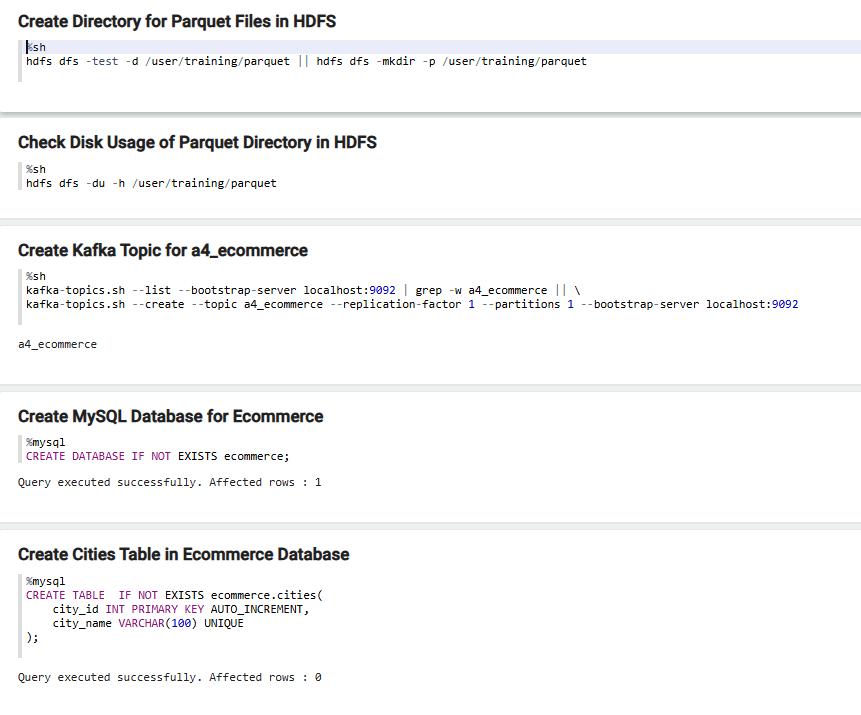
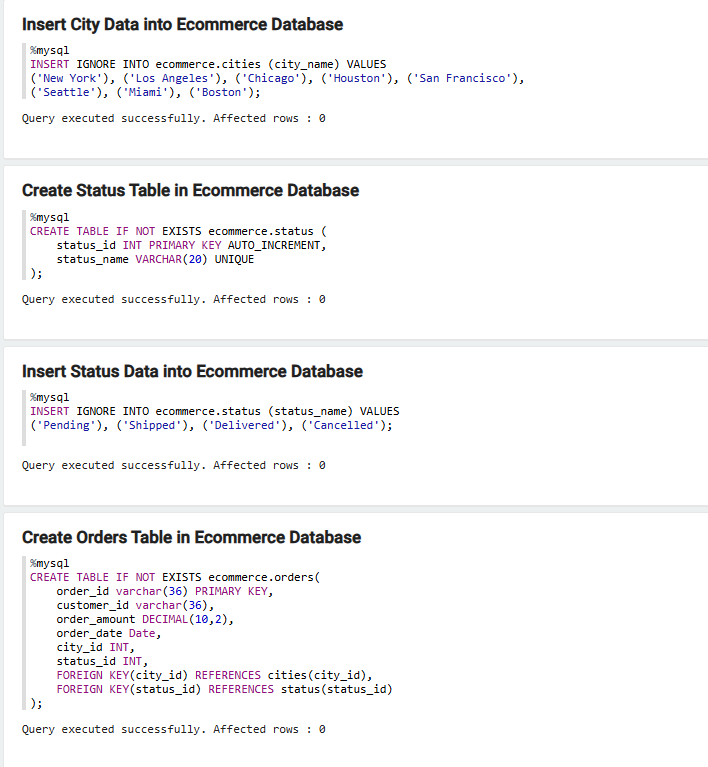
**Team Assignment CHANGE DATA CAPTURE AND REAL-TIME ANALYSIS**

#### **David Liu (Task 1 & 2) ; Vilok Guliani (Task 3 & 4)**

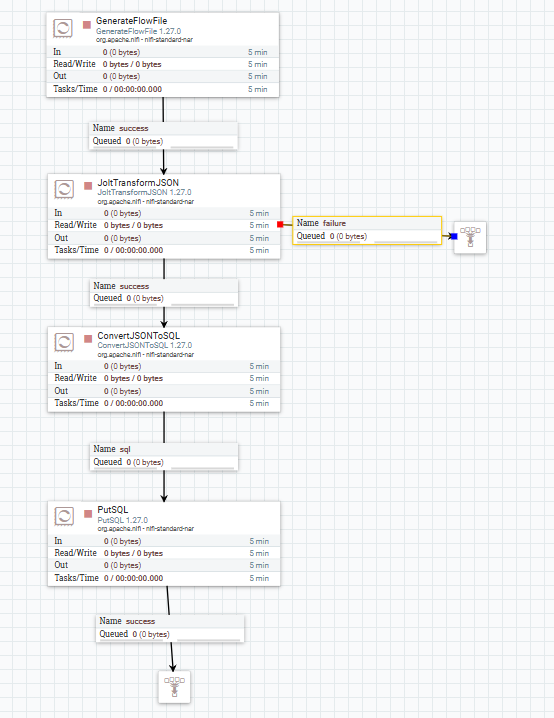
#### **Task 1 Fake Data Generation (NiFi → MySQL) (7 pts)**

