Quick Start on Demo Environment

February 4, 2016

Contents

1 Prepare Environment

Since two client scripts are written in Python, you should setup basic python environment. And install following two dependencies:

```
sudo pip install requests click prettytable
```

2 Ingestion

First, create a table "demo_table"

```
python ingestion_client.py create \
    -h 52.27.67.237 \
    -u demo -a VbJSMjQ2MTIyNjY2MjJSOlHKTAEKleom \
    -d demodb \
    -t demo_table
```

Then, check the status if the ingestion agent is up.

```
python ingestion_client.py status \
    -h 52.27.67.237 \
    -u demo -a VbJSMjQ2MTIyNjY2MjJSOlHKTAEKleom \
    -d demodb \
    -t demo_table
```

Now, we can POST the data chunk by chunk. Here, we use 10 lines for each request.

```
python ingestion_client.py postfile \
    -h 52.27.67.237 \
    -u demo -a VbJSMjQ2MTIyNjY2MjJSOlHKTAEKleom \
    -d demodb \
    -t demo_table \
    -f wc.sample.json \
    --lines 10
```

3   Create/Update Schema

And, you can also check the code in "create_update_schema.py" for the schema definition.

```
python create_update_schema.py \
    -h 52.27.67.237 \
    -u demo -a VbJSMjQ2MTIyNjY2MjJSOlHKTAEKleom \
    -d demodb \
    -t demo_table
```

4   Query Example

You can query using "query.py"

4.1   Hive Engine

```
python query.py \
    -h 52.27.67.237 \
    -e hive \
    -d default \
    -q "SELECT uuid,ssid,source_host FROM demo_table limit 10"
```

The output will look like this:

```
host:52.27.67.237, port:18789, engine:hive, db:default, query:SELECT uuid,ssid,source_h
+------------------------------------+---------------------------+------------------
|                uuid                |           ssid            |      source_host
+------------------------------------+---------------------------+------------------
| fc2afe6d-e715-4585-a1f9-27776c4bfc7b | 913klbjvregia61mve4g01l6l7 | cax-web03.aws.com
| 573f1314-fd4d-4a04-a919-2132961bfee2 | lvped4ges659e7f8933lkrlmo4 | cax-web03.aws.com
| 74645fbe-82e0-4408-a37d-dd8d96564ab2 | 0hlatnug02m6om1r3iugaeqqh6 | cax-web03.aws.com
```

```
| aab85085-1779-4056-a7bd-bbce747fdfff | msccisclv1d30jepd06o4nsur7 | cax-web02.aws.com
| e5fc8dfd-9459-4ae0-9aa5-30e2f3f20d3a | 52qs5dg93o498t6hlk1ni7snn2 | cax-web04.aws.com
| 6ead7616-6bd8-4cc1-861e-c5be2e6a7202 | b9dvcc6frffo6116o3l0n7atq5 | cax-web03.aws.com
| 2157966b-5c2e-4913-9bad-0613c07c8116 | d58ptgb43tgni69seoe44e1216 | cax-web02.aws.com
| 06c8d39d-bccd-4ab6-b56d-9c0829e55426 | 4fd022k4b4n3i15ha8q5b1qmd3 | cax-web01.aws.com
| f119a607-c76d-4b93-8f52-d4f5c32ee648 | hgkoot1e1hgpj7lb8q93tvafb0 | cax-web03.aws.com
| 4fd380eb-21cb-4aa4-9332-4d4b8fb19f72 | dgk5on6k2bqbm0cav5441t0as3 | cax-web04.aws.com
+--------------------------------------+----------------------------+------------------
Query DONE
```

4.2 MPP Engine

```
python query.py \
    -h 52.27.67.237 \
    -e mpp \
    -d demo \
    -q "SELECT uuid,ssid,source_host FROM demo_table limit 10"
```

The output will look like this:

```
host:52.27.67.237, port:18789, engine:mpp, db:demo, query:SELECT uuid,ssid,source_host
Query results now, status:Processing
Query results now, status:Processing
+--------------------------------------+----------------------------+------------------
|                 uuid                 |            ssid            |    source_host
+--------------------------------------+----------------------------+------------------
| 01146ab6-aadb-423d-81b8-ea0051ddadda | eq751c4d1i21tgagfc5dfh21e6 | cax-web04.aws.com
| f0583231-a531-498a-8356-d285cc603d3f | dieg9ok0j2qmvgpdntqpt85to0 | cax-web04.aws.com
| cd5febc6-db26-499a-bb13-98ed4f2a3091 | qhon2877qnag3r14efhhqfd3u2 | cax-web03.aws.com
| b69c49e9-0220-431f-a490-128eef90348c | vgko93g93qd6qufql1fgf3os13 | cax-web02.aws.com
| b4b39e53-da3b-42dc-8871-7147fec26506 | imq9av4n9arsiru200kutkdl31 | cax-web04.aws.com
| 99216648-2b1f-49d8-9f7e-c94d8ffe84a2 | n4fm84lj3ij8f4seps1o4ksbs2 | cax-web03.aws.com
| a1ac0528-d1c8-4ea0-bf0c-2d166328f613 | 2k2p6ma4vspqmk8ksa9uit0b82 | cax-web01.aws.com
| de1a7c69-9753-4083-847b-a367940b6f99 | igrpuc3l21gki198jaii0sqk82 | cax-web04.aws.com
| f3146326-5fe2-4613-814f-d2e70fb6cbfb | orkde6qnhqs9ns4cr2rp2els81 | cax-web02.aws.com
| 95790330-2ab2-4388-b82b-389f6816f386 | igeo5ml8d67l1nd0kb3ov5kgn0 | cax-web05.aws.com
+--------------------------------------+----------------------------+------------------
Query DONE
```

5   Example Queries

5.1   MPP

```sql
SELECT COUNT(*)
FROM demo_table

SELECT COUNT(e) AS n_e,ssid
FROM demo_table
GROUP BY ssid
ORDER BY n_e;
```

5.2   Hive

```sql
SELECT COUNT(*)
FROM demo_table

SELECT COUNT(e) AS n_e,ssid
FROM demo_table
GROUP BY ssid
ORDER BY n_e;
```