NATIONAL ENGINEERING CENTER
University of the Philippines
Diliman, Quezon City

5.0 Designing Dimension Tables

Eugene Rex L. Jalao, Ph.D.

Associate Professor

Department Industrial Engineering and Operations Research

University of the Philippines Diliman

Module 2 of the Business Intelligence and Analytics Track of UP NEC and the UP Center of Business Intelligence

Outline for This Training

1. Introduction to Data Warehousing
2. DW Lifecycle and Project Management
   - Case Study on DW PM
3. Dimensional Modeling
4. Designing Fact Tables
5. **Designing Dimension Tables**
   - **Case Study on Dimension Modeling**
6. Extraction Transformation and Loading
   - Case Study on ETL Planning
7. Transformation and Loading Methodologies
   - Case Study on ETL

Outline for this Session

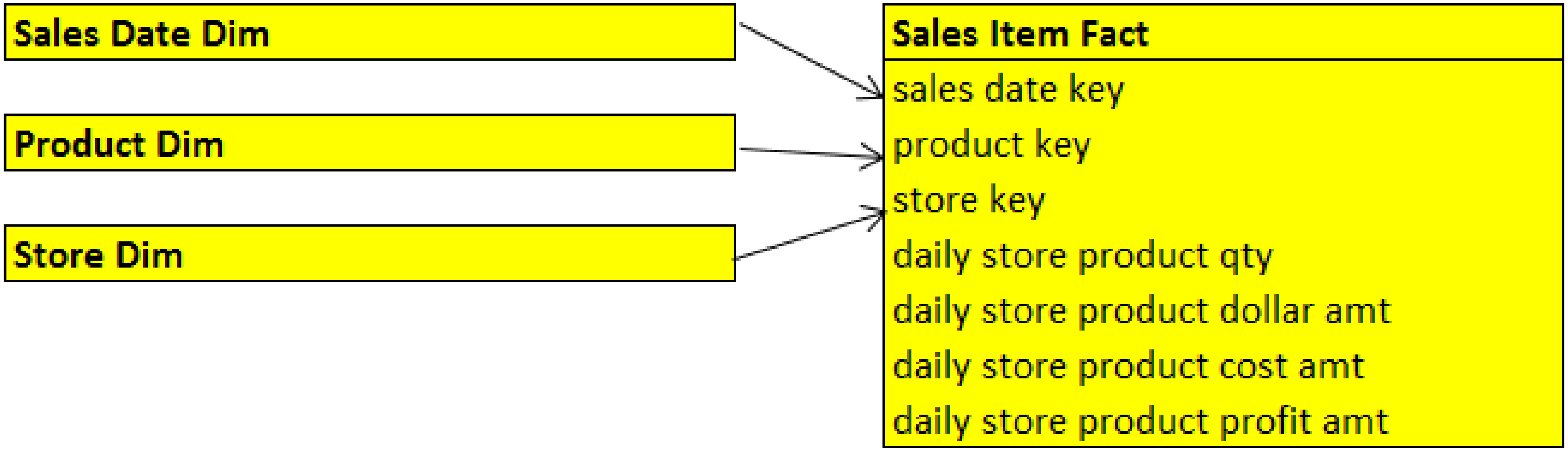
- Introduction to Dimension Tables
- Dimension Table Contents
- Dimension Table Types
- Generating Dimension Models
- Case Example
- Case Study

Dimension Tables

- A dimensional model divides the world into measurements and context.
- Context intuitively divided into clumps called dimensions.
- Dimensions describe the “who, what, where, when, why, and how” of the facts.
- They can also identified as the “by” words in a business question that asks for a report.
  - Example: “I’d like a report that lists sales by store by product by quarter.

Sample Dimension Model

- One row per product per day per store



Sales Date Dim

Product Dim

Store Dim

Sales Item Fact

sales date key
product key
store key
daily store product qty
daily store product dollar amt
daily store product cost amt
daily store product profit amt

Dimension Tables

- Contain the parameters by which the fact table measures are analyzed
  - amount sold is analyzed by day, month, quarter, or year
  - amount sold on sunny days vs. rainy days
  - inventory quantity analyzed by warehouse by product
  - profit analyzed by product, category, department, store, district, or region

Dimension Table Traits

- Provide the context to the fact table measures they describe
- Contain descriptors of the business (nouns)
- Utilize business terminology
- Many large columns
- Contain textual and discrete data
- Are usually smaller than fact tables

Dimension Table Traits

- Have a single column surrogate primary key (called the warehouse dimension key)
- Are joined to a fact table through a foreign key reference to their primary key
- Can contain one or more hierarchies
- The hierarchies are de-normalized into the dimension tables

Generic Dimension and Fact Tables

Dimension Tables

Persons:

- Customer
- Employee
- Supplier
- Competitor
- Applicant
- Alliance

Places:

- Facility/Building
- Demographic Area
- Location

Things:

- Product/Service
- Part/Component
- Material/Supply
- Physical Asset/Equipment

Concepts:

- Organizational Unit
- Job
- G/L Account
- Financial Statement
- Tax Rate Table
- Policy/Procedure
- Law/Regulation
- Contract
- Strategy
- Course
- Campaign/Promotion
- Audit
- Project

Fact Tables

Events:

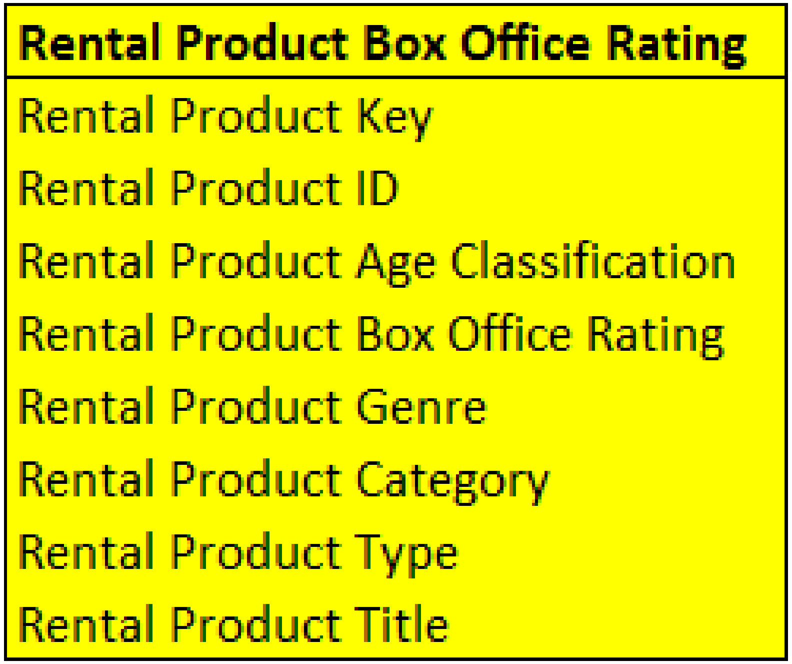
- Customer Order
- Shipping Order
- Customer Invoice
- Requisition
- Purchase Order
- Receipt
- Supplier Invoice
- Disbursement
- Incident
- Work Order
- Test/Inspection

Outline for this Session

- Introduction to Dimension Tables
- **Dimension Table Contents**
- Dimension Table Types
- Generating Dimension Models
- Case Example
- Case Study

The Anatomy of a Dimension

- Which attribute is the primary key?
- Which is the natural key?
- Which are the detail attributes?
- Which are the analytical attributes?
- What hierarchies do you see?

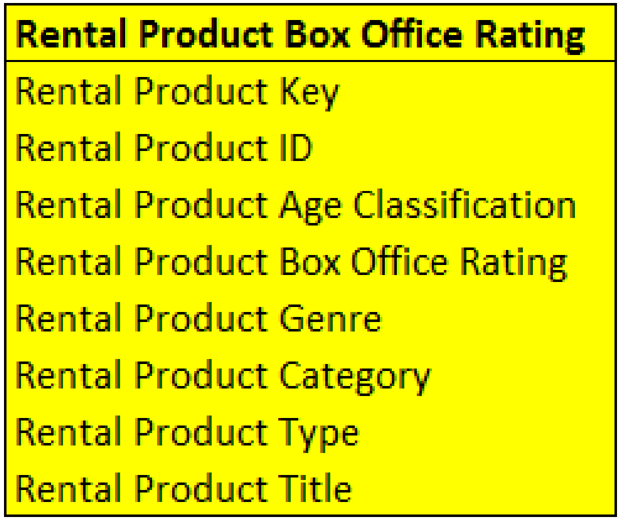


Rental Product Box Office Rating
Rental Product Key
Rental Product ID
Rental Product Age Classification
Rental Product Box Office Rating
Rental Product Genre
Rental Product Category
Rental Product Type
Rental Product Title

Rental Product Dimension Anatomy

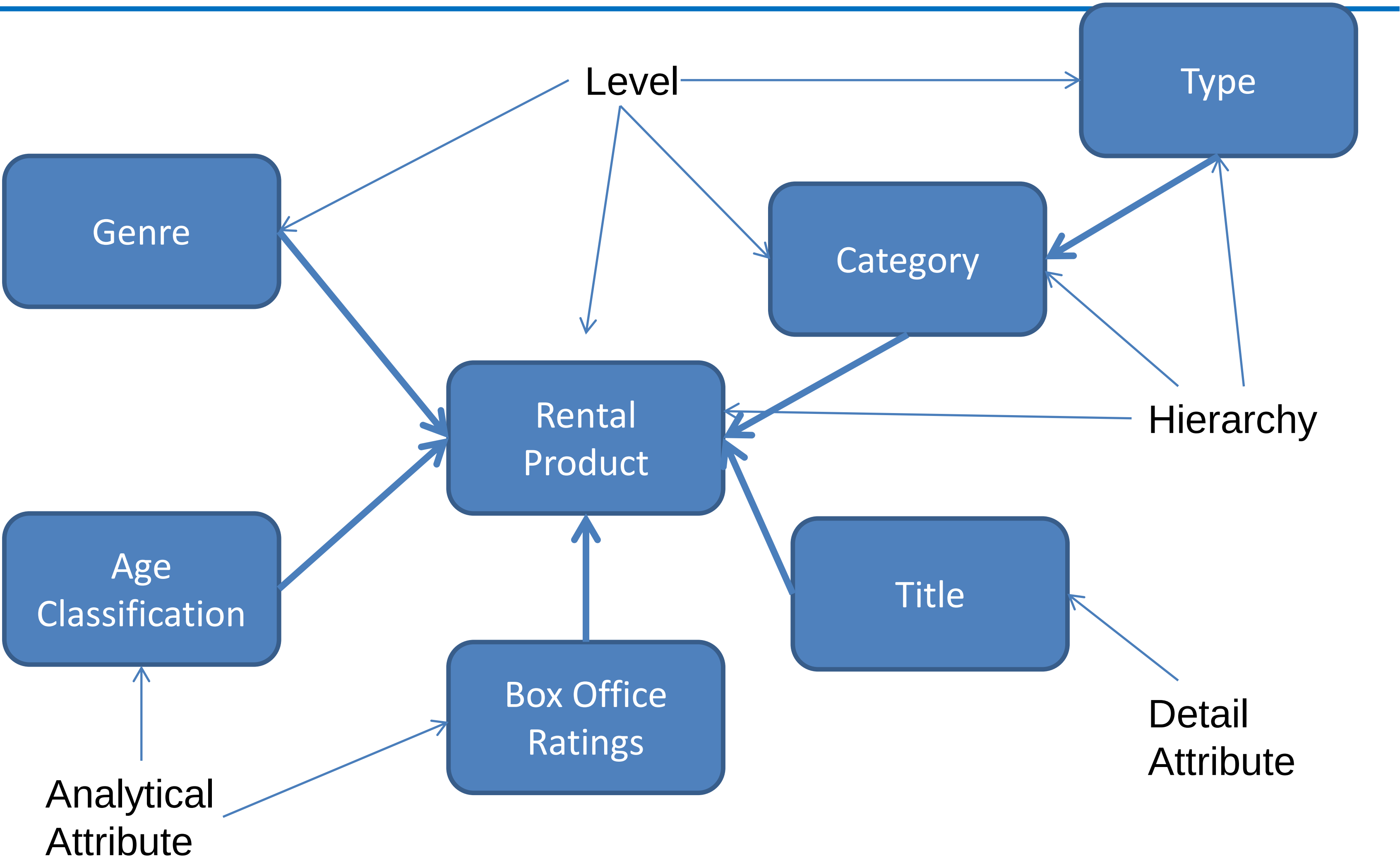
- Primary key
  - Rental Product Key
- Natural key
  - Rental Product ID
- Detail attributes
  - Rental Product ID
  - Rental Product Title
- Hierarchies
  - Product < Category < Type
  - Product < Genre

- Analytical attributes
  - Rental Product Age Classification
  - Rental Product Box Office Rating

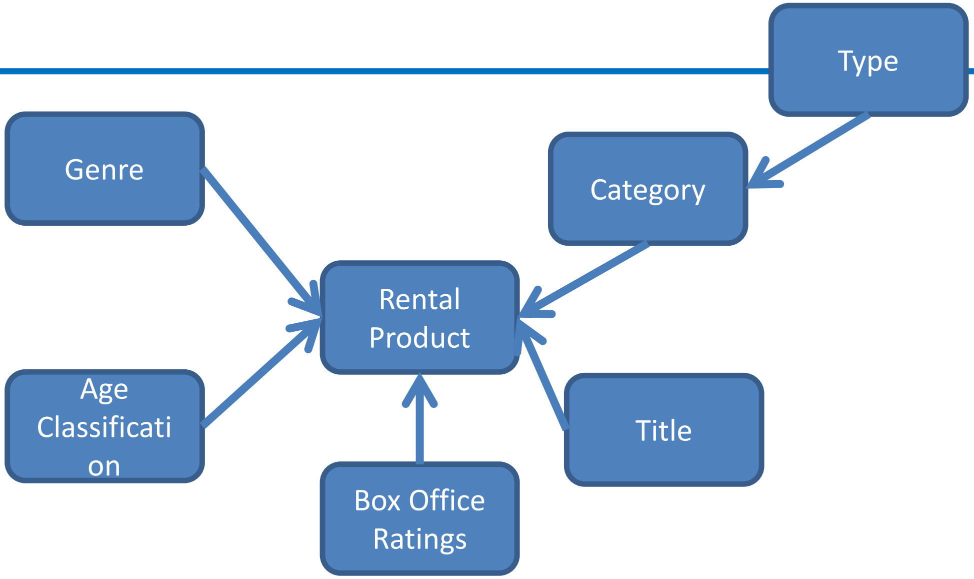


Rental Product Box Office Rating
Rental Product Key
Rental Product ID
Rental Product Age Classification
Rental Product Box Office Rating
Rental Product Genre
Rental Product Category
Rental Product Type
Rental Product Title

