Introducing Dimensional Modelling

- What is dimensional modeling?
- Logical design technique for structuring data so that its intuitive to business users and delivers fast query performance.
- Divides world into measurement s and context
- Measurements captured by business processes
- Measurements are numeric values as facts
- Facts are surrounded by textual context
- Context is divided into logical groups called dimensions
- Dimensions describe who, what, when, where, why and how
- Dimensional models stored in a relational database are star schemas
- Dimensional models stored in multidimensional online analytical processing structures are called cubes

- What about normalized modeling?
- Data is divided into many discrete entities each becomes a table in rdbms
- Beneficial for transaction processing
- Used in ERP
- 3NF models as ER models
- Fact tables are normalized to 3Nf in a dimensional model

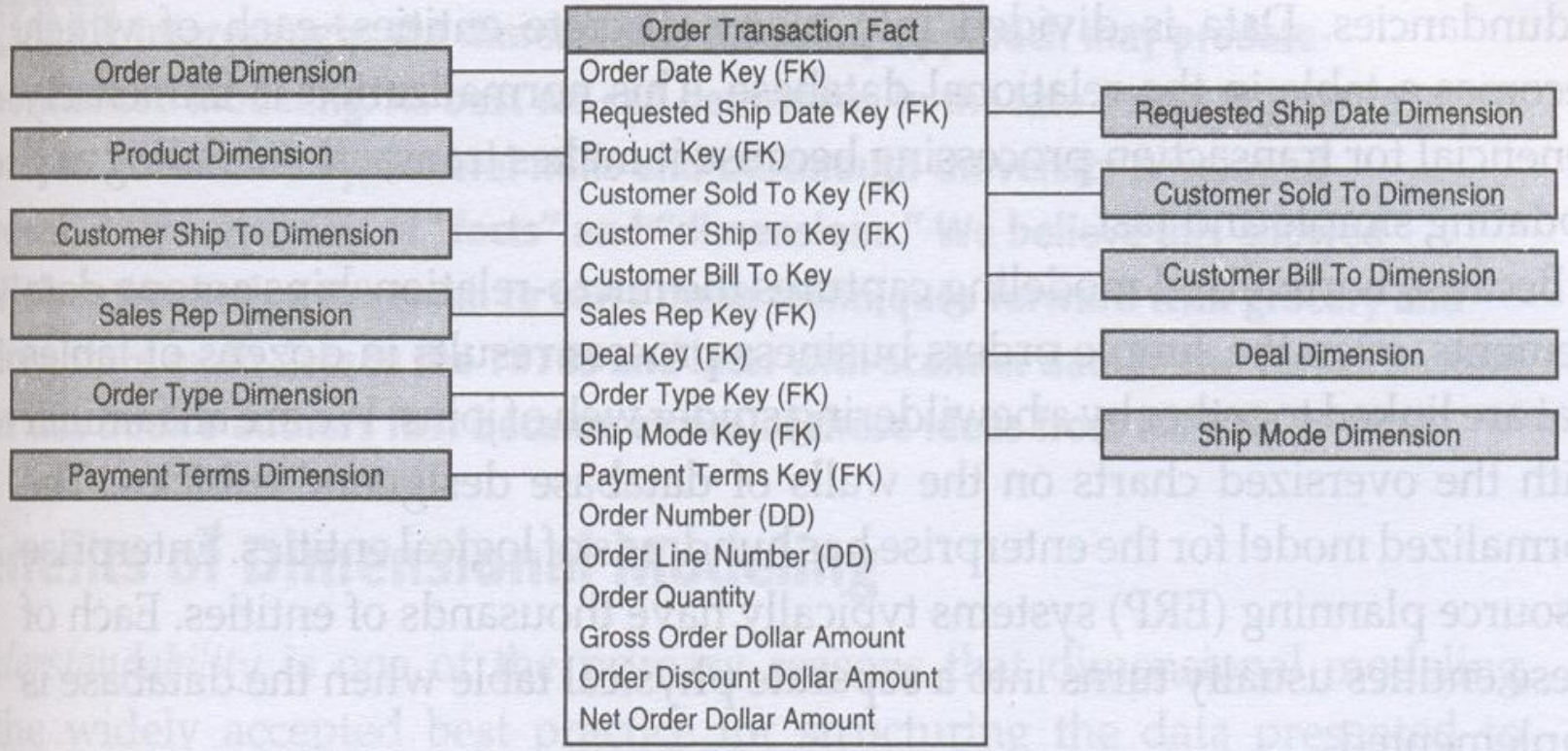


Figure 6-1 Dimensional model of the orders business process for a manufacturer.

