
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

FSI315

Build a capital markets data hub with Amazon FinSpace

Vinny Saulys

Principal Solutions Architect
Amazon FinSpace

Tim Griesbach

Head of Product
Amazon FinSpace

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Data challenges for large enterprises

The data hub and a capital markets data hub

FinSpace as the capital markets data hub

Diving deeper

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data challenges in a large enterprise

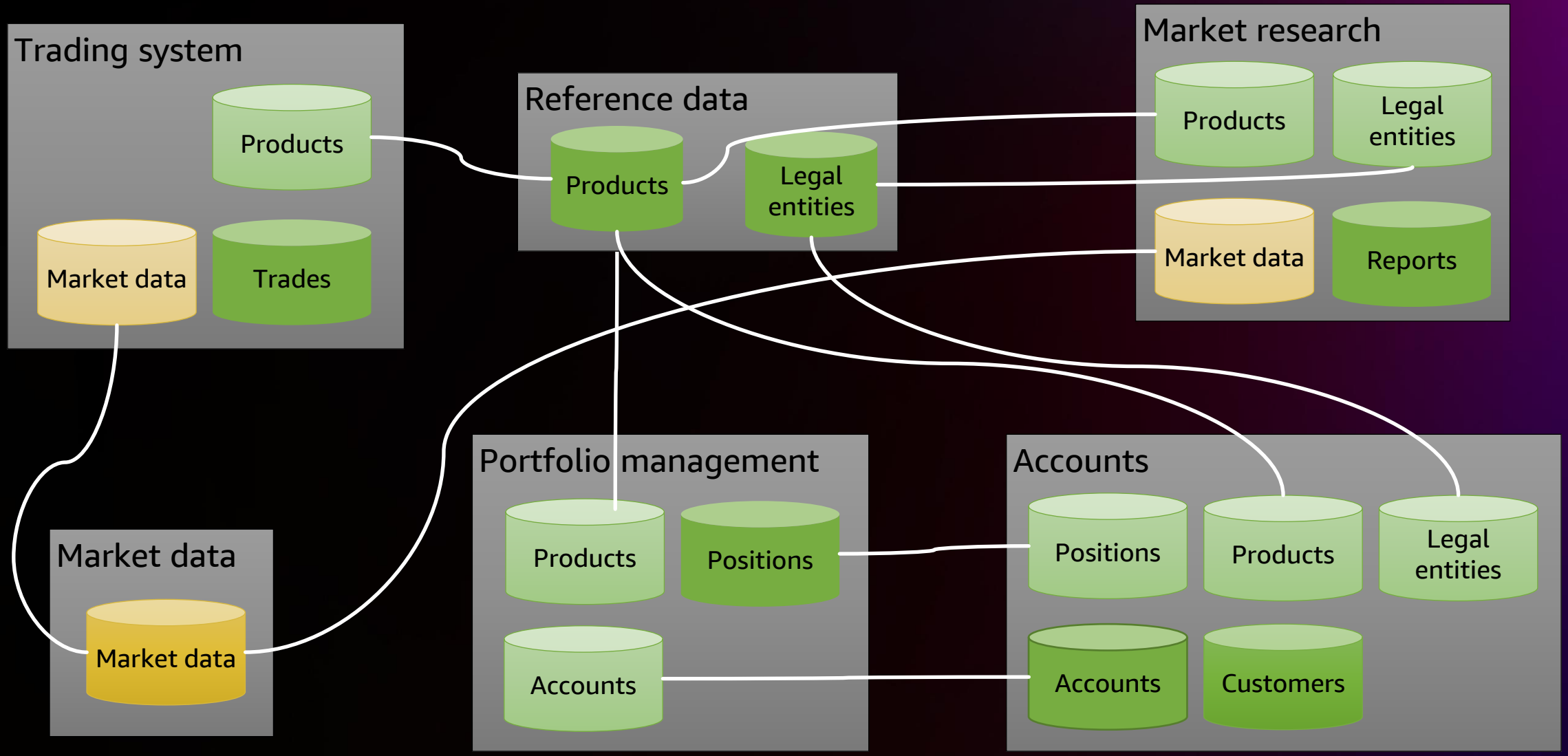
- Many systems produce data that can be useful to analysis
  - What is the right source and is it current?
  - Data is often copied between systems
- How to share analytic results (more data) with others in the enterprise
- Point to point sharing is inefficient, and copied data can get stale
- Companies are facing hybrid data challenges as well
- Data is split across cloud and on-premises environments

Question: What data challenges are you facing today?

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Capital markets producers and consumers



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Solution: Data hub

- Connects data producers (applications, processes, and teams) with data consumers (other applications, processes, and teams)
- Centralized data repository to allow seamless data connectivity
  - Find the data you need with minimal effort
- Reuse the data for numerous reasons
- Hubs enable efficiency, scale, and agility

Question: What solutions are you using today?

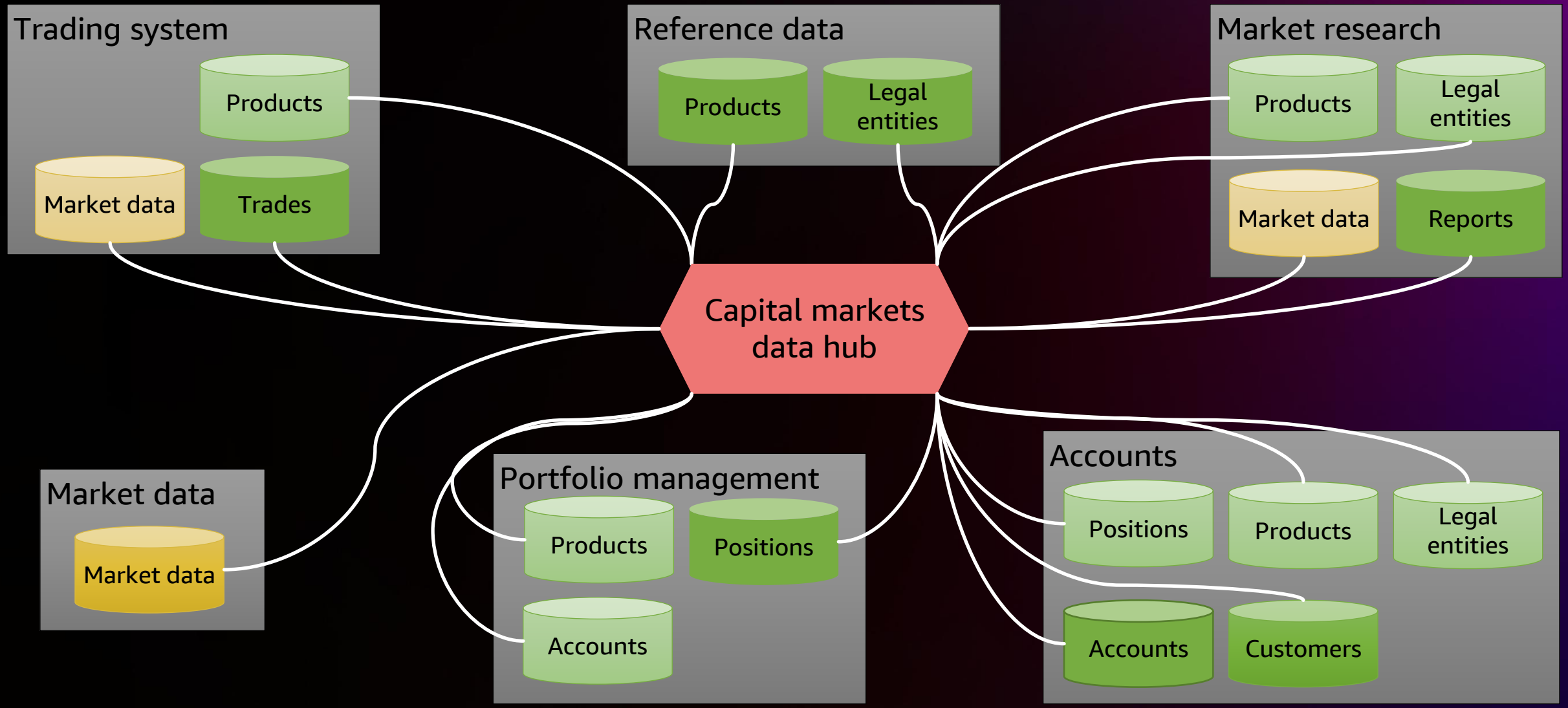
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

"This is not a technology, but an approach to more effectively determine where, when, and for whom data needs to be mediated, shared, and then linked and/or persisted. A data hub is a logical architecture which enables data sharing by connecting producers of data (applications, processes, and teams) with consumers of data (other applications, process, and teams). Endpoints interact, provisioning data into it or receiving data from it, and the hub provides a point of mediation and governance and visibility to how data is flowing across the enterprise."

