
Cloud Data Governance and Catalog Custom Scanner DBT with Snowflake case study

Sachin Jain - Principal Solutions Architect

Rashmi P - Solutions Architect

Cloud Data Governance and Catalog



Single solution for data intelligence



Automated to save time on manual tasks



Start immediately with a cloud-native solution



Put accurate information at your team's fingertips

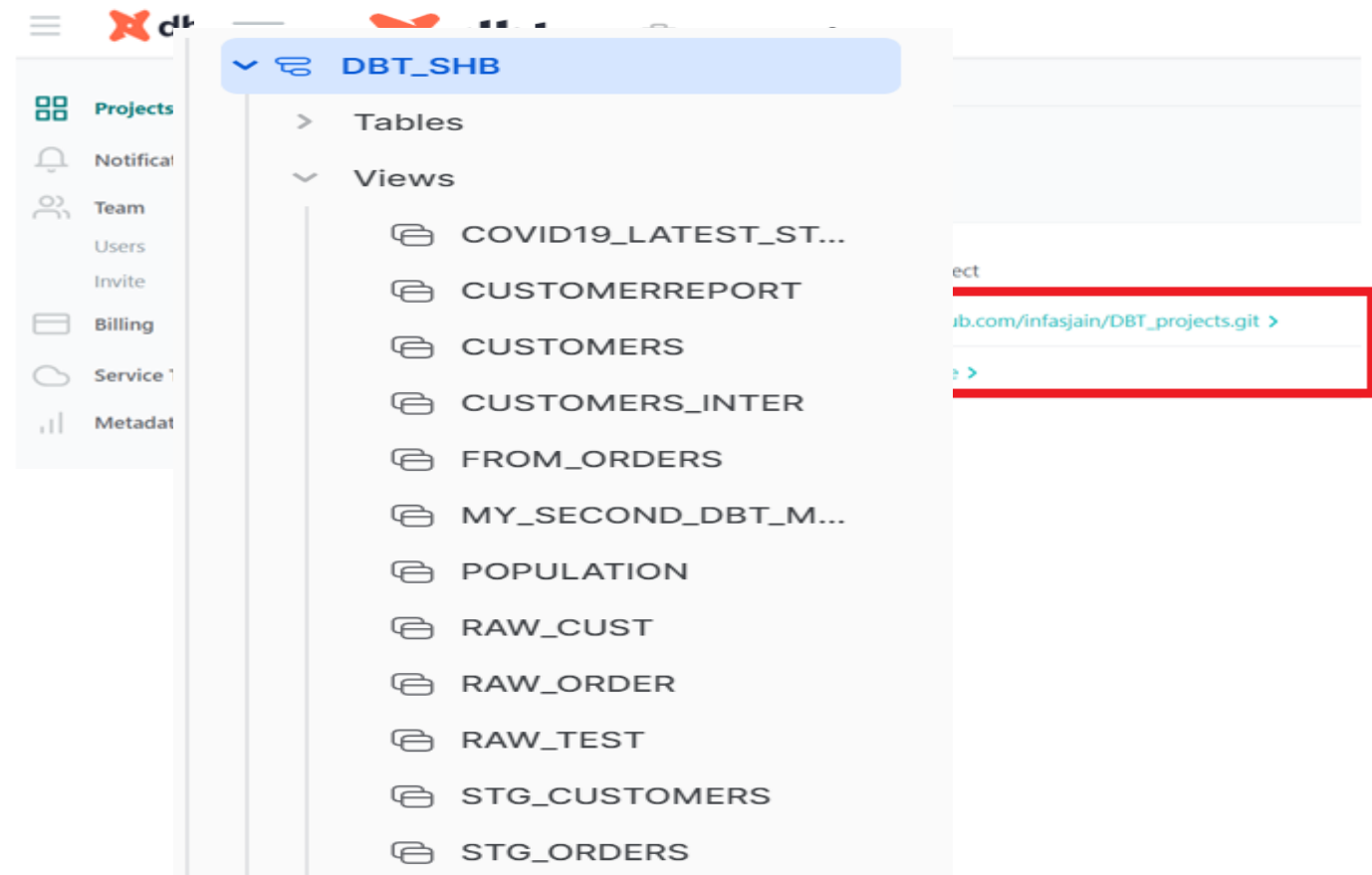
DBT and Snowflake

Data Building Tool is a command-line tool that enables data analysts and engineers to transform data in their warehouses simply by writing select statements.

