

1. HBase Table Creation :

Table Name : **nytlc**

Column Family 1 : **trip_details**

Column Family 2 : **invoice_details**

Column Family Name	Column Name
trip_details	vendorid
	tpep_pickup_datetime
	tpep_dropoff_datetime
	passenger_count
	trip_distance
	ratecodeid
	store_and_fwd_flag
	pulocationid
	dolocationid
invoice_details	payment_type
	fare_amount
	extra
	mta_tax
	tip_amount
	tolls_amount
	improvement_surcharge
	total_amount
	congestion_surcharge
	airport_fee

Hbase code to create table:

```
create 'nytlc', 'trip_details', 'invoice_details'
```

```

hbase:002:0> create 'nytlc', 'trip_details', 'invoice_details'
Created table nytlc
Took 1.3587 seconds
=> Hbase::Table - nytlc
hbase:003:0> list
TABLE
nytlc
1 row(s)
Took 0.0241 seconds
=> ["nytlc"]
hbase:004:0>

```

Verify the Hbase table created using describe command:

describe 'nytlc'

```

hbase:004:0> describe 'nytlc'
Table nytlc is ENABLED
nytlc
COLUMN FAMILIES DESCRIPTION
(NAME => 'invoice_details', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0')
(NAME => 'trip_details', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0')
2 row(s)
Quota is disabled
Took 0.1459 seconds
hbase:005:0>

```

2. Sqoop Command to ingest data from mysql to HBase table – nytlc

Part 1: For Column Family – trip_details

```

sqoop import --connect jdbc:mysql://demodb.c3fsmo4st2kd.us-east-1.rds.amazonaws.com:3306/car_db --username admin -P --table ny_taxi_log --columns "rowid,vendorid,tpep_pickup_datetime,tpep_dropoff_datetime,passenger_count,trip_distance,ratecodeid,store_and_fwd_flag,pulocationid,dolocationid" --hbase-table nytlc --column-family trip_details --hbase-row-key rowed

```

```

[root@ip-172-31-42-99 ~]# sqoop import --connect jdbc:mysql://demodb.c3fsmo4st2kd.us-east-1.rds.amazonaws.com:3306/car_db --username admin -P --table ny_taxi_log --columns "rowid,vendorid,tpep_pickup_datetime,tpep_dropoff_datetime,passenger_count,trip_distance,ratecodeid,store_and_fwd_flag,pulocationid,dolocationid" --hbase-table nytlc --column-family trip_details --hbase-row-key rowed
Warning: /usr/lib/sqoop/./hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
2023-04-26 03:45:15,530 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
Enter password:

```

```

2023-04-14 00:40:53,427 INFO mapreduce.Job: map 100% reduce 0%
2023-04-14 00:40:54,432 INFO mapreduce.Job: Job job_1681427039642_0001 completed successfully
2023-04-14 00:40:54,614 INFO mapreduce.Job: Counters: 33
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=1345404
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=454
    HDFS: Number of bytes written=0
    HDFS: Number of read operations=4
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=0
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=5
    Other local map tasks=5
    Total time spent by all maps in occupied slots (ms)=350271840
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=7297330
    Total vcore-milliseconds taken by all map tasks=7297330
    Total megabyte-milliseconds taken by all map tasks=11208698880
  Map-Reduce Framework
    Map input records=18880595
    Map output records=18880595
    Input split bytes=454
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=27457
    CPU time spent (ms)=980980
    Physical memory (bytes) snapshot=1147944960
    Virtual memory (bytes) snapshot=12496433152
    Total committed heap usage (bytes)=852492288
    Peak Map Physical memory (bytes)=642469888
    Peak Map Virtual memory (bytes)=3133796352
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=0
2023-04-14 00:40:54,621 INFO mapreduce.ImportJobBase: Transferred 0 bytes in 3,752.096 seconds (0 bytes/sec)
2023-04-14 00:40:54,629 INFO mapreduce.ImportJobBase: Retrieved 18880595 records.
[root@ip-172-31-73-152 HBase]#

```

Verify the import by getting the details of last record with rowkey '18880595' using get command - get 'nytlc', '18880595'

