IIT School of Applied Technology

ILLINOIS INSTITUTE OF TECHNOLOGY

**information technology & management**

# 526 Data Warehousing

March 2, 2016

Week 7 Presentation

# Week 07 Topic: Dimensional Modeling: More Dimension Patterns and Considerations



- We will cover
  - Design Workshop #2: Enterprise Data Warehouse Bus Matrix
  - Fact Attributes or Dimension Attributes?
  - Dealing with Rapidly Changing Monster Dimensions: Mini Dimensions
  - Outriggers
  - Resolving Multivalued Relationships using Bridge Tables
  - Design Workshop #3: Design Review Exercise

# Design Workshop #2:
# Enterprise Data Warehouse Bus Matrix



Handouts

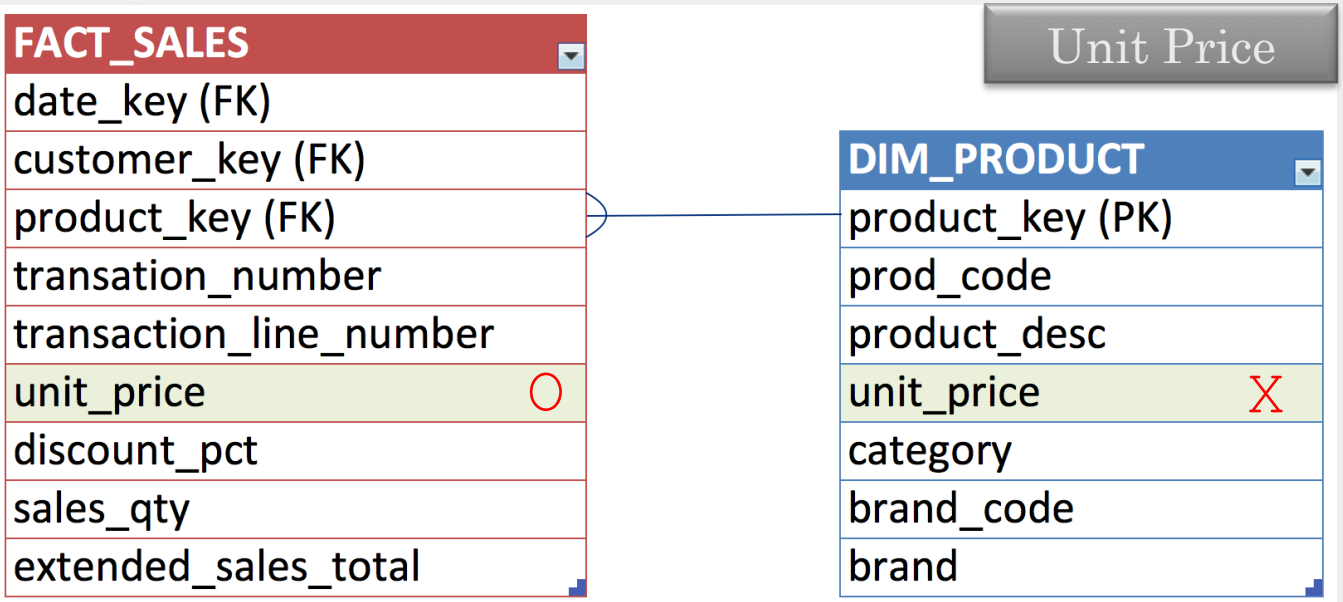
(also available on BB)

Content › Week 06 Introduction to Assignment 01 › **Class Excercise**

# Fact Attributes or Dimension Attributes?

- Fact Attributes
  - Numerical measurements
  - Pertain implicit time series of observations
  - Participate in numerical computations (sum, averages, etc.)
- Dimension Attributes
  - Textual descriptors
  - Targets of constraints
  - Provide the content of "row headers" (grouping columns)

