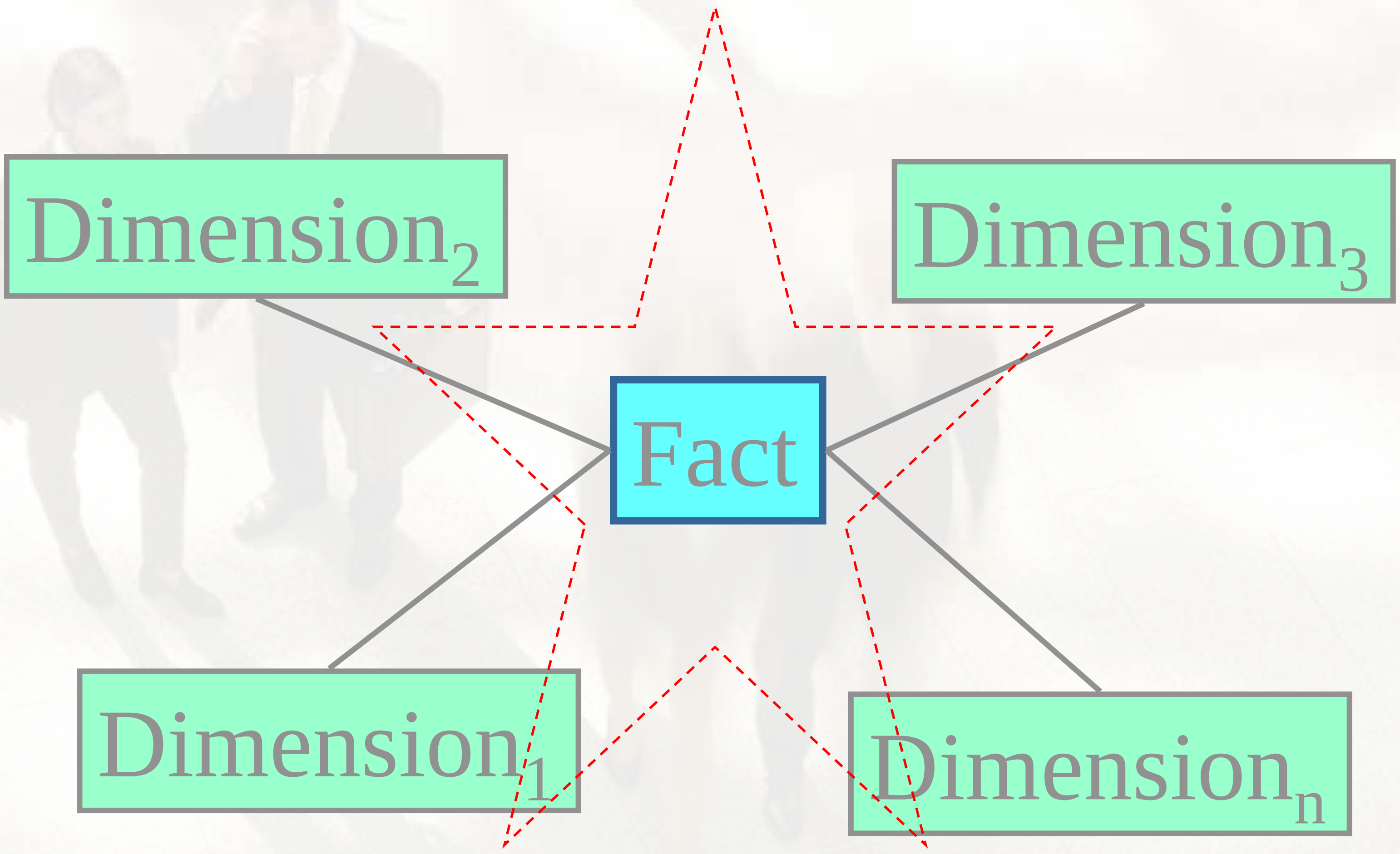
Lection 7

Dimensional modeling

Learning Objectives

- Revision of Retails Star Schema
- Inventory Models
- Semi-additive facts
- Data Warehouse Bus Architecture
- Conformed dimensions

Star for Retail



Star for Retail

Time

time_k	day	month	quarter	year	season
T01	1	Jan.	1Q	1999	New Years
T02	2	Jan.	1Q	1999	New Years
T03	3	Jan.	1Q	1999	Normal
T04	4	Jan.	1Q	1999	Normal
T05	5	Jan.	1Q	1999	Normal

Customer

customer_k	name	age	gender	education	income	address
9901	David	30	Male	Graduate	80,000	34 Greystone, Austin TX 78730
9902	Nathan	55	Male	Graduate	100,000	23 Wood #21, Houston TX 71010
9903	Jane	55	Female	College	95,000	12 Central, New York NY 10200
9904	Mary	23	Female	High_School	60,000	9 King, Buffalo NY 11200
9905	Steve	25	Male	High_School	55,000	11 Main, San Antonio TX 70340