Gartner

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Source: https://bit.ly/gartnerdatahub

Connecting data producers and consumers



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Capital markets data hub

Focus on data consumption and production for capital markets customers

Key features:

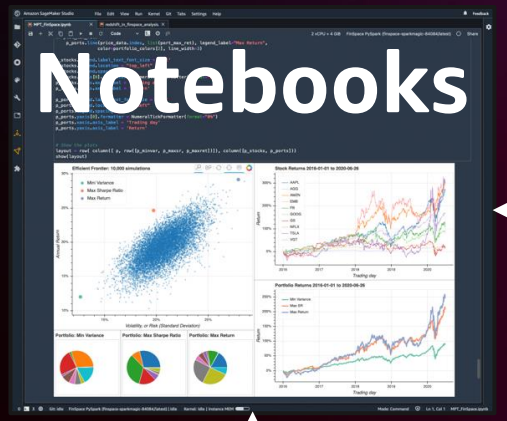
- Connects data producers with data consumers
  - Data vendors, API providers, including internal systems
  - Data warehouses, data lakes, relational databases (on-premises and in cloud)
- Analytics: perform analyses and easily share results
  - Scale and agility with capital markets data such as time series and related computation needs
- Governance and audit of all data (consumed and produced)

Question: What data or services (consuming or producing data) do you use?

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

FinSpace: Capital markets data hub



Data vendors

AWS Data Exchange

Amazon S3

Amazon Redshift

FinSpace APIs

Data lake

Goldman Sachs

GSFCD APIs

Data vendors

Snowflake Marketplace

Snowflake (Multi-cloud)

Vendor marketplaces

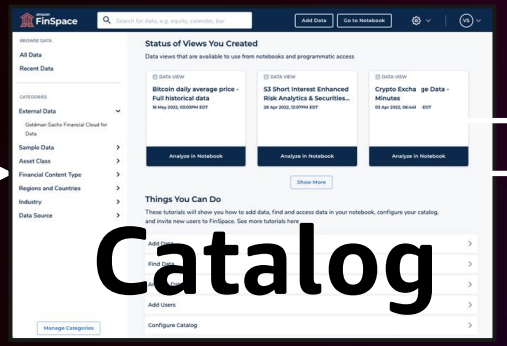
Vendor APIs

RDBMS (On-premises or cloud)

FinSpace APIs

Analytics

APACHE Spark



AWS Lake Formation

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Diving deeper: Where to next?

- How we cataloged tables from databases and then used the data in FinSpace
  - Solution outline
  - Creating and populating the attribute set from the data collected from the source database
- How the AWS Glue Data Connector works
  - Roles for permissions, AWS Glue Data Catalog metadata pulled into FinSpace
- How the GSFCD data connector works
  - Where is the API key kept? How is the API initialized with the key?
- How does data sharing work?
  - What is done to enable the sharing to the account's Lake Formation?

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Vinny Saulys
vssaulys@amazon.com

Tim Griesbach
awstim@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

FinSpace overview

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon FinSpace: Overview



Data management and analytics service that reduces the time to organize, prepare, and access data needed for financial analysis

- Instantly access data from a centralized data store
- Easily find and share data
- Prepare time series data at petabyte scale
- Perform historical point-in-time analysis
- Enforce your data access rules

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon FinSpace: Architecture



Data sources

First-party data

Internal data sources



Enterprise data lake



Business line databases

Second-party data

Third-party data

External data sources



Third-party data feeds



Alternative data



Amazon FinSpace

Interfaces

- Web UI
- Notebooks
- Python SDK

Data management

- Optimized views
- Data versioning
- Storage

Data organization

- Search and browse
- Business data catalog
- Metadata

Prepare and analyze

- Machine learning
- Time series analytics
- Managed Spark clusters

Governance and audit

- Data access management
- Resource access control
- Audit reporting



Data consumers



Data analysts



Business analysts



Data scientists



Applications, engines, and algorithms

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Blogs

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Blog posts: Amazon RedShift and Snowflake



- Analyze daily trading activity using transaction data from Amazon Redshift in Amazon FinSpace
- Making financial data accessible for analytics: How Amazon FinSpace can integrate with Snowflake

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data connectors

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Populating producers in the catalog



1. Attribute sets capture data about a database source
   - Database name, table name, and source type
2. Swept all tables in a database and recorded relevant information in catalog
   - Example: positions_tb, portfolio_db, RedShift
3. When requesting data from source, API uses the metadata to load the relevant driver and accesses the data from the correct table and database

See blog posts re: Amazon RedShift and Snowflake

We've now expanded and simplified with data connectors

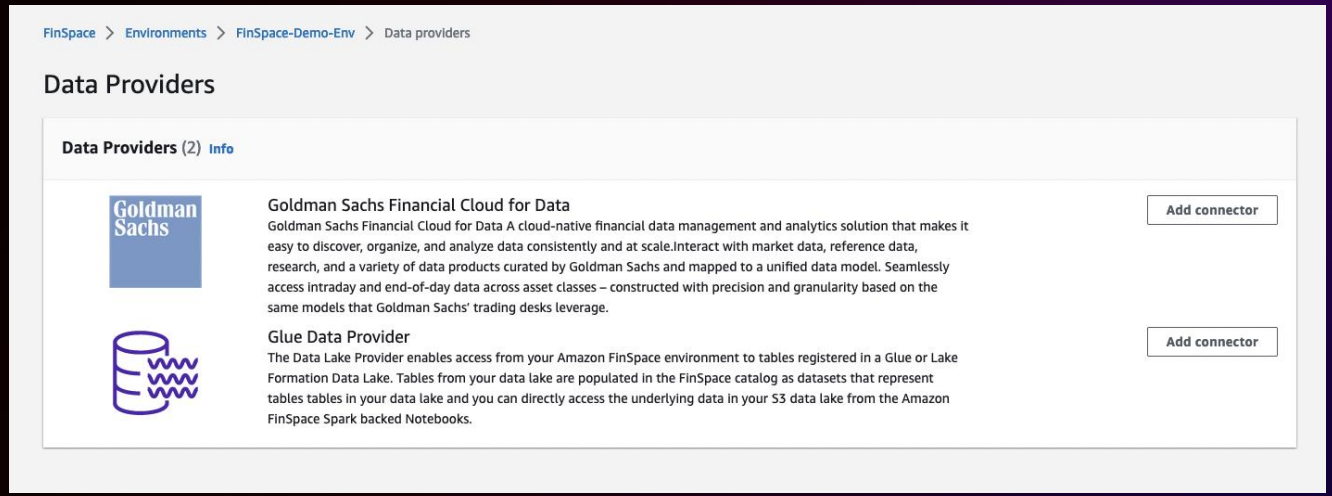
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data connectors to data producers

- AWS Glue Data Provider: Supports AWS Glue and Lake Formation
- Goldman Sachs Financial Cloud for Data (GSFCD)

