IIT School of Applied Technology

ILLINOIS INSTITUTE OF TECHNOLOGY

information technology & management

526 Data Warehousing

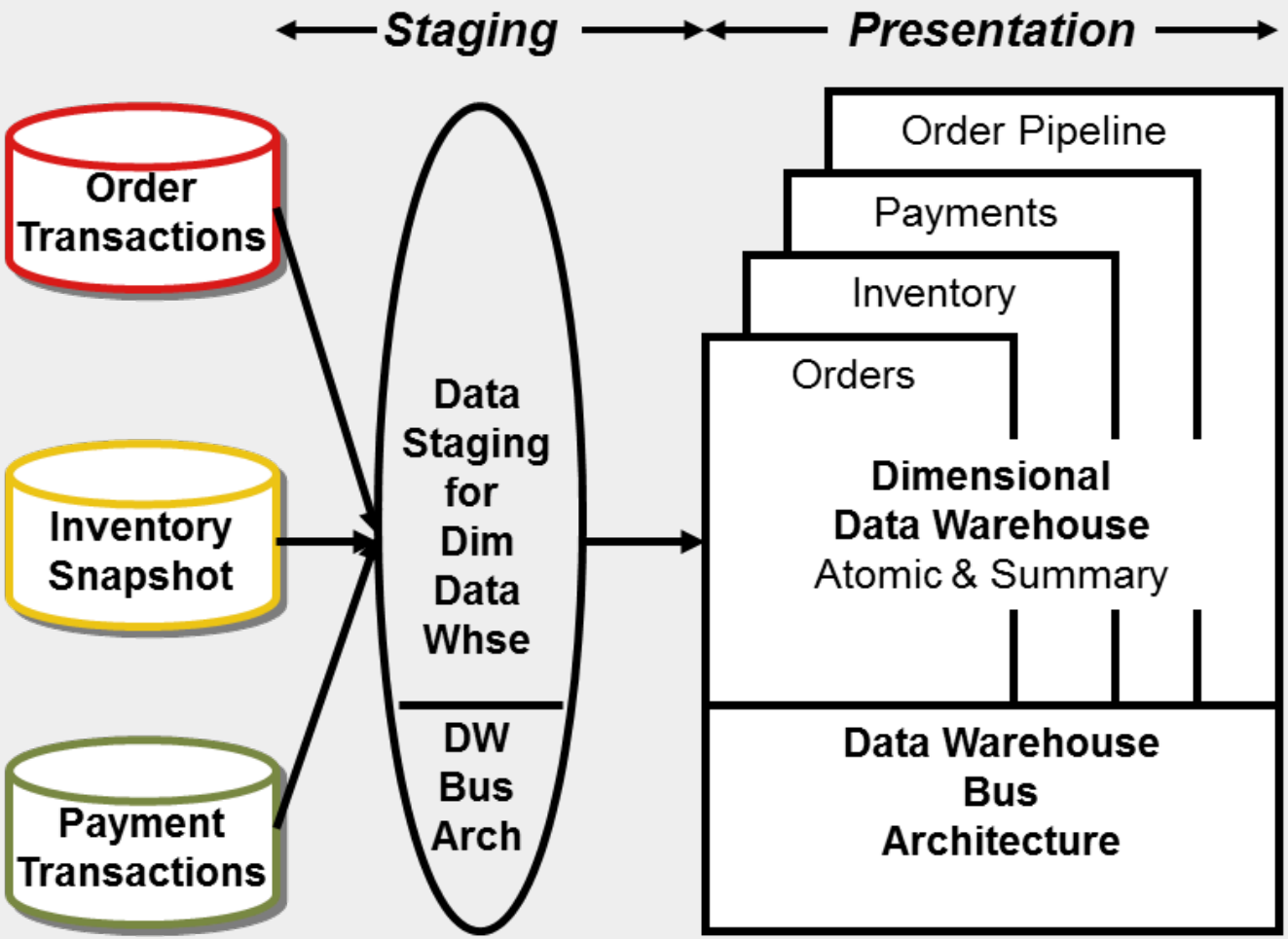
February 17, 2016

Week 5 Presentation

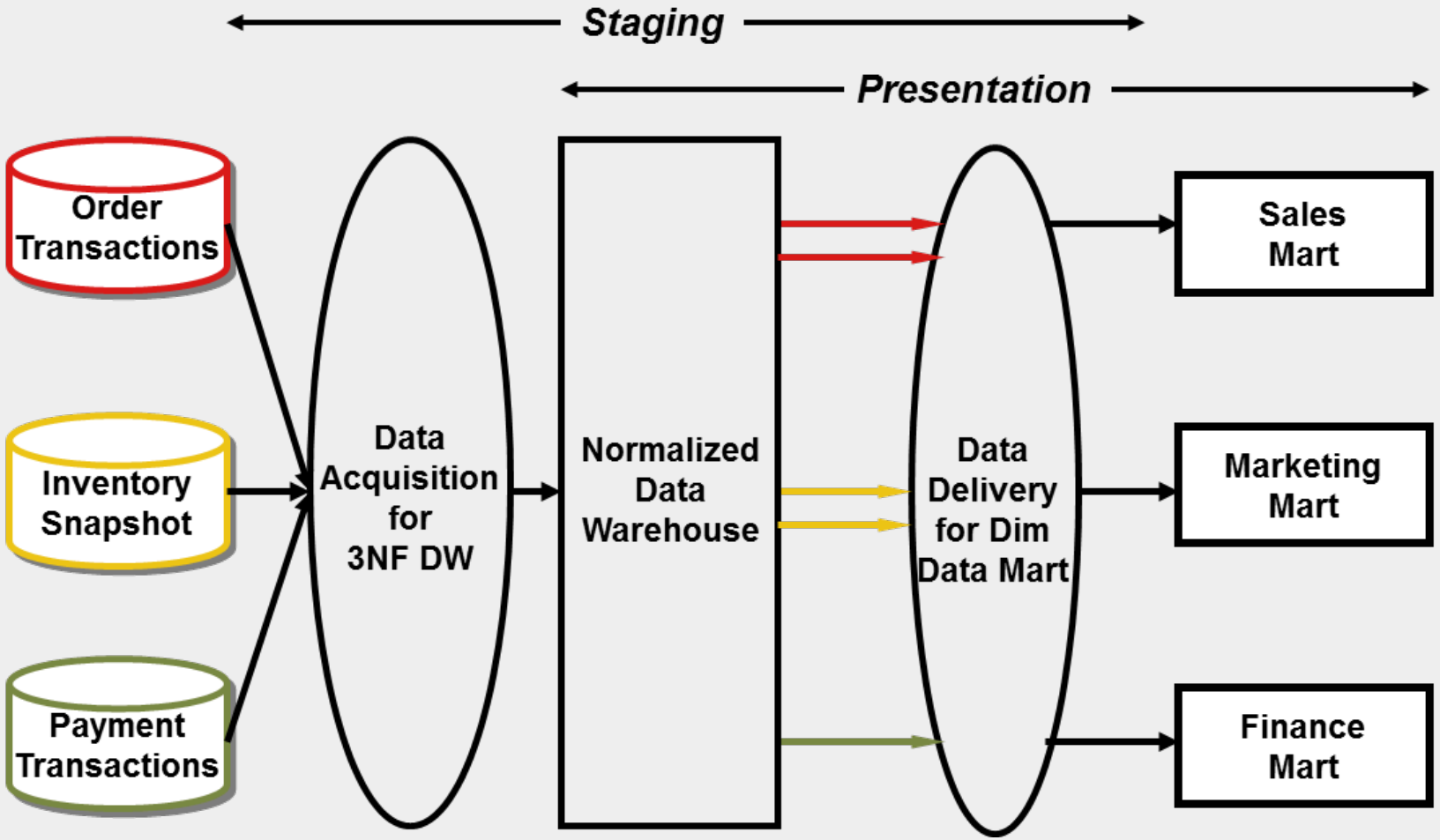
Week 05 Topic: Dimensional Modeling: Integration Via Conformed Dimensions