**Prepping the Mysql database for NIFI data generation**

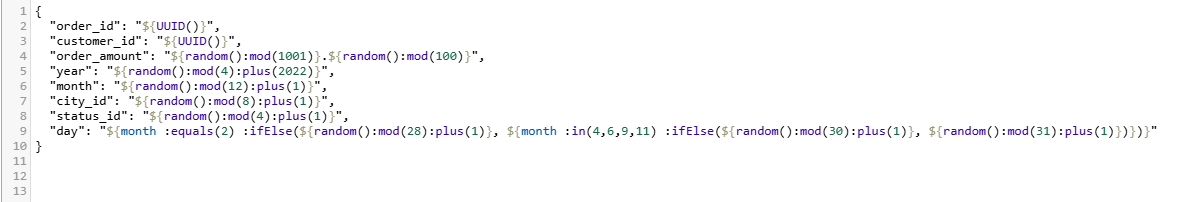




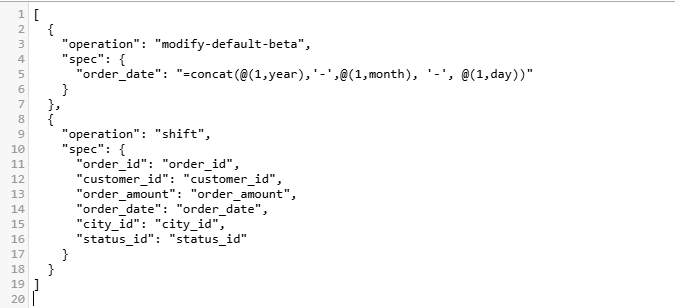
**NIFI Data Generation**



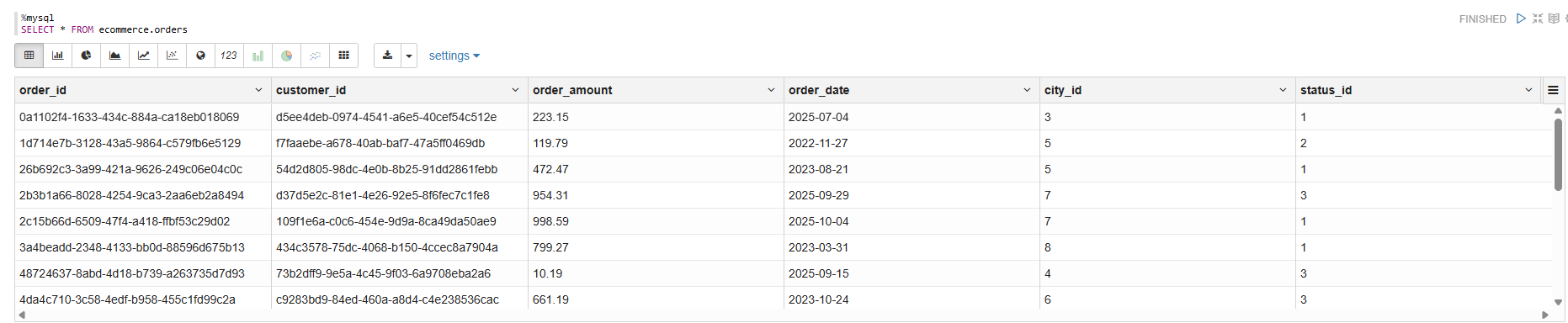
**Data Generation NIFI code :**



**JOLT Transform code:**



#### **Result in Mysql database**



#### 

#### 

#### 

#### 

#### 

#### 

#### 

#### 

#### 

#### 

#### 

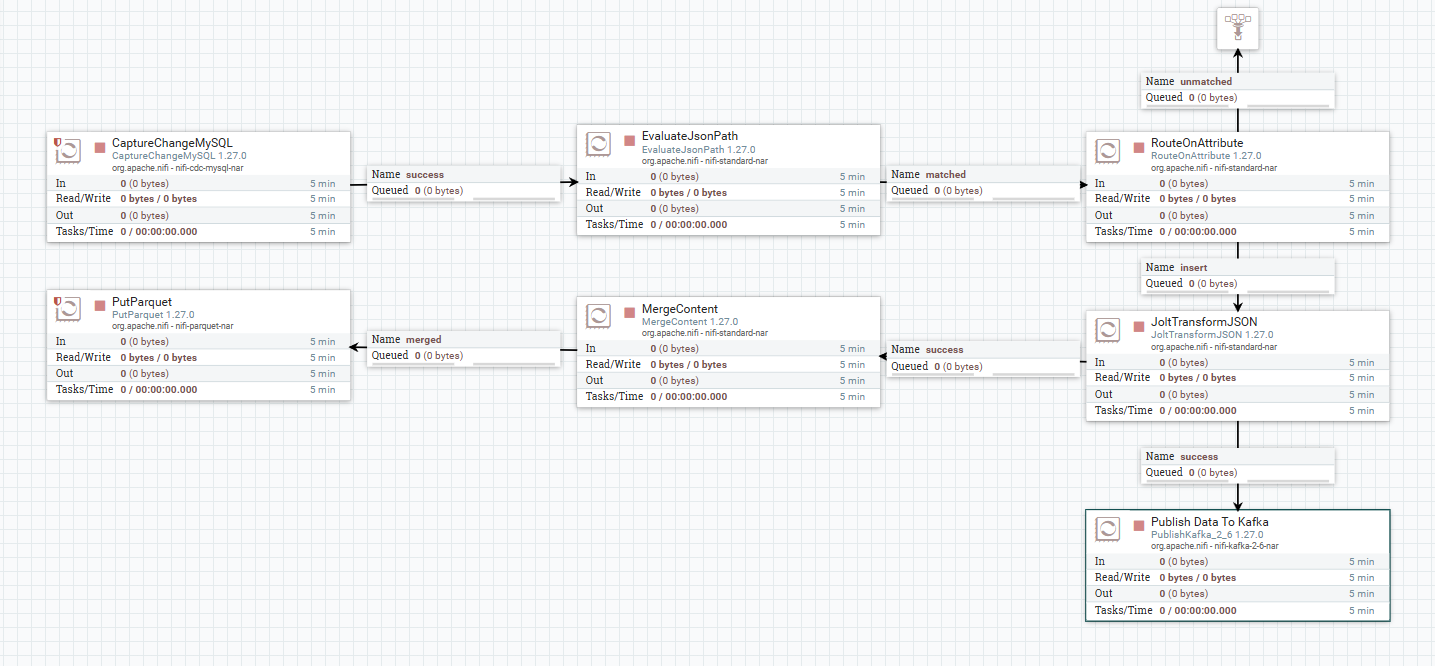
#### 

#### 

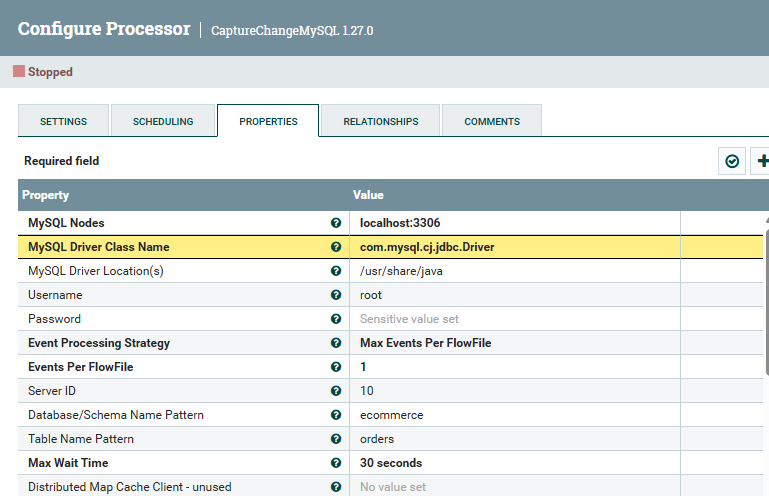
#### 

#### 

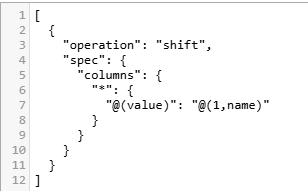
#### **Task 2 Data Access with Hive Change Data Capture (MySQL → Kafka via NiFi) (6 pts)**



