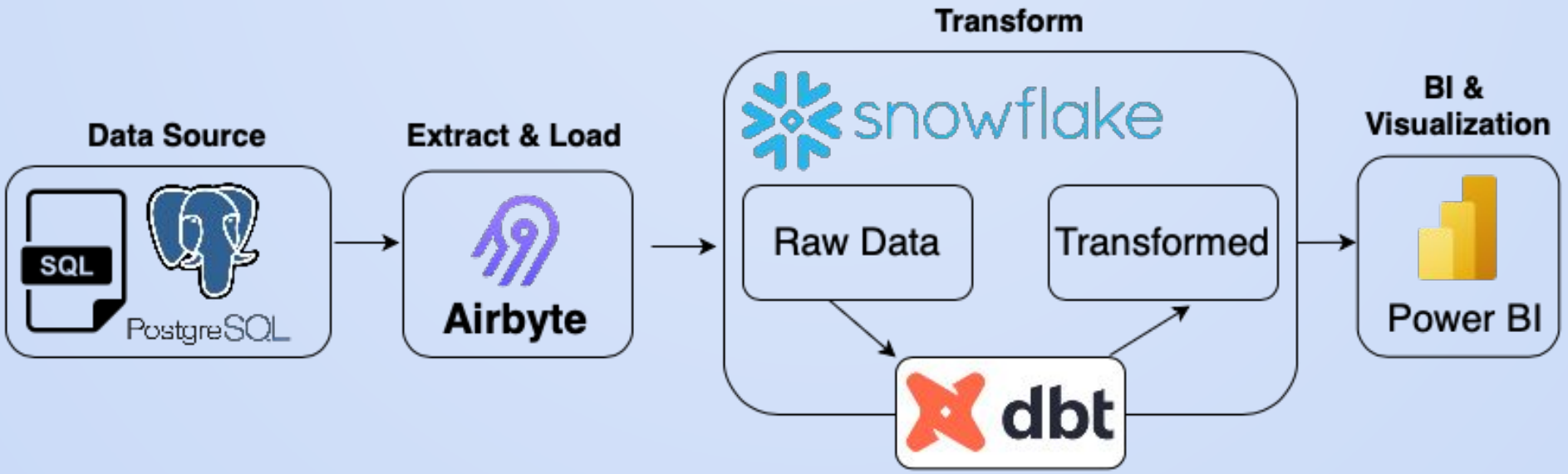
GROUP 3 PROJECT 2

NORTHWIND DATA WAREHOUSE

PROJECT CONTEXT

- Project Context
  - To help Northwind Store track orders, shipments, inventory and customer information with an ELT
- Goals
  - Business solutions relating to revenue, inventory level, shipment costs, order volume
  - This data pipeline would be valuable to Northwind to increase efficiency and lower costs

SOLUTIONS ARCHITECTURE DIAGRAM (ELT)

