IIT School of Applied Technology

ILLINOIS INSTITUTE OF TECHNOLOGY

information technology & management

526 Data Warehousing

March 9, 2016

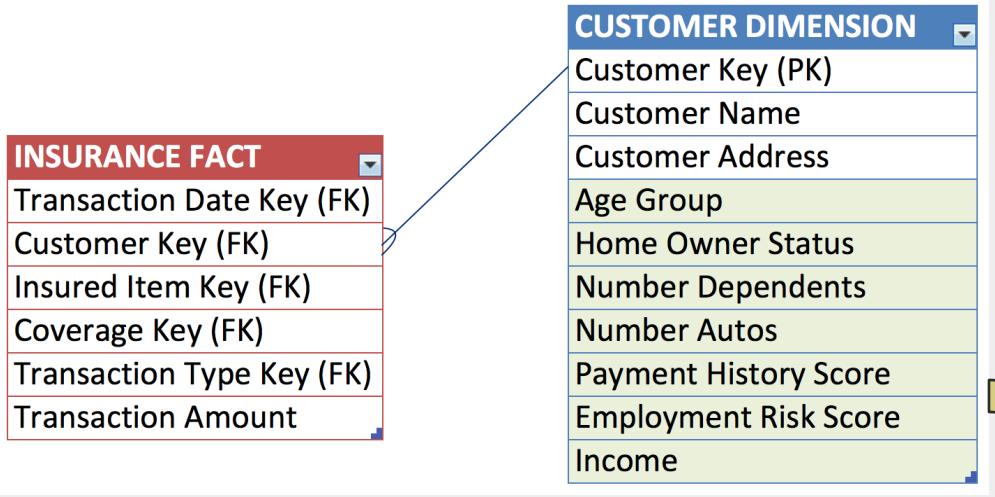
Week 8 Presentation

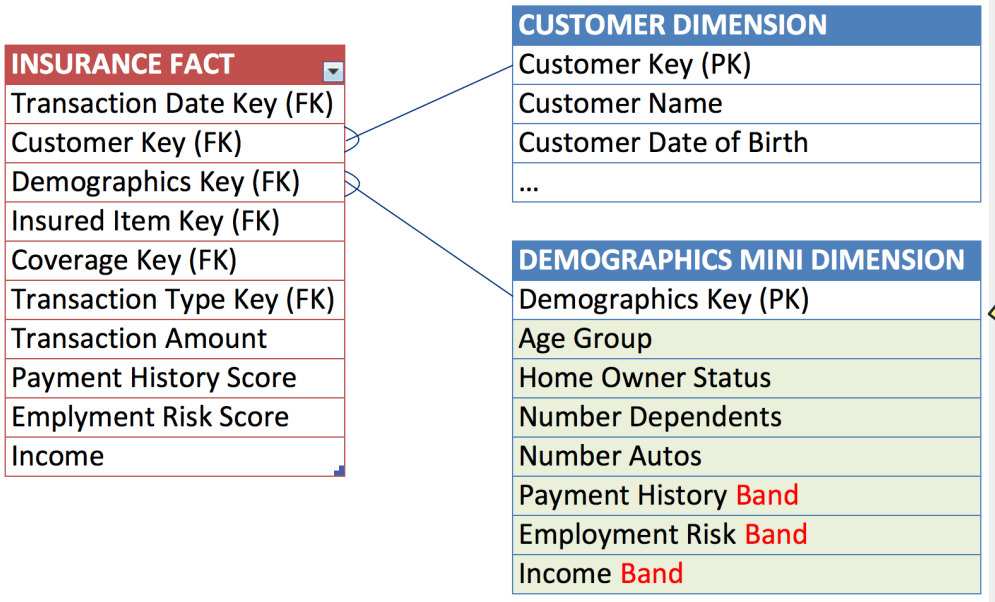
Week 08 Topic: Dealing with Hierarchies

- We will cover
  - Mini Dimensions Revisited
  - Resolving Multivalued Relationships using Bridge Tables
  - Design Workshop #3: Design Review Exercise
  - Post-Spring Break Agenda Preview

