Collibra

THE DATA INTELLIGENCE CONFERENCE

DATA CITIZENS 22

Collibra + Snowflake: Better together with data lineage

Sheeri Cabral
Senior Product Manager
Collibra

Paul Gancz
Partner Solutions Architect
Snowflake

Sheeri Cabral

Senior Product Manager

Collibra

Paul Gancz

Partner Solutions Architect

Snowflake

Agenda

1. Data intelligence starts with the data
2. Business challenges
3. Collibra + Snowflake
4. Snowflake lineage with Collibra
5. Demo
6. Q&A

Data intelligence starts with the data

Personas

Business Analyst | Data Steward | Data Engineer | Data Privacy Manager | Data Scientist

System of Engagement

Team productivity

salesforce — *GTM*

servicenow — *IT & Ops*

Collibra — *Data*

ATLASSIAN — *Engineering*

workday. — *HR*

Data Tools

Individual productivity

ETL / ELT | Data Modeling | Data Science | Dashboards & Analytics

Data Infrastructure

Compute & storage

Data Warehouse | Data Lake | Database / OLTP | Streaming | Query engines | Real time Analytics

Hyperscalers

Business challenges

Limited visibility

Lack a comprehensive view of data

No data provenance

Unclear where and how data flows within and between systems

Lack of trust

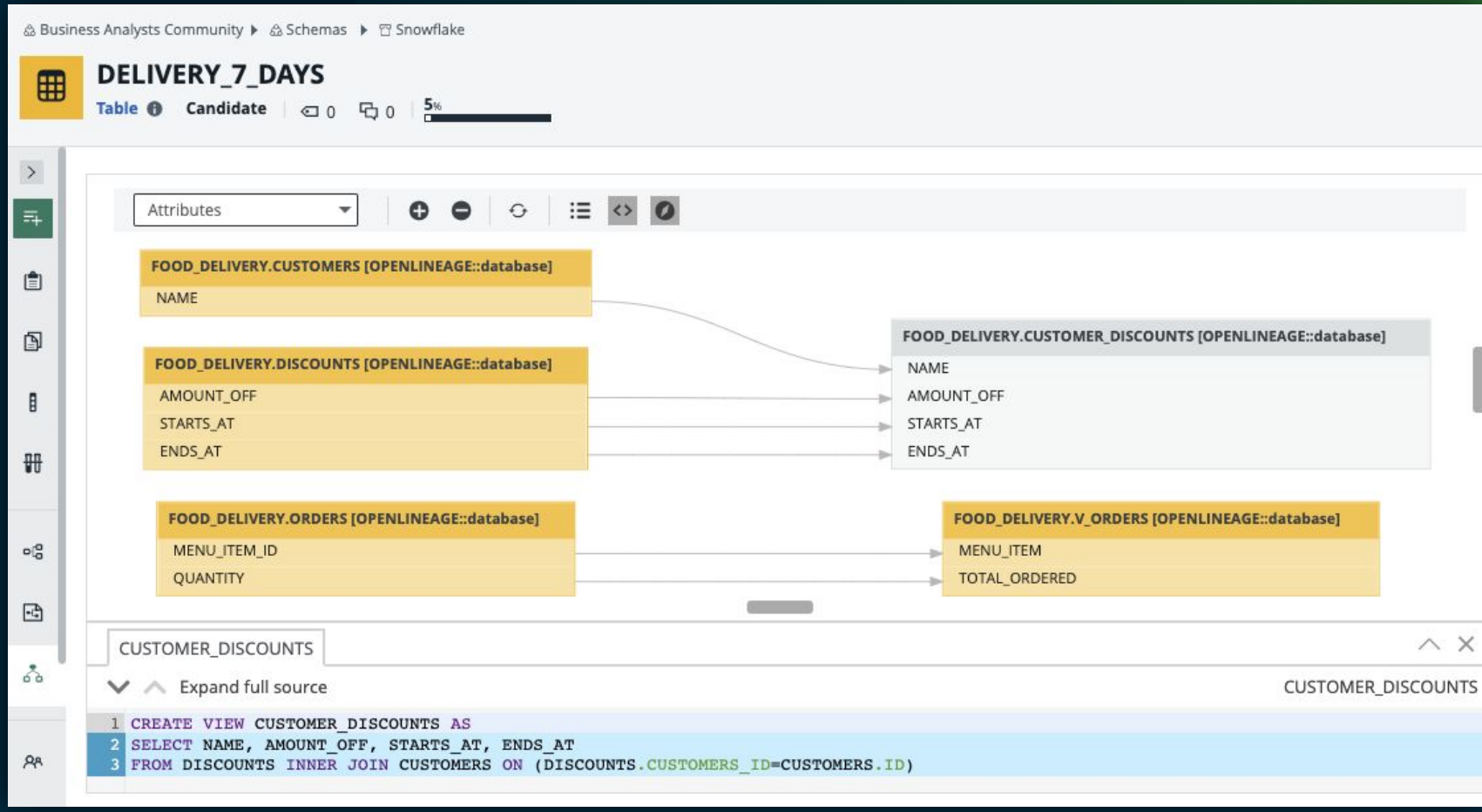
Low confidence that data can be trusted and ready to use