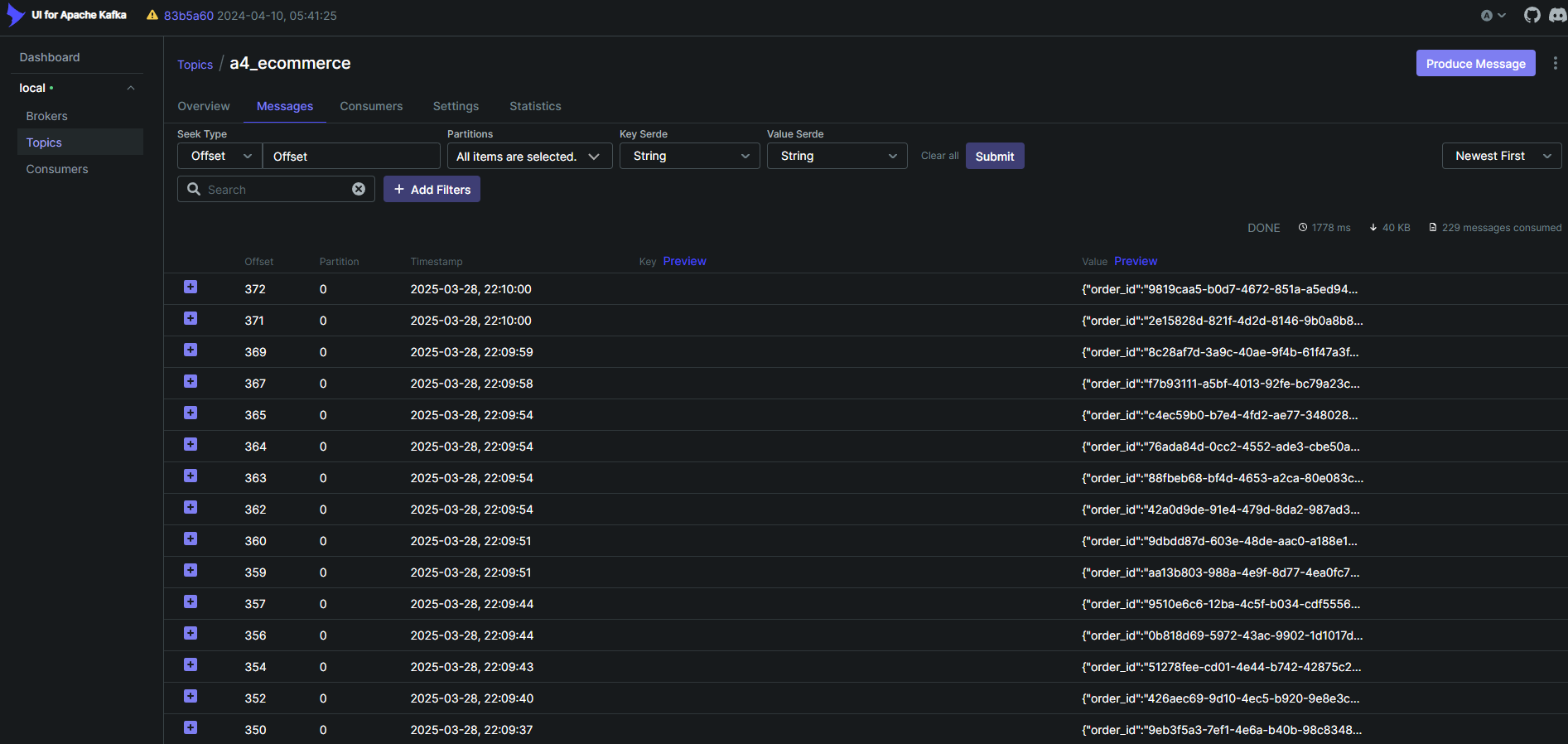
**CDC Setup**

****

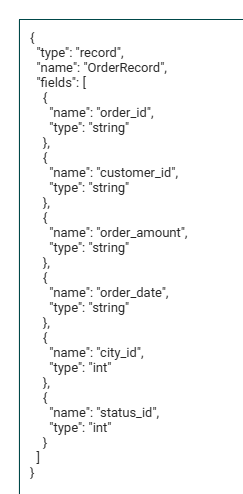
**JOLT Transform Code:**



**Publishing to Kafka**



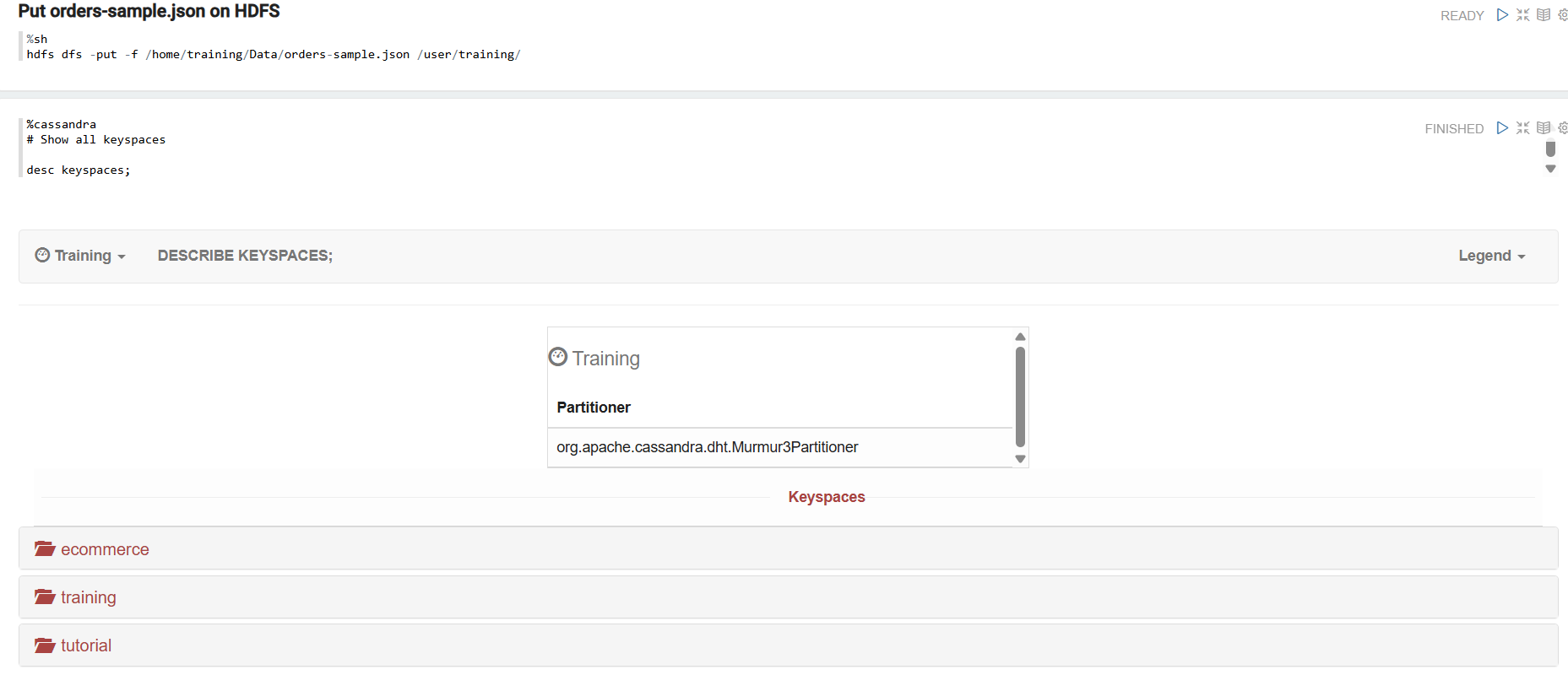
**Schema Required for writing into Avro -> Parquet**