Benefits of dimensional modelling

- Understandability
- Query Performance

Dimensional modeling primer

- Fact tables
- Fact represents performance measurement
- Nearly every fact is numeric
- Confirmed fact
- Fact table keys
- Fact table granularity

Dimension Tables

- -attributes serve two purposes
- -1)Query filtering
- -2)query result set labeling
- Power of DW proportional to quality and depth of dimension attributes
- -attributes are textual fields
- Attributes serve
- -Verbose(labels)
- -Descriptive
- -Complete
- -Discretely valued(take only one value in each row)
- -Quality assured(no misspellings, impossible values)
- Most dimensional models have 8 to 15 dimensional tables

Dimensional Table keys

- Primary keys are simple integers assigned in sequence starting with 1
- Primary keys are called as surrogate keys
- As join fields between fact and dimension tables
- 4 byte key represent 4GB integers
- Advantages of using surrogate keys
- Performance
- -compact surrogate keys translate into better performance
- Buffer from operational key management practices
- -due to limited retention needs operational keys may be reused but it creates problem in DW
- Mapping to integrate disparate sources
- -same entity is assigned different natural keys by different sources systems

- Handle unknown or not applicable conditions
- Track changes in dimension attribute values
- -slowly changing dimensions
- -assign next available key

Conformed Dimensions

- -shared across data warehouse environment
- -joining to multiple fact tables representing various business processes
- two types
- -1) basic type are absolutely identical with same keys, attribute name, attribute definitions and domain values
- -ex product dimension referenced by sales order fact table is identical to product dimension joined to inventory facts
- -2) Dimensions conform when one is perfect subset of a more detailed dimension table (for ex.refer next slide fig)