ITMD - 526

Dealing with Monster Dimensions: Mini-Dimension to the Rescue





Break off the hot attributes into their own separate mini dimension

It has one row for each possible combination of the attributes

Value bands are used 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

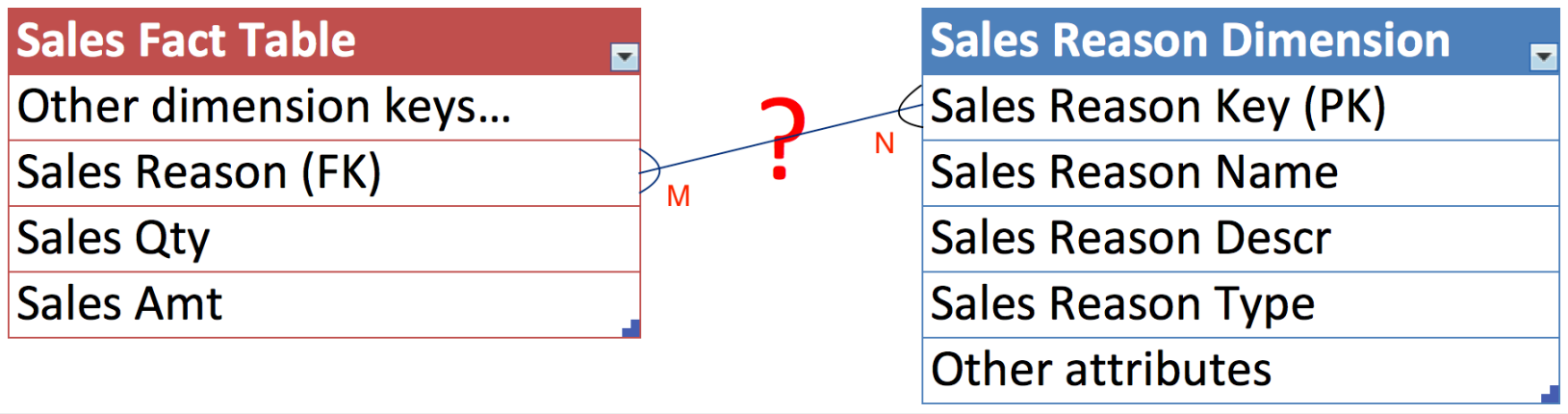
Dimension Value Band Example

Sugar Level	Daily Calories Burned	Age Group
50-65	0-500	0 to 10
66-80	501-1000	11 to 20
81-95	1001-1500	21 to 30
96-100	1501-2000	31 to 40
101-115	2001-2500	41 to 50
116-130	2501-3000	51 to 60
131-145	3001-3500	61 to 70
146-160	3501-4000	71 to 80

Source: http://www.webdetails.pt/pentaho/api/repos/BioMe/webapp/index.html?userid=pentaho&password=demo#/analytics/refine

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*



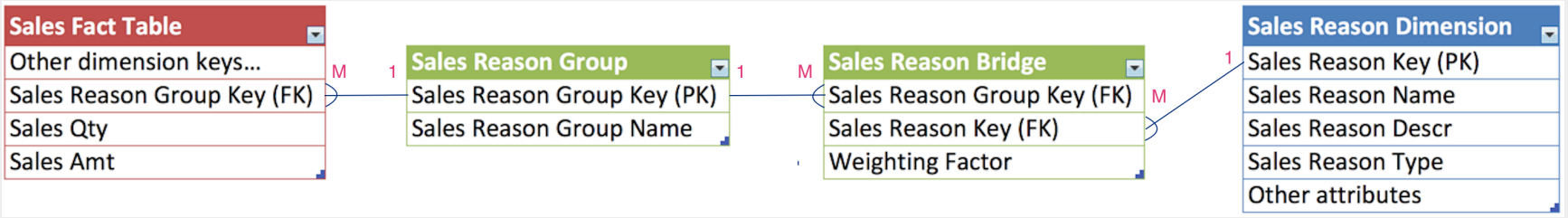
Resolving Multivalued Relationships Using Bridge Tables