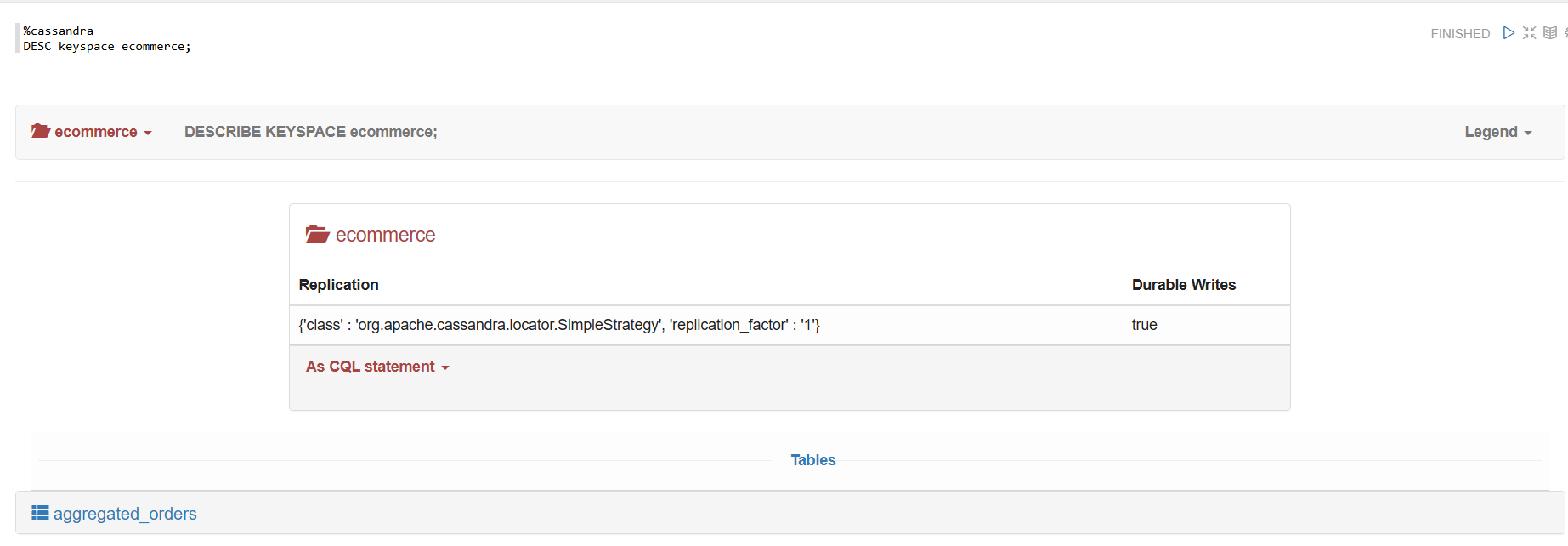
****

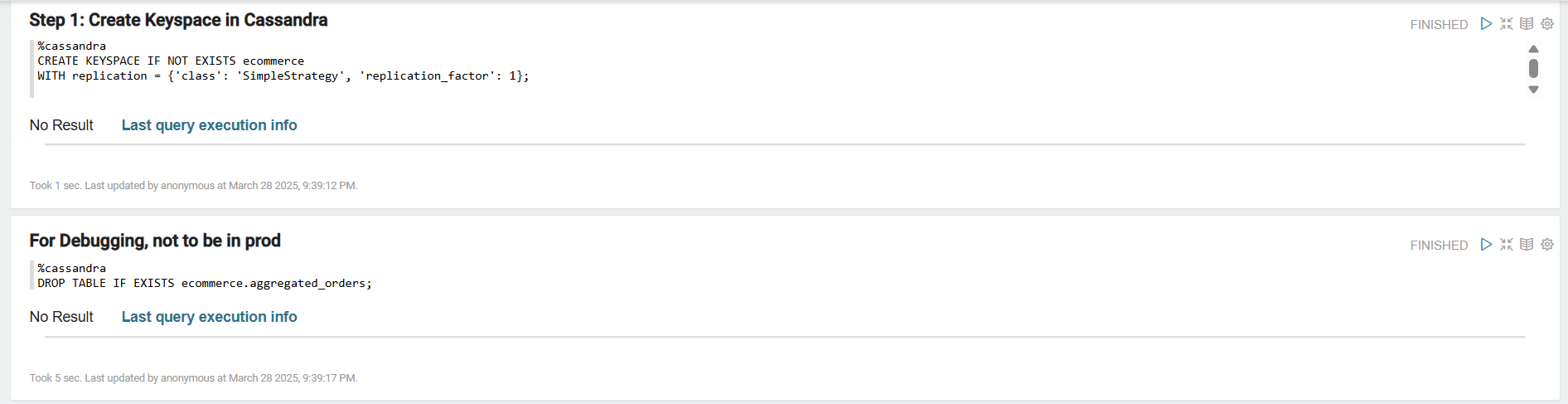
#### 

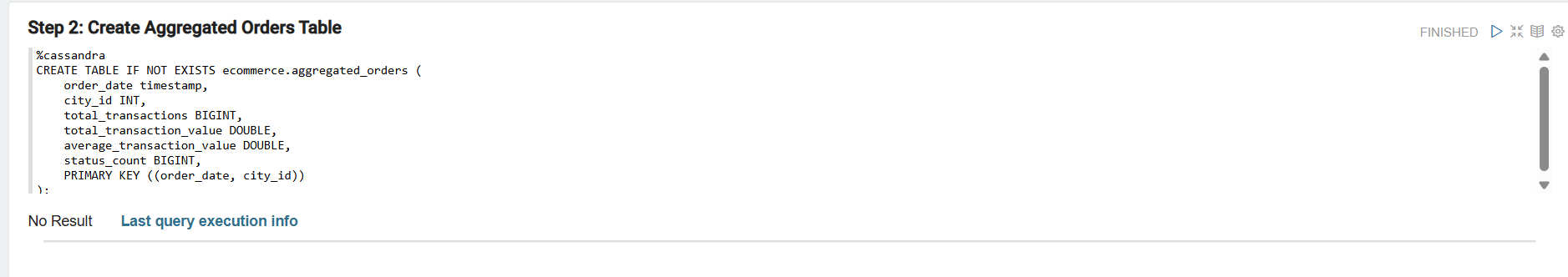
#### 

#### **Task 3 Real-Time Data Processing (Kafka → Spark Streaming → Cassandra) (9 points)**

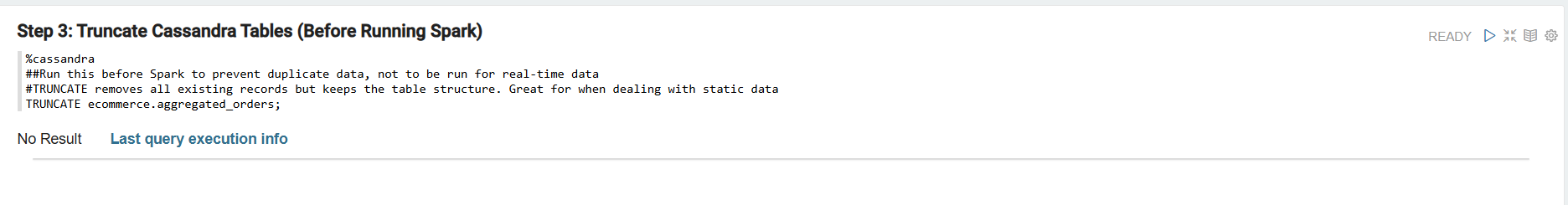






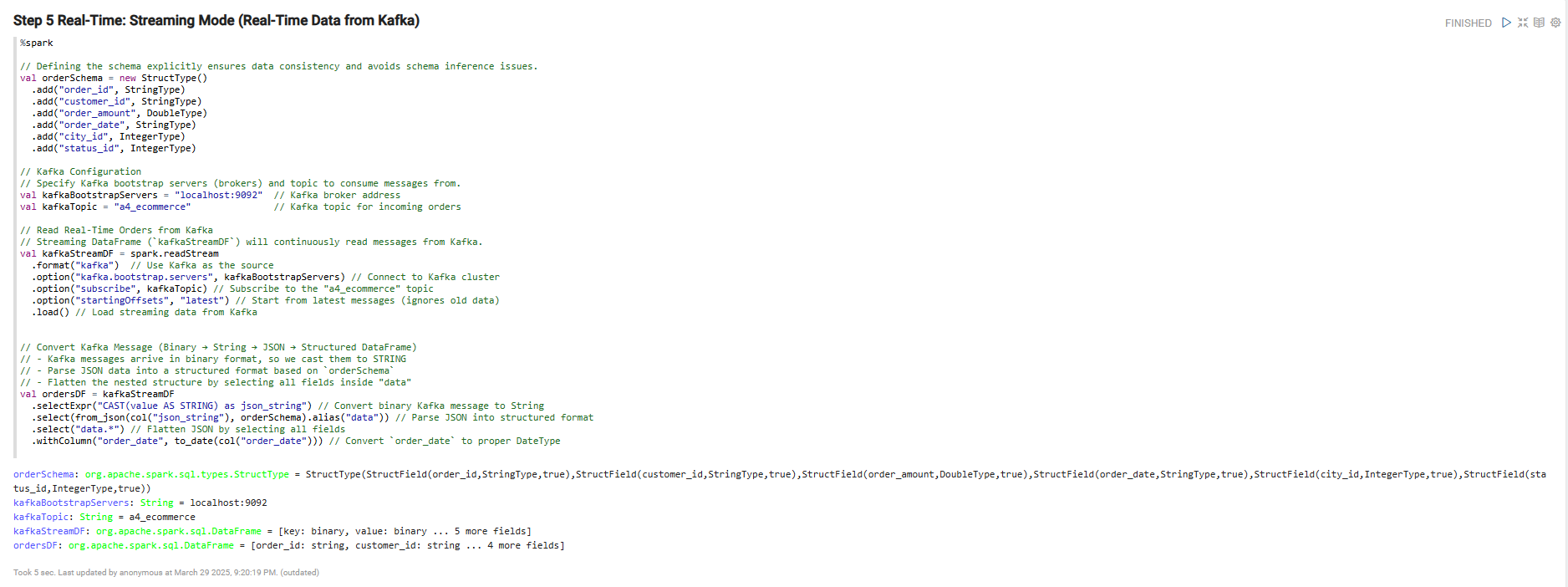


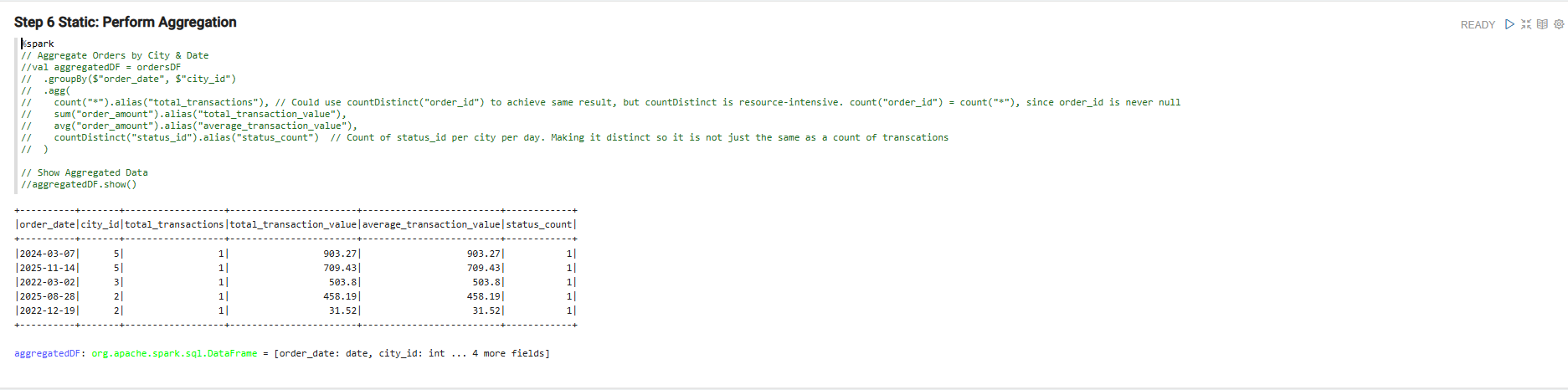
#### 

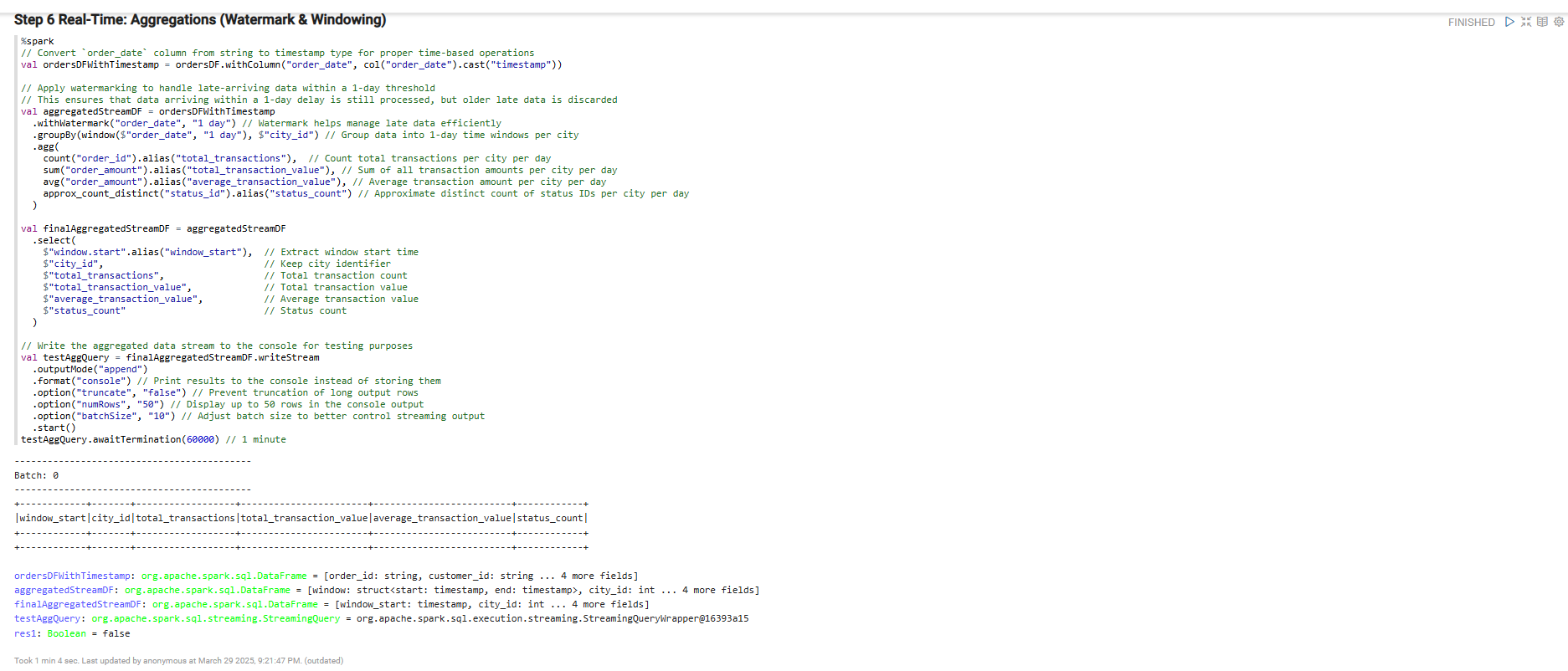




