Dimensional Modeling 101

Presented by: Michael Davis

CEO OmegaSoft,LLC

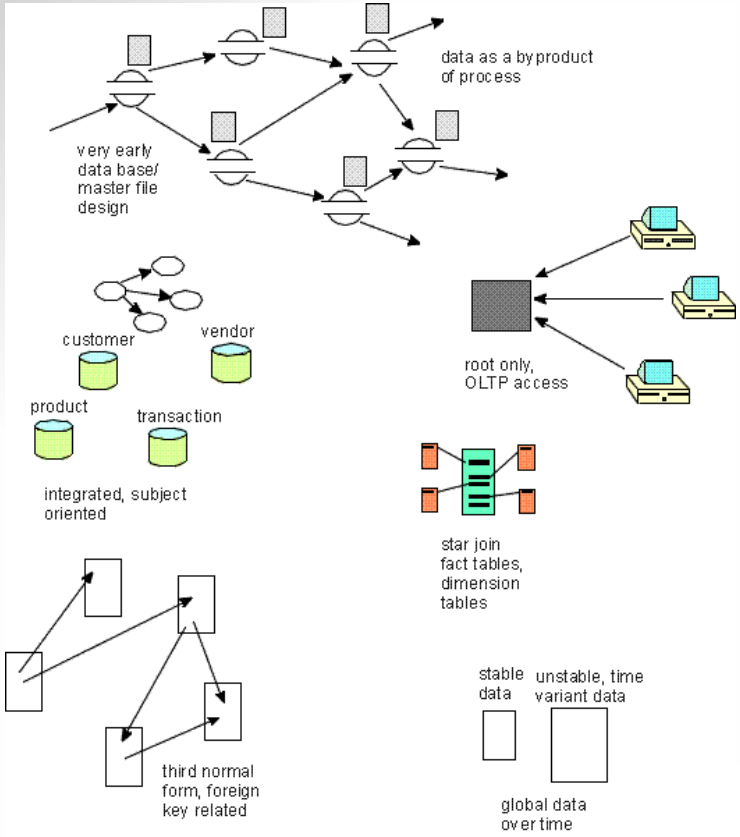
Agenda

- ***Brief history of Database Design***
- ***Dimension Modeling Terminology***
- ***Case study overview***
- ***4 step Dimensional Modeling Process***
- ***Additional Data Warehousing concepts***
  - Slowly changing dimensions
  - Snowflaking
- ***Question & Answer***

Brief history of Database Design

- ***Evolution of Database design***
  - By-Product of a Function
  - Subject Oriented
- ***Modern Database design***
  - Relational approach (Ted Codd)
  - Data Warehousing design

Approaches to Database Design



Different approaches to data base design over the years.

Figure 1

Dimensional Modeling vs. Normalized Modeling

- ***Normalized database modeling***
  - Anomalies
    - Insertion Anomaly
    - Deletion Anomaly
    - Update Anomaly
  - Performance Issues in a Data Warehouse
- ***Dimensional modeling***
  - Benefits
    - Easy to Comprehend
    - Query Performance

Dimensional Design Process

- ***4 Step Dimensional Modeling Process***
  - Identify the business process
  - Declare the grain
  - Define the dimensional tables
  - Define the fact table

Case Study

- ***Overview Case study***
  - Adaptation of a 3-D science fiction feature film that is now playing in theaters
  - 3-D rain forest populated with extremely tall blue people and three- legged mammals
  - Box office record
  - Nominated for an Academy award
  - Unattianium
  - 12 stores, divided in 4 regions, 100 products
  - Tribal managers are interested in Profitability of rain forest stores.

Dimensional modeling terminology

- ***What is a Grain ?***
  - The granularity of the data stored in a Data Warehouse
- ***What is a Fact table ?***
  - The integral table in a Dimensional model. A fact table contains measurements or facts.
- ***What is a Dimension table ?***
  - Supporting tables in a Dimensional model. Contains textual information.
- ***What is a Surrogate key ?***
  - A meaningless sequential value assigned to each row of a Dimension table.