Rental Product Dimension



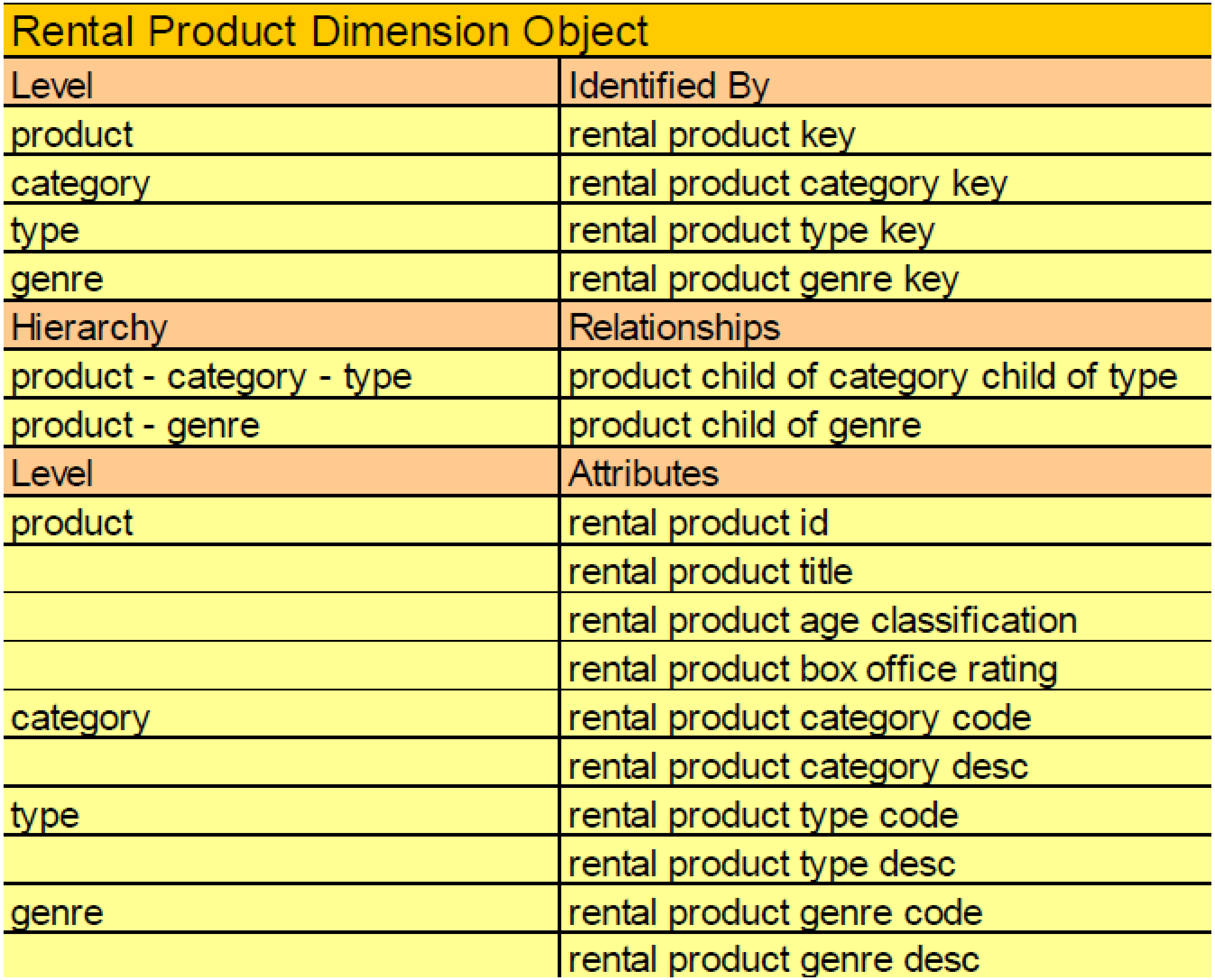
Rental Product Dimension



Rental Product Dimension Table

Column Name	Description	Sample Data
ental product key	data warehouse assigned surrogate key identifying an individual title	1', '2'
ental product id	operational system identifier for an individual title	A2345FGD'
ental product title	film title	The Gods Must be Crazy'
ental product age classification	a classification as to how long the title has been available	New Release', Recent Release', etc.
ental product box office rating	a classification as to how well the title did in movie theaters	Super Hit', 'Hit', Above Avg', etc.
ental product category key	data warehouse assigned surrogate key for product categories	1', '2'
ental product category code	product category code assigned by operational system	VHS', 'DVD', 'PS', 'NI'
ental product category desc	product category description	Video Tape', DVD Disk', 'Play Station', 'Nintendo'
ental product type key	data warehouse assigned surrogate key for product types	1', '2'
ental product type code	product type code from operational system	MV', 'GM'
ental product type desc	product type description	Movie', Game'
ental product genre key	data warehouse assigned surrogate key for product genres	1', '2'
ental product genre code	product genre code	CM', 'DR', 'AD'
ental product genre desc	product genre description	Comedy', 'Drama', 'Adventure Game'

Dimension Object Definition



Rental Product Dimension Object	
Level	Identified By
product	rental product key
category	rental product category key
type	rental product type key
genre	rental product genre key
Hierarchy	Relationships
product - category - type	product child of category child of type
product - genre	product child of genre
Level	Attributes
product	rental product id
	rental product title
	rental product age classification
	rental product box office rating
category	rental product category code
	rental product category desc
type	rental product type code
	rental product type desc
genre	rental product genre code
	rental product genre desc

What is a surrogate key?

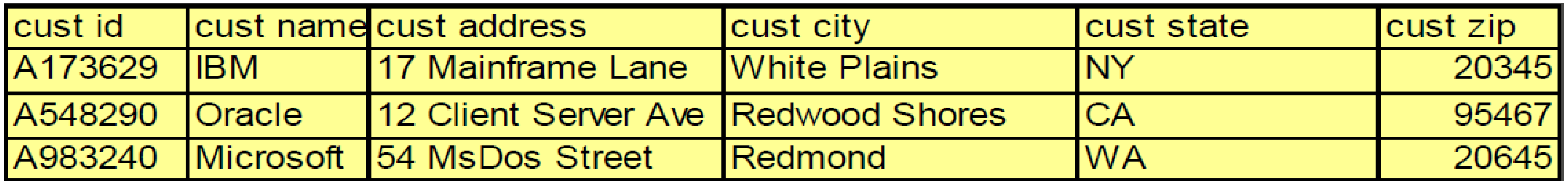
- A surrogate key is a system assigned primary key.
- When the first row is added to a dimension, the system automatically assigns a key of 1 to the row.
- As each additional row is added, the system automatically increments the key by 1.
- It’s meaningless, but essential as a foreign key in fact tables
- Important: Retain source system primary key as unique identifier to use as lookup argument during ETL process and for report headers

Warehouse Dimension Keys

- Single column surrogate keys
  - Provide key control within the data warehouse
  - Substantially improve performance
  - Enable one method of tracking attribute history
  - Facilitate exception references from a fact table
- Implemented in every dimension, even date and time dimensions

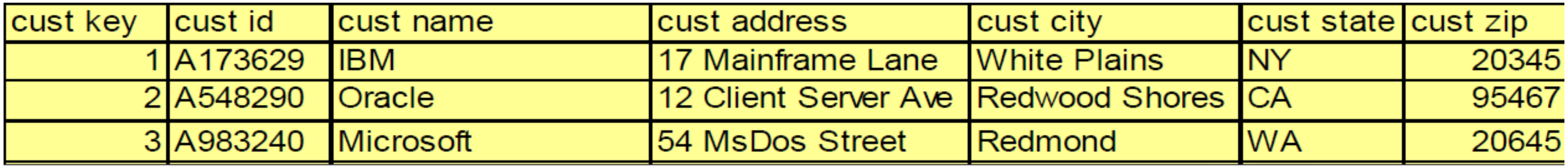
Example

- Legacy Data



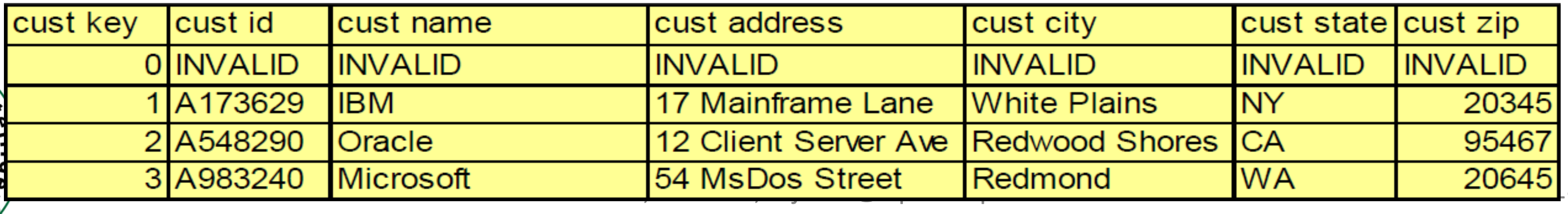
cust id	cust name	cust address	cust city	cust state	cust zip
A173629	IBM	17 Mainframe Lane	White Plains	NY	20345
A548290	Oracle	12 Client Server Ave	Redwood Shores	CA	95467
A983240	Microsoft	54 MsDos Street	Redmond	WA	20645

- Within Data Warehouse, Data Marts, Analytical Application Stores



cust key	cust id	cust name	cust address	cust city	cust state	cust zip
1	A173629	IBM	17 Mainframe Lane	White Plains	NY	20345
2	A548290	Oracle	12 Client Server Ave	Redwood Shores	CA	95467
3	A983240	Microsoft	54 MsDos Street	Redmond	WA	20645

- Invalid Rows

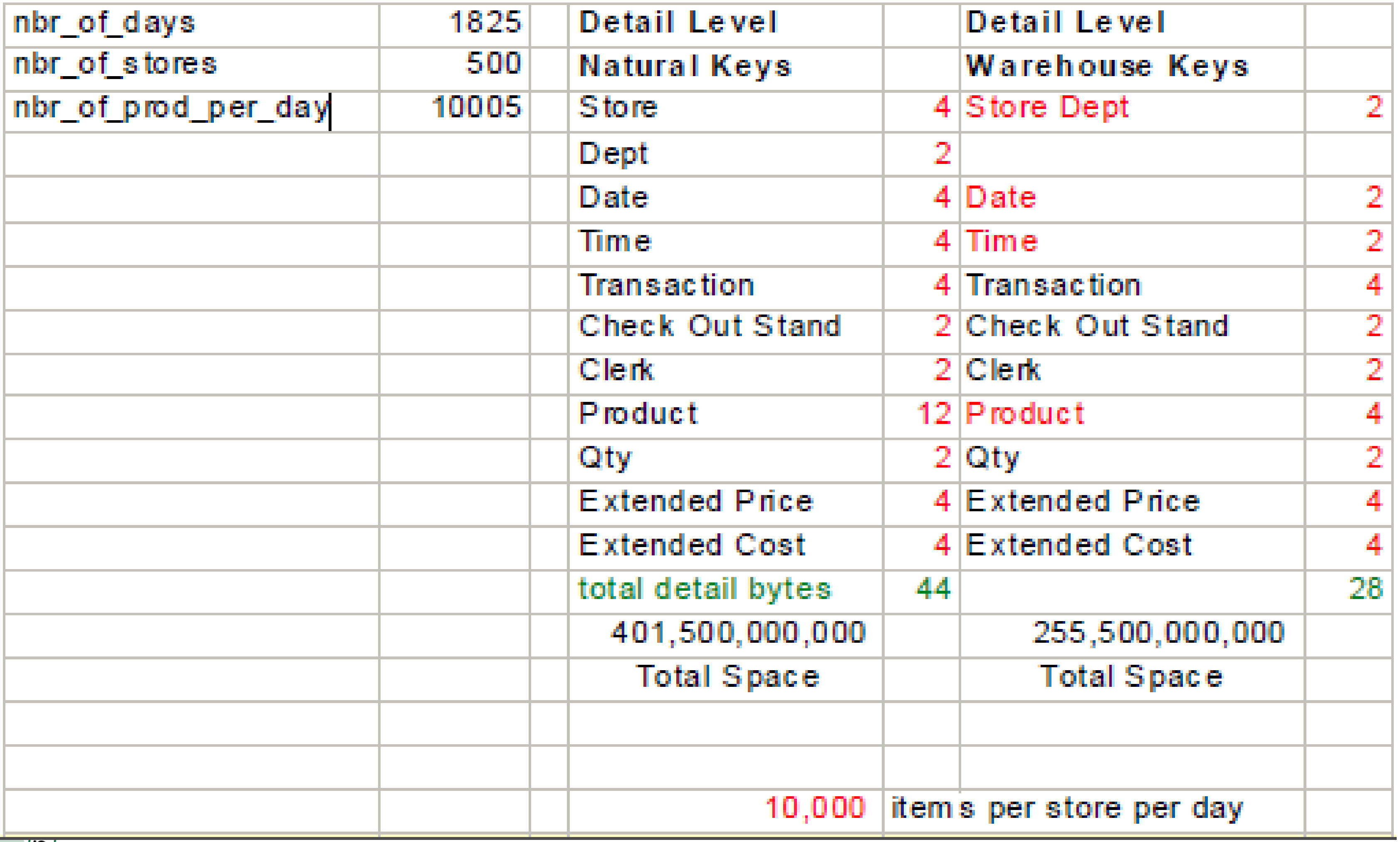


cust key	cust id	cust name	cust address	cust city	cust state	cust zip
0	INVALID	INVALID	INVALID	INVALID	INVALID	INVALID
1	A173629	IBM	17 Mainframe Lane	White Plains	NY	20345
2	A548290	Oracle	12 Client Server Ave	Redwood Shores	CA	95467
3	A983240	Microsoft	54 MsDos Street	Redmond	WA	20645

Benefits of Surrogate Keys

- Provide Key Control
  - Maintain dimension key control from within the Data Warehouse environment
    - Isolation from the operational system
    - Strategic vs. Operational perspective
- Substantially Improve Performance using a single column primary key
- These keys are the foreign key references which are carried in the fact tables
- Substantially reducing fact table sizes

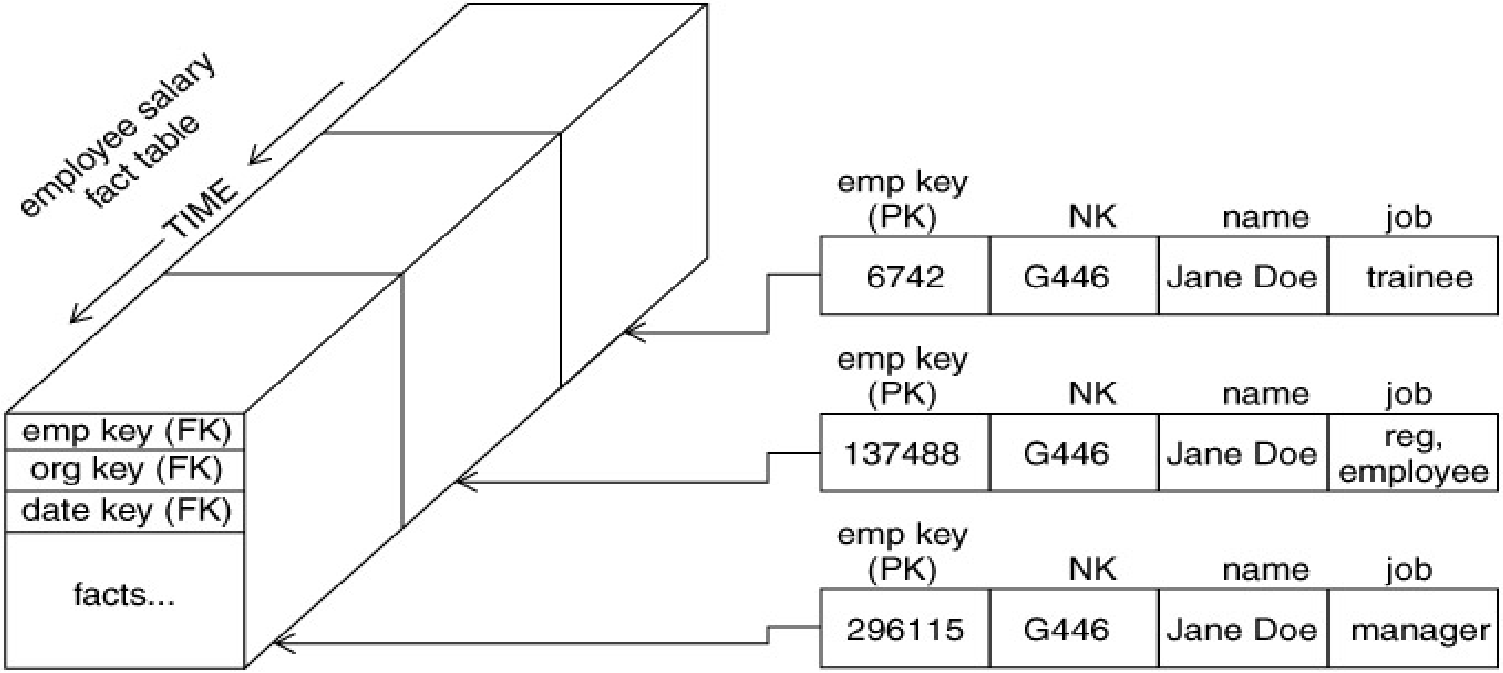
Warehouse Key Usage – Data Warehouse Size Effect



nbr_of_days	1825		Detail Level		Detail Level	
nbr_of_stores	500		Natural Keys		Warehouse Keys	
nbr_of_prod_per_day	10005		Store	4	Store Dept	2
			Dept	2		
			Date	4	Date	2
			Time	4	Time	2
			Transaction	4	Transaction	4
			Check Out Stand	2	Check Out Stand	2
			Clerk	2	Clerk	2
			Product	12	Product	4
			Qty	2	Qty	2
			Extended Price	4	Extended Price	4
			Extended Cost	4	Extended Cost	4
			total detail bytes	44		28
			401,500,000,000		255,500,000,000	
			Total Space		Total Space	
			10,000	items per store per day		

