InstaCart Data Warehouse

Sarthak Killedar
Rupali Shyam
Dilisha Naidu
Sayantani Bhattacharjee
Lakshmi Ravichandran
Sanzil Madye

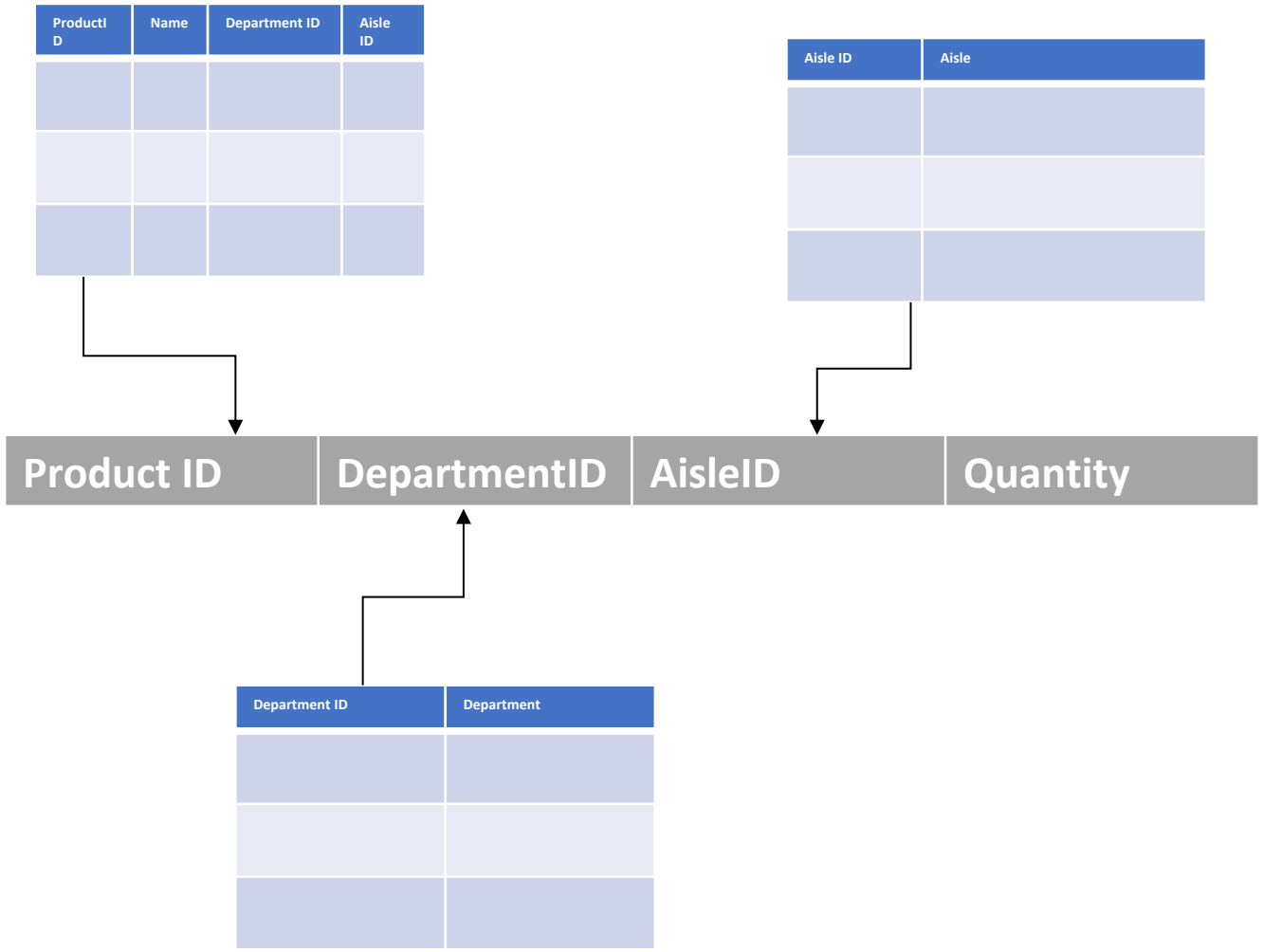
We make use of the instaCart dataset to perform certain analysis on the data. This can help gain useful insight on the working and performance of a business.