Impacting your business with lineage

Get a **complete, context-rich view** of where data comes from and trust that it is relevant and trustworthy

Trace data flows and show relationships, so users can quickly **understand your data's history**

Predict the impact of potential changes with lineage at the table, column, transformation and SQL query level



Collibra & Snowflake partnership

- **Snowflake Elite Partner**
  Since 2020
- **Snowflake Ventures**
  Investment
- **Snowflake Partner Network Competency Program**
  Healthcare and Financial Services

Snowflake platform
Under the hood

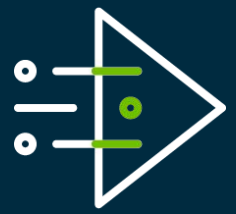
DATA

Unstructured
Semi-Structured
Structured
Streaming

OUTCOMES

Insights
Predictions
Monetization
Data Products

Why the Snowflake data cloud?



Fast for any Workload

Run virtually any number or type of job across users and data volumes quickly and reliably.

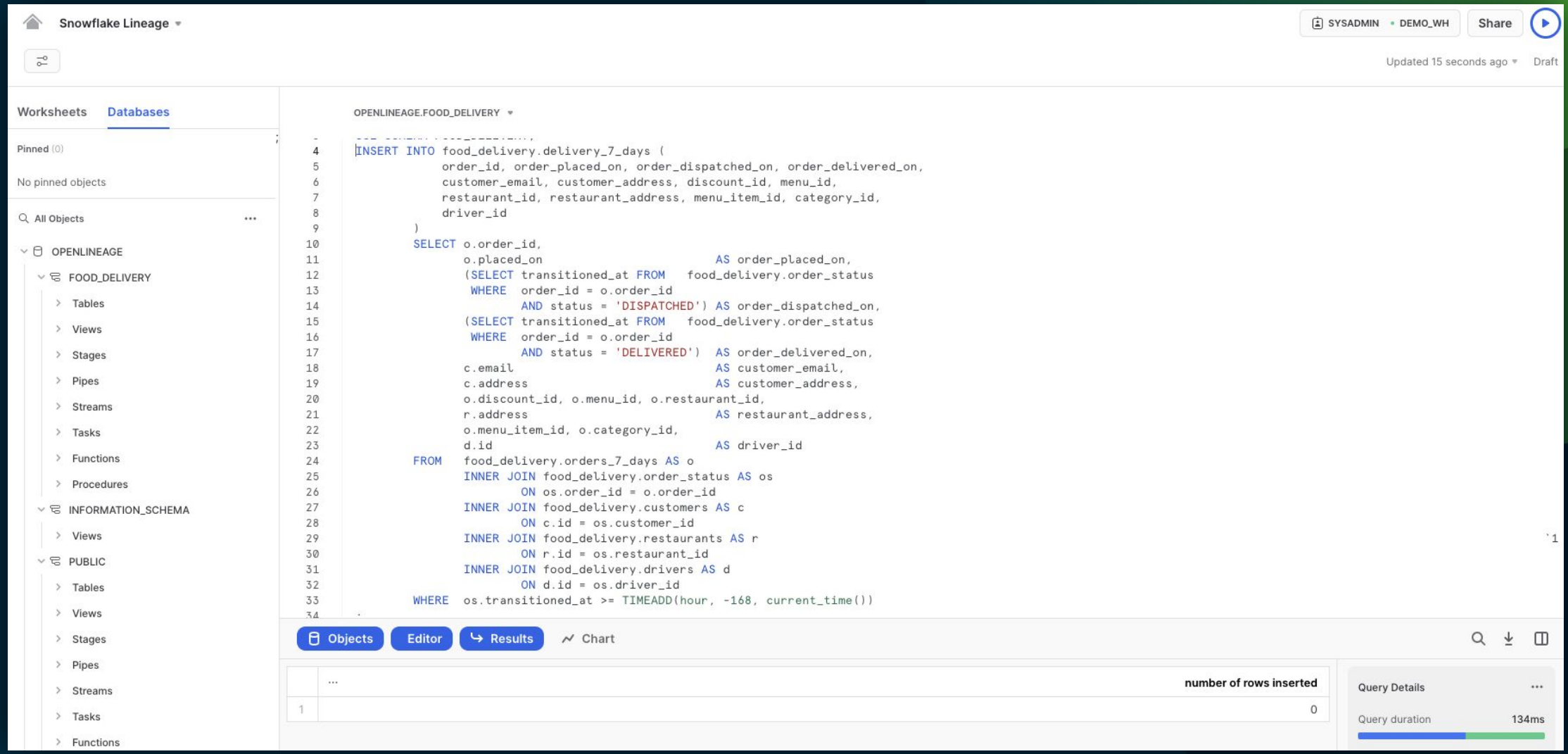
It just works

Replace manual with automated to operate at scale, optimize costs, and minimize downtime.

Connected to what matters

Extend access and collaboration across teams, workloads, clouds, and data, seamlessly and securely.

Snowflake in action

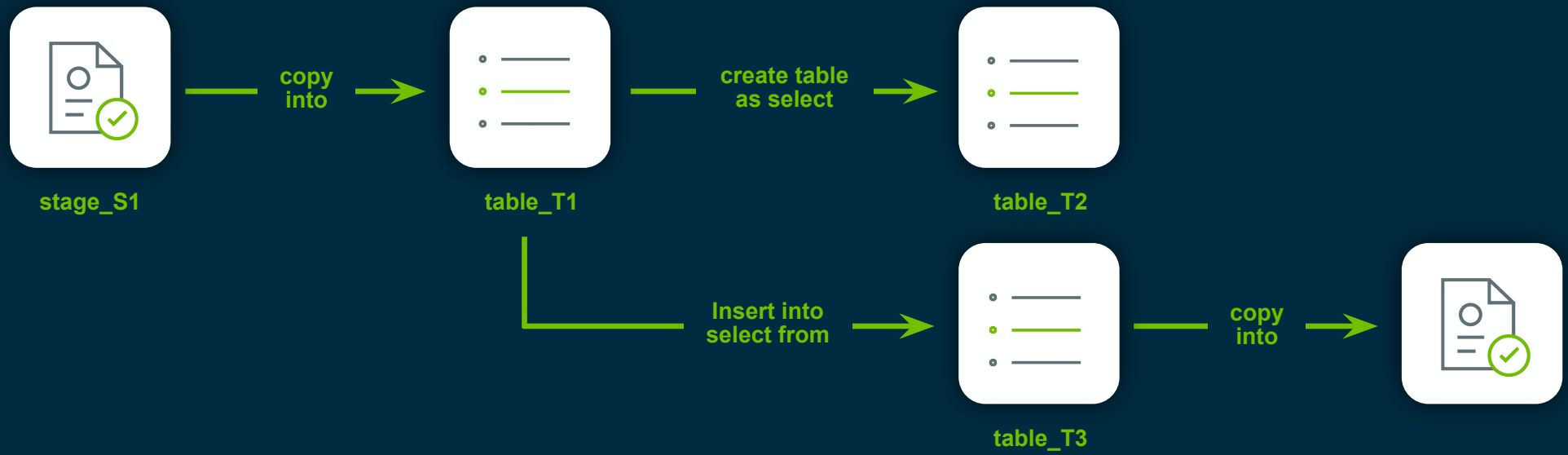


DATA CITIZENS 22

Demo

```
  "sources" : [
    {
      "id" : "snowOL",
      "type" : "DatabaseSnowflake",
      "mode" : "SQL-API",
      "collibraSystemName" : "Snowflake",
      "hostname" : "collibr[redacted].snowflakecomputing.com",
      "auth" : {
        "type" : "Basic",
        "username" : "CERTIFICATION"
      },
      "customConnectionProperties": "role=SYSADMIN",
      "databaseNames" : [ "OPENLINEAGE", "CERTIFICATION" ],
      "warehouse": "DEMO_WH"
    }
  ]
}
Sheeris-MacBook-Pro ~ ; JAVA_OPTS='--add-opens java.base/java.nio=ALL-UNNAMED' bin/lineage-harvester full-sync
```

