Supply Chain Analysis

•••

DSC-202: Data Management for Data Science

– guided by Dr. Amarnath Gupta

Team Members

Shreyasi Nath

A69034237

Varadraj Bartakke

A69034734

Bhaavya Naharas

A69034420

Application Context

Future-Ready Platform

Dual-Database Architecture

Scalability & Flexibility

Integrated Data Sources

Industry Relevance

Decision-Making & Analytics

Complex Supply Chain Environment

Application achievements

Real-Time Decision Support

Optimized Resource Allocation

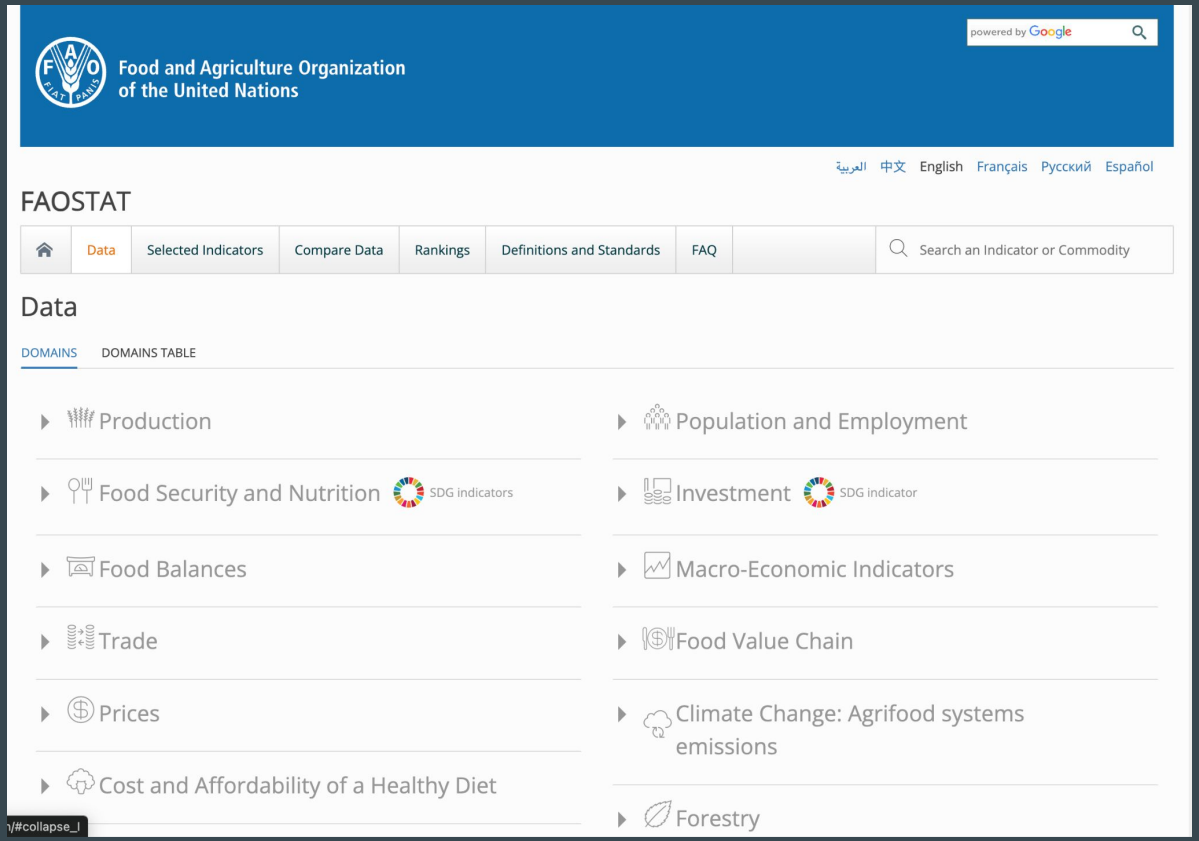
Enhanced Supply Chain Visibility

Accurate Inventory Tracking

Integrated Data Insights

Predictive Analytics

Data Source


Food and Agriculture Organization of the United Nations
powered by Google
العربية 中文 English Français Русский Español
FAOSTAT
Data
Selected Indicators
Compare Data
Rankings
Definitions and Standards
FAQ
Search an Indicator or Commodity
Data
DOMAINS
DOMAINS TABLE
Production
Food Security and Nutrition
SDG indicators
Food Balances
Trade
Prices
Cost and Affordability of a Healthy Diet
Population and Employment
Investment
SDG indicator
Macro-Economic Indicators
Food Value Chain
Climate Change: Agrifood systems emissions
Forestry

Data Sources

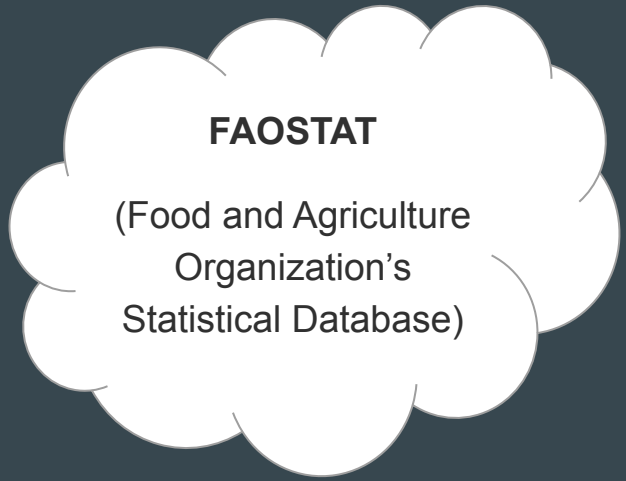
Role of data source

PRODUCTION

- Production of items/commodities
- Yearly data from 1990 to 2023
- Elements like Yield, Harvest, Crops, etc.