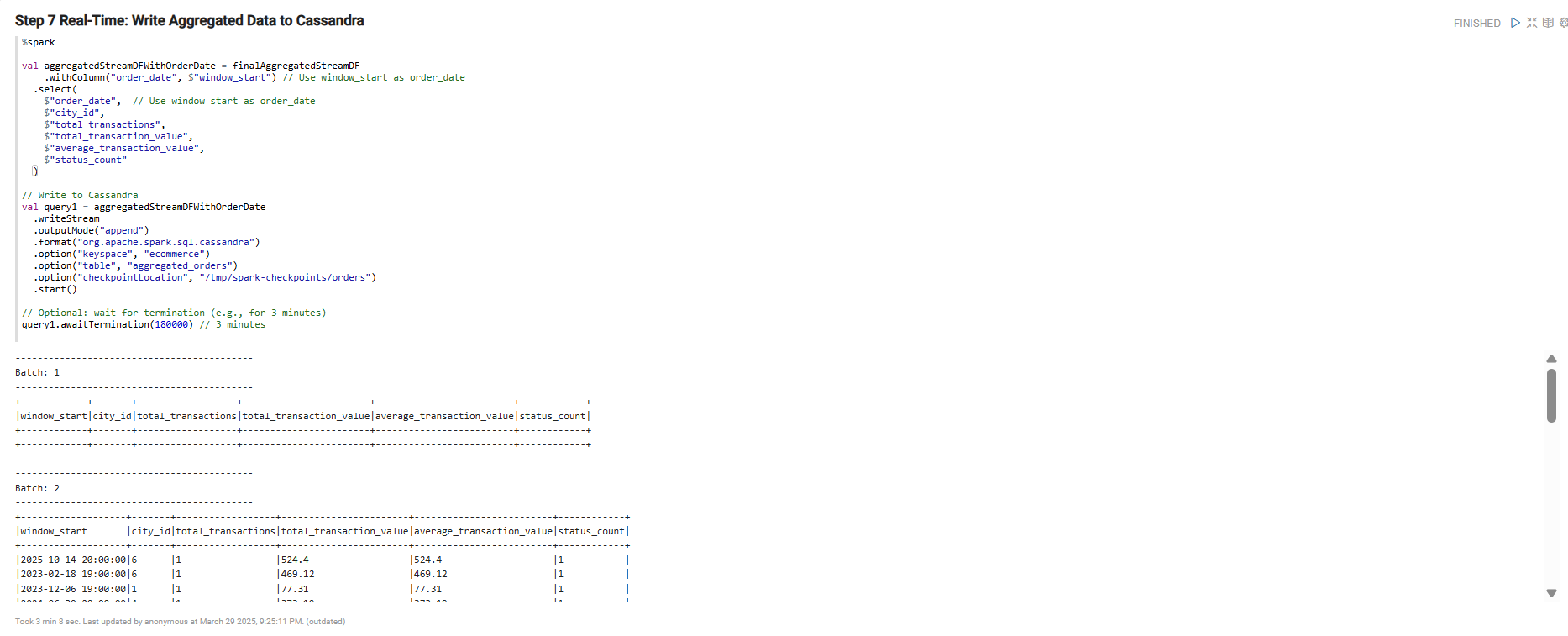








#### 

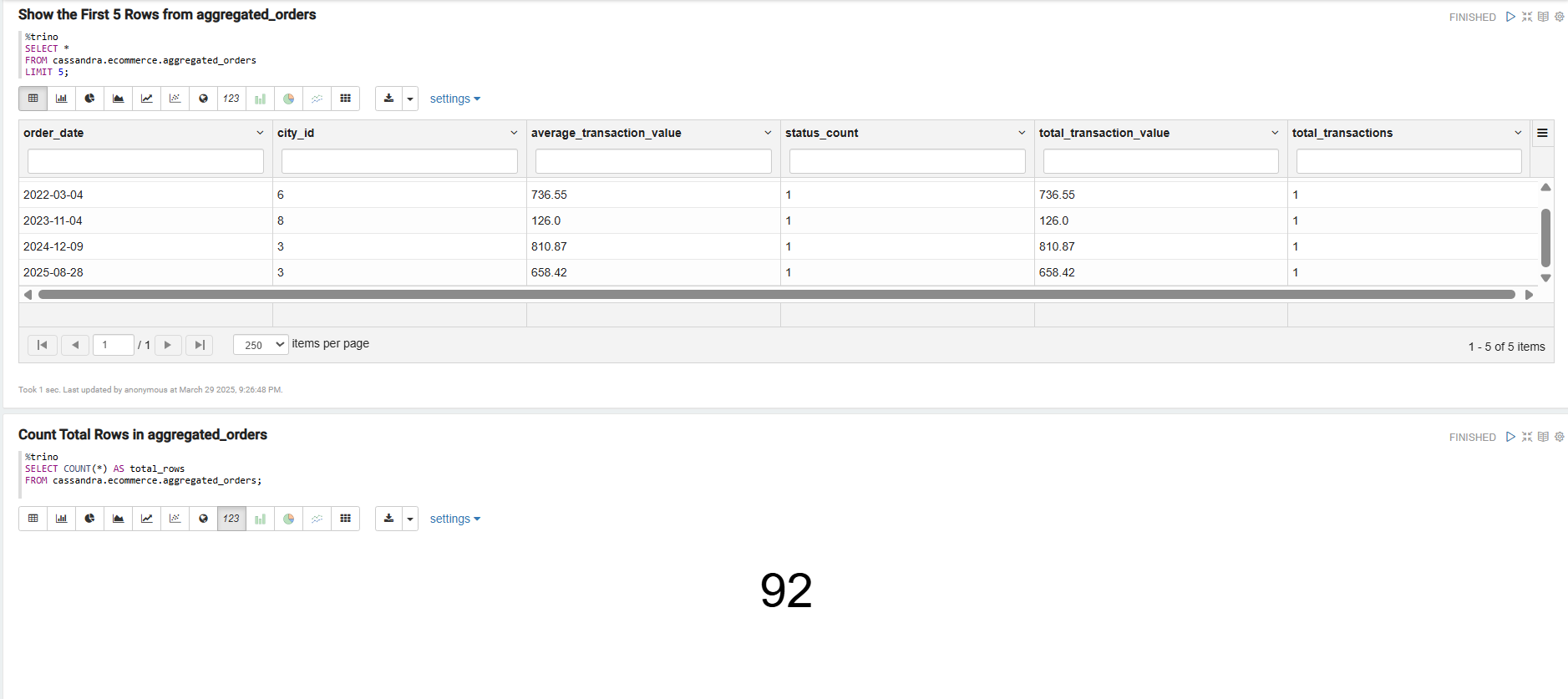


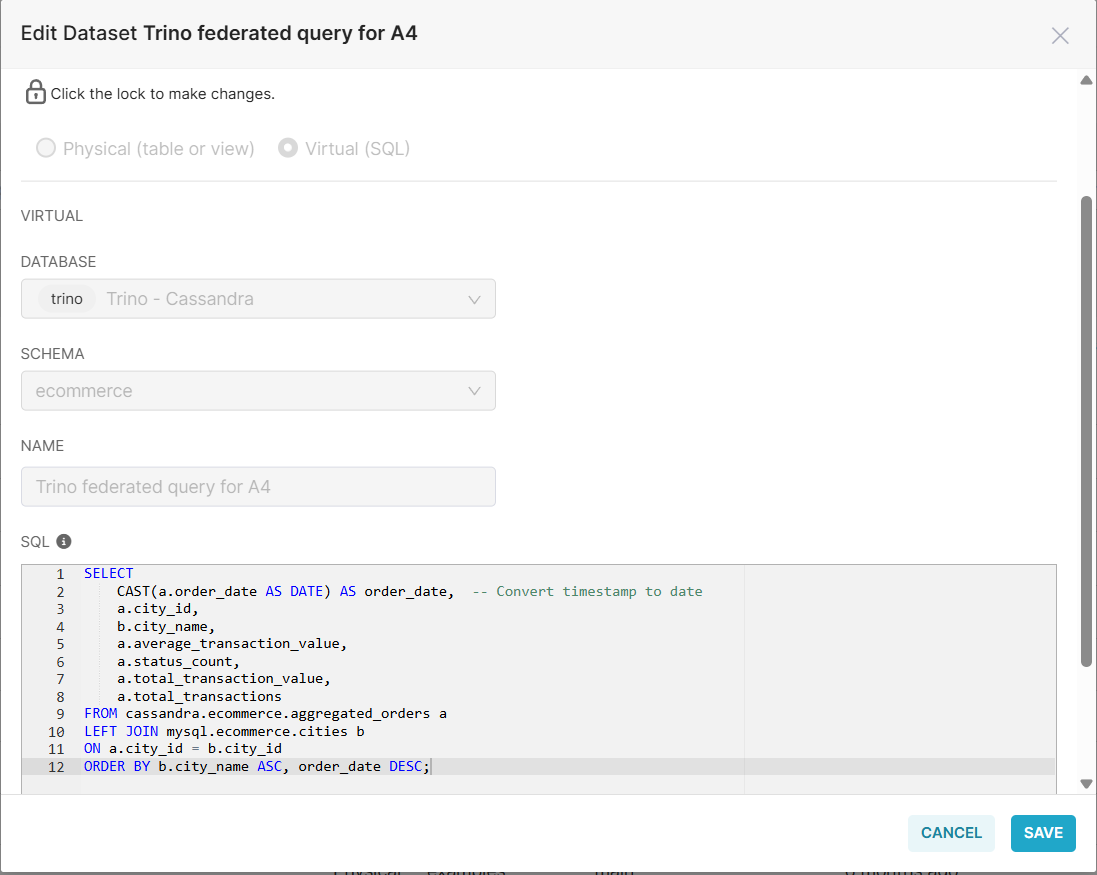
#### **Task 4 Data Analysis & Visualization (Cassandra → Trino → Superset) (8 points)**