As required by the project, the analysis is done using three different databases, relational DB, MongoDB and Neo4j.

Since the data is available in terms of tables/relations, the relational database would be an ideal choice.

We make use of 5 dimensional table such as Orders, Products, Aisle, Departments and Order Details to create the Fact Tables.

Product Fact Table




```
CREATE TABLE [dbo].[ProductStarFact](
    [ProductId] [int] NOT NULL,
    [DepartmentId] [int] NOT NULL,
    [AisleId] [int] NOT NULL,
    [Quantity] [int] NOT NULL,
CONSTRAINT [PK_ProductStarFact] PRIMARY KEY
CLUSTERED ([ProductId] ASC)
```

MongoDb Structure

Products Collection

product_id	product_name	department [department_name]	aisle [aisle_name]

Order-Product Collection

order_id	product_id	re-ordered	add_to_cart

Orders Collection

order_id	user_id	order_dow	order_hour_of_day

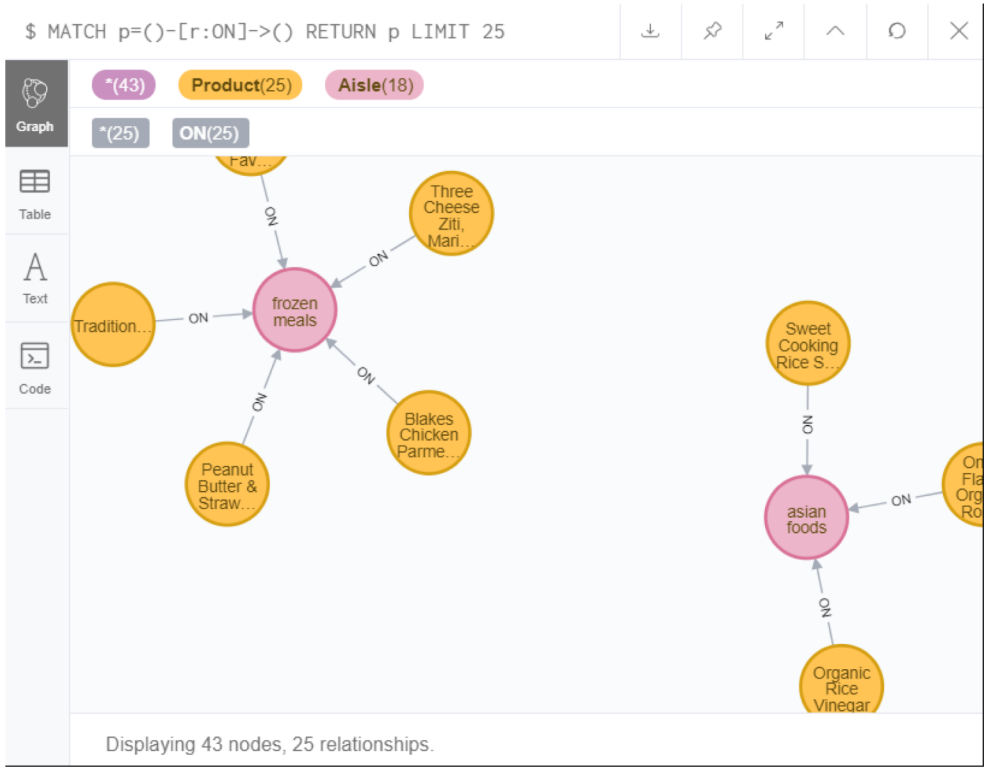
Neo4J (Graph Database)

- For the right use case, relational databases are powerful tools. But today's users are asking for *more* than an RDBMS can handle. More features, more data, more speed and – most importantly – *more connections.*
- Property graph model and Cypher query language makes it easy to understand a database.
- Neo4j delivers the lightning-fast read and write performance you need, while still protecting your **data integrity**. It is the only **enterprise-strength graph database** that combines **native graph** storage, **scalable** architecture optimized for speed, and **ACID** compliance to ensure predictability of relationship-based queries.