FOOD BALANCES

- Production and Consumption of commodities
- Indicators, Flags, Elements
- Yearly data from 1990 to 2023



URL: FAOSTAT Datasets

Data Sources

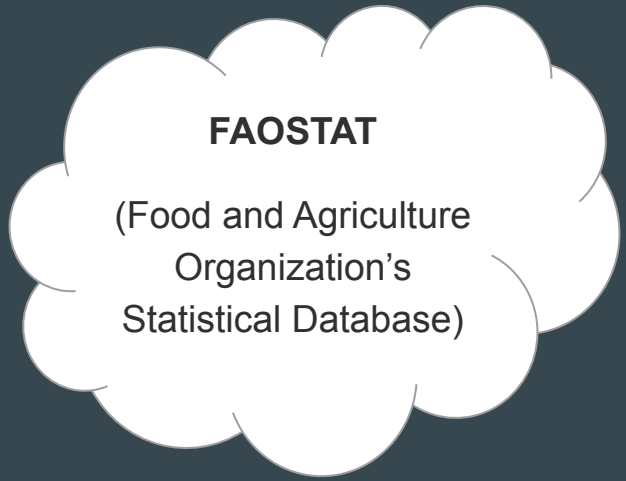
Role of data source

TRADE

- Trade Matrix between countries
- Import & Export quantities of products
- Trade Indicators

PRICES

- Price metrics in USD and SLC.
- Exchange rate data
- Producer and Consumer Prices



URL: FAOSTAT Datasets

Data Sources

Structure and content of data source

Price Data (price.csv)

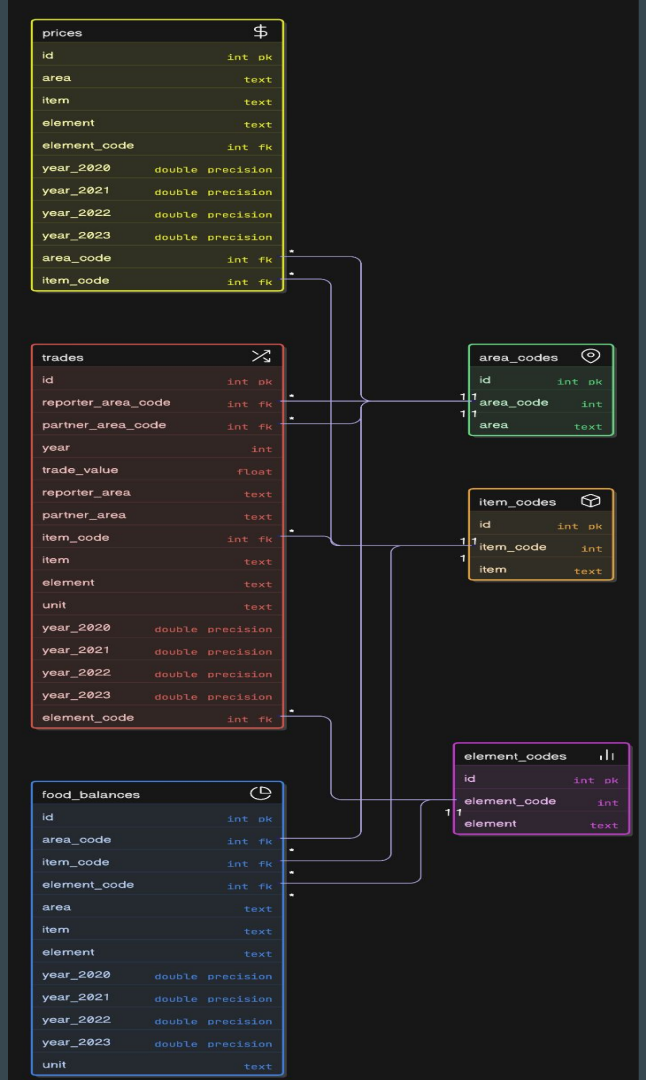
- **Structure:**
  - **Attributes:** area, area_code, item, item_code, element, element_code, unit.
  - **Time Series:** Columns for multiple years (from 1991 up to 2023, with a focus on recent years 2020–2023).
- **Content:**
  - Captures detailed price information for a wide range of products.
  - Consistent naming conventions facilitate historical and comparative analyses.

Trade Data (trade.csv)

- **Structure:**
  - **Attributes:** reporter_area, reporter_area_code, partner_area, partner_area_code, item, item_code, element, element_code, unit.
  - **Time Series:** Annual trade values for 2020 to 2023.
- **Content:**
  - Documents bilateral trade statistics and the movement of goods.
  - Combines geographic and product-specific data to reveal supply chain interconnections.

