# 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!
➢Normalization isn't appropriate for analytics
➢Analytics often use hierarchies
➢Record size is important for analytics
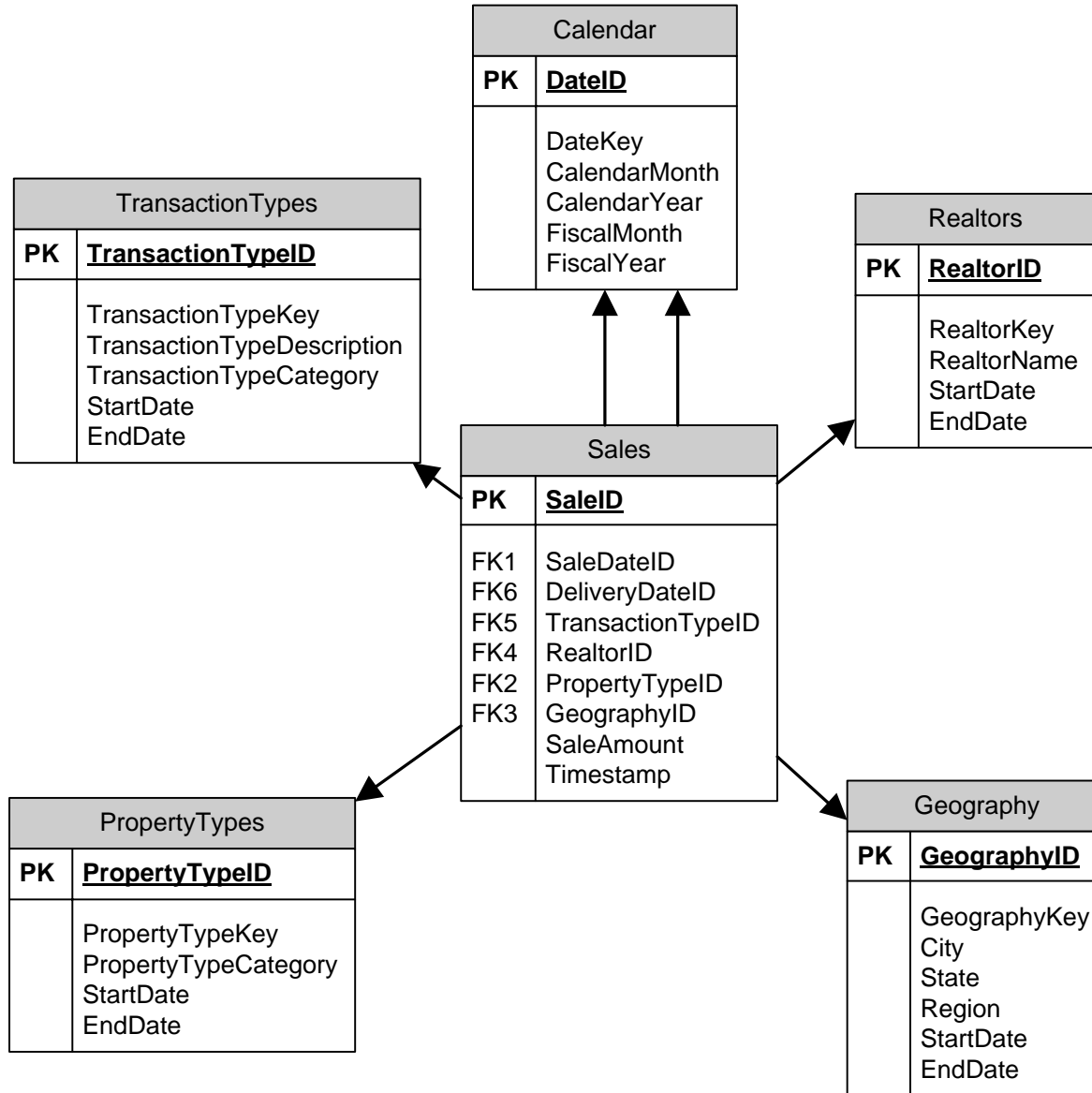➢Two kinds of analytic data
  ➢Facts (aka. Measures)
  ➢Dimensions

# Facts & Dimensions

➢Useful to carve up complex data
➢The foundation for analytical design
   ➢Enables something (Fact) to be described by something else (Dimension)
➢Facts and Dimensions _must_ be used together!
➢Provide a framework of organizing data
   ➢Star schema
➢Support for tracking history
   ➢Slowly changing dimensions
   ➢Source systems often don't track history