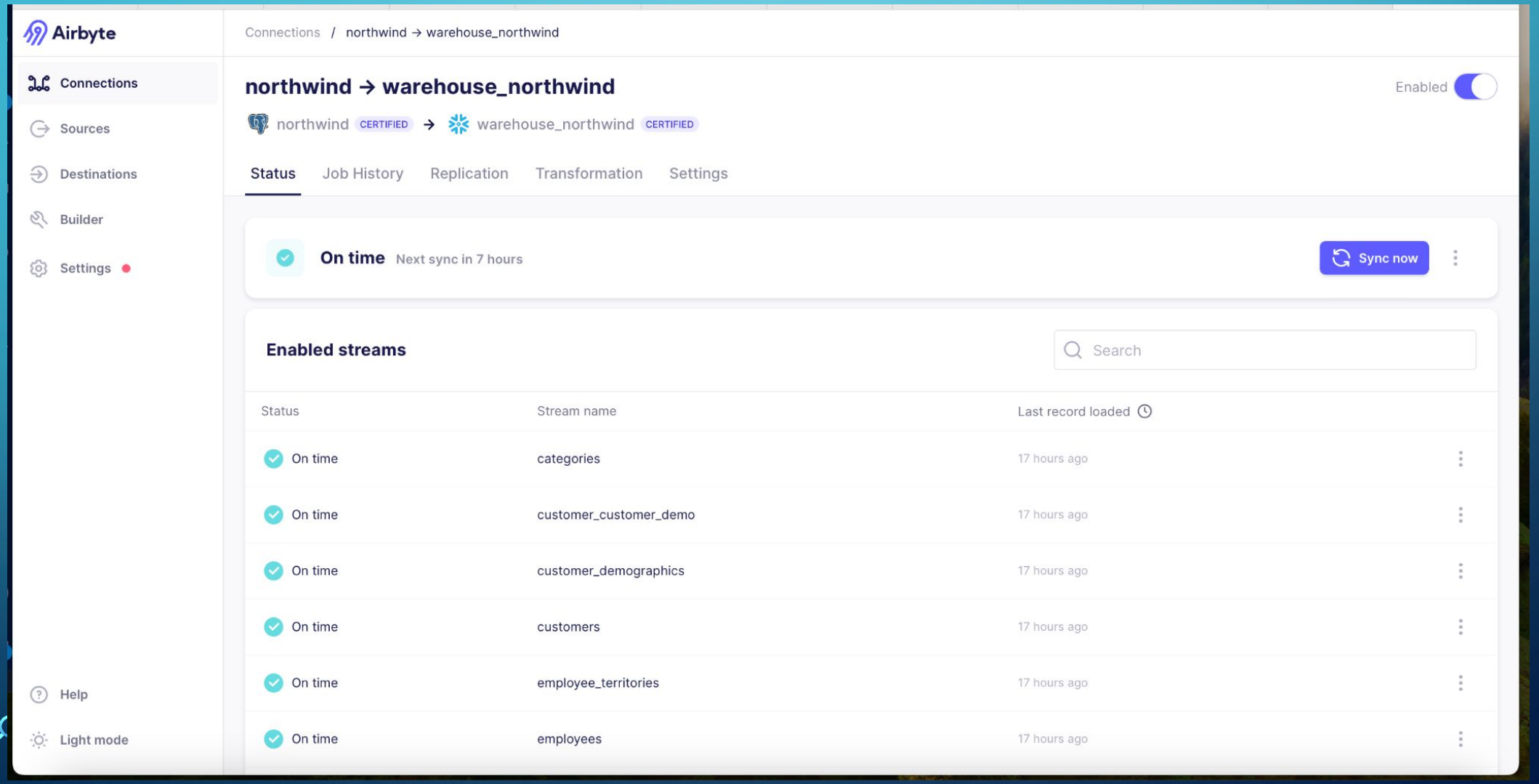


docker

aws

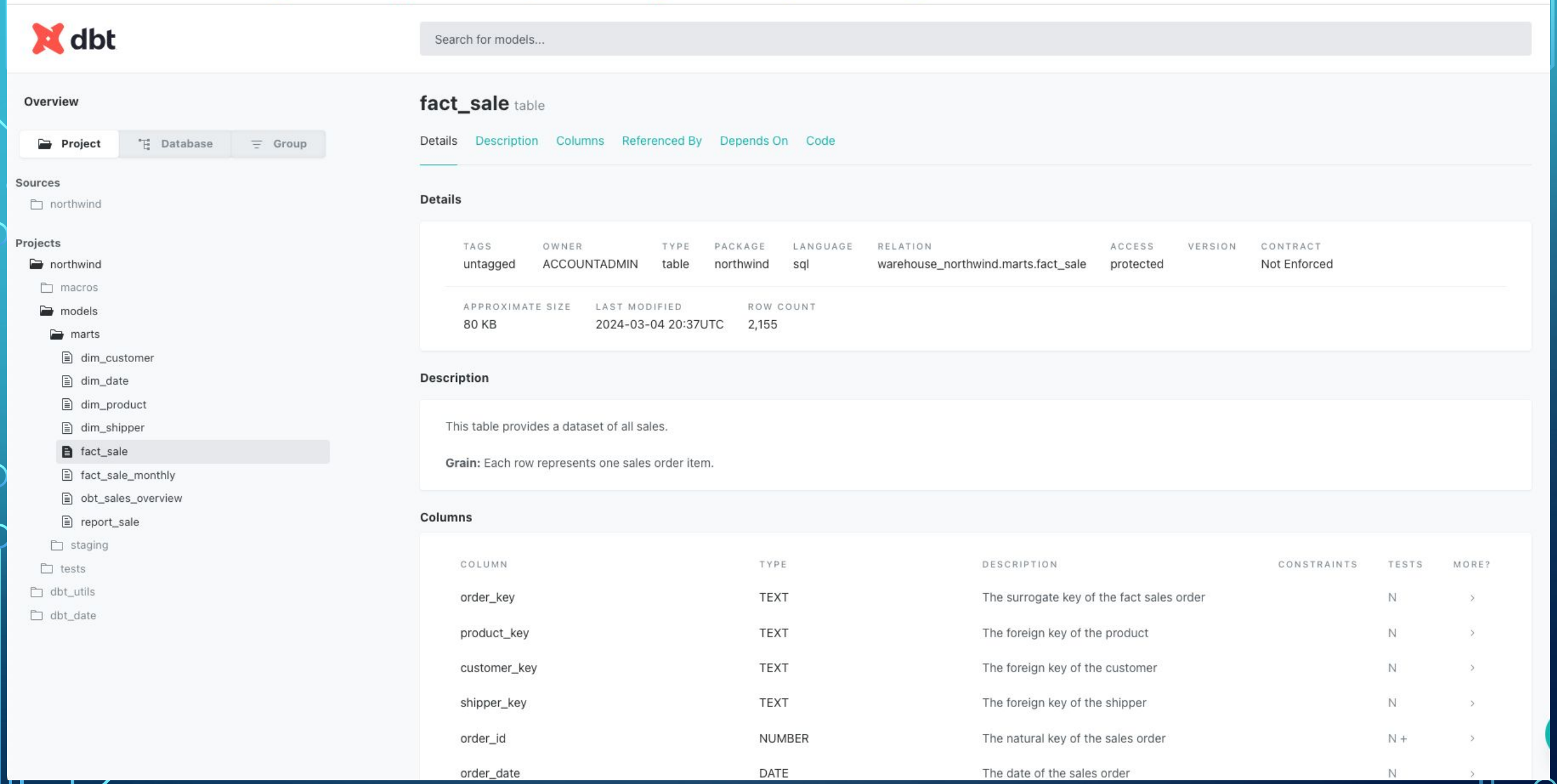
EXTRACT & LOAD: AIRBYTE

- Load SQL file into Postgres, extract & load using Airbyte into Snowflake



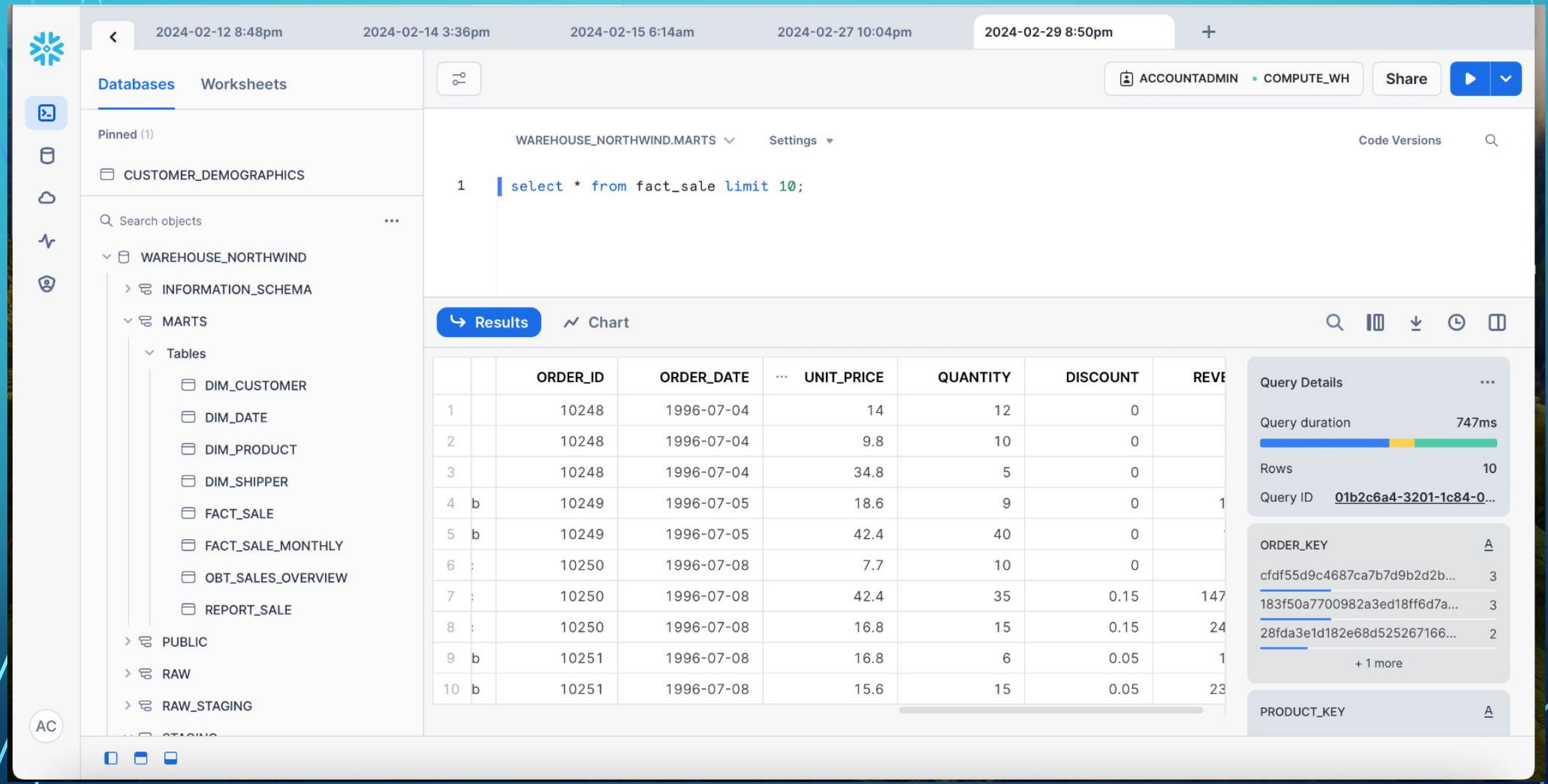
TRANSFORMATIONS: DBT TO SNOWFLAKE

- Aggregate functions
  - “SUM” to get total revenue and freight costs
  - “COUNT” to get total number of employees and orders
- Table for monthly summaries of sales using “SUM” and “GROUP BY”
- Renaming columns using “AS”
- Calculations like unit price * quantity to find revenue
- Table joins to create dimensions and facts in data model
- Window function “PARTITION BY” to calculate all orders shipped by each shipper and total orders made by each customer
- Casting string values to date types

