# Data Warehousing
## CS199 - ACC
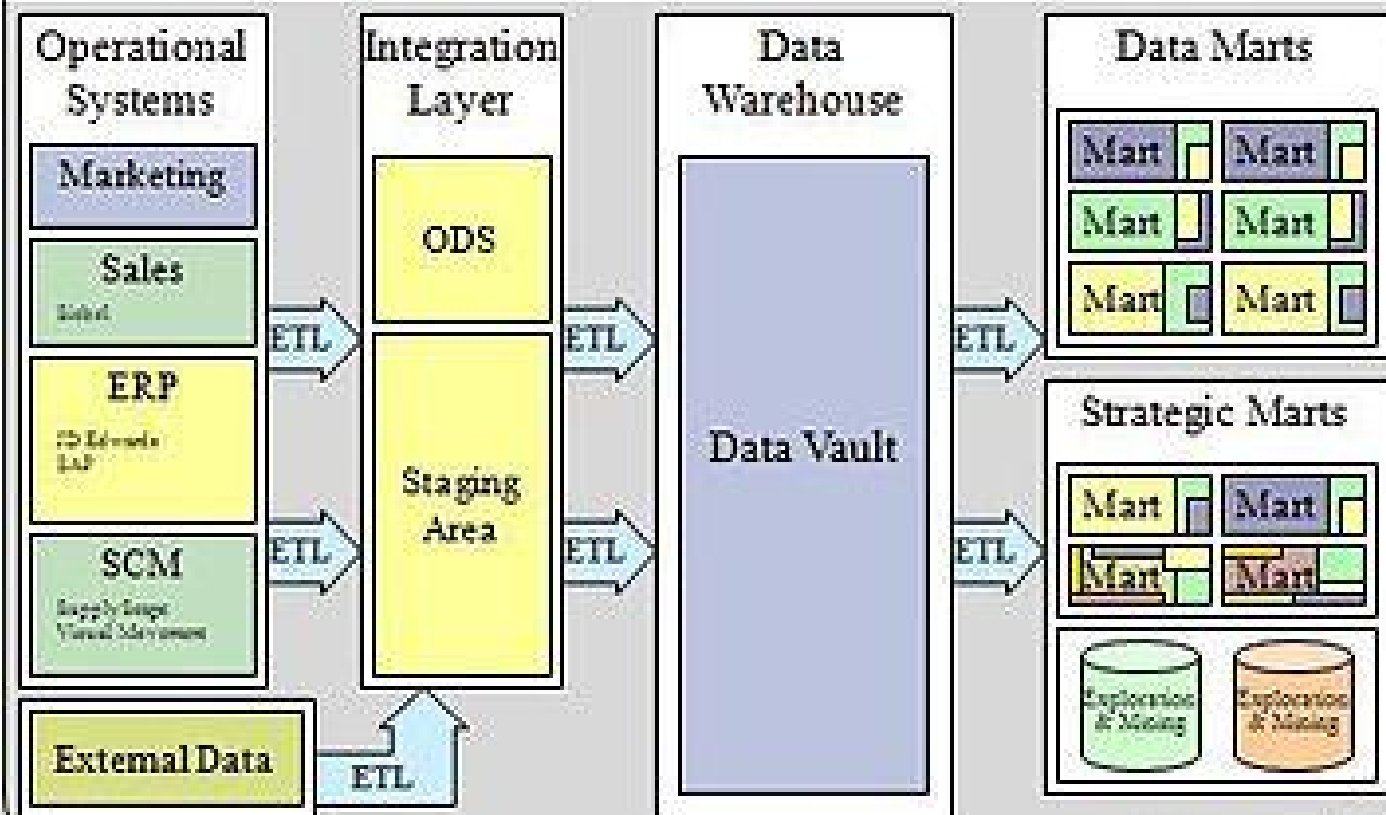
Prof. Robert J. Brunner

Ben Congdon
Tyler Kim
Bhuvan Venkatesh

# Why?

- To store and aggregate large sums of data in suitable format for problem-solving and decision-making

- To get data quickly

# Data Warehouse

| Operational Systems | Integration Layer | Data Warehouse | Data Marts |
|---|---|---|---|