Sales

time_k	product_k	store_k	customer_k	amount	cost
T01	Q33	A01	9901	50	25
T01	K21	B01	9902	60	40
T01	Q33	A03	9903	10	5
T02	K21	A02	9903	50	20
T02	Q33	A01	9902	100	50
T02	K21	A03	9901	5	3
T02	K21	B02	9904	20	15
T04	K21	A02	9901	3	1
T05	P67	A01	9905	4	3
T05	K21	A02	9905	10	3

product_k	name	brand	subcategory	category
K21	Power_Clean	Cleaners	detergent	house_goods
Q33	Coke_Classic	Cola	carbonated	soft_drink
Q34	Coke_Diet	Cola	carbonated	soft_drink
P67	Sprite	Coolers	carbonated	soft_drink

Product

store_k	name	zip	region	state	manager	phone
A01	Farwest	78700	Travis	TX	Brown	512-345-6678
A02	Anderson	78700	Travis	TX	Molly	512-342-3356
A03	Koneig	79220	Austin	TX	Jones	512-399-1245
B01	South	10020	Soho	NY	Jane	212-245-4563
B02	Central	10032	Midtown	NY	Marvin	212-362-2278

Store

Star for Retail

ETL: Avoid normalization

Joining with Code Tables

Customer_Desc

cid	name	age	sex	education	income	address
9901	David	30	1	1	80,000	34 Greystone, Austin TX 78730
9902	Nathan	55	1	1	100,000	23 Wood #21, Houston TX 71010
9903	Jane	55	2	2	95,000	12 Central, New York NY 10200
9904	Mary	23	2	3	60,000	9 King, Buffalo NY 11200
9905	Steve	25	1	3	55,000	11 Main, San Antonio TX 70340

Education_Code

Code	Literal
1	Graduate
2	College
3	High School
4	Others

Gender_Code

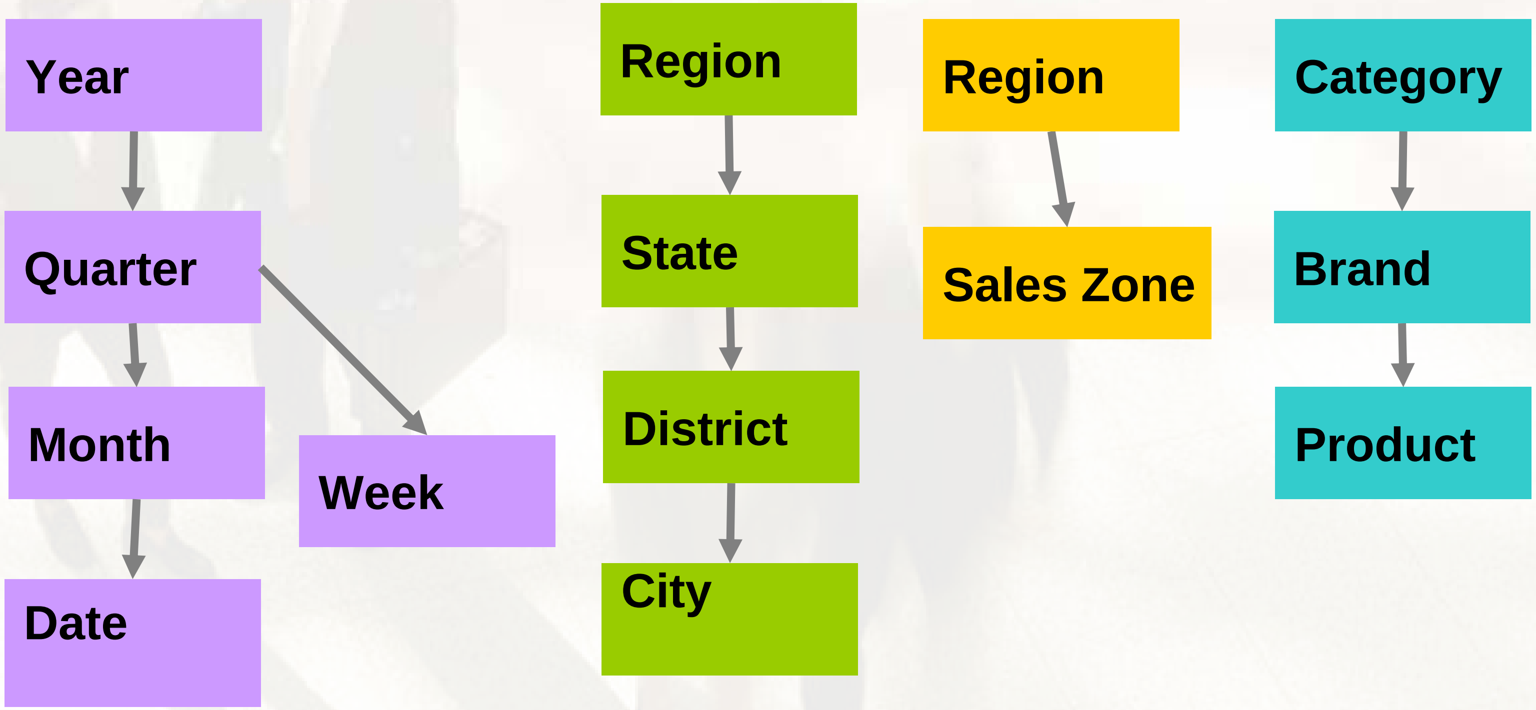
Code	Literal
1	Male
2	Female

Customer

customer_k	name	age	gender	education	income	address
9901	David	30	Male	Graduate	80,000	34 Greystone, Austin TX 78730
9902	Nathan	55	Male	Graduate	100,000	23 Wood #21, Houston TX 71010
9903	Jane	55	Female	College	95,000	12 Central, New York NY 10200
9904	Mary	23	Female	High_School	60,000	9 King, Buffalo NY 11200
9905	Steve	25	Male	High_School	55,000	11 Main, San Antonio TX 70340