We are about to begin…

- ***3-D Glasses***
- ***Your featured film is brought to you by ….***

Theater Rules

- ***Please turn your Cell phones Off***
- ***No talking (audience)***
- ***Questions after presentation***

Step 1: Identify the business process to model

- ***A business process is a collection of related, structured activities or tasks that produce a specific service or product for a particular customer or customers.***
- ***Examples***
  - Purchasing
  - Sales
  - Marketing
- ***Apply to Case Study***

Step 2: Determine the grain

- ***Key Elements of the Grain***
  - Represents a row in a Fact table
  - Atomic level
  - Aggregate data
- ***Examples***
  - Line item on a sales receipt
  - Bill of Sale
  - line item on a bill
- ***Grain of our Case study***
  - Sales transaction
  - Sales by Day by Product by Store

Step 3: Define the Dimensions

- ***What are Dimensions ?***
  - Descriptive
  - Report headers
  - “Fat” tables
- ***Process of Defining Dimensions***
  - Derived from the grain
  - Using Interrogative pronouns: Who, What, Where, When
- ***Dimensions of our Case study***

Date Dimension

- ***Only Dimension that can be pre-loaded***
- ***Represents a particular day, subdivided in its respective parts***
- ***Contains a row for each day of the period the Data Warehouse is covering***
  - Examples
    - Date Dimension covering 10 years, would have 3,650 rows

Date Dimension

Date Dimension	
PK	**Date SK**
	Date Day of Week Month Year Quarter ....etc

DATE SK	1
DATE	3/9/2009
DAY of Week	Tuesday
MONTH	March
YEAR	2009
QUARTER	First Quarter

Product Dimension

- ***Contains a row for each product in our retail stores***
- ***Master list of Products***

Product Dimension

Product Dimension	
PK	**Product SK**
	Product Description Brand Description Category Description UPC Code Height Weight Color Shape ....etc

Product SK	1
Product Description	Gritty on the surface
Category Description	Clay
UPC Code	12345
Height	4.5
Weight	16
Color	Gray
Shape	Oval

Store Dimension

- ***Contains a row for every store in our rain forest***
- ***Descriptive attributes***

Store Dimension

Store Dimension	
PK	**Store SK**
	Store Name Store Manager Store Street Address Store Region Store County Store Square Footage ....etc

STORE SK	1
STORE MANAGER	Jake S.
STORE STREET ADDRESS	237 Quake Forest Hills
STORE REGION	West Forest
STORE COUNTY	Pandora 1
STORE SQUARE FOOTAGE	10000

Step 4: Define the Facts

- ***Integral part in a data warehouse***
  - Form the center of a Star Schema design
- ***Characteristics***
  - Usually numeric
- ***Types of Facts***
  - Additive
    - Additive facts can be aggregated by addition. i.e., Sales Dollar amount
  - Semi-Additive
    - Semi-additive Facts can be aggregated along some dimensions
  - Non-Additive
    - Non-additive facts cannot be added. i.e., Averages

