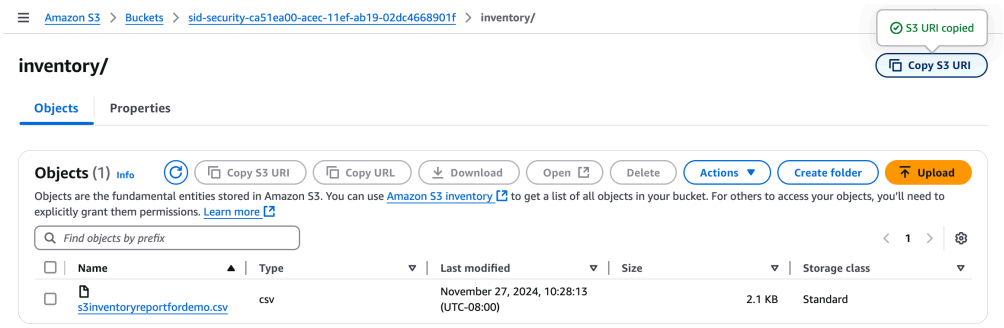
Setting up Amazon Athena

From the AWS console in the top search bar, search and select `s3`.

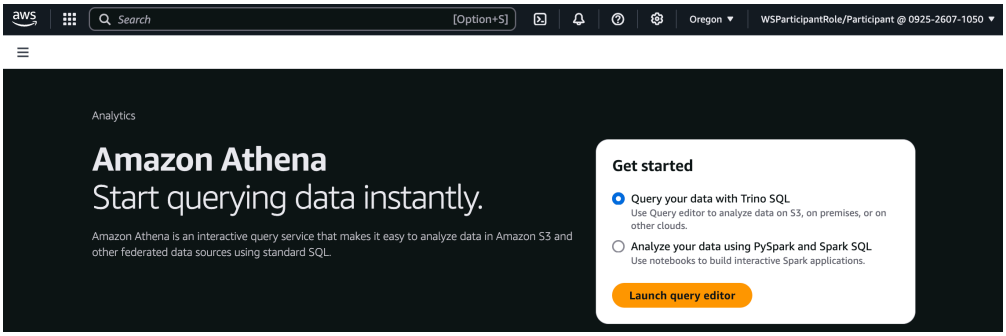
- Click the bucket name starting with `sid-security-xxxxxxxx` and click on the "inventory" prefix.
- Click the `Copy S3 URI` in the upper right, we will use this later when creating the Athena table.

You should see a screen similar to:

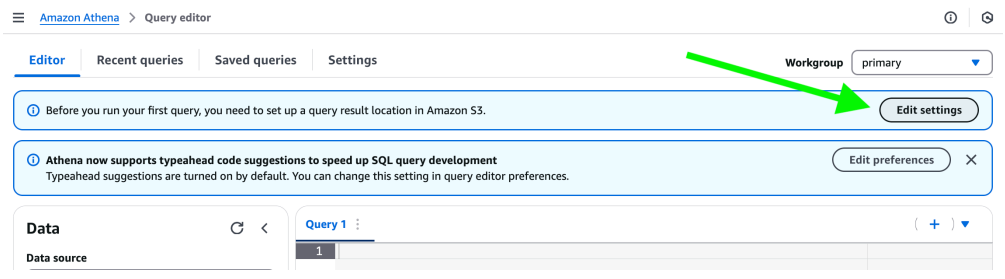


From the AWS console in the top search bar, search and select `athena`.

- You should see a Get Started page with “Query your with Trino SQL”, select the radio button and click `Launch query editor` at that page to continue, as shown in the following image:



Since this is the first time we are using Amazon Athena, we will need to configure an S3 bucket to store the query results. Next, click `Edit settings` and `Browse S3`, we will use the same bucket starting with `sid-security-xxxxxxxx` click `Choose` and `Save`



You will end up at the `Settings` tab, click the `Editor` tab before we move on.

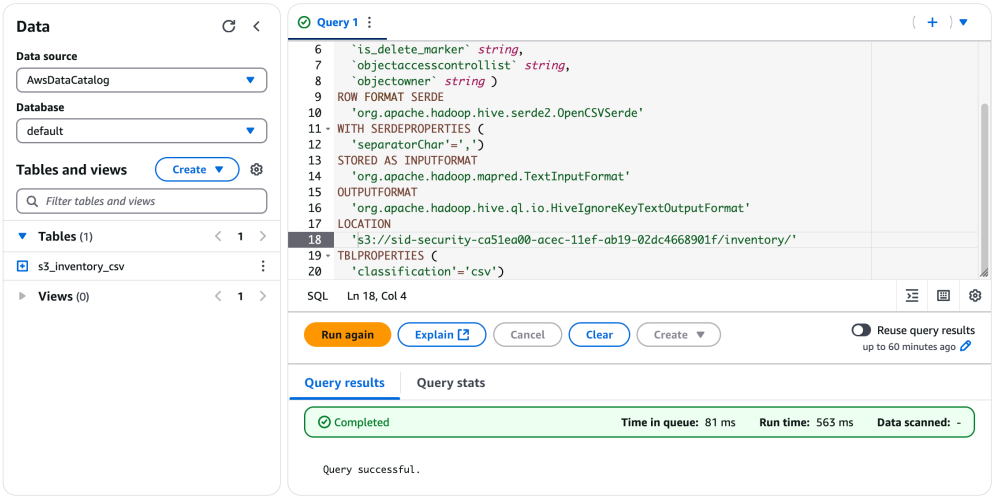
Now we can create a table in order to query the Amazon S3 Inventory results. We will use the SQ statement below to create a table named s3_inventory_csv in the default Athena database. You r replace the Amazon S3 Inventory location that we copied earlier in the following `LOCATION` string w the appropriate values for your configuration. Copy and past the follow into the query window, sinc this is our first query it should be labeled `Query1`

```
CREATE EXTERNAL TABLE `s3_inventory_csv`(
  `bucket` string,
  `key` string,
  `version_id` string,
  `is_latest` string,
  `is_delete_marker` string,
  `objectaccesscontrollist` string,
  `objectowner` string )
ROW FORMAT SERDE
  'org.apache.hadoop.hive.serde2.OpenCSVSerde'
WITH SERDEPROPERTIES (
  'separatorChar'=',')
STORED AS INPUTFORMAT
  'org.apache.hadoop.mapred.TextInputFormat'
OUTPUTFORMAT
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION
  's3://sid-security-xxxx/inventory/'
TBLPROPERTIES (
  'classification'='csv')
```

Note that for the LOCATION field, we use the Amazon S3 Inventory location we copied previously .

Click `Run` to create the table. You should see a green "Completed" message if you scroll down under `Query results.`

Once the SQL has been run, there will be a tabled named, s3_inventory_csv in your list of tables. We will use this table/schema to enable us to query the S3 Inventory report in the following section.



Before we continue, let's take a minute to look at the schema by clicking on the plus sign next to the table named `s3_inventory_csv` this will expand the selection to show the different columns. If you remember we added two of the ACL fields `Object Owner` and `Object ACL` which you can see, along with other fields. If we were to add other fields to the Inventory report we could extend the schema to match as well.

In the next section we will start querying the S3 Inventory Report with Amazon Athena.

Previous Next