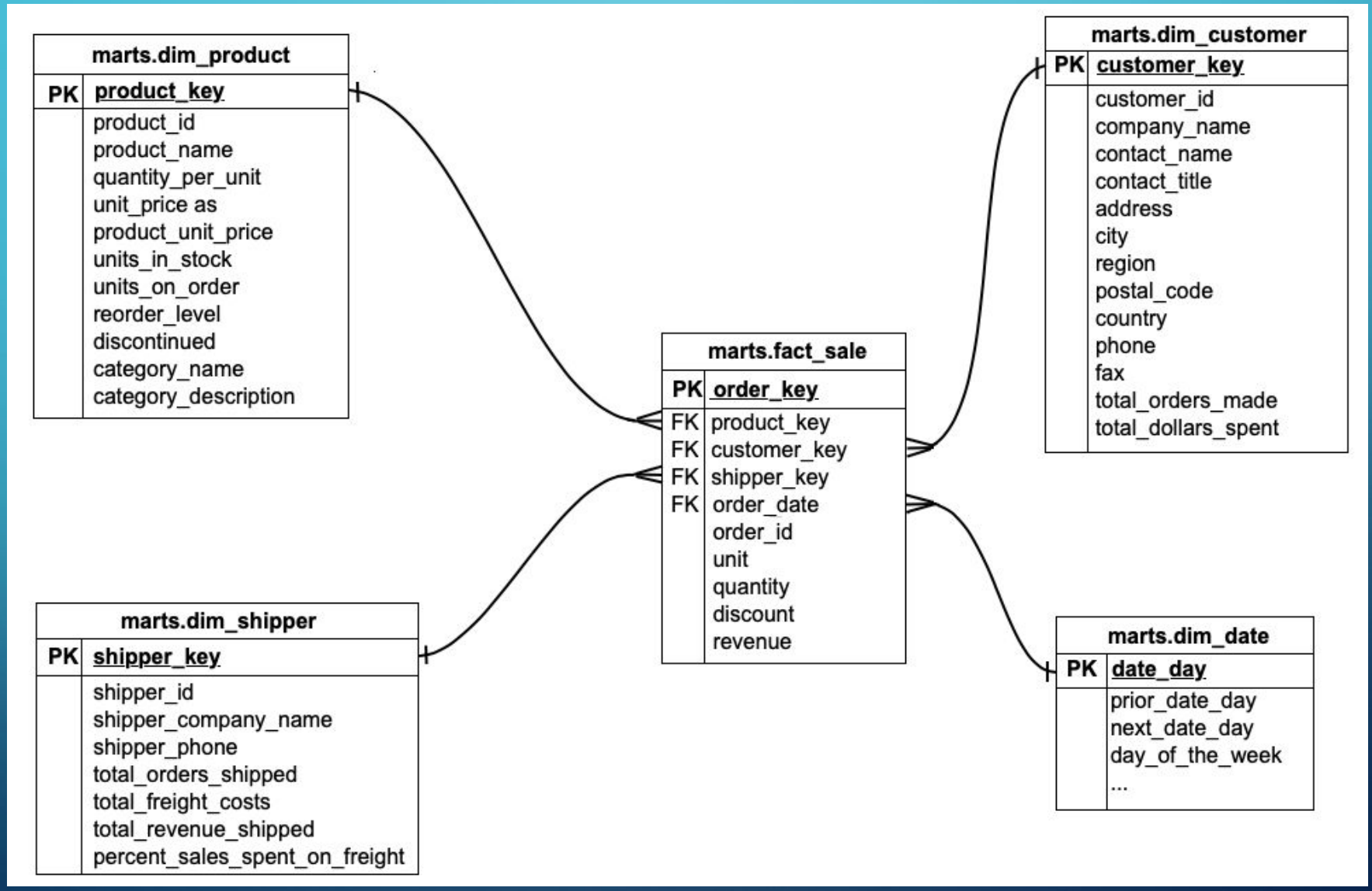

dbt
Search for models...
Overview
Project
Database
Group
Sources
northwind
Projects
northwind
macros
models
marts
dim_customer
dim_date
dim_product
dim_shipper
fact_sale
fact_sale_monthly
obt_sales_overview
report_sale
staging
tests
dbt_utils
dbt_date
fact_sale table
Details
Description
Columns
Referenced By
Depends On
Code
Details
TAGS
untagged
OWNER
ACCOUNTADMIN
TYPE
table
PACKAGE
northwind
LANGUAGE
sql
RELATION
warehouse_northwind.marts.fact_sale
ACCESS
protected
VERSION
CONTRACT
Not Enforced
APPROXIMATE SIZE
80 KB
LAST MODIFIED
2024-03-04 20:37UTC
ROW COUNT
2,155
Description
This table provides a dataset of all sales.
Grain: Each row represents one sales order item.
Columns
COLUMN
TYPE
DESCRIPTION
CONSTRAINTS
TESTS
MORE?
order_key
TEXT
The surrogate key of the fact sales order
N
product_key
TEXT
The foreign key of the product
N
customer_key
TEXT
The foreign key of the customer
N
shipper_key
TEXT
The foreign key of the shipper
N
order_id
NUMBER
The natural key of the sales order
N +
order_date
DATE
The date of the sales order
N

DATA MODELS

- Fact Tables
  - fact_sale_monthly: Summary of all sale orders rolled up to the month level.
  - fact_sale: All sale orders at the individual transaction level
- Dimension Tables
  - dim_customers: Customer records including their Customer ID and company info
  - dim_date: Provides all dates including previous dates compared to current row
  - dim_products: Provides info about products and their category
  - dim_shippers: Provides summary of shippers and their company info
- One Big Table
  - report_sale: Provides one table that includes all orders by customer, product, shipper, and category.