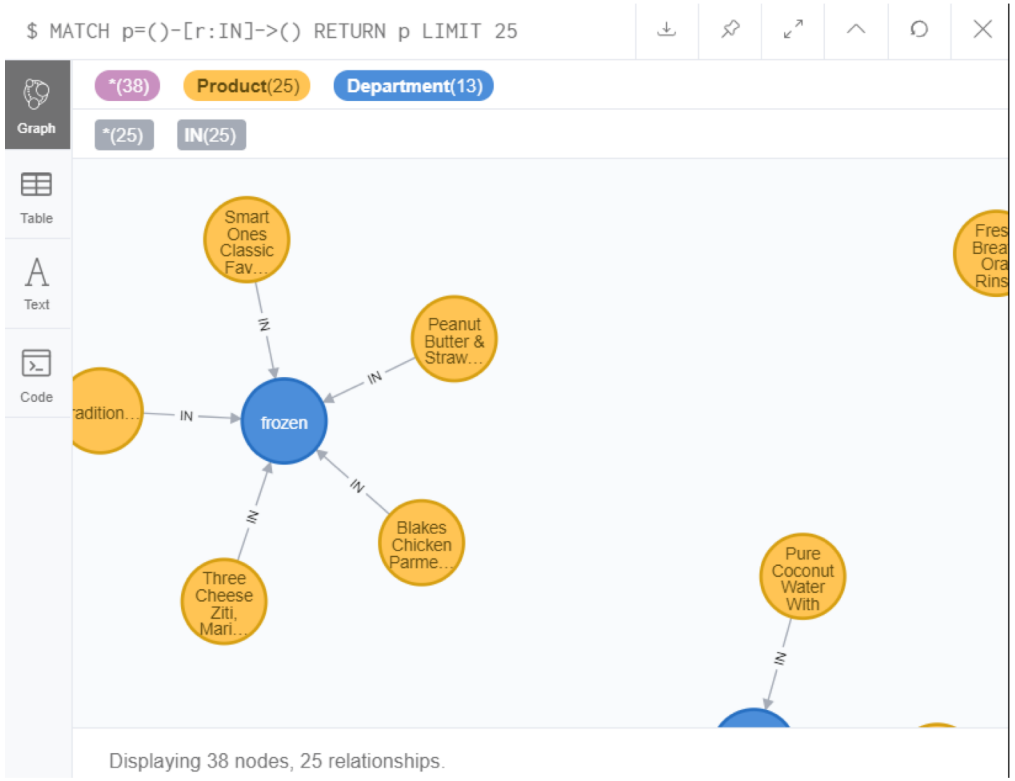
- Competitive advantage in business using Neo4j like ebay's competitive advantage in same-day delivery.

