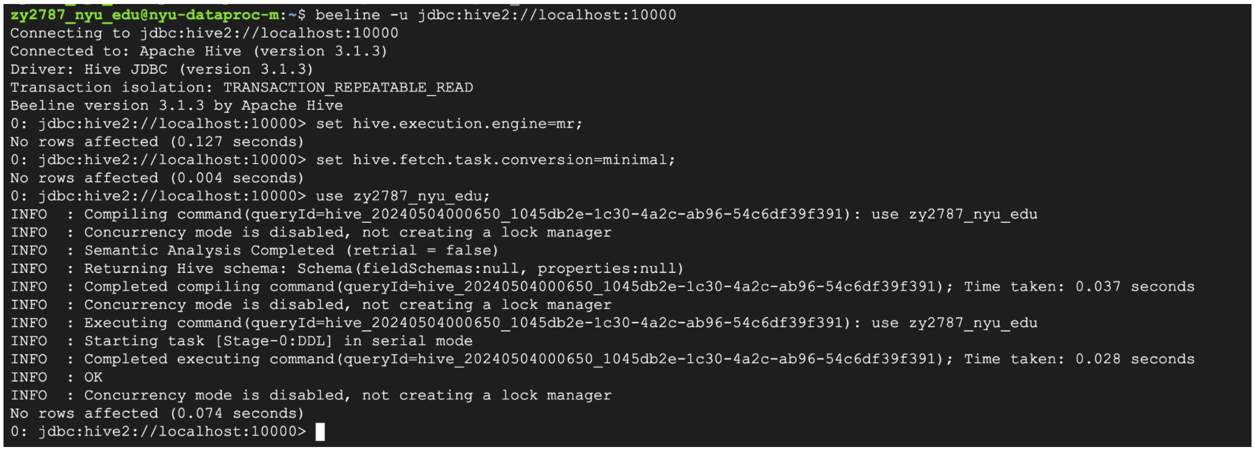
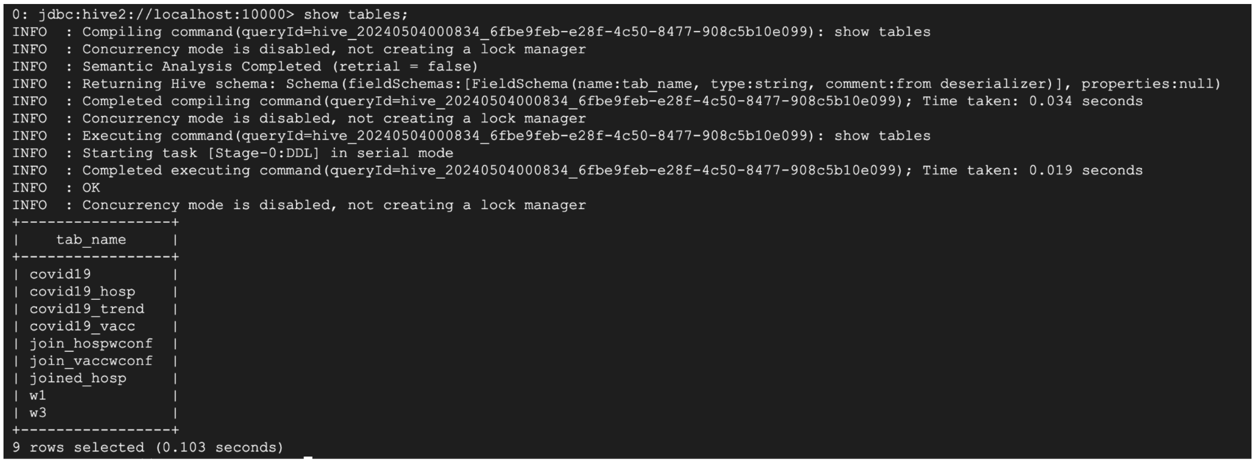
This script is the command that we used for Hive-generating tables.

Connect to Hive:



```
zy2787_nyu_edu@nyu-dataproc-m:~$ beeline -u jdbc:hive2://localhost:10000
Connecting to jdbc:hive2://localhost:10000
Connected to: Apache Hive (version 3.1.3)
Driver: Hive JDBC (version 3.1.3)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.3 by Apache Hive
0: jdbc:hive2://localhost:10000> set hive.execution.engine=mr;
No rows affected (0.127 seconds)
0: jdbc:hive2://localhost:10000> set hive.fetch.task.conversion=minimal;
No rows affected (0.004 seconds)
0: jdbc:hive2://localhost:10000> use zy2787_nyu_edu;
INFO  : Compiling command(queryId=hive_20240504000650_1045db2e-1c30-4a2c-ab96-54c6df39f391): use zy2787_nyu_edu
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20240504000650_1045db2e-1c30-4a2c-ab96-54c6df39f391); Time taken: 0.037 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20240504000650_1045db2e-1c30-4a2c-ab96-54c6df39f391): use zy2787_nyu_edu
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20240504000650_1045db2e-1c30-4a2c-ab96-54c6df39f391); Time taken: 0.028 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
No rows affected (0.074 seconds)
0: jdbc:hive2://localhost:10000>
```