Sales Transaction Fact table

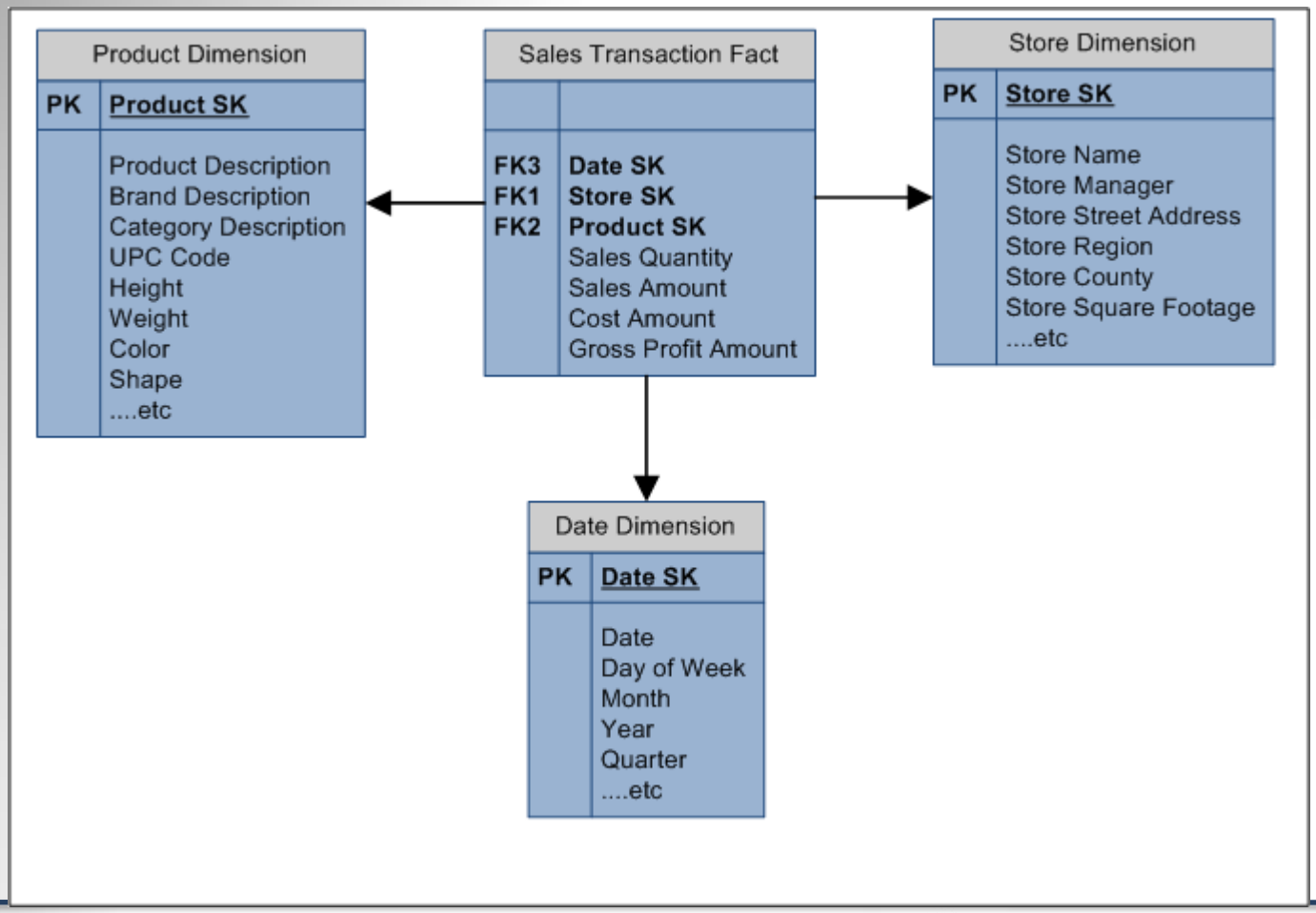
Sales Transaction Fact
Date SK **Store SK** **Product SK** Sales Quantity Sales Amount Cost Amount Gross Profit Amount ....etc

Date SK	1
Store SK	1
Product SK	1
Sales Quantity	10
Sales Amount	750
Cost Amount	250
Gross Profit Amount	500

Fact Tables vs. Dimensional Tables

- ***Fact tables***
  - Narrow in size
    - Small number of columns
  - Contain mostly numeric data
- ***Dimensional tables***
  - Wide in size
    - Many columns
  - Contain textual Information

Putting it all together



Data Model in Action

Date SK	Date	Day of Week	Month	Year	Quarter
1	3/9/2009	Tuesday	March	2009	First Quarter
2	2/9/2009	Monday	March	2009	First Quarter
3	6/15/2009	Monday	June	2009	Second Quarter
4	12/31/2009	Thursday	December	2009	Fourth Quarter
5	1/12/2010	Tuesday	January	2010	First Quarter