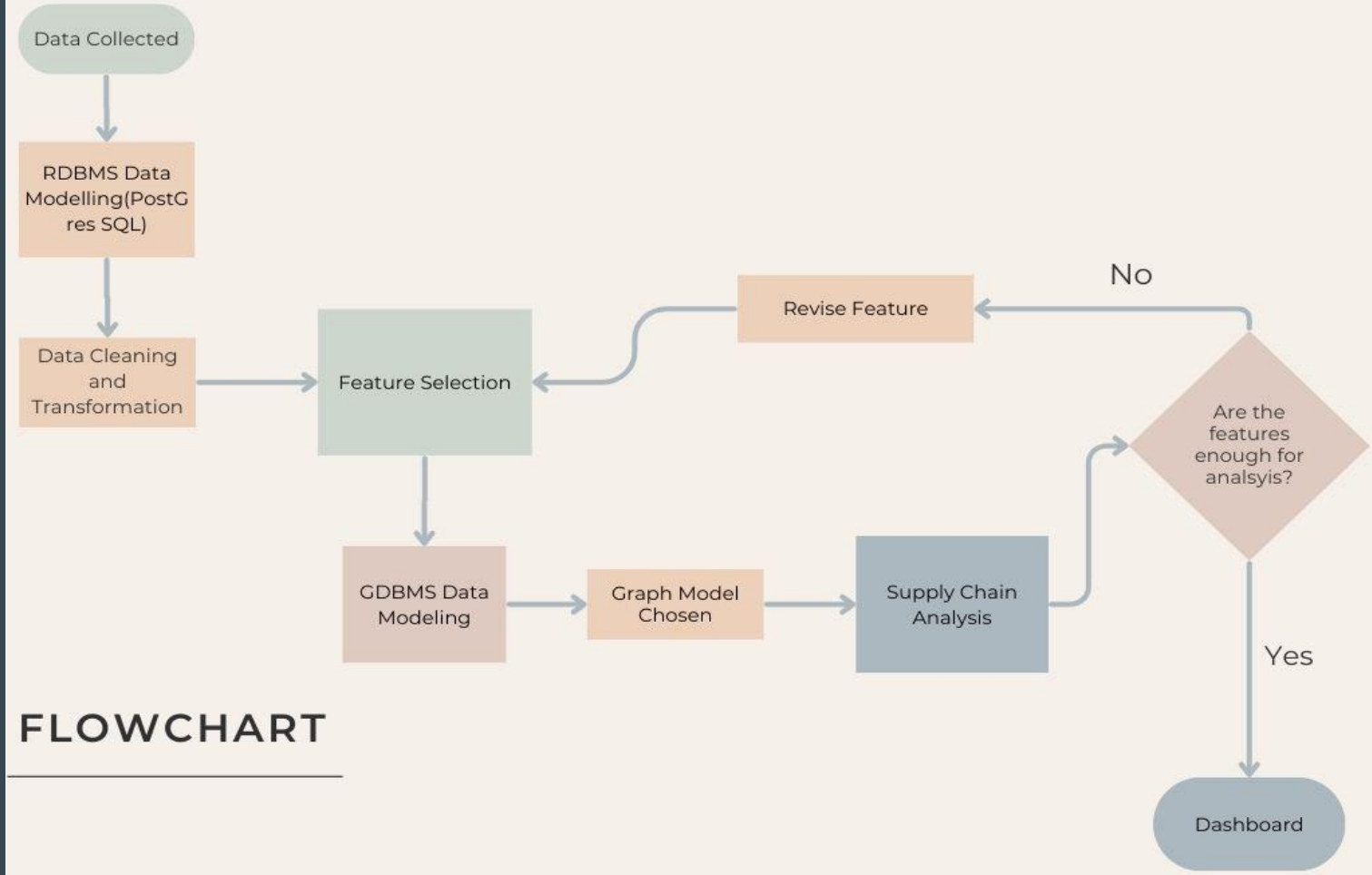
Food Balances Data (foodbalance.csv)

- **Structure:**
  - **Attributes:** area_code, area, item_code, item, element_code, element, unit.
  - **Time Series:** Annual records for the years 2020 to 2023.
- **Content:**
  - Represents the quantitative balance of food commodities.
  - Supports the assessment of production, consumption, and stock levels to manage food inventory effectively.



Methodology

Technical Steps Taken:



Methodology

1. RDBMS

Relational Data Model

- **Structured Schema Design:**
  - Defined tables for Price Data, Trade Data, and Food Balances.
  - Each table includes attributes for region/country codes, products, and time series data (years).
- **Normalization:**
  - Ensured minimal data redundancy by normalizing entities like area, item, and element.

Feature Selection

- **Key Variables Identification:**
  - Selected critical variables like trade volume, price trends, and commodity balances.
- **Dimensionality Reduction:**
  - Removed extraneous columns to focus on impactful features.
- **Data Integration:**
  - Combined insights across Price, Trade, and Food Balances datasets to enable comprehensive inventory analysis.

Preprocessing

- **Data Cleaning:**
  - Handled missing values, standardized units, and corrected inconsistent formats.
- **Data Transformation:**
  - Converted raw data into time series formats.
  - Aggregated and normalized data to prepare for further analysis.

Logical Combination of Data Sources

- Joined tables via common keys (e.g., area_code, item_code) to create unified views.
- Created summary tables and views to support queries on inventory levels, pricing trends, and trade relationships.

```sql
CREATE TABLE prices (
    area_code    INTEGER,
    area         TEXT,
    item_code    INTEGER,
    item         TEXT,
    element      TEXT,
    element_code INTEGER,
    unit         TEXT,
    year_2020        DOUBLE PRECISION,
    year_2021        DOUBLE PRECISION,
    year_2022        DOUBLE PRECISION,
    year_2023        DOUBLE PRECISION,
    PRIMARY KEY (area_code, item_code, element_code),  -- Unique per country-product-element
    FOREIGN KEY (area_code) REFERENCES area_codes(area_code),
    FOREIGN KEY (item_code) REFERENCES item_codes(item_code),
    FOREIGN KEY (element_code) REFERENCES element_codes(element_code)
);

--food bal load

INSERT INTO postgres.public.area_codes (
    area_code,
    area)
select area_code,
    area
from production_cle_area_code
where area IN ('United States of America','Ecuador','Uruguay','Canada','Brazil');

INSERT INTO postgres.public.item_codes (
    item_code,
    item)
select item_code, item
from production_cle_item_code
where  item IN ('Apples','Oats','Spinach','Chickens','Wheat');

INSERT INTO element_codes (
    element_code,
    element)
select element_code, element
from production_cle_elements
--where  element IN ('Import quantity','Import value','Export quantity','Export value','Production','Food supply quantity (tonnes)','Producer Price (USD/tonne)')
 ;
```

