

AWS

Services

Search

Output configuration

Info

AWS Glue

×

Getting started

ETL Jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

▼ Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

► Data Integration and ETL

► Legacy pages

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

Step 2

[Choose data sources and classifiers](#)

Step 3

[Configure security settings](#)

Step 4

**Set output and scheduling**

Step 5

Review and create

Target database

spotifystreamscan

Clear selection

Add database

Table name prefix - optional

fact\_popularity

Maximum table threshold - optional

This field sets the maximum number of tables the crawler will generate. In the event that this number is surpassed, the crawl will fail with an error. If not set, the crawler will automatically generate the number of tables depending on the data schema.

Type a number greater than 0

Advanced options

S3 schema grouping

Create a single schema for each S3 path

By default, when a crawler defines tables for data stored in S3, it considers both data compatibility and schema similarity. Select this check box to group compatible schemas into a single table definition across all S3 objects under the provided include path. Other criteria will still be considered to determine proper grouping.

Table level - optional

The value must be a positive integer that indicates table location (the absolute level in the dataset). The level for the top level folder is 1. For example, for the path mydataset/a/b/, if the level is set to 3, the table is created at location mydataset/a/b.

Enter text

When the crawler detects schema changes in the data store, how should AWS Glue handle table updates in the data catalog?

Update the table definition in the data catalog

Add new columns only

Ignore the change and don't update the table in the data catalog

New partitions are added, but existing partitions are not updated

Update all new and existing partitions with metadata from the table

Partitions inherit metadata properties — such as their classification, input format, output format, SerDe information, and schema — from their parent table.

How should AWS Glue handle deleted objects in the data store?

Delete tables and partitions from the data catalog

Ignore the change and don't update the table in the data catalog

Mark the table as deprecated in the data catalog

**AWS Glue**

- Getting started
- ETL jobs
  - Visual ETL
  - Notebooks
  - Job run monitoring
- Data Catalog tables
- Data connections
- Workflows (orchestration)

▼ Data Catalog

- Databases
  - Tables
  - Stream schema registries
  - Schemas
- Connections
- Crawlers
  - Classifiers
- Catalog settings

► Data Integration and ETL

► Legacy pages

- What's New
- Documentation
- AWS Marketplace

☑ Enable compact mode

☑ Enable new navigation

---

**Crawler successfully starting**  
The following crawler is now starting: "Star Schema Crawler"

## Star Schema Crawler

Last updated (UTC)  
May 8, 2024 at 21:02:29

[Run crawler](#) [Edit](#) [Delete](#)

### Crawler properties

Name <b>Star Schema Crawler</b>	IAM role <a href="#">glue-crawler-role</a>	Database <b>spotifystreamscan</b>	State <b>READY</b>
Description This crawler accesses the databates pulled from Spotify API and creates the star schema for our datawarehouse.	Security configuration -	Lake Formation configuration -	Table prefix <b>fact_popularity</b>
Maximum table threshold -			

► Advanced settings

Crawler runs | Schedule | Data sources | Classifiers | Tags

### Schedule Info

Status <b>🟢 Scheduled</b>	Frequency <b>Daily</b>	Time <b>At 09:02 PM</b>
------------------------------	---------------------------	----------------------------