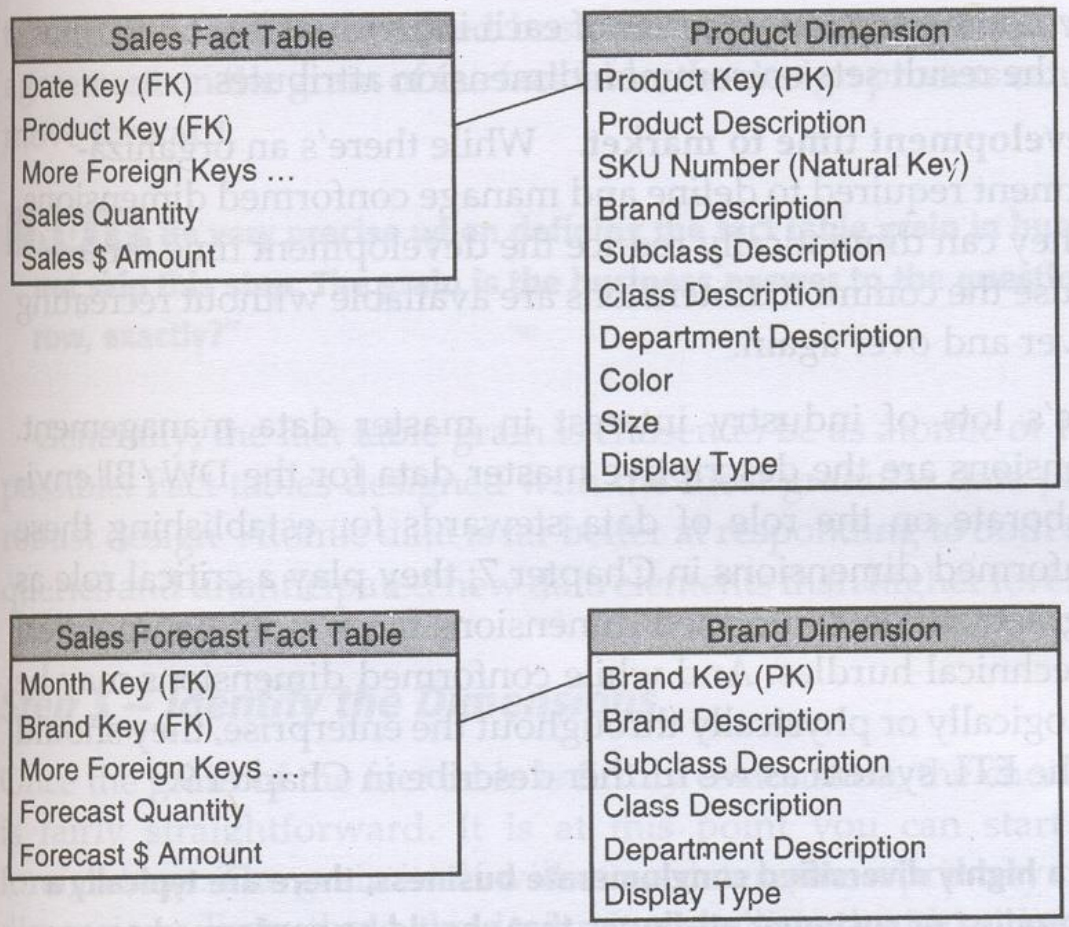


Figure 6-2 Conformed detailed and shrunken dimension tables corresponding to fact tables at different granularity.

Four step Dimensional Design Process

- Step 1 Choose Business Process
- Step 2 declare the Grain
- Step 3 identify the dimensions
- Syep 4 Identify the facts

Enterprise Data warehouse Bus Architecture

- Planning Crisis
- Manager is to understand content and location
- Every data item must be understood
- DW/BI manager develops independent solution
- Faces problem in integration
- To solve above problem we need Bus Architecture
- Row corresponds to business process
- Column corresponds to dimension of business

Value Chain Implications

- Raw material purchasing
- Raw material delivery
- Raw material inventory
- Bill of materials
- Manufacturing
- Shipping to warehouse
- Finished goods inventory
- Customer orders
- Shipping to customers
- Invoicing
- Payments
- returns