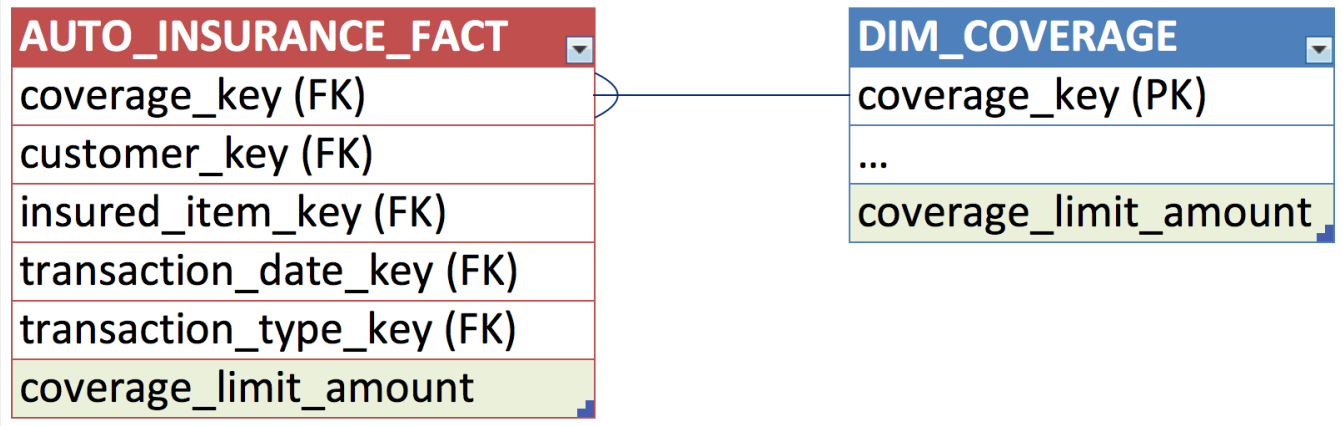
# Fact Attributes or Dimension Attributes? (cont'd)




- The unit price varies over time and over location
- It is a rather rapidly changing, not a good fit for a SCD Type 2 dimension attribute
- It is not a good row header item as it is a continuous (not discrete) value
- Thus, it is a fact attribute