Star for Retail

Hierarchies



Case study: Inventory

- Value chain
  - Consisting of organization's key business processes
  - The flow of an organization's primary activities
  - Provides high-level insight into the overall enterprise DW
- Movement of products

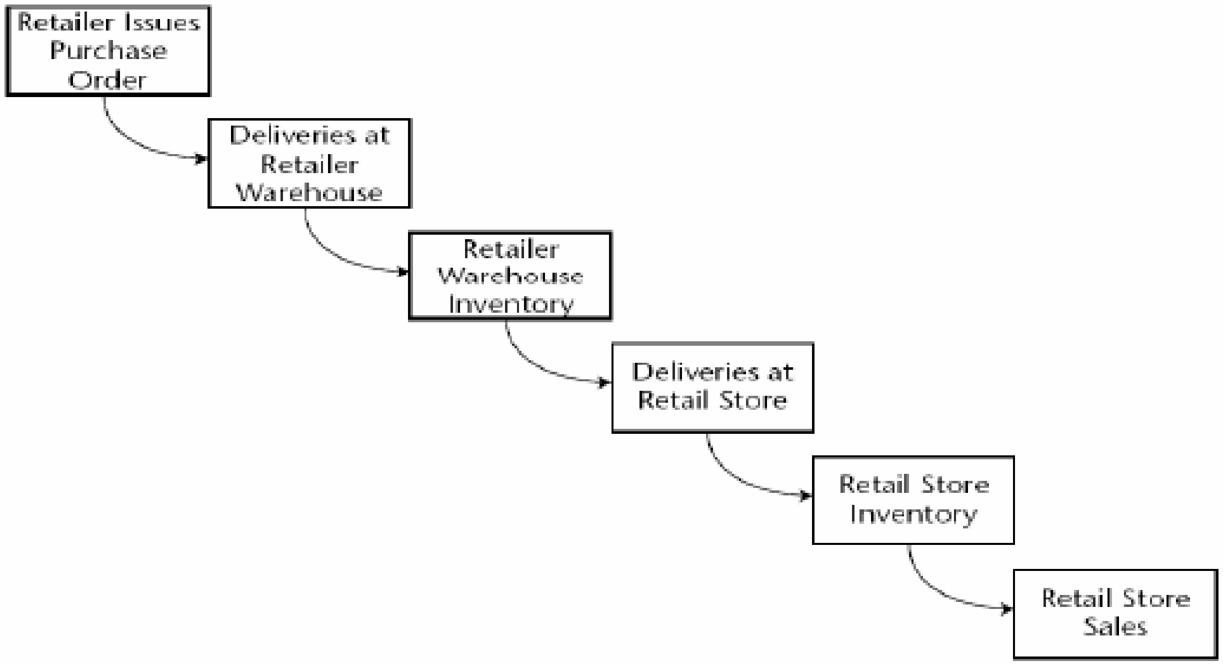


Figure 3.1 Subset of a retailer's value chain.

Inventory

- Inventory Periodic Snapshot
- Inventory Transactions

Design Decisions

- **Choosing the process.**

  Selecting the subjects from the information packages for the first set of logical structures to be designed.
- **Choosing the grain.**

  Determining the level of detail for the data structures.
- **Identifying and conforming the dimensions.**

  Choosing the business dimensions (such as product, market, time, etc.) to be included in the first set of structures .
- **Choosing the facts.**

  Selecting the metrics or units of measurements (such as product sale units, dollar sales, dollar revenue, etc.) included in set of structures.
- Choosing the duration of the database.

  Determining how far back in time you should go for historical data.

Inventory Periodic Snapshot

- Optimized inventory levels
  - Minimize out-of-stocks
  - Reduces overall inventory carrying costs
- Dimensional model
  - Dimensions
    - Date, product, store
- Simple dimensional design

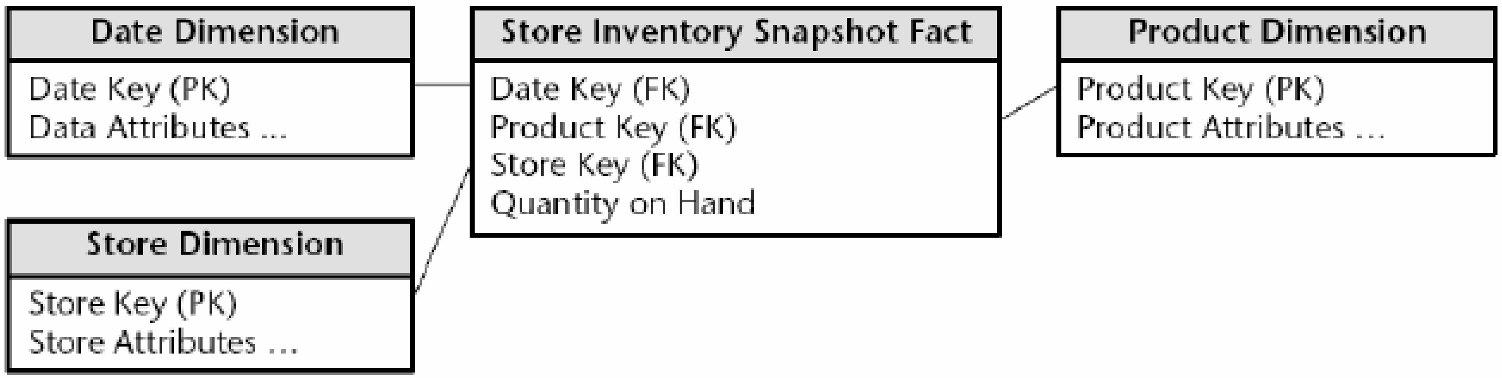


Figure 3.2 Store inventory periodic snapshot schema.

