A Perfect Ten

The Data Model
Thank you to our Sponsors

Global Partner
Microsoft Azure
PASS

Gold Sponsors
Dallas DBAs
Profisee
vmware

Silver Sponsors
IDERA
SentryOne
Main Street DBAs
ProcureSQL
melissa

Bronze Sponsor
SQL Grease
Consultant Corner

Blog Sponsor
SQLCoop
Leslie Andrews

▪ Database Architect @ City of Albuquerque Police Department
▪ Database Architect @ Genesis Healthcare
▪ Implementation Specialist/Trainer @ epaCube
▪ Software Developer/Manager @ Bernalillo County Metropolitan Court

▪ Frequent SQL Saturday Speaker
▪ Speaker Idol @ Summit 2018
▪ Idera ACE Class of 2019

@landrews5807  www.lace.technology  leslie@lace.technology
Today’s Objectives

Comparison of Third Normal Form and Dimensional Models
Leave with ability to look at databases critically
The Perfect Ten
Winner
2018 Westminster Best In Show

Zsa-zsa
Winner
2018 World’s Ugliest Dog Contest

The Right Model for the Job

A Relational Model: Third Normal Form

- Edgar Codd
  - The relational model for database management is an approach to managing data using a structure and language consistent with first-order predicate logic where all data is represented in terms of tuples, grouped into relations. A database organized in terms of the relational model is a relational database.
The Right Model for the Job

A Relational Model: Third Normal Form

- OLTP – Online Transactional Processing
- Operational Systems that do the Business of the Organization
- Many Create/Read/Update/Delete operations completed quickly
The Right Model for the Job

A Dimensional Model: Star Schema

- Ralph Kimball
  - Dimensional modeling includes a set of methods, techniques and concepts for use in data warehouse design. The approach focuses on identifying the key business processes within a business and modelling and implementing these first before adding additional business processes, a bottom-up approach.

https://en.wikipedia.org/wiki/Dimensional_modeling
The Right Model for the Job

A Dimensional Model: Star Schema

- OLAP – Online Analytical Processing
- Low volume of transactions
- Reporting of aggregates
- Support cube analysis
  - Aka “Slicing & Dicing”
The Right Model for the Job

Other Models

▪ Graph (SQL 2017)
  ▪ Nodes & Edges to define relationships
▪ Document
  ▪ i.e. Lotus Notes
▪ Entity-Attribute-Value
Creating a Perfect 10
Database normalization is the process of restructuring data in order to reduce data redundancy and improve data integrity.

**Normal Forms**

- **1st**
  - Ensure that the values in each column of a table are atomic.

- **2nd**
  - All the non-key attributes must be dependent on the whole key.

- **3rd**
  - The table should not contain transitive dependencies.

Normalized Questionnaire Model
null
Many To Many
How To Demo
**Reporting Department**

**Case Number:** 0000000000

**Crash Date (MM/DD/YY):**

**Military Time:**

**City Occurred In:**

**Location:**

- **Feet:**
- **Miles:**

**Crash Occurred On:**

- **Roadway:**
- **Off Roadway:**

**Crash Classification:**

- **Overturned:**
- **Other N-Cr:**
- **Pedestrian:**
- **Other Vehicle:**
- **Vehicle on Other Rdwy:**
- **Parked Vehicle:**
- **Vehicle Object:**
- **Driver Object:**