Date Dimension

Pro	Product Description	Brand Des	Category I	UPC Code	Height	Weight	Color	Shape
1	Gritty on the surface	Gritty	Clay	12345	4.5	16	Gray	Oval
2	Gritty on the surface	Gritty	Clay	22244	2	9	Dark Gray	Oval
3	Gritty on the surface	Gritty	Clay	22225	2	5	Black	Circle
4	Small pieces of Elements	Loamy	Metallic	55255	8.5	10	Silver	Square
5	Slippery on the surface	Slick	Sandstone	82378	10.5	25	Brown	Diamond

Product Dimension

St	Store N	Store Mar	Store Street Address	Store Region	Store Cou	Store Square Footage
1	Store 1	Jake S.	237 Quake Forest Hills	West Forest	Pandora 1	10,000
2	Store 2	Mike D.	1112 Main Lake Mountain	West Forest	Pandora 1	10,250
3	Store 3	Destiny K.	50 Yellow Lane	East Forest	Pandora 2	10,000
4	Store 4	Laniesha S.	150 Waterfront Avenue	East Forest	Pandora 2	10,000
5	Store 5	Angelia N.	2 Nile River Avenue	North Forest	Pandora 3	25,000

Store Dimension

Data Model in Action

- ***Sales Transaction Fact***

Date SK	Store SK	Product SK	Sales Quantity	Sales Amount	Cost Amount	Gross Profit Amount
1	1	1	10	750	250	500
2	1	5	25	1500	500	1000
5	2	4	75	5000	2500	2500
5	2	3	50	3000	1000	2000
5	4	2	35	1750	750	1000

Total Sales by Store Report

Store	Region	Sales Quantity	Total Sales	Total Profit
Store 1	West Forest	35	$2,250	$1,500
Store 2	**West Forest**	**125**	**$8,000**	**$4,500**
Store 4	East Forest	35	$1,750	$1,000

Data Model in Action

Date SK	Store SK	Product SK	Sales Quantity	Sales Amount	Cost Amount	Gross Profit Amount
1	1	1	10	750	250	500
2	1	5	25	1500	500	1000
5	**2**	**4**	**75**	**5000**	**2500**	**2500**
5	**2**	**3**	**50**	**3000**	**1000**	**2000**
5	4	2	35	1750	750	1000

Total Sales by Region

- ***Drilling up and Drilling down***

Region	Sales Quantity	Total Sales	Total Profit
West Forest	**155**	**$10,250**	**$6,000**
East Forest	35	$1,750	$1,000

Additional Data Warehousing concepts

- ***Slowly changing Dimensions***
- ***Different types***
  - Type 1
    - Attribute is overwritten
  - Type 2
    - A new dimension record with new surrogate key created.
  - Type 3
    - A new column is added to dimension table.
- ***Snowflaking***

The end…

- ***Email Address:***
  - mdavis@omegasoftgroup.com

Questions and Answers

- ***Got Questions ?***

References

- *Paul Lane. (2009, August). Retrieved 2009, from Oracle Database Data Warehousing Guide, 11g Release 2 (11.2): http://www.dba.bg/forum/oracle11g_r2/E11882_01/server.112/e10810.pdf*
- *Ralph Kimball, M. R. (2008). The Data warehouse Lifecycle Toolkit. Indianapolis: Wiley Publishing, Inc.*
- *Ralph Kimball, M. R. (2002). The Data warehouse Toolkit Second Edition. New York: John Wiley and Sons, Inc.*
- *Inmon, W. H. (n.d.). A BRIEF HISTORY OF DATABASE DESIGN. Retrieved February 2010, from Corporate Information Factory: http://www.inmoncif.com/view/36*
- *http://en.wikipedia.org/wiki/Business_process*