PARQUET - Read From Parquet Files Using Materialized Views

> Note: The following Table of Contents should exist in all documents

Goal

The goal of this document is to provide very simple way to read from an existing parquet file using Materialized views inside incorta.

In Example below, Materialized view is created to read the parquet file of SALES.SALES table provided by Sample Tenant Demo.

Solution

1. Ensure Spark is up & running and incorta is properly integrated with Spark.
2. Place the parquet file in shared folder or path, where it can be accessed by incorta
3. Login to incorta with user having permission to manage schemas (i.e. Super user, Schema Manager, or Analyzer)
4. Go to Schemas & Session Variables Page
5. Create new or Edit an Existing Schema
6. Add New Materialized view table as shown below
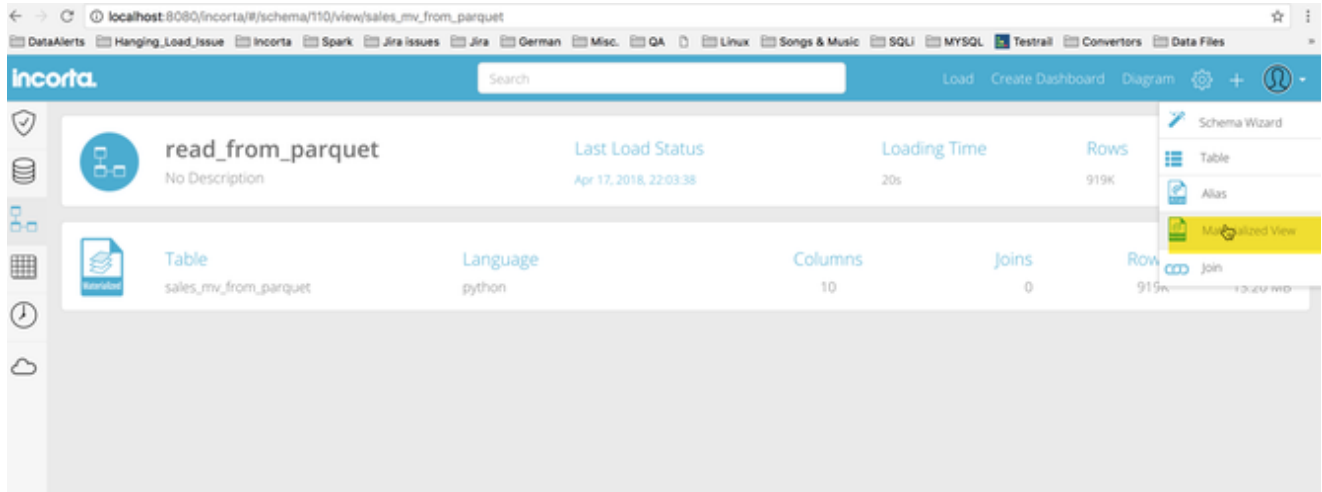
6. Choose Language as Python. Then add code as shown below. Please note that

- **parquetFile** is a variable name that can be set to anything.
- Please provide **correct full path of parquet files when using method**: **sqlContext.read.parquet or method: spark.read.parquet**

Simple read sqlContext.read.parquet

```
# direct read the parquet content by defining a variable and adding the
parquet path. No extra processing is required
parquetFile = sqlContext.read.parquet("/Users/emanmoussa
/incorta_rel211_official/tenants/Demo/parquet/SALES/SALES/part-000000")
# Save the result
save(parquetFile)
```

Or:

### Using sqlContext.read.parquet then processing result and saving it

```
#read the parquet content by defining a variable and adding the parquet
path
parquetFile = sqlContext.read.parquet("/Users/emanmoussa
/incorta_rel211_official/tenants/Demo/parquet/SALES/SALES/part-000000")
# Parquet files can also be registered as tables and then used in SQL
statements.
parquetFile.registerTempTable("parquetFile")
#Add your query. Below is a simple select from the parquet. You can
select specific fields if you need, or use specific functions
selectedSales= sqlContext.sql("SELECT * FROM parquetFile")
# Save the result
save(selectedSales)
```
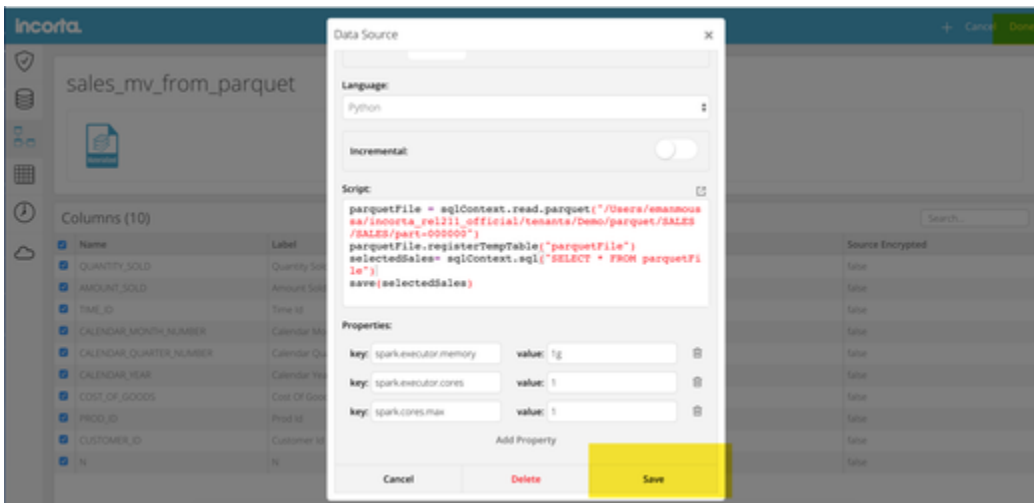
Or:

Simple read using spark.read.parquet

```
#simple read using spark.read.parquet
df2 = spark.read.parquet('/home/incorta/IncortaAnalytics_3.1.7/tenants
/Demo/parquet/SALES/SALES/part-000000')
save(df2)
```
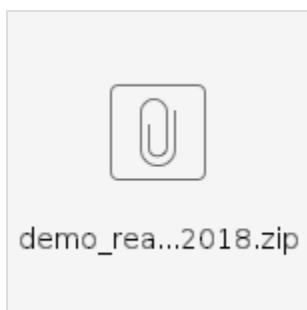
7. Optional: Add MV properties if needed, example is below
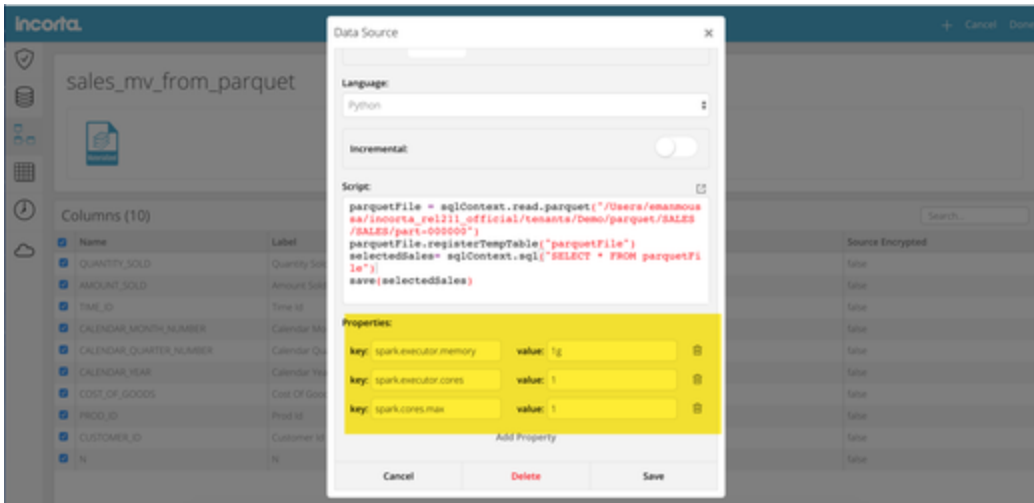
8. Click on Save on the Data Source pop-up then Done on Top right

Note: Sample Schema is attached below. You will need to modify it before use as per steps 6-8.

demo_rea...2018.zip

References:

please refer to spark programming guide & Choose Spark as language: https://spark.apache.org/docs/1.6.1/sql-programming-guide.html