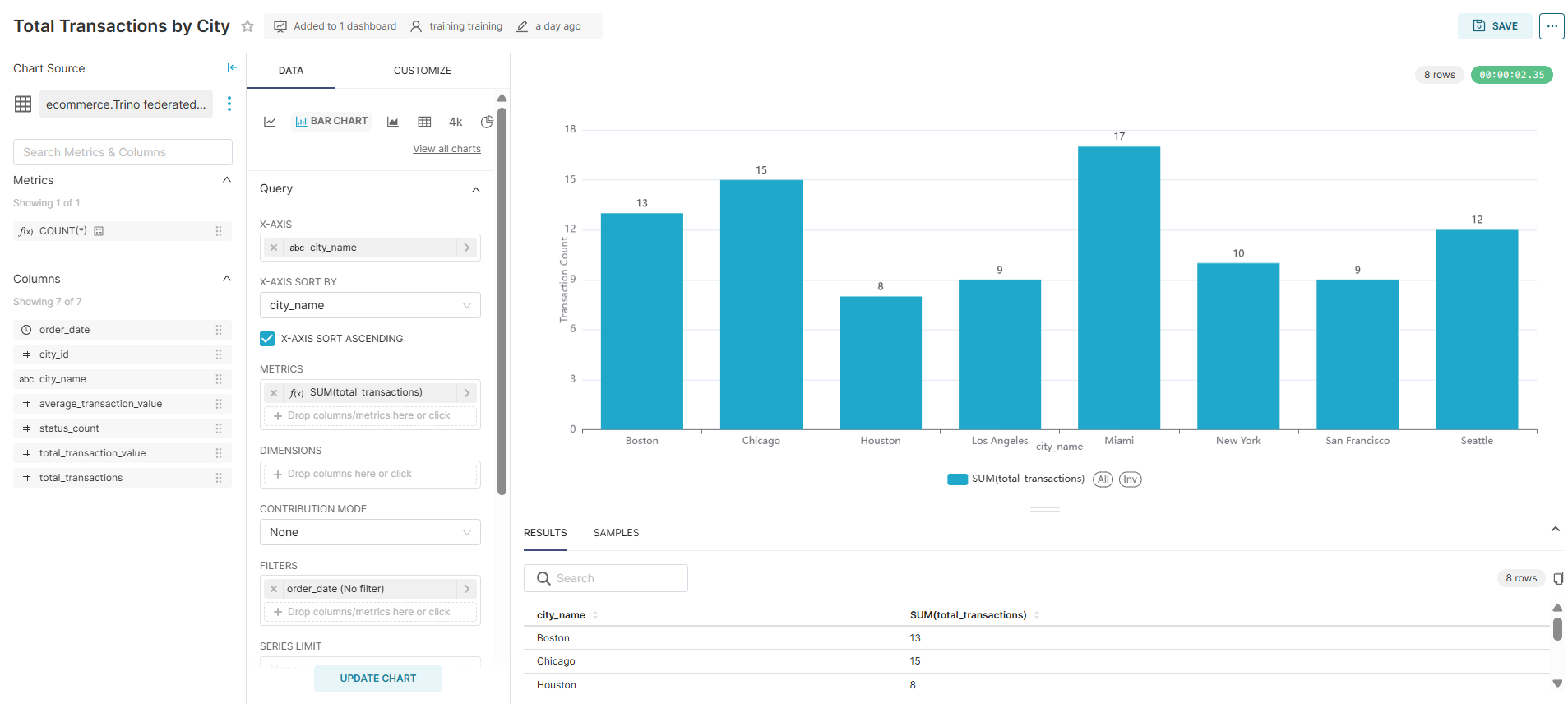
**Trino**

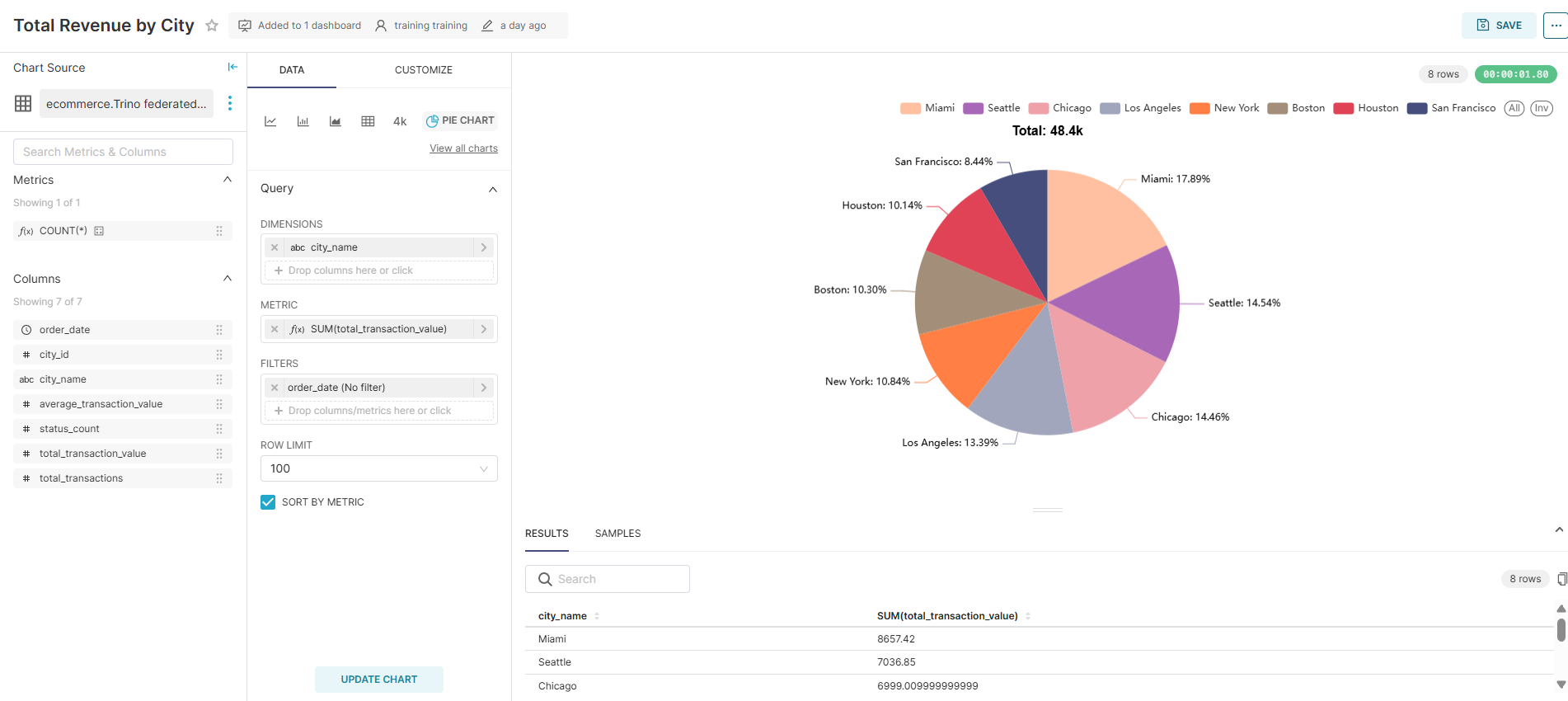
Use Trino to query the aggregated data in Cassandra.

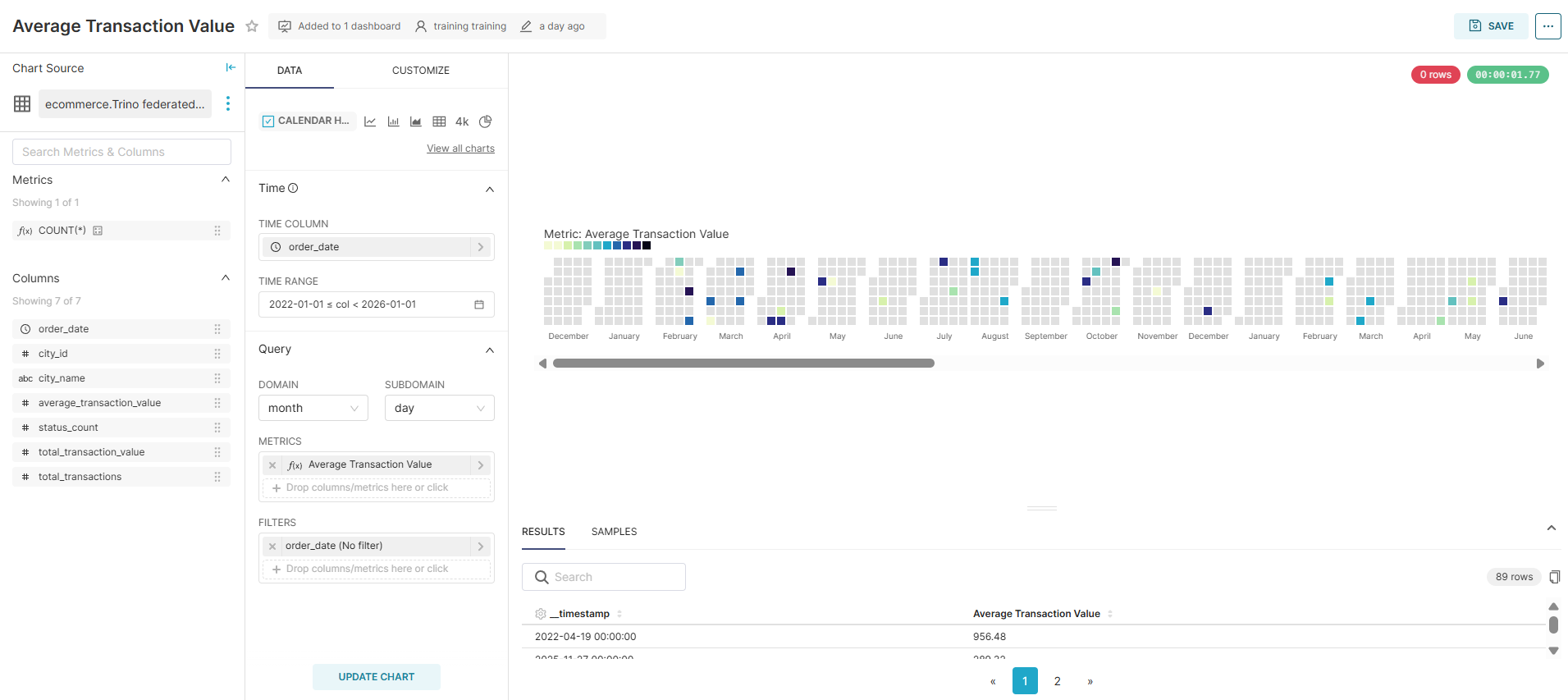
* Show the first 5 rows from the table.
* Show a row count of the current data in the table.