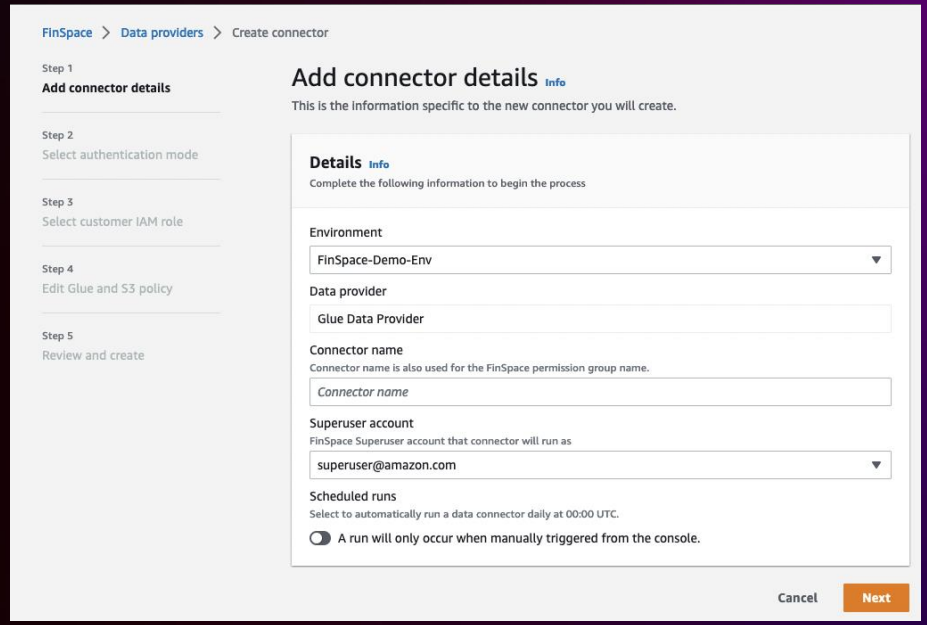
Configure and FinSpace does the rest



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Glue Data connector

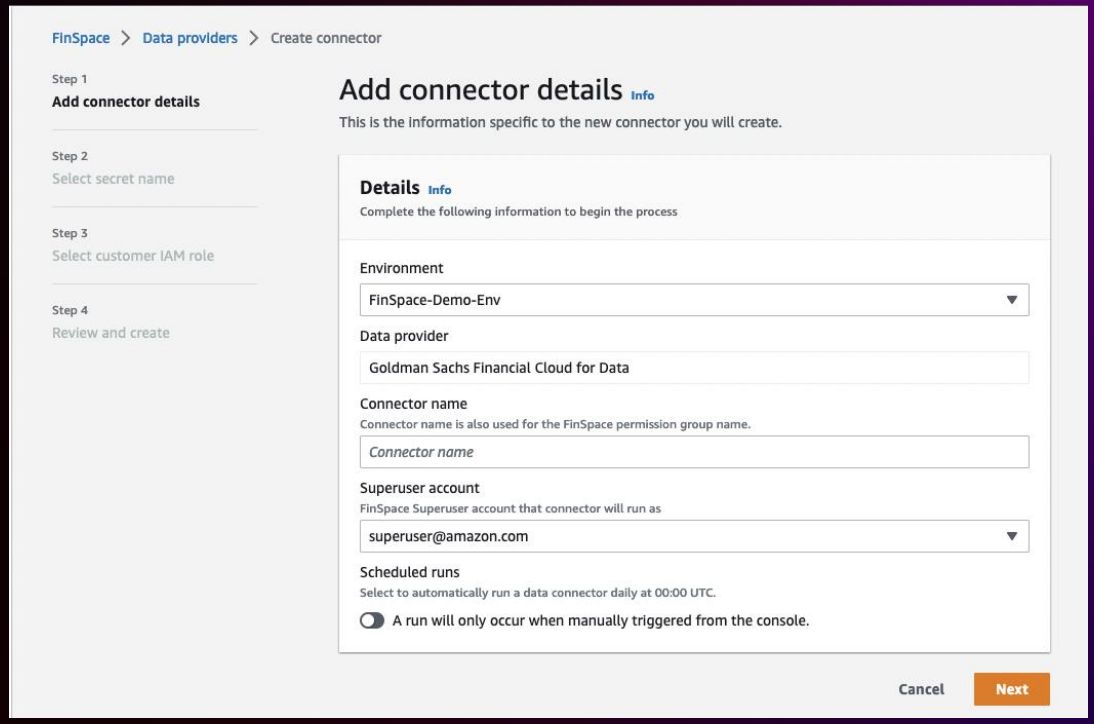
- Two sources
  - AWS Glue
  - Lake Formation
- Easily set up from the console
- Sources are read-only from FinSpace
- System syncs source tables to FinSpace datasets
- Further describe datasets in FinSpace
- Search and locate datasets from FinSpace



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Goldman Sachs Financial Cloud for Data

- Announced at re:Invent 2021
- Easy set up from console
  - Provide your GSFCD API key
- Catalog kept in sync with available data



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

FinSpace data sharing

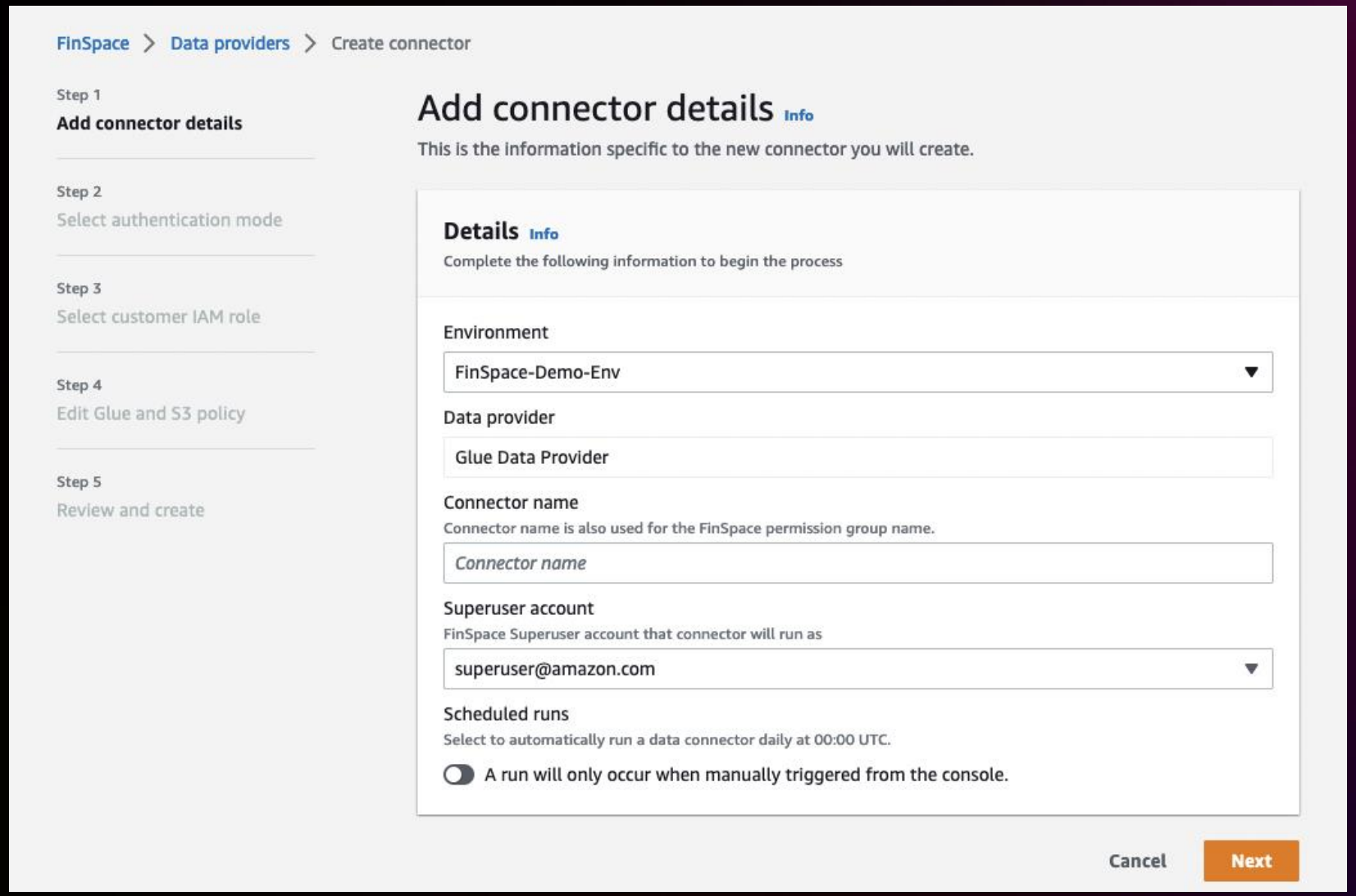
- FinSpace as data producer to other consumers
- Data consumers: Lake Formation
- Enables analytics such as Amazon RedShift Spectrum, Amazon Athena, and Amazon QuickSight
- Enabled as a service request, shares to the account’s Lake Formation instance