Inventory Periodic Snapshot

- Additional attributes
  - Product dimension
    - Minimum reorder quantity
    - Descriptors of each product stock keeping unit (SKU)
  - Store dimension
    - Selling square-footage
    - Frozen and refrigerated storage square footages

Inventory Periodic Snapshot

- Semiadditive Facts
  - Numeric fact that can be added along some dimensions but not others
  - Example
    - Inventory levels
      - Additive across products or stores
      - But cannot be additive across date
  - Complexity of inventory calculation
    - Cannot use the SQL AVG function
      - No standard functionality that would compute the average over just the date dimension
    - Solutions
      - With an embedded SQL
      - By querying the date dimension separately and storing the resulting value

Inventory Periodic Snapshot

	Mon	Tue	Wed	Thu	Fri
Prod A	1	1	2	2	1
Prod B	2	1	2	2	1
SumDate	3	2	4	4	2
TotalSum					**15**

AVG = TotalSum / 10 = 15 / 10 = 1.5

AVG_DATE = TotalSum / 5 = 15 / 5 = 3

Inventory Periodic Snapshot

- Dense snapshot tables
  - Inventory levels are measured frequently
    - To avoid out-of-stock situation
    - Example
      - 60,000 products * 100 store * 14 row width = 84MB
      - A year's worth of daily snapshots >= 30GB
  - To reduce the snapshot frequencies
    - 1,095 snapshots during a 3-year period

      => 208 snapshots(60 daily + 148 weekly snapshots in two separate fact tables)

Inventory Transactions

- Inventory transactions at the warehouse
  - Receive product
  - Place product into inspection hold
  - Release product from inspection hold
  - Return product to vendor due to inspection failure
  - Place product in bin
  - Authorize product for sale
  - Pick product from bin
  - Package product for shipment
  - Ship product to customer
  - Receive product from customer
  - Return product to inventory from customer return
  - Remove product from inventory

Inventory Transactions

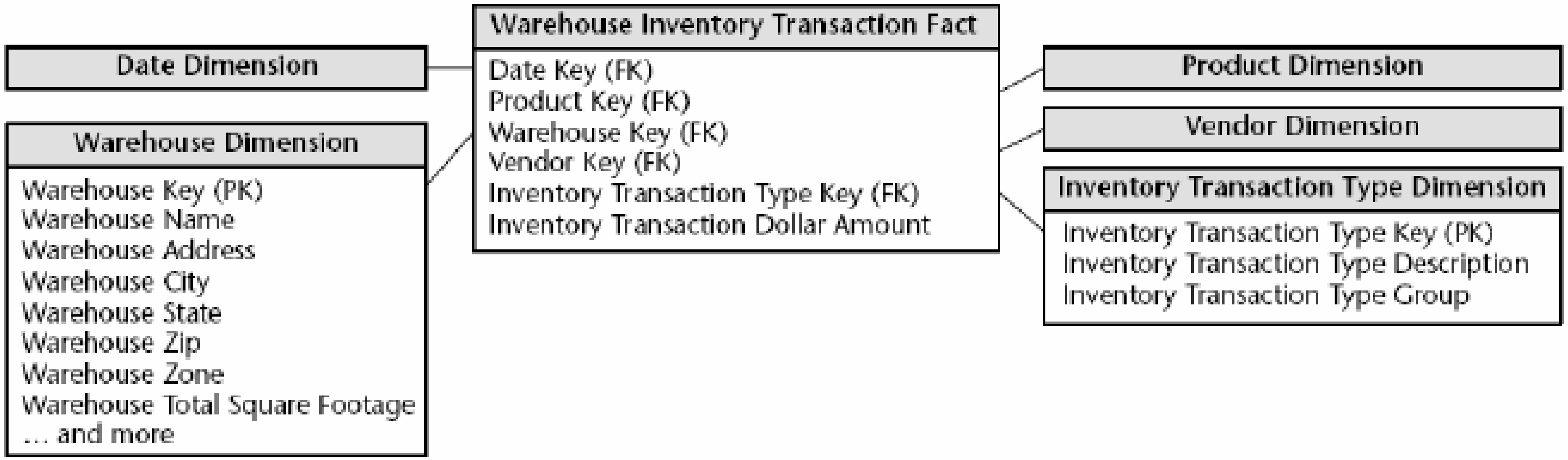


Figure 3.4 Warehouse inventory transaction model.