- We will cover
  - Kimball's and Inmon's Architectures Comparison Revisited
  - Denormalization to Avoid Snowflake/Centipede
  - Default/Dummy Row in Dimension Tables
  - Design Workshop #2: Header/Line Item Transactions
  - Conformed Dimensions
  - Enterprise Data Warehouse Bus Architecture
  - Enterprise Data Warehouse Bus Matrix

Kimball Architecture Revisited

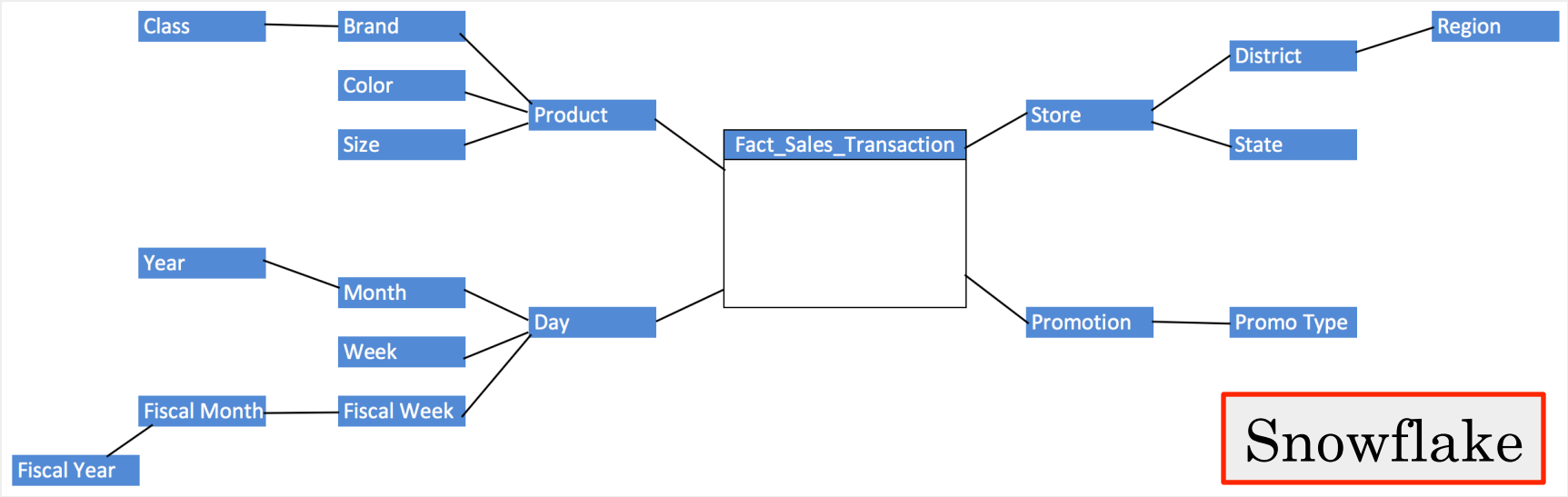


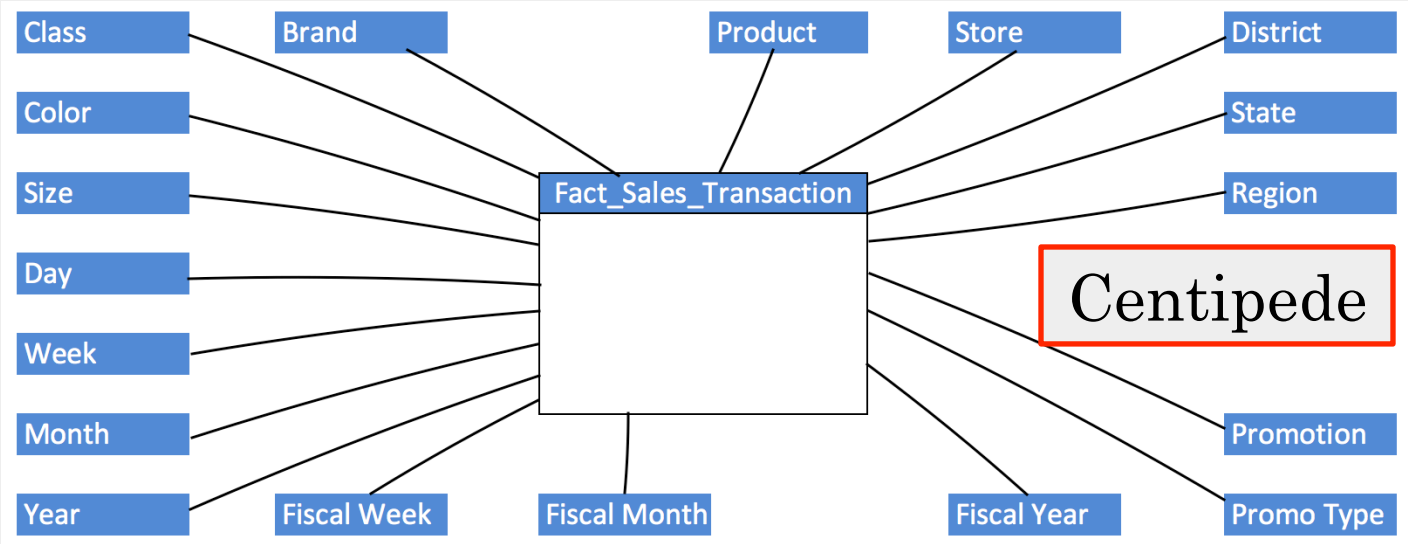
Simplified Hub-and-Spoke Corp Info Factory (CIF) Architecture Revisited



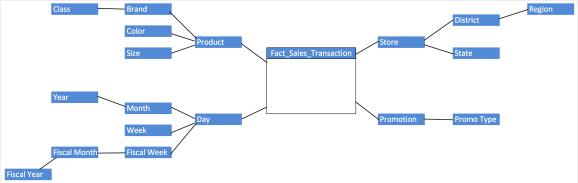
Source: http://www.kimballgroup.com/2004/03/differences-of-opinion/

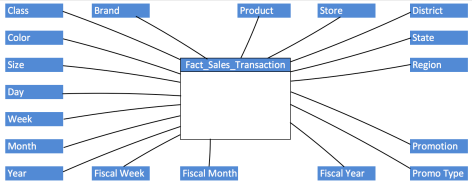
Avoid Snowflake or Centipede Model via Denormalization