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data connectors setup

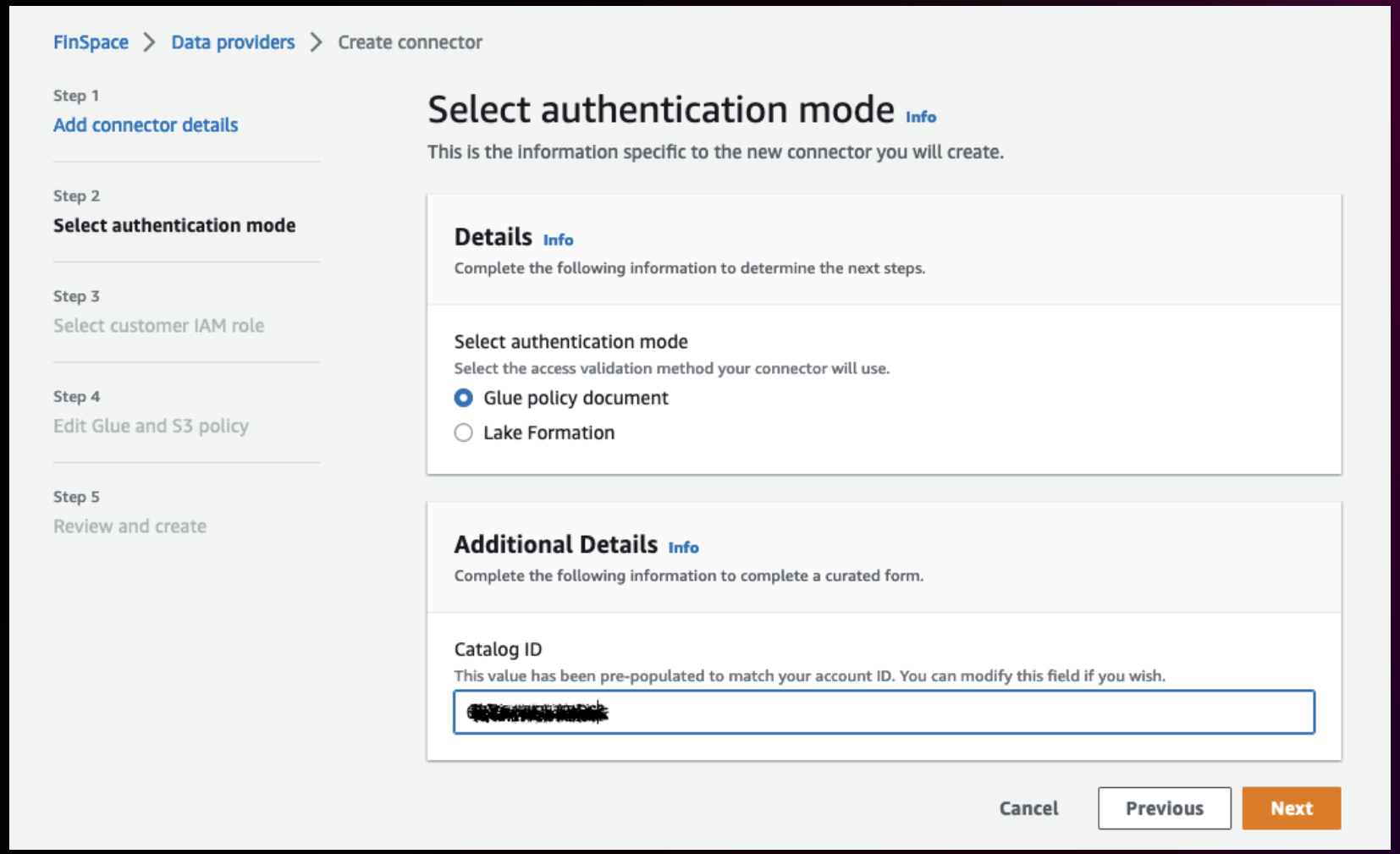
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Glue connector screenshots



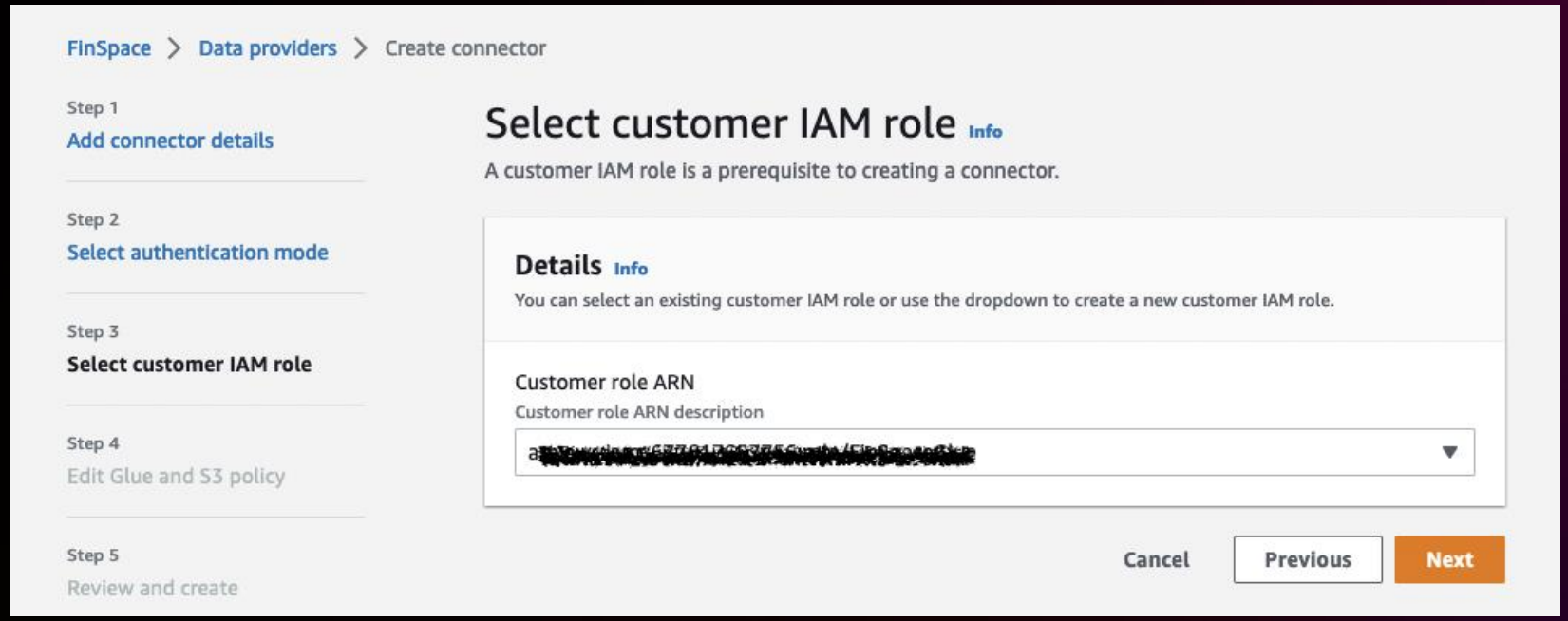
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Glue connector screenshots



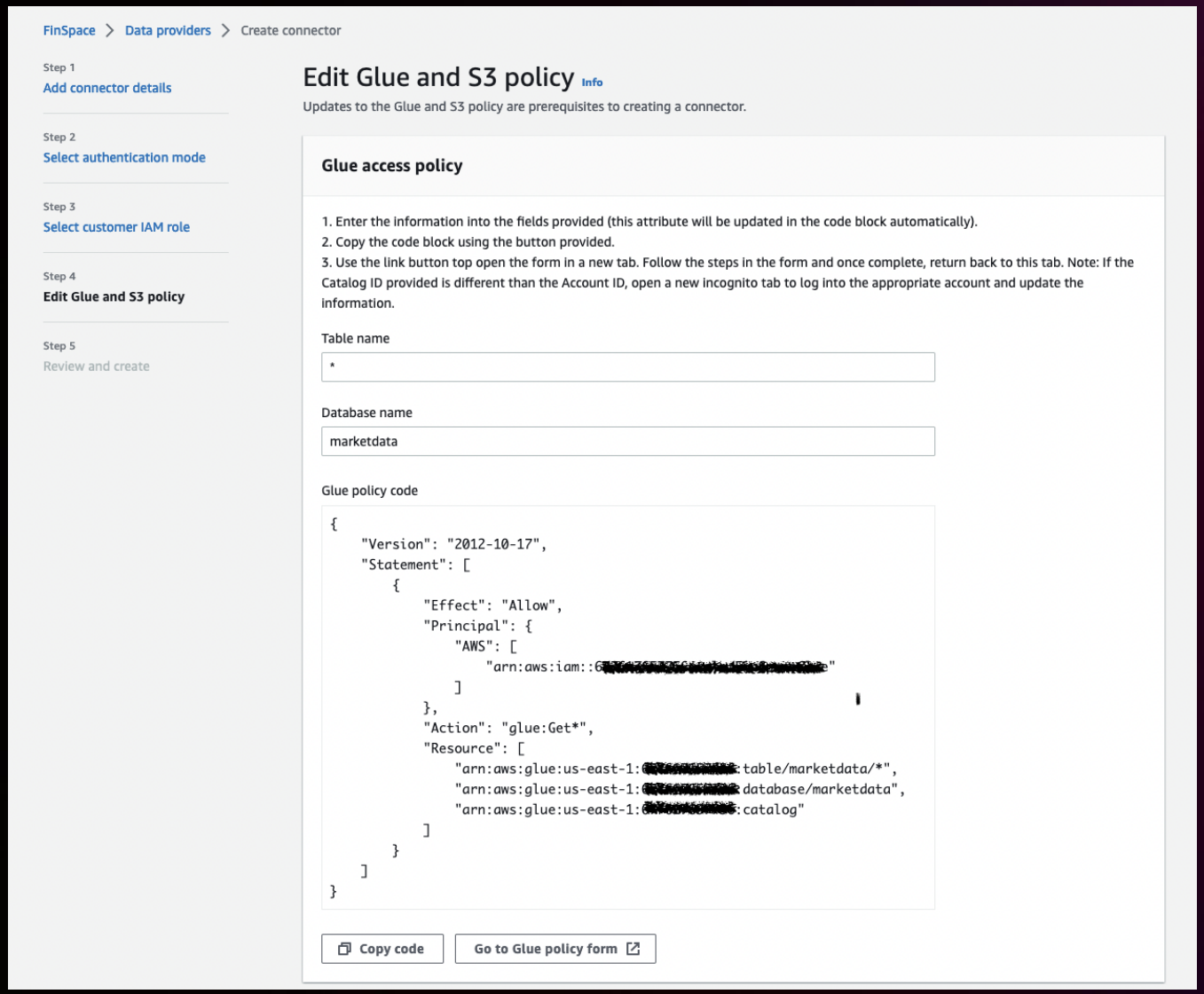
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Glue connector screenshots