# Fact Attributes or Dimension Attributes? (cont'd)

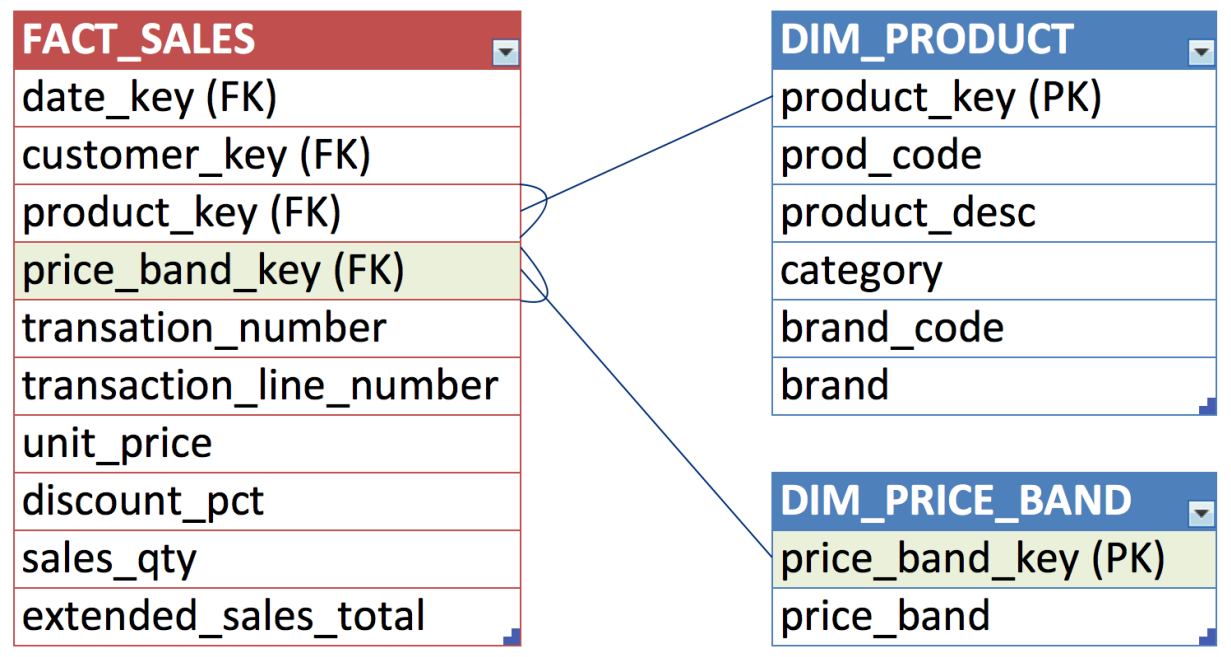
Coverage Limit Amount




- The coverage limit participates as a query constrain
- It is generally a discrete value ($300k, $400k, ....)
- It slowly changes over time
- It participates in computations such as sum, average on policies and coverage

# Fact Attributes or Dimension Attributes? (cont'd)

- For a truly continuous numeric dimension attribute, a value band can be an excellent alternative
  - unit price, GRE score, credit score, etc.



# Dealing with Rapidly Changing Monster Dimensions: SCD Type 2 Revisited



- Changes in dimensions arrive
  - Far less frequently than fact table measurements → Slow changing
  - Type 2: Insert a new dimension row with the new data and new effective date

| Supplier_Key | Supplier_Code | Supplier_Name | Supplier_State | Start_Date | End_Date |
|---|---|---|---|---|---|
| 123 | ABC | Acme Supply Co | CA | 2000-01-01 | 2004-12-21 |

Type 2

| Supplier_Key | Supplier_Code | Supplier_Name | Supplier_State | Start_Date | End_Date |
|---|---|---|---|---|---|
| 123 | ABC | Acme Supply Co | CA | 2000-01-01 | 2004-12-21 |
| 124 | ABC | Acme Supply Co | IL | 2004-12-22 | 2999-12-31 |

| Customer_Key | Customer_ID | Customer_Birth | Supplier_State | Start_Date | End_Date |
|---|---|---|---|---|---|
| 332 | C01 | 1977-08-01 | CA | 2000-01-01 | 2004-12-21 |

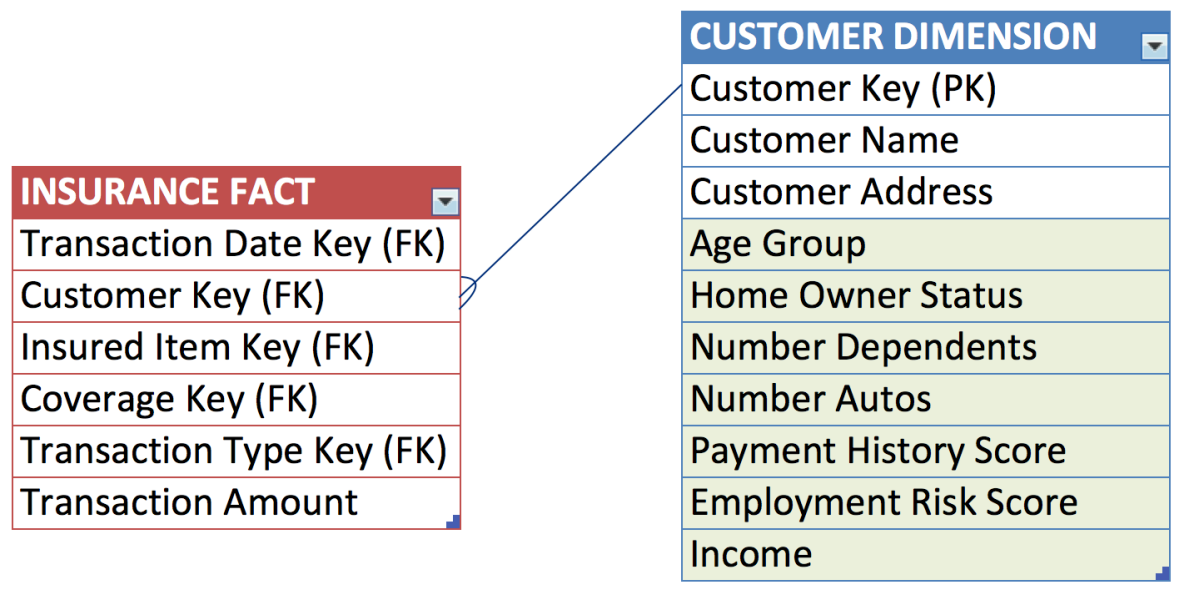
Type 1 Type 2

| Customer_Key | Customer_ID | Customer_Birth | Supplier_State | Start_Date | End_Date |
|---|---|---|---|---|---|
| 332 | C01 | 1977-09-01 | CA | 2000-01-01 | 2004-12-21 |
| 333 | C01 | 1977-09-01 | IL | 2004-12-22 | 2999-12-31 |