Track Attribute History

- Enable one method of tracking dimension attribute changes
  - Type 2 – Slowly Changing Dimension
    - Not to be used for all dimension t



emp key (PK)	NK	name	job
6742	G446	Jane Doe	trainee

emp key (PK)	NK	name	job
137488	G446	Jane Doe	reg, employee

emp key (PK)	NK	name	job
296115	G446	Jane Doe	manager

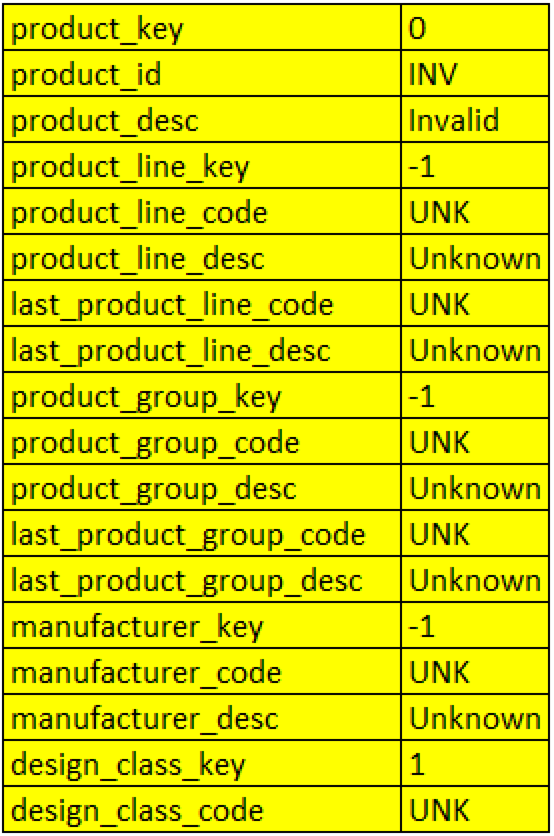
Exception Condition Dimension Table Rows

- Indicate that the row in the fact table referenced an exception condition
- 0 – the fact table row had an invalid legacy id for this dimension (Invalid)
- -1 – The fact table row should reference a value for this dimension, but the value is unknown (Missing Mandatory)
- -2 – The fact table row is not applicable for this dimension (Missing Optional)

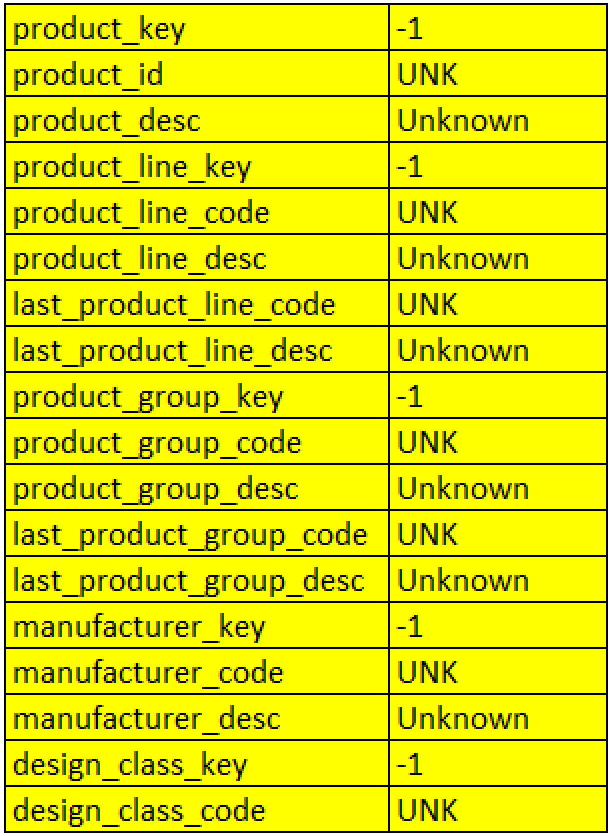
Examples

- Invalid reference from fact table
  - The sale of a product whose product ID is not in the dimension table
- Unknown reference from fact table
  - The sale of a product whose product ID is missing
- The fact table row is not applicable to this dimension
  - The sale of a product that is not on promotion

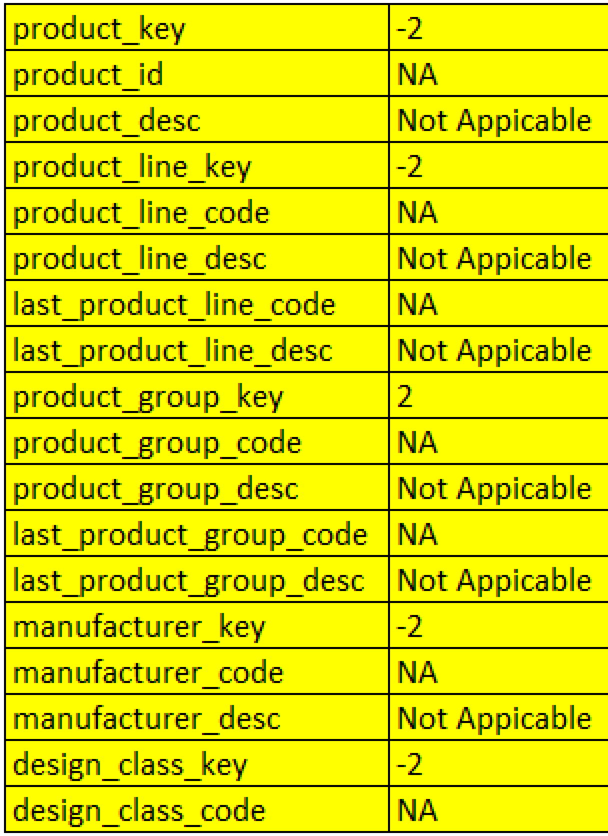
Default Dimension Rows



product_key	0
product_id	INV
product_desc	Invalid
product_line_key	-1
product_line_code	UNK
product_line_desc	Unknown
last_product_line_code	UNK
last_product_line_desc	Unknown
product_group_key	-1
product_group_code	UNK
product_group_desc	Unknown
last_product_group_code	UNK
last_product_group_desc	Unknown
manufacturer_key	-1
manufacturer_code	UNK
manufacturer_desc	Unknown
design_class_key	1
design_class_code	UNK



product_key	-1
product_id	UNK
product_desc	Unknown
product_line_key	-1
product_line_code	UNK
product_line_desc	Unknown
last_product_line_code	UNK
last_product_line_desc	Unknown
product_group_key	-1
product_group_code	UNK
product_group_desc	Unknown
last_product_group_code	UNK
last_product_group_desc	Unknown
manufacturer_key	-1
manufacturer_code	UNK
manufacturer_desc	Unknown
design_class_key	-1
design_class_code	UNK



product_key	-2
product_id	NA
product_desc	Not Appicable
product_line_key	-2
product_line_code	NA
product_line_desc	Not Appicable
last_product_line_code	NA
last_product_line_desc	Not Appicable
product_group_key	2
product_group_code	NA
product_group_desc	Not Appicable
last_product_group_code	NA
last_product_group_desc	Not Appicable
manufacturer_key	-2
manufacturer_code	NA
manufacturer_desc	Not Appicable
design_class_key	-2
design_class_code	NA

Outline for this Session

- Introduction to Dimension Tables
- Dimension Table Contents
- **Dimension Table Types**
- Generating Dimension Models
- Case Example
- Case Study

Dimension Table Classifications and Examples

- Date based
  - Date Dim
  - Month Dim
- Time based
  - Time Dim
- Business Entities
  - Store Dim
  - Customer Dim
- Analytical Profiles
  - Customer Demographics Dim
- Correlated Entities
  - Finance Profile Dim
- Versions of Business Entities
  - Product Version Dim
- Flags and Indicators
  - Transaction Profile Dim
- Degenerate Dimensions

Date Based Dimensions

- Sale Date, Sales Calendar Month, Sales Calendar Quarter, Sales Calendar Year, Sales Fiscal Month, Sales Fiscal Quarter, Sales Fiscal Year
- Start Date
- End Date
- Arrival Date
- Departure Date
- Order Date
- Ship Date

The Date Dimension Family

- Implement the family and name each dimension for its granularity
  - Date Dim
  - Week Dim
  - Month Dim
  - Quarter Dim
  - Year Dim
- Use at least one character column for date
- Put in all attributes that simplify analysis
- Enable all date functions (add, subtract, etc)

Sample Date Dimension Table

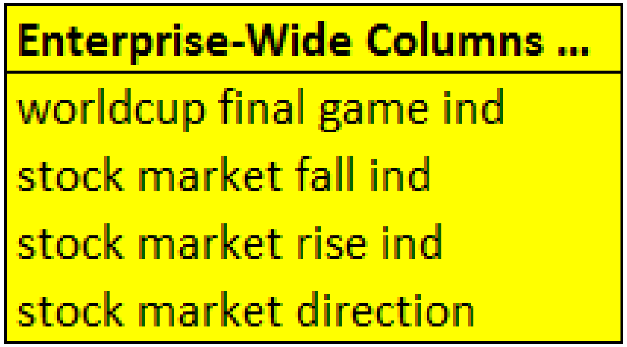
Fiscal Period Columns …
,day_nbr_in_fiscal_quarter
,day_nbr_in_fiscal_year
,week_nbr_in_fiscal_year
,month_nbr_in_fiscal_year
,fiscal_quarter_yyyyq
,fiscal_quarter_name
,fiscal_quarter_abrv
,fiscal_quarter_nbr_in_year
,fiscal_quarter_nbr_overall

Season / Holiday Columns …
holiday ind
holiday name
business season name
calendar season name

Date Dim
day_date_key
,day_date
,day_date_yyyymmdd
,day_of_week_name
,day_of_week_abrv
,day_nbr_in_week
,day_nbr_in_month
,day_nbr_in_quarter
,day_nbr_in_year
,day_nbr_overall
,weekday_ind
,weekend_ind
,week_yyyyww
,week_nbr_in_year
,week_nbr_overall
,month_yyyymm
,month_name
,month_abrv
,month_nbr_in_year
,month_nbr_overall
,quarter_yyyyq
,quarter_name
,quarter_abrv
,quarter_nbr_in_year
,quarter_nbr_overall
,year_yyyy

Enterprise-Wide Events

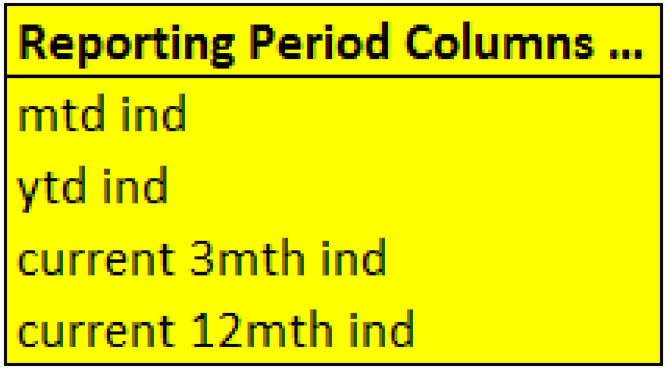
- Special Event Description
  - World Cup final game
  - Significant stock market fall
  - Significant stock market rise
  - Interest rate raise
  - Interest raise drop
  - Gasoline price increase
  - Gasoline price decrease



Enterprise-Wide Columns ...
worldcup final game ind
stock market fall ind
stock market rise ind
stock market direction

Support Reporting Periods

- MTD IND (Month to Date Indicator)
- YTD IND (Year to Date Indicator)
- Current 3MTH IND
- Current 12MTH IND



Reporting Period Columns ...
mtd ind
ytd ind
current 3mth ind
current 12mth ind

Simpler, less error-prone filters

- SELECT sum(sales) FROM fact <join statatements> WHERE date BETWEEN **2/1/2008** AND **4/31/2008**
- vs
- SELECT sum(sales) FROM fact <join statatements> WHERE current_3mth_ind = **'YES'**

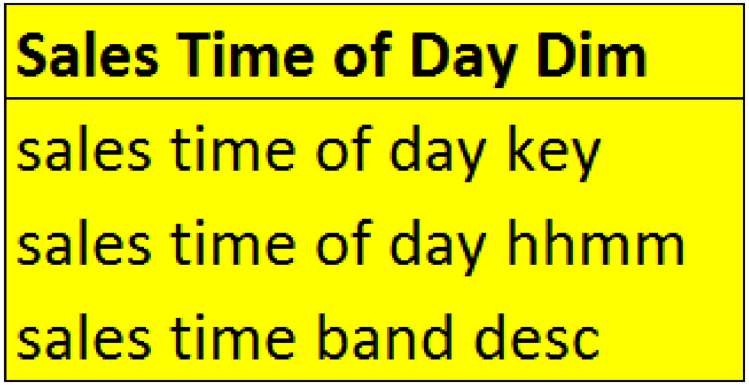
- Easy to extend 'Current 3Mths' or 'Not Current 3Mths'

Date Dimension Physical Implementation

- Custom views for the business dimension roles it plays
  - Generic (Date Dim)
    - Date
    - Month
  - Specific Business Dimension (Sales Date Dim)
    - Sales Date
    - Sales Month
- Implemented using Role-Playing Dimensions

Time Based Dimension

- One row for each minute in a day



Sales Time of Day Dim
sales time of day key
sales time of day hhmm
sales time band desc

The Time Dimension

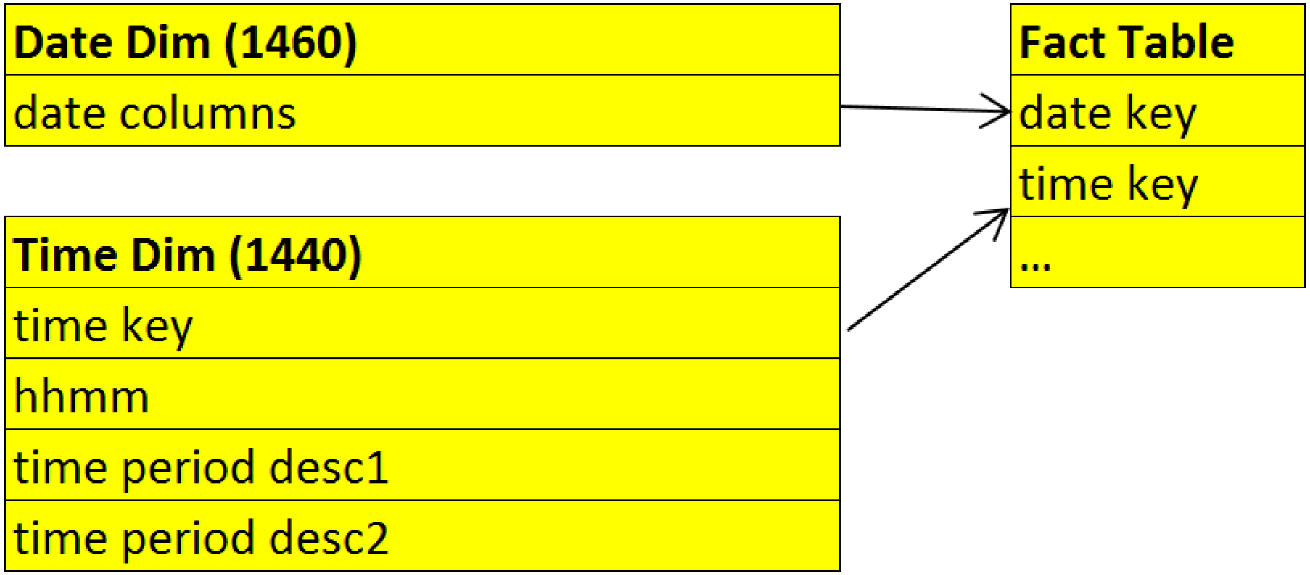
- Objective: To determine the staffing requirements for tellers in a bank at each point in a day
- Should the Day Dimension go to an hour or second granularity?