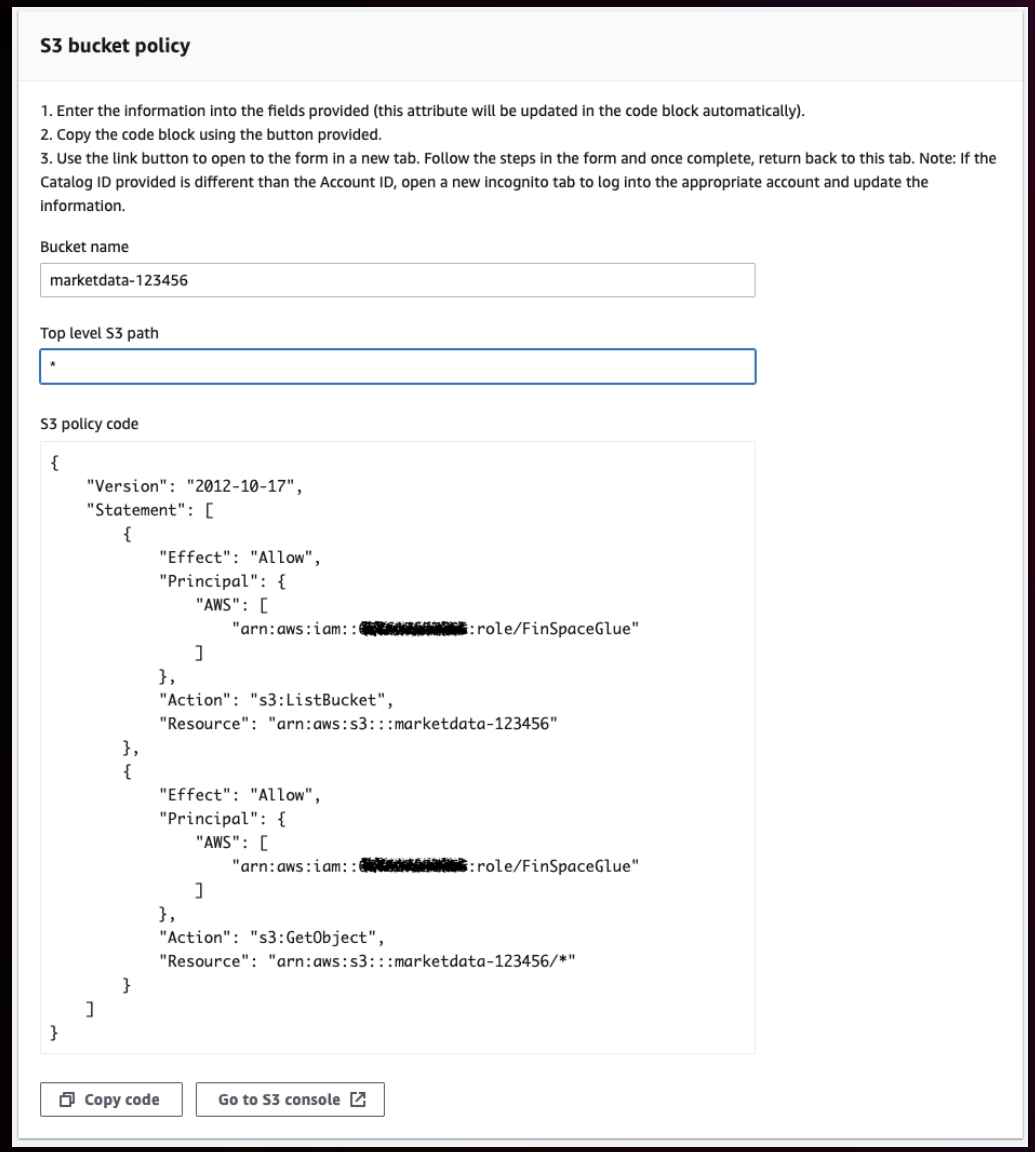
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Glue connector screenshots



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Glue connector screenshots



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Glue connector notes

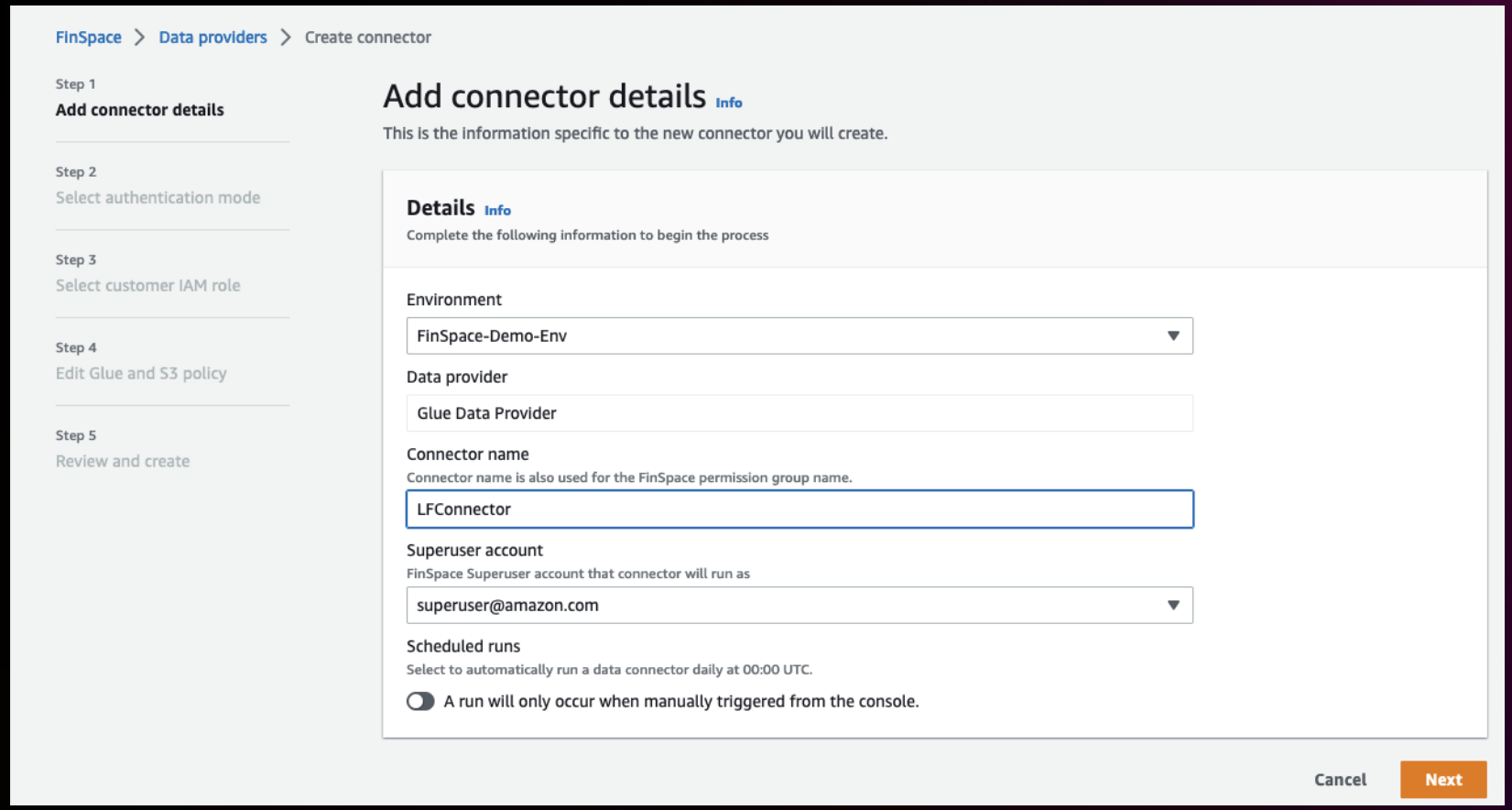


- Allow FinSpace to query AWS Glue Data Catalog
  - Trust policy allows FinSpace principal to query catalog
  - Specifically allow ‘get’ for specific database(s) and table(s)
  - Can wildcard the tables
- Grant access to Amazon S3 data
  - Amazon S3 policy grants FinSpace principal access to bucket
  - Actions: GetObject and ListBucket
  - Location of table data of the AWS Glue Data Catalog