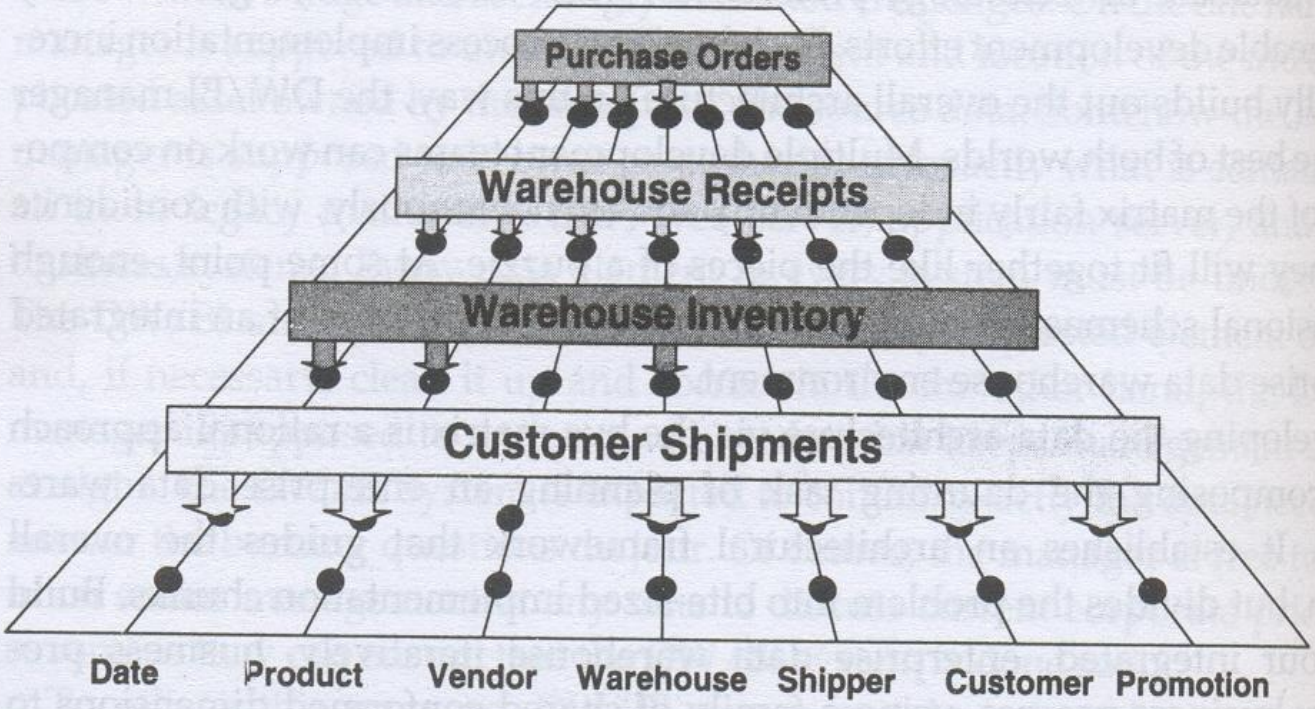


Figure 6-4 Diagram of data warehouse bus with conformed dimension interfaces.

	Date	Raw Material	Supplier	Plant	Product	Shipper	Warehouse	Customer	Sales Rep	Promotion Deal
Raw Material Purchasing	X	X	X	X		X				
Raw Material Delivery	X	X	X	X		X				
Raw Material Inventory	X	X	X	X						
Bill of Materials	X	X		X	X					
Manufacturing	X	X	X	X	X					
Shipping to Warehouse	X			X	X	X	X			
Finished Goods Inventory	X			X			X			
Customer Orders	X				X	X		X	X	X
Shipping to Customer	X				X	X	X	X	X	X
Invoicing	X				X		X	X	X	X
Payments	X				X			X	X	X
Returns	X				X	X		X	X	X

Figure 6-5 Bus matrix for manufacturing supply chain.

Common Matrix Mishaps

- Departmental rows
- Report centric or too narrowly defined rows
- Overly generalized columns
- Separate columns for each level of hierarchy

- Date and Time dimension
- Surrogate Date keys
- Time of Day

Slowly changing Dimensions

- Customer Key Name State
- 1001 Christina Illinois
- Type 1: The new record replaces the original record. No trace of the old record exists.
- Customer Key Name State
- 1001 Christina California
- Advantages:
- - This is the easiest way to handle the Slowly Changing Dimension problem, since there is no need to keep track of the old information.

- **Disadvantages:**
- - All history is lost. By applying this methodology, it is not possible to trace back in history. For example, in this case, the company would not be able to know that Christina lived in Illinois before.
- **Usage:**
- About 50% of the time.
- **When to use Type 1:**
- Type 1 slowly changing dimension should be used when it is not necessary for the data warehouse to keep track of historical changes.

Type 2 SCD

- Customer Key Name State
- 1001 Christina Illinois

- Customer Key Name State
- 1001 Christina Illinois
- 1005 Christina California

- **Advantages:**
- - This allows us to accurately keep all historical information.

- **Disadvantages:**
- - This will cause the size of the table to grow fast. In cases where the number of rows for the table is very high to start with, storage and performance can become a concern.
- - This necessarily complicates the ETL process.

- **Usage:**
- About 50% of the time.

- **When to use Type 2:**
- Type 2 slowly changing dimension should be used when it is necessary for the data warehouse to track historical changes.

Type 3 Slowly Changing Dimension

- Customer Key Name Original State Current State Effective Date
- 1001 Christina Illinois California 15-JAN-2003

- Advantages:
- - This does not increase the size of the table, since new information is updated.
- - This allows us to keep some part of history.

- Disadvantages:
- - Type 3 will not be able to keep all history where an attribute is changed more than once. For example, if Christina later moves to Texas on December 15, 2003, the California information will be lost.

- Usage:

- Type 3 is rarely used in actual practice.
- When to use Type 3:
- Type III slowly changing dimension should only be used when it is necessary for the data warehouse to track historical changes, and when such changes will only occur for a finite number of time.