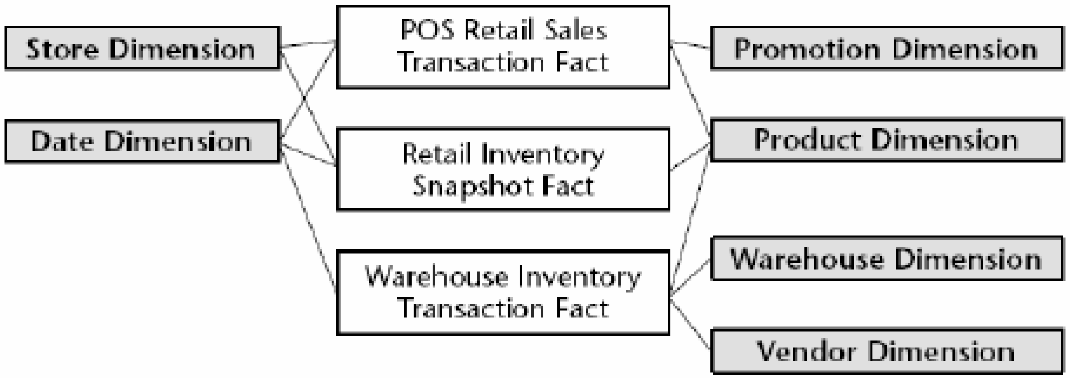
- Characteristics
  - The most detailed information available about inventory
    - It mirrors fine scale inventory manipulations
  - Useful for measuring the frequency and timing of specific transaction types

Inventory Transactions

- Question example
  - How many times have we placed a product into an inventory bin on the same day we picked the product from the same bin at a different time?
  - How many separate shipments did we receive from a given vendor, and when did we get them?
  - On which products have we had more than one round of inspection failures that caused return of the product to the vendor?
- Disadvantage
  - It is impractical to use this table as the sole basis
    - It is too cumbersome and impractical
    - For broad data warehouse questions that span dates or products

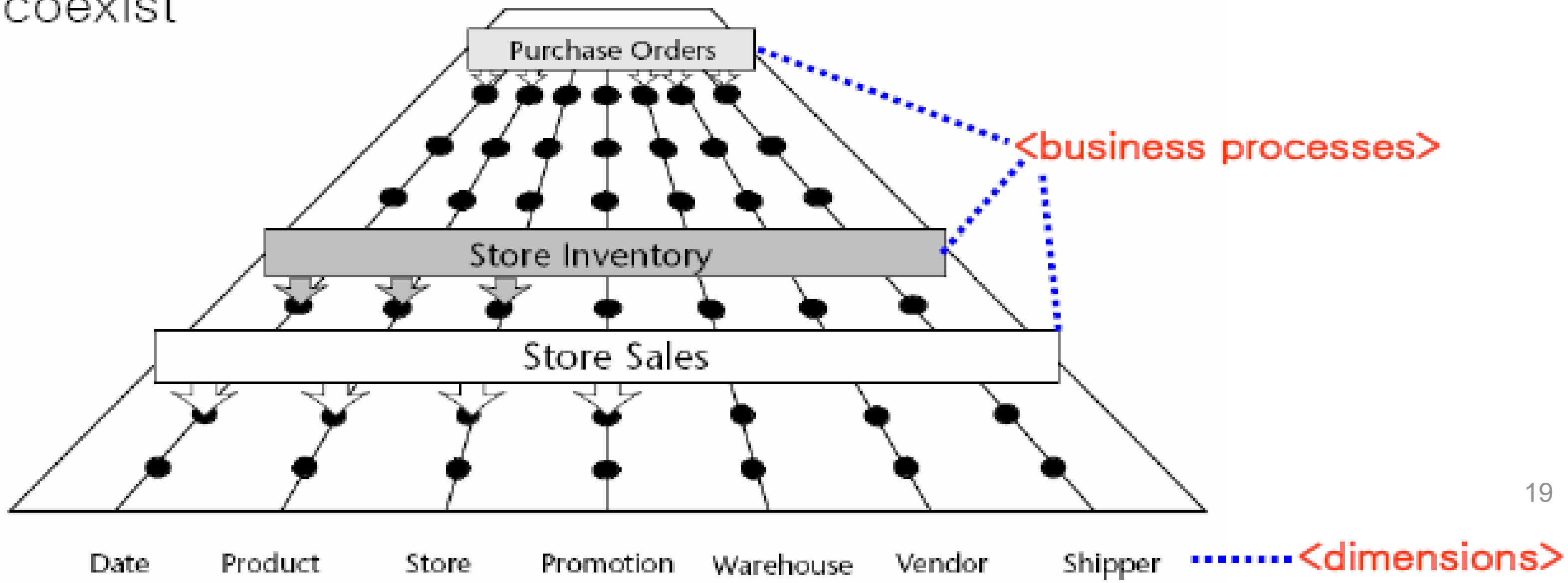
Value Chain Integration

- Needs
  - To better evaluate performance
  - To better leverage scarce resource and gain efficiencies
- Common dimensions
  - At each process, the models share several common dimensions
    - Date, product, and store
  - It is critical to designing data marts that can be integrated
- Drill across
  - The linkage that use multipass SQL to query each data mart separately, and outer join the query results based on a common dimension attribute



Data Warehouse Bus Architecture

- For long-term data warehouse success
  - Need to use architected, incremental approach to build the enterprise's warehouse
  - Bus
    - Common structure to which everything connects and from which everything derives power
- A standard data warehouse bus architecture
  - The separate data marts can be plugged together and usefully coexist