Exported:

1. food_balances.csv
2. trades.csv
3. prices.csv

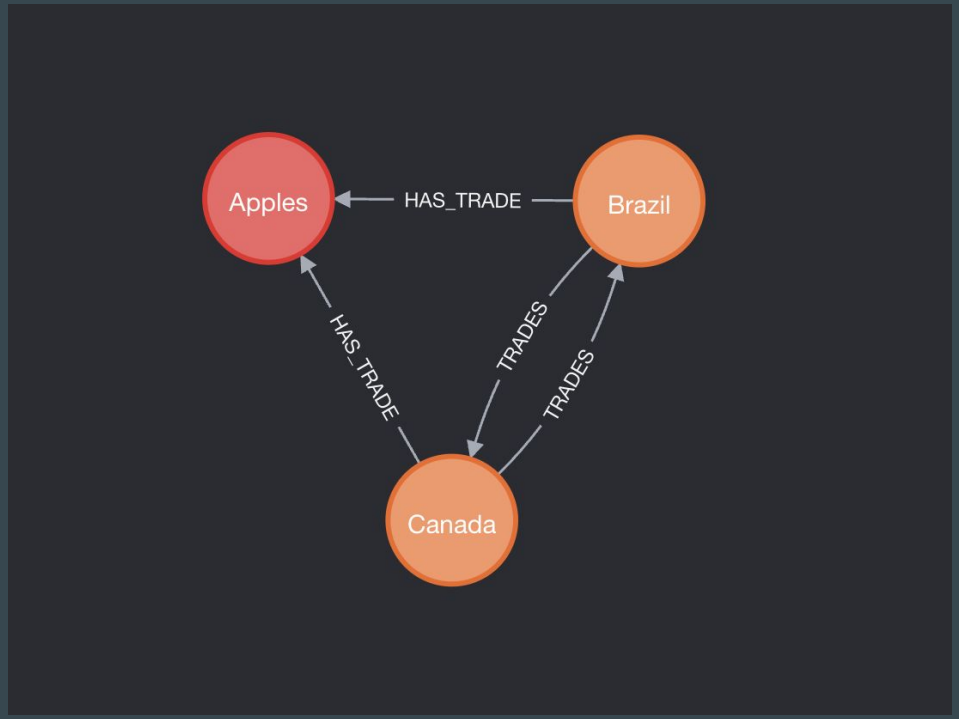
```sql
CREATE TABLE food_balances (
    area_code    INTEGER,
    area         TEXT,
    item_code    INTEGER,
    item         TEXT,
    element_code INTEGER,
    element      TEXT,
    unit         TEXT,
    year_2020    DOUBLE PRECISION,
    year_2021    DOUBLE PRECISION,
    year_2022    DOUBLE PRECISION,
    year_2023    DOUBLE PRECISION,
    PRIMARY KEY (area_code, item_code, element_code),  -- Ensures unique country-product-element
    FOREIGN KEY (area_code) REFERENCES area_codes(area_code),  -- Links to area table
    FOREIGN KEY (item_code) REFERENCES item_codes(item_code),  -- Links to item table
    FOREIGN KEY (element_code) REFERENCES element_codes(element_code) -- Links to element table
);

-- Table for Trade (Import & Export Data)
CREATE TABLE trades (
    reporter_area_code INT,
    reporter_area      TEXT,
    partner_area_code  INT,
    partner_area       TEXT,
    item_code          INT,
    item              TEXT,
    element_code       INT,
    element           TEXT,
    unit              TEXT,
    year_2020         DOUBLE PRECISION,
    year_2021         DOUBLE PRECISION,
    year_2022         DOUBLE PRECISION,
    year_2023         DOUBLE PRECISION,
    PRIMARY KEY (reporter_area_code, partner_area_code, item_code, element_code),
    FOREIGN KEY (reporter_area_code) REFERENCES area_codes(area_code),
    FOREIGN KEY (partner_area_code) REFERENCES area_codes(area_code),
    FOREIGN KEY (item_code) REFERENCES item_codes(item_code),
    FOREIGN KEY (element_code) REFERENCES element_codes(element_code)
);
```

Methodology

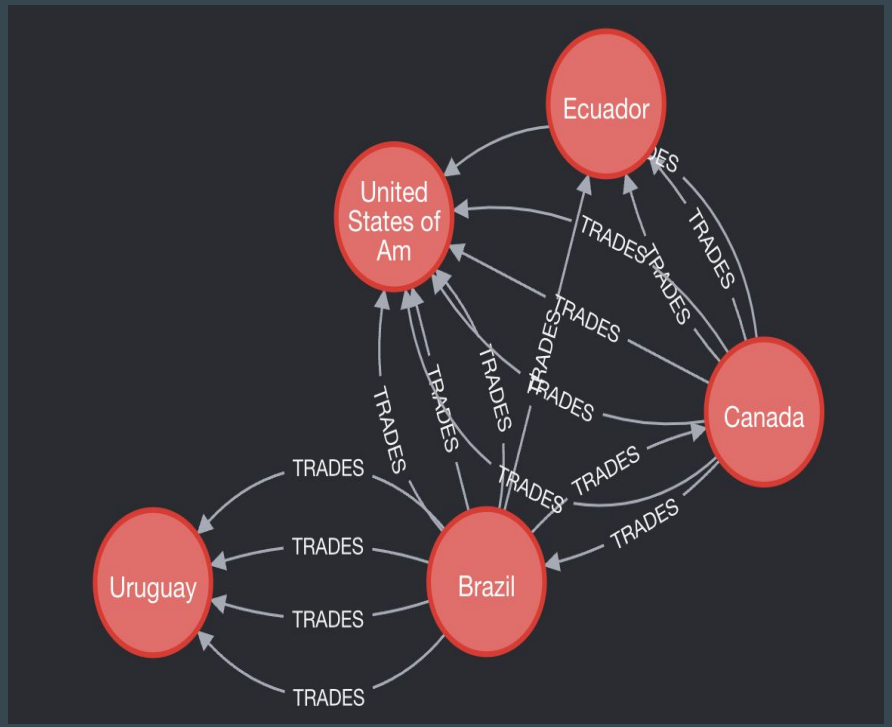
Graph Data Model (Old vs New)

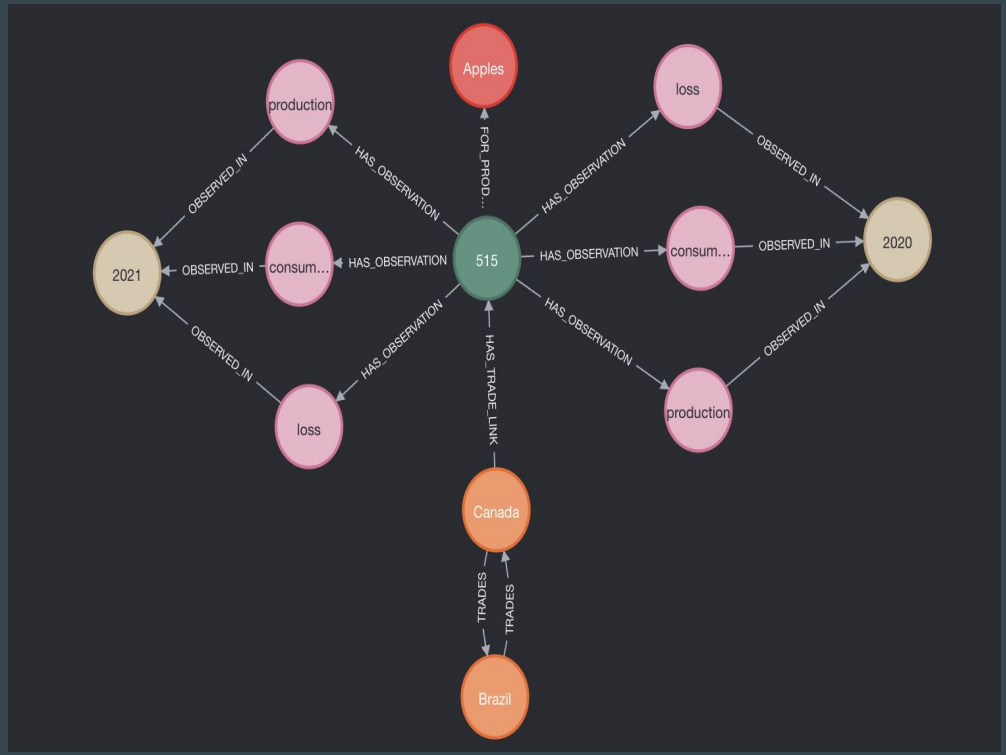
Old Model	New Model
Nodes: Country, Product Relationships: 1. (Country)-[:TRADES {export, unit}]->(Country) 2. (Country)-[:HAS_TRADE {production,consumption, loss, unit, year}]->(Country)	Nodes: Country, Product, Year, Trade Link, Element Relationships: 1. (Country)-[:TRADES {year, volume, price}]->(Country) 2. (Country)-[:HAS_TRADE_LINK]->(TradeLink Node)-[:HAS_OBSERVATION]->(ELEMENT)-[:OBSERVED_IN]->(YEAR) 3. (Country)-[:HAS_TRADE_LINK]->(TradeLink Node)-[:FOR_PRODUCTION]->(PRODUCT)

Old Model



New Model


Ecuador
United States of Am
Canada
Uruguay
Brazil
TRADES


Apples
production
loss
2021
consum...
515
consum...
2020
loss
production
Canada
Brazil
FOR_PROD...
HAS_OBSERVATION
OBSERVED_IN
HAS_TRADE_LINK
TRADES

Methodology- End Result Structure

DBMS Integration:

We have combined two Database Management Systems- RDBMS (here, PostgreSQL) and GDBMS (here, Neo4j), to form an integrated environment. The RDBMS manages structured data like prices, trades and food balances, while the GDBMS focuses on the complex relationships among countries, items, prices and traders.

Unified Dashboards & Advanced Queries:

We used this dual-layered data architecture to create interactive dashboards, merging key insights from both the relational and graph databases. Additionally, we ran Cypher queries to tackle complex real-world challenges, such as tracking price fluctuations and shortest path graphs, ensuring a comprehensive and flexible solution.

Demonstration

- RDBMS Query
- Node-relationship
- Cypher Queries(simple-complex)
- Dashboards

Cypher Queries

Answering real world problems

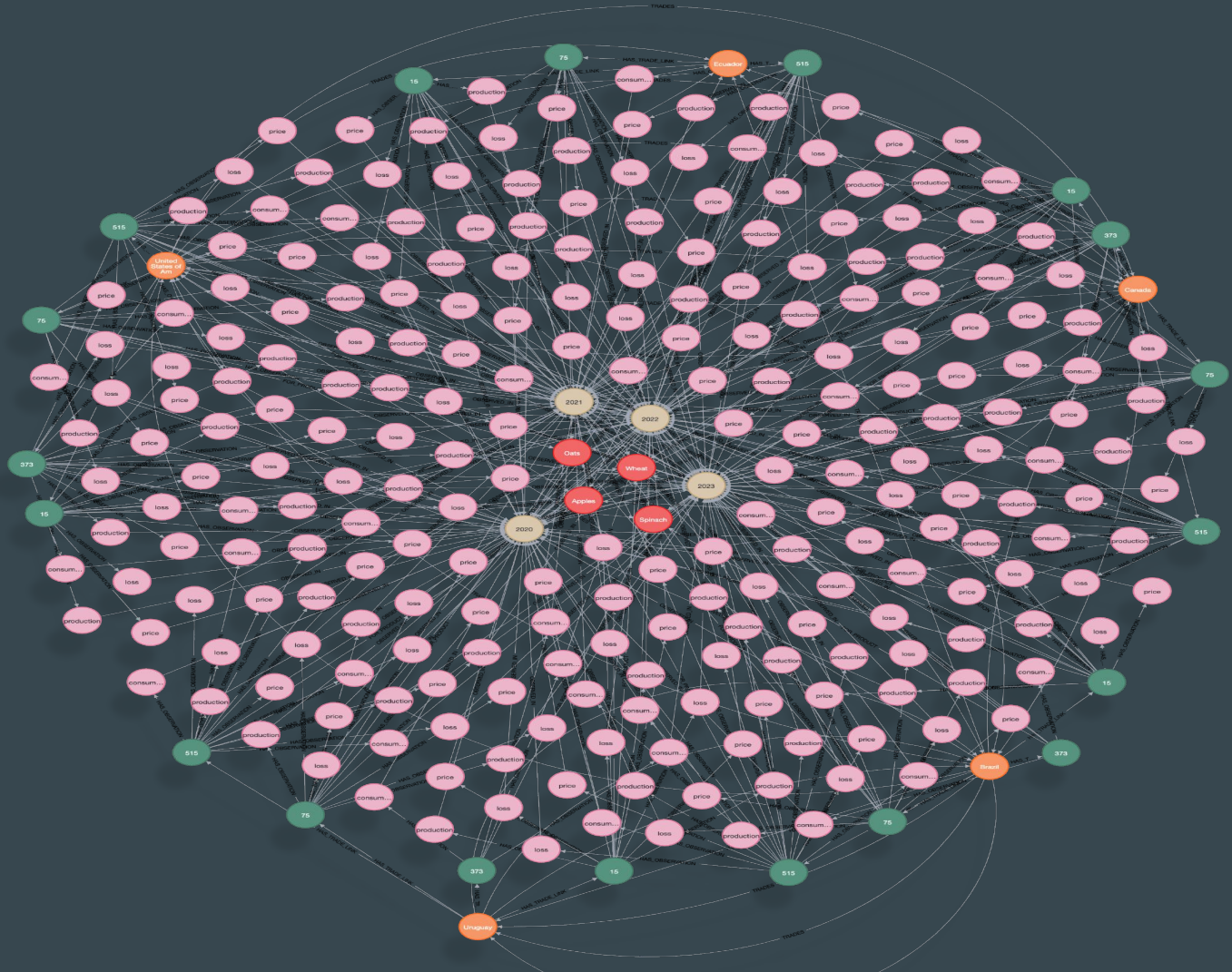
```
MATCH (c:Country)-[:HAS_TRADE_LINK]→(tl:TradeLink)-[:FOR_PRODUCT]→(p:FoodProduct)
MATCH (tl)-[:HAS_OBSERVATION]→(o:Observation)-[:OBSERVED_IN]→(y:Year)
RETURN c.name, p.name, o.type, o.value, y.year
LIMIT 20;
```

```
MATCH (c1:Country)-[r:TRADES]→(c2:Country)
RETURN c1.name, c2.name, r.product_name, r.export_2023
LIMIT 10;
```

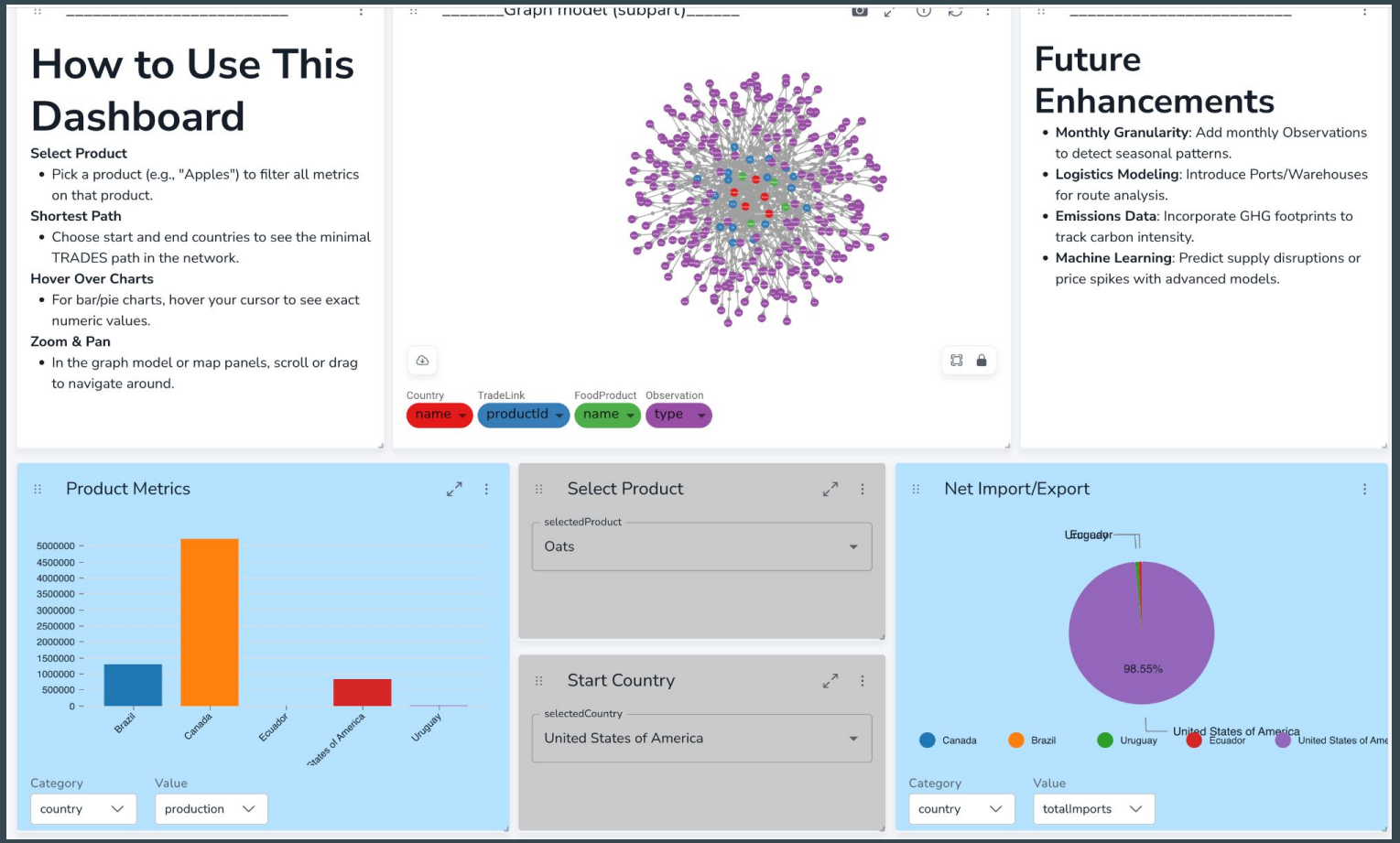
```
LOAD CSV WITH HEADERS FROM 'file:///trades.csv' AS row
MERGE (c:Country {
  area_code: toInteger(trim(row.reporter_area_code)),
  name: trim(row.reporter_area)
})
MERGE (c2:Country {
  area_code: toInteger(trim(row.partner_area_code)),
  name: trim(row.partner_area)
});
```

```
MATCH (c:Country {name: "Brazil"})-[r:TRADES]→(partner:Country)
OPTIONAL MATCH (c)-[:Htl_TRADE_LINK]→(tl:TradeLink)-[:FOR_PRODUCT]→(p:FoodProduct)

OPTIONAL MATCH (tl)-[:HAS_OBSERVATION]→(prod:Observation {type: "production"})
OPTIONAL MATCH (tl)-[:HAS_OBSERVATION]→(loss:Observation {type: "loss"})
OPTIONAL MATCH (tl)-[:HAS_OBSERVATION]→(cons:Observation {type: "consumption"})

OPTIONAL MATCH (company:Company)-[:INVOLVED_IN]→(tl)
// Companies related to trade links

OPTIONAL MATCH (company)-[:PRODUCES]→(prod)
OPTIONAL MATCH (company)-[:LOSSES]→(loss)
```

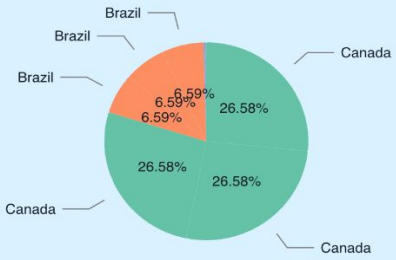
Result


Ecuador
Canada
Brazil
Uruguay
Oats
Wheat
Apples
Spinach
2020
2021
2022
2023
price
production
consum...
loss

Dashboard


How to Use This Dashboard
Select Product
• Pick a product (e.g., "Apples") to filter all metrics on that product.
Shortest Path
• Choose start and end countries to see the minimal TRADES path in the network.
Hover Over Charts
• For bar/pie charts, hover your cursor to see exact numeric values.
Zoom & Pan
• In the graph model or map panels, scroll or drag to navigate around.
Country
name
TradeLink
productId
FoodProduct
name
Observation
type
Future Enhancements
• Monthly Granularity: Add monthly Observations to detect seasonal patterns.
• Logistics Modeling: Introduce Ports/Warehouses for route analysis.
• Emissions Data: Incorporate GHG footprints to track carbon intensity.
• Machine Learning: Predict supply disruptions or price spikes with advanced models.
Product Metrics
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
Brazil
Canada
Ecuador
Uruguay
Category
country
Value
production
Select Product
selectedProduct
Oats
Start Country
selectedCountry
United States of America
Net Import/Export
98.55%
Canada
Brazil
Uruguay
Ecuador
Category
country
Value
totalImports

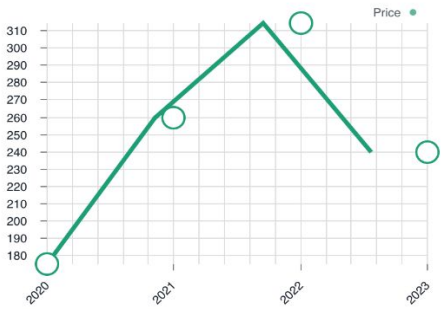
Alternate supplier of a product



Category: AlternateSupplier

Value: ProductionQuantity

Price fluctuations



X-value: Year

Y-value: Price

Trade ban

```
Query returned no data.
```

Import&Export of all countries all products

country	product	totalExports	totalImports
Canada	Wheat	1,030,422	33,219
Canada	Oats	427,975	0
Brazil	Wheat	274,707	19,472
Canada	Apples	38,116	0
Canada	Spinach	18,744	0

Rows per page: 5 1–5 of 22

Start Country

startcountry: United States of America

End Country

endcountry: Canada

Shortest trade path

Country: name

Net Revenue

ExportingCountry	Revenue	Unit
Canada	1,030,422	1000 USD
Brazil	274,707	1000 USD
United States of A...	116,369	1000 USD
Uruguay	39,759	1000 USD

Rows per page: 5 1–4 of 4

Select Product

chosenProduct

Wheat

._____.

"Visualization sells insights, not pictures."

Product Growth Rate

Product	Year	TotalConsump...	GrowthRate
Spinach	2,023	358,630.09	20.34
Apples	2,022	5,196,721.32	0.65
Spinach	2,022	298,010.32	-1.71
Apples	2,023	4,953,547.45	-4.68
Apples	2,021	5,162,988.69	-8.61

Rows per page: 5 1–5 of 6

CURRENT LIMITATIONS

Limited Advanced Analytics

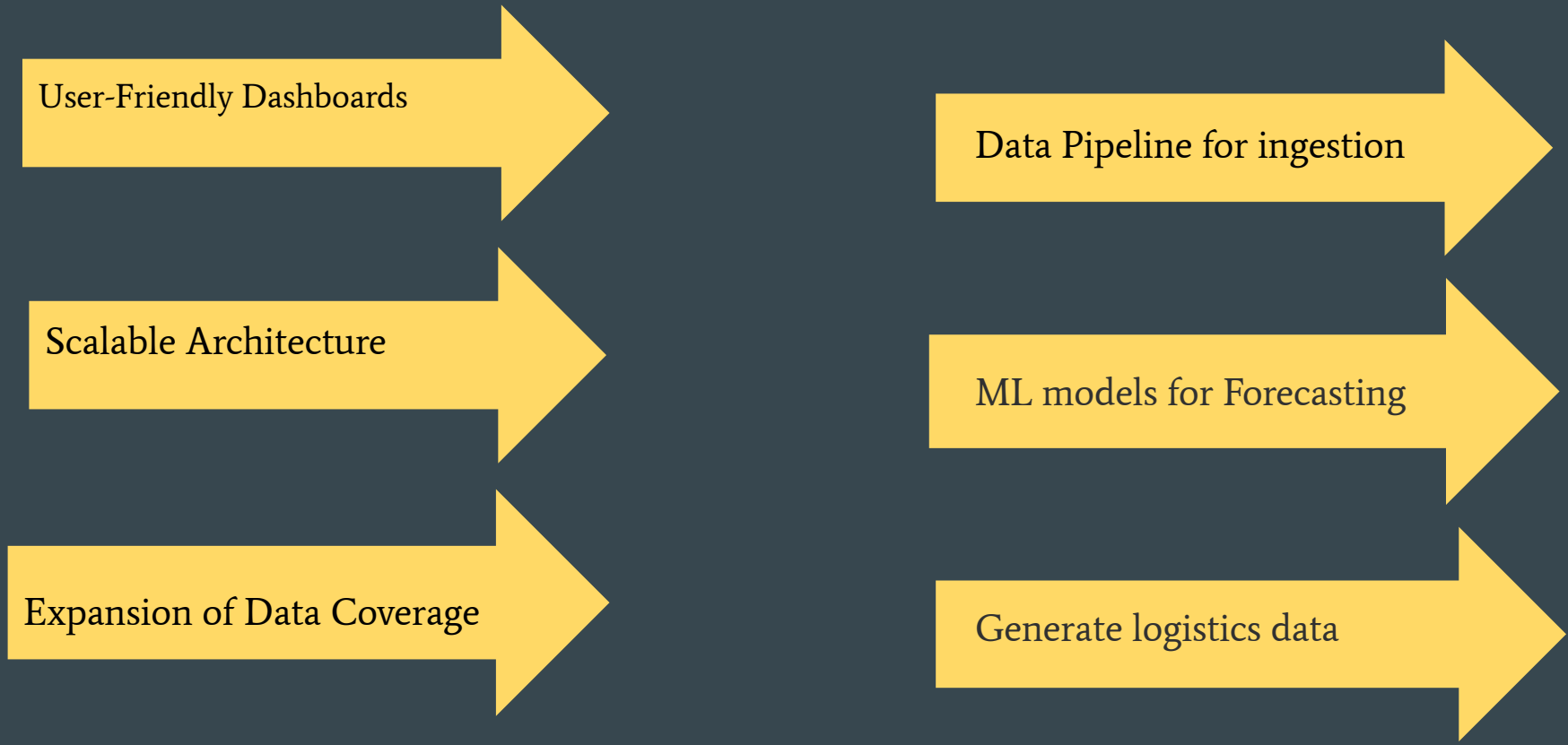
Complexity of Multi-DBMS Integration

Lack of logistics data

Partial Dataset Utilization

Data Refresh and Updates

FUTURE SCOPE



THANK YOU

...

LET'S DISCUSS