Day Dimension Sizes for Different Granularities

- Date_Dim = 1460 (4 Years)
- Hour_Dim = 24
- Minute_Dim = 60
- Second_Dim = 60

- Day_Hour_DIM = 35,040
- Day_Min_DIM = 2,102,400
- Day_Second_DIM = 126,144,000

Sample Time Dimension



Date Dim (1460)

date columns

Time Dim (1440)

time key

hhmm

time period desc1

time period desc2

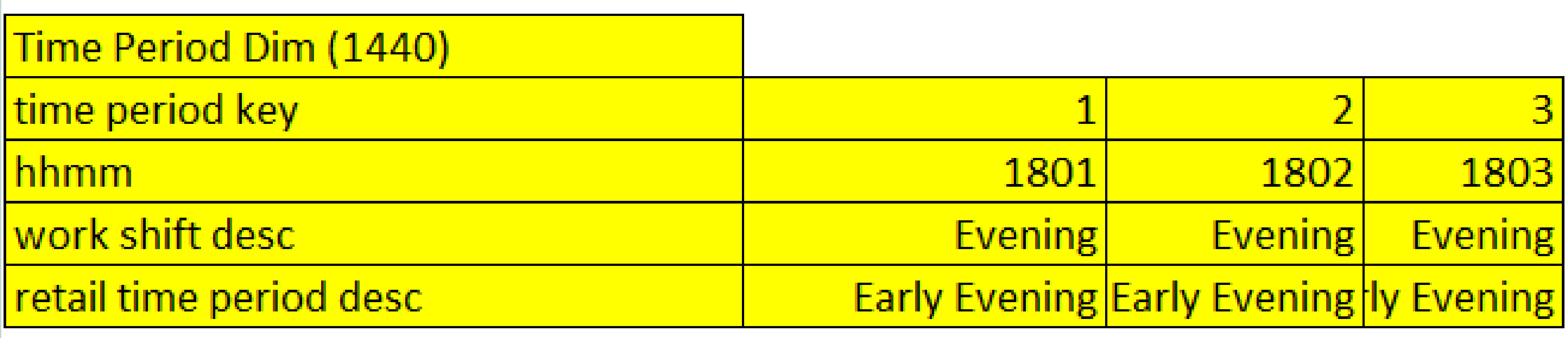
Fact Table

date key

time key

...

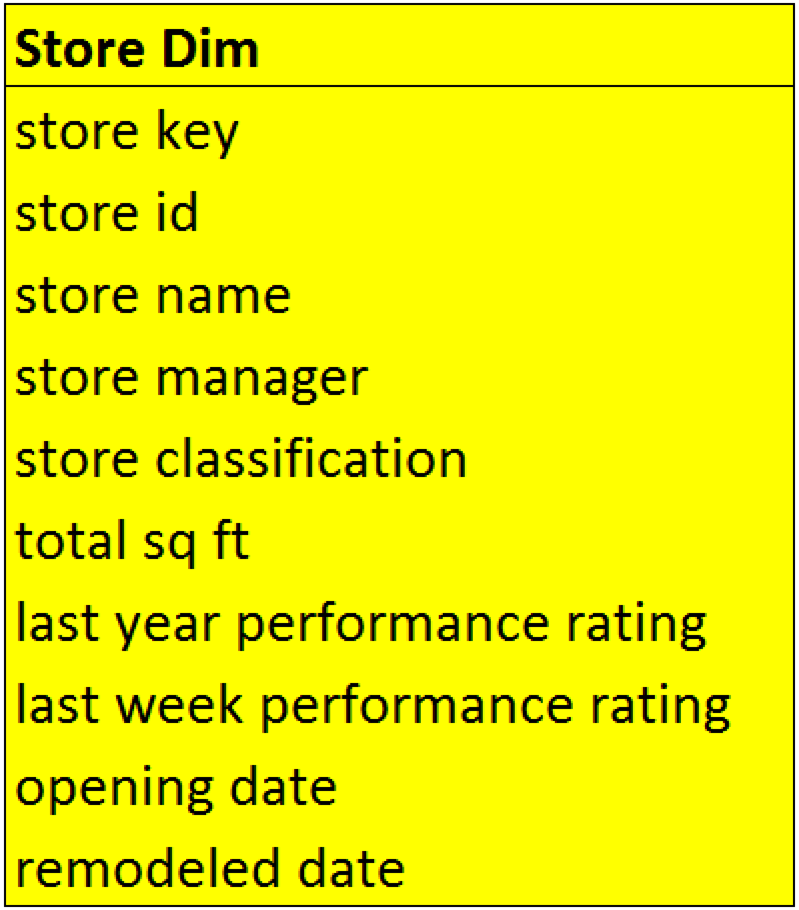
Sample Time Period Data



Time Period Dim (1440)			
time period key	1	2	3
hhmm	1801	1802	1803
work shift desc	Evening	Evening	Evening
retail time period desc	Early Evening	Early Evening	ly Evening

Business Entity Dimension

- One row for each store.



Store Dim
store key
store id
store name
store manager
store classification
total sq ft
last year performance rating
last week performance rating
opening date
remodeled date

Detail and Analytical Profile Dimensions

- One row for each instance of the business entity

Customer Detail Dim
customer key
customer billing id
customer service id
customer full name
customer full street address
customer salutation
customer salutation type
customer first and middle names
customer first name
customer middle name
customer last name
customer name origin
customer home telephone country code
customer home telephone area code
customer home telephone number
customer fax country code
customer fax area code
customer fax number
customer week telephone country code
customer week telephone area code
customer week telephone number
customer email address
customer email domain
customer secondary postal code
customer current geo profile key
customer current dmg profile key
customer current scoring profile key

Customer Demographics Profile
customer dmg key
customer gender
customer age range
customer ethnicity
customer household type

Customer Geographic Profile
customer geo key
customer country
customer region
customer city
customer county
customer state
customer postal code type
customer primary postal code

Customer Scoring Profile
customer scoring profile key
customer duration range
total purchase amount range
last 3 month purchase amount range
customer segment

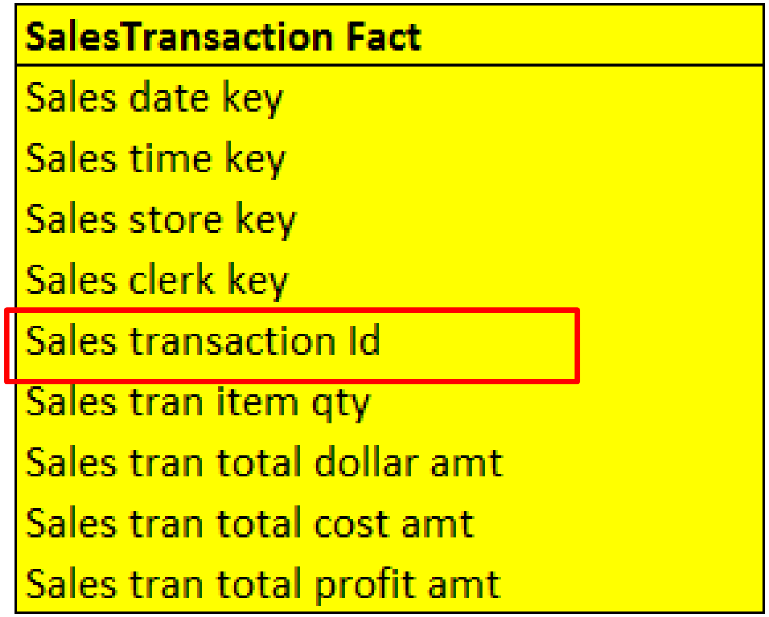
Version Dimension

- Keeps track of all history of changes in a business entity
- One row for each change date for each product.

Product Version Dimension
product_key
product_id
product_desc
product_line_key
product_line_code
product_line_desc
last_product_line_code
last_product_line_desc
product_group_key
product_group_code
product_group_desc
last_product_group_code
last_product_group_desc
manufacturer_key
manufacturer_code
manufacturer_desc
design_class_key
design_class_code
design_class_desc
effective_date
current_row_ind

Degenerate Dimension

- Transaction id is a degenerate dimension
- All transaction attributes have already been attached to the fact table



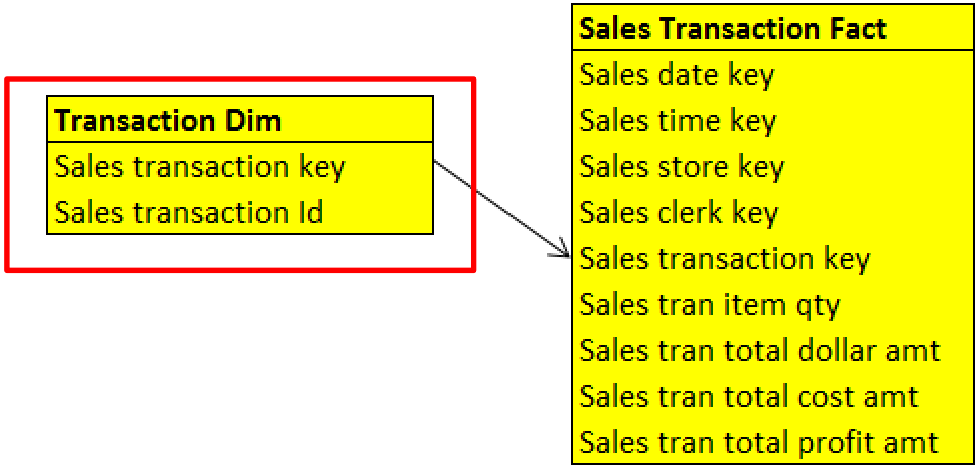
SalesTransaction Fact
Sales date key
Sales time key
Sales store key
Sales clerk key
Sales transaction Id
Sales tran item qty
Sales tran total dollar amt
Sales tran total cost amt
Sales tran total profit amt

Large Space
Needed!

Degenerate Dimension: Alternative Model

- Make a dimension for transaction
  - Protects against reuse of transaction ids
  - Reduces size of critical analytical space
  - Retain smaller key in fact table which is seldom used
  - Will serve as a dimension to a transaction level fact table if built
  - It will only be used if there is a need to bring back individual detail
  - Good container for text comment field

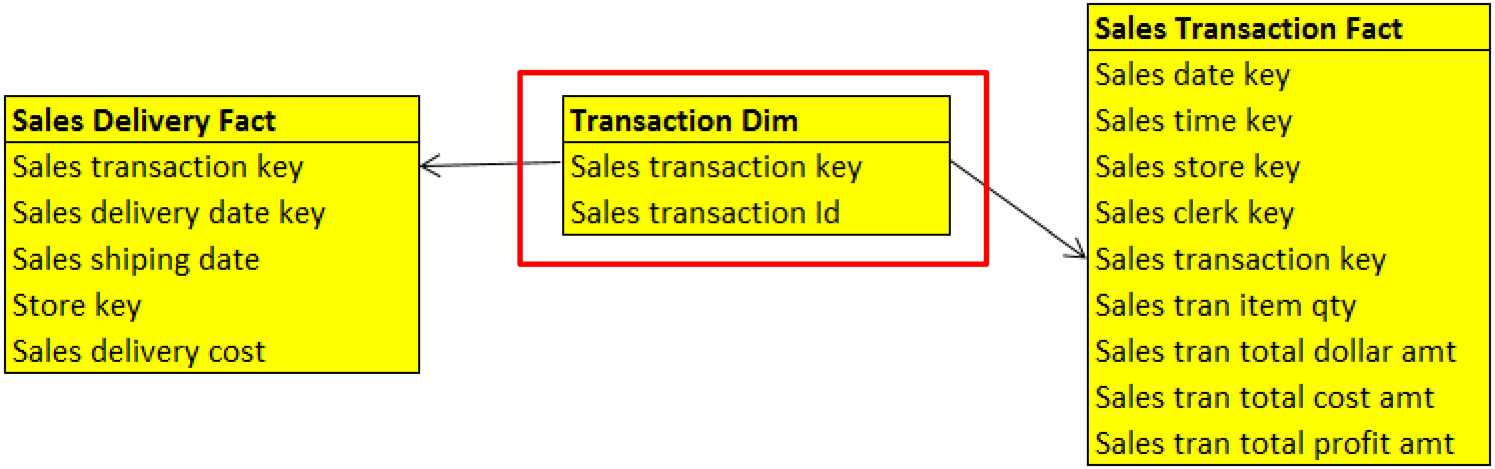
Implementation of a Degenerate Dim



Transaction Dim
Sales transaction key
Sales transaction Id

Sales Transaction Fact
Sales date key
Sales time key
Sales store key
Sales clerk key
Sales transaction key
Sales tran item qty
Sales tran total dollar amt
Sales tran total cost amt
Sales tran total profit amt

- Can Also Handle Multiple Connections



Sales Delivery Fact
Sales transaction key
Sales delivery date key
Sales shiping date
Store key
Sales delivery cost

Transaction Dim
Sales transaction key
Sales transaction Id

Sales Transaction Fact
Sales date key
Sales time key
Sales store key
Sales clerk key
Sales transaction key
Sales tran item qty
Sales tran total dollar amt
Sales tran total cost amt
Sales tran total profit amt

What About?

- Weather
- Promotions
- Ads
- Displays

Condition Dimensions

- Conditions that may effect fact table activity dependent on date and another dimension
- Cannot be handled in the date dimension
- Examples
  - Local events
  - Promotions
  - Weather Conditions

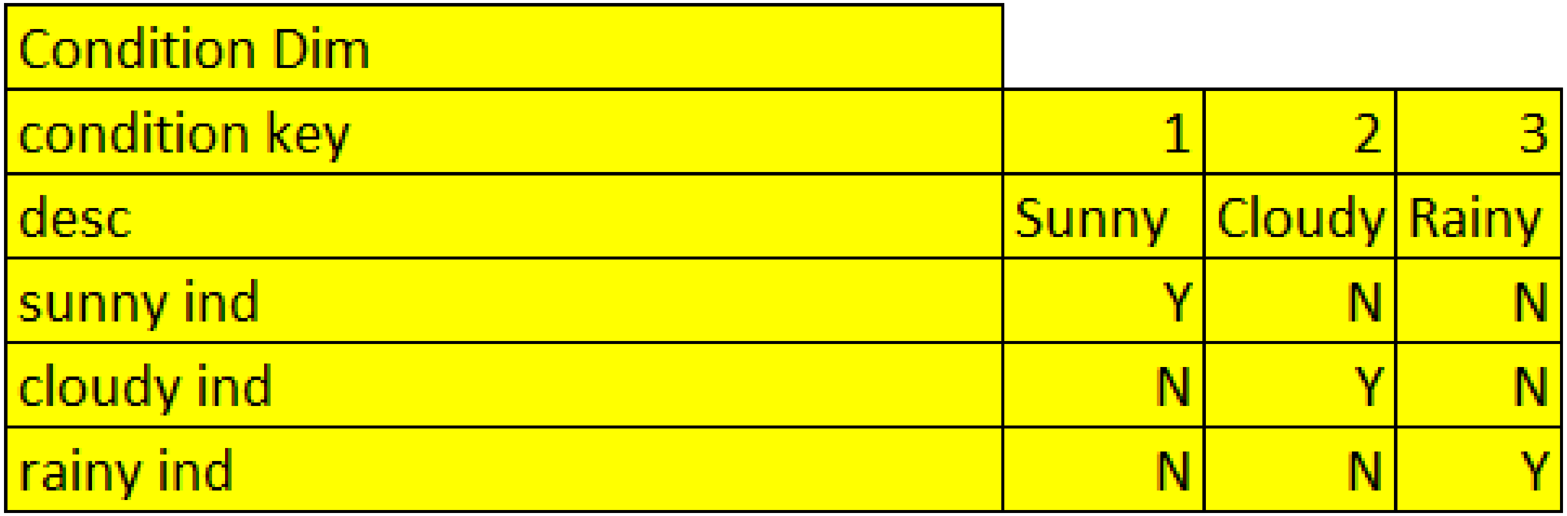
Local Weather Dimension

- Condition whose value is determined by day and store
- Often need system to capture
- Added value for conditions specific to your business

Consider Using Indicators

- Make values into column names
  - Very helpful for usability
- Use character data type
  - No ones and zeros
- Need standardized method of handling
  - Several choices

Proposed Method

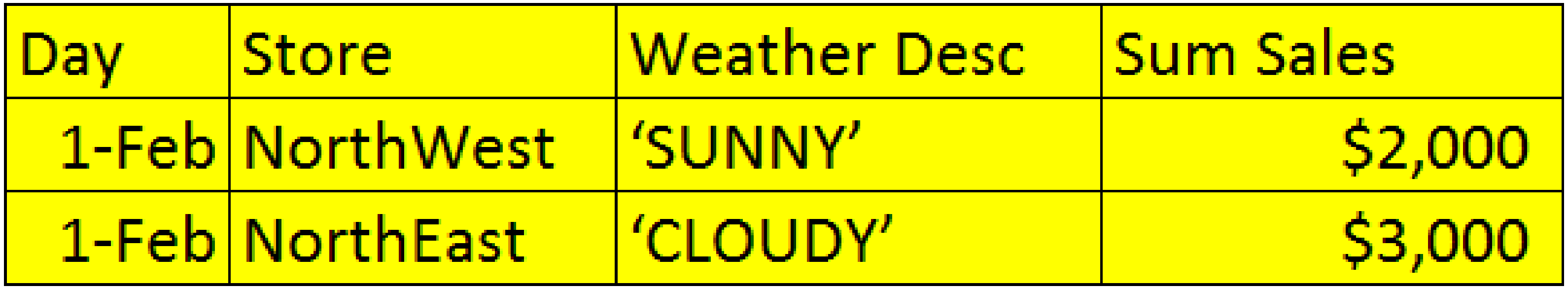


Condition Dim			
condition key	1	2	3
desc	Sunny	Cloudy	Rainy
sunny ind	Y	N	N
cloudy ind	N	Y	N
rainy ind	N	N	Y