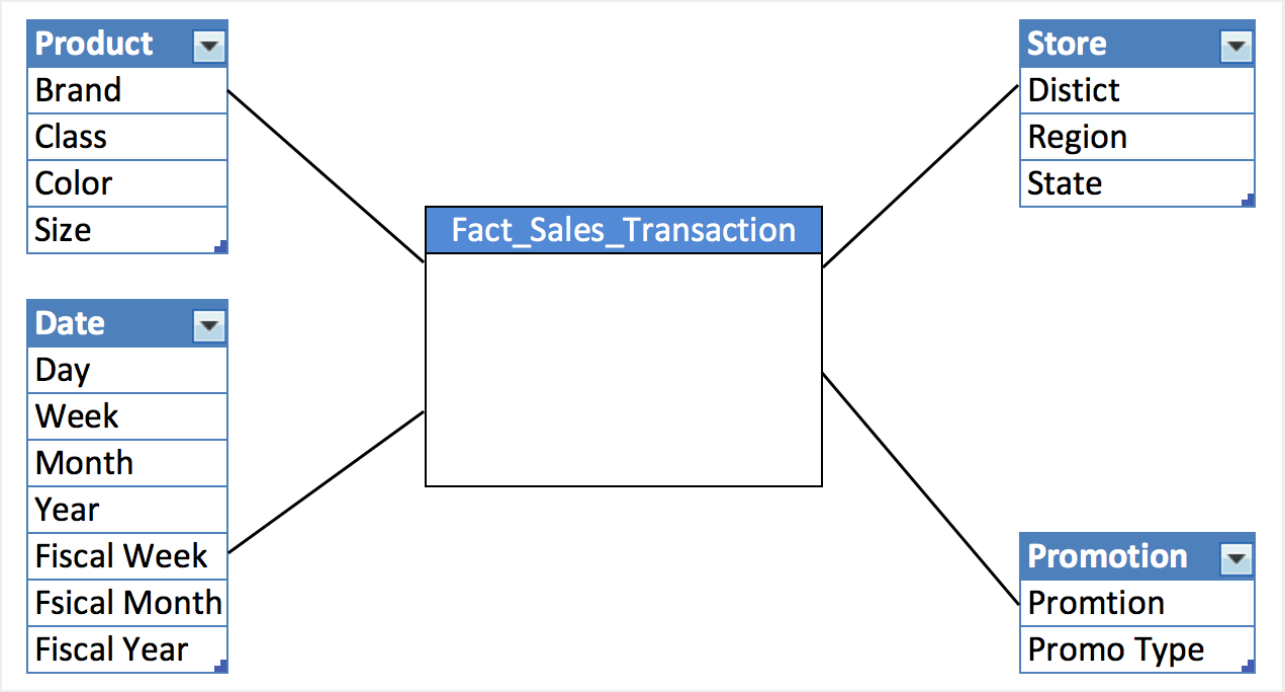


Avoid Snowflake or Centipede Model via Denormalization (cont'd)





Denormalization



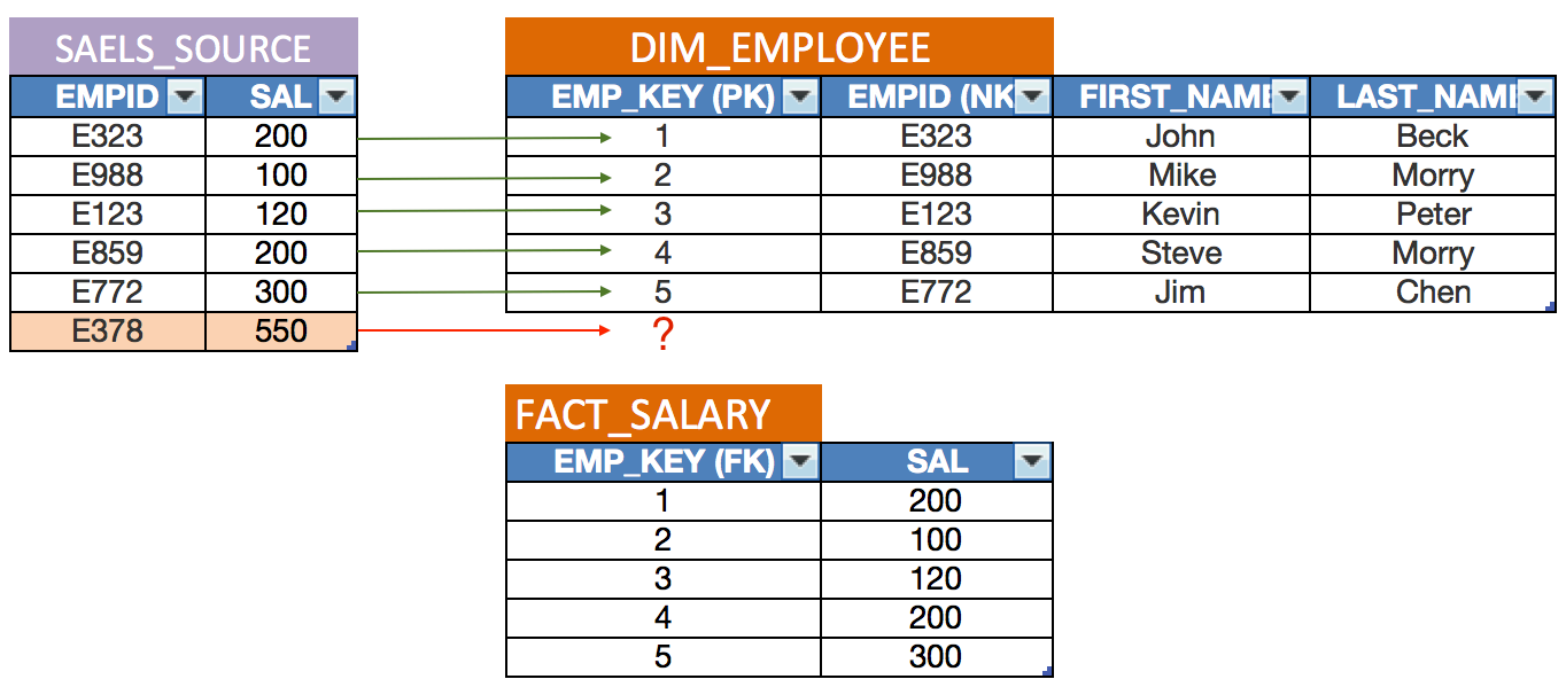
ITMD - 526

Dealing with Nulls Revisited

- NULL dimension attributes
  - Strongly discouraged to avoid unexpected query results (e.g. invalidating index strategy)
  - Use default values instead – N/A, Unknown, Invalid, To be determined,…
- NULL facts
  - Use ONLY IF it truly means N/A, Unknown, and Invalid, not zero

Default/Dummy Row in Dimension Tables

- Every foreign key in a fact table should reference a dimension row
- Default/dummy row prevents missing rows in building fact tables