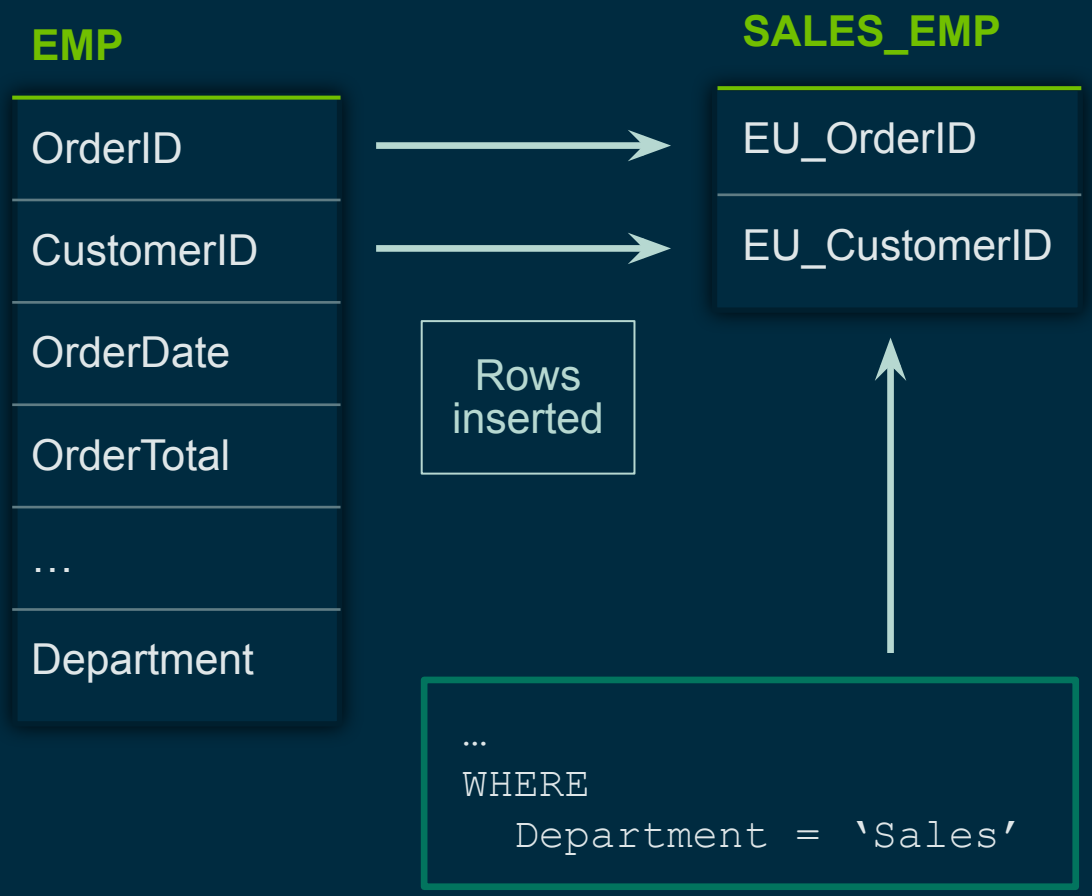
Snowflake access history (writes)
Know the lineage of data



PATH	TARGET_NAME	TARGET_DOMAIN	TARGET_COLUMNS
stage_S1–>table_T1	table_T1	TABLE	[“CONTENT”]
stage_S1–>table_T1–>table_T2	table_T2	TABLE	[“ID”,”NAME”]
stage_S1–>table_T1–>table_T3	table_T3	TABLE	[“NAME”,”ID”]
stage_S1–>table_T1–>table_T3->stage_S2	stage_S2	STAGE	[]

