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

PRT069

SPONSORED BY SELECT STAR

# Automated data discovery: Break free from hours of manual data hunting

Shinji Kim

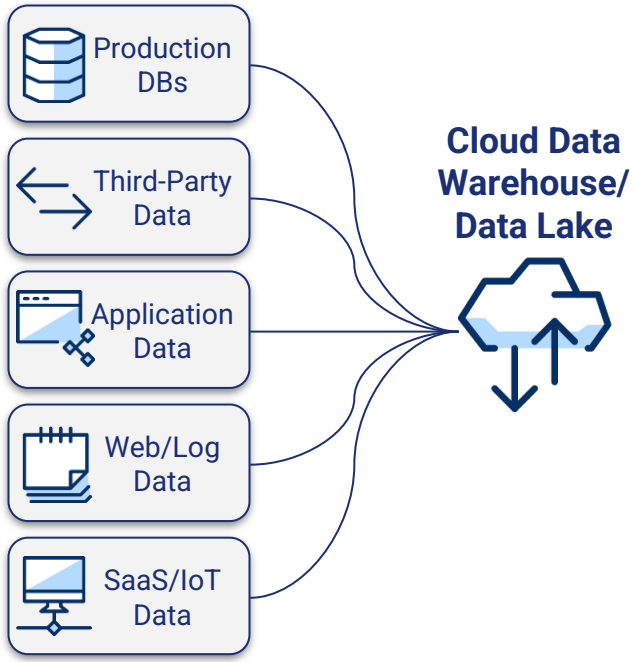
Founder and CEO
Select Star

aws

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

# Changes to the modern data stack have led organizations to have more data, and more data users, than ever before

Explosion in data volume and data sources going to CDWs



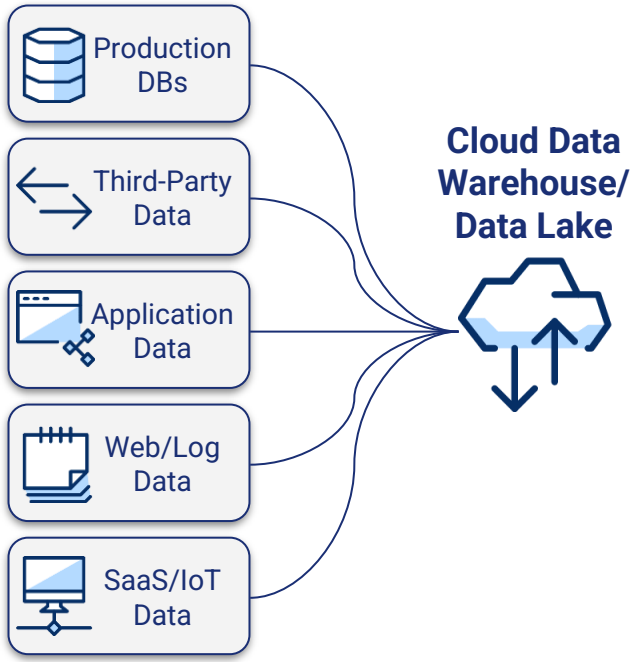

© Select Star



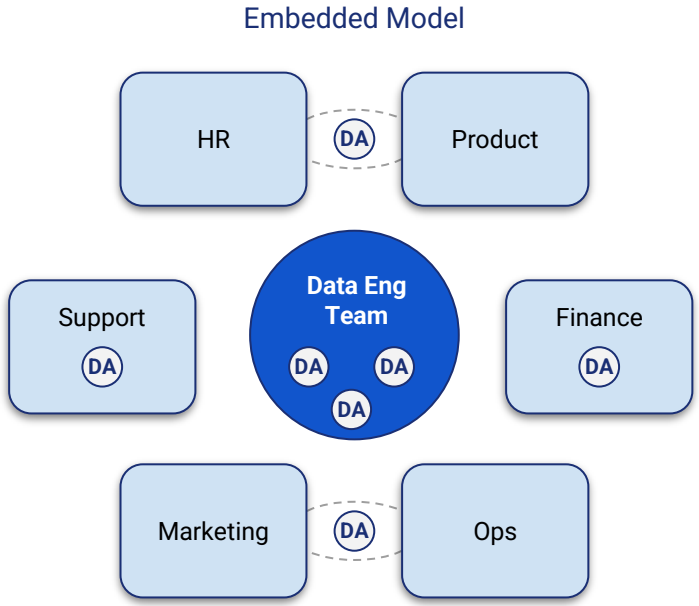
PRIVATE & CONFIDENTIAL

# Changes to the modern data stack have led organizations to have more data, and more data users, than ever before

## Explosion in data volume and data sources going to CDWs




## Decentralization of data ownership



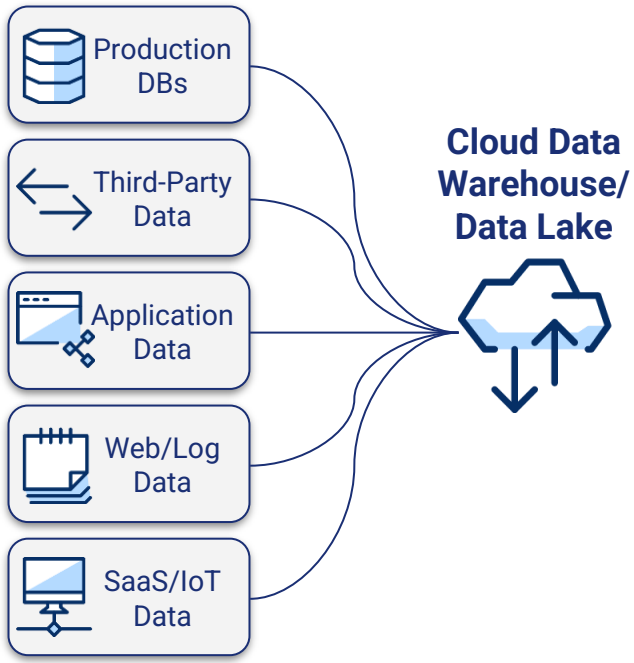

© Select Star



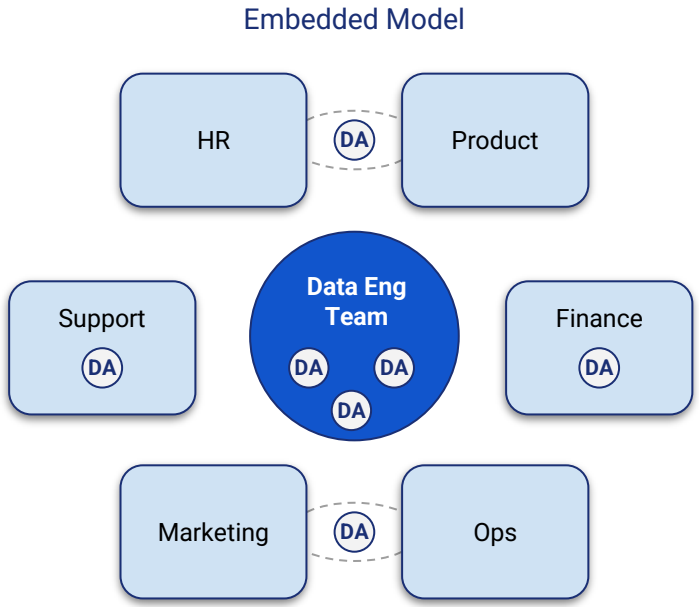
PRIVATE & CONFIDENTIAL

# Changes to the modern data stack have led organizations to have more data, and more data users, than ever before

## Explosion in data volume and data sources going to CDWs




## Decentralization of data ownership




## Democratization of data access




© Select Star



PRIVATE & CONFIDENTIAL

# Finding and understanding data is a big challenge in many organizations today

**Marketing Analyst:** Where can I find the customer engagement numbers?

**Product Manager:** Where did the data come from?



**General Manager:** How did you calculate this metric?

**Data Scientist:** Do we have this data?

**Finance Analyst:** Why does my revenue number look different than yours?

PRIVATE & CONFIDENTIAL

© Select Star

# Some of the existing solutions are resource intensive and/or don’t meet today’s needs

## Wiki/Documentation

**Outdated Information:**

Manual documentation often can’t keep up with the changes in the DWH/BI tool

## Custom DIY/Open Source

**Heavy Engineering Investment:**

6–12 months to productionize a custom tool that will require continuous investment for upkeep

## Enterprise Data Catalog

**Poor UX and Manual Workflow:**

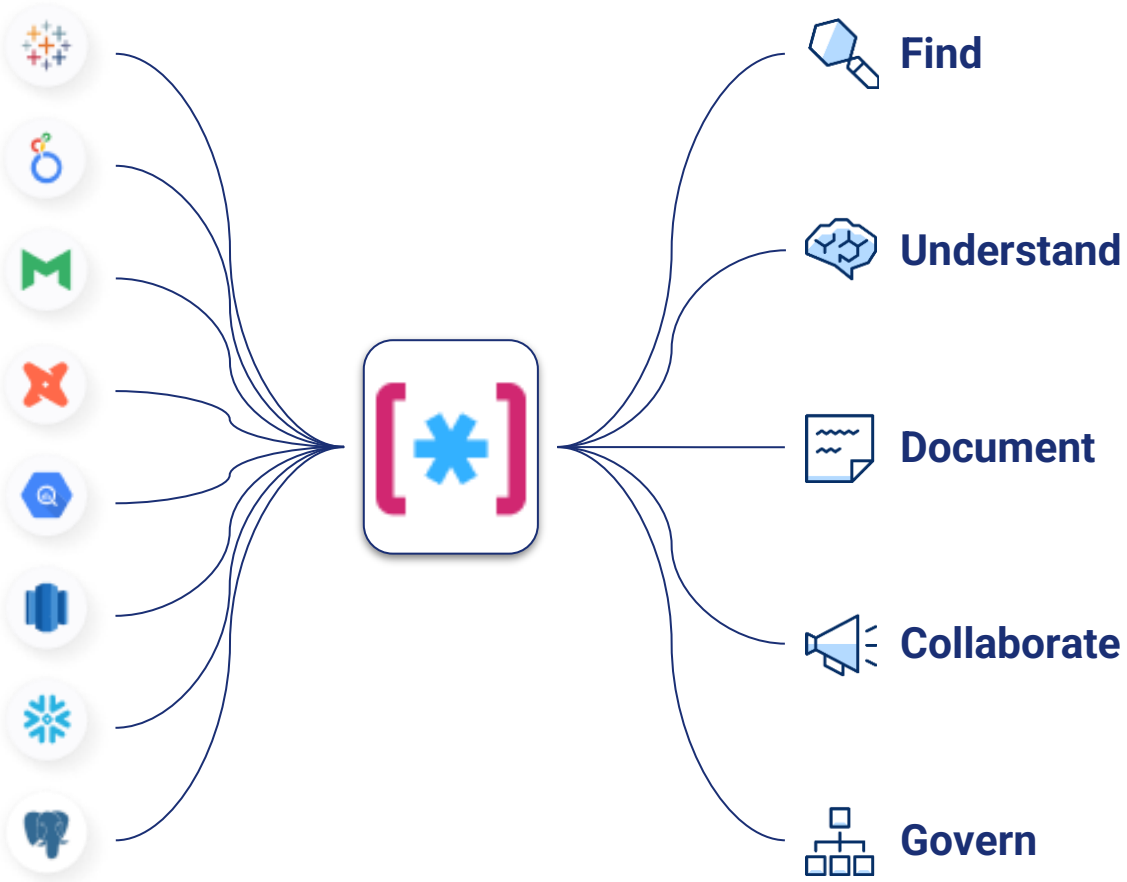
Upfront financial investment, dedicated data governance team with training required

© Select Star

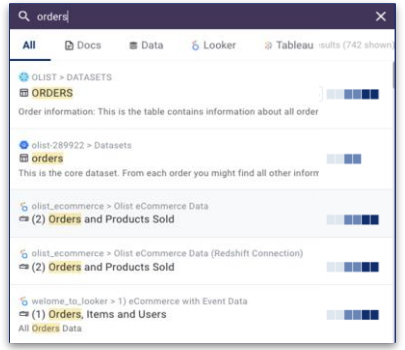


PRIVATE & CONFIDENTIAL

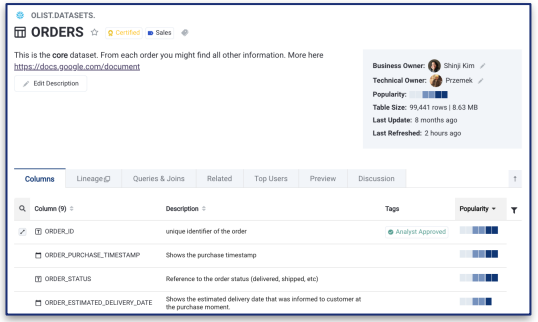
# Data discovery creates a single source of truth for data within your organization, saving time and resources while increasing trust



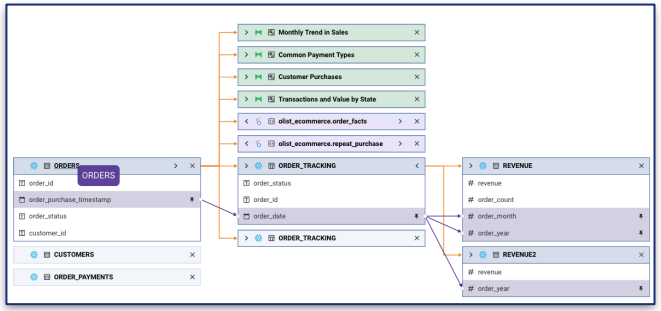

*Universal Search*




*Automated Insights and Documentation*




*Column-Level Lineage*



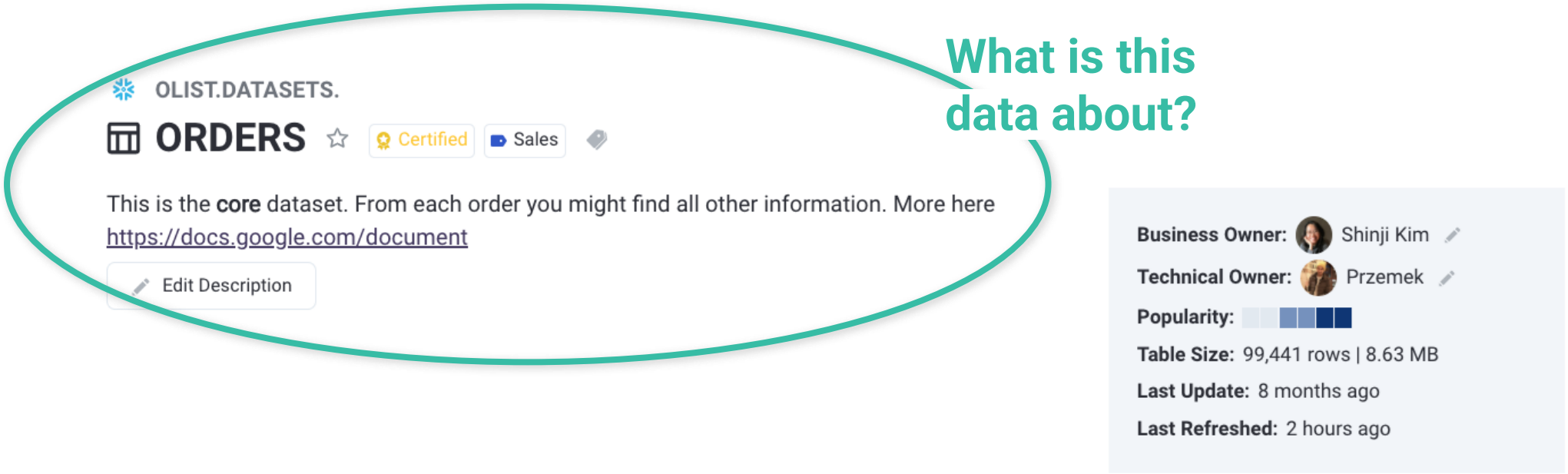

© Select Star



PRIVATE & CONFIDENTIAL

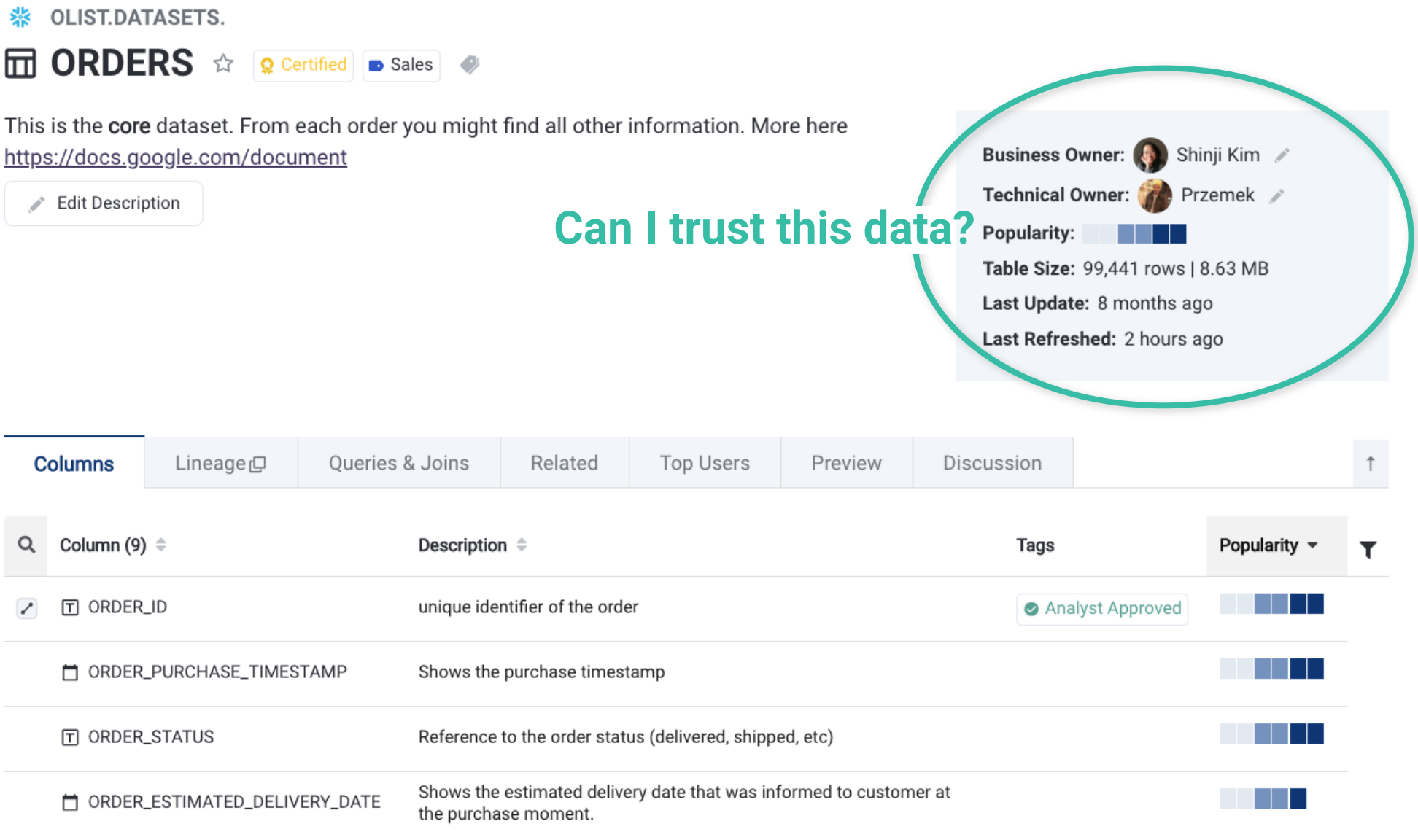
# Select Star: Automated data discovery and governance

What is this data about?



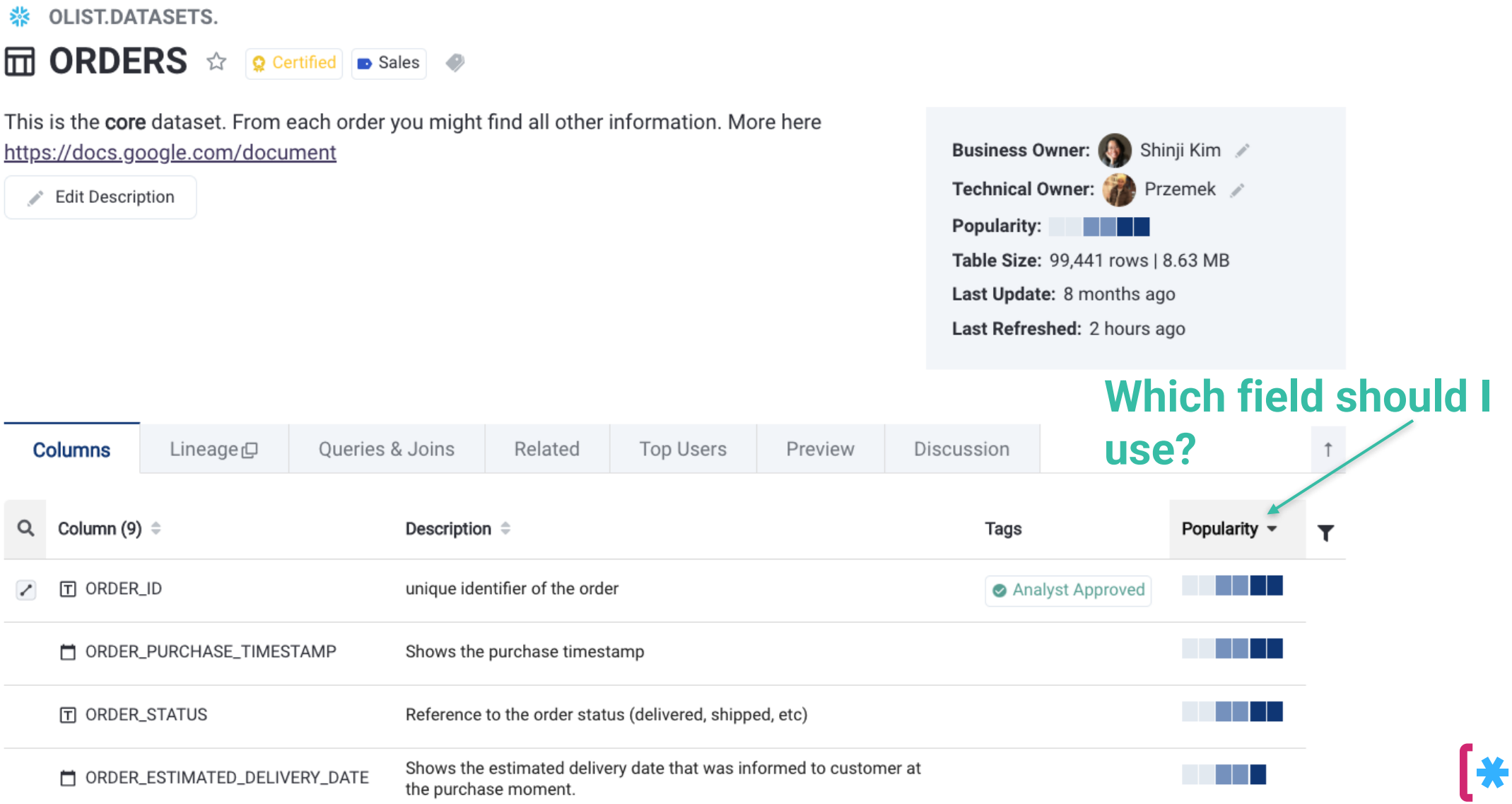

© Select Star

# Select Star: Automated data discovery and governance



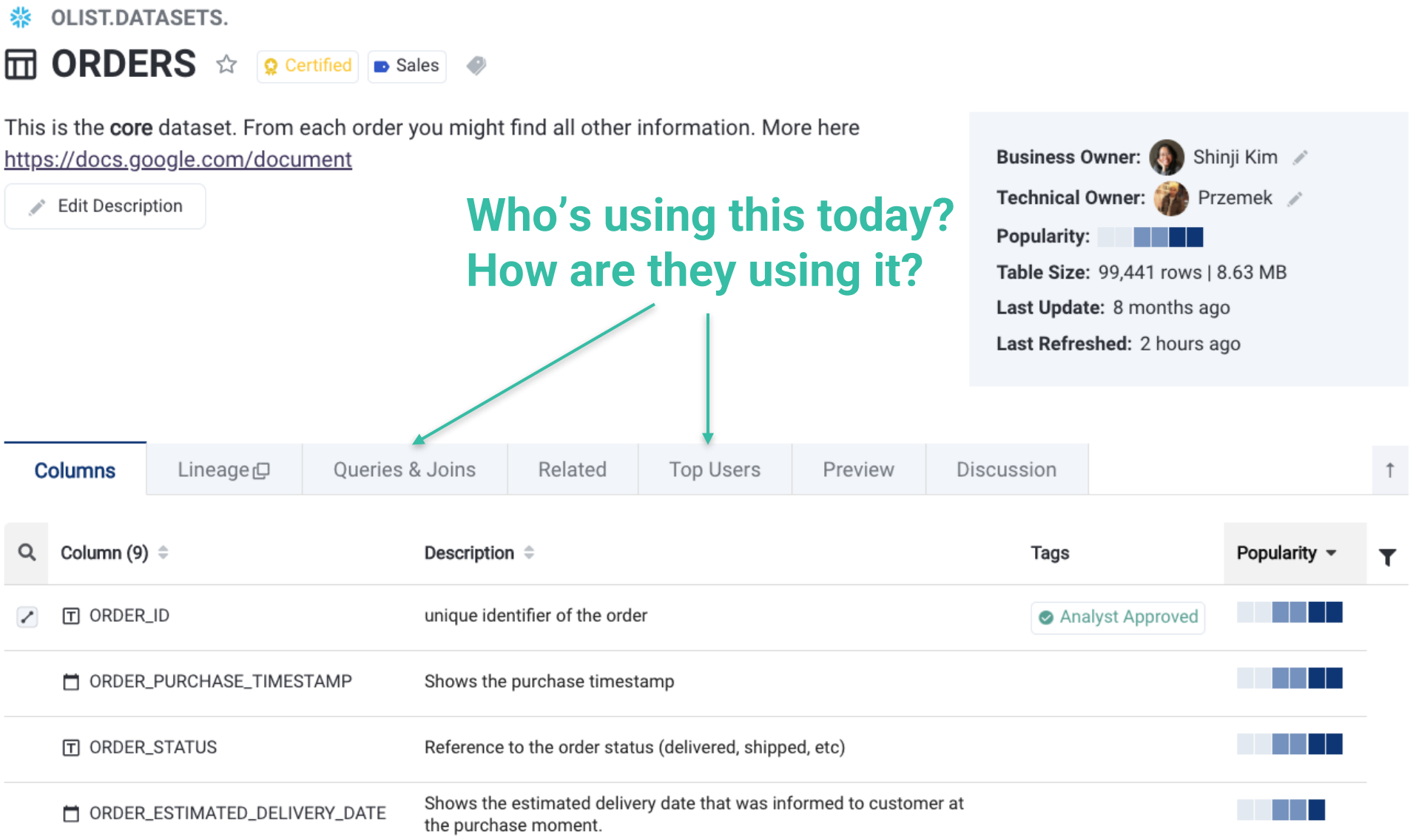

© Select Star

# Select Star: Automated data discovery and governance

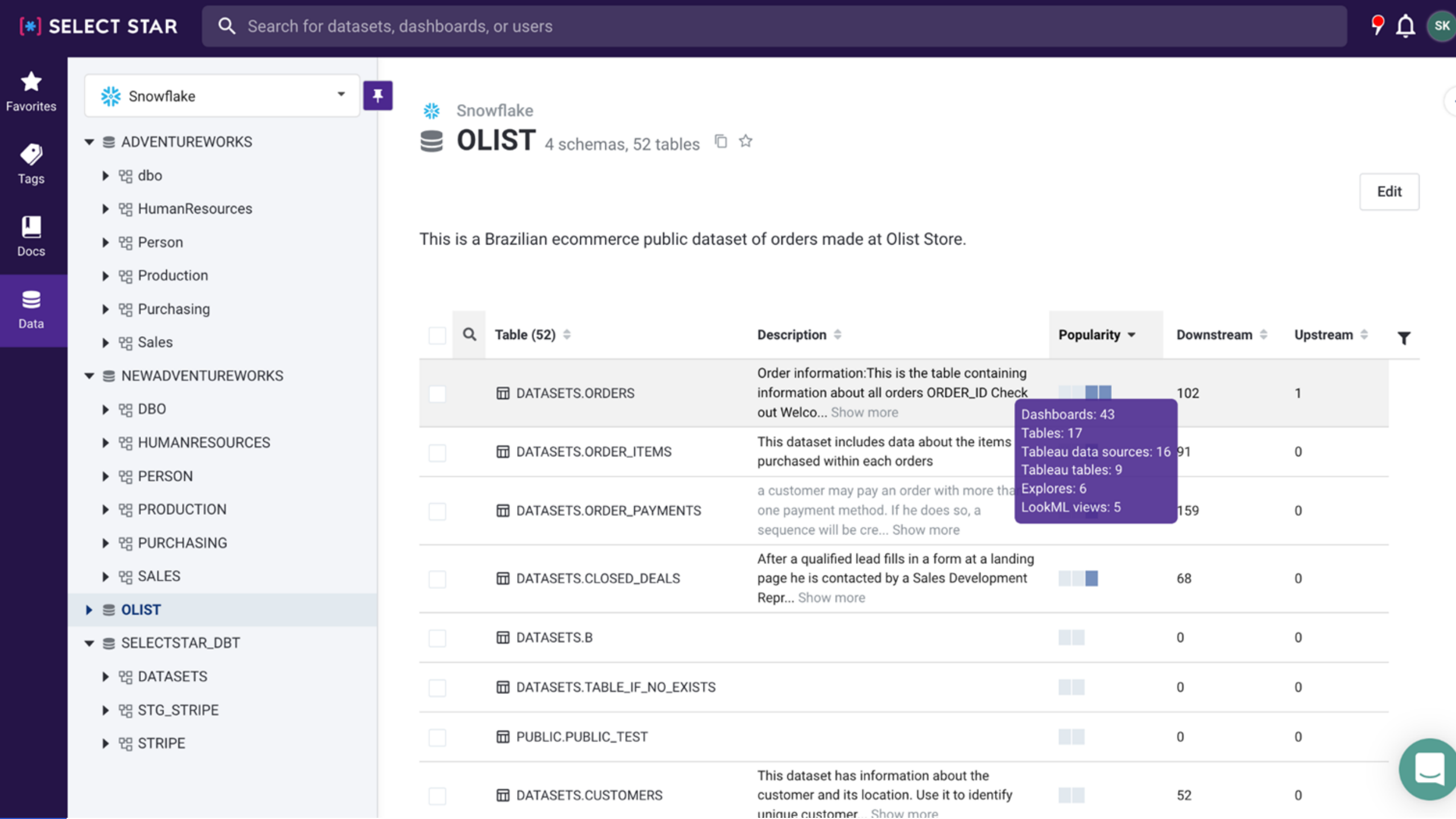



© Select Star

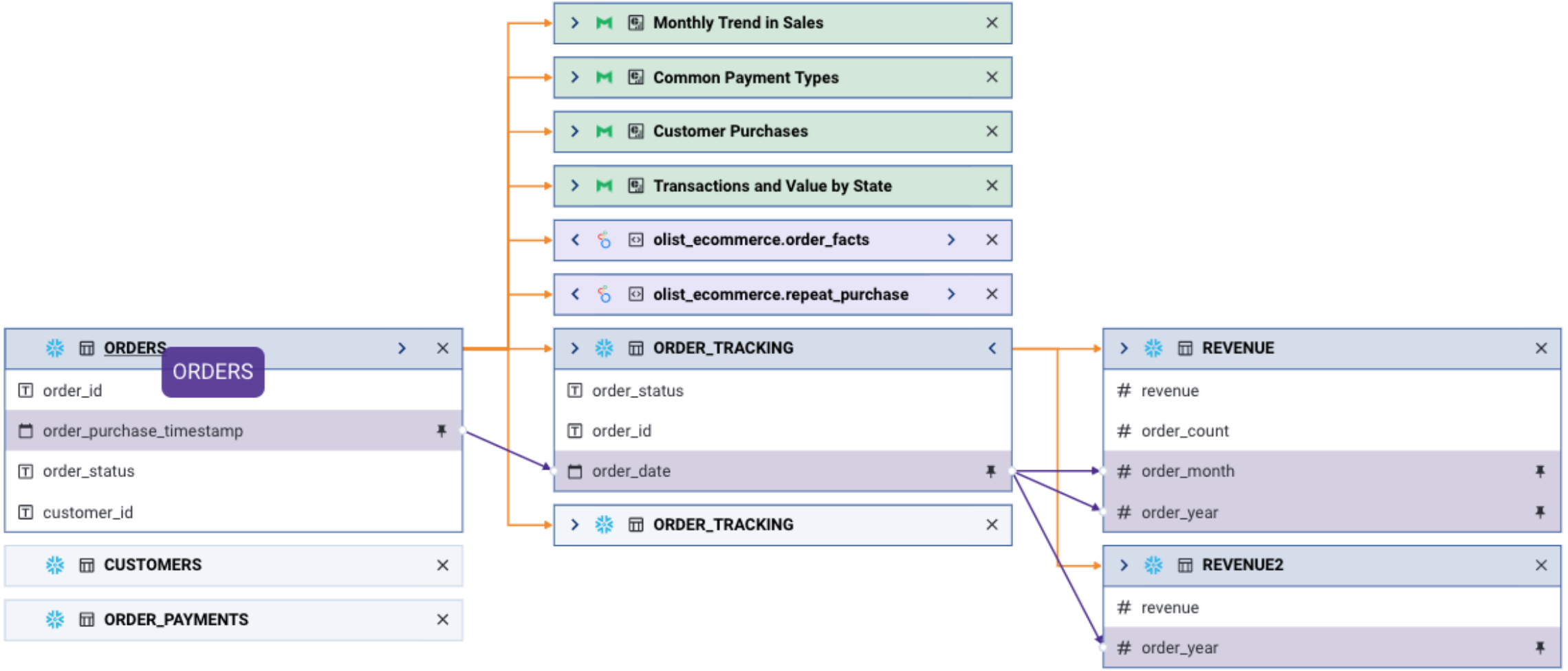
# Select Star: Automated data discovery and governance




© Select Star


SELECT STAR
Search for datasets, dashboards, or users
SK
Favorites
Tags
Docs
Data
Snowflake
ADVENTUREWORKS
dbo
HumanResources
Person
Production
Purchasing
Sales
NEWADVENTUREWORKS
DBO
HUMANRESOURCES
PERSON
PRODUCTION
PURCHASING
SALES
OLIST
SELECTSTAR_DBT
DATASETS
STG_STRIPE
STRIPE
Snowflake
OLIST 4 schemas, 52 tables
Edit
This is a Brazilian ecommerce public dataset of orders made at Olist Store.
Table (52)
Description
Popularity
Downstream
Upstream
DATASETS.ORDERS
Order information:This is the table containing information about all orders ORDER_ID Check out Welco... Show more
102
1
Dashboards: 43
Tables: 17
Tableau data sources: 16
Tableau tables: 9
Explores: 6
LookML views: 5
DATASETS.ORDER_ITEMS
This dataset includes data about the items purchased within each orders
91
0
DATASETS.ORDER_PAYMENTS
a customer may pay an order with more tha one payment method. If he does so, a sequence will be cre... Show more
159
0
DATASETS.CLOSED_DEALS
After a qualified lead fills in a form at a landing page he is contacted by a Sales Development Repr... Show more
68
0
DATASETS.B
0
0
DATASETS.TABLE_IF_NO_EXISTS
0
0
PUBLIC.PUBLIC_TEST
0
0
DATASETS.CUSTOMERS
This dataset has information about the customer and its location. Use it to identify unique customer... Show more
52
0

# Select Star: Automated data discovery and governance

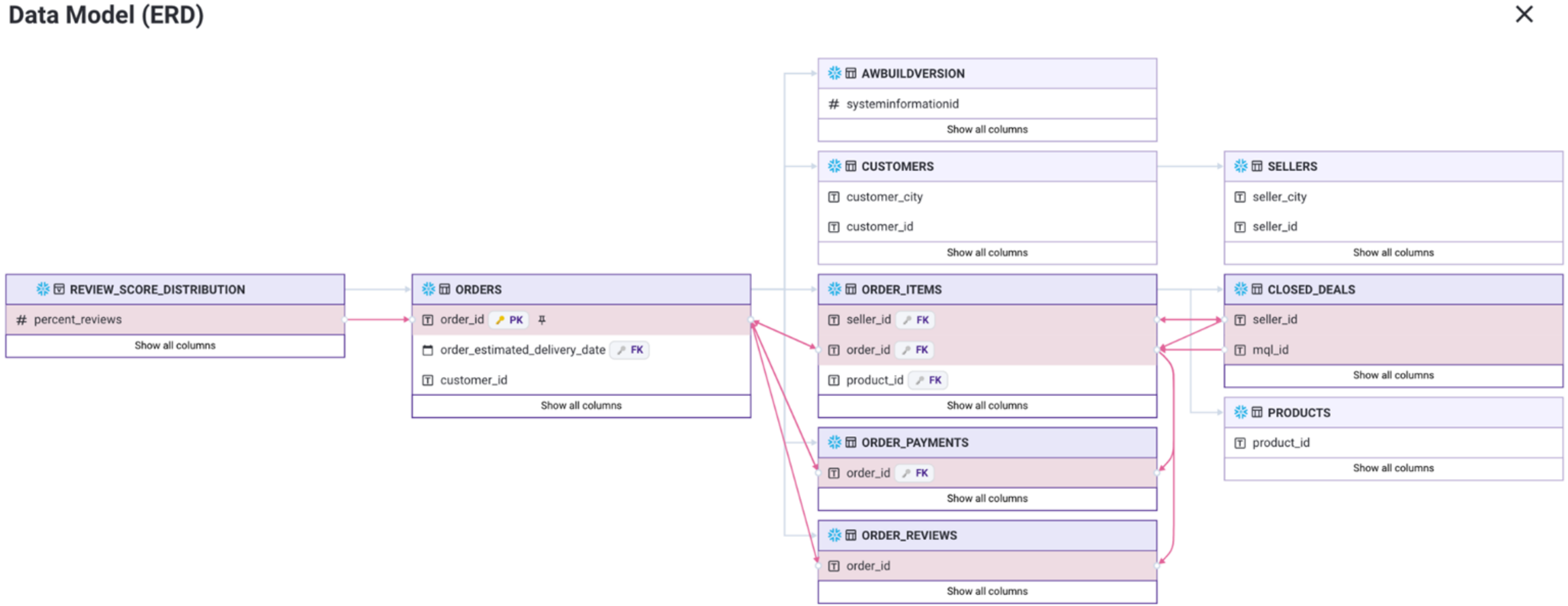



© Select Star



PRIVATE & CONFIDENTIAL

# Select Star: Automated data discovery and governance




© Select Star



PRIVATE & CONFIDENTIAL

# A growing number of customers rely on Select Star for their data discovery needs

















































© Select Star



PRIVATE & CONFIDENTIAL

# What to consider and evaluate when looking for data discovery tools?



## Fully Automated Approach

How long does it take to integrate and get it running with your own data?

## Continuous Metadata Analytics

Automated data lineage and usage models provide insights on who's using what and how each dataset is being used inside the organization.

## Simple and Easy To Use

Is it designed not just for the engineers and analysts, but also for Ops, PMs, and business stakeholders to use?

## System of Record for Company Data

Beyond physical metadata, what's the support like for docs, metrics, and collaboration? These allow you to use it as a true knowledge base of organization data for everyone.

© Select Star



PRIVATE & CONFIDENTIAL

# Thank you!

Shinji Kim
hello@selectstar.com
@selectstarhq
@shinjikim



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

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