Data Warehouse Bus Architecture

- Characteristics
  - A rational approach to decomposing the enterprise data warehouse planning task
    - Design a master suite of standardized dimensions and facts
    - Implementation of separate data marts
    - Separate data marts come on line, they fit together
  - Is independent of technology and the database platform

Data Warehouse Bus Architecture

- Data Warehouse Bus Matrix
  - The tool we use to create, document, and communicate the bus architecture

BUSINESS PROCESSES	Date	Product	Store	Promotion	Warehouse	Vendor	Contract	Shipper
	COMMON DIMENSIONS							
Retail Sales	X	X	X	X				
Retail Inventory	X	X	X					
Retail Deliveries	X	X	X					
Warehouse Inventory	X	X			X	X		
Warehouse Deliveries	X	X			X	X		
Purchase Orders	X	X			X	X	X	X

Data Warehouse Bus Architecture

- First-level data marts
  - Derived from a single primary source system
  - The rows of the matrix
- Consolidated data marts
  - Derived from more complex multisource marts
  - More difficult to implement
    - ETL effort grows alarmingly with each additional major source
- Advantage
  - Very powerful device
    - Planning
      - Defining the overall data architecture for the warehouse
      - Prioritize which dimensions should be tackled first
    - Communication
      - Visually conveys the entire plan at once

Data Warehouse Bus Architecture

1) Establish Scope: biz proc., goals & grain (Look around)

- Identify business process
- Understand analysis goals
  - Report automation?
  - List of Biz Questions
  - Scope/source of measures
- Identity state of & relationships to corp. DW work:
  - Process in Bus Matrix?
  - Shared Dims?
  - Data already in DW?
  - Surrounding Biz procs?
- Define the Grain
  - Level/Frequency
  - Describe in detail
    - The event OR
    - Key measures

2) Develop/Define initial Logical Dim. Model (Look over)

- Identify initial set of Dimensions
- For each Dimension
  - Verify it applies
  - Identify Attributes
  - Look for hierarchies
  - Meaningful names
  - Clear definitions
  - Identify SME/biz owner
- Identify the Facts:
  - Numeric-additive 1st
  - Indiv. Counts & Amts…
  - Think intellectual capital
    - Calculations?
    - Derivations?
  - Meaningful names
  - Clear definitions

3) Refine Logical Dimensional Model (Look in & out)

- Focus shifts to internal design (Look in: content)
  - Domain Analysis
  - Define Valid Value Set
  - Source-2-Target Maps
  - Handle none & bad!
  - ID Change Strategy
- Look in: Right Dims?
  - Role Playing
  - Look for clusters
  - Correlation?/Change freq?
  - Look at total #
    - Combine? (junk)
- Look out:
  - Revisit Corp Bus Matrix
  - Re-use Shared Dims
  - New Shared Dims?

 28 Dimensional Modeling - A Whirlwind Tour of How & Why

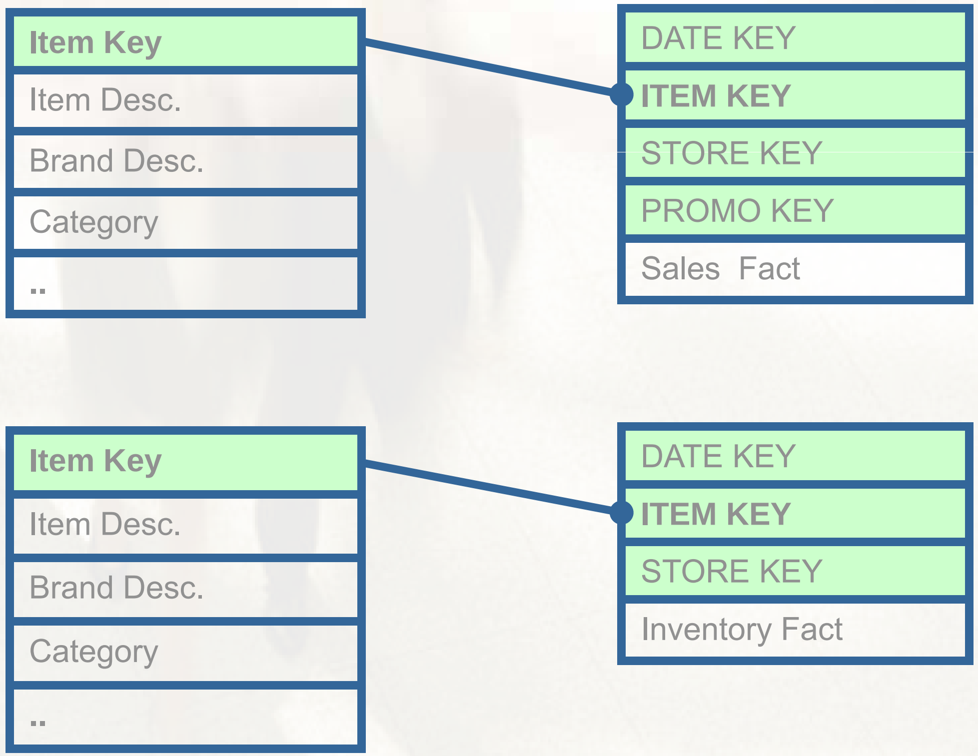
Data Warehouse Bus Architecture

Conformed dimensions

- **Option 1:** Identical dimensions with the same keys, labels, definitions and values

