



**IIT School of Applied Technology**

ILLINOIS INSTITUTE OF TECHNOLOGY

**information technology & management**

# **526 Data Warehousing**

February 3, 2016

Week 3 Presentation

# Updates

- New folder structure in BB
  - By Week
  - Readings and Other Resources
  - Discussion forum
- Glossary
  - Can be very useful for preparing the final
- TA Walk-In hour
  - Preeti Tambade / [ptambade@hawk.iit.edu](mailto:ptambade@hawk.iit.edu)
  - Every Monday from 11 am - 12 pm, in Perlstein Hall - Room 223

# Week 03 Topic: Dimensional Modeling: Basic Fact Tables Techniques

- We will cover
  - Fact Tables Revisited
  - Three Types of Fact Tables
  - Degenerate Dimensions
  - 4-Step Dimensional Design Process
  - Design Workshop #1

# Fact Tables Revisited

- Contains the **measurements** or facts about a business process
- Stay true to the grain
- Build up from **the lowest grain possible**
- Use **conformed facts** (identical field name for identical technical definition)
- Three types of Fact tables
  - **Transaction Fact**
  - **Periodic Snapshot Fact**
  - **Accumulating Snapshot Fact**

# Fact Tables Revisited (Cont'd)

## : Physical Table Elements

- Contains the following elements
  - **Primary** key – surrogate key: not common (only if ETL tools require)
  - **Foreign keys** to dimensions
  - **Degenerate dimensions**
    - Transaction indicators or flags
  - **Measure** or metrics
    - Transaction amounts

Retail Sales Facts
Date Key (FK)
Product Key (FK)
Store Key (FK)
Promotion Key (FK)
Customer Key (FK)
Clerk Key (FK)
Transaction #
Sales Dollars
Sales Units

# Surrogate Key

- One of the basic elements of data warehouse design
- **Every join** between dimension tables and fact tables **should be based on surrogate keys**, not natural keys
- Essential for Slowly Changing Dimension Type 2
- Essential when no natural key is available
- Improves join performance and saves storage space

# Three Types of Fact Tables:

## Transaction Fact

- The grain corresponds to a measurement taken at **a single instant**
  - e.g. The grocery store beep
- **Unpredictably sparse or dense**
- Can be enormous, with the largest containing many billions of records.

Transaction Grain
Date (FK)
Product (FK)
Store (FK)
Customer (FK)
Cashier (FK)
Manager (FK)
Promotion (FK)
Weather (FK)
Basket (FK)
<i>time of day</i>
<i>quantity</i>
<i>extended price</i>

# Three Types of Fact Tables:

## Periodic Snapshot Fact

- The grain corresponds to **a predefined span of time**
  - e.g. a financial reporting period
- **All of the reporting entities appears in each snapshot**
- Is predictably dense and can get large as well
  - e.g. 2 million accounts for 10 years' of monthly snapshot will generate 2.4 billion fact rows

Periodic Snapshot Grain
Month (FK)
Account (FK)
Branch (FK)
Household (FK)
<i>balance</i>
<i>fees paid</i>
<i>interest earned</i>
<i>transaction count</i>



# Three Types of Fact Tables: Accumulating Snapshot Fact

- The grain corresponds to **well-defined predictable processes**
  - e.g. Order processing and college admissions
- Fact **rows are revisited and overwritten** as the process progresses through its steps from beginning to end
- Is much smaller than the other two types

Accumulating Snapshot Grain
Order Date (FK)
Ship Date (FK)
Delivery Date (FK)
Payment Date (FK)
Return Date (FK)
Warehouse (FK)
Customer (FK)
Promotion (FK)
Order Status (FK)
<i>quantity</i>
<i>extended list price</i>
<i>discounts</i>
<i>extended net price</i>

# Three Types of Fact Tables: Comparisons

	Periodic Snapshot	Transaction	Accumulating Snapshot
Time period represented	Regular predictable intervals	Point in time	Indeterminate time span, typically short lived
Grain	One row per period	One row per transaction event	One row per life
Table loads	Insert	Insert	Insert and update
Row updates	Not revisited	Not revisited	Revisited whenever activity
Date dimension	End-of-period	Transaction date	Multiple dates for standard milestones
Facts	Performance for predefined time interval	Transaction activity	Performance over finite time