aws

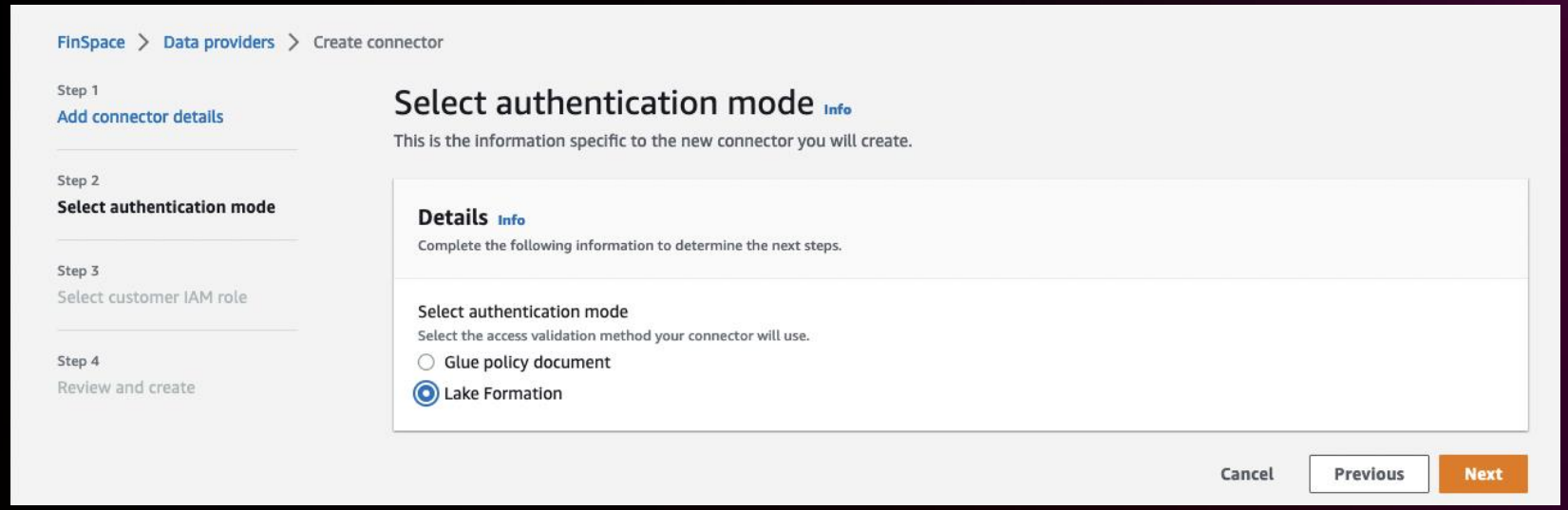
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