SAELS_SOURCE

EMPID	SAL
E323	200
E988	100
E123	120
E859	200
E772	300
E378	550

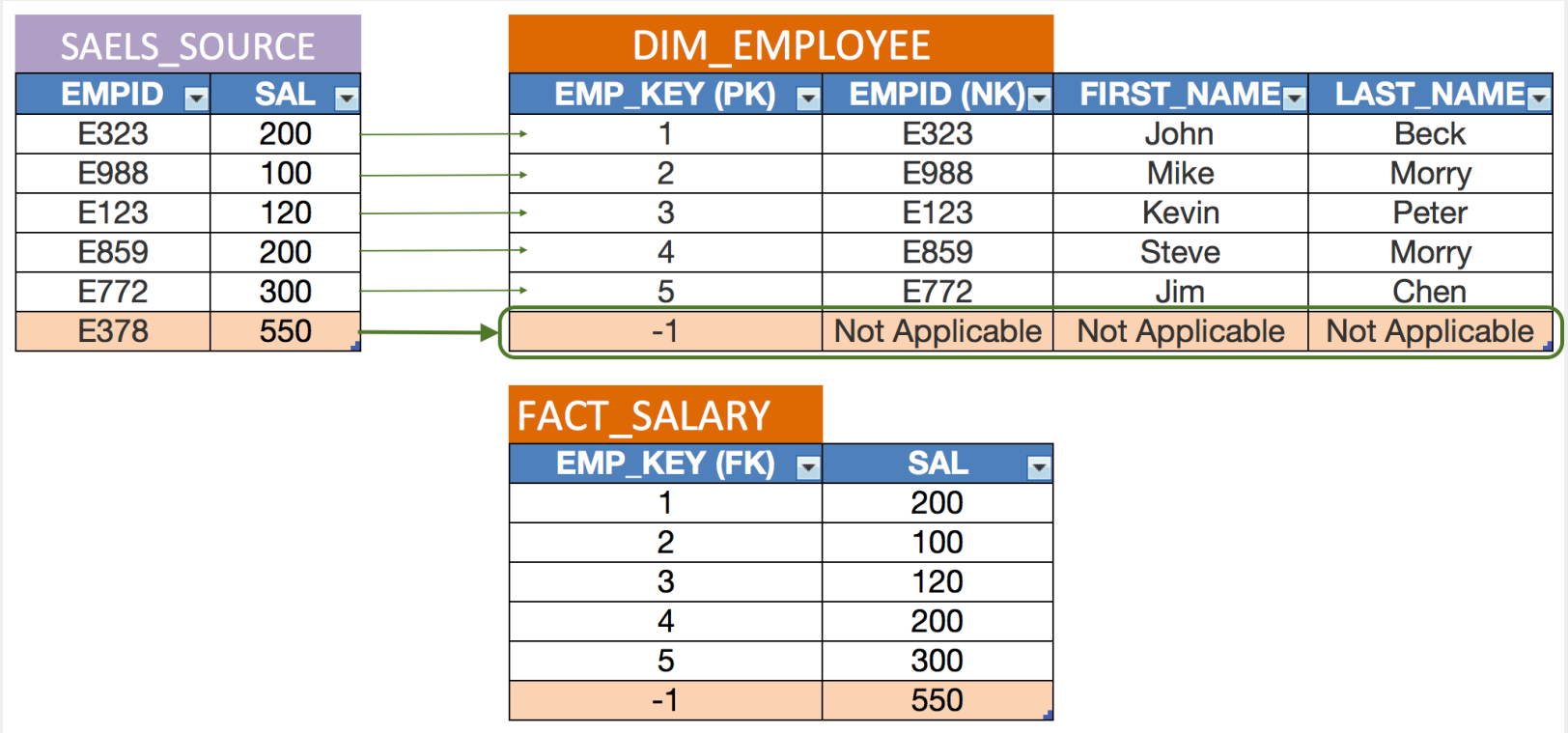
DIM_EMPLOYEE

EMP_KEY (PK)	EMPID (NK	FIRST_NAM	LAST_NAM
1	E323	John	Beck
2	E988	Mike	Morry
3	E123	Kevin	Peter
4	E859	Steve	Morry
5	E772	Jim	Chen
?			

FACT_SALARY

EMP_KEY (FK)	SAL
1	200
2	100
3	120
4	200
5	300

Default/Dummy Row in Dimension Tables (cont'd)



SAELS_SOURCE

EMPID	SAL
E323	200
E988	100
E123	120
E859	200
E772	300
E378	550

DIM_EMPLOYEE

EMP_KEY (PK)	EMPID (NK)	FIRST_NAME	LAST_NAME
1	E323	John	Beck
2	E988	Mike	Morry
3	E123	Kevin	Peter
4	E859	Steve	Morry
5	E772	Jim	Chen
-1	Not Applicable	Not Applicable	Not Applicable

FACT_SALARY

EMP_KEY (FK)	SAL
1	200
2	100
3	120
4	200
5	300
-1	550

- Default values for dimension attributes:
  "Missing Value", "Not Happened Yet", "Domain Violation", "Not Applicant", etc.

Design Workshop #2

ITMD - 526

INVOICE

Kitchen Bath & Beyond
1100 Dono Street
E Subway, CA 44321
(233) 233-1232

Invoice No.	Invoice Date
2034	3/1/15
Ship By Date	**Actual Ship Date**
2/25/15	2/28/15
Order Type	**Payment Terms**
Telephone	Net 30 Days

BILL TO Customer

Erick Johnson
2012 E Evandel Ave.
Walt, CT, 55493
(234) 234-3234
ejohnson-dummy@mail.com

SHIP TO Customer

Customer Service Repair
PO Box 456
Two Tower, NY 45920 USA

#	Product	Quantity	Unit Price	Extended Price
1	Fenix LD 20 6 Level High Performance Cree LED Flas	4	54.94	219.76
2	Coleman Cooler Hinges	5	5.99	29.95
3	Duracell Daylite 80-Lumens, 3-Watt LED Flashlight	1	28.24	28.24
				-
				-
Thank you for your business!		**SUBTOTAL**		**277.95**
		FREIGHT CHARGE		**11.81**
		TOTAL	**$**	**289.76**

Warehouse:

Midwest Distribution Center
2522 Dundee Street
West Niles, NY 12243

Ship Mode: 3-Day Ground

Invoice Remarks:
Watch for our new product catalog coming soon!

Design Workshop #2 (Cont'd)

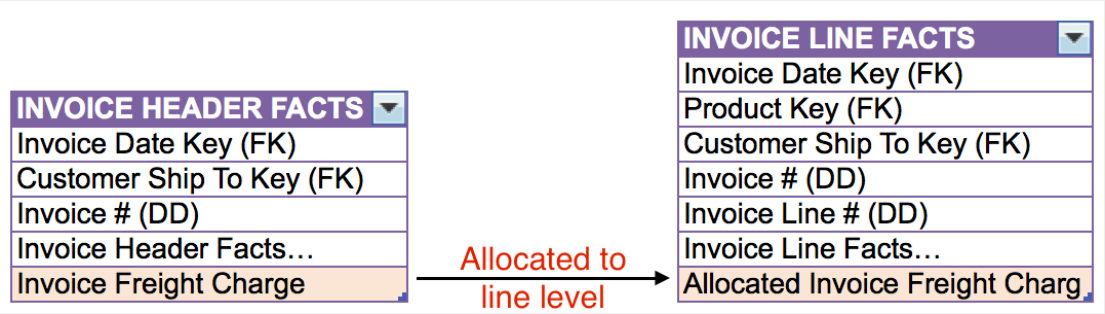
Complete the 4-step process for designing dimensional models.