2024-02-12 8:48pm
2024-02-14 3:36pm
2024-02-15 6:14am
2024-02-27 10:04pm
2024-02-29 8:50pm
Databases
Worksheets
ACCOUNTADMIN
COMPUTE_WH
Share
Pinned (1)
CUSTOMER_DEMOGRAPHICS
Search objects
WAREHOUSE_NORTHWIND
INFORMATION_SCHEMA
MARTS
Tables
DIM_CUSTOMER
DIM_DATE
DIM_PRODUCT
DIM_SHIPPER
FACT_SALE
FACT_SALE_MONTHLY
OBT_SALES_OVERVIEW
REPORT_SALE
PUBLIC
RAW
RAW_STAGING
WAREHOUSE_NORTHWIND.MARTS
Settings
Code Versions
1 select * from fact_sale limit 10;
Results
Chart
ORDER_ID
ORDER_DATE
UNIT_PRICE
QUANTITY
DISCOUNT
REVE
1 10248 1996-07-04 14 12 0
2 10248 1996-07-04 9.8 10 0
3 10248 1996-07-04 34.8 5 0
4 b 10249 1996-07-05 18.6 9 0 1
5 b 10249 1996-07-05 42.4 40 0
6 10250 1996-07-08 7.7 10 0
7 10250 1996-07-08 42.4 35 0.15 147
8 10250 1996-07-08 16.8 15 0.15 24
9 b 10251 1996-07-08 16.8 6 0.05 1
10 b 10251 1996-07-08 15.6 15 0.05 23
Query Details
Query duration 747ms
Rows 10
Query ID 01b2c6a4-3201-1c84-0...
ORDER_KEY
cfdf55d9c4687ca7b7d9b2d2b... 3
183f50a7700982a3ed18ff6d7a... 3
28fda3e1d182e68d525267166... 2
+ 1 more
PRODUCT_KEY
AC

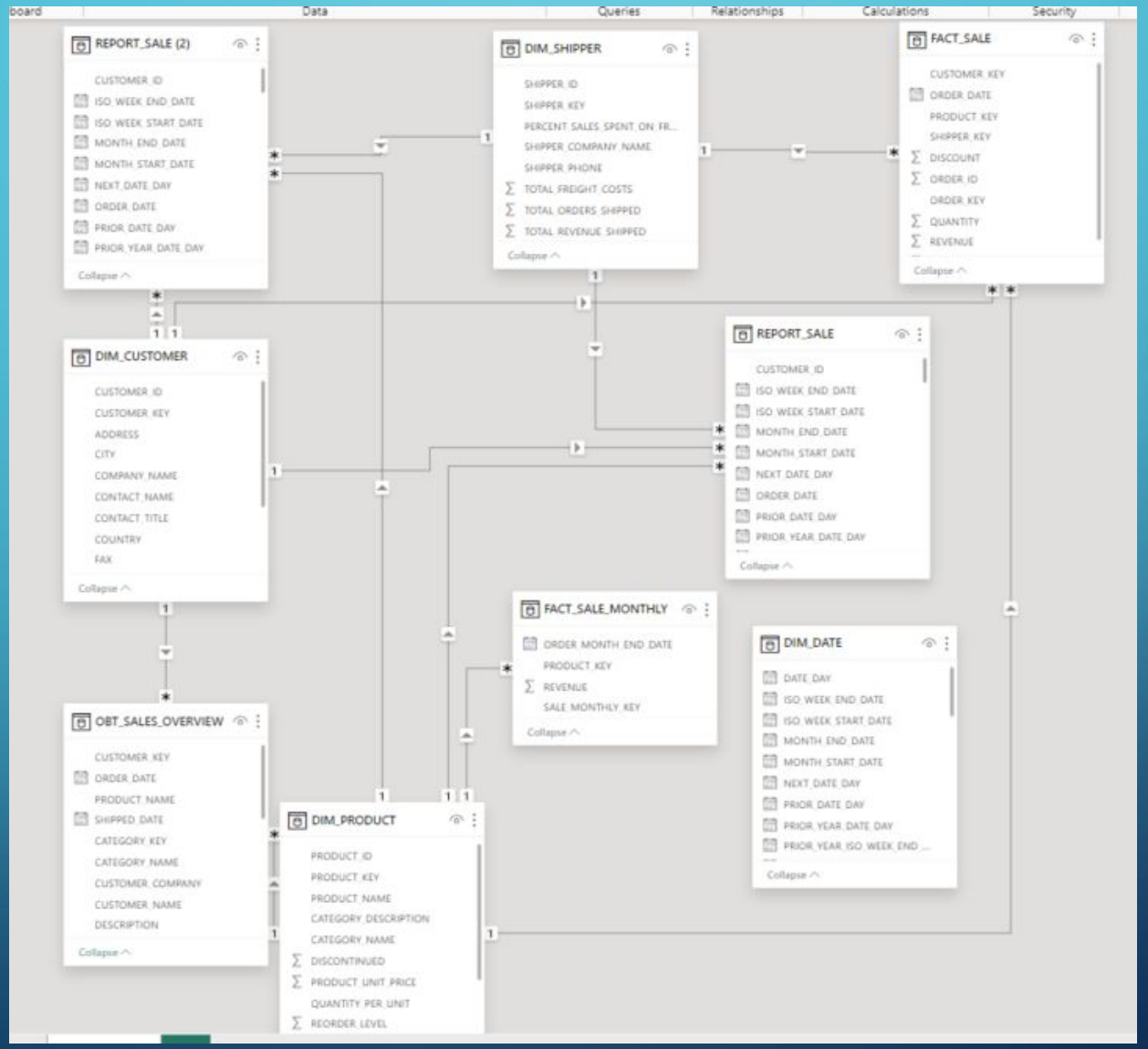
ENTITY RELATIONSHIP DIAGRAM



DATA QUALITY TESTS

- Custom Tests
  - Assert order: Ensure quantity and unit price is greater than 0
  - Assert employee id: Greater than 0 and not greater than the total number of employees
- Generic Tests
  - Column greater than: Ensure price is more than custom value
  - Column less than: Ensure discount is less than custom value
  - Min word count: Address and full name have 2 words minimum
  - Min character count: Ensure order id has a minimum character count of x