- Only practical if there are a limited/finite number of values

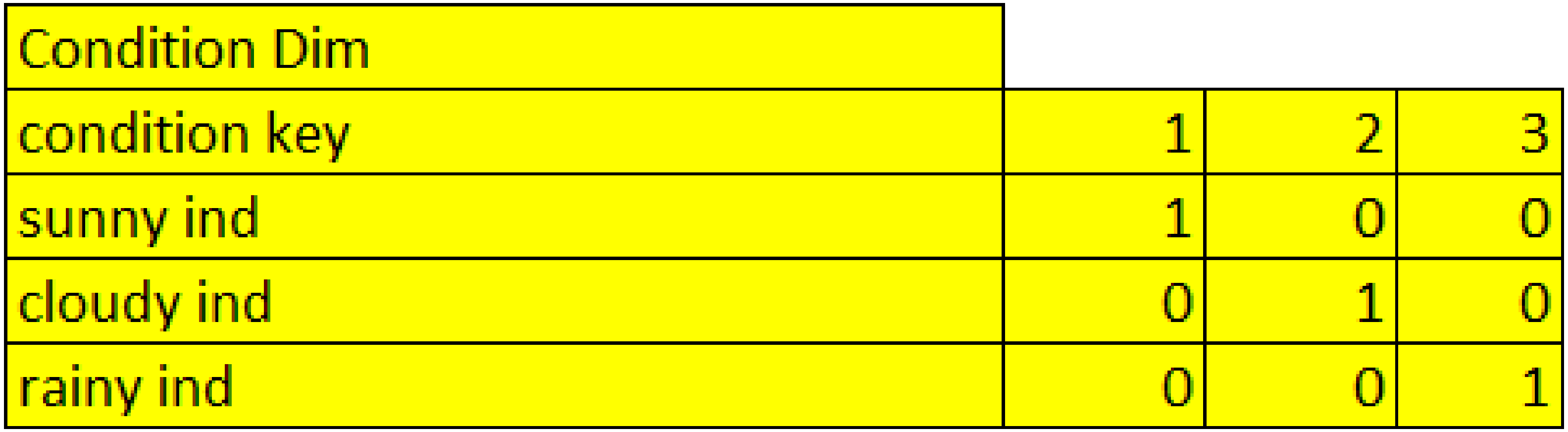
Usage

```
SELECT Day, Store, Weather_Desc,
SUM(Sales)
FROM Weather
WHERE Weather_Desc = ‘SUNNY’ OR
Weather_Desc = ‘CLOUDY’ ;
```



Day	Store	Weather Desc	Sum Sales
1-Feb	NorthWest	‘SUNNY’	$2,000
1-Feb	NorthEast	‘CLOUDY’	$3,000

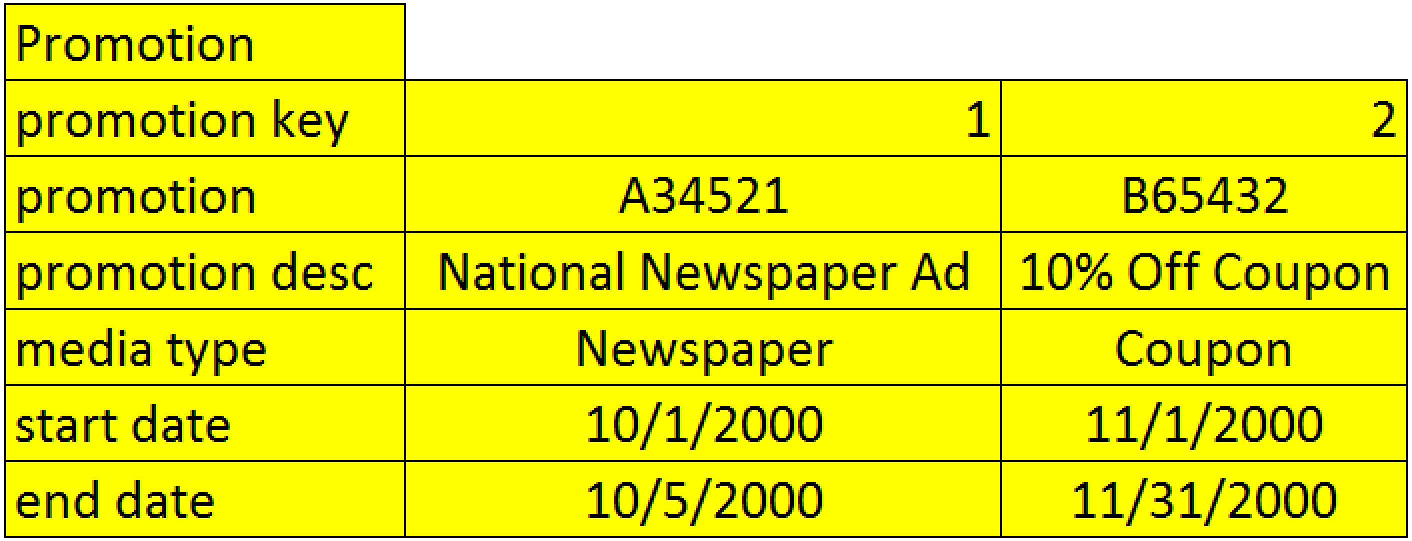
Avoid



Condition Dim			
condition key	1	2	3
sunny ind	1	0	0
cloudy ind	0	1	0
rainy ind	0	0	1

The Promotion Dimension

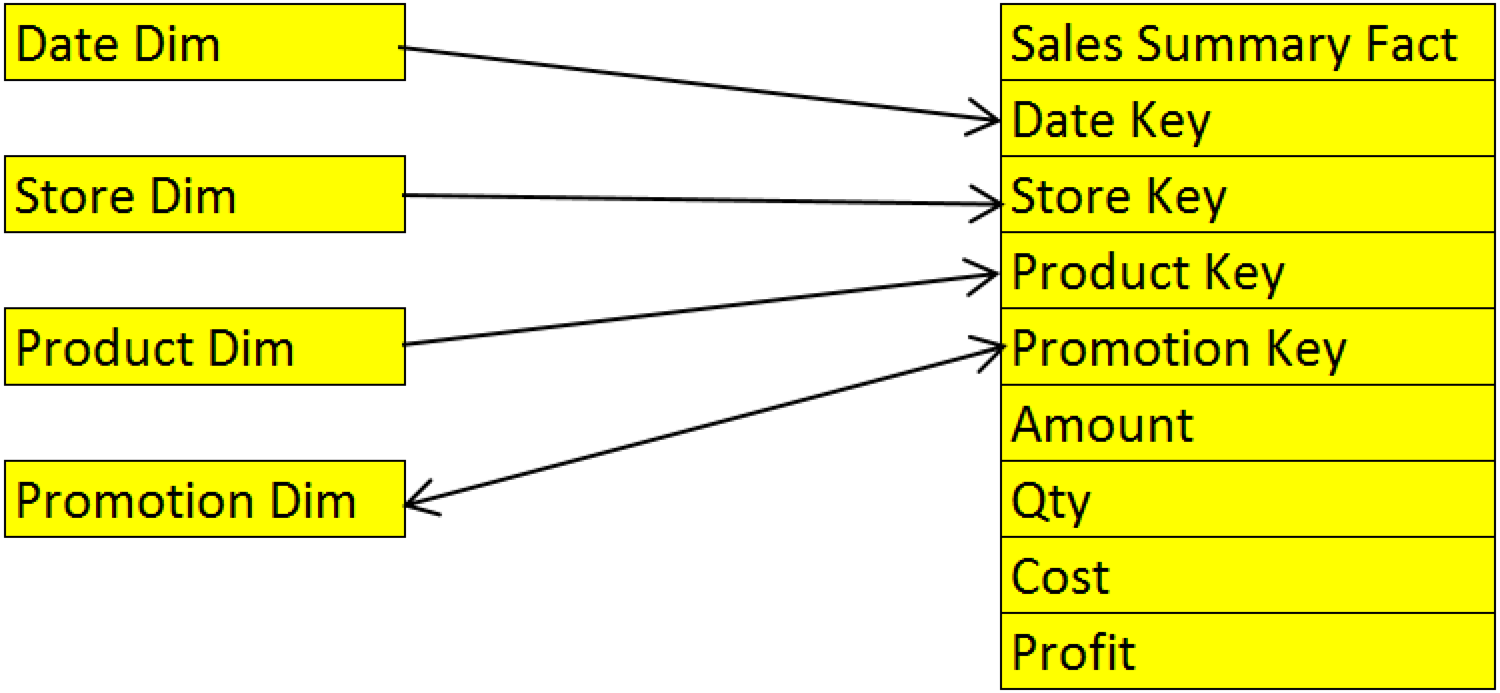
- Another condition or causal dimension
- Hopefully promotions affect product sales



Promotion		
promotion key	1	2
promotion	A34521	B65432
promotion desc	National Newspaper Ad	10% Off Coupon
media type	Newspaper	Coupon
start date	10/1/2000	11/1/2000
end date	10/5/2000	11/31/2000

The Problem

- Products may have a combination of promotions in effect at the same time



Date Dim

Store Dim

Product Dim

Promotion Dim

Sales Summary Fact

Date Key

Store Key

Product Key

Promotion Key

Amount

Qty

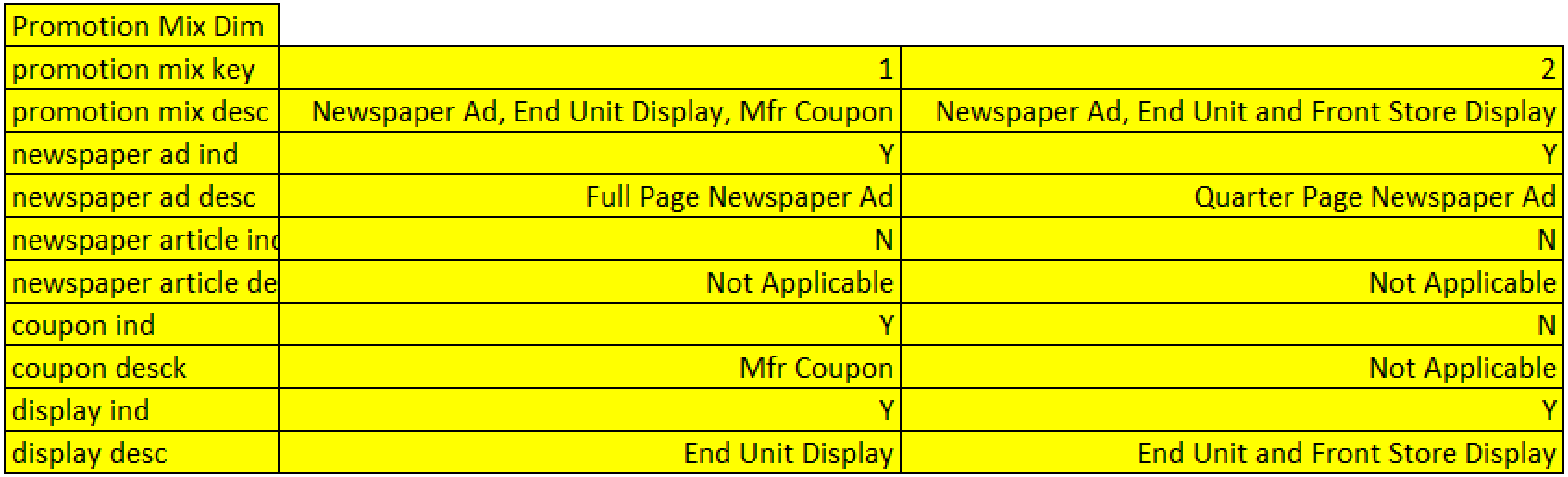
Cost

Profit

Multi-valued Dimensions

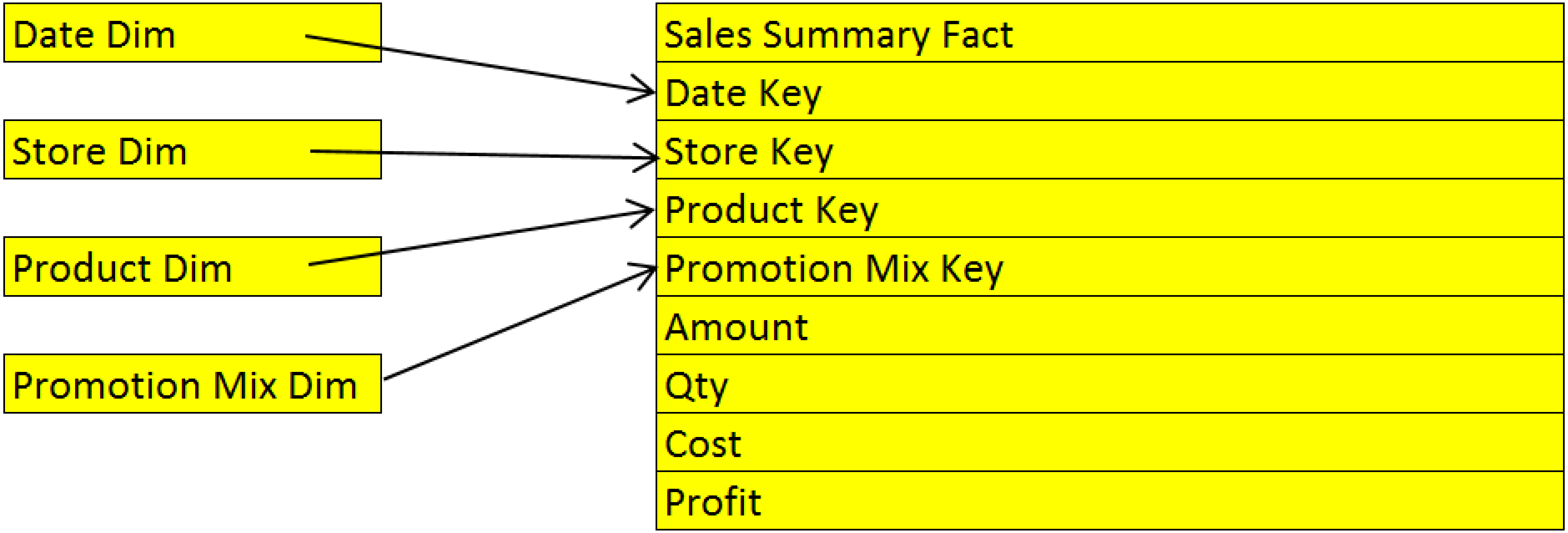
- Dimensions where the dimension table takes on more than one value for an individual fact table row
- One solution is to convert the multiple rows into one row
- Usually Called: Mix Dimension table
- One row for each different mix of values encountered in an individual fact table row

The Promotion Mix Dimension



Promotion Mix Dim		
promotion mix key	1	2
promotion mix desc	Newspaper Ad, End Unit Display, Mfr Coupon	Newspaper Ad, End Unit and Front Store Display
newspaper ad ind	Y	Y
newspaper ad desc	Full Page Newspaper Ad	Quarter Page Newspaper Ad
newspaper article ind	N	N
newspaper article de	Not Applicable	Not Applicable
coupon ind	Y	N
coupon desck	Mfr Coupon	Not Applicable
display ind	Y	Y
display desc	End Unit Display	End Unit and Front Store Display