- Business Process:
- Grain:
- Dimensions:
- Facts:

Draw dimensional star schema:

Allocating Header Facts

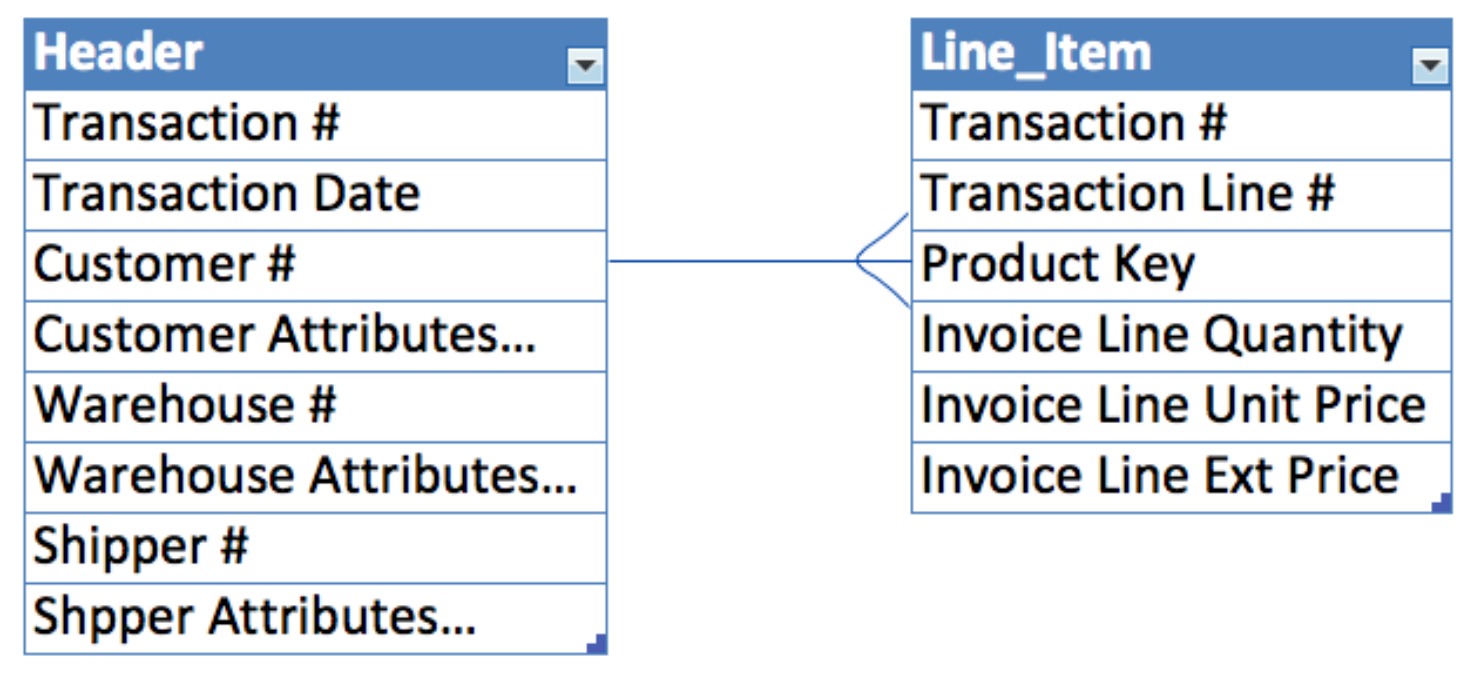
- Recommended
  - Option 1: Allocate to line-grained fact table



  - Option 2: Leave unallocated fact in header-grain fact table
- Not recommended
  - Include unallocated fact on every line fact row
  - Include unallocated fact on first or last fact row
  - Store in transaction dimension

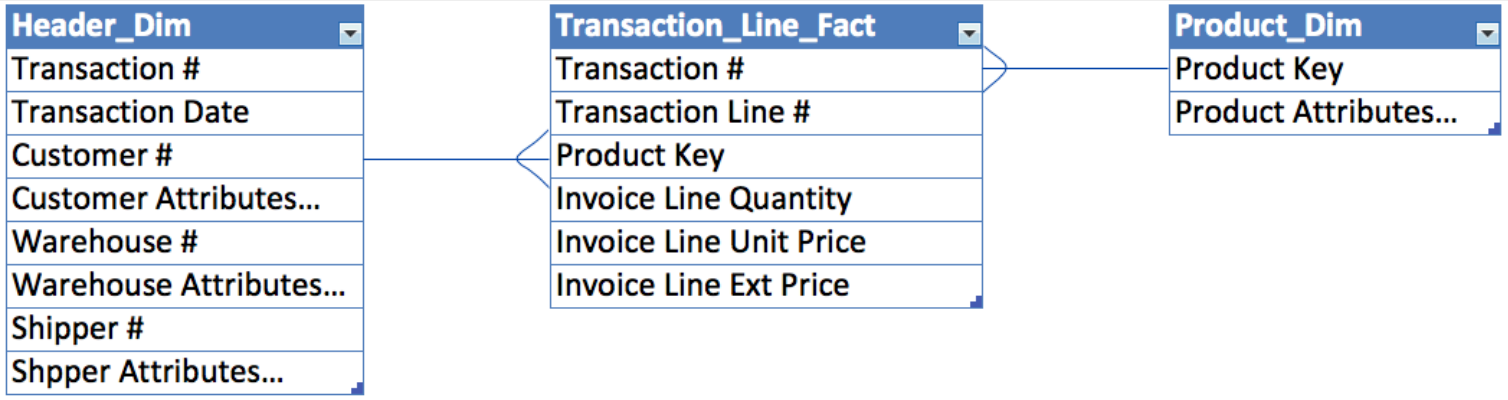
Patterns to Avoid when Modeling Header/Line Item Transactions

- ERD in OLTP Source System:



Patterns to Avoid when Modeling Header/Line Item Transactions (cont'd)