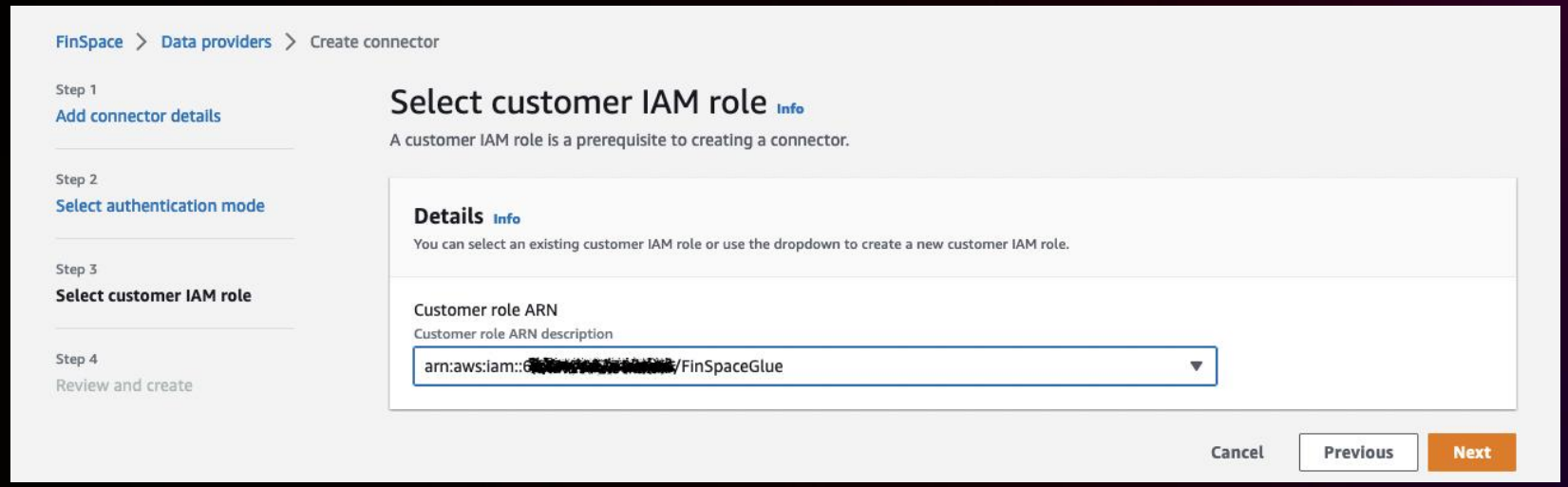
Lake Formation connector screenshots



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

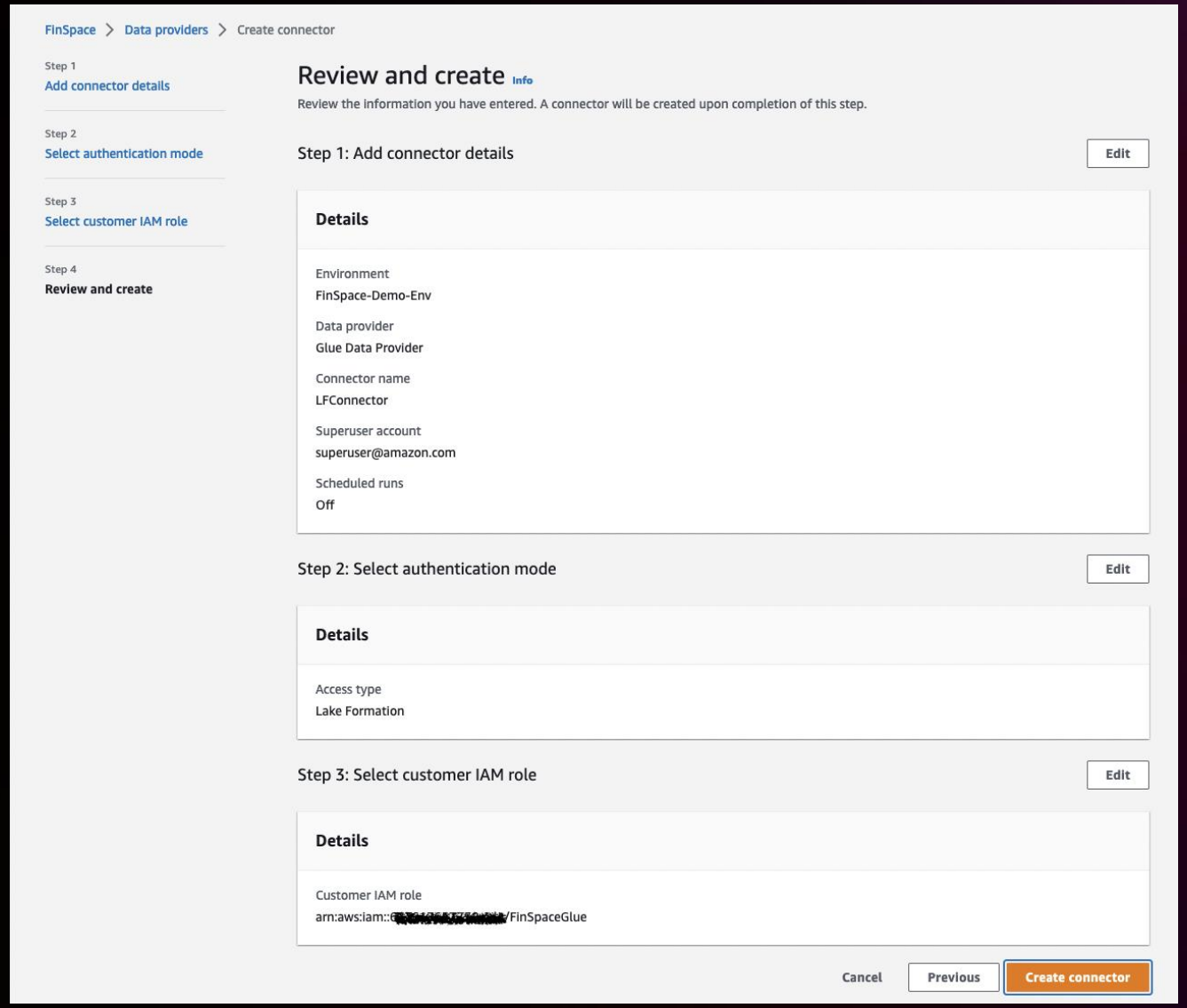
Lake Formation connector screenshots





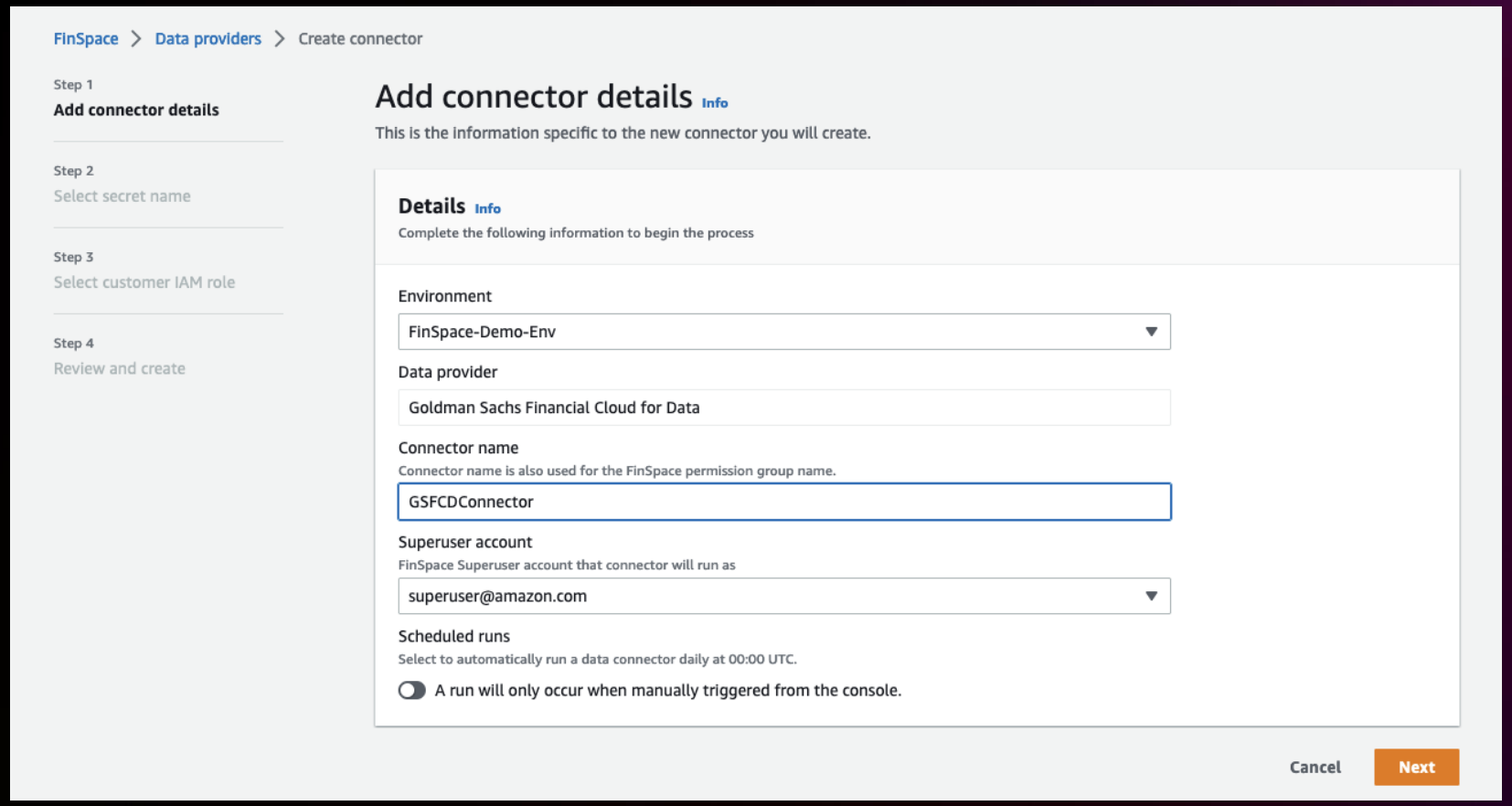
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lake Formation connector screenshots



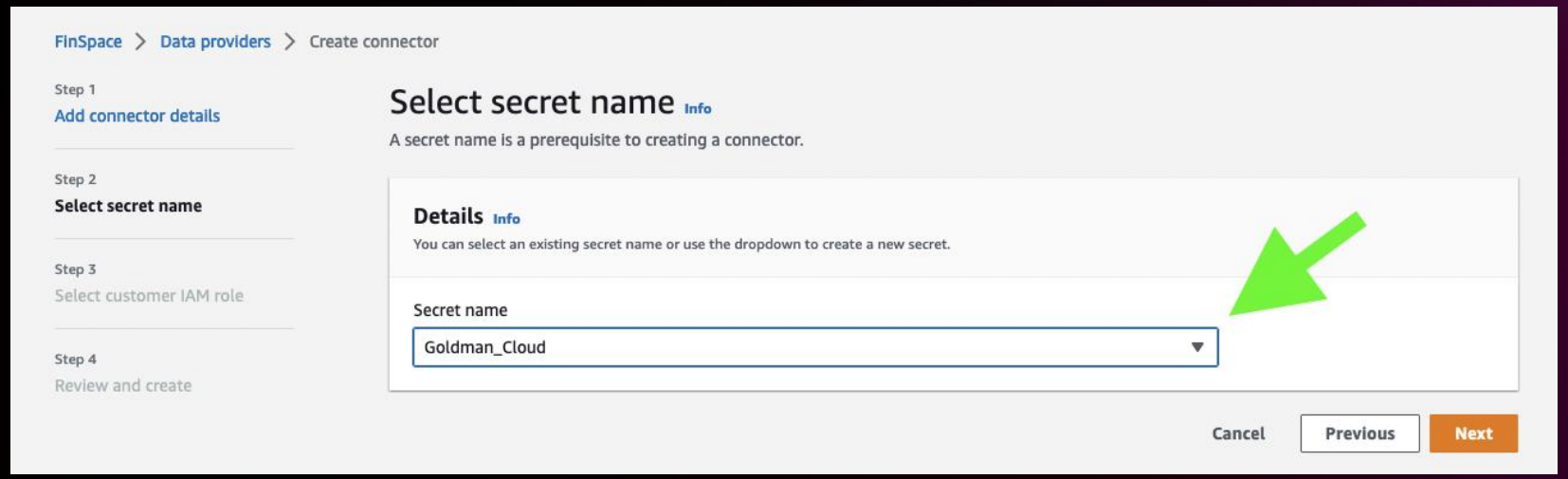
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