New Schema



Date Dim

Store Dim

Product Dim

Promotion Mix Dim

Sales Summary Fact

Date Key

Store Key

Product Key

Promotion Mix Key

Amount

Qty

Cost

Profit

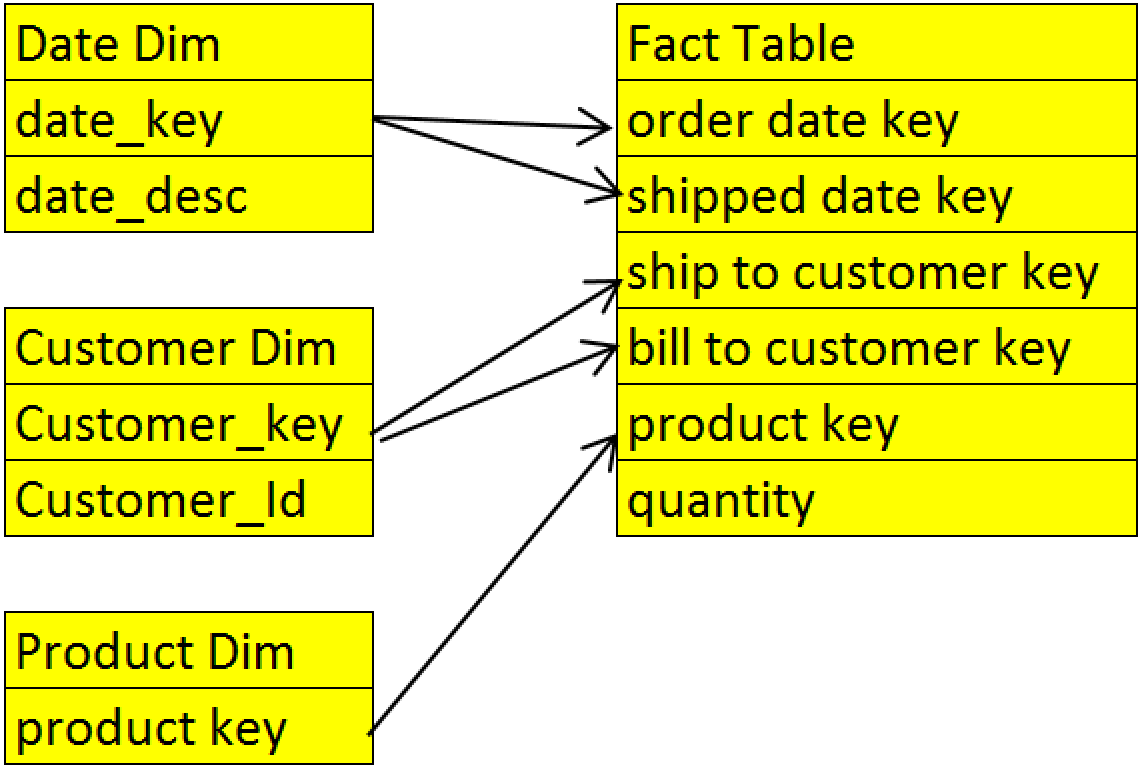
Provides Valuable Analyses

- Average daily sales or profit when product
  - Is on display plus ad plus coupon
  - Is on display plus coupon
  - Is on display plus ad
  - Is on display only
  - Is on ad plus coupon
  - Is on ad only
  - Coupon only
  - Not on promotion

Role-Playing Dimensions

- Entity taking on different roles or uses for the same entity
- Examples
  - Order date
  - Ship date
  - Bill to customer
  - Ship to customer
  - Flight origination airport
  - Flight destination airport

Wrong Way to Model



Date Dim
date_key
date_desc

Customer Dim
Customer_key
Customer_Id

Product Dim
product key

Fact Table
order date key
shipped date key
ship to customer key
bill to customer key
product key
quantity

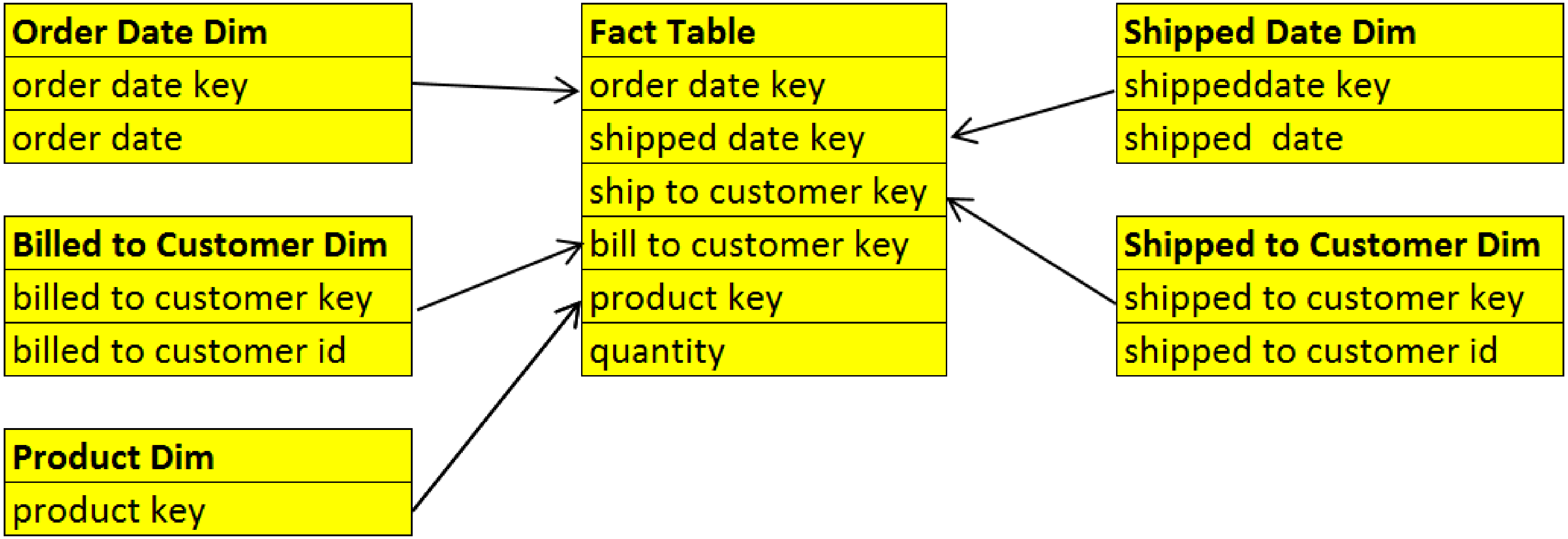
Drag and Drop Consequences

- Report Desired
  - Bill to customer id
  - Shipped to customer id
  - Order date
  - Shipped date
  - Quantity

- Drag and drop result
  - Customer id
  - Customer id
  - Day date
  - Day date

- The column headings are the same and indistinguishable from each other as to which is which.

Dimensional Implementation Logical View



Order Date Dim
order date key
order date

Billed to Customer Dim
billed to customer key
billed to customer id

Product Dim
product key

Fact Table
order date key
shipped date key
ship to customer key
bill to customer key
product key
quantity

Shipped Date Dim
shippeddate key
shipped date

Shipped to Customer Dim
shipped to customer key
shipped to customer id

Dimensional Physical Implementation

- Physical base table
  - Date Dim
- Views*
  - Order Date Dim
  - Ship Date Dim
- View
  - is the result set of a stored query on the data, which the database users can query just as they would in a persistent database collection object.
  - Is not part of the physical schema

SQL Code Using Views for Role-Playing Dimensions

- Date dimension

  -
    ```
    CREATE TABLE DATE (date_key, day_of_week, month, …)
    ```
- Order Date view