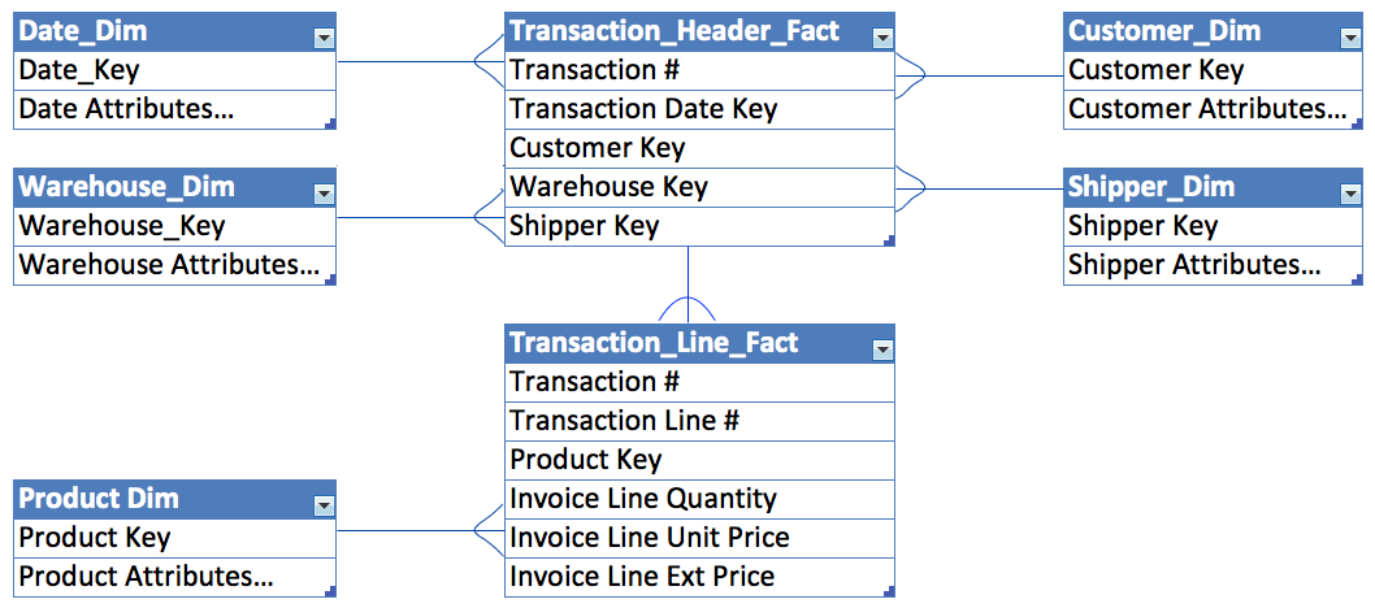
- Pattern #1 to Avoid



- The transaction header dimension will get very large
- Business user's requests will have to go through this very large dimension
- This can get even larger when applying SCD Type 2 (preserving change history)

Patterns to Avoid when Modeling Header/Line Item Transactions (cont'd)

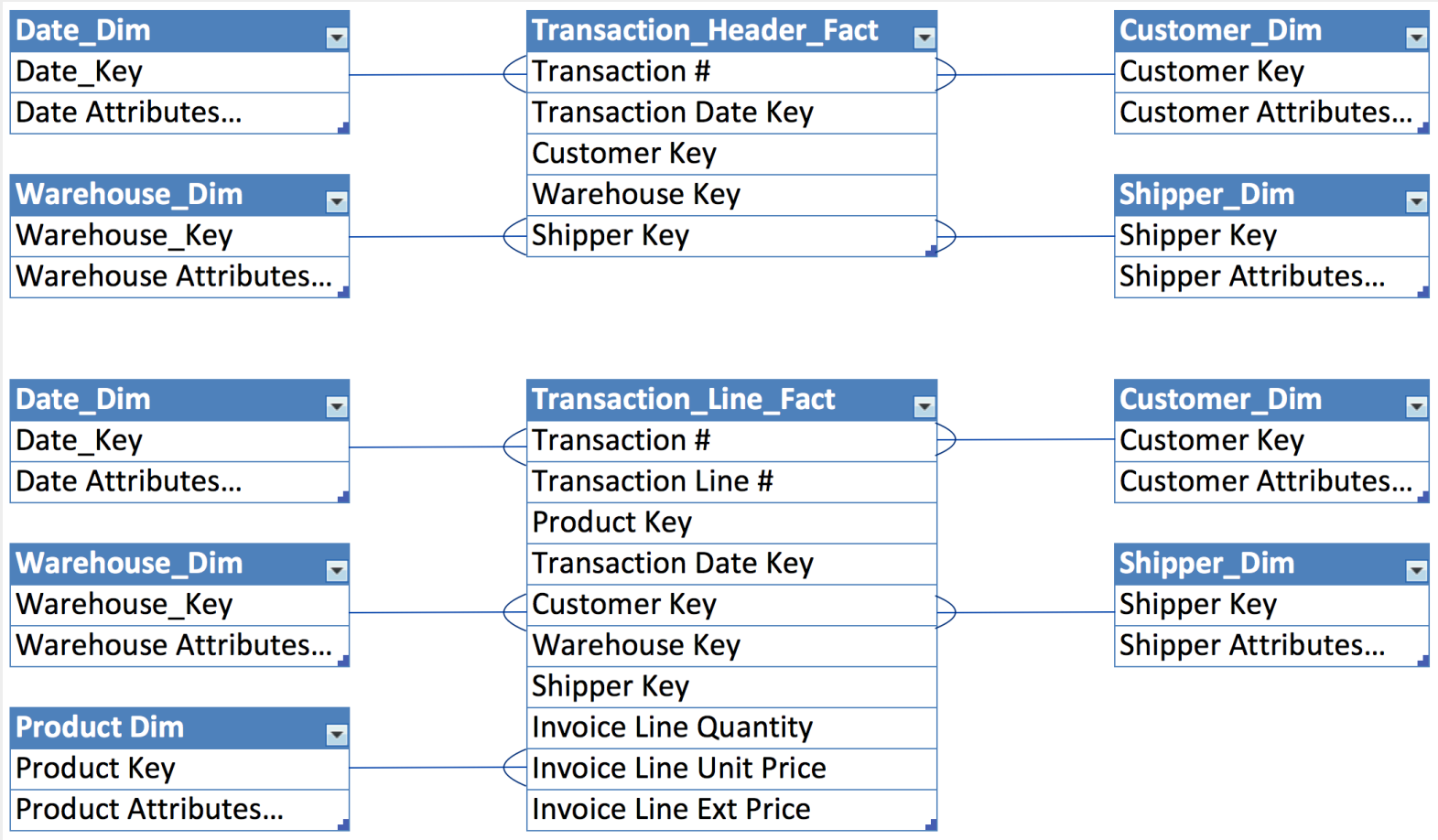
- Pattern #2 to Avoid



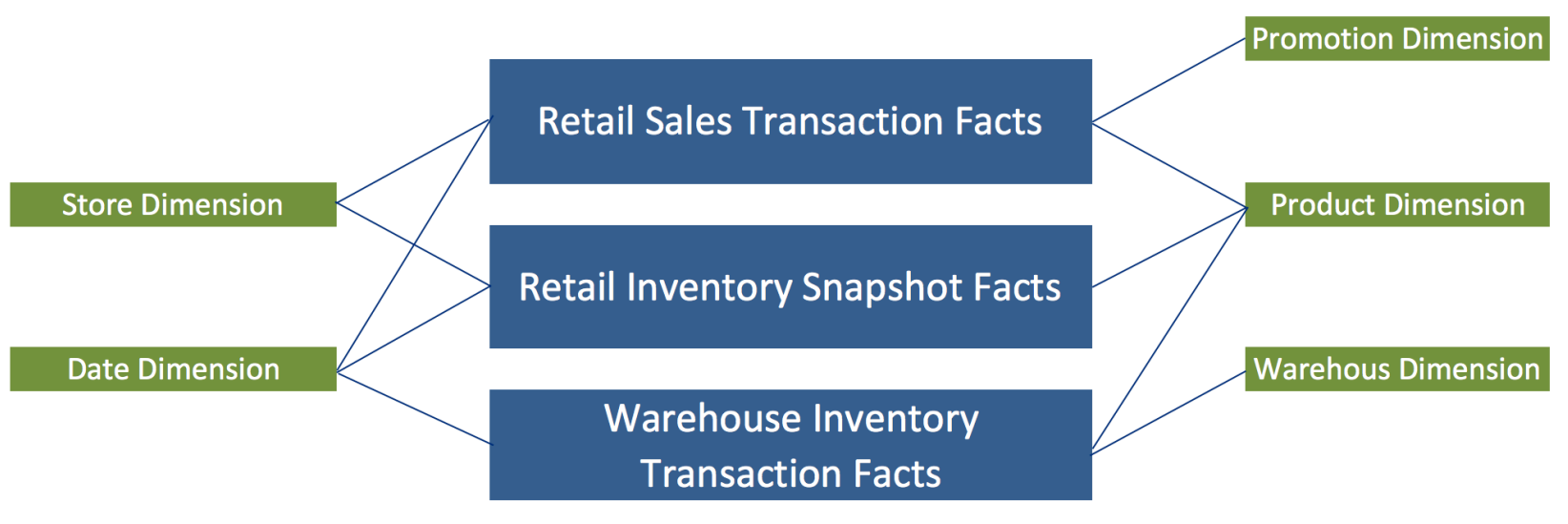
  - It is fine to build Header fact for header metrics, but Line fact table should inherit header dimensions