Mini dimensions

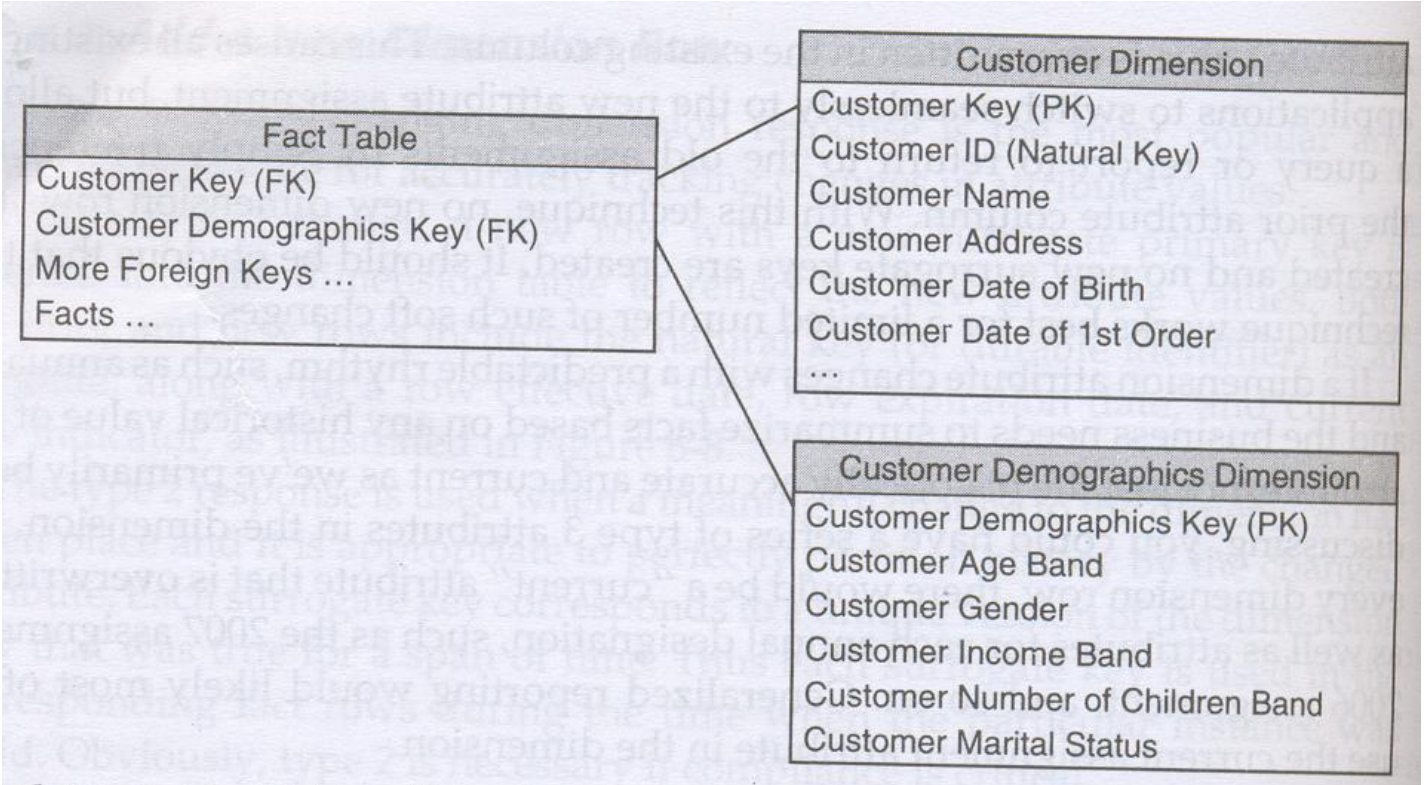


Figure 6-9 Fact table with a mini-dimension to track customer demographic changes.