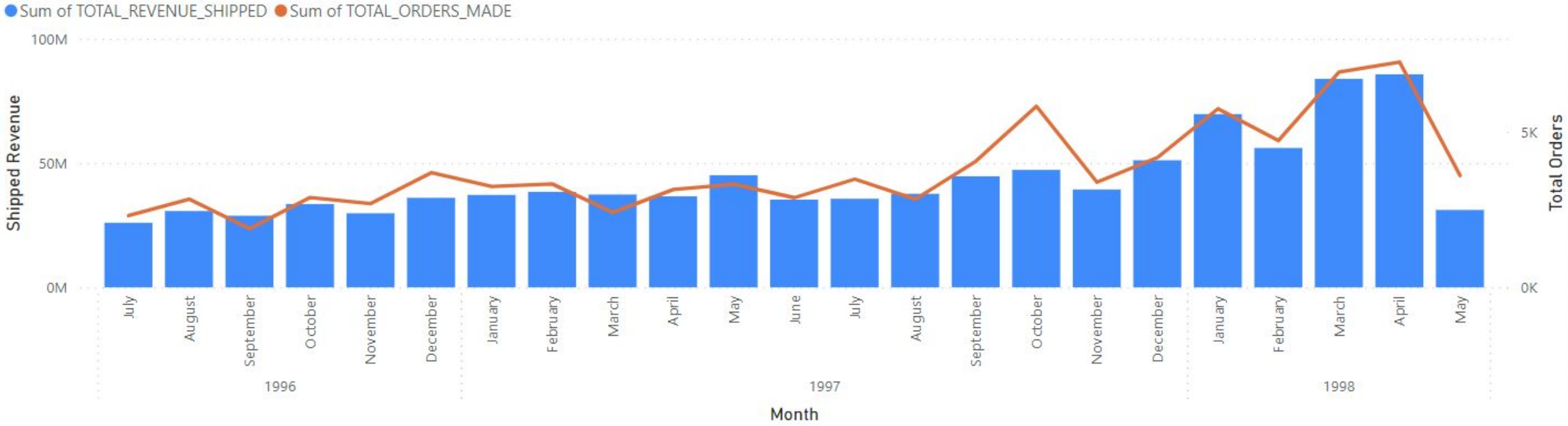
SEMANTIC MODELING WITH POWER BI



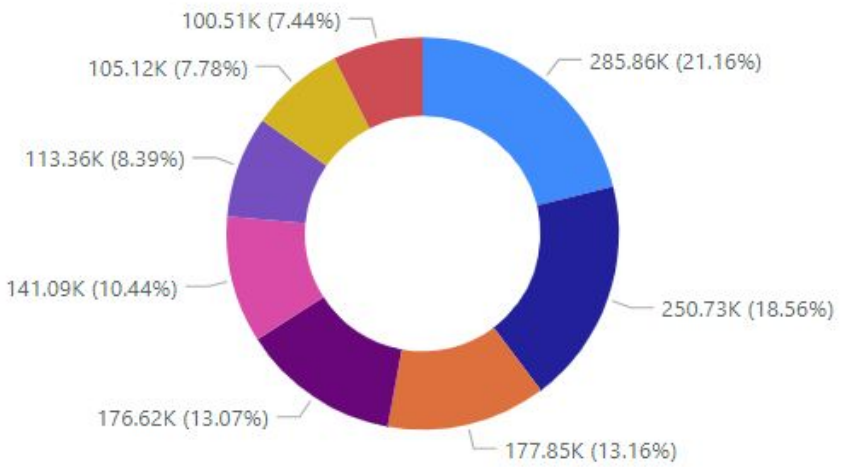
SEMANTIC LAYER METRICS

- Total Orders Made Monthly & Total Orders Shipped Monthly
  - Sum of monthly shipped revenue overlayed with sum of monthly orders made
- Revenue by Product Category
  - Percent of total revenue by product category revenue
- Supplier Orders by Product Category
  - A summary of supplier's orders by product category

