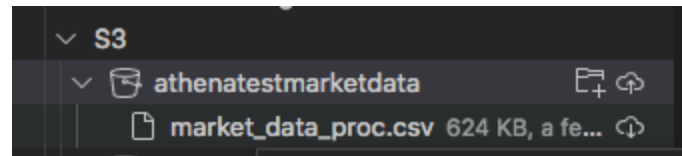


Amazon Athena

- Amazon Athena es un servicio de consultas interactivo que facilita el análisis de datos en Amazon S3 con SQL estándar.
- Athena no tiene servidor, de manera que no es necesario administrar infraestructura y solo paga por las consultas que ejecuta.

- Añadimos ficheros csv con el mismo formato a un bucket.



- Creamos una tabla en aws atheneas:


Analytics

Amazon Athena

Start querying data instantly.

Amazon Athena is an interactive query service that makes it easy to analyze data in Amazon S3 and other federated data sources using standard SQL.

How it works




Begin querying your data

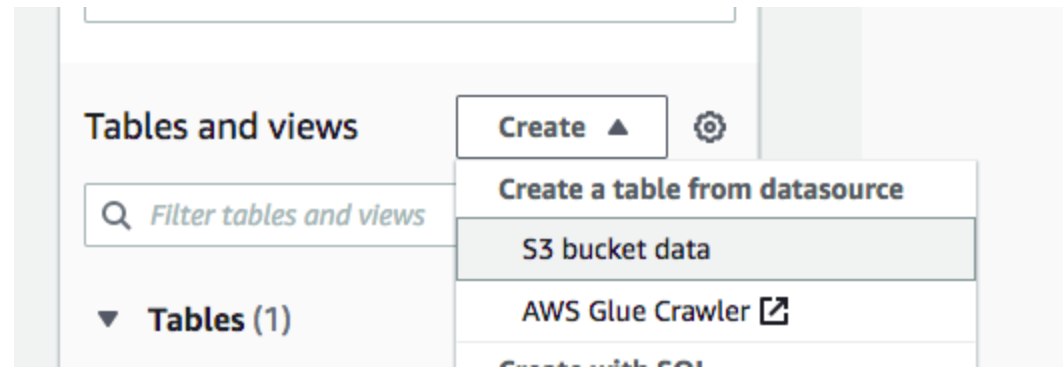
Discover the query editor and start querying right away.

[Explore the query editor](#)

Pricing

Paris - Data scanned	\$7 per TB
----------------------	------------

[Cost calculator](#) 



Dataset

Location of input data set

🔍 s3://athenatestmarketdata



View

Browse S3

Input the path to the data set you want to process on Amazon S3. For example if your data is stored at s3://input-data-set/logs/1.csv, please enter s3://input-data-set/logs/. If your data is already partitioned, e.g. s3://input-data-set/logs/year=2004/month=12/day=11/ just input the base path s3://input-data-set/logs/

Encryption [Información](#)

Choose this option if the underlying data is encrypted in Amazon S3.

☐ Encrypted data set

Data format

Data format

CSV



Column details

Column name must be single words that start with a letter or a digit. Certain advanced column types (namely, structs) are not exposed in this interface.

Column name

VALOR

Column type

string ▼

Remove

Column name

VOLUMEN

Column type

int ▼

Remove

Column name

PRECIO

Column type

float ▼

Remove

Column name

TIME

Column type

timestamp ▼


Remove


Add a column

Bulk add columns

Database

test_db ▼

Tables and views Create ▼ 

 *Filter tables and views*

▼ **Tables (1)** < 1 >

- [-] **marke_data** ⋮
 - valor ⋮
string
 - volumen ⋮
int
 - precio ⋮
float
 - time ⋮
timestamp

▼ **Views (0)** < 1 >

SAN	1	5.41	2018-03-19 09:00:18.002
SAN	44	5.41	2018-03-19 09:00:18.002
SAN	20	5.41	2018-03-19 09:00:18.002
SAN	45	5.41	2018-03-19 09:00:18.002
SAN	40	5.41	2018-03-19 09:00:18.002
SAN	10	5.41	2018-03-19 09:00:18.002
SAN	367	5.41	2018-03-19 09:00:18.002
SAN	396	5.41	2018-03-19 09:00:18.002
SAN	223	5.41	2018-03-19 09:00:18.002

- Podemos hacer consultas desde python con pyathena.
- Se instala con:

```
pip install pyathena
```

- Necesitamos un nuevo bucket de staging.
- Podemos hacer una consulta con:

```
from pyathena import connect
import pandas as pd

conn = connect(s3_staging_dir='s3://atheneastaging', region_name='eu-west-3')
df = pd.read_sql('SELECT * FROM "test_db"."marke_data" limit 100', conn)
print(df)
```

DEMO

Ejercicio

- Usa Amazon Athena para el fichero market_data_proc.csv y realiza una consulta desde la interfaz.
- Realiza la misma consulta desde python.