

# What are the different layers of a data warehouse?

DATA WAREHOUSING CONCEPTS



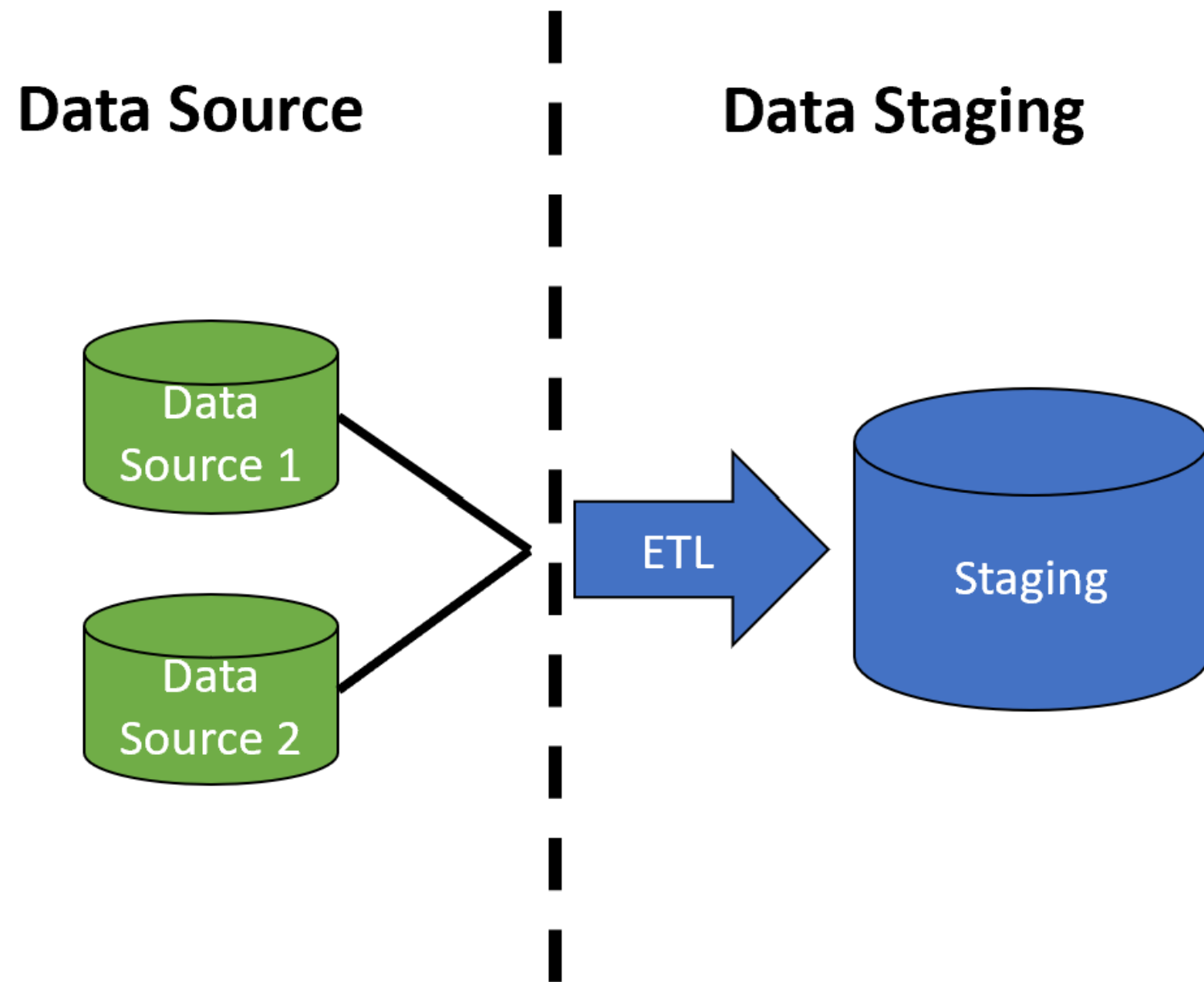
**Aaren Stubberfield**  
Data Scientist

# Layer overview - data source

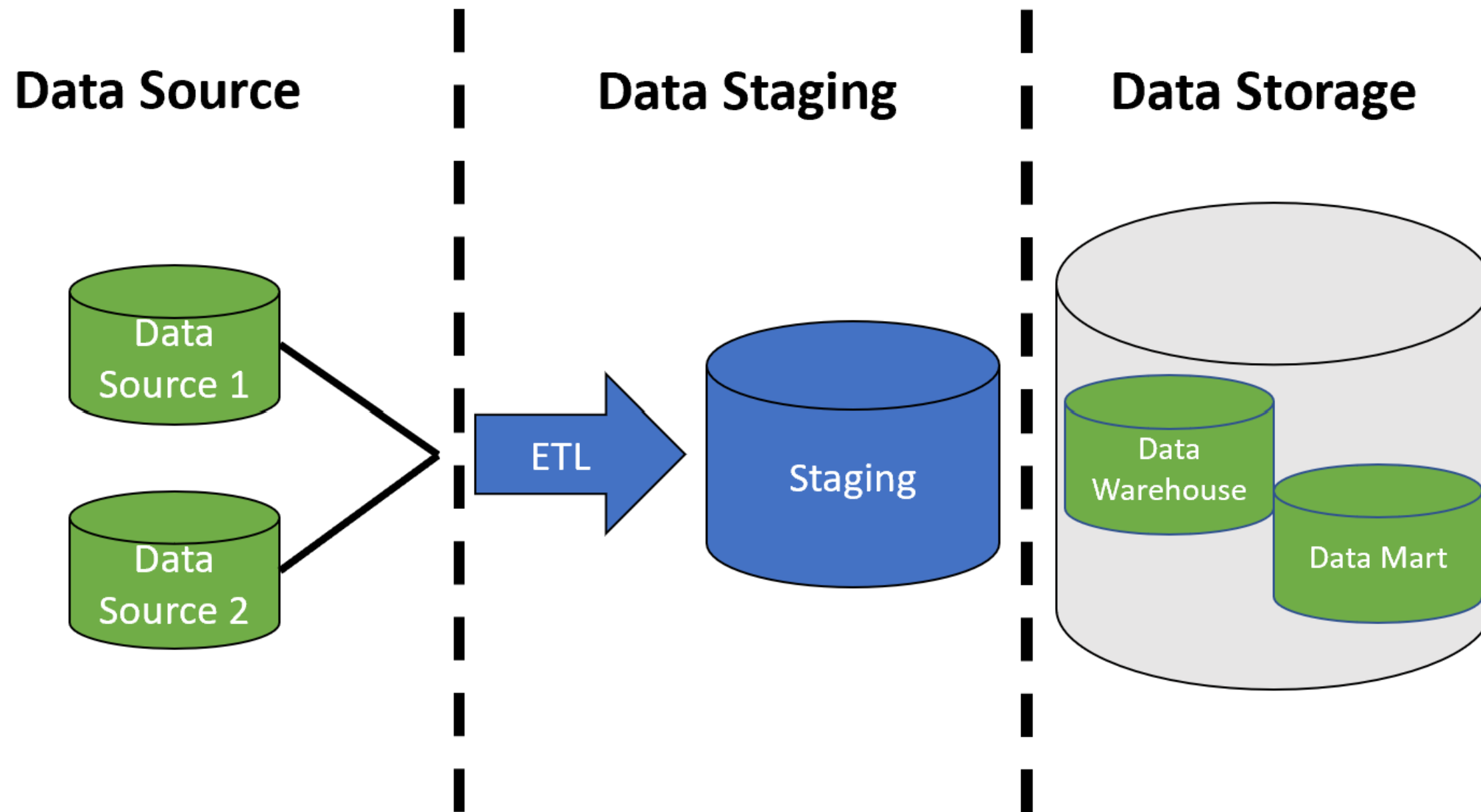
## Data Source



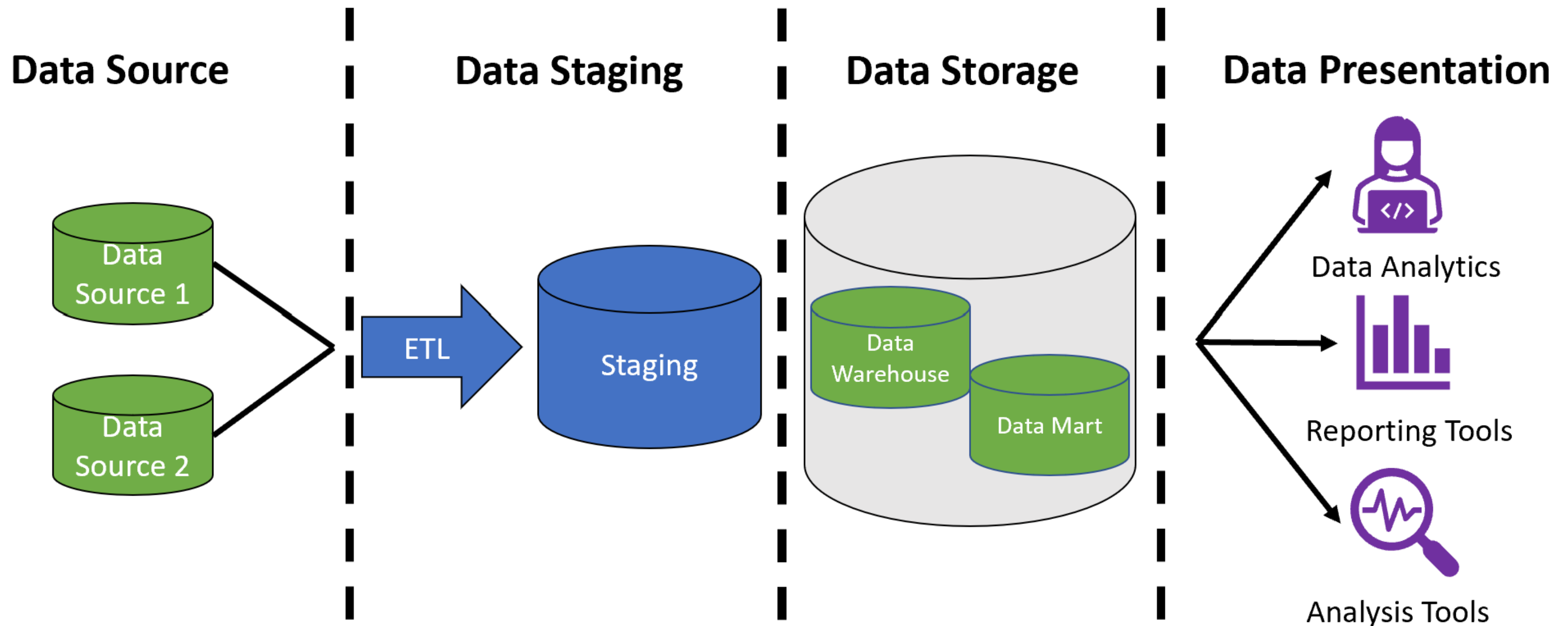
# Layer overview - data staging



# Layer overview - data storage



# Layer overview - data presentation



# Data source layer

- All data sources for data warehouse
- Examples of data sources:
  - Transactional database
  - Log files
  - Spreadsheets

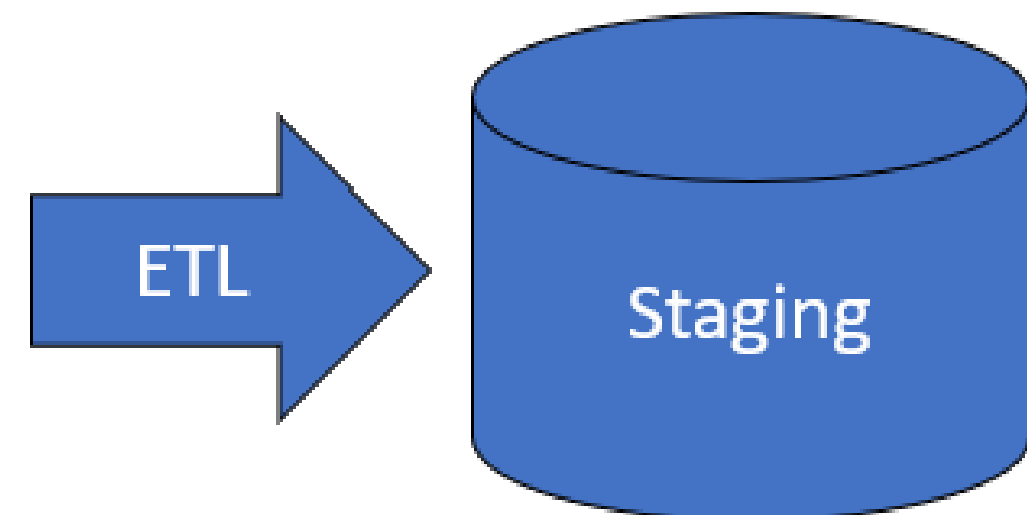
## Data Source



# Data staging layer

- Layer extracts, transform, and clean data through ETL process
- Contains ETL process and storage tables

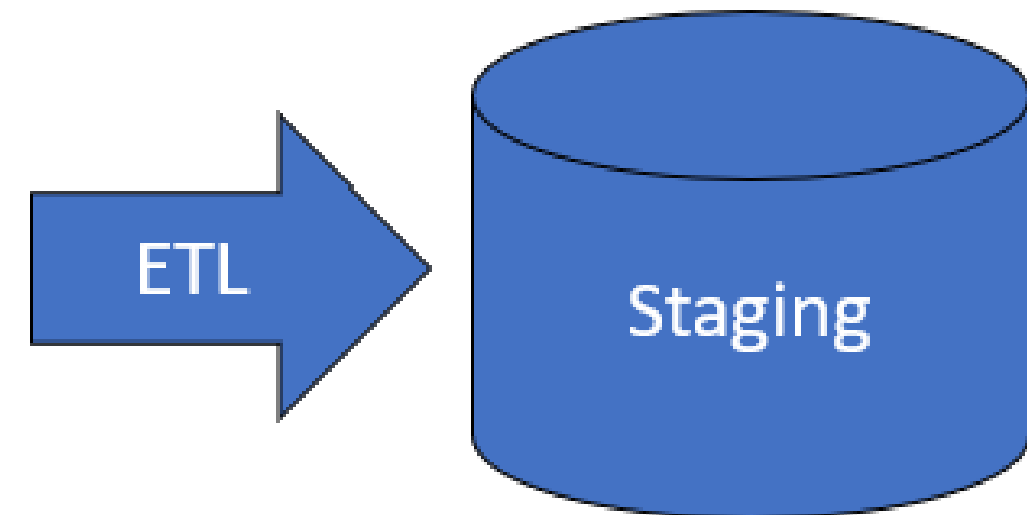
## Data Staging



# ETL process within data staging layer

- Extracted
- Business rules applied and cleaned
- Staging database often used
- Must be able to extract valid data
- Batch / full loading

## Data Staging

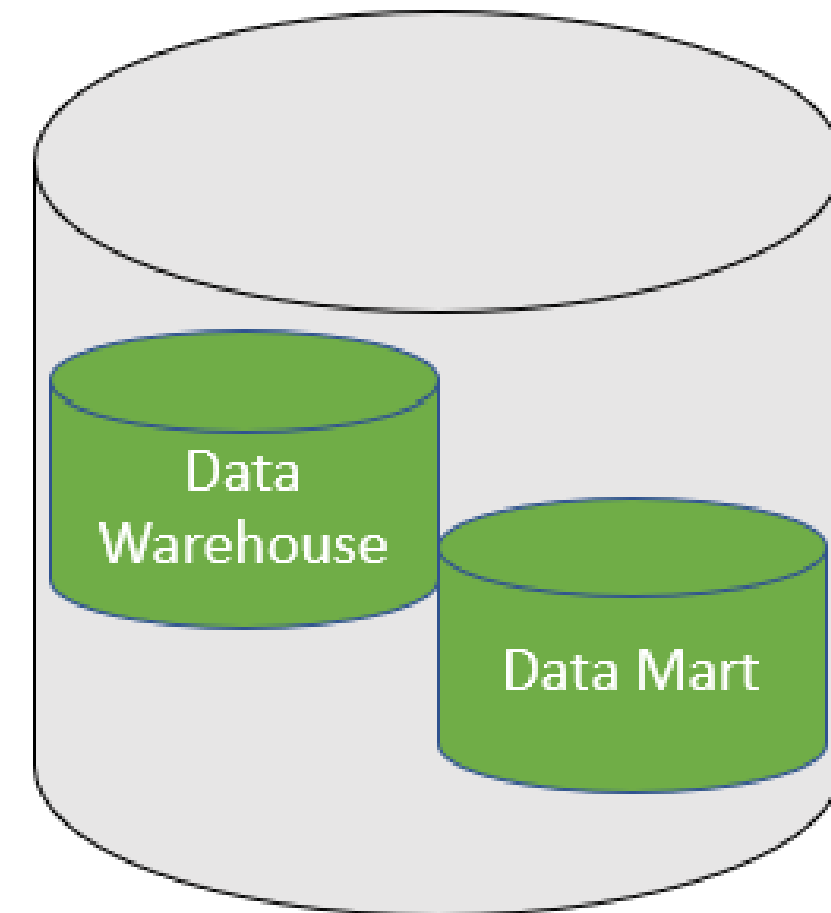




# Data storage layer

- Data is stored in warehouse and data marts
  - *Data warehouse -> Data mart*
  - *Data mart -> Data warehouse*

## Data Storage



# Data presentation layer

- Users interact with stored data
- Users:
  - Use BI (Business Intelligence) tools
  - Use data mining tools
  - Create direct queries

## Data Presentation



Data Analytics

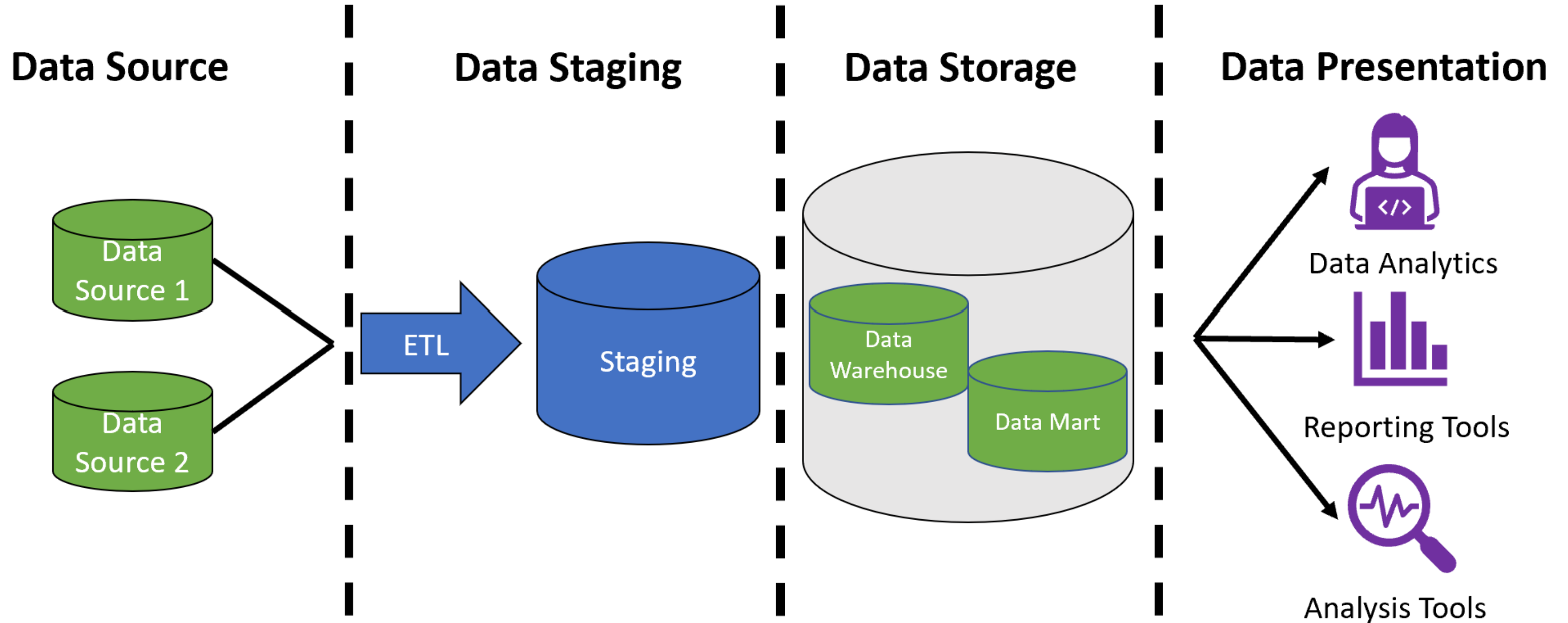


Reporting Tools



Analysis Tools

# Summary



**Let's practice!**  
DATA WAREHOUSING CONCEPTS

# The presentation layer

DATA WAREHOUSING CONCEPTS



**Aaren Stubberfield**  
Data Scientist

# Presentation layer tools

- Users interact with the presentation layer
  - Area of constant development

## Presentation Layer Groups:

- Automated reporting/dashboarding tools
- BI/data analytics
- Direct queries

## Data Presentation



Data Analytics



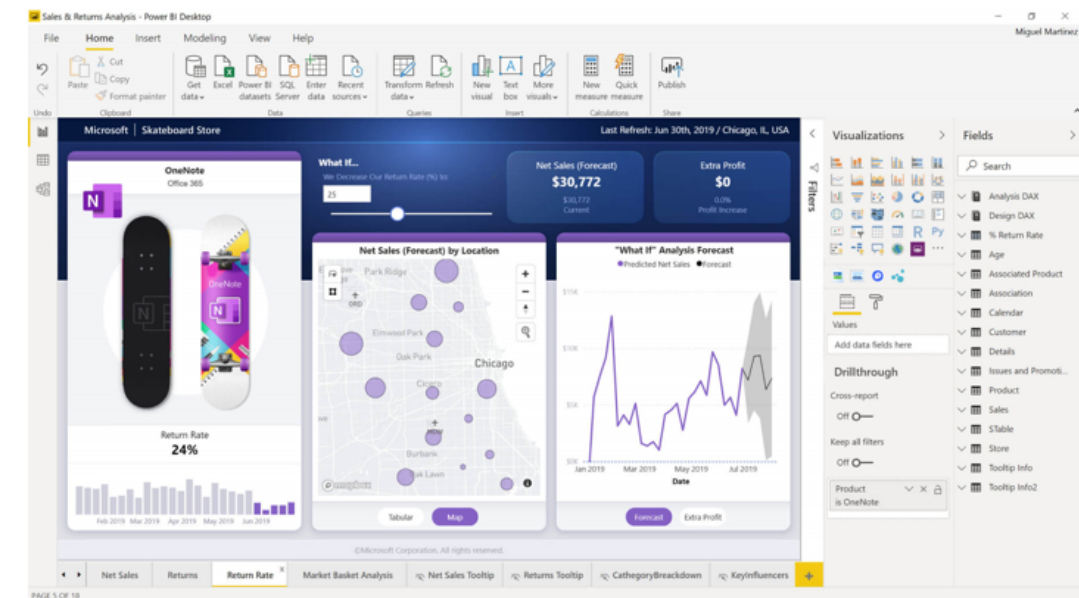
Reporting Tools



Analysis Tools

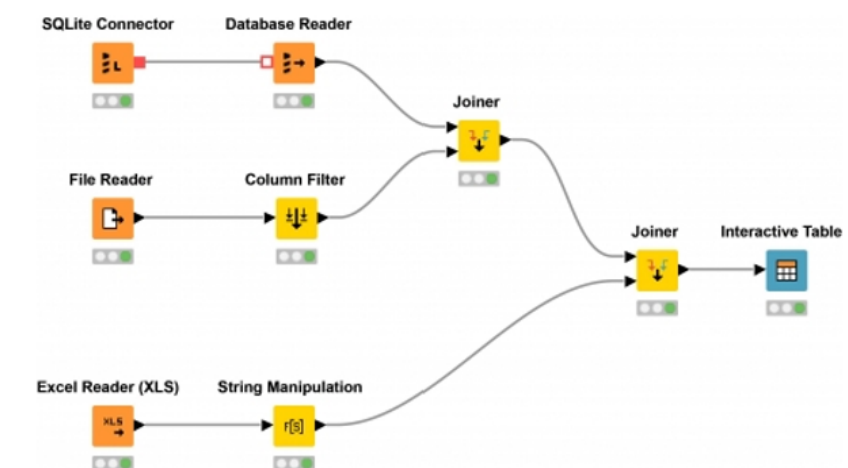
# Automated reporting/dashboarding

- Goal:
  - Create reports needed for decision making
  - Create dashboards using historical data
- Users:
  - Analysts
  - Citizen Data Scientist



# BI/data analytics

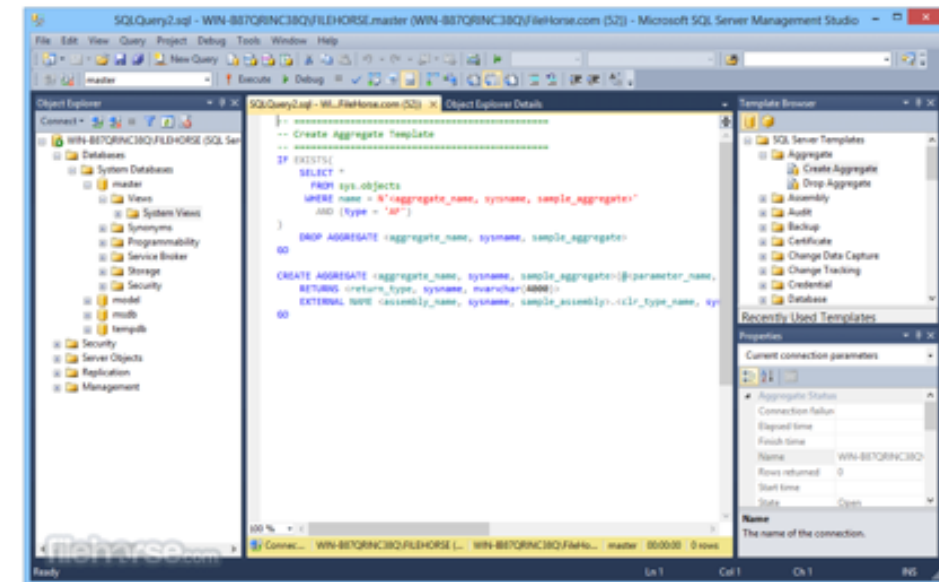
- Goal:
  - Tools for exploration
  - Looking for patterns
- Users:
  - Analysts
  - Data Scientist





# Direct queries

- Goal:
  - Sophisticated tools for exploration
- Users:
  - Analysts
  - Data Scientist
  - Data Engineer



**Let's practice!**  
DATA WAREHOUSING CONCEPTS

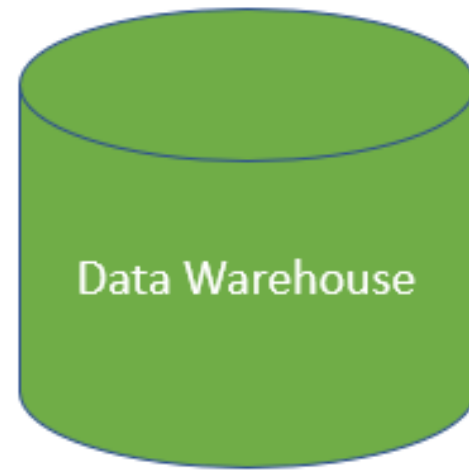
# Data warehouse architectures

DATA WAREHOUSING CONCEPTS



**Aaren Stubberfield**  
Data Scientist

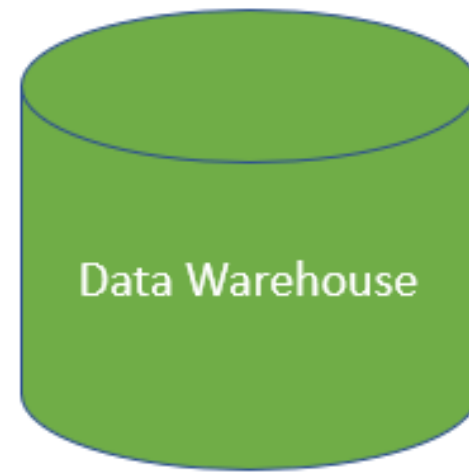
# Inmon - top-down



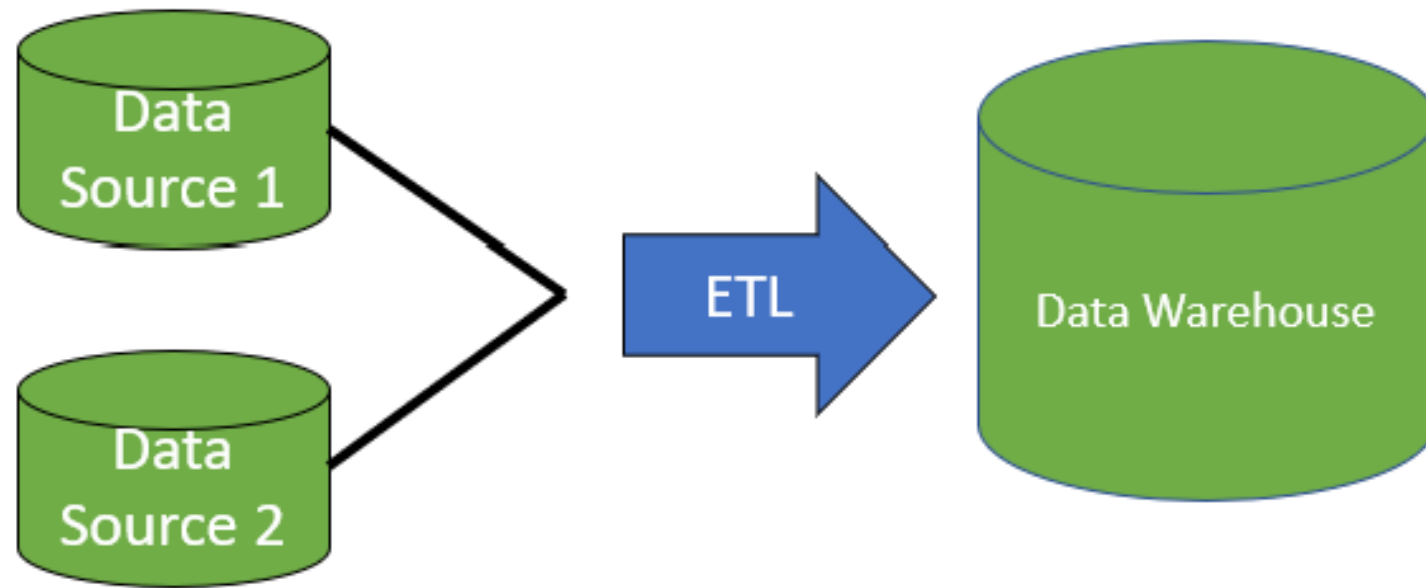
# Inmon - top-down

Must decide:

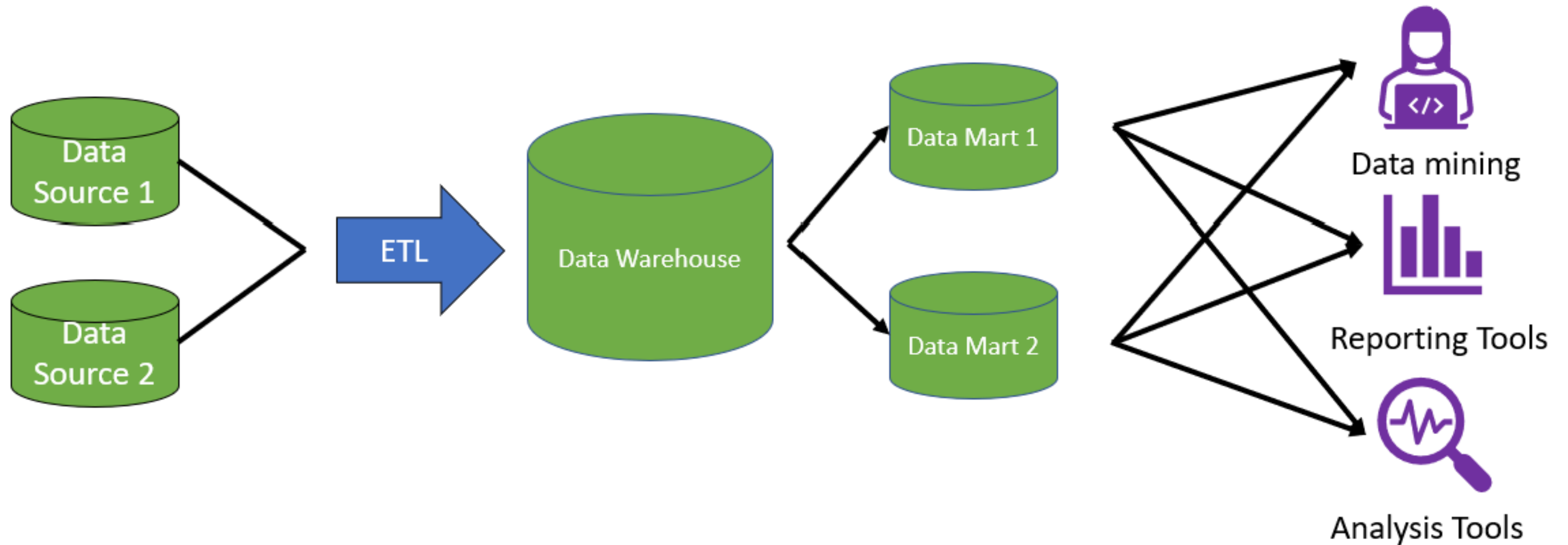
- On all data definitions, cleaning, and business rules
- Before any data enters warehouse



# Inmon - top-down



# Inmon - top-down



# Pros and cons of top-down

## Advantages:

- Single source of truth for organization
- Normalization = less storage
- Easy to change data marts to support reporting changes

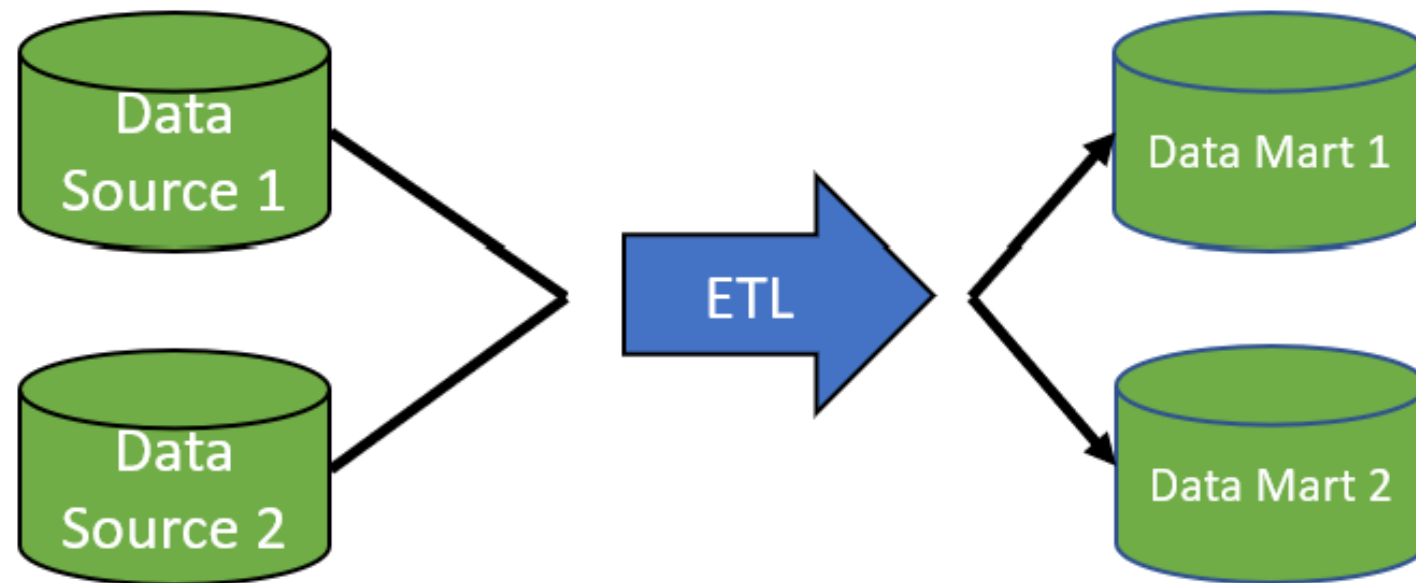
## Disadvantages:

- More joins = slower response time
- Lengthy upfront work
  - Higher startup cost

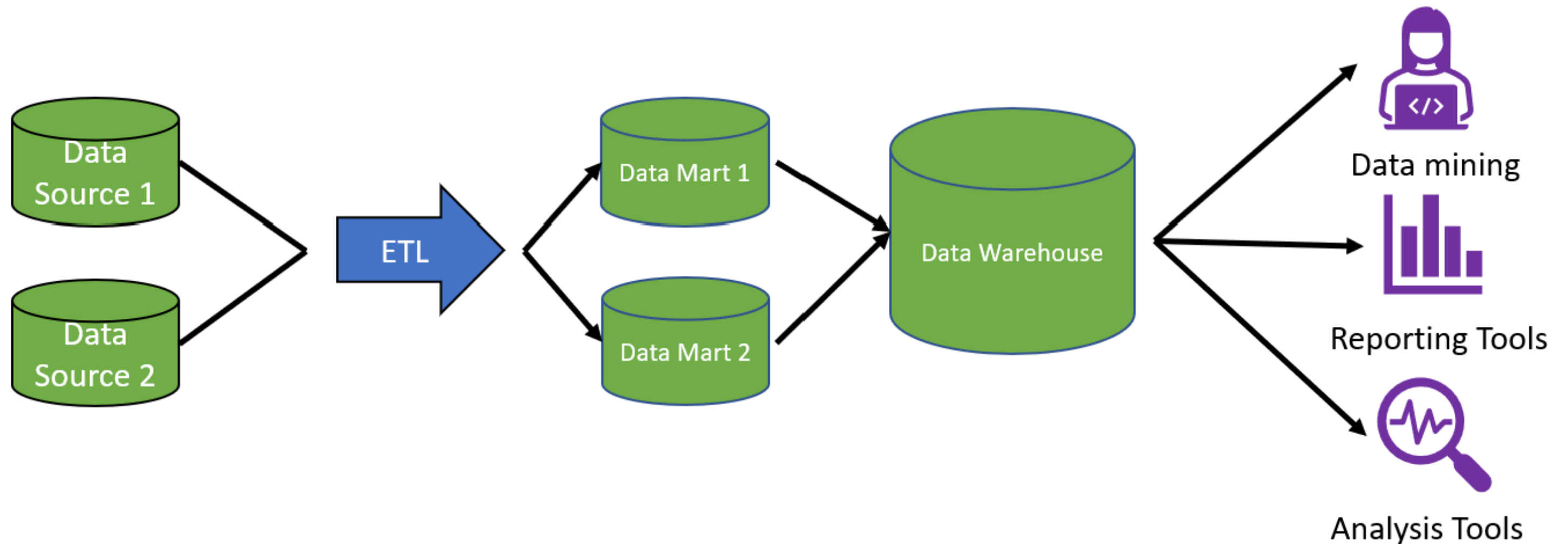


# Kimball - bottom-up

- Denormalizes data
- Focus on departmental data mart
- Data moves directly from ETL to data marts



# Kimball - bottom-up



# Pros and cons of bottom-up

## Advantages:

- Upfront development speed
  - Lower startup cost
- Denormalized = user friendly

## Disadvantages:

- Increased ETL processing time
- Greater possibility of duplicate data
- Ongoing development needed

**Let's practice!**  
DATA WAREHOUSING CONCEPTS

# OLAP and OLTP systems

DATA WAREHOUSING CONCEPTS



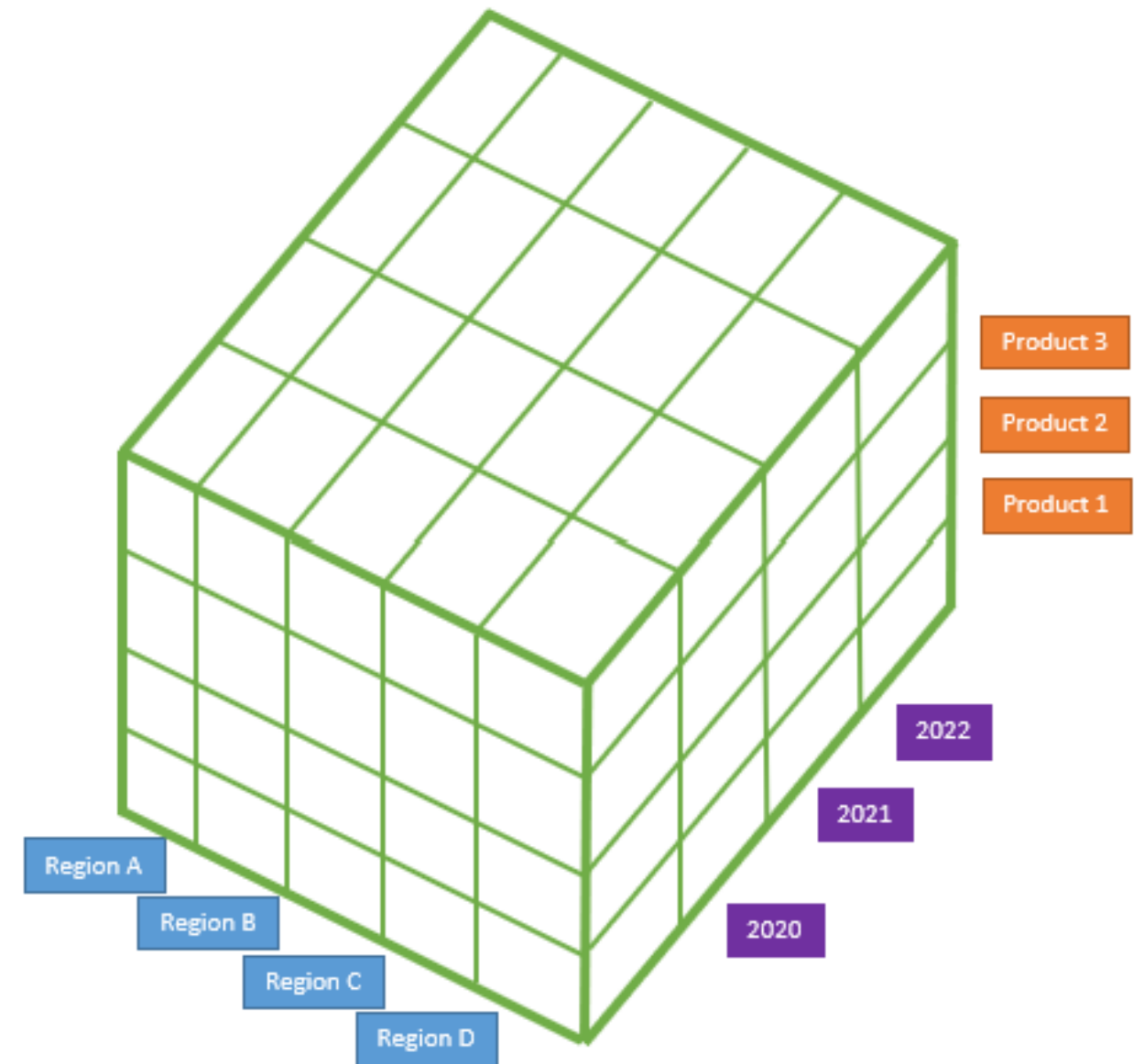
**Aaren Stubberfield**  
Data Scientist

# OLAP systems

- OLAP (*online analytical processing*)
- Designed to support analysis of large amounts of data
- Example dimensional organization:
  - country, state, city
  - years, months, days
- OLAP reorganizes data into multidimensional format

# OLAP cube

- OLAP cube key to OLAP system
- Faster processing vs. traditional relational databases
- Hypercubes have more than three dimensions



# OLTP

- OLTP (*online transaction processing*)
- Designed for processing simple database queries
- Used in source systems to data warehouse



<sup>1</sup> Photo by Rodnae-Productions on Pexels



# Example for a credit card company

## OLTP:

- System tracks customer's purchase
- Processes large amounts of simple database updates to account balances

## OLAP:

- Designed for analyzing purchase data
- Data organized by multiple dimensions



# Summary

Differences	OLAP	OLTP
Optimization:	Complex read-only queries for analysis	Simple queries
Data Representation:	Multidimensional	Rows and columns

**Let's practice!**  
DATA WAREHOUSING CONCEPTS