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

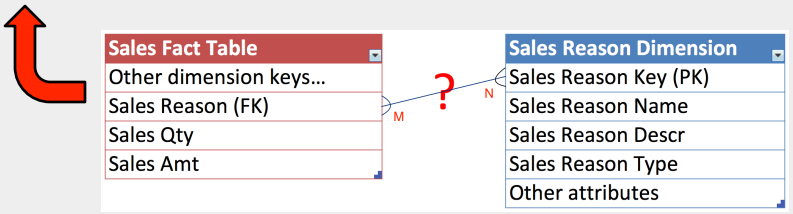
Resolving Multivalued Relationships Using Bridge Tables

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

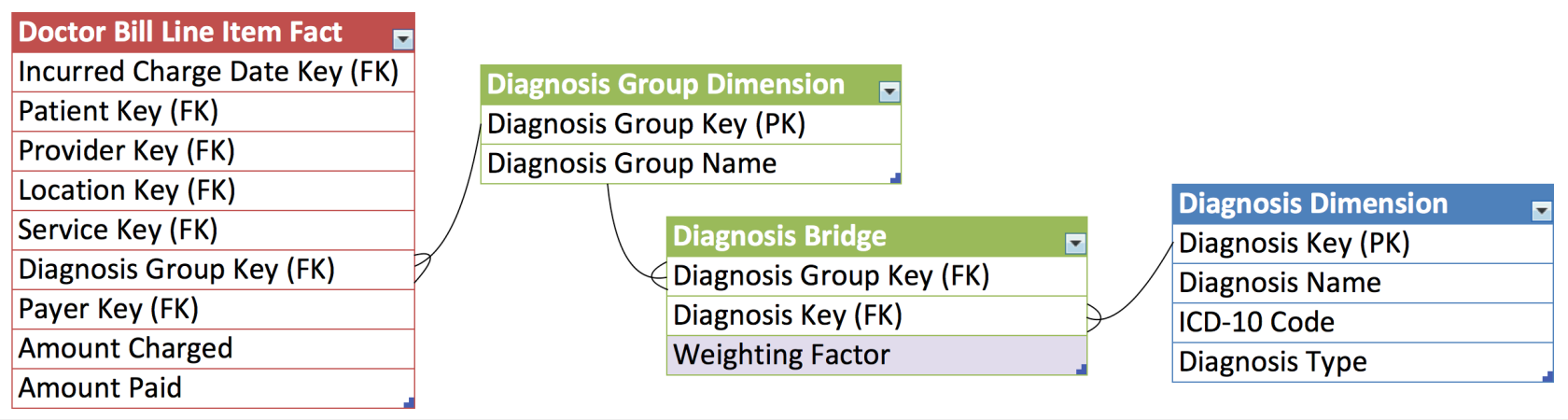
Sample rows from Sales Reason Dimension

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

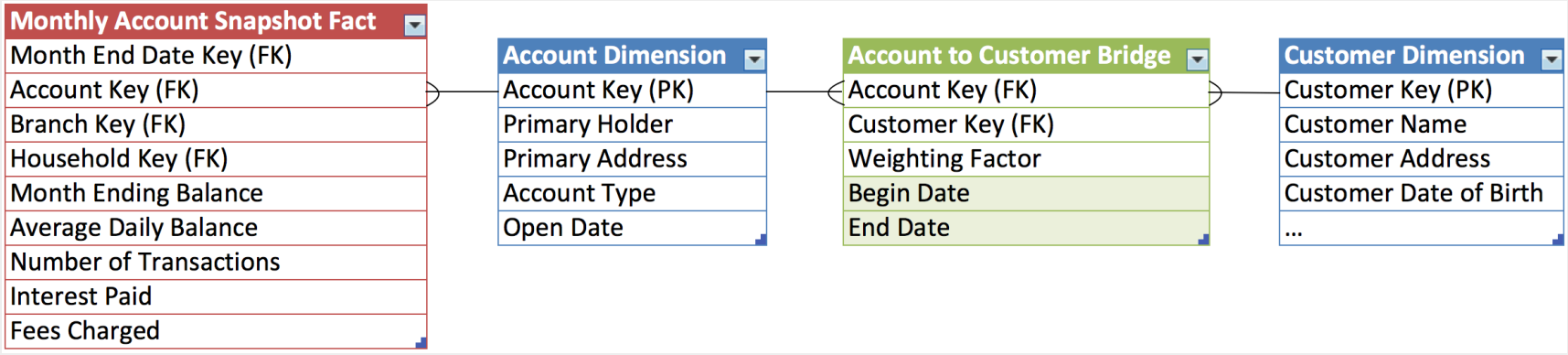
Multivalued Diagnosis Bridge



- 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)

