Viewpoint
How can I query the Insight Mongo database using SQL?

This article will help you write SQL queries against the INSIGHT mongo database using Apache Drill. Before you start you need to connect Viewpoint reports to the insight database as explained in http://kb.adaptivecomputing.com/phpmyfaq/index. php?action=artikel&cat=5&id=220&artlang=en

Once you have followed the steps to connect Drill to the insight database you can run SQL queries. This article assumes that you have named the drill connection to the Insight mongo database "mongo2" as described in the previous article.

Open the drill web interface. By default Drill runs on port 8047. For example:

http://localhost:8047

Click the "Query" in the menu at the top of the Drill page. To start with try a simple query just to make sure the connection is set up. For the "jobs" collection a simple query would be as follows: Be sure to add the LIMIT 5 clause. If you don't it will return ALL the documents in collection. That will be time consuming and will put stress on the database.

```
###########################################
SELECT * FROM mongo2.insight.jobs LIMIT 5
###############################################
```

You should see a table appear with columns like _id, processorSecondsDedicated, commandFile, etc. There will be more columns than fit on a web page but this will ensure that things are working before trying more complicated queries.

Next try a more complicated query that only shows selected fields in the jobs collection.

```
##############################################
SELECT
  j.name as name,
  j.states.state as state,
  j.dates.completedDate as completedDate,
  j.credentials.jobClass as jobClass
```

```
FROM mongo2.insight.jobs j
LIMIT 5
##########################################
```

This should show a table of 5 jobs with columns for the name, state, completedDate, and user columns. Here are some more examples

To show only jobs in a Completed state run in the "batch" jobClass add a WHERE clause like the following.

```
##########################################

SELECT
  j.name as name,
  j.states.state as state,
  j.dates.completedDate as completedDate,
  j.credentials.jobClass as jobClass
FROM mongo2.insight.jobs j
WHERE
  j.credentials.jobClass = 'batch' AND
  j.states.state = 'Completed'
LIMIT 20

##########################################
```

To show only jobs that have completed in a certain time window, say between January 1 and February 1 of 2020, you can use the following SQL. Note that you can change whether the jobs will be sorted by completion date in ascending or descending order using the ORDER BY clause and passing in either ASC or DESC as shown below.

```
##########################################
SELECT
  j.name as name,
  j.states.state as state,
  j.dates.completedDate as completedDate,
  j.credentials.jobClass as jobClass
FROM mongo2.insight.jobs j
WHERE
  j.dates.completedDate > '2020-01-01T00:00:00.000-04:00' AND
  j.dates.completedDate < '2020-02-01T00:00:00.000-04:00' AND
  j.states.state = 'Completed'
ORDER BY
  j.dates.completedDate DESC
LIMIT 20
```

Viewpoint

##

You can also use SQL aggregate functions and GROUP BY clauses. For example, to count how many jobs are in each state

##

```
SELECT
j.states.state as state,
COUNT(j.states.state) AS jobCount
FROM mongo2.insight.jobs j
WHERE
j.dates.startDate > '2020-01-01T00:00:00.000-04:00' AND
j.dates.startDate < '2020-02-01T00:00:00.000-04:00'
GROUP BY j.states.state
LIMIT 10
```

##

So far we have only queried the "jobs" collection. To see what other collections are available and what other fields are in the jobs collection you can use the mongo command line client to view the database schema. First determine the username and password needed to connect to the insight mongo database. This found in /opt/insight/etc/config.groovy on your insight host.

##
```
[root]# vim /opt/insight/etc/config.groovy...
mongo.username="insight_user"
mongo.password="changeme!"
```
##

One the host where mongo is running (the mongo with the insight database) there should be a mongo command line client already installed. Use this client to open a connection to the insight database. For example:

##
```
[root]# mongo -u insight_user -p 'changeme!' insight
```

Viewpoint

```
> show collections
################################################
```

From here you can list the collections insight the database.

```
################################################
> show collections
insight_run
job_samples
job_state_journal
jobs <-------------- This is the collection we have been querying
last_handled_message
node_history
node_life_cycle_journal
node_samples
nodes
...
workload_view
################################################
```

These collections can be referenced in drill by using the following as a table name (assuming the drill connection to insight mongo is called "mongo2")

mongo2.insight.<collection name>

You'll notice in the examples above using the jobs collection the table was named

mongo2.insight.jobs

Next lets see what fields are available in a collection. From the mongo prompt show a document from the collection you want to see. For example, to see a document from the jobs collection run "db.jobs.findOne()" as shown below.

```
################################################
> db.jobs.findOne()
```

Viewpoint

```
{
"_id" : ObjectId("5d4c771686c5ba9fa914dbab"),
"processorSecondsDedicated" : 158.7,
"commandLineArguments" : null,
...
"dates" : {
...
"completedDate" : ISODate("2019-08-08T19:19:34Z"),
...
},
...
"name": "0",
...
"credentials" : {
"qos" : null,
"jobClass" : "batch",
"qosRequested" : null,
"user" : "testuser1491518208",
"account" : null,
"group" : "testuser1491518208"
},
"states":{
...
"state":"Completed"
}
...
}
```

##

Notice the relationship of this document to the SQL queries from above. Top level fields like "name" can be referenced like this

SELECT j.name FROM mongo2.insight.jobs j ...rest of query

Nested fields like state, which is within states, can be referenced using a period "." like this "states.state" or "dates.completedDate"

SELECT j.states.state FROM mongo2.insight.jobs j ...rest of query

WARNING: In Drill SQL the words "user", "group", and "count" are reserved word and will cause a syntax error unless surrounded with back ticks. Unfortunately these are fields in the insight database. For some reason this knowledge base article is using the WRONG BACK TICK character. So the following example should work but

won't

SELECT name, j.credentials.`user` FROM mongo2.insight.jobs LIMIT 5

When you run it you'll see
a org.apache.drill.common.exceptions.UserRemoteException: VALIDATION ERROR:
At line 1, column 14: Table 'j' not found SQL Query null [Error Id:
36faa01c-5304-4688-8faa-e102d51b5b86 on ultron-01:31010]

To get it to run replace the back ticks around user (e.g. `user`) by erasing the `
character and pressing the back tick key on your keyboard.. On English US
keyboards this is the key that shares the ~ character

SELECT name, j.credentials.[BACK TICK]user[BACK TICK] FROM mongo2.insight.jobs
LIMIT 5

Unique solution ID: #1222
Author: Nate Seeley
Last update: 2020-02-03 23:20