Sales Schema

Inventory Schema

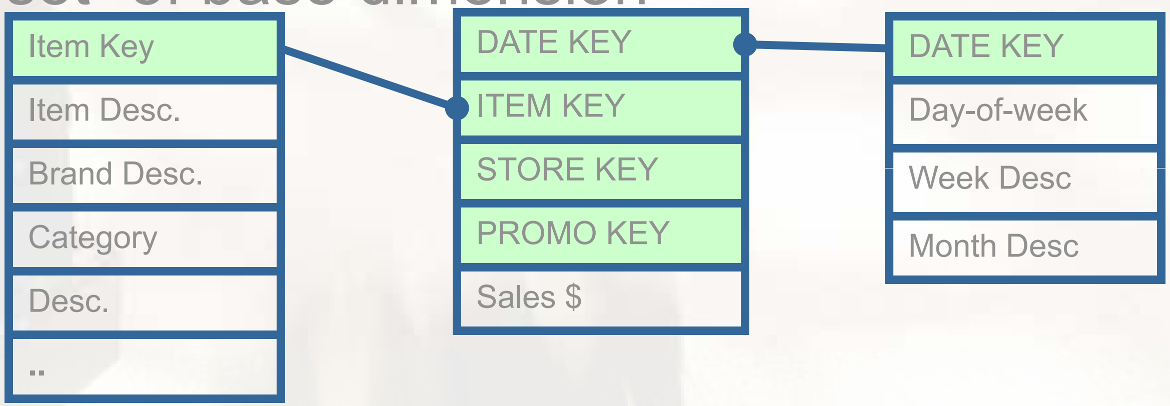


Data Warehouse Bus Architecture

Conformed dimensions

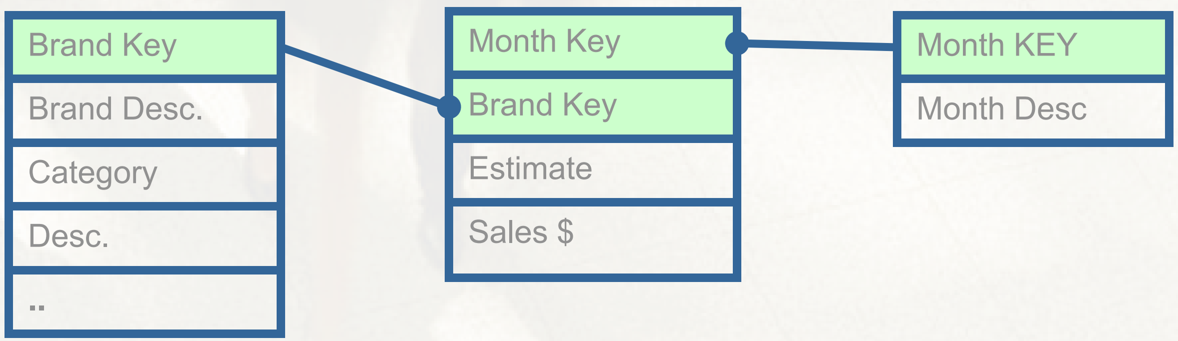
- **Option 2:** "Subset" of base dimension