# Degenerate Dimension

- A **dimension key** in the fact table that **does not have its own dimension table**, because all the interesting attributes have been placed in analytic dimension
- Does not join to a corresponding dimension
- Becomes a **part of the fact table's primary key**
- Commonly occurs when the fact table's grain is a single line item transaction

Retail Sales Facts
Date Key (FK)
Product Key (FK)
Store Key (FK)
Promotion Key (FK)
Cashier Key (FK)
Payment Method Key (FK)
<b>POS Transaction # (DD)</b>
Sales Quantity
Regular Unit Price
Discount Unit Price
Net Unit Price
Extended Discount Dollar Amount
Extended Sales Dollar Amount
Extended Cost Dollar Amount
Extended Gross Profit Dollar Amount

# The Grain Revisited

- An event that creates a fact record
- **Start at the lowest**, most atomic grain
  - Atomic data is the most expressive data and versatile to unexpected requests
- Avoid **mixed granularity**
- A clear definition of grain provides rich information about dimensions

# 4-Step Dimensional Design Process

## ➤ Identity the Business Process

- Source of “measurements”
- e.g. taking orders, invoicing, receiving payments, handling service calls, etc.

## ➤ Identity the Grain

- What does 1 row in fact table represent/mean?
- Lowest atomic grain delivers most flexibility

## ➤ Identity the Dimensions

- Descriptive context, true to the grain

## ➤ Identify the Facts

- Numeric additive measurements, true to the grain

# 4-Step Dimensional Design Process

## 1. Identity the Business Process

- Source of “measurements”
- e.g. taking orders, invoicing, receiving payments, handling service calls, etc.

## 2. Identity the Grain

- What does 1 row in fact table represent/mean?
- Lowest atomic grain delivers most flexibility

## 3. Identity the Dimensions

- Descriptive context, true to the grain

## 4. Identify the Facts

- Numeric additive measurements, true to the grain

# 4-Step Dimensional Design Process: An Example

## 1. Identify Business Process

- Claim Billing

## 2. Identify the Grain

- A line item of a doctor's bill

## 3. Identify Dimensions

- Date (of treatment)
- Doctor (maybe called “provider”)
- Patient
- Procedure
- Primary Diagnosis

# 4-Step Dimensional Design Process: An Example (Cont'd)

## 3. Identify Dimensions (cont'd)

- Location (presumably the doctor's office)
- Billing Organization (an organization the doctor belongs to)
- Responsible Party (either the patient, or the patient's legal guardian)
- Primary Payer (often an insurance plan)
- Secondary Payer (maybe the responsible party's spouse's insurance plan) and quite possibly others.

## 4. Identify the Facts

- Billed Amount



# Design Workshop #1

Complete the 4-step process for designing dimensional models.

➤ Business Process:

➤ Grain:

➤ Dimensions:

➤ Facts:

K & B Beyond 1100 Dono Street E Subway, CA 44321 (233) 233-1232		
Store: 00921 Cashier: 98889/Julie		
Kelsyus Floating Cooler 1005 26.99-5.00 PROMO	<u>2@21.99</u>	43.98
Petzl E97 PP Tikka Plus 2 Headlamp, 1 1025 39.94-10.00 PROMO	<u>1@29.94</u>	59.88
Light My Fire Spork 4-Pack 1022	<u>1@13.99</u>	13.99
Coleman WeatherMaster 8 Tent 1049 206.99-50.00 PROMO	<u>1@156.99</u>	156.99
TOTAL		274.84
AMOUNT TENDERED CASH		274.84
ITEM COUNT		6
-----		
Transaction: 567		2/3/15 14:21
-----		
Thank hyou for shopping at K & B Beyond		
37475384729288200000		

# Design Workshop #1 (Cont'd)

Draw dimensional star schema

# Week 03 Topic: Dimensional Modeling: Basic Fact Tables Techniques

Questions?