Hybrid SCD

PRODUCT KEY	Product Description	Product Code	Historical Department	Current Department	Effective Date	Expiration Date	Current Row Ind
12345	IntelliKidz 1.0	ABC999-Z	Education	Critical Thinking	2/15/2007	5/31/2007	Not current
25984	IntelliKidz 1.0	ABC999-Z	Strategy	Critical Thinking	6/1/2007	12/31/2007	Not current
34317	IntelliKidz 1.0	ABC999-Z	Critical Thinking	Critical Thinking	1/1/2008	12/31/9999	Current

Figure 6-10 Hybrid slowly changing dimension with historical and current attributes.

Role playing Dimensions

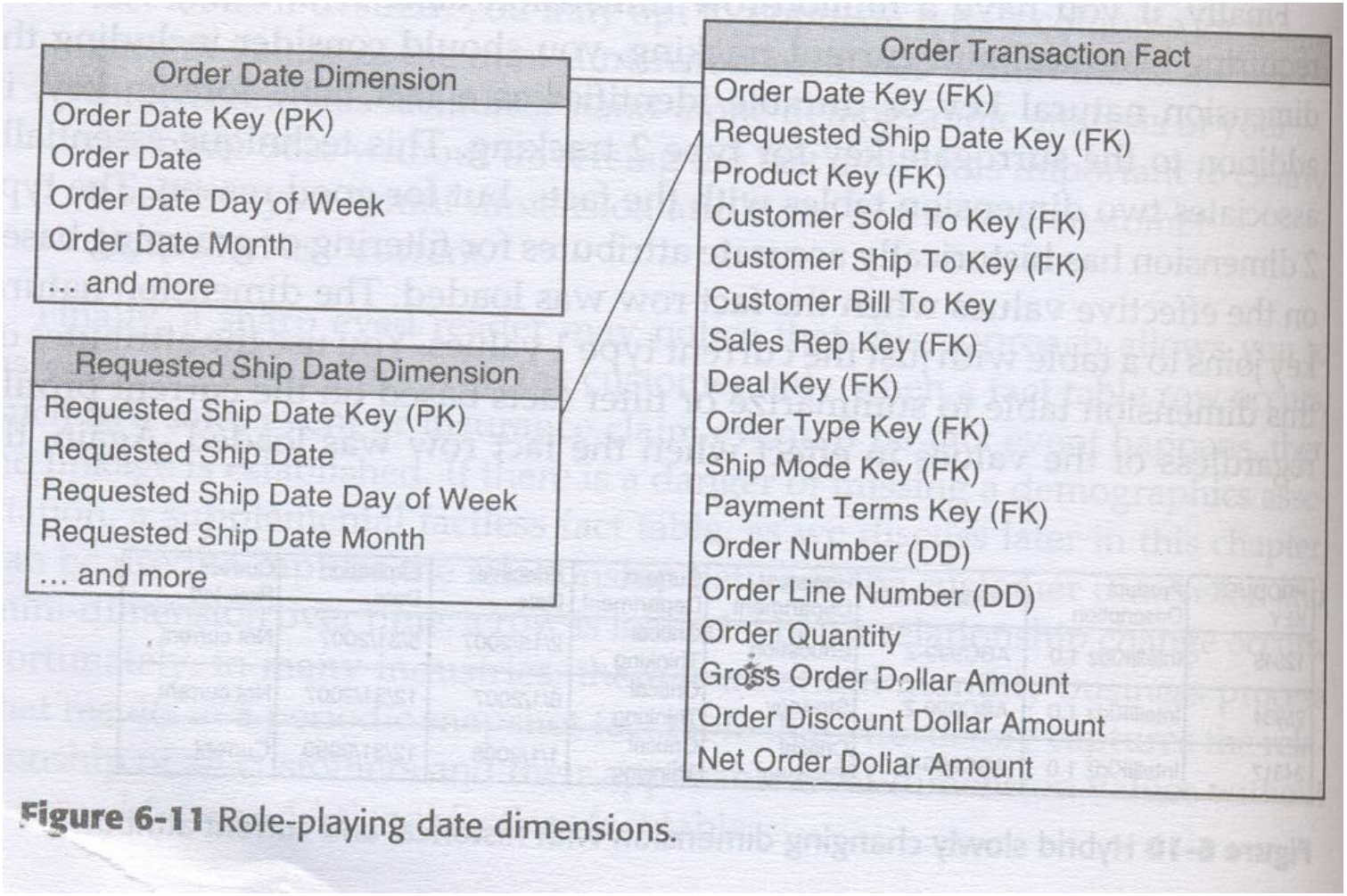


Figure 6-11 Role-playing date dimensions.