Patterns to Avoid when Modeling Header/Line Item Transactions (cont'd)

- Improved pattern (not optimal yet)

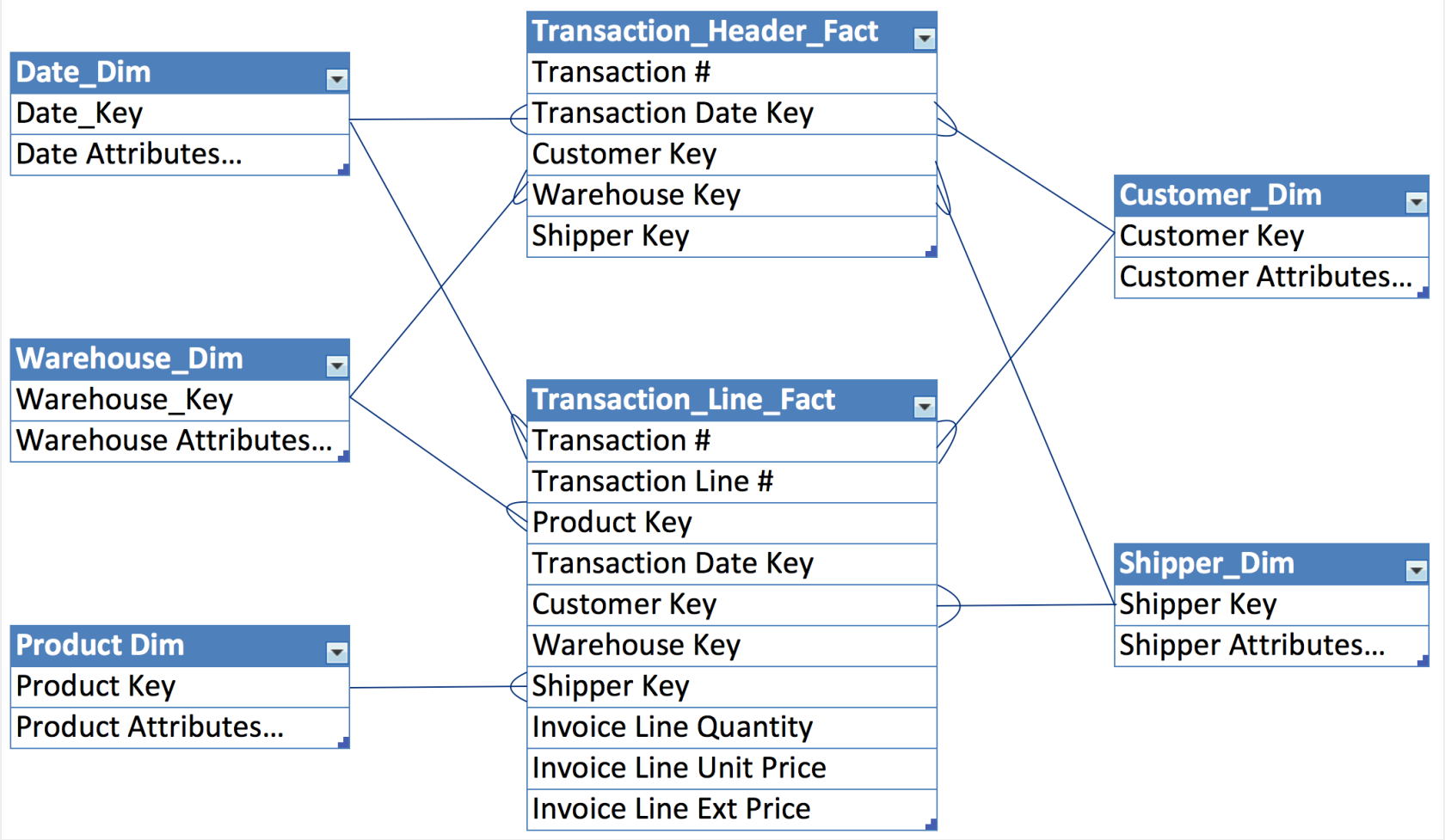


Conformed Dimensions Tables Shared by Fact Tables



- Each business process typically is represented by one or more fact tables
- Using shared, common dimensions is absolutely critical for data marts integrated seamlessly

Patterns to Avoid when Modeling Header/Line Item Transactions Revisited



Conformed Dimensions

- Identical dimensions conform
- Shrunken rollup dimensions conform
  - Domain values of conform columns must match

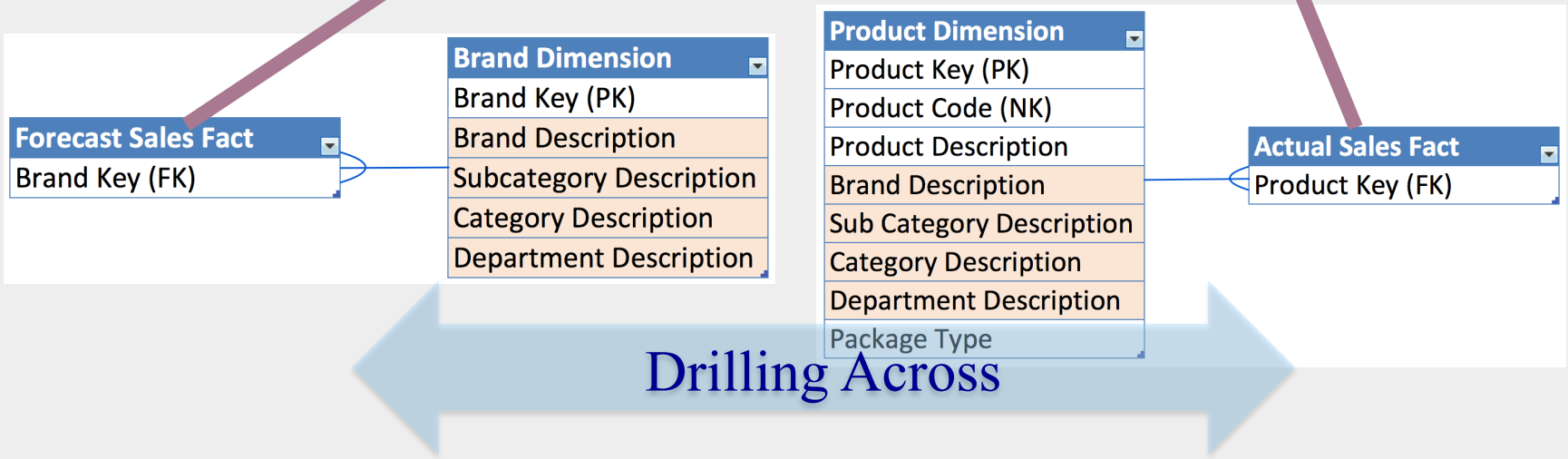
Product Dimension
Product Key (PK)
Product Code (NK)
Product Description
Brand Description
Sub Category Description
Category Description
Department Description
Package Type