Put Four Cleaned dataset into hive and turn it into 4 tables:



```
0: jdbc:hive2://localhost:10000> show tables;
INFO  : Compiling command(queryId=hive_20240504000834_6fbe9feb-e28f-4c50-8477-908c5b10e099): show tables
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tab_name, type:string, comment:from deserializer)], properties:null)
INFO  : Completed compiling command(queryId=hive_20240504000834_6fbe9feb-e28f-4c50-8477-908c5b10e099); Time taken: 0.034 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20240504000834_6fbe9feb-e28f-4c50-8477-908c5b10e099): show tables
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20240504000834_6fbe9feb-e28f-4c50-8477-908c5b10e099); Time taken: 0.019 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
+-----------------+
|    tab_name     |
+-----------------+
| covid19         |
| covid19_hosp    |
| covid19_trend   |
| covid19_vacc    |
| join_hospwconf  |
| join_vaccwconf  |
| joined_hosp     |
| w1              |
| w3              |
+-----------------+
9 rows selected (0.103 seconds)
```

Generate joined tables from 4 different tables created.
Joint Result 1:
sumconfirmed + Culmulative_Hosp

zy2787_nyu_edu@nyu-dataproc-m:~$ hadoop fs -mkdir join_hospWconfirm

0: jdbc:hive2://localhost:10000> create external table joined_hospwconfirm (dateKey string, location string, sumconfirmed int, sumrecovered int, new_hospitalized_patients int, cumulative_hospitalized_patients int, new_intensive_care_patients int, cumulative_intensive_care_patients int)
. . . . . . . . . . . . . . . . > row format delimited fields terminated by ','
. . . . . . . . . . . . . . . . > location '/user/zy2787_nyu_edu/join_hospWconfirm/‘;

```
Insert into table joined_hospwconfirm
select a.dateKey, a.location, a.sumconfirmed, a.sumrecovered, c.new_hospitalized_patients,
c.cumulative_hospitalized_patients, c.new_intensive_care_patients,
c.cumulative_intensive_care_patients

FROM covid19 a inner join covid19_hosp c on a.dateKey = c.dateKey AND a.location = c.location;
```

Joint result 2:
Confirmed + vaccined

```
zy2787_nyu_edu@nyu-dataproc-m:~$ hadoop fs -mkdir join_vaccWconf
```

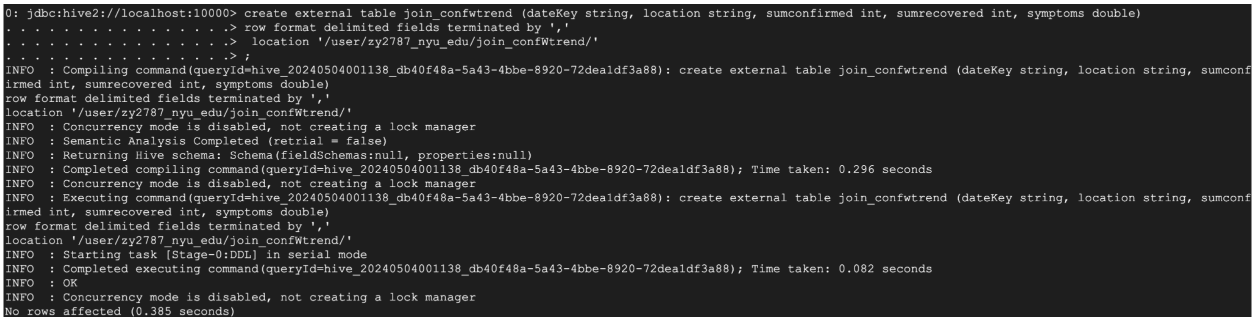
```
0: jdbc:hive2://localhost:10000> create external table join_vaccwconf (dateKey string, location
string, sumconfirmed int, cumulative_persons_fully_vaccinated int)
. . . . . . . . . . . . . . . . > row format delimited fields terminated by ','
. . . . . . . . . . . . . . . . > location '/user/zy2787_nyu_edu/join_vaccWconf/‘;
```

```
Insert into table join_vaccwconf
select a.dateKey, a.location, a.sumconfirmed, b.cumulative_persons_fully_vaccinated
FROM covid19 a inner join covid19_vacc b on a.dateKey = b.dateKey AND a.location = b.location;
```