# Star Schema

**Calendar**

| PK | **DateID** |
|----|-----------|
|    | DateKey |
|    | CalendarMonth |
|    | CalendarYear |
|    | FiscalMonth |
|    | FiscalYear |

**TransactionTypes**

| PK | **TransactionTypeID** |
|----|----------------------|
|    | TransactionTypeKey |
|    | TransactionTypeDescription |
|    | TransactionTypeCategory |
|    | StartDate |
|    | EndDate |

**Realtors**

| PK | **RealtorID** |
|----|--------------|
|    | RealtorKey |
|    | RealtorName |
|    | StartDate |
|    | EndDate |

**Sales**

| PK | **SaleID** |
|----|-----------|
| FK1 | SaleDateID |
| FK6 | DeliveryDateID |
| FK5 | TransactionTypeID |
| FK4 | RealtorID |
| FK2 | PropertyTypeID |
| FK3 | GeographyID |
|     | SaleAmount |
|     | Timestamp |

**PropertyTypes**

| PK | **PropertyTypeID** |
|----|-------------------|
|    | PropertyTypeKey |
|    | PropertyTypeCategory |
|    | StartDate |
|    | EndDate |

**Geography**

| PK | **GeographyID** |
|----|----------------|
|    | GeographyKey |
|    | City |
|    | State |
|    | Region |
|    | StartDate |
|    | EndDate |

# Dimensions

- **Define how data can be analyzed**
  - Think *fact* by *dimension*
    - Sales by **year**
    - Defects by **factory**
    - Grades by **school**
- **Characteristics**
  - Descriptive (usually aren't numbers)
  - Indicate grouping
- **Can be hierarchical**
  - Day, month, year
  - Store, city, county, state
- **Dimensions without matching facts are useless!**
- **Facts will need foreign keys from dimensions**

# Dimension Maintenance

- **Dimensions can change over time**
- **Can house history**
  - Slowly Changing Dimension
- **Dimensions require maintenance**
  - When new descriptions are found
  - When descriptions change
- **Surrogate *and* natural keys needed**
  - Business recognizes natural keys
  - Surrogate key enables tracking changes
- ***Don't over do it!***
  - Folks get lost
  - Cube size increases almost exponentially
  - It's more maintenance

# Dimension Structure

➤ **Hierarchical**
    ➤ Dimensions can be organized like an outline
    ➤ A single dimension can have several hierarchies!
➤ **Examples**
    ➤ Geography
    ➤ Organization
    ➤ Calendar
        ➤ Gregorian
        ➤ Fiscal
        ➤ 454
➤ **Enable drill down**
    ➤ Provide known levels of aggregation

# 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**
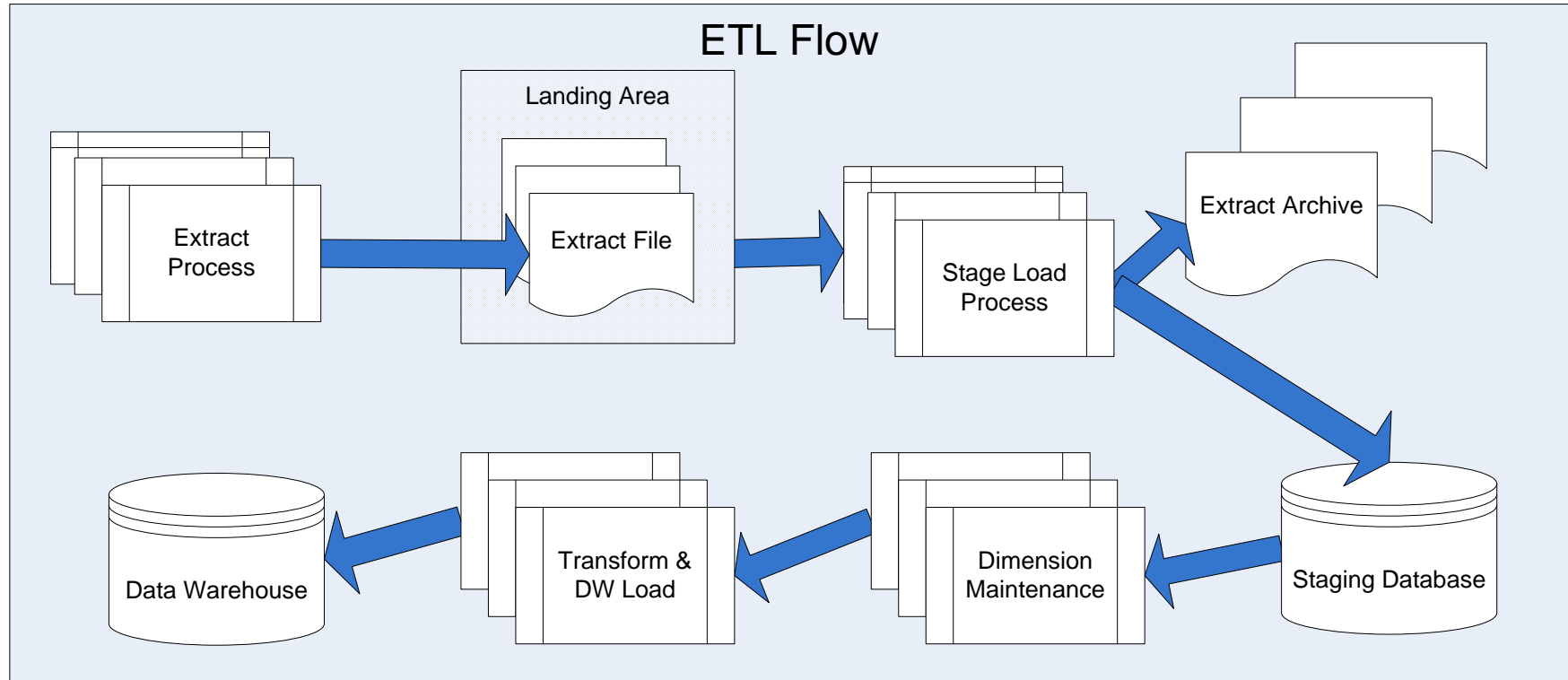  - Inventory
  - Balance Sheet

