folder1
[cloudera@quickstart ~]$ hadoop fs -put Desktop/card_trans.csv /data
[cloudera@quickstart ~]$ sqoop export \
> --connect jdbc:mysql://quickstart.cloudera:3306/banking \
> --username root \
> --password cloudera \
> --table card_transactions \
> --export-dir /data/card_trans.csv \
> --fields-terminated-by ','
```

We can see the table is populated in mysql

```
mysql> select * from card_transactions limit 10;
+----------------+------------------+--------------+---------+----------+----------------+---------------------+--------+
| transaction_id | card_id          | member_id    | amount  | postcode | pos_id         | transaction_dt      | status |
+----------------+------------------+--------------+---------+----------+----------------+---------------------+--------+
|              1 | 348702330256514  | 37495066290  | 9084849 |    33946 | 614677375609919 | 11-02-2018 00:00:00 | GENUIN |
|              2 | 348702330256514  | 37495066290  |  330148 |    33946 | 614677375609919 | 11-02-2018 00:00:00 | GENUIN |
|              3 | 348702330256514  | 37495066290  |  136052 |    33946 | 614677375609919 | 11-02-2018 00:00:00 | GENUIN |
|              4 | 348702330256514  | 37495066290  | 4310362 |    33946 | 614677375609919 | 11-02-2018 00:00:00 | GENUIN |
|              5 | 348702330256514  | 37495066290  | 9097094 |    33946 | 614677375609919 | 11-02-2018 00:00:00 | GENUIN |
|              6 | 348702330256514  | 37495066290  | 2291118 |    33946 | 614677375609919 | 11-02-2018 00:00:00 | GENUIN |
|              7 | 348702330256514  | 37495066290  | 4900011 |    33946 | 614677375609919 | 11-02-2018 00:00:00 | GENUIN |
|              8 | 348702330256514  | 37495066290  |  633447 |    33946 | 614677375609919 | 11-02-2018 00:00:00 | GENUIN |
|              9 | 348702330256514  | 37495066290  | 6259303 |    33946 | 614677375609919 | 11-02-2018 00:00:00 | GENUIN |
|             10 | 348702330256514  | 37495066290  |  369067 |    33946 | 614677375609919 | 11-02-2018 00:00:00 | GENUIN |
+----------------+------------------+--------------+---------+----------+----------------+---------------------+--------+
10 rows in set (0.00 sec)
```

Sqoop Job Failure example

Refer Document 3.

1. Here we will try to find the reason of the Job failure.

2. Also we need to make sure the any failure should not impact our actual table.

Example of a working export using staging table

Create a Normal Table

```
CREATE TABLE card_transactions (
 card_id BIGINT,
 member_id BIGINT,
 amount INT,
 postcode INT,
 pos_id BIGINT,
 transaction_dt varchar(255),
 status varchar(255),
 PRIMARY KEY (card_id, transaction_dt));
```

Create a Staging Table

```
CREATE TABLE card_transactions_stage (
 card_id BIGINT,
 member_id BIGINT,
 amount INT,
 postcode INT,
 pos_id BIGINT,
 transaction_dt varchar(255),
 status varchar(255),
 PRIMARY KEY (card_id, transaction_dt) );
```

Move the file in HDFS & run Sqoop Export

hadoop fs -put Desktop/**card_transactions_new.csv** /data

```
sqoop export \
--connect jdbc:mysql://quickstart.cloudera:3306/banking \
--username root \
--password cloudera \
--table card_transactions \
--staging-table card_transactions_stage \
--export-dir /data/card_transactions_new.csv \
--fields-terminated-by ','
```

Sqoop export job completed successfully

```
Bytes written=0
20/04/14 16:04:32 INFO mapreduce.ExportJobBase: Transferred 14.9736 KB in 27.2157 s
econds (563.389 bytes/sec)
20/04/14 16:04:32 INFO mapreduce.ExportJobBase: Exported 62 records.
20/04/14 16:04:32 INFO mapreduce.ExportJobBase: Starting to migrate data from stagi
ng table to destination.
20/04/14 16:04:32 INFO manager.SqlManager: Migrated 62 records from `card_transacti
ons_stage` to `card_transactions`
[cloudera@quickstart ~]$
```

We have learnt Sqoop export

Happy Learning!!!

Follow US

Trainer	**Mr. Sumit Mittal**
Phone	**9108179578**
Email	**trendytech.sumit@gmail.com**
Website	**https://trendytech.in/courses/big-data-online-training/**
LinkedIn	**https://www.linkedin.com/in/bigdatabysumit/**
Twitter	**@BigdataBySumit**
Instagram	**bigdatabysumit**
Facebook	**https://www.facebook.com/trendytech.in/**
Youtube	**TrendyTech**