Relations

ON relation

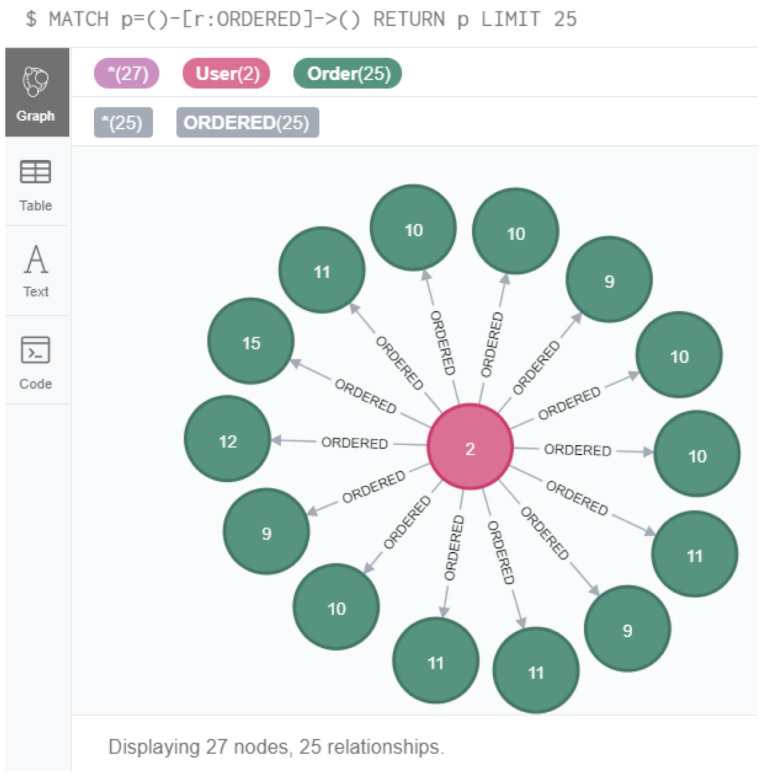


IN relation

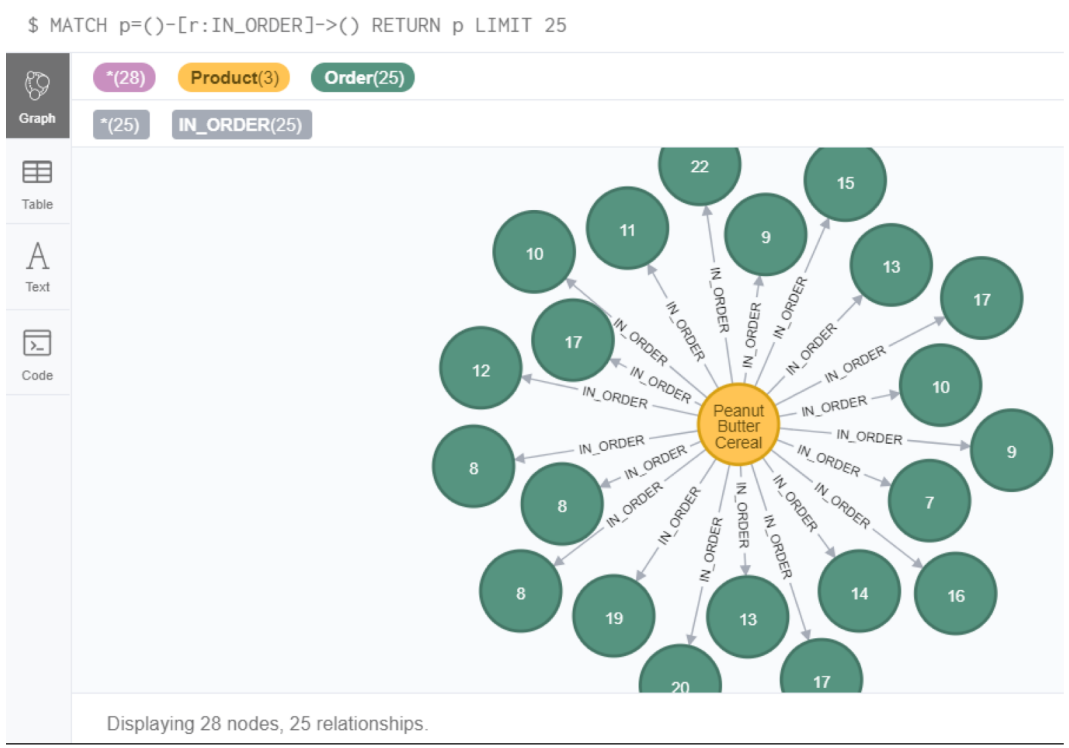


Relations -contd.