  -
    ```
    CREATE VIEW ORDER_DATE (order_date_key, order_day_of_week, order_month, …) AS SELECT date_key, day_of_week, month, … FROM DATE
    ```
- Ship Date view

  -
    ```
    CREATE VIEW SHIP_DATE (ship_date_key, ship_day_of_week, ship_month, …) AS SELECT date_key, day_of_week, month, … FROM DATE
    ```

Logical Dimension Table

Dim
date_key
date_desc
date_yyyymmdd
day_of_week_name
day_of_week_abrv
day_nbr_in_week
day_nbr_in_month
day_nbr_in_quarter
day_nbr_in_year
day_nbr_overall
weekday_ind
weekend_ind
week_yyyyww
week_nbr_in_year
week_nbr_overall
month_yyyymm
month_name
month_abrv
month_nbr_in_year
month_nbr_overall
quarter_yyyyq
quarter_name
quarter_abrv
quarter_nbr_in_year
quarter_nbr_overall
year

create view ORDER_DATE_DIM
(order_date_key
order_date
order_date_yyyymmdd
order_day_of_week_name
order_day_of_week_abrv
order_day_nbr_in_week
order_day_nbr_in_month
order_day_nbr_in_quarter
order_day_nbr_in_year
order_day_nbr_overall
order_weekday_ind
order_weekend_ind
order_week_yyyyww
order_week_nbr_in_year
order_week_nbr_overall
order_month_yyyymm
order_month_name
order_month_abrv
order_month_nbr_in_year
order_month_nbr_overall
order_quarter_yyyyq
order_quarter_name
order_quarter_abrv
order_quarter_nbr_in_year
order_quarter_nbr_overall
order_year)

create view SHIP_DATE_DIM
(ship_date_key
ship_date
ship_date_yyyymmdd
ship_day_of_week_name
ship_day_of_week_abrv
ship_day_nbr_in_week
ship_day_nbr_in_month
ship_day_nbr_in_quarter
ship_day_nbr_in_year
ship_day_nbr_overall
ship_weekday_ind
ship_weekend_ind
ship_week_yyyyww
ship_week_nbr_in_year
ship_week_nbr_overall
ship_month_yyyymm
ship_month_name
ship_month_abrv
ship_month_nbr_in_year
ship_month_nbr_overall
ship_quarter_yyyyq
ship_quarter_name
ship_quarter_abrv
ship_quarter_nbr_in_year
ship_quarter_nbr_overall
ship_year)

Outline for this Session

- Introduction to Dimension Tables
- Dimension Table Contents
- Dimension Table Types
- **Generating Dimension Models**
- Case Example
- Case Study

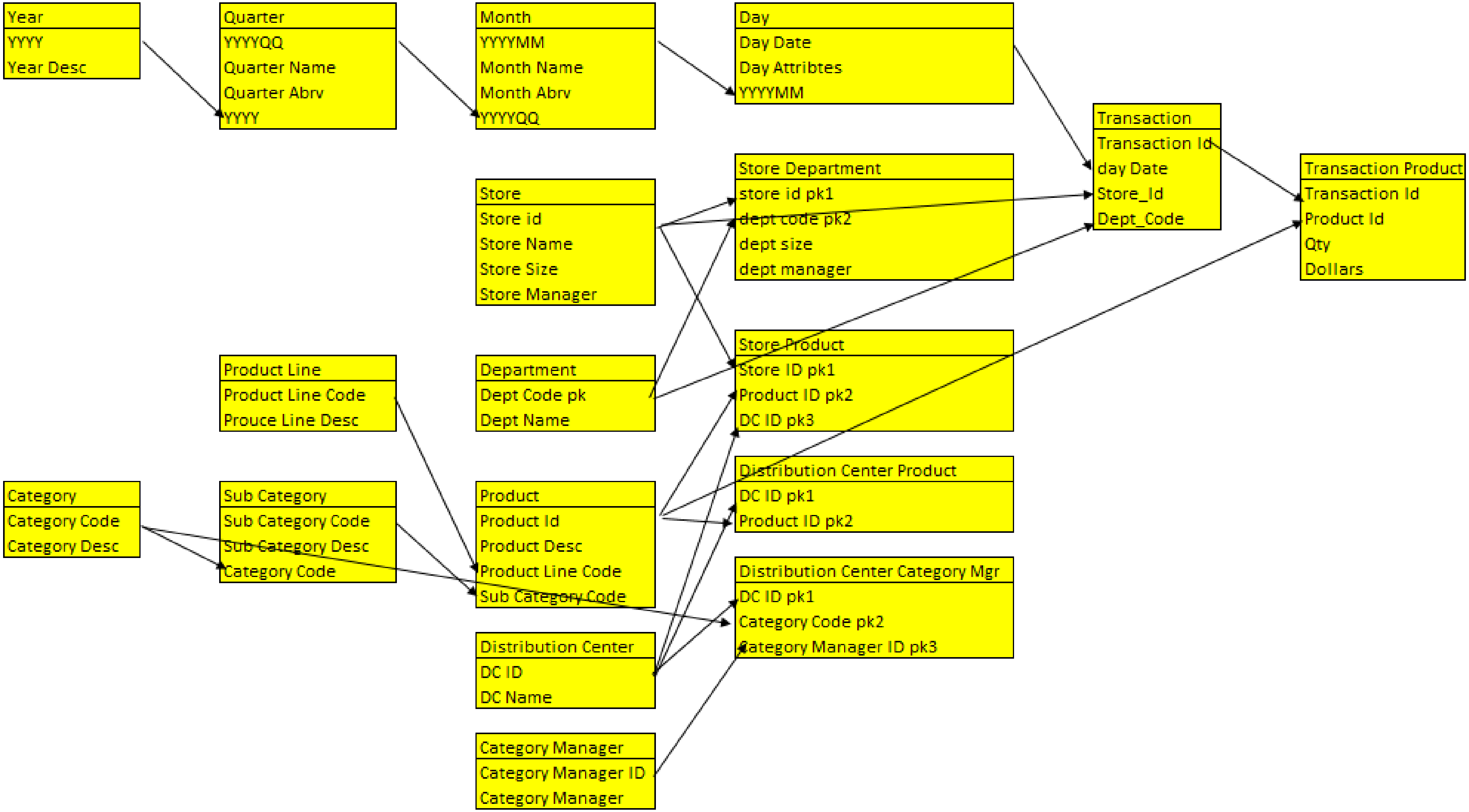
Dimensional Normal Form

- A creative and practical approach originated by Mike Schmitz to design Dimension Table Families
  - Fact tables are highly normalized for maintainability and flexibility
  - Dimensions have their hierarchies de-normalized into them for usability and performance
  - The schema is limited to two levels
    - A single first level or central highly normalized table called a fact table
    - Multiple second level tables called dimension tables linked to the first level table in primarily one to many relationships

Dimension Table Families

- Business dimensions should be modeled in 3NF reflecting the true hierarchical relationships embedded in them
- Each embedded dimension should be implemented as a separate dimension table with the upper level dimensions de-normalized into them
- Only one dimension table in a dimension family is attached to any one fact table

Example: 3NF ~ 169 Gb



Year
YYYY
Year Desc

Category
Category Code
Category Desc

Quarter
YYYYQQ
Quarter Name
Quarter Abrv
YYYY

Product Line
Product Line Code
Prouce Line Desc

Sub Category
Sub Category Code
Sub Category Desc
Category Code

Month
YYYYMM
Month Name
Month Abrv
YYYYQQ

Store
Store id
Store Name
Store Size
Store Manager

Department
Dept Code pk
Dept Name

Product
Product Id
Product Desc
Product Line Code
Sub Category Code

Distribution Center
DC ID
DC Name

Category Manager
Category Manager ID
Category Manager

Day
Day Date
Day Attribtes
YYYYMM

Store Department
store id pk1
dept code pk2
dept size
dept manager

Store Product
Store ID pk1
Product ID pk2
DC ID pk3

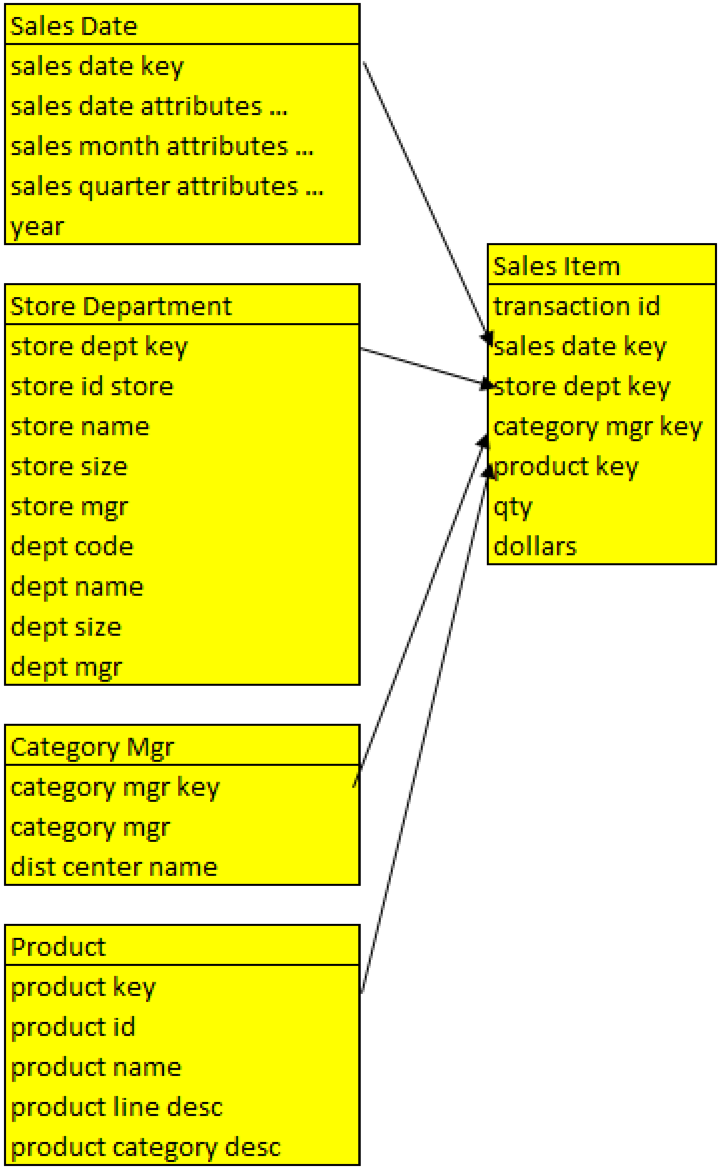
Distribution Center Product
DC ID pk1
Product ID pk2

Distribution Center Category Mgr
DC ID pk1
Category Code pk2
Category Manager ID pk3

Transaction
Transaction Id
day Date
Store_Id
Dept_Code

Transaction Product
Transaction Id
Product Id
Qty
Dollars

Dimensional Normal Form Schema



Sales Date
sales date key
sales date attributes …
sales month attributes …
sales quarter attributes …
year

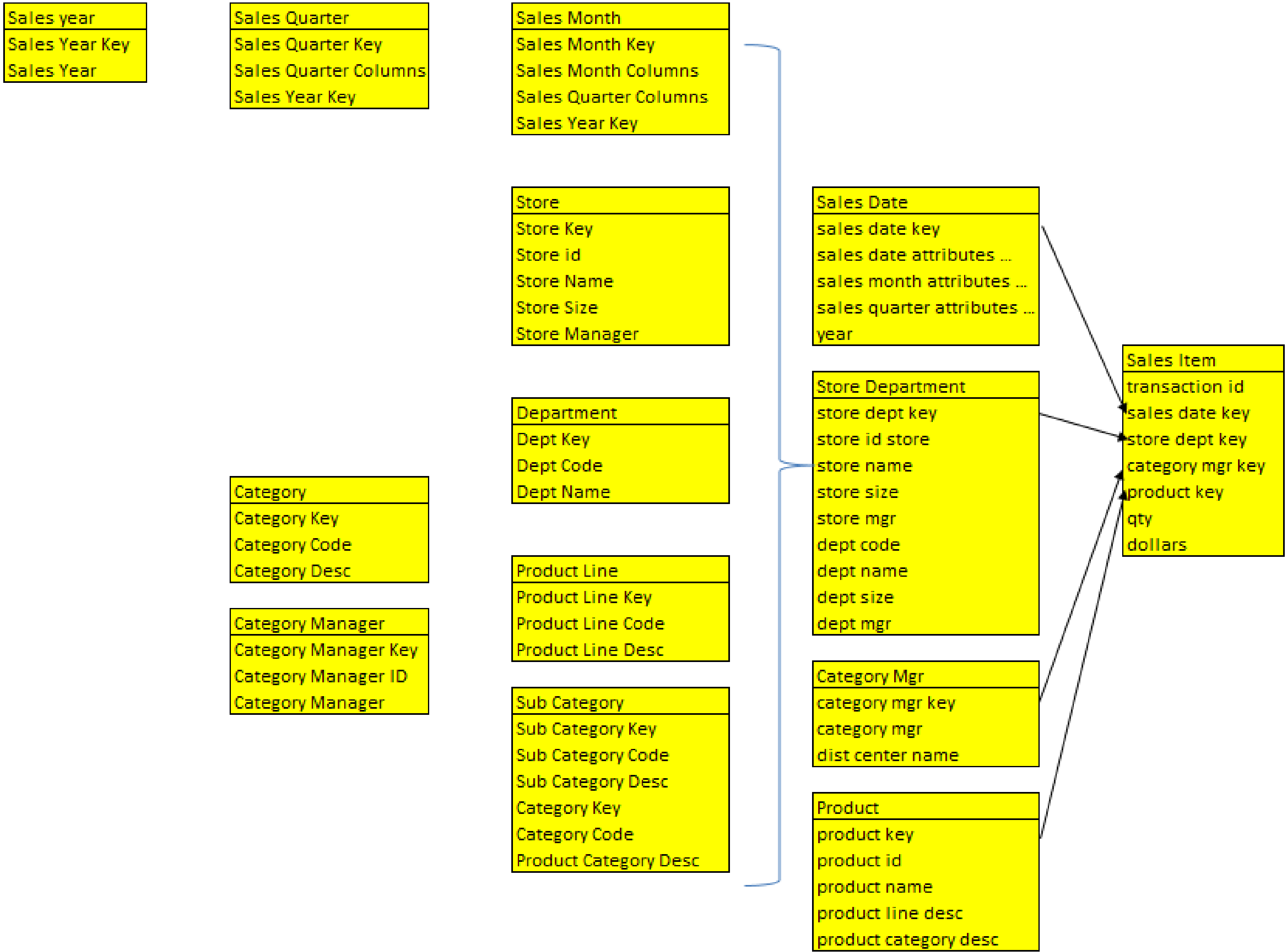
Store Department
store dept key
store id store
store name
store size
store mgr
dept code
dept name
dept size
dept mgr

Category Mgr
category mgr key
category mgr
dist center name

Product
product key
product id
product name
product line desc
product category desc

Sales Item
transaction id
sales date key
store dept key
category mgr key
product key
qty
dollars

Dimensional Normal Form with Dimension Families



Sales year
Sales Year Key
Sales Year

Sales Quarter
Sales Quarter Key
Sales Quarter Columns
Sales Year Key

Sales Month
Sales Month Key
Sales Month Columns
Sales Quarter Columns
Sales Year Key

Store
Store Key
Store id
Store Name
Store Size
Store Manager

Department
Dept Key
Dept Code
Dept Name

Category
Category Key
Category Code
Category Desc

Category Manager
Category Manager Key
Category Manager ID
Category Manager

Product Line
Product Line Key
Product Line Code
Product Line Desc

Sub Category
Sub Category Key
Sub Category Code
Sub Category Desc
Category Key
Category Code
Product Category Desc

Sales Date
sales date key
sales date attributes …
sales month attributes …
sales quarter attributes …
year

Store Department
store dept key
store id store
store name
store size
store mgr
dept code
dept name
dept size
dept mgr

Category Mgr
category mgr key
category mgr
dist center name

Product
product key
product id
product name
product line desc
product category desc

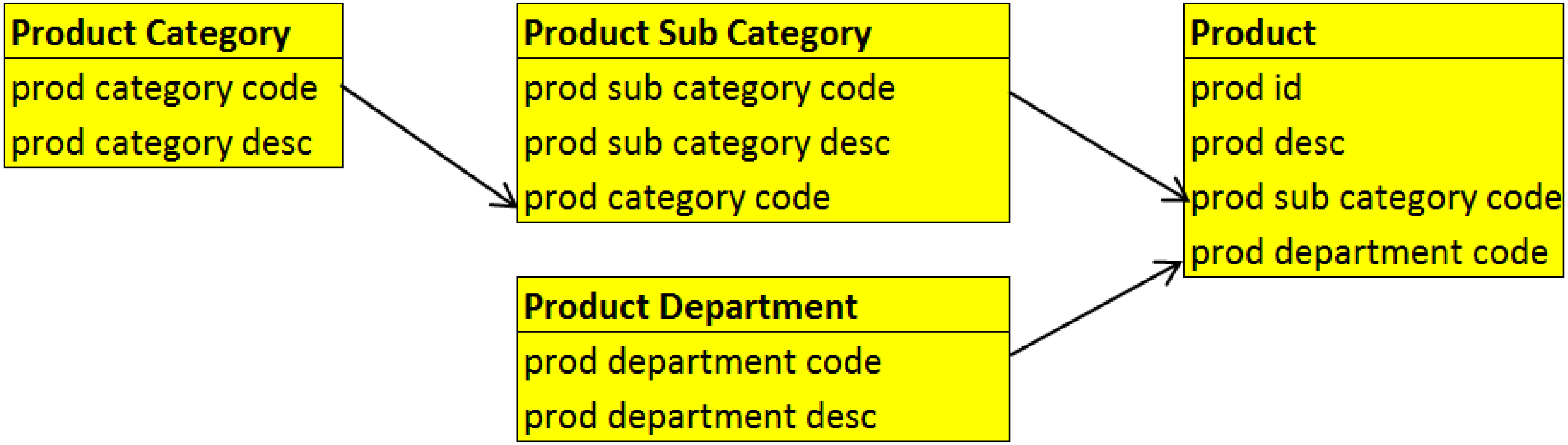
Sales Item
transaction id
sales date key
store dept key
category mgr key
product key
qty
dollars

Space is not a Problem

- 3rd Normal Form
  - – 169,363,742,293 bytes = ~169 GB
- 3nf – De-Normalized for performance (into the fact table)
  - – 227,760,000,000 bytes = ~228 GB

Product Dimension Family Normalized Model

- Dimensional Normal Form (DNF) Step 1



Product Category
prod category code
prod category desc

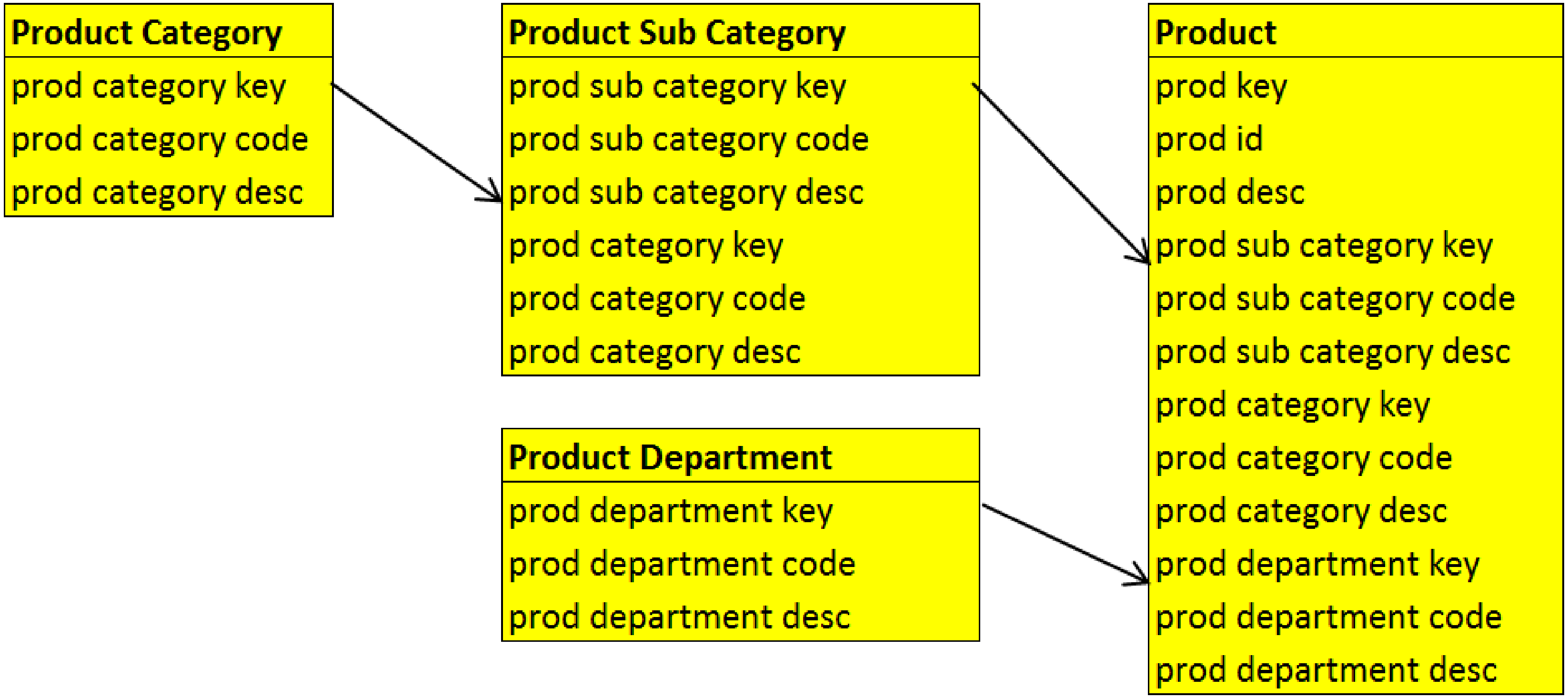
Product Sub Category
prod sub category code
prod sub category desc
prod category code

Product Department
prod department code
prod department desc

Product
prod id
prod desc
prod sub category code
prod department code

Product Dimension Family Denormalized Model

- Dimensional Normal Form (DNF) Step 2



Product Category
prod category key
prod category code
prod category desc