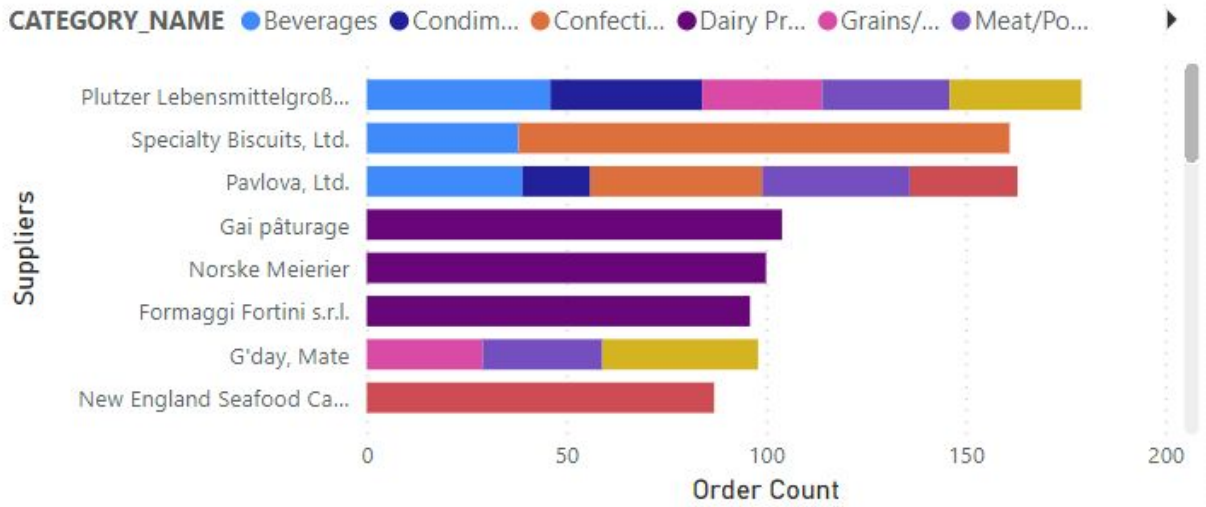
Revenue by Month



Revenue by Category

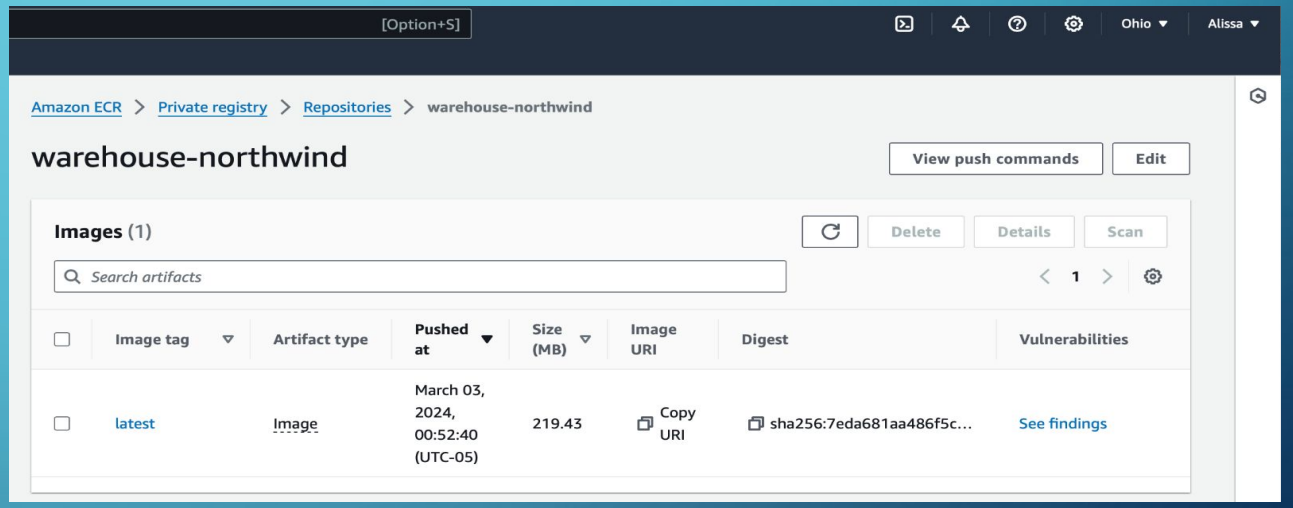


Supplier Orders by Category

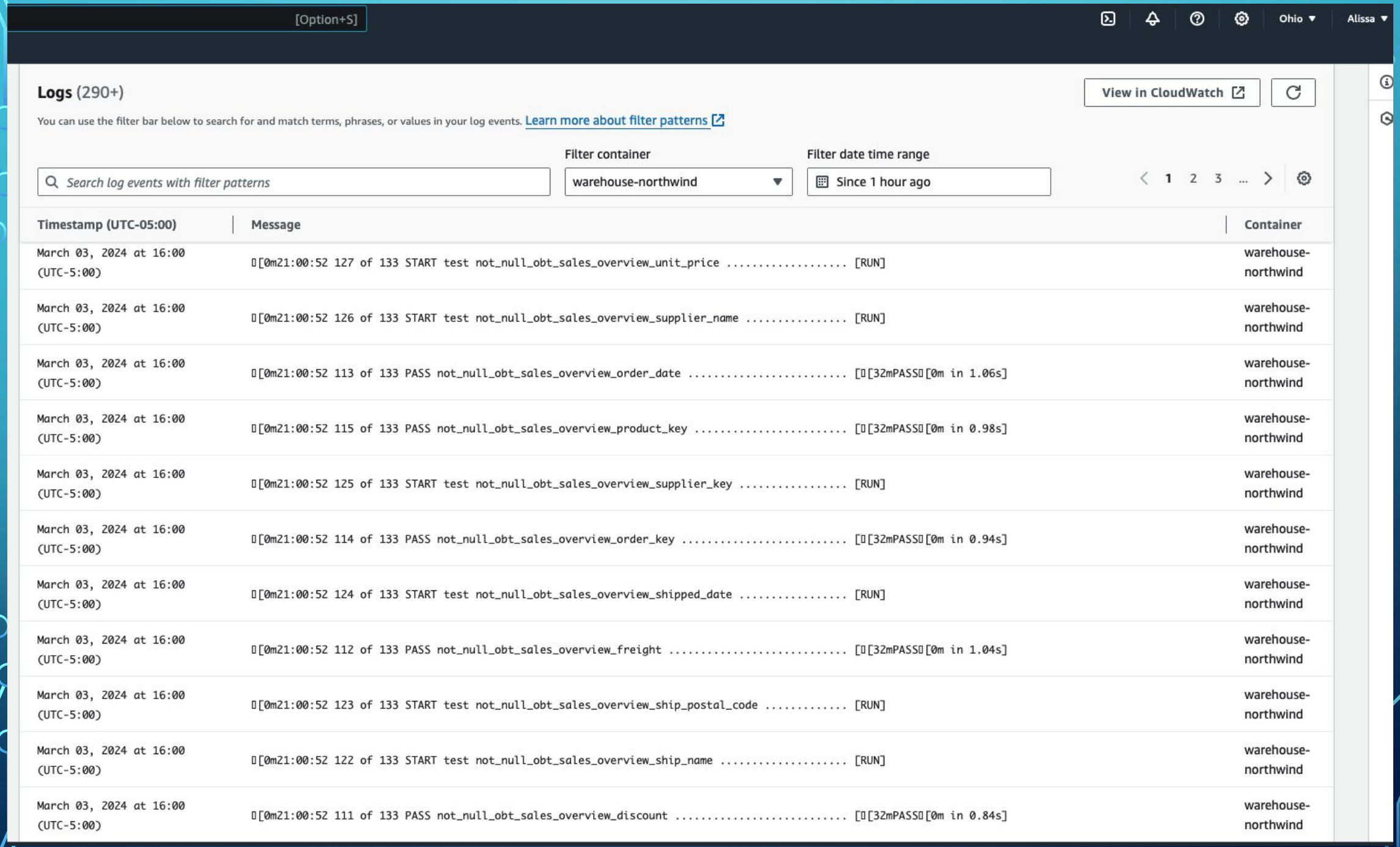


CLOUD DEPLOYMENT AND ORCHESTRATION

- Created and built a docker image locally
- Pushed into ECR (Elastic Container Registry) repository in AWS
- Deploy Docker Image from ECR via ECS (Elastic Container Service) with AWS



- Create new ECS Fargate Cluster
- Create new Task definition with snowflake environment variables
- Create new task in cluster and schedule task accordingly


[Option+S]
Ohio
Alissa
Logs (290+)
View in CloudWatch
You can use the filter bar below to search for and match terms, phrases, or values in your log events. Learn more about filter patterns
Filter container
Filter date time range
Search log events with filter patterns
warehouse-northwind
Since 1 hour ago
1 2 3 ...
Timestamp (UTC-05:00)
Message
Container
March 03, 2024 at 16:00 (UTC-5:00)
[0m21:00:52 127 of 133 START test not_null_obt_sales_overview_unit_price ..... [RUN]
warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)
[0m21:00:52 126 of 133 START test not_null_obt_sales_overview_supplier_name ..... [RUN]
warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)
[0m21:00:52 113 of 133 PASS not_null_obt_sales_overview_order_date ..... [[32mPASS[0m in 1.06s]
warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)
[0m21:00:52 115 of 133 PASS not_null_obt_sales_overview_product_key ..... [[32mPASS[0m in 0.98s]
warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)
[0m21:00:52 125 of 133 START test not_null_obt_sales_overview_supplier_key ..... [RUN]
warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)
[0m21:00:52 114 of 133 PASS not_null_obt_sales_overview_order_key ..... [[32mPASS[0m in 0.94s]
warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)