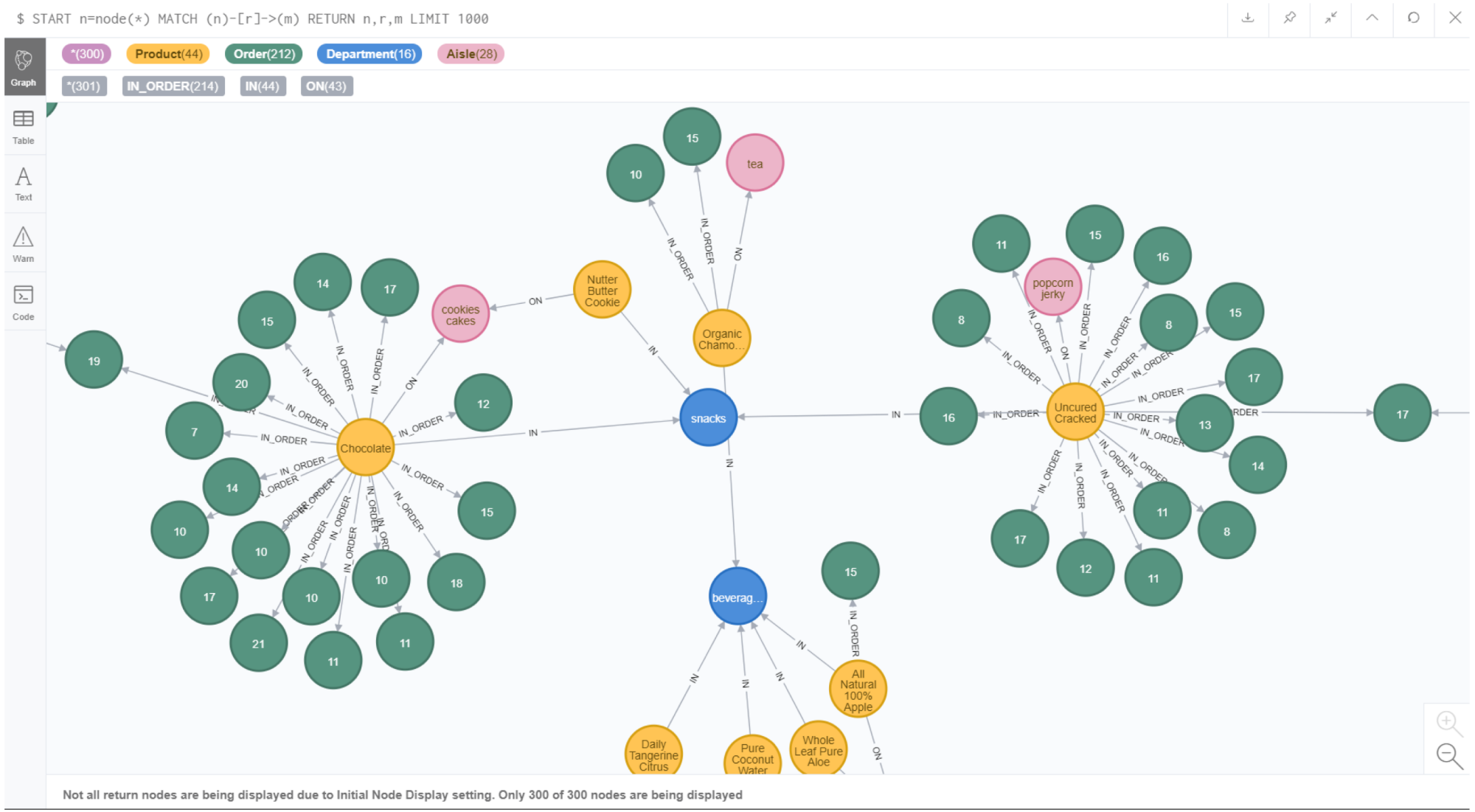
ORDERED



IN-ORDER



Graph Model


$ START n=node(*) MATCH (n)-[r]->(m) RETURN n,r,m LIMIT 1000
Graph
*(300)
Product(44)
Order(212)
Department(16)
Aisle(28)
*(301)
IN_ORDER(214)
IN(44)
ON(43)
Table
Text
Warn
Code
Chocolate
snacks
Uncured Cracked
Organic Chamo...
Nutter Butter Cookie
cookies cakes
tea
popcorn jerky
beverag...
All Natural 100% Apple
Daily Tangerine Citrus
Pure Coconut Water
Whole Leaf Pure Aloe
IN_ORDER
IN
ON
Not all return nodes are being displayed due to Initial Node Display setting. Only 300 of 300 nodes are being displayed