Git repository and Snowflake Data warehouse Configuration

DBT – YML and SQL files

Snowflake Tables and Views



DBT Lineage Graph

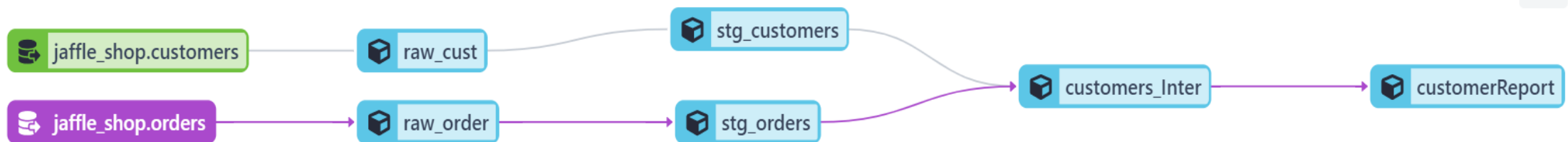
Source YML File

```
1  version: 2
2
3  sources:
4    - name: jaffle_shop
5      database: RAW
6      tables:
7        - name: orders
8        - name: customers
9
10   - name: stripe
11     tables:
12       - name: payments
```

SQL file

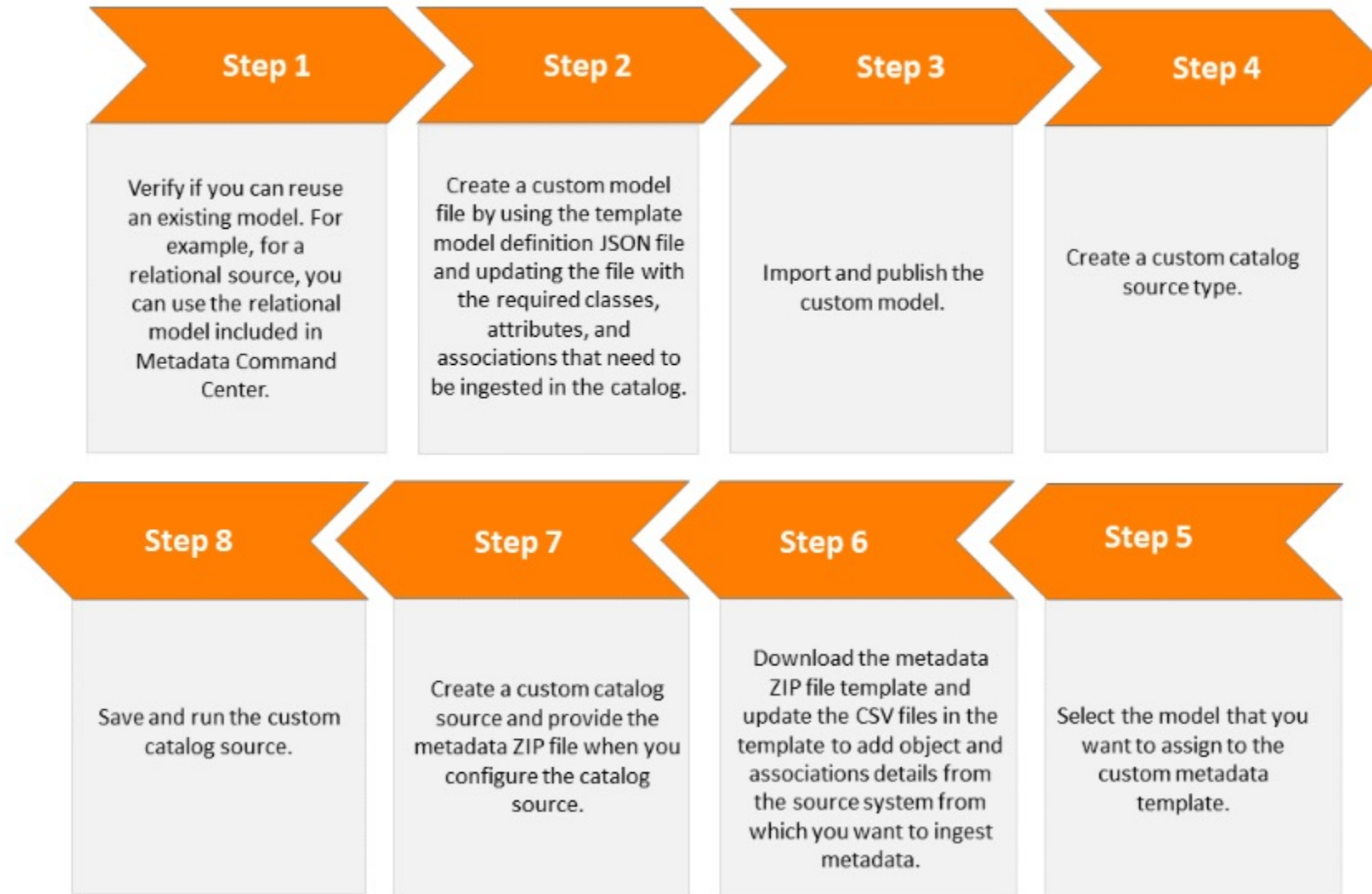
```
customerReport.sql
1  with customer as (
2
3    select * from {{ ref('customers_Inter') }}
4
5  ),
6
7  final as (
8
9    select
10      customers_Inter.customer_id,
```