[0m21:00:52 124 of 133 START test not_null_obt_sales_overview_shipped_date ..... [RUN]
warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)
[0m21:00:52 112 of 133 PASS not_null_obt_sales_overview_freight ..... [[32mPASS[0m in 1.04s]
warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)
[0m21:00:52 123 of 133 START test not_null_obt_sales_overview_ship_postal_code ..... [RUN]
warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)
[0m21:00:52 122 of 133 START test not_null_obt_sales_overview_ship_name ..... [RUN]
warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)
[0m21:00:52 111 of 133 PASS not_null_obt_sales_overview_discount ..... [[32mPASS[0m in 0.84s]
warehouse-northwind

762c5b9d02d44cc69d378afbfcee5ac1

Stop

Task stopped at: 2024-03-03T21:01:18.445Z
Essential container in task exited

Configuration | Logs | Networking | Volumes (0) | Tags

Logs (290+)

View in CloudWatch

You can use the filter bar below to search for and match terms, phrases, or values in your log events. Learn more about filter patterns

Search log events with filter patterns

Filter container: warehouse-northwind

Filter date time range: Since 1 hour ago

1 2 3 ...

Timestamp (UTC-05:00)	Message	Container
March 03, 2024 at 16:00 (UTC-5:00)	[0m21:00:55	warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)	[0m21:00:55 Done. PASS=133 WARN=0 ERROR=0 SKIP=0 TOTAL=133	warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)	[0m21:00:55	warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)	[0m21:00:55 [32mCompleted successfully[0m	warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)	[0m21:00:55	warehouse-northwind
March 03, 2024 at 16:00 (UTC-5:00)	[0m21:00:55 Finished running 22 table models, 111 tests in 0 hours 0 minutes and 15.87 seconds (15.87s).	warehouse-northwind
March 03, 2024 at 16:00		warehouse-

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences