Lecture 18:

- Looker Studio / Traditional Data Warehouse Concepts

Announcements:

- HW-3 out
- Exam 1 Friday
Basic Terminology used in Looker Platform

Based on Traditional Data Warehouse Concepts ...

Measures: values of the “facts” or “observations” in a data warehouse
  • e.g., number of items sold and purchase price of each transaction
  • the attributes that would go in an SQL aggregate

Dimensions: provide context for the facts
  • e.g., customer, product, location, time
  • attributes that would go in a GROUP BY clause

Metrics in Looker are measures ... where “metrics measure dimensions”
Basic Terminology used in Looker Platform

**Star Schema**: A traditional approach for structuring a warehouse

- Fact table, with foreign keys to dimension tables
- One dimension table per fact-table dimension attribute
- e.g., location dimension with store, city, state, region, country attributes

![Diagram of Star Schema](image)

Simple “textbook” example
**Basic Terminology used in Looker Platform**

**Snowflake Schema**: Normalized dimension tables

- Fact table, with foreign keys to dimension tables (one per dimension att)
- e.g., fact table references product table (via product id)
- product table has id and product name
- but also references to brand and category tables

---

**Simple “textbook” example**
Basic Terminology used in Looker Platform

**Roll Up**: aggregating

- aggregating at a coarser-grained level (along dimensions)
- e.g., average sales by city to average sales by state

**Drill Down**: disaggregating

- going from more aggregated to less aggregated (along dimensions)
- in Looker Studio, typically means going to actual data

**Slicing**: subsetting by filtering on one dimension

- e.g., looking only at one region

**Dicing**: subsetting by filtering on two or more dimensions

- e.g., only looking at specific regions and product types
Basic Terminology used in Looker Platform

**Pivot**: “rotate” tables around a column
- similar to a group by, but where each group becomes a column
- e.g., average sales by state (where each state is an attribute)

**Looks**: stand-alone visualization ... Looker specific
- what are created in the Explorer section of Looker Studio
- corresponding SQL automatically generated

**Dashboard**: Also referred to as a **Report**
- contain one or more visualizations
- support selections, filters, etc., across visualizations
A Maxim:
“Any successful data analysis system will (eventually be asked to) support SQL”

Plan for the middle 3rd of the semester ... 

- High-level architectures for scalable data (query) processing
- Including “traditional” query processing basics (review, for newer archs)
- Popular file formats for data analytics (column-based, hybrid)
- Spark (Spark SQL, Dataproc) and Dremel (BigQuery) architectures

... then move on to Data Mining / ML approaches and data pipelines (last 3rd)

(*) Including MapReduce ... e.g., Apache Hive and Pig
Summary – Things to Know

- Fact tables, Measures vs Dimensions
- Star and Snowflake schemas and differences
- Drill down, Roll up (Drill up), Slice, Dice
- Pivot tables