All 9 columns are uploaded in column family – trip_details

```

hbase:029:0> get 'nytlc','18880595'
COLUMN                                CELL
trip_details:dolocationid             timestamp=2023-04-14T00:38:04.185, value=193
trip_details:passenger_count          timestamp=2023-04-14T00:38:04.185, value=1
trip_details:pulocationid             timestamp=2023-04-14T00:38:04.185, value=193
trip_details:ratecodeid               timestamp=2023-04-14T00:38:04.185, value=1
trip_details:store_and_fwd_flag        timestamp=2023-04-14T00:38:04.185, value=N
trip_details:tpep_dropoff_datetime    timestamp=2023-04-14T00:38:04.185, value=2017-02-28 20:45:55.0
trip_details:tpep_pickup_datetime     timestamp=2023-04-14T00:38:04.185, value=2017-02-28 20:45:42.0
trip_details:trip_distance             timestamp=2023-04-14T00:38:04.185, value=0.00
trip_details:vendordid                 timestamp=2023-04-14T00:38:04.185, value=2
1 row(s)
Took 0.0362 seconds
hbase:030:0>

```

Part 2: For Column Family – invoice_details

```
sqoop import --connect jdbc:mysql://demodb.c3fsmo4st2kd.us-east-1.rds.amazonaws.com:3306/car_db --username admin -P --table ny_taxi_log --columns "rowid,payment_type,fare_amount,extra,mta_tax,tip_amount,tolls_amount,improvement_surcharge,total_amount,congestion_surcharge,airport_fee" --hbase-table nytlc --column-family invoice_details --hbase-row-key rowed
```

```
[root@ip-172-31-42-99 ~]# sqoop import --connect jdbc:mysql://demodb.c3fsmo4st2kd.us-east-1.rds.amazonaws.com:3306/car_db --username admin -P --table ny_taxi_log --columns "rowid,payment_type,fare_amount,extra,mta_tax,tip_amount,tolls_amount,improvement_surcharge,total_amount,congestion_surcharge,airport_fee" --hbase-table nytlc --column-family invoice_details --hbase-row-key rowed
Warning: /usr/lib/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
2023-04-26 03:48:06,757 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
Enter password:
```

```
hbase:065:0> get 'nytlc','18880595'
COLUMN                                CELL
invoice_details:airport_fee            timestamp=2023-04-14T01:53:47.316, value=0.0
invoice_details:congestion_surcharge   timestamp=2023-04-14T01:53:47.316, value=0.0
invoice_details:extra                  timestamp=2023-04-14T01:53:47.316, value=0.5
invoice_details:fare_amount            timestamp=2023-04-14T01:53:47.316, value=2.5
invoice_details:improvement_surcharge  timestamp=2023-04-14T01:53:47.316, value=0.3
invoice_details:mta_tax                timestamp=2023-04-14T01:53:47.316, value=0.5
invoice_details:payment_type           timestamp=2023-04-14T01:53:47.316, value=2
invoice_details:tip_amount              timestamp=2023-04-14T01:53:47.316, value=0.0
invoice_details:tolls_amount           timestamp=2023-04-14T01:53:47.316, value=0.0
invoice_details:total_amount           timestamp=2023-04-14T01:53:47.316, value=3.8
trip_details:dolocationid              timestamp=2023-04-14T00:38:04.185, value=193
trip_details:passenger_count           timestamp=2023-04-14T00:38:04.185, value=1
trip_details:pulocationid              timestamp=2023-04-14T00:38:04.185, value=193
trip_details:ratecodeid                timestamp=2023-04-14T00:38:04.185, value=1
trip_details:store_and_fwd_flag         timestamp=2023-04-14T00:38:04.185, value=N
trip_details:tpep_dropoff_datetime     timestamp=2023-04-14T00:38:04.185, value=2017-02-28 20:45:55.0
trip_details:tpep_pickup_datetime      timestamp=2023-04-14T00:38:04.185, value=2017-02-28 20:45:42.0
trip_details:trip_distance             timestamp=2023-04-14T00:38:04.185, value=0.00
trip_details:vendordid                 timestamp=2023-04-14T00:38:04.185, value=2
1 row(s)
Took 0.0107 seconds
hbase:066:0>
```