Designing for Analytics
Facts, Dimensions and the star schema

Robert Hatton
Email: rob@convitali.com
Website: www.convitali.com

Designing for analytics

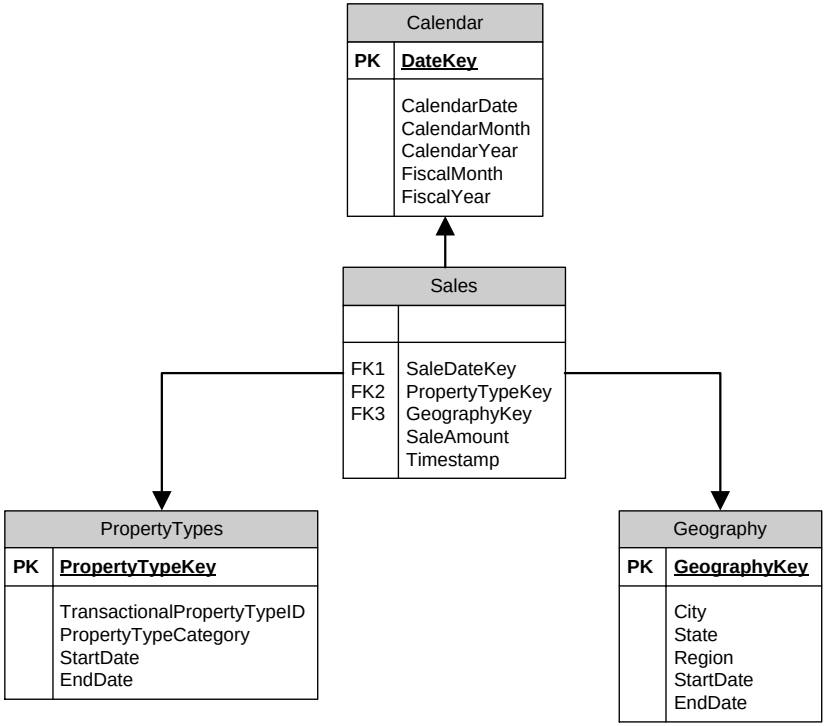
- Analytic and Transactional design are different!
  - Formal normalization isn't always best for analytics
  - Analytics often use hierarchies
  - Record size is important for analytics
  - Two kinds of analytic data
    - Facts (aka. Measures)
    - Dimensions

Facts & Dimensions

- The foundation for analytical design
- The intellectual knife to carve up complex data
- Provide a framework of organizing data
  - Star schema
  - Slowly changing dimensions
    - Source systems often don't track history

Designing for Analytics

Star Schema



Dimensions

- Define how data can be analyzed
  - Think ***fact*** by ***dimension***
    - Sales by year
    - Defects by factory
    - Grades by school
- Set up dimensions first
  - Everything else is either a fact or junk
  - Dimensions without matching facts are useless!
  - Facts will need foreign keys from dimensions

Dimensions

- Characteristics
  - Descriptive
    - That means they usually aren't numbers!
  - Indicate grouping
- Can be hierarchical
  - Day, month, year
  - Store, city, county, state

Dimensions

- Can house history
  - Slowly Changing Dimension
- Require maintenance
  - When new descriptions are found
  - When descriptions change
- Surrogate ***and*** natural keys needed
  - Surrogate key minimizes space & speeds queries
  - Business recognizes natural keys

Designing for Analytics

Dimensions

- ***Don't over do it!***
  - Folks get lost
  - Cube size increases almost exponentially
  - It's more maintenance
- Consider multiple databases
  - Data Marts
  - Target to audience

Dimensions

- Hierarchies
  - Examples
    - Calendar
    - Geography
    - Organization
- Enable drill down
  - Provide known levels of aggregation
- Single vs. multi table
  - Complexity vs. economy

Facts (aka Measures)

- Characteristics
  - Quantitative
  - Part of a continuum
  - Not generic or descriptive
    - Usually not text
- Requires a matching Dimension!
  - It's junk without a way to describe it
- Documented at point in time
  - Facts don't change
  - Works with Slowly Changing Dimensions

Fact Types

- Aggregate
  - Sales
  - Income Statement
  - Production
- Snapshot
  - Inventory
  - Balance Sheet

Facts &Dimensions

- No totally accurate definition
  - Grey area
    - Textual facts: Acid, neutral, base
  - Perversions
    - Degenerate Dimensions
      - Fact is the description
        - Example: Invoice Number
    - Factless Facts
      - Relationships between Facts & Dimensions
        - Example: Number of students taking a test

Facts &Dimensions

- Virtual Dimensions
  - Enumeration
    - Acid, neutral, base matches 1, 0,-1
  - Bucketization (Descretization)
    - Range of values matches bucket value
    - Created from directly from facts

Facts &Dimensions

- Granularity
  - Level of detail
    - Coarse, fine or atomic
  - Impacts both Facts and Dimensions
  - Can be decreased, but not increased
    - Source systems usually don't maintain history
  - Lean toward fine granularity
    - But don't explode the database

THANKS!

Robert Hatton
Email: rob@convitali.com
Website: www.convitali.com