**Operational Systems**
- Marketing
- Sales
- ERP
- SCM
- External Data

**Integration Layer**
- ODS
- Staging Area

**Data Warehouse**
- Data Vault

**Data Marts**
- Mart, Mart, Mart, Mart, Mart, Mart

**Strategic Marts**
- Mart, Mart, Mart, Mart

ETL

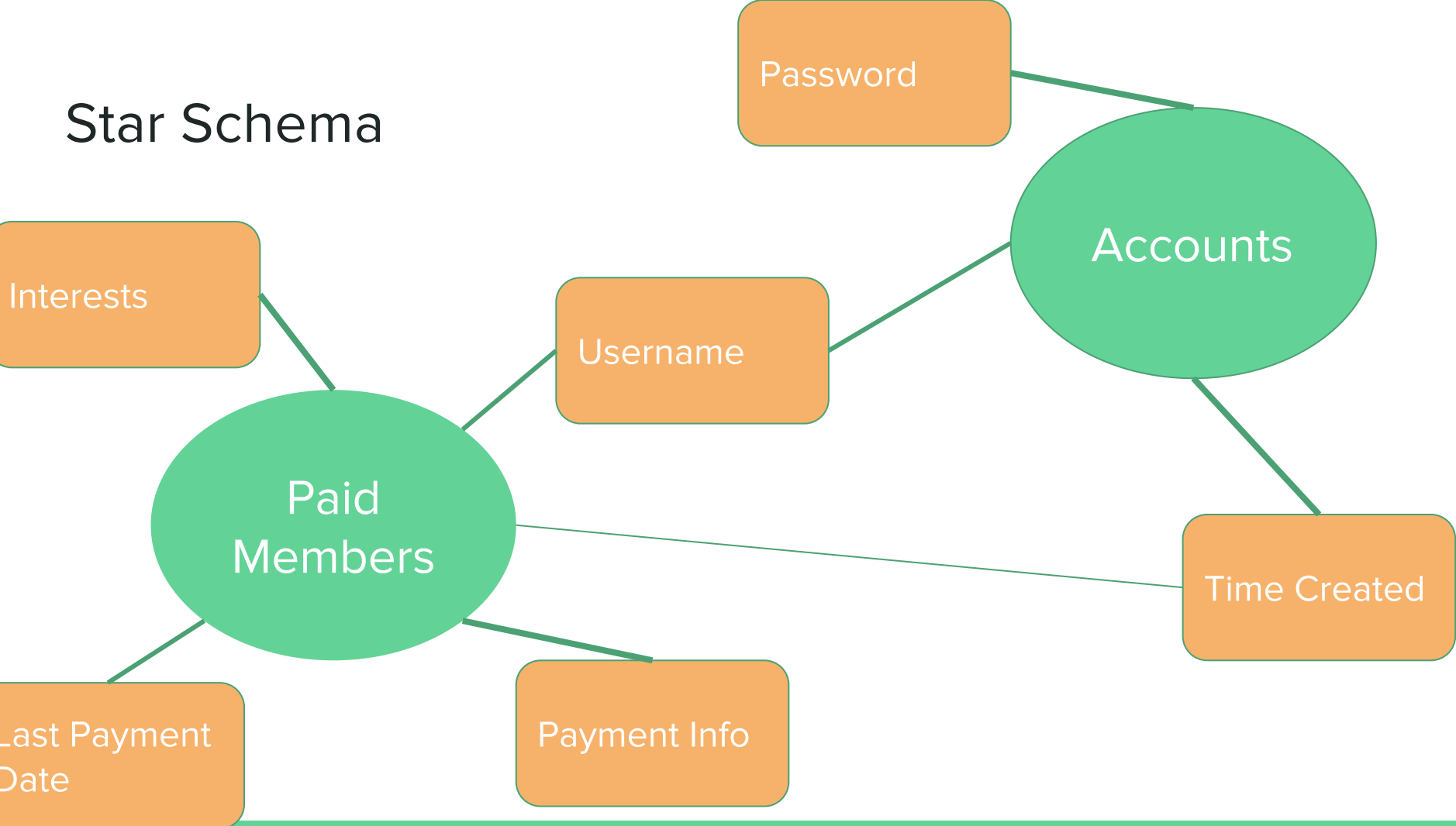# Three Stages in Data Warehouse

- ## Extraction

    - Get the data

- ## Transformation

    - Make it useful

- ## Loading

    - Save it to the warehouse

# Data Marts

Subsets of the Data Warehouse

- Independent Data Pools
  - Separations of queries
- Simpler for different audience/users
- Make it easier to solve smaller problems

Star Schema

# Google BigQuery

- **Managing Data**
    - Can create/delete based on a JSON-encoded schema
    - Import data encoded as CSV or JSON from Google Storage
- **Query**
    - Standard SQL Dialect - returns JSON
    - Max length result of 128 MB or unlimited - large query results enabled
- **Integration**
    - Can be used from Google Apps / Spreadsheets, or any languages with REST API / client libraries

s e v e r l e s s