Lineage Graph

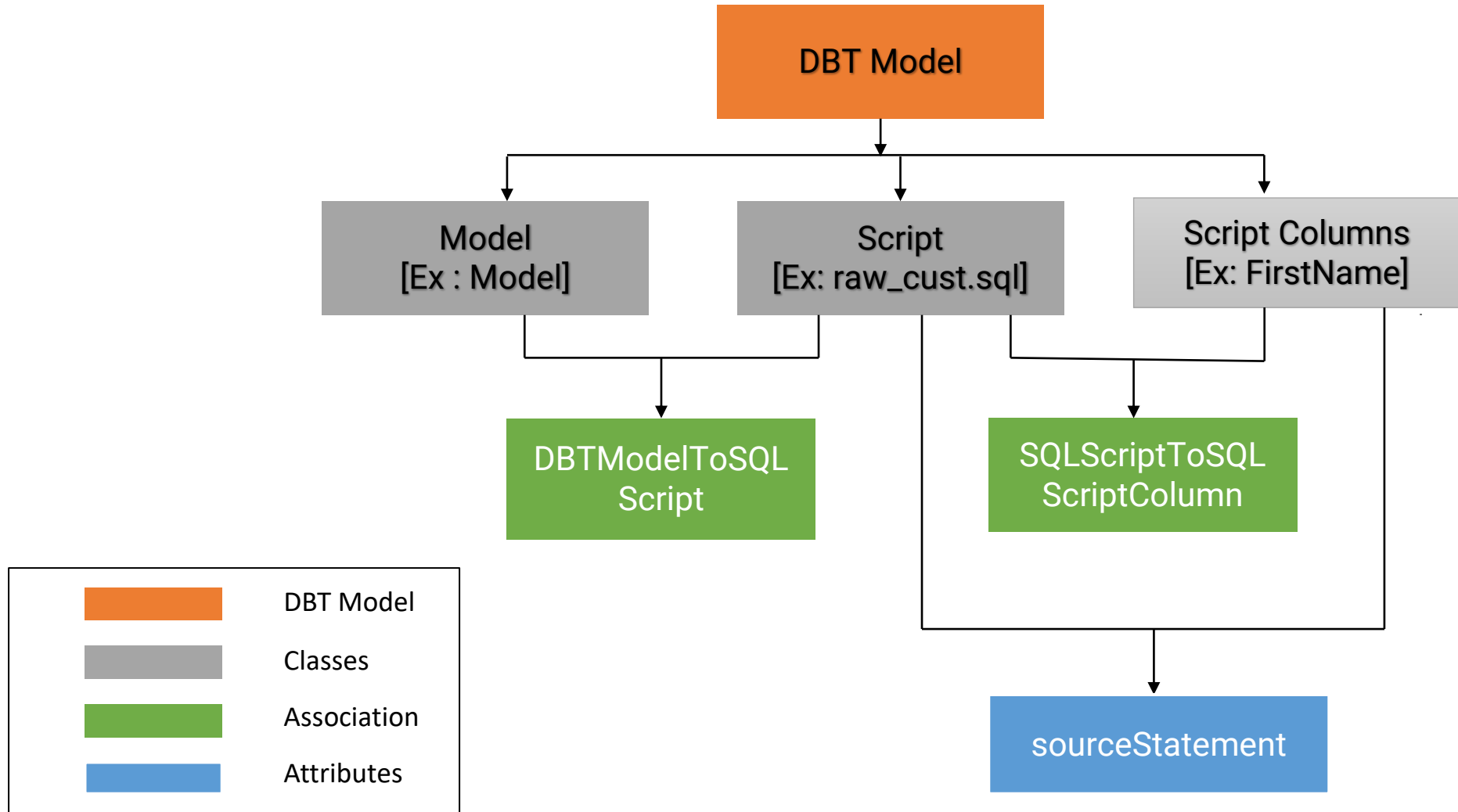


Custom Scanner

CDGC – Custom scanner steps



DBT – Custom Scanner Model and Associations



Model Definition File

Model file:

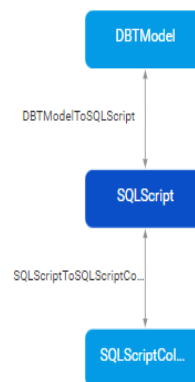
- 1.Package [custom.dbtmodel]
- 2.Classes [DBTModel , SQLScript ,SQLScriptColumn]
- 3.Association [DBTModelToSQLScript,SQLScriptToSQLScriptColumn]
- 4.Attributes [sourceStatement]

Summary: SQL Script

Properties

Associations

Attributes



Model Details

Classes

DBT Model

SQL Script

Column

Attributes

PrimaryKeyColumn

taskType

htmlLink

calculationComplexity

controlConditions

expression

sourceStatement

Associations

ModelToScript

ScriptToColumn

Metadata Definition File

custom.dbtmodel.DBTModel	Microsoft Excel Comma S...	1 KB	No	1 KB	36%	6/21/2022 3:06 PM
custom.dbtmodel.SQLScript	Microsoft Excel Comma S...	1 KB	No	1 KB	39%	6/21/2022 3:07 PM
custom.dbtmodel.SQLScriptColumn	Microsoft Excel Comma S...	1 KB	No	1 KB	47%	6/21/2022 3:10 PM
links	Microsoft Excel Comma S...	1 KB	No	1 KB	59%	6/21/2022 3:16 PM

custom.dbtmodel.SQLScript.csv

	A	B	C	D	E	F	G
1	core.externalId	core.name	core.description	core.business	core.business	core.reference	custom.dbtmodel.sourceStatement
2	DataScript1	raw_cust.sql	This has data flow from Customer to raw_cust				"select id as customer_id, first_name, last_name from raw.jaffle_shop.customers"

custom.dbtmodel.DBTModel.csv

	A	B	C	D	E	F	G	H
1	core.externalId	core.name	core.description	core.businessDescription	core.business	core.reference		
2	DBTMODEL	MODEL	THIS IS A MODEL IN DBT MODEL					

Links.csv

	A	B	C
1	Source	Target	Association
2	Resource	DBTModel	core.ResourceParentChild
3	DBTModel	DataScript1	custom.dbtmodel.DBTModelToSQLScript
4	DataScript1	DataTaskSource1OutputColumn1	custom.dbtmodel.SQLScriptToSQLScriptColumn
5	DataScript1	DataTaskSource1OutputColumn2	custom.dbtmodel.SQLScriptToSQLScriptColumn
6	DataScript1	DataTaskSource1OutputColumn3	custom.dbtmodel.SQLScriptToSQLScriptColumn

custom.dbtmodel.SQLScriptColumn.csv

	A	B	C	D	E	F	G	H	I	J	K
1	core.externalId	core.name	core.description	core.business	core.business	core.reference	custom.dbtmodel.sourceStatement				
2	DataTaskSource1OutputColumn1	CUSTOMER_ID									
3	DataTaskSource1OutputColumn2	FIRST_NAME									
4	DataTaskSource1OutputColumn3	LAST_NAME									

Custom Catalog Source

New

Choose the asset type you want to create

Find

Catalog Source

Customization

Data Classification

Lookup Table



Teradata Database

A relational database management system (RDBMS) solution for data warehousing applications

Custom Catalog Sources (4)



DBT_Snowflake

This is a placeholder created for DBT Model



demo_cst



Demo_DBTModel



sachin_dbt

New Catalog Source

< Back

Next >

Save

1 Registration

2 Configuration

3 Schedule

Description:

Connection Information

Metadata Source Type:

DBT_Snowflake

Source Type: * ?

☒ Upload

☐ Provide a Local Path

[Download Template](#)

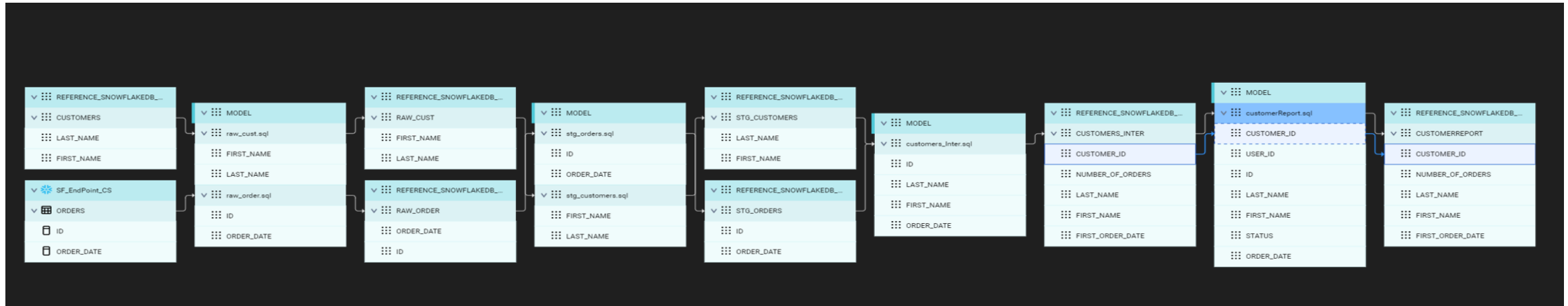
File Details: * ?

Select a file in the ZIP format

Browse

Lineage and Connection Assignment

- *Scan the Snowflake resource for connection assignment as the reference assets are replaced by actual physical assets*



References

- Cloud Data Governance and Catalog: [Click Here](#)
- Introduction and Getting Started: [Click Here](#)
- Governance: [Bulk Upload Business Metadata](#)
- Custom Scanner: [Click Here](#)
- Metadata Command Center:
 - [Catalog Scanner Configuration](#)
 - [Creating Catalog Source](#)
 - [Configure Runtime Environments](#)

DEMO

CDGC – Custom Scanner Case Study