**Analysis Code:**

**Vehicle No. 1**

<table>
<thead>
<tr>
<th>Drivers Full Name (Last, First, Middle)</th>
<th>Address</th>
<th>Driver’s License Number</th>
<th>State</th>
<th>Type</th>
<th>Status</th>
<th>Restrictions</th>
<th>Endorsements</th>
<th>Expires</th>
<th>Interlock</th>
<th>Chilblite</th>
<th>Zip Code</th>
<th>Phone</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Date of Birth (MM/DD/RR)</th>
<th>Occupation</th>
<th>Seat Pos.</th>
<th>Age</th>
<th>Sex</th>
<th>Race</th>
<th>Injuries</th>
<th>Off Code</th>
<th>On Code</th>
<th>Field Intox</th>
<th>Acting Intox</th>
<th>Ejected</th>
<th>Ejected</th>
<th>Elapsed</th>
<th>Mid Trans</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Occupant’s Name (Last, First, Middle)</th>
<th>Occupant’s Address (City, State, Zip)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Vehicle Yr.</th>
<th>Vehicle Make</th>
<th>Color</th>
<th>Body Style</th>
<th>Cargo Body Type</th>
<th>Vehicle Use (1)</th>
<th>Vehicle Use (2)</th>
<th>Towed?</th>
<th>Towed to</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>DOT #</th>
<th>Intake Center Code</th>
<th>Towed By</th>
<th>Towed To</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Number of Axles</th>
<th>Vehicle Weight Rating</th>
<th>Gross Combination Weight Rating</th>
<th>Hazard</th>
<th>Hazard Name</th>
<th>AND</th>
<th>1 digit #</th>
<th>Hazmat Released?</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Carrier’s Name</th>
<th>Carrier’s Address</th>
<th>Carrier’s Zip</th>
<th>Owner’s Name</th>
<th>Owner’s Company Name</th>
<th>Owner’s Address</th>
<th>Owner’s Zip</th>
<th>Owner’s Telephone</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Insured By: (Name of Company)</th>
<th>Policy Number</th>
<th>Trailer or Towed Vehicles (1)</th>
<th>Type</th>
<th>Year</th>
<th>Make</th>
<th>License Yr.</th>
<th>License State</th>
<th>License Number</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Trailer or Towed Vehicles (2)</th>
<th>Type</th>
<th>Year</th>
<th>Make</th>
<th>License Yr.</th>
<th>License State</th>
<th>License Number</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Trailer or Towed Vehicles (3)</th>
<th>Type</th>
<th>Year</th>
<th>Make</th>
<th>License Yr.</th>
<th>License State</th>
<th>License Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash Report Number</td>
<td>0000000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case Number</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### APPARENT CONTRIBUTING FACTORS

**Event**
- Excessive Speed
- Speed too fast for conditions
- Failed to yield right of way
- Passed stop sign
- disregarded traffic signal
- driveway left of center
- Improper overtaking
- Avoided no contact vehicle
- Avoided no contact other
- Cell Phone
- Texting
- Low Visibility due to smoke
- High speed pursuit

**Drivew/ Pedcyclist Physical Cond.**

<table>
<thead>
<tr>
<th>Q1</th>
<th>D1</th>
<th>D2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Driver/Pedestrian Action**
- At Intersection
  - P1
  - P2
  - P1 P2
  - With Signal
  - Against Signal
  - No Signal
  - Crossing
  - From egress
  - Obstruction
  - Standing
  - No Crosswalk
  - Walking Against Traffic
  - Impatience
  - Uncontrolled
  - Other
  - Walking Across
  - Obstruction
- Not at Intersection
  - P1
  - P2
  - P1 P2
  - walk
  - against
  - others
  - playing in road

**Driver/Pedestrian Surity (Check 1 or more each with X)**
- Consumed Alcohol
- Consumed a Controlled Substance
- Had Not Consumed Alcohol
- Sobriety Unknown
- Consumed Medication
- Tested by Instrument
- Breath Test Administered
- Blood Test Administered
- Standard Field Sobriety Test Administered
- Refused Test

**Event**
- Going straight
- Overtaking/Pausing
- Right Turn
- Left Turn
- U Turn
- Start from parking
- Slowing
- Backing
- Other

**Road Number**
- 1 Lane
- 2 Lanes
- 3 Lanes
- 4 + Lanes
- Undivided
- Unpaved
- Physical Divider
- Painted
- Other
- Other

**Road Design**
- Level
- Hike
- On Grade
- Dip
- Step
- Way Stop
- Heaters
- No Controls
- Other

**Road Surface**
- Paved Center & Edgeline
- Impaired
- Field Sign
- Traffic Signals
- Stop Sign
- No Passing Zone
- R.S. Gate
- Way Stop
- Heaters
- No Controls
- Other

**Road Cond.**
- Dry
- Wet
- Snow
- Ice
- Loose Material
- Paved

**Lighting**
- Daylight
- Dawn
- Dusk
- Dark
- Other
- Not Stated
Identify the Entities

- Accident/Incident/Event
- People
- Vehicles
- Trailers
Identify the Relationships

- One to Many Vehicles per Accident
  - Property as Other Involved
- Zero to Many Witnesses per Accident
- One Owner per Vehicle
- Zero to One Driver per Vehicle
  - Pedestrian/Cyclist as Other Involved
- Zero to Many Occupants per Vehicle
- Zero to Many Towed/Trailers per Vehicle
Basic Model
The dimensional model separates business process data into fact tables, which hold the measurable, quantitative data about a business, and dimensions which are descriptive attributes related to fact data.