# Dealing with Rapidly Changing Monster Dimensions: Monster Dimensions



- Imagine an insurance company with a big customer dimension (e.g. 30 million) with rapidly changing demographics (in green)

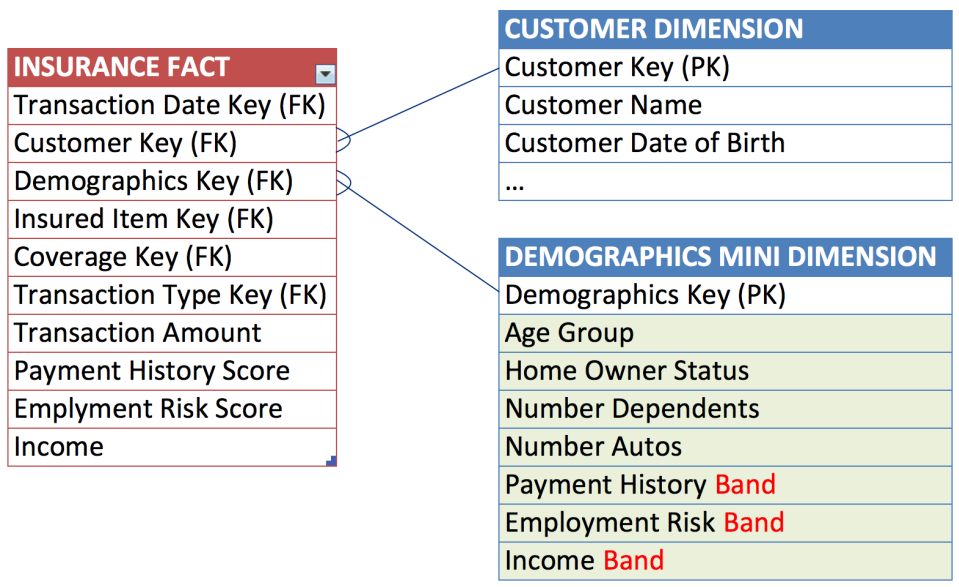



- The dimension table size can be easily doubled within a short period making this a rapidly changing monster dimension

# Dealing with Rapidly Changing Monster Dimensions: Monster Dimensions (cont'd)



- The solution is to break off the hot attributes into their own separate mini dimension




- The mini dimension contains one row for each possible combination of the attributes
- Value bands are used in the mini-dimension to reduce the number of rows overall

# Dealing with Rapidly Changing Monster Dimensions: Monster Dimensions (cont'd)



**Customer dimension sample row:**

| Customer Key | Customer Name | Date of Birth |
|---|---|---|
| 123456 | John Smith | 1984-02-10 |

**Demographics mini-dimension sample row:**

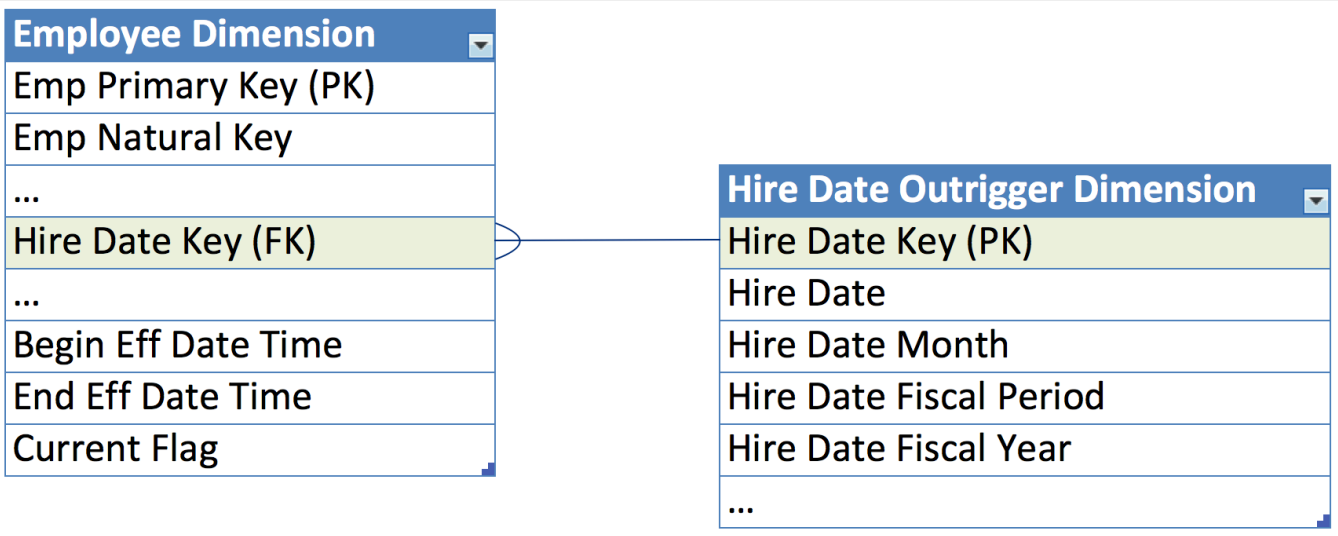
| Demographics Key | Age Group | Income Band |
|---|---|---|
| 1 | 25-29 | $50,000 - $59,999 |
| 2 | 30-34 | $50,000 - $59,999 |
| 3 | 30-34 | $60,000 - $69,999 |

**Fact table sample row:**

| Transaction Date Key | Customer Key | Demographics Key |
|---|---|---|
| 20140131 | 123456 | 1 |
| 20140228 | 123456 | 2 |
| 20140331 | 123456 | 2 |
| 20140430 | 123456 | 3 |

# Outriggers

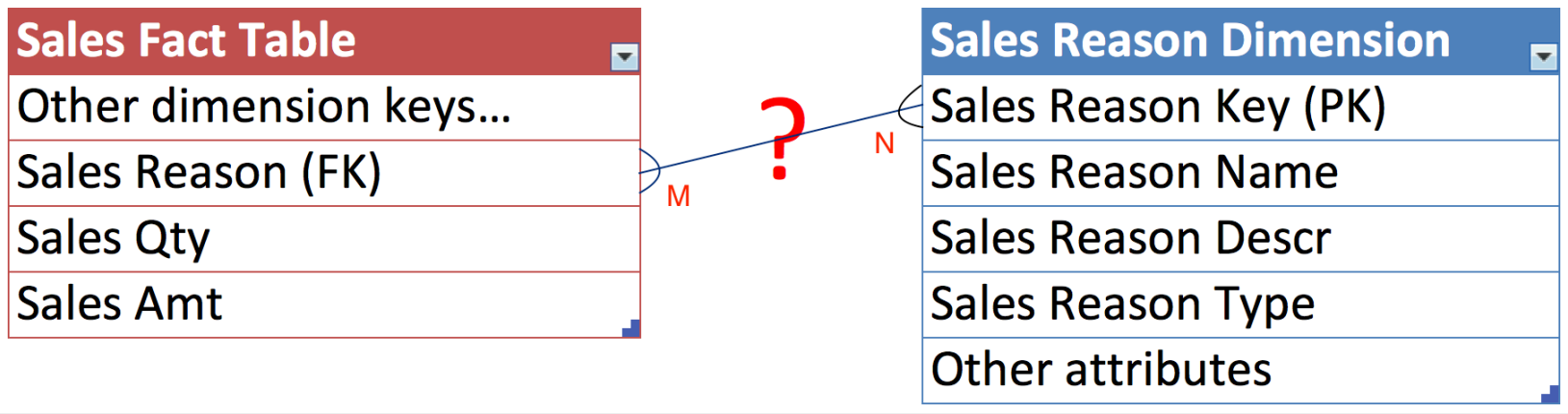
- Dimension tables joined to other dimension tables
- In this case, a date dimension serves as an outrigger to the employee dimension via role-playing
- Outriggers are acceptable in moderation but should be viewed as the exception rather than the rule



# Resolving Multivalued Relationships Using Bridge Tables



- In a classic dimensional schema, each dimension attached to a fact table has a single value consistent with the fact table's grain
- But there are a number of situations in which a dimension is legitimately *multivalued*



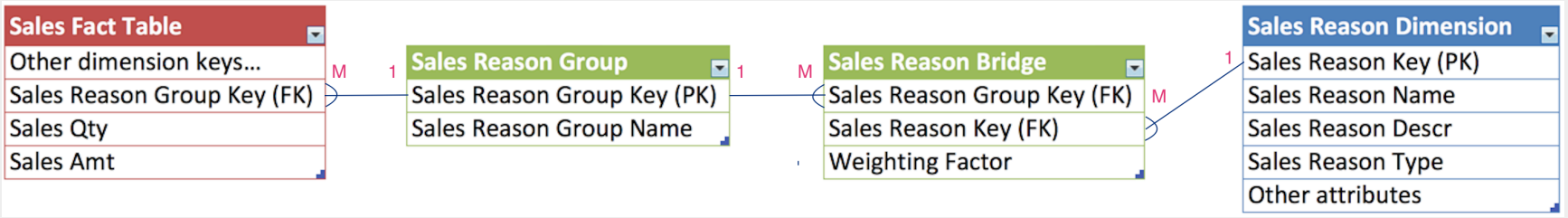
# Multivalued Dimension Examples



- Many sales reasons on a single transaction
- Many customers in a bank account
- Many diagnoses at the time of a treatment
- Many witnesses to an accident
- Many options on a car

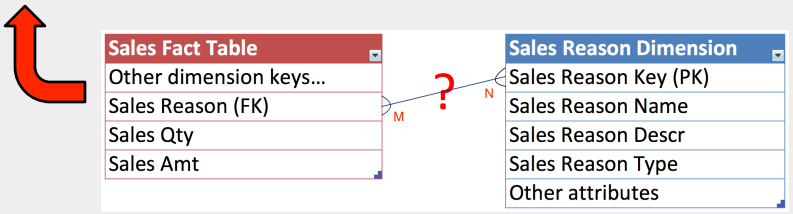
# Resolving Multivalued Relationships Using Bridge Tables

## Multivalued Sales Reasons Bridge




**Sample rows from Sales Reason Group**

| Sales Reason Group Key | Sales Reason Group Name |
|---|---|
| 101 | Product Quality |
| 102 | Product Quality; Promotion |
| 103 | Product Quality; Convenient Location |




**Sample rows from Sales Reason Bridge**

| Sales Reason Group Key | Sales Reason Key | Weighting Factor |
|---|---|---|
| 101 | 1 | 1 |
| 102 | 2 | 0.5 |
| 102 | 3 | 0.5 |
| 103 | 1 | 0.5 |
| 103 | 2 | 0.5 |

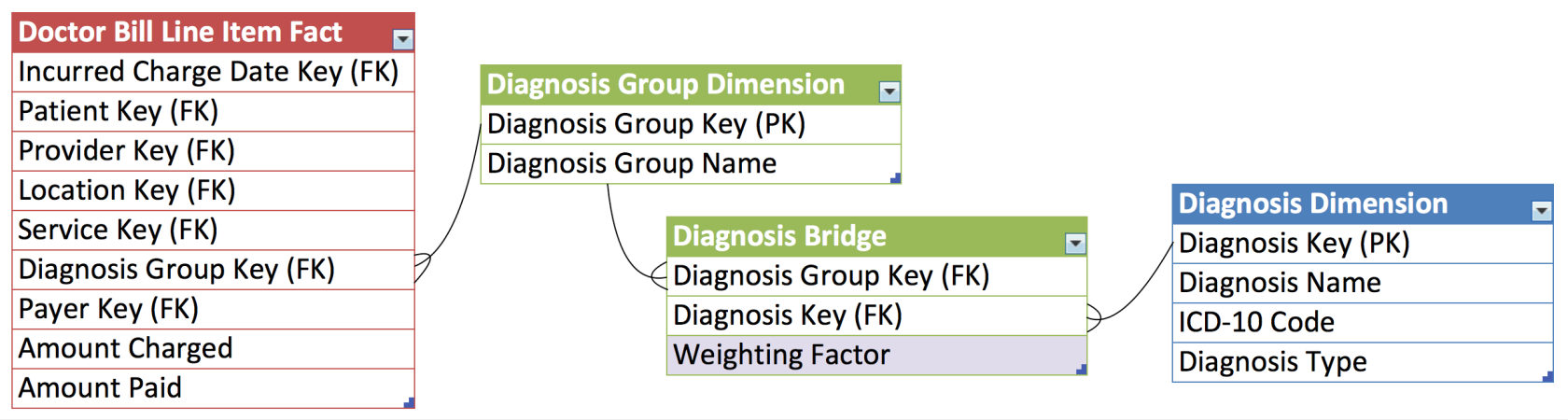
**Sample rows from Sales Reason Dimension**

| Sales Reason Key | Sales Reason Name | Other attributes |
|---|---|---|
| 1 | Product Quality | ... |
| 2 | Promotion | ... |
| 3 | Convenient Location | ... |

The Sales Reason Group table may be required by your modeling tool to resolve FK/PK relationships. It provides no useful information at query time and is often omitted

# Resolving Multivalued Relationships Using Bridge Tables

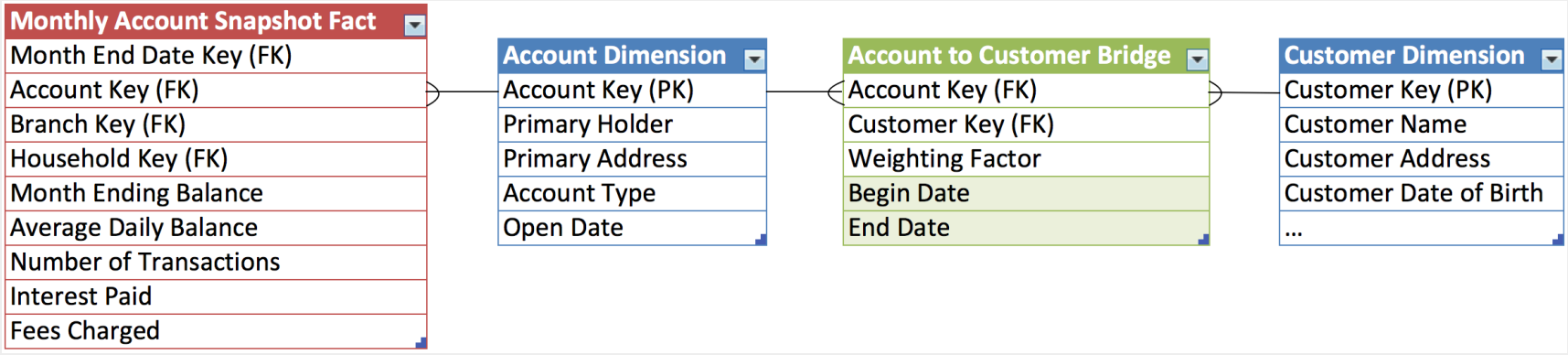
## Multivalued Diagnosis Bridge




- The weighting factor is an explicit allocation
- Records in the Diagnosis Group Dimension can be made for each patient, but in this case it seems reasonable to re-use diagnosis groups, especially for out patient treatments where many groups would be repeated

# Resolving Multivalued Relationships Using Bridge Tables

## Bank Account to Customer Bridge




- Associate customers to accounts where these have a many-to-many relationship
- Query account balances by individual customer or groups of customers
- Show account balances correctly weighted (prorated) by individual customers to avoid double counting
- Show account balances by customer "impact" (un-weighted)