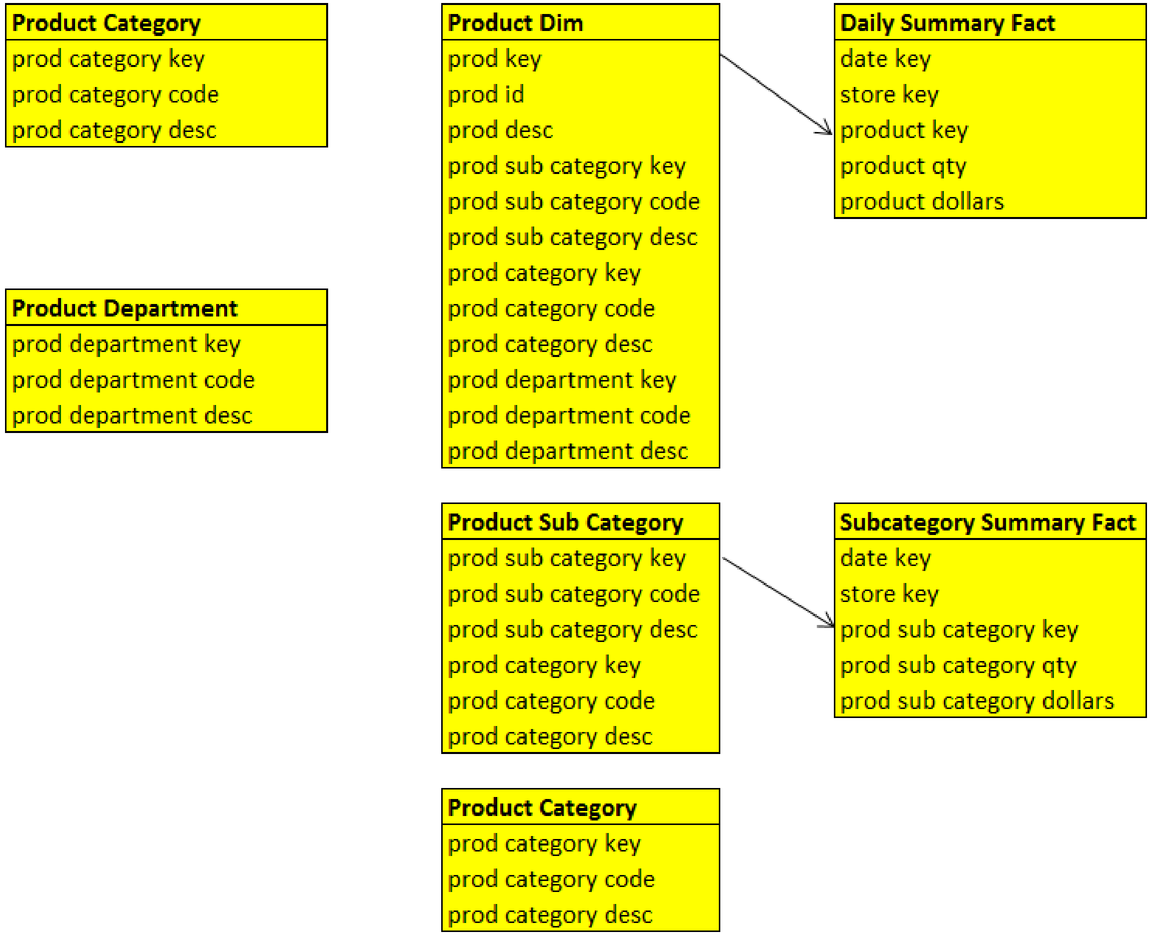
Product Sub Category
prod sub category key
prod sub category code
prod sub category desc
prod category key
prod category code
prod category desc

Product Department
prod department key
prod department code
prod department desc

Product
prod key
prod id
prod desc
prod sub category key
prod sub category code
prod sub category desc
prod category key
prod category code
prod category desc
prod department key
prod department code
prod department desc

Product Dimension Family Usage
Product Dim Daily Summary Fact

- Dimensional Normal Form (DNF) Step 3



Product Category
prod category key
prod category code
prod category desc

Product Department
prod department key
prod department code
prod department desc

Product Dim
prod key
prod id
prod desc
prod sub category key
prod sub category code
prod sub category desc
prod category key
prod category code
prod category desc
prod department key
prod department code
prod department desc

Daily Summary Fact
date key
store key
product key
product qty
product dollars

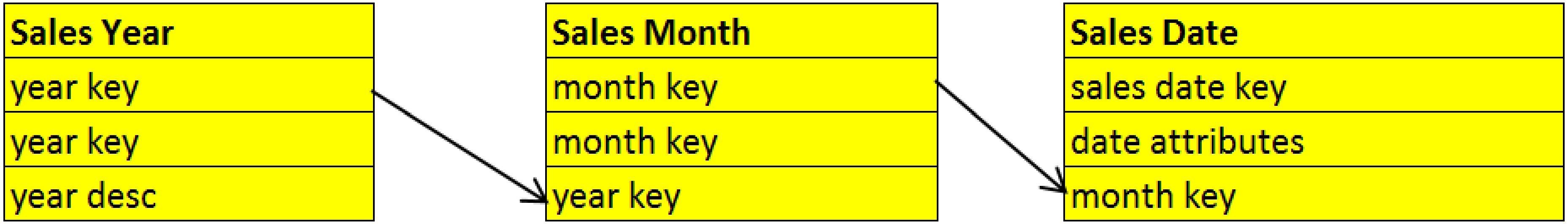
Product Sub Category
prod sub category key
prod sub category code
prod sub category desc
prod category key
prod category code
prod category desc

Subcategory Summary Fact
date key
store key
prod sub category key
prod sub category qty
prod sub category dollars

Product Category
prod category key
prod category code
prod category desc

What is Snowflaking?

- To use normalized tables in the dimensional model.
- Break dimension hierarchies into normalized tables connected by foreign key – primary key relationships



Sales Year
year key
year key
year desc

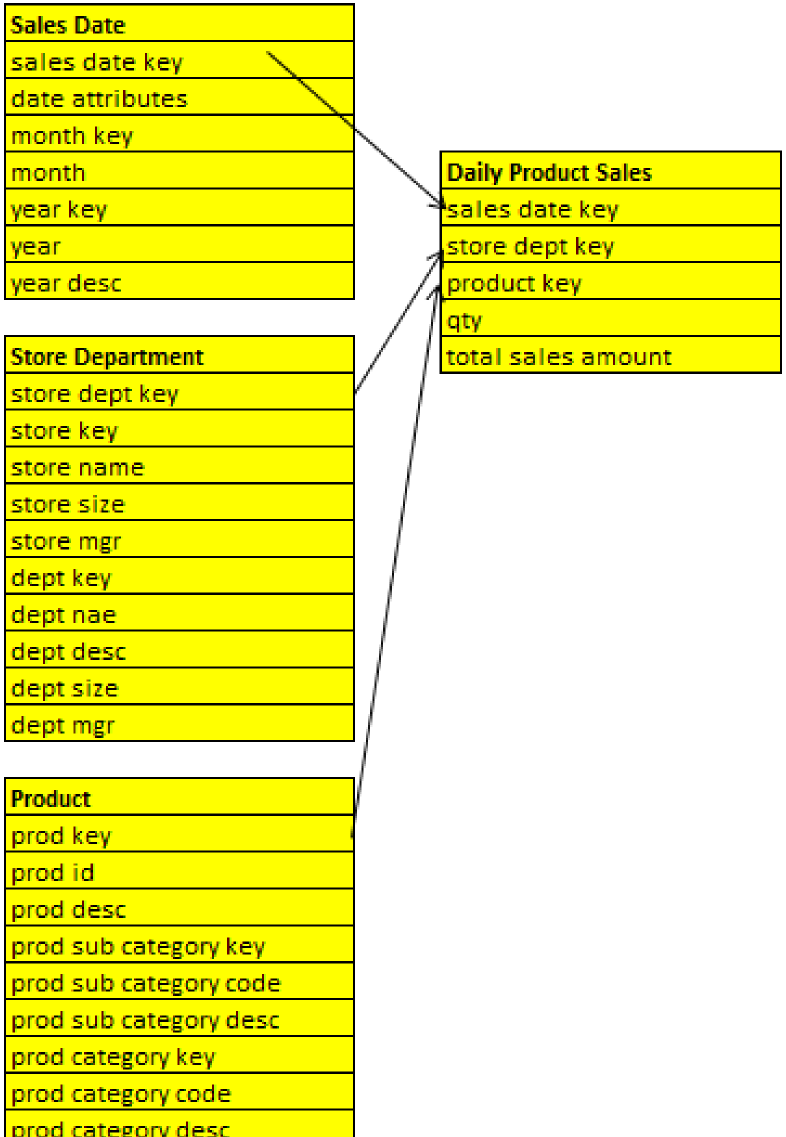
Sales Month
month key
month key
year key

Sales Date
sales date key
date attributes
month key

Why is this Bad ?

- **Joins**
- **Joins**
- **Joins**
- **Joins**
- Every join costs something and one extra join may cause the database optimizer to choose a bad algorithm

Dimension Table Solution



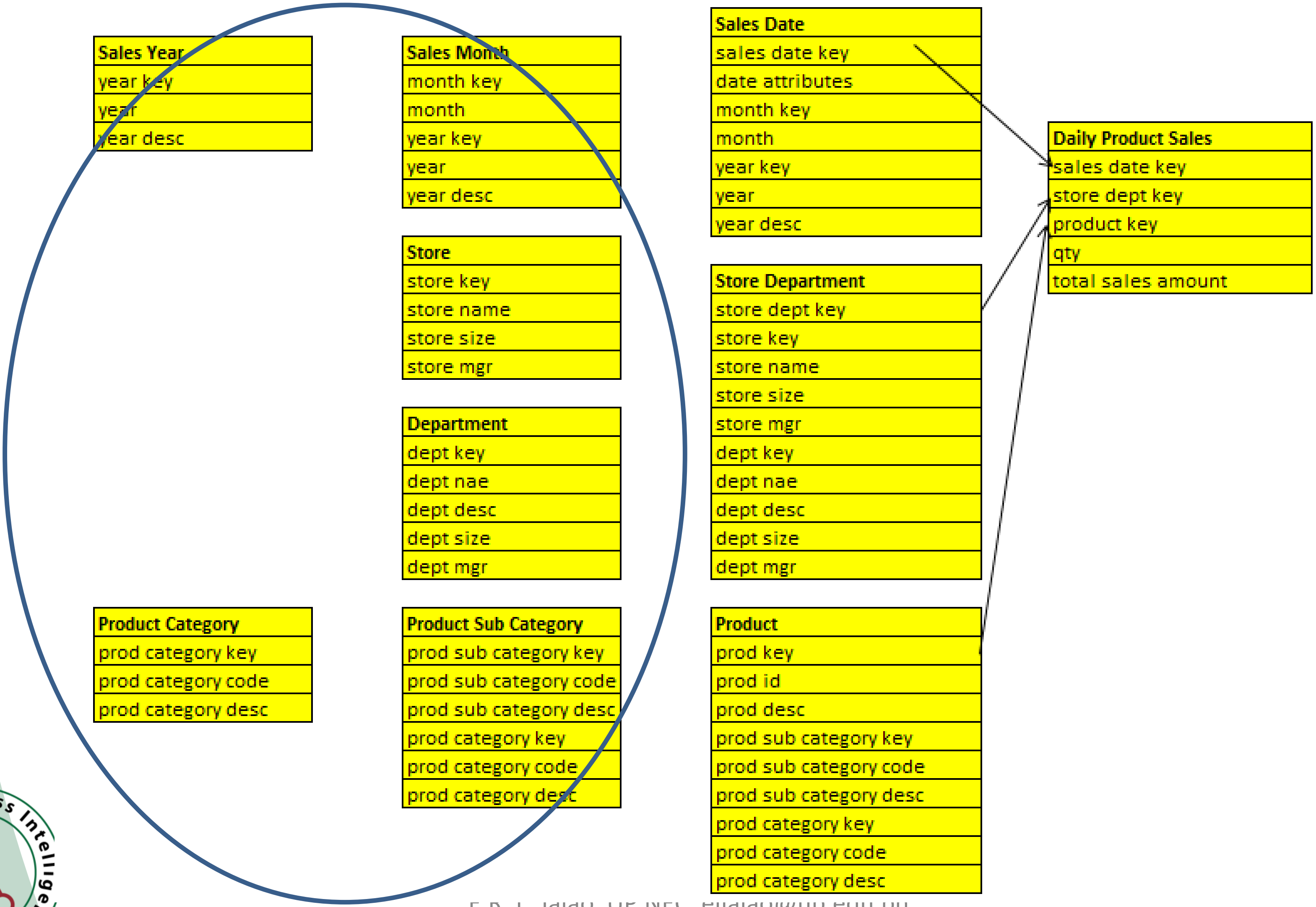
Sales Date
sales date key
date attributes
month key
month
year key
year
year desc

Store Department
store dept key
store key
store name
store size
store mgr
dept key
dept nae
dept desc
dept size
dept mgr

Product
prod key
prod id
prod desc
prod sub category key
prod sub category code
prod sub category desc
prod category key
prod category code
prod category desc

Daily Product Sales
sales date key
store dept key
product key
qty
total sales amount

DNF Complete Solution



Sales Year
year key
year
year desc

Sales Month
month key
month
year key
year
year desc

Store
store key
store name
store size
store mgr

Department
dept key
dept nae
dept desc
dept size
dept mgr

Product Category
prod category key
prod category code
prod category desc

Product Sub Category
prod sub category key
prod sub category code
prod sub category desc
prod category key
prod category code
prod category desc

Sales Date
sales date key
date attributes
month key
month
year key
year
year desc

Store Department
store dept key
store key
store name
store size
store mgr
dept key
dept nae
dept desc
dept size
dept mgr

Product
prod key
prod id
prod desc
prod sub category key
prod sub category code
prod sub category desc
prod category key
prod category code
prod category desc

Daily Product Sales
sales date key
store dept key
product key
qty
total sales amount

Outline for this Session

- Introduction to Dimension Tables
- Dimension Table Contents
- Dimension Table Types
- Generating Dimension Models
- **Case Example**
- Case Study

Dimensional Model Exercise: Hotel Business Overview

- Maitutulog Mo Kaya Hotel (MMK Hotel)
- Composed of 500 hotels
- Three property types (luxury, economy, budget)
- Different room types
- Want to maximize utilization
- Want to maximize profit

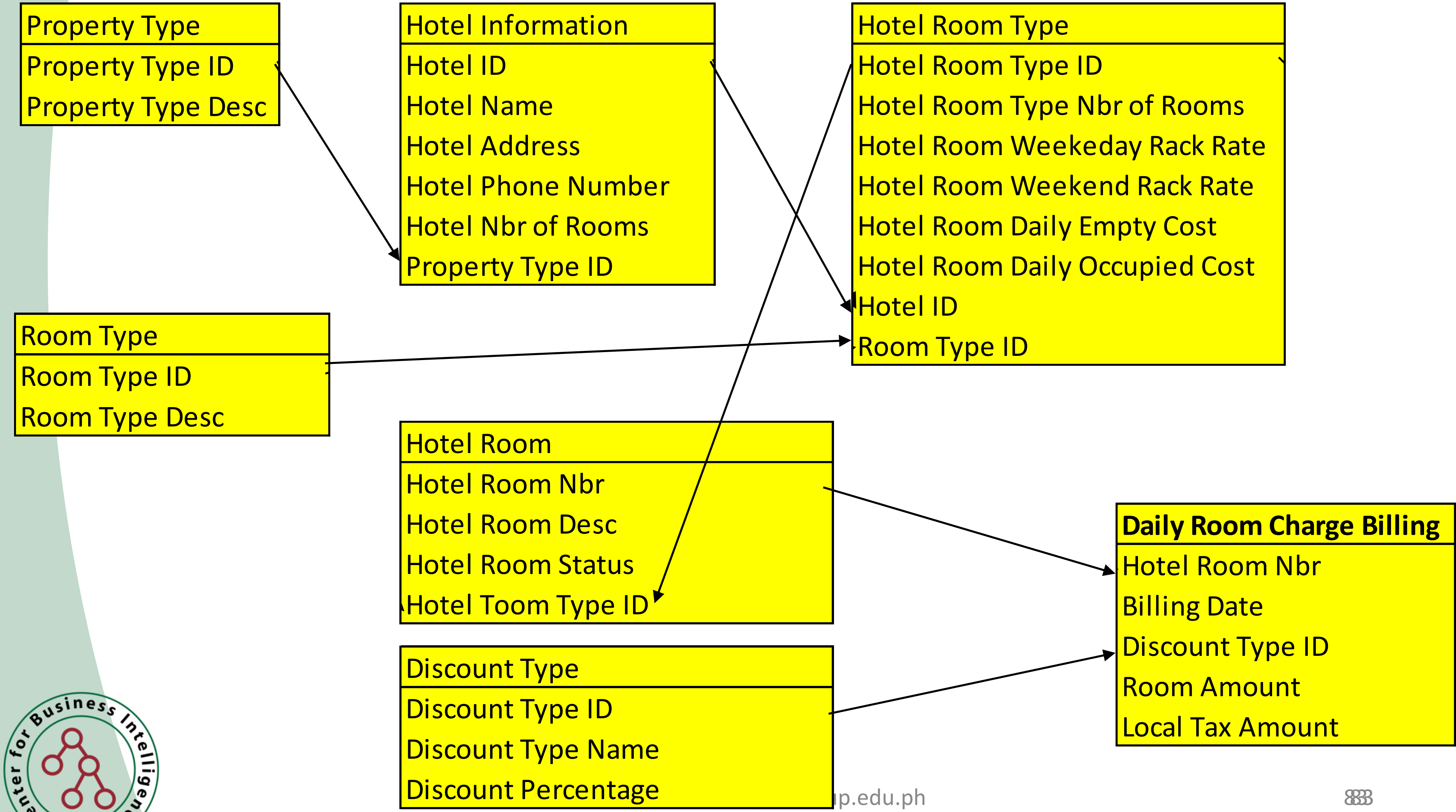
Business Strategic Planning

- Determine the most profitable hotel type and room type mixes
- Can I just build more of the most profitable hotel types or can I just put in more of the most profitable room types
- Is there a saturation point (not enough luxury hotel customers or people wanting luxury suites available in the area) that would make utilization and profit go down
- What is the competition doing that will cause a change in my occupancy rates
- How do rates, discounts, promotions, campaigns, and advertising effect my profit

Daily Room Type Profitability Analysis

- What room types have the highest profitability and which have the lowest profitability across the chain, by property type?
- Which hotels have room type profitability different from the norm?
- How does weekend profitability compare with weekday?
- How does weekday profitability differ by day?

Hotel Property Management System
Source Data 3NF



Property Type
- Property Type ID
- Property Type Desc

Hotel Information
- Hotel ID
- Hotel Name
- Hotel Address
- Hotel Phone Number
- Hotel Nbr of Rooms
- Property Type ID

Hotel Room Type
- Hotel Room Type ID
- Hotel Room Type Nbr of Rooms
- Hotel Room Weekeday Rack Rate
- Hotel Room Weekend Rack Rate
- Hotel Room Daily Empty Cost
- Hotel Room Daily Occupied Cost
- Hotel ID
- Room Type ID

Room Type
- Room Type ID
- Room Type Desc

Hotel Room
- Hotel Room Nbr
- Hotel Room Desc
- Hotel Room Status
- Hotel Toom Type ID

Discount Type
- Discount Type ID
- Discount Type Name
- Discount Percentage

Daily Room Charge Billing
- Hotel Room Nbr
- Billing Date
- Discount Type ID
- Room Amount
- Local Tax Amount

Challenge

- Build the LDMs for Daily Room Billing and Daily Utilization and Profitability

Outline for this Session

- Introduction to Dimension Tables
- Dimension Table Contents
- Dimension Table Types
- Generating Dimension Models
- Case Example
- **Case Study**

Case Study 2

- Northwind Database

Outline for this Session

- Introduction to Dimension Tables
- Dimension Table Contents
- Dimension Table Types
- Generating Dimension Models
- Case Example
- Case Study

References

- Kimball, Ralph, Margy Ross, Warren Thornthwaite, Joy Mundy, and Bob Becker, *The Data Warehouse Life Cycle Toolkit, Second Edition*, Wiley, 2008, ISBN 978-0-470-14977-5
- Schmitz, Michael D. UCI Irvine Data Warehousing Notes (2014), High Performance Data Warehousing
- Simon, Alan. CIS 391 PPT Slides
- Jeltema ,Bernie, UCI Irvine Data Warehousing Notes (2014), Strategic Frameworks, Inc.