result 3:
Joint Confirmed + trend

```
zy2787_nyu_edu@nyu-dataproc-m:~$ hadoop fs -mkdir join_confWtrend
```

```
0: jdbc:hive2://localhost:10000> create external table join_confwtrend (dateKey string, location
string, sumconfirmed int, sumrecovered int, symptoms double)
. . . . . . . . . . . . . . . . > row format delimited fields terminated by ','
. . . . . . . . . . . . . . . . > location '/user/zy2787_nyu_edu/join_confWtrend/‘
```



```
0: jdbc:hive2://localhost:10000> create external table join_confwtrend (dateKey string, location string, sumconfirmed int, sumrecovered int, symptoms double)
. . . . . . . . . . . . . . . . .> row format delimited fields terminated by ','
. . . . . . . . . . . . . . . . .>  location '/user/zy2787_nyu_edu/join_confWtrend/'
. . . . . . . . . . . . . . . . .> ;
INFO  : Compiling command(queryId=hive_20240504001138_db40f48a-5a43-4bbe-8920-72dea1df3a88): create external table join_confwtrend (dateKey string, location string, sumconf
irmed int, sumrecovered int, symptoms double)
row format delimited fields terminated by ','
location '/user/zy2787_nyu_edu/join_confWtrend/'
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20240504001138_db40f48a-5a43-4bbe-8920-72dea1df3a88); Time taken: 0.296 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20240504001138_db40f48a-5a43-4bbe-8920-72dea1df3a88): create external table join_confwtrend (dateKey string, location string, sumconf
irmed int, sumrecovered int, symptoms double)
row format delimited fields terminated by ','
location '/user/zy2787_nyu_edu/join_confWtrend/'
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20240504001138_db40f48a-5a43-4bbe-8920-72dea1df3a88); Time taken: 0.082 seconds
INFO  : OK
INFO  : Concurrency mode is disabled, not creating a lock manager
No rows affected (0.385 seconds)
```

```
Insert into table join_confwtrend
select a.dateKey, a.location, a.sumconfirmed, a.sumrecovered, d.symptoms
FROM covid19 a inner join covid19_trend d on a.dateKey = d.dateKey AND a.location = d.location;
```

Joint Result 4:

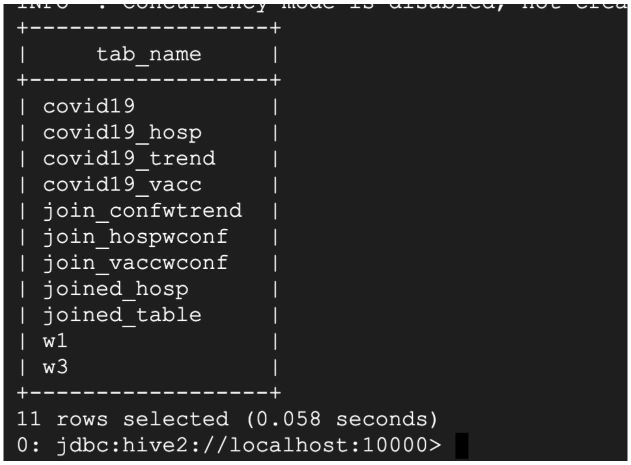
Join all four tables together:

```
0: jdbc:hive2://localhost:10000> create external table joined_table (dateKey string, location string,
sumconfirmed int, sumrecovered int, cumulative_vaccine_doses_administered int,
cumulative_persons_fully_vaccinated int, cumulative_persons_vaccinated int, new_hospitalized_patients
int, cumulative_hospitalized_patients int, new_intensive_care_patients int,
cumulative_intensive_care_patients int, symptoms double)
. . . . . . . . . . . . . . . . > row format delimited fields terminated by ','
. . . . . . . . . . . . . . . . > location '/user/zy2787_nyu_edu/joinedTableAll/‘
```

```
SELECT a.dateKey, a.location, a.sumconfirmed, a.sumrecovered,
b.cumulative_vaccine_doses_administered, b.cumulative_persons_fully_vaccinated,
b.cumulative_persons_vaccinated, c.new_hospitalized_patients,
c.cumulative_hospitalized_patients, c.new_intensive_care_patients,
c.cumulative_intensive_care_patients, d.symptoms

FROM covid19 a inner join covid19_vacc b on a.dateKey = b.dateKey AND a.location =
b.location inner join covid19_hosp c on a.dateKey = c.dateKey AND a.location = c.location inner
join covid19_trend d on a.dateKey = d.dateKey AND a.location = d.location;
```

In summary, these are all the tables we created:



```
INFO  : Concurrency mode is disabled, not crea
+------------------+
|     tab_name     |
+------------------+
| covid19          |
| covid19_hosp     |
| covid19_trend    |
| covid19_vacc     |
| join_confwtrend  |
| join_hospwconf   |
| join_vaccwconf   |
| joined_hosp      |
| joined_table     |
| w1               |
| w3               |
+------------------+
11 rows selected (0.058 seconds)
0: jdbc:hive2://localhost:10000>
```

covid19 = epidemiology data
covid19_hosp = hospitality data
codiv19_trend = search trend data
codiv19_vacc = vaccination data