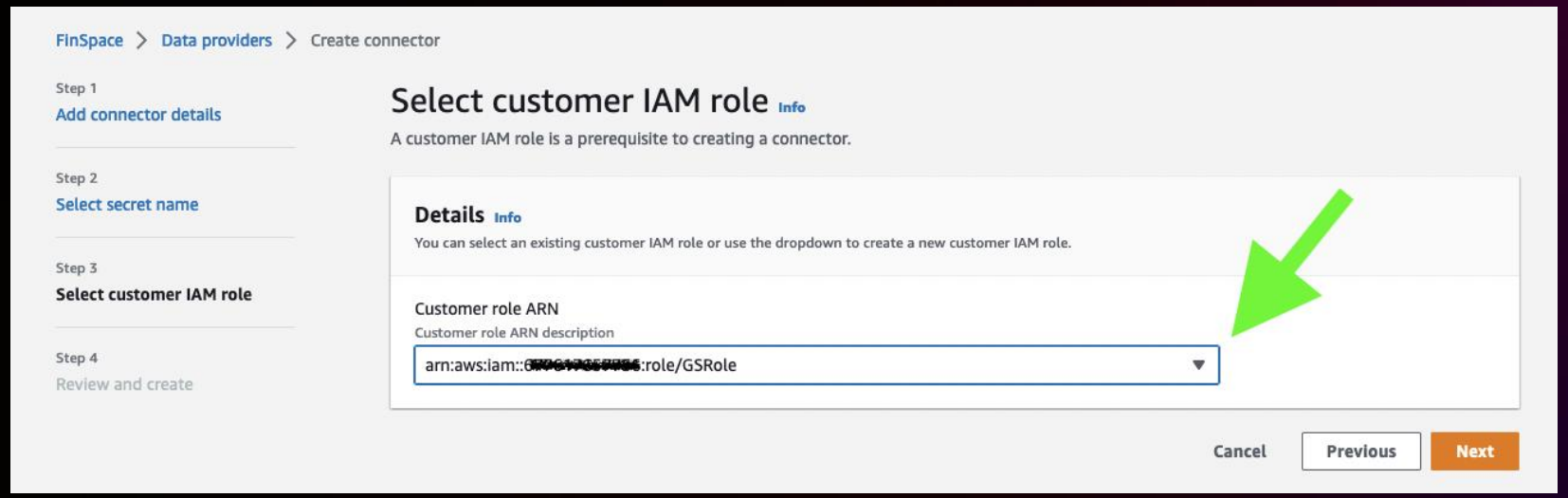
GSFCD connector screenshots: Step 1



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

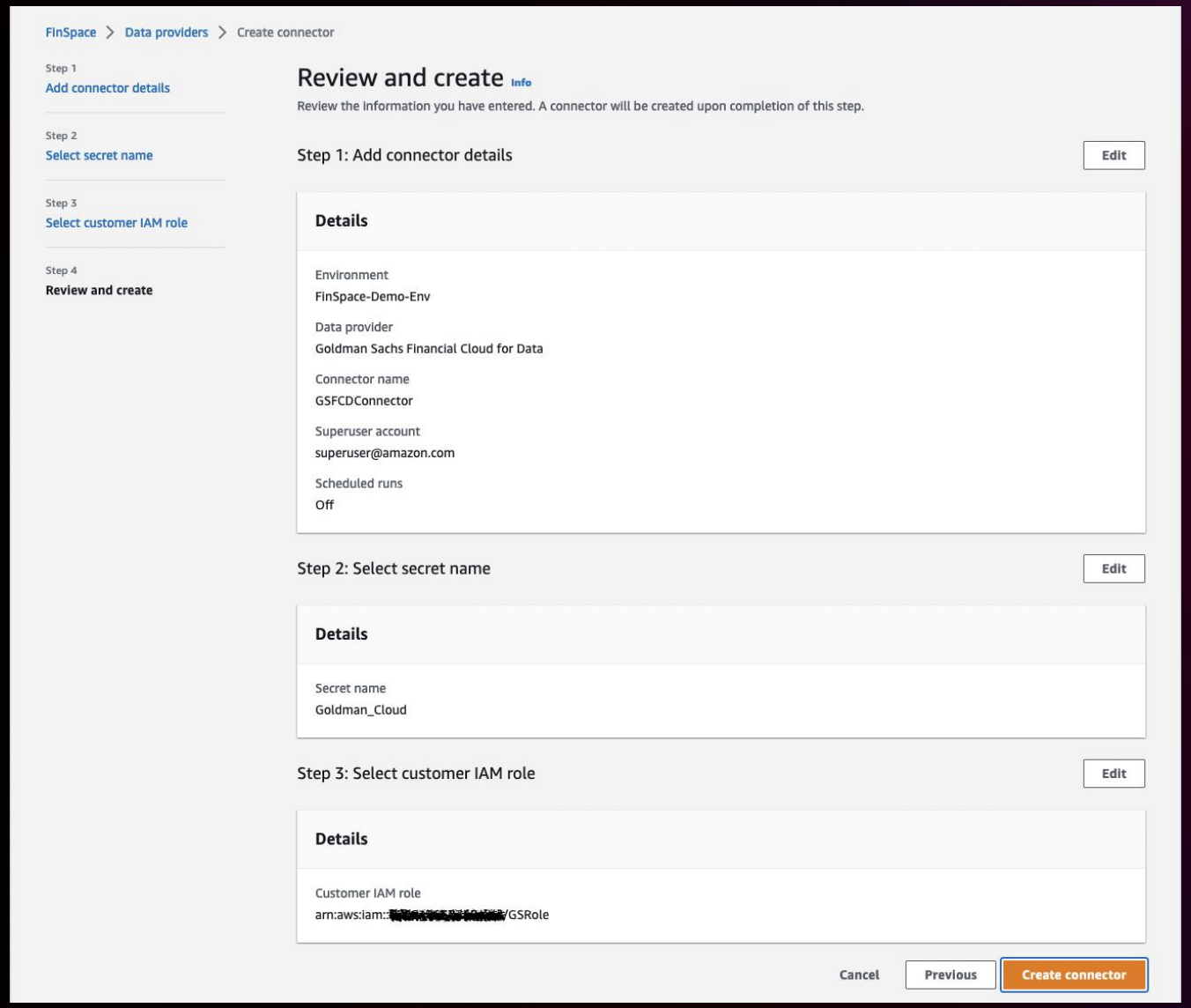
GSFCD connector screenshots: Step 2 and 3





© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

GSFCD connector screenshots: Step 4



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data connectors notes

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lake Formation connector notes



- Underlying Amazon S3 location of Lake Formation is registered as a Lake Formation data lake location
  - AWS Documentation
- Underlying Amazon S3 data lake location is registered with the non-default role
  - AWS Documentation
- Enable external data filtering on Lake Formation
  - Add the FinSpace customer account as an allowed account and “finspace” as an allowed tag

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data lake connector notes



- Permission group is created, name founded on the connector name
  - Format: “<Connector Name> Group”
  - Example: DataLakeConnectorOne creates the group “DataLakeConnectorOne Group”

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

GSFCD connector: Notes



- AWS Secrets Manager holds the API keys
  - Key has client id and secret (2 values)
- Connector uses role to access secret
  - Create policy, attach role to policy

- FinSpace permission group is created, name founded on the connector name
  - Format: “<Connector Name> Group”
  - Example: GSConnectorOne creates the group “GSConnectorOne Group”

- Grant access to data from GSFCD by adding users to the group

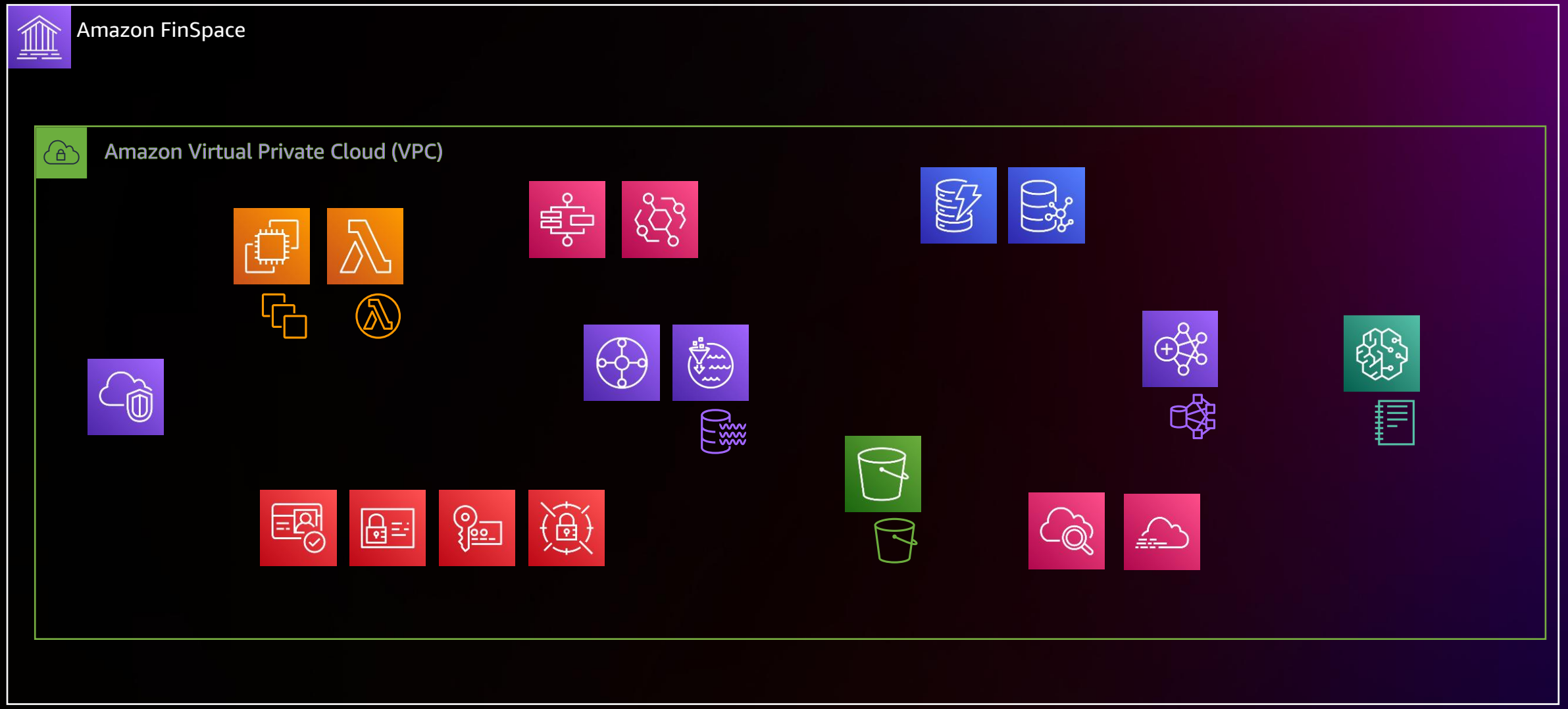
aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

GSFCD connector: Policy

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "secretsmanager:GetSecretValue",
            "Resource": "arn:aws:secretsmanager:us-east-1:123456789012:secret:Oct2022Secret-Yph8th"
        },
        {
            "Effect": "Allow",
            "Action": "kms:Decrypt",
            "Resource": "arn:aws:kms:us-east-1:123456789012:key/1b3dd08d-a956-4d25-8969-85c47ae3650e"
        }
    ]
}
```

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS services used by FinSpace



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.