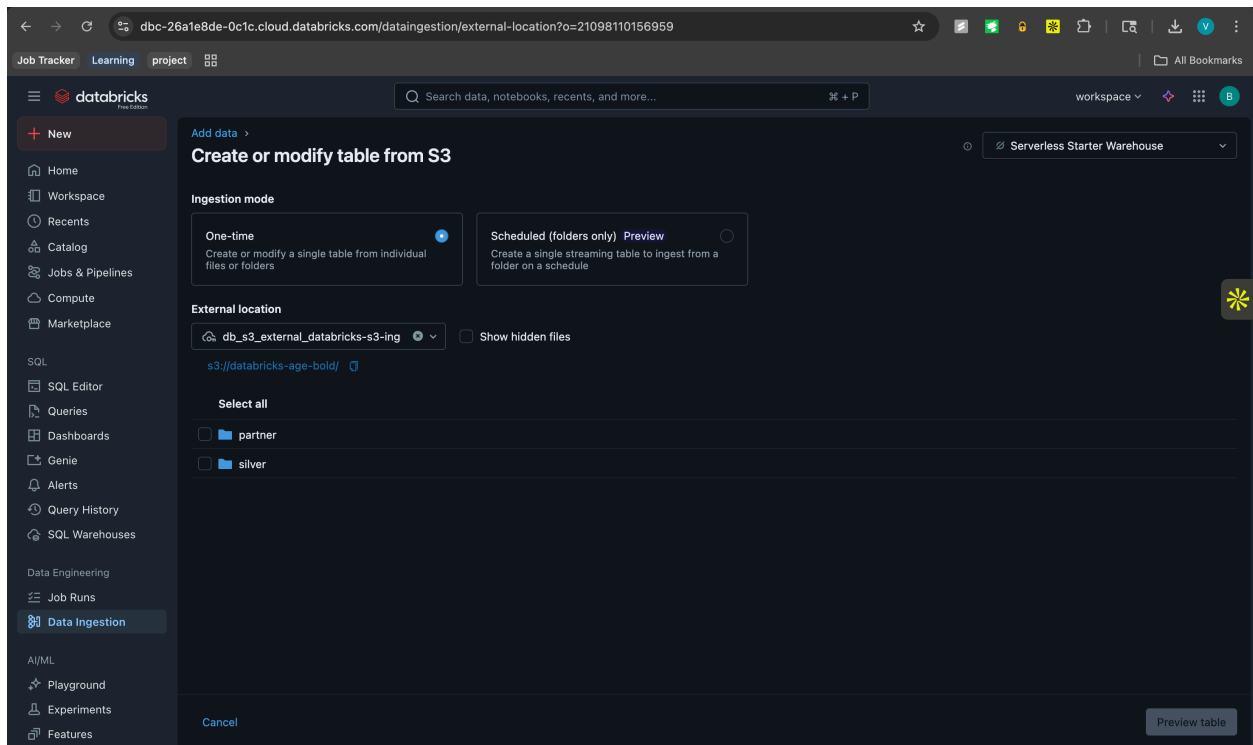


## 1. Uploaded the files into S3

The screenshot shows the AWS S3 console interface. The browser address bar displays the URL: `us-east-2.console.aws.amazon.com/s3/buckets/databricks-age-bold?region=us-east-2&prefix=partner/&showversions=false`. The console header includes the AWS logo, a search bar, and navigation links for Job Tracker, Learning, and project. The main content area shows the 'partner/' bucket with two tabs: 'Objects' and 'Properties'. The 'Objects' tab is active, displaying a list of objects. The list includes two folders: 'acme/' and 'bettercare/'. The table has columns for Name, Type, Last modified, Size, and Storage class. The 'acme/' folder is listed with a size of '-' and a storage class of '-'. The 'bettercare/' folder is also listed with a size of '-' and a storage class of '-'. The console footer includes links for CloudShell, Feedback, and Console Mobile App, along with copyright information for Amazon Web Services, Inc. or its affiliates.

Name	Type	Last modified	Size	Storage class
acme/	Folder	-	-	-
bettercare/	Folder	-	-	-

## 2. Connected the Databricks to S3 using External S3 location and added files in it as raw layer



3. Queried the data from ingested files into dataframes and created a final curated table for downstream applications

dbc-26a1e8de-0c1c.cloud.databricks.com/editor/notebooks/2772387274727454?o=21098110156959#command/6816786062084360

Job Tracker Learning project

All Bookmarks

New

Home

Workspace

Recents

Catalog

Jobs & Pipelines

Compute

Marketplace

SQL

SQL Editor

Queries

Dashboards

Genie

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

AI/ML

Playground

Experiments

Features

Workspace

Catalog

Type to search...

For you

All

My organization

workspace

default

eligibility\_gold\_unified

external\_id

first\_name

last\_name

dob

email

phone

partner\_code

information\_schema

system

cert\_prep\_catalog

main

Delta Shares Received

samples

pyspark Data\_Processing x +

Python Tabs: ON main Last edit was 21 minutes ago

08:35 PM (8s) 9

Python

```
)

print("\n** Gold table published: {GOLD_TABLE}")
spark.sql("SELECT partner_code, COUNT(*) AS cnt FROM {GOLD_TABLE} GROUP BY partner_code").show()

spark.sql("SELECT * FROM {GOLD_TABLE} ORDER BY partner_code, external_id").show(
truncate=False)

> See performance (4)
```

```
** Gold table published: eligibility_gold_unified

+-----+-----+
|partner_code|cnt|
+-----+-----+
|ACME| 2|
|BETTERCARE| 2|
+-----+-----+

+-----+-----+-----+-----+-----+-----+
|external_id|first_name|last_name|dob|email|phone|partner_code|
+-----+-----+-----+-----+-----+-----+
|1234567890A|John|Doe|1955-03-15|john.doe@email.com|555-123-4567|ACME|
|9876543210B|Jane|Smith|1948-07-22|jane.smith@email.com|555-987-6543|ACME|
|BC-001|Alice|Johnson|1965-08-10|alice.j@test.com|555-222-3333|BETTERCARE|
|BC-002|Charlie|Brown|1972-03-25|charlie.b@test.com|555-444-5555|BETTERCARE|
+-----+-----+-----+-----+-----+-----+

[Shift+Enter] to run and move to next cell
[Cmd+Shift+P] to open the command palette
```