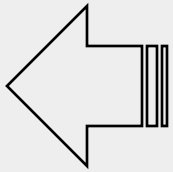
ITMD - 526

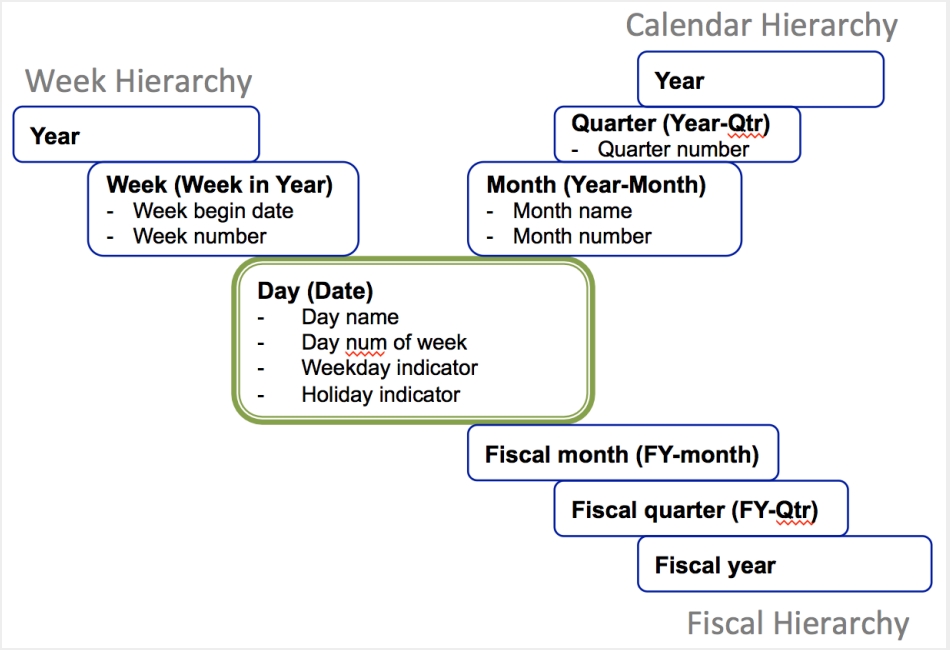
Dealing with Hierarchies

Fixed Depth Positional Hierarchies

- A series of many-to-one relationships
- The hierarchy levels become separate positional attributes in a dimension table
- A dimension can have multiple fixed depth hierarchies in it

Date Dimension
Day
Week
Month
Quarter
Year
Fiscal Month
Fiscal Quarter
Fiscal Year



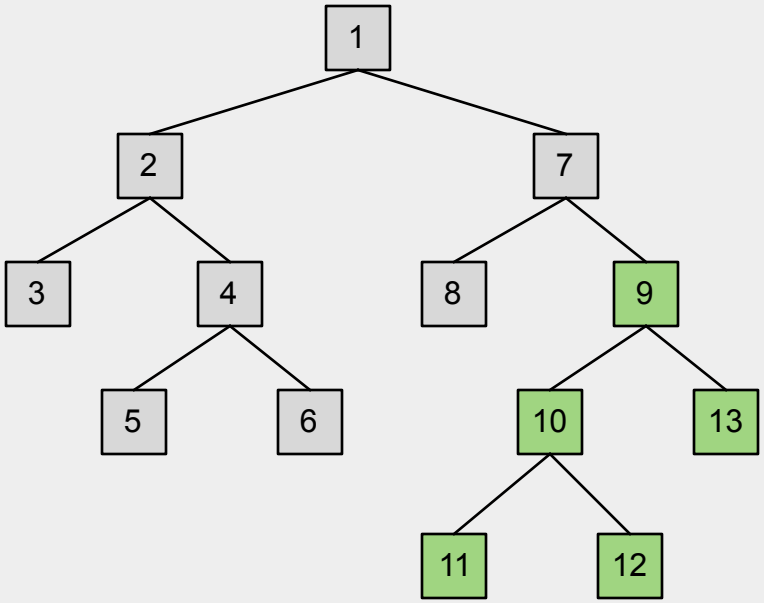


ITMD - 526

Dealing with Hierarchies

Ragged/Variable Depth Hierarchies

- Ragged hierarchies of indeterminate depth are difficult to model and query in a relational database
- A specially constructed bridge table to the rescue



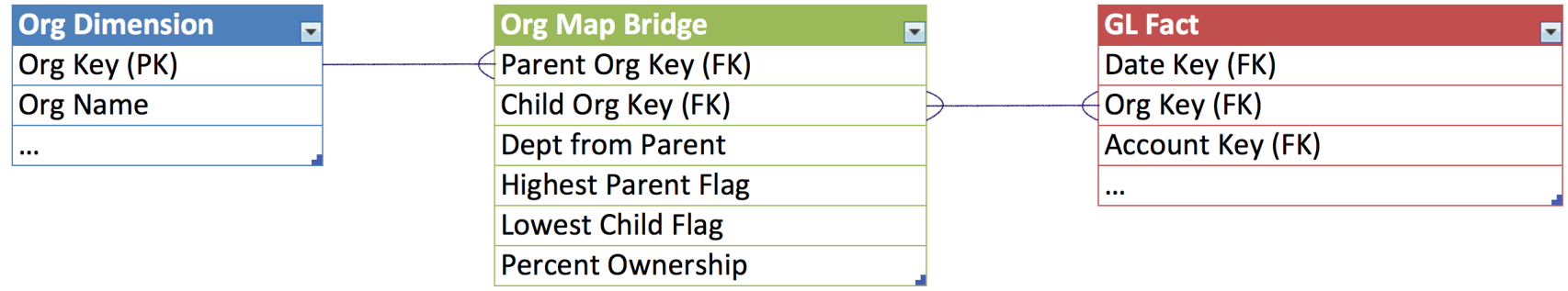
Parent Key	Child Key	Depth From Parent	Highest Parent Flag	Lowest Child Flag
9	9	0	FALSE	FALSE
9	10	1	FALSE	FALSE
9	11	2	FALSE	TRUE
9	12	2	FALSE	TRUE
9	13	1	FALSE	TRUE
10	10	0	FALSE	FALSE
10	11	1	FALSE	TRUE
10	12	1	FALSE	TRUE
11	11	0	FALSE	TRUE
12	12	0	FALSE	TRUE
13	13	0	FALSE	TRUE

ITMD - 526

Dealing with Hierarchies (cont'd)

Ragged/Variable Depth Hierarchies

- Child element references to the fact
- Parent element references to the dim




```
SELECT SUM(f.balance * m.percent_ownership)
  FROM fact_gl f,
       bridget_org_map b,
       dim_org d
 WHERE d.org_name        = 'Divisional Rollup'
   AND f.org_key         = b.child_org_key
   AND d.org_key         = b.parent_org_key
```

Design Workshop #3: Design Review Exercise
Identify Potential Design Flaws

- Guidelines to Identify Design Flaws
  - What's the Grain?
  - Mixed-Grain or Textual Facts?
  - Dimension Descriptors and Decodes?
  - Explicit Date Dimension?
  - Surrogate Keys?
  - Handling of Hierarchies?
  - Slowly Changing Dimension Strategies?
  - Well-Understood Business Requirements?

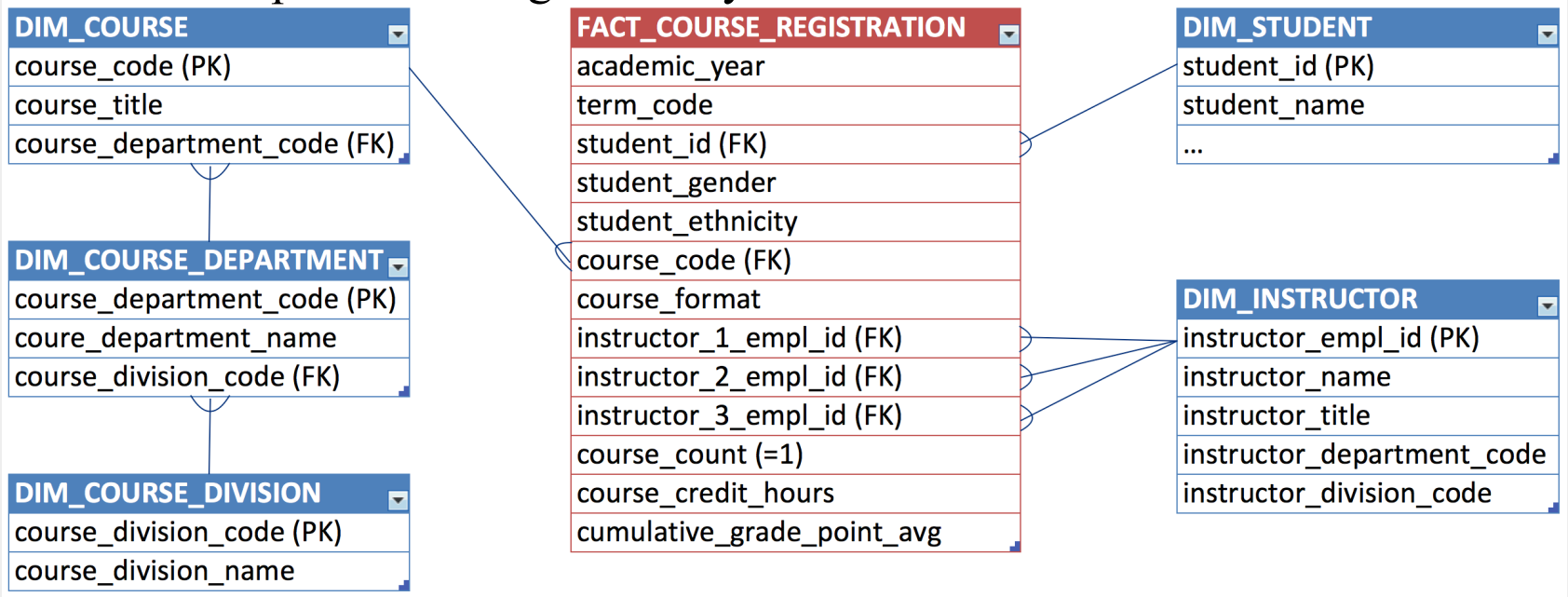
Source: http://www.kimballgroup.com/2003/10/fistful-of-flaws/

Design Workshop #3: Design Review Exercise

Identify Potential Design Flaws (cont'd)

Business Process: Student/Course snapshot

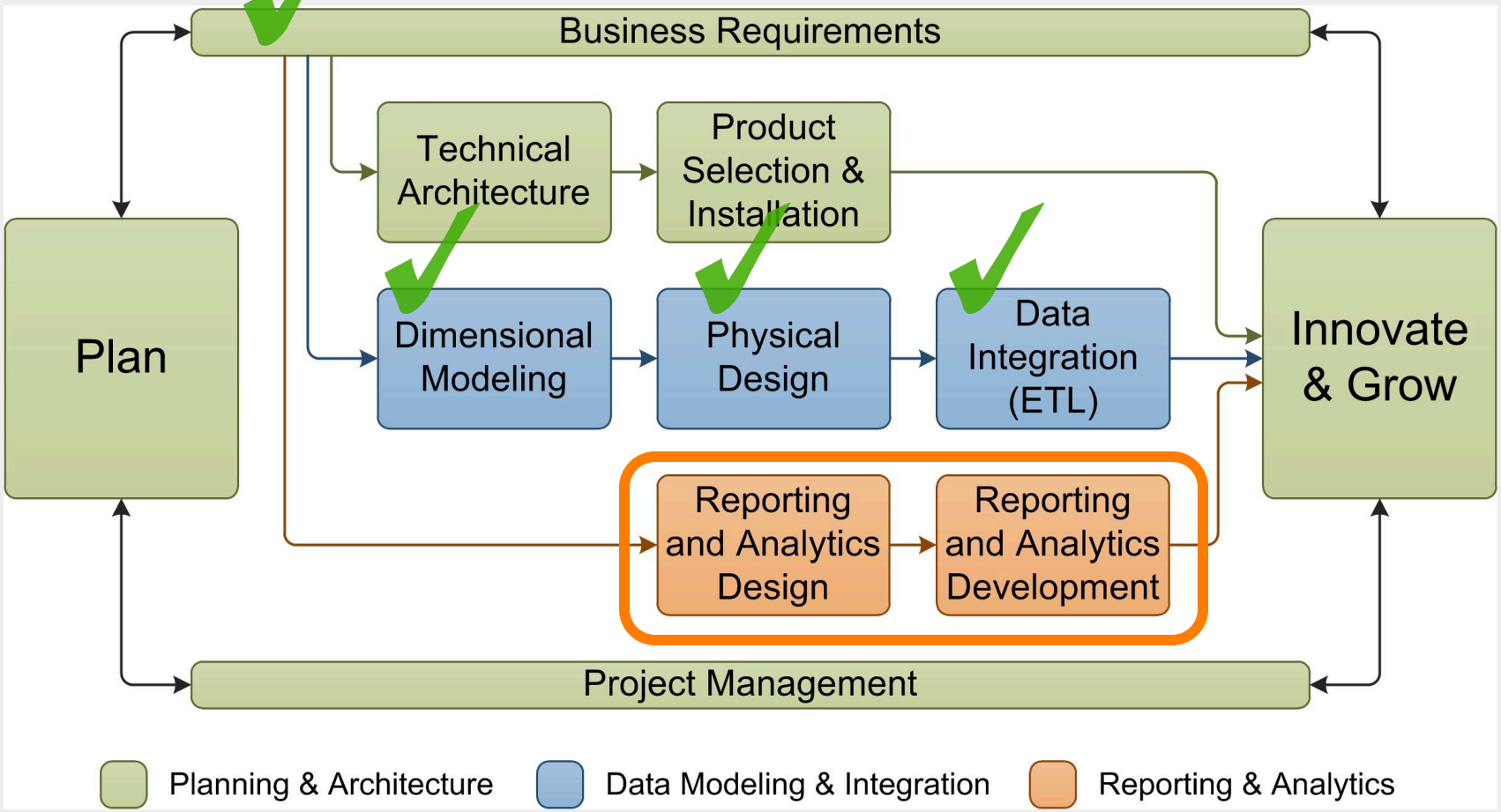
Grain: 1 row per course registered by student for on each term



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

Kimball Lifecycle Approach



What to Expect After Spring Break

- More Dimensional Modeling Tips & Exercises
- SQL Optimization for DW/BI
- BI Reporting and Analytics Development (Mondrian, Saiku, Tableau, Pentaho BI Server, etc.)
- The Future of DW/BI
- Assignment 02
- Final Exam
  - Important: Revisit past weeks' Readings & Glossary on BB

Week 08 Topic: Dealing with Hierarchies

Questions?