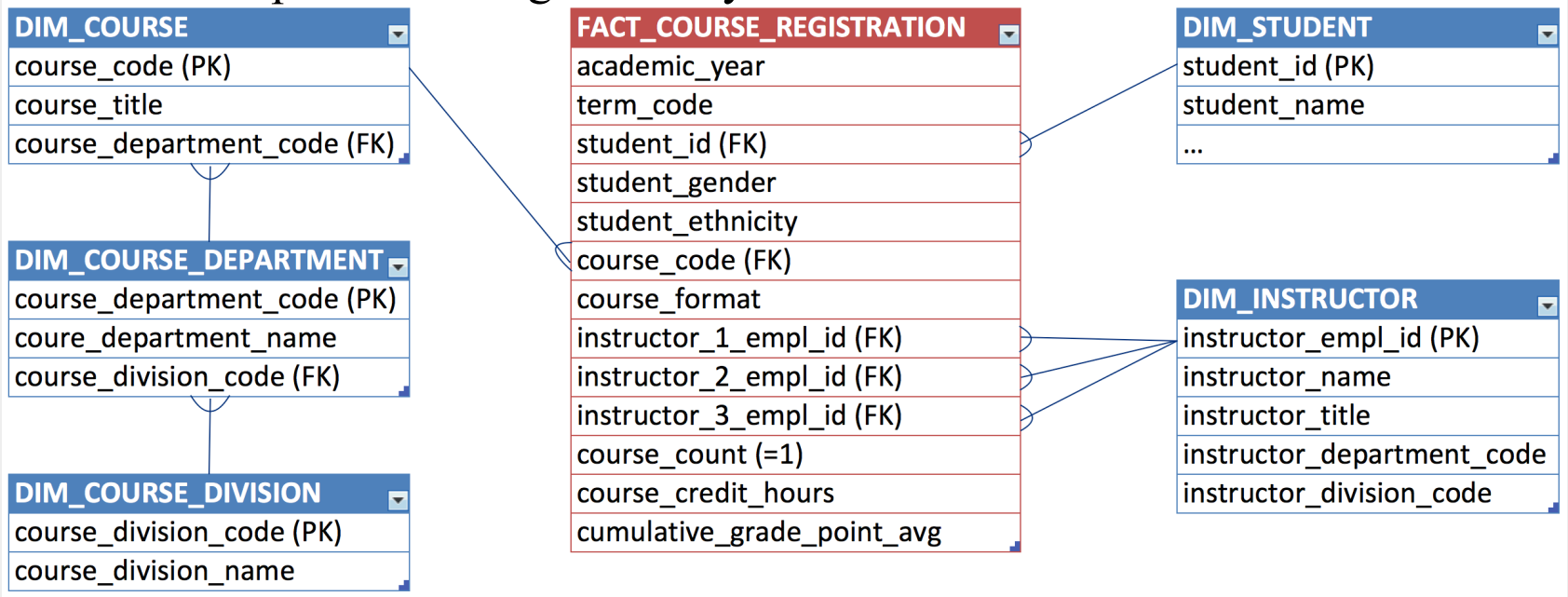
# Design Workshop #3: Design Review Exercise

# Identify Potential Design Flaws



Business Process: Student/Course snapshot

Grain: 1 row per course registered by student for on each term




Sample fact rows:

| Academic Year | Term Code | Student ID | Student Gender | Student Ethnicity | Course Code | Course Format | Instructor 1 Eml ID | Instructor 2 Eml ID | Instructor 3 Eml ID | Couse Count | Course Credit Hours | Student Cum Grade Point |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2014-2015 | FALL | 1234 | F | H | ECON101 | LECT | SR123 | | | 1 | 4 | 3.50 |
| 2014-2015 | FALL | 1234 | F | H | GOVT201 | LECT | PW456 | BB789 | | 1 | 4 | 3.50 |
| 2014-2015 | FALL | 1234 | F | H | CHEM103 | LAB | KS246 | NR468 | | 1 | 6 | 3.50 |
| 2014-2015 | FALL | 1234 | F | H | YOGA101 | SEM | KV680 | | | 1 | 2 | 3.50 |
| 2014-2015 | SPRING | 1234 | F | H | GOVT102 | LECT | SR123 | PW456 | | 1 | 4 | 3.55 |

# Week 07 Topic: Dimensional Modeling: More Dimension Patterns and Considerations



Questions?