Junk Dimensions

Invoice Indicator Key	Payment Terms	Order Mode	Ship Mode
1	Net 10	Telephone	Freight
2	Net 10	Telephone	Air
3	Net 10	Fax	Freight
4	Net 10	Fax	Air
5	Net 10	Web	Freight
6	Net 10	Web	Air
7	Net 15	Telephone	Freight
8	Net 15	Telephone	Air
9	Net 15	Fax	Freight
10	Net 15	Fax	Air
11	Net 15	Web	Freight
12	Net 15	Web	Air
13	Net 30	Telephone	Freight
14	Net 30	Telephone	Air
15	Net 30	Fax	Freight
16	Net 30	Fax	Air
17	Net 30	Web	Freight
18	Net 30	Web	Air
19	Net 45	Telephone	Freight
20	Net 45	Telephone	Air
21	Net 45	Fax	Freight
22	Net 45	Fax	Air
23	Net 45	Web	Freight
24	Net 45	Web	Air

Figure 6-12 Sample rows from a junk dimension.

Snowflaking and outriggers

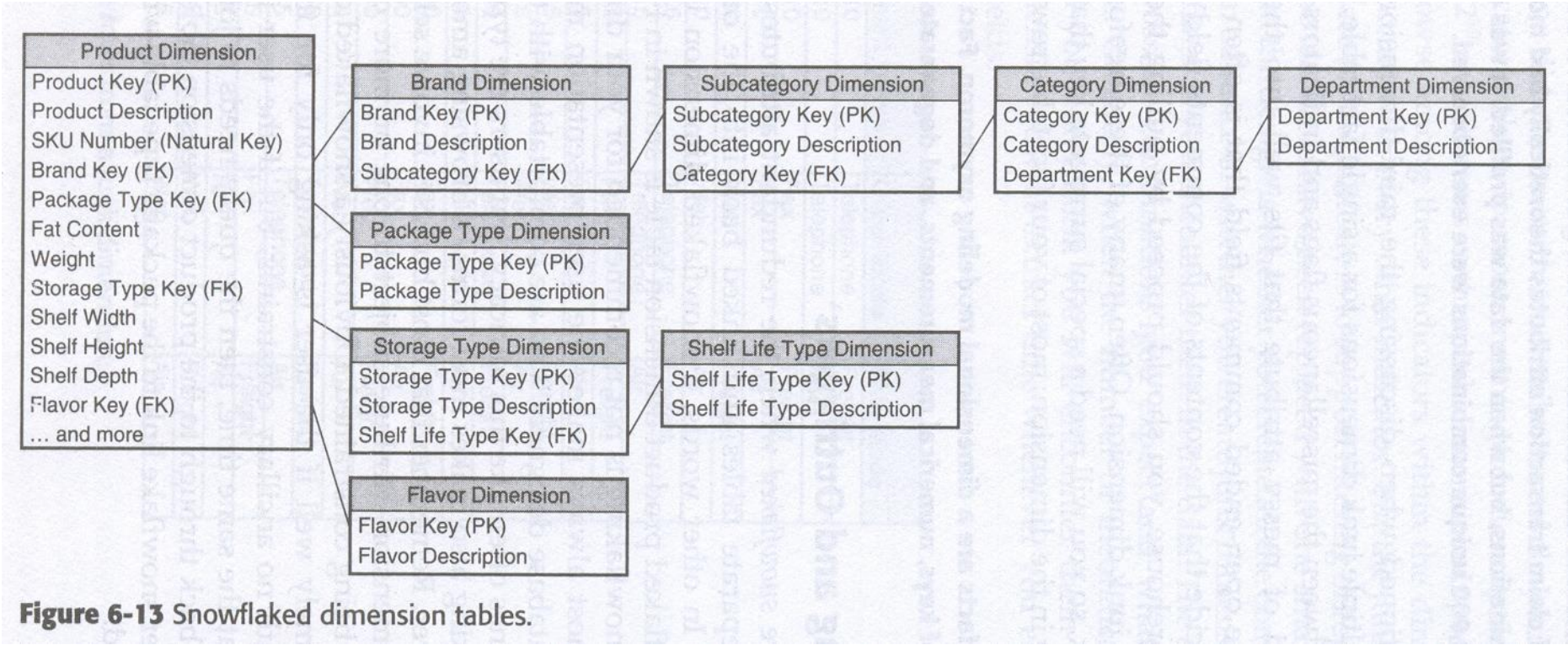
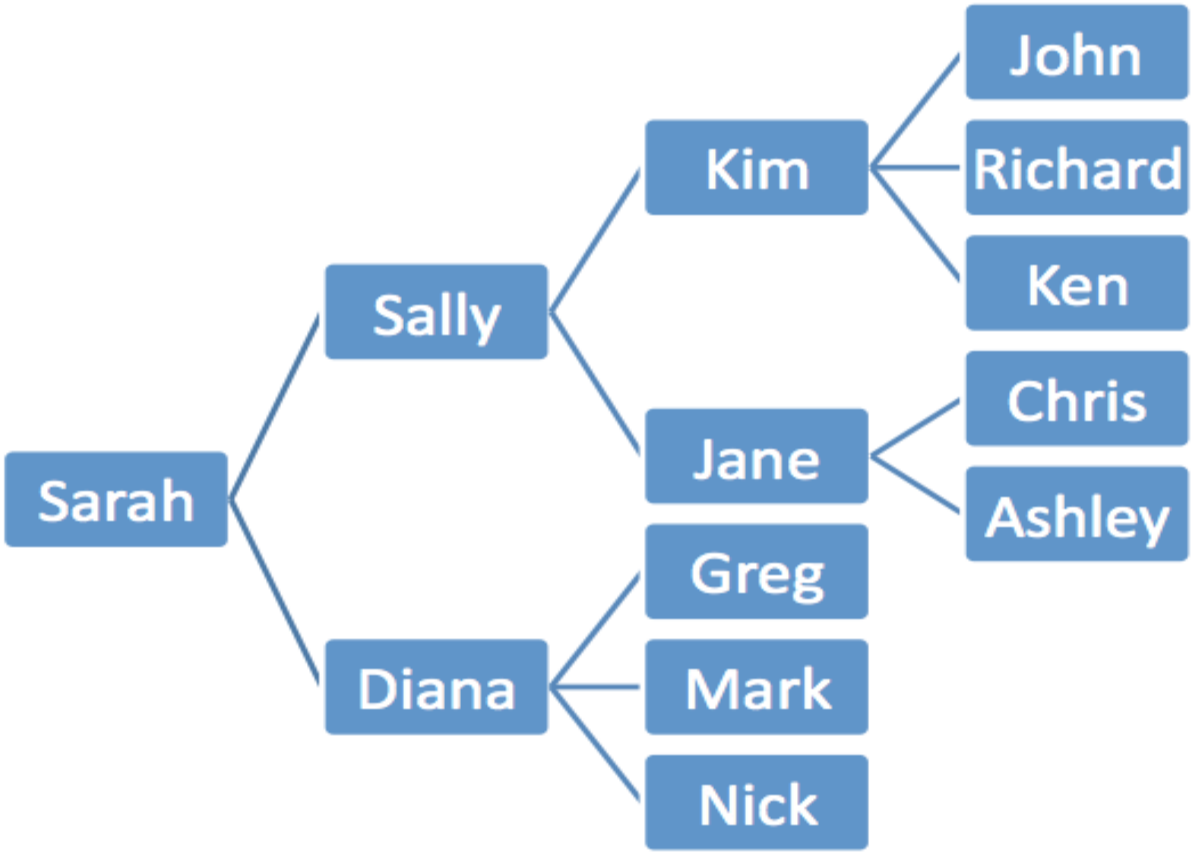


Figure 6-13 Snowflaked dimension tables.