Conforms

Brand Dimension
Brand Key (PK)
Brand Description
Subcategory Description
Category Description
Department Description

Drilling Across Fact Tables via Conformed Dimension Tables

Brand Description	Forecast Sales $ Amount	Actual Sales $ Amount2
UBQTous	$ 1,853.00	$ 1,832.00
SmartIOT	$ 9,932.00	$ 10,234.00
WiggleWorm	$ 3,486.00	$ 2,345.00



- Open separate connection to each source
- Assemble each answer set
- Merge answer sets on conformed row headers

Enterprise Data Warehouse Bus Architecture

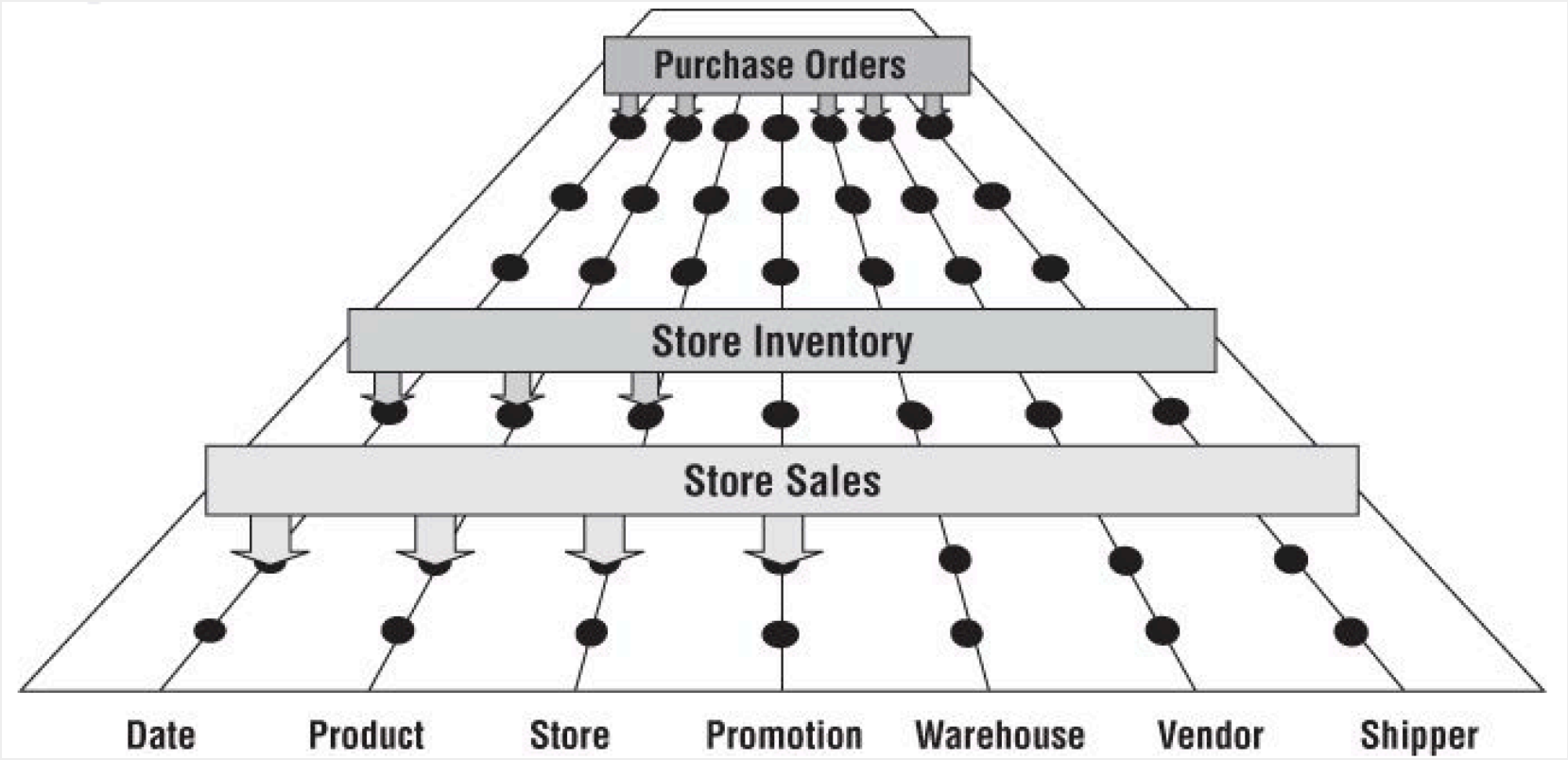


Figure 4.9 Enterprise data warehouse bus with shared dimensions.
Kimball, Ralph; Ross, Margy (2013-07-01). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Wiley. Kindle Edition.

Enterprise Data Warehouse Bus Architecture (cont'd)

- The Enterprise Data Warehouse Bus Architecture provides a standardized master set of conformed dimensions and conformed facts used through the data warehouse
- It is analogous to the bus in your computer, providing a standard interface that allows many different kinds of devices to connect to your computer and co-exist
- Conformed dimensions are standard dimensions that are shared among dimensional models.

Enterprise Data Warehouse Bus Architecture (cont'd)

- The use of conformed dimension is the central technique for building an enterprise data warehouse from a set of dimensional models
- As the separate dimensional models are developed, they plug into the Bus, fitting together like pieces of the puzzle
- Isolated data marts that cannot be tied together are disastrous. Stovepipe data marts merely perpetuate incompatible views of the business

Kimball, Ralph; Ross, Margy. Dimensional Modeling in Depth (41-42). Kimball University.

Enterprise Data Warehouse Bus Matrix

Figure 4.10 Sample enterprise data warehouse bus matrix for a retailer.
Kimball, Ralph; Ross, Margy (2013-07-01). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Wiley. Kindle Edition.

BUSINESS PROCESSES	Date	Product	Warehouse	Store	Promotion	Customer	Employee
Issue Purchase Orders	X	X	X				
Receive Warehouse Deliveries	X	X	X				X
Warehouse Inventory	X	X	X				
Receive Store Deliveries	X	X	X	X			X
Store Inventory	X	X		X			

- Rows translate into fact tables
- Columns represent common dimensions used across the enterprise. Mark the intersections where the dimensions are relevant to the business processes. The resulting matrix will be surprising dense

Enterprise Data Warehouse Bus Matrix (cont'd)

- Sharing conformed dimensions across the data warehouse is absolutely critical
  - Ensures consistent definition of common data
  - Ensures consistent row/column heading labels and roll-ups
  - Ensures consistent "values" for consistently defined dimensions and attributes
  - Reduce time to market
  - Support integration and drilling across fact tables

- Committing to use conformed dimensions is a business policy. It represents more political challenges than technical hurdles

Week 05 Topic: Dimensional Modeling: Integration Via Conformed Dimensions

Questions?