# Facts & Dimensions

- **No totally accurate definition**
- **Grey area**
  - Textual facts: Acid, neutral, base
- **Perversions**
  - Degenerate Dimensions
    - Fact is the description (Invoice Number)
  - Factless Facts
    - Relationships between Facts & Dimensions
      - Example: Number of students taking a test
- **Granularity**
  - Level of detail
    - Coarse, fine or atomic
  - Impacts both Facts and Dimensions
  - Can be decreased, but not increased
  - Lean toward fine granularity
    - But don't explode the database

# Extract Transform Load

ETL Flow

Landing Area

Extract Process → Extract File → Stage Load Process → Extract Archive

Stage Load Process → Staging Database

Staging Database → Dimension Maintenance → Transform & DW Load → Data Warehouse

# Building Dimensions From Lookup Tables

➢ **Create the dimension tables**
  ➢ Include Business Key and Surrogate key
  ➢ Type 1 or Type 2 Slowly Changing Dimension
➢ **Update translation table**
➢ **Populate dimension from source lookup table**
  ➢ Only fills bottom level of hierarchy
➢ **Update hierarchy**
  ➢ Fill in the upper levels of hierarchy

# Free Form Dimension Sources

## Source Facts

| Table name: salesTransactions | | | | |
|---|---|---|---|---|
| Property Type | Zip | Sale Date | Price | **Notes** |
| 101 | 34654 | 2/12/2012 | $50348 | **cash** |
| 103 | 92704 | 2/18/2012 | $187930 | **Financed purchase** |
| 109 | 60130 | 2/23/2012 | $1948392 | **Cash purchase** |
| 101 | 60122 | 2/23/2012 | $62956 | **Lease** |
| 103 | 32654 | 2/25/2012 | $38723 | **cash** |

## Destination Dimension

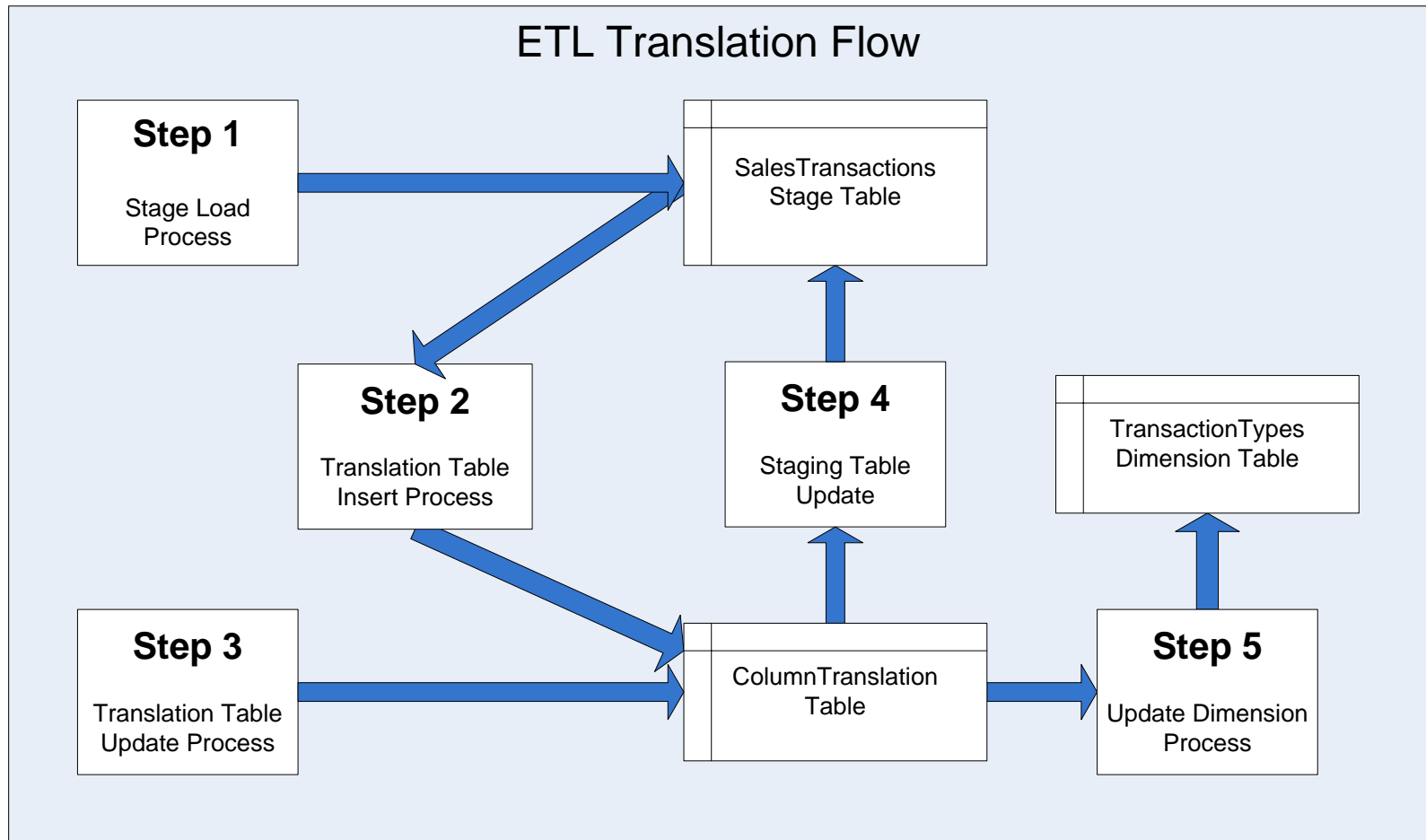| Table name: transactionTypes | | |
|---|---|---|
| TransactionTypeKey | *transactionType* | transactionCategory |
| 1 | *Cash* | Non-financed |
| 2 | *Mortgage* | financed |
| 3 | *Lease* | financed |

# Building Dimensions From
# Free-Form Columns

- **Create the dimension tables**
  - Include Business Key and Surrogate key
  - Type 1 or Type 2 Slowly Changing Dimension
- *Update translation table*
- **Populate dimension from the *translation* table**
  - Only fills bottom level of hierarchy
- *Use translation table to update Fact table*
  - *Change free-form descriptions to dimension members*
- **Update hierarchy**
  - Fill in the upper levels of hierarchy

# ETL Translation Flow



ETL Translation Flow

**Step 1**

Stage Load Process

SalesTransactions Stage Table

**Step 2**

Translation Table Insert Process

**Step 4**

Staging Table Update

TransactionTypes Dimension Table

**Step 3**

Translation Table Update Process

ColumnTranslation Table

**Step 5**

Update Dimension Process

# Transformation and Load

**Source Transaction Table**

| Transaction Type | Property Type | Geography | Realtor | SaleDate | Price |
|---|---|---|---|---|---|
| Cash | SF | 34654 | RJH | 2/12/2012 | $50348 |
| Mortgage | MF | 92704 | NAS | 2/18/2012 | $187930 |
| Lease | FP | 60073 | ELK | 2/23/2012 | $1948392 |

**Destination Fact Table**

| Transaction TypeID | Property TypeID | Geography ID | Realtor ID | SaleDate ID | Price |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 20120212 | $50348 |
| 2 | 2 | 2 | 2 | 20120218 | $187930 |
| 3 | 3 | 3 | 3 | 20120223 | $1948392 |

# THANKS!

Robert Hatton

Email: rob@convitali.com

Blog: www.convitali.com/BI