Sales Schema



Item key	Item Desc	Brand Desc	Category Desc
0001	Cheerios 10oz	Cheerios	Cereal

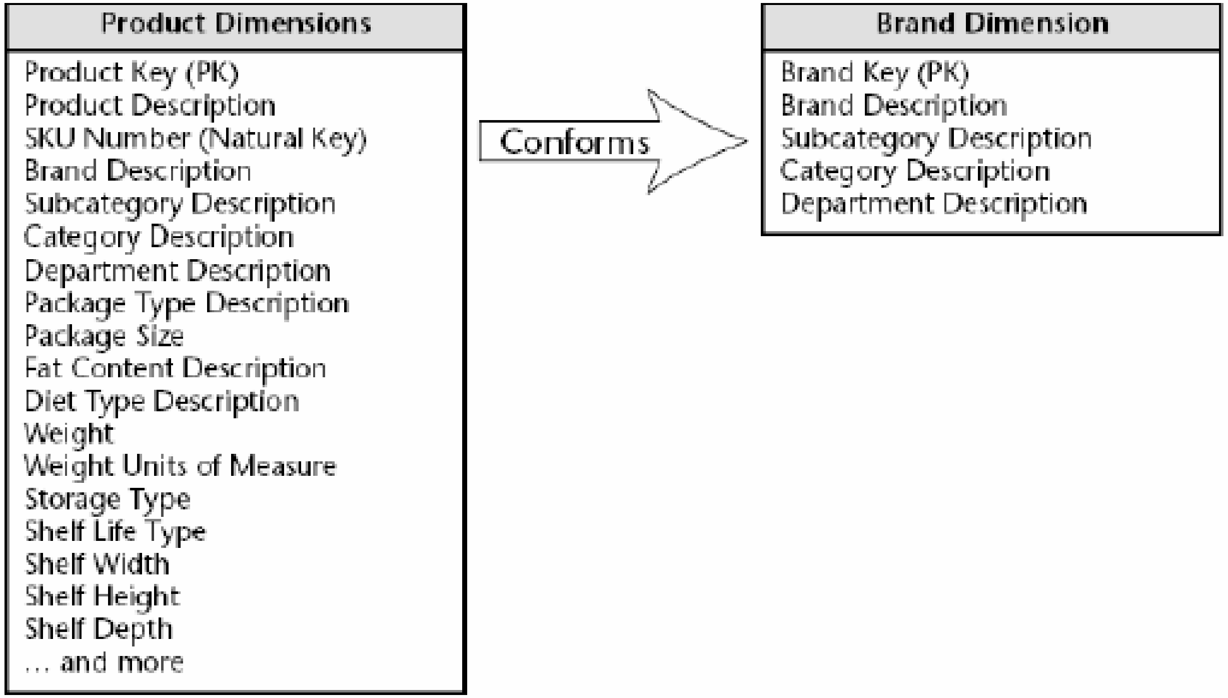
Forecast Schema



Brand key	Brand Desc	Category Desc
1001	Cheerios	Cereal

Data Warehouse Bus Architecture

- Conformed dimensions
  - Subset of the most granular, detailed dimension
  - Roll-up dimensions conform to the base-level atomic dimension

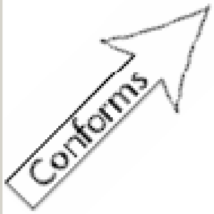


Data Warehouse Bus Architecture

- Example rollup dimension

Product Dimension

product key	description	SKU number	brand	category	department
30	Athletic Drink	48530259240	Big Can	Drinks	Grocery
8	Fizzy Light	33411763259	Big Can	Drinks	Grocery
9	Fizzy Classic	95946398896	Big Can	Drinks	Grocery
10	Athletic Drink	88602993232	Big Can	Drinks	Grocery
28	Fizzy Light	92822703206	Big Can	Drinks	Grocery
29	Fizzy Classic	74695428497	Big Can	Drinks	Grocery
48	Fizzy Light	59632819867	Big Can	Drinks	Grocery
49	Fizzy Classic	64758233722	Big Can	Drinks	Grocery
50	Athletic Drink	63998140597	Big Can	Drinks	Grocery
52	Clear Refresher	26124581284	National Bottle	Drinks	Grocery
51	Strong Cola	78532224693	National Bottle	Drinks	Grocery
11	Strong Cola	59015963215	National Bottle	Drinks	Grocery
12	Clear Refresher	94794170004	National Bottle	Drinks	Grocery
31	Strong Cola	10478516528	National Bottle	Drinks	Grocery
32	Clear Refresher	89835195915	National Bottle	Drinks	Grocery
16	Salty Corn	80323441322	American Corn	Food	Grocery
56	Salty Corn	21628878100	American Corn	Food	Grocery
36	Salty Corn	54983505685	American Corn	Food	Grocery
37	Dried Grits	11184804406	American Corn	Food	Grocery
38	Power Chips	51364643658	American Corn	Food	Grocery
17	Dried Grits	15536655574	American Corn	Food	Grocery
57	Dried Grits	55681968175	American Corn	Food	Grocery
58	Power Chips	43992125296	American Corn	Food	Grocery
18	Power Chips	44513822387	American Corn	Food	Grocery
27	Sweet Tooth	10787621276	Chewy Industries	Food	Grocery
7	Sweet Tooth	51770124461	Chewy Industries	Food	Grocery



Brand rollup Dimension

brand key	brand	category	department
1	Big Can	Drinks	Grocery
2	National Bottle	Drinks	Grocery
3	American Corn	Food	Grocery
4	Chewy Industries	Food	Grocery

Data Warehouse Bus Architecture

- Conformed dimension subsetting
  - Two dimensions are the same level of detail
  - One represents only a subset of rows

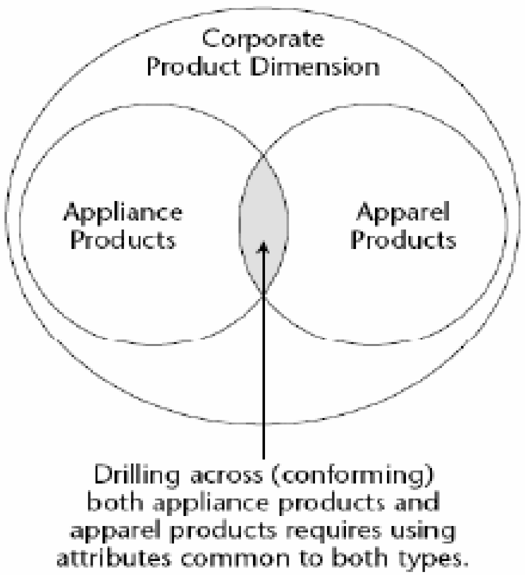


Figure 3.10 Conforming dimension subsets at the same granularity.

- Dimension authority
  - Responsibility for each conformed dimension
    - for defining, maintaining, and publishing

Data Warehouse Bus Architecture

- Conformed Facts
  - If facts exist in more than one place
    - then they must have the same name, units, and definition
  - If two facts are different
    - then give them different names

Summary

- Dimensional models for the three complementary view of inventory
  - Inventory Periodic Snapshot
  - Inventory Transactions
  - Inventory Accumulating Snapshot
- introduced key concepts
  - The data warehouse bus architecture and matrix
  - Conformed dimensions, the bus and the matrix