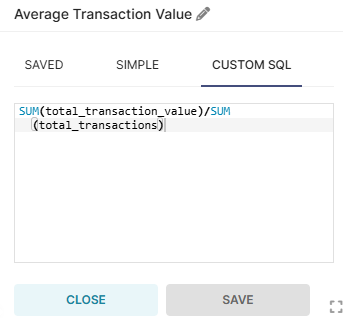
****

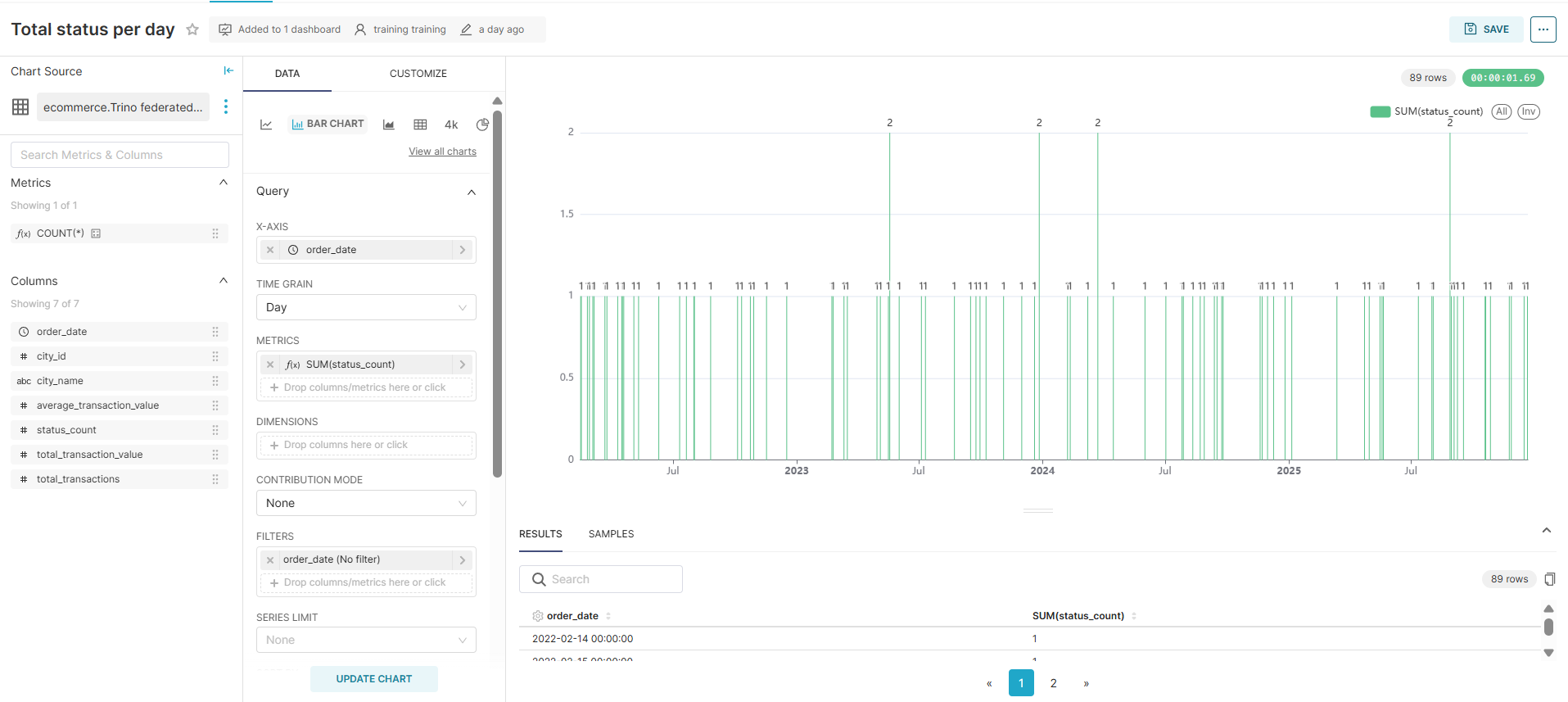


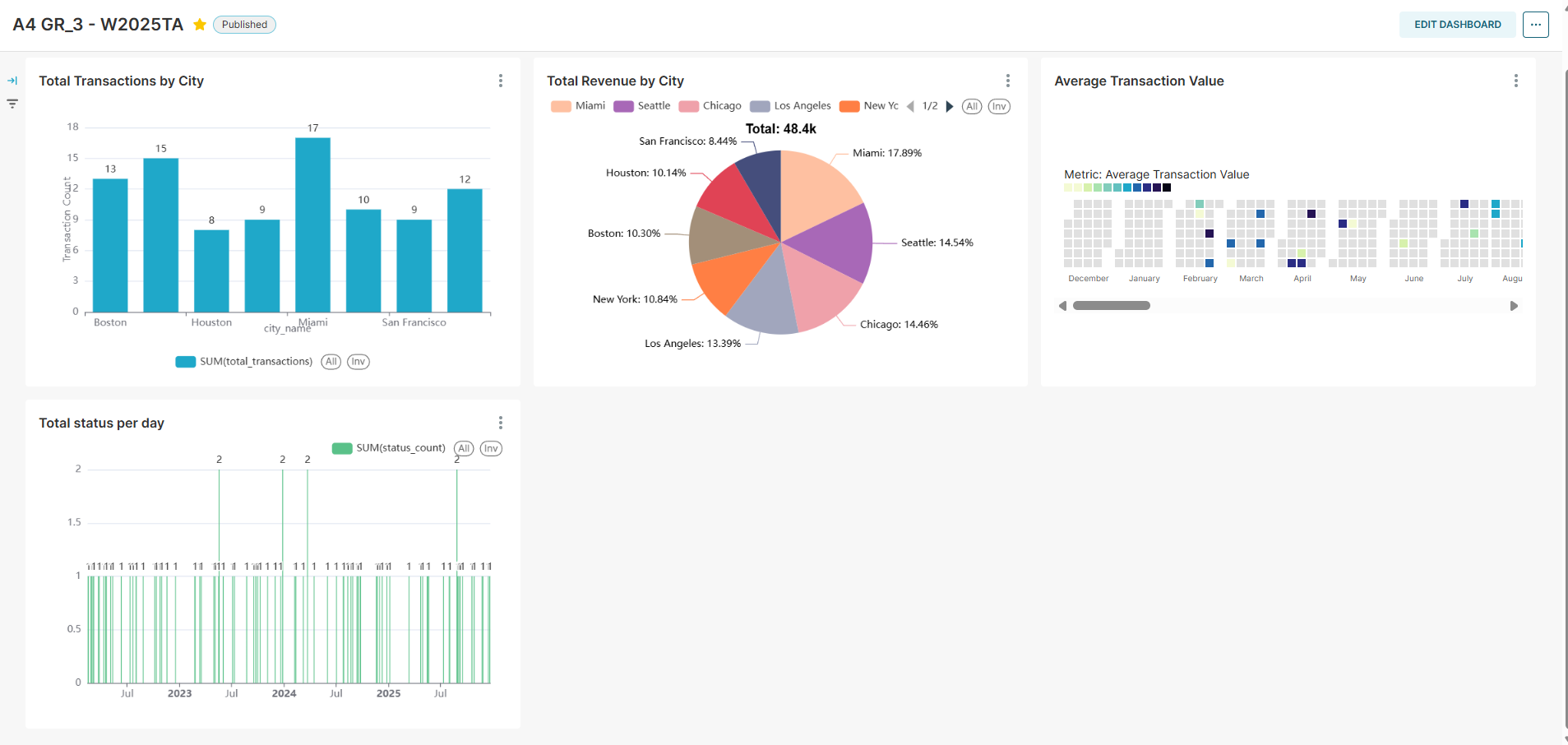
****

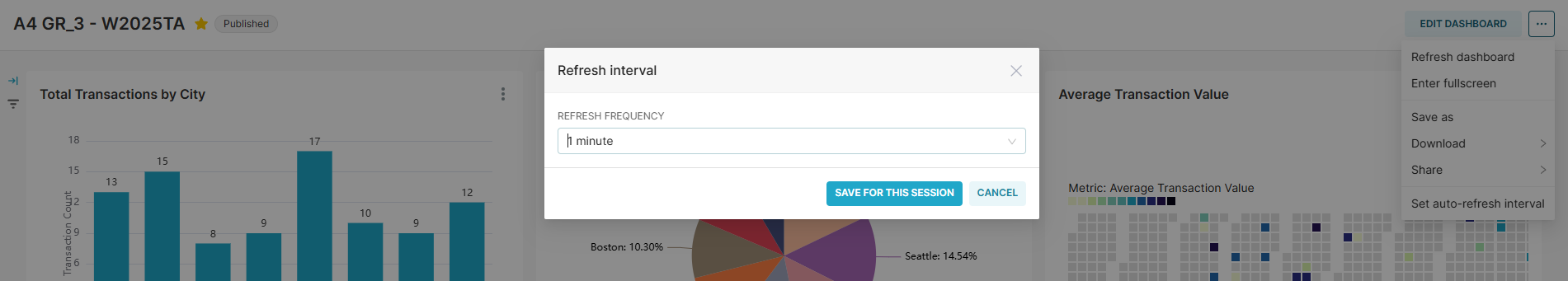
****

****

****

****

****

****