Column lineage

Track flow of sensitive columns, recorded in Access History

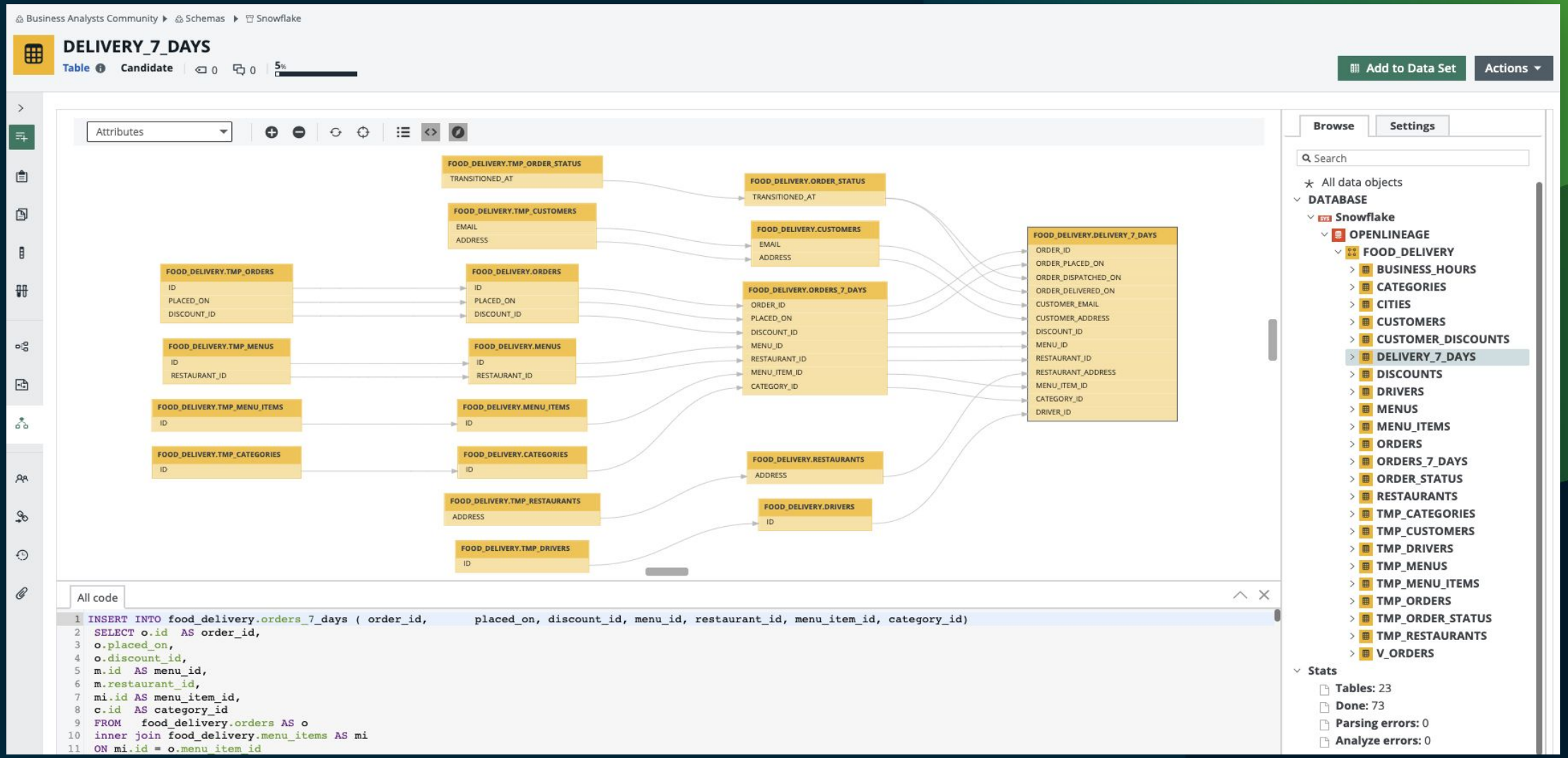


Access History (Current)

Source	Target
EMP	SALES_EMP

Just Released!

Source	Target
EMP.OrderID	SALES_EMP.EU_OrderID
EMP.CustomerID	SALES_EMP.EU_CustomerID



Questions?

Special thanks:
Raja Balakrishnan
Daphnee Geyer
Mihail Istrate

THE DATA INTELLIGENCE CONFERENCE

DATA CITIZENS 22

Next generation of Snowflake lineage

Snowflake SQL-API Now in public beta!

- Dedicated Snowflake scanner for integration
- View technical lineage including column-level lineage and transformations done by:
  - Views
  - Storage procedures
  - Queries/streams
  - Snowpipes

Snowflake SQL - GA

- Leverages SQL scanner for Snowflake integration
- Stored procedures are not parsed
- Scanner is one-directional