# Building the Dimensions

## Bus Matrix

<table>
<thead>
<tr>
<th>BUSINESS PROCESSES</th>
<th>COMMON DIMENSIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Date</td>
</tr>
<tr>
<td>Issue Purchase Orders</td>
<td>X</td>
</tr>
<tr>
<td>Receive Warehouse Deliveries</td>
<td>X</td>
</tr>
<tr>
<td>Warehouse Inventory</td>
<td>X</td>
</tr>
<tr>
<td>Receive Store Deliveries</td>
<td>X</td>
</tr>
<tr>
<td>Store Inventory</td>
<td>X</td>
</tr>
<tr>
<td>Retail Sales</td>
<td>X</td>
</tr>
<tr>
<td>Retail Sales Forecast</td>
<td>X</td>
</tr>
<tr>
<td>Retail Promotion Tracking</td>
<td>X</td>
</tr>
<tr>
<td>Customer Returns</td>
<td>X</td>
</tr>
<tr>
<td>Returns to Vendor</td>
<td>X</td>
</tr>
<tr>
<td>Frequent Shopper Sign-Ups</td>
<td>X</td>
</tr>
</tbody>
</table>
Nothing but the Facts

Grain

- Defines what level of detail is observed for a particular event
- Multiple Fact tables for the same information at a different Grain
Nothing but the Facts

- Foreign Keys to Dimensions
- Measures
  - There can be multiple measures in a single fact table
- Types of Measures
  - Additive
  - Semi-Additive
  - Non-Additive
Dimensions

- Slowly Changing
  - Depending on the business requirement, should an attribute’s history of changes be preserved in the data warehouse?
    - Type 1
    - Type 2

- Rapidly Changing
Dimensions

- **Degenerate**
  - Attribute is stored in the Fact Table, not a separate dimension.

- **Role Playing**
  - Same dimension key joined to multiple fields in the fact table

- **Shrunken**
  - Aka Snowflake – subset of another dimension

- **Static**
  - Attributes not extracted from the data source
Dimensions

- Junk
  - Combination of Unrelated attributes
- Inferred
  - A surrogate key when dimension record not ready
- Conformed
  - A dimension used in multiple locations
Star Schema

Never more than a single join away from dimensions – all the attributes in very wide tables
How To Demo

State of New Mexico
Uniform Crash Report
## Filling a Dimension

<table>
<thead>
<tr>
<th>ID</th>
<th>Towed</th>
<th>Due to disabling</th>
<th>Severity</th>
<th>Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td>Yes</td>
<td>Yes</td>
<td>Heavy</td>
<td>Disabled</td>
</tr>
<tr>
<td>1002</td>
<td>Yes</td>
<td>No</td>
<td>Heavy</td>
<td>Disabled</td>
</tr>
<tr>
<td>1003</td>
<td>No</td>
<td>Yes</td>
<td>Heavy</td>
<td>Disabled</td>
</tr>
<tr>
<td>1004</td>
<td>No</td>
<td>No</td>
<td>Heavy</td>
<td>Disable</td>
</tr>
<tr>
<td>1005</td>
<td>Yes</td>
<td>Yes</td>
<td>Heavy</td>
<td>Functional</td>
</tr>
<tr>
<td>1006</td>
<td>Yes</td>
<td>No</td>
<td>Heavy</td>
<td>Functional</td>
</tr>
<tr>
<td>1007</td>
<td>No</td>
<td>Yes</td>
<td>Heavy</td>
<td>Functional</td>
</tr>
<tr>
<td>1008</td>
<td>No</td>
<td>No</td>
<td>Heavy</td>
<td>Functional</td>
</tr>
</tbody>
</table>
Possible Dimensions

- Vehicle
  - VIN, Year, Model, Make, Color, Body Style, License Plate State Abbreviation, License Plate State Name, License Plate Number, Plate Expiration Date, Number of Axles, Weight Rating, HazMat Placard 4 digit Number, HazMat Name, HazMat 1 digit Number, HazMat Released, Posted Speed, Safe Speed, Left Scene of Crash, Direction Vehicle Headed, Weather, Road Condition, Road Character, Grade of Road
Possible Dimensions

- Accident Details (Junk Dimension)
  - Incident Number/Crash Report Number, Case Number, CAD Number, NMDOT Number, Crash Classification, Hit and Run, Fatal Injury, On Private Property, Lighting, Weather

- Occupant
  - Person Name, Drivers License State Abbreviation, Drivers License State Name, Drivers License Number, Type Of License, License Restrictions, License Endorsement, License Expiration Date, Interlock License, Date of Birth, Occupation, Seat Position
Using a Third Normal Form Model for an Analysis Process
Bad Modelling that leads to incorrect aggregates

Zsa-zsa of Star Schema?
A Perfect Ten
Your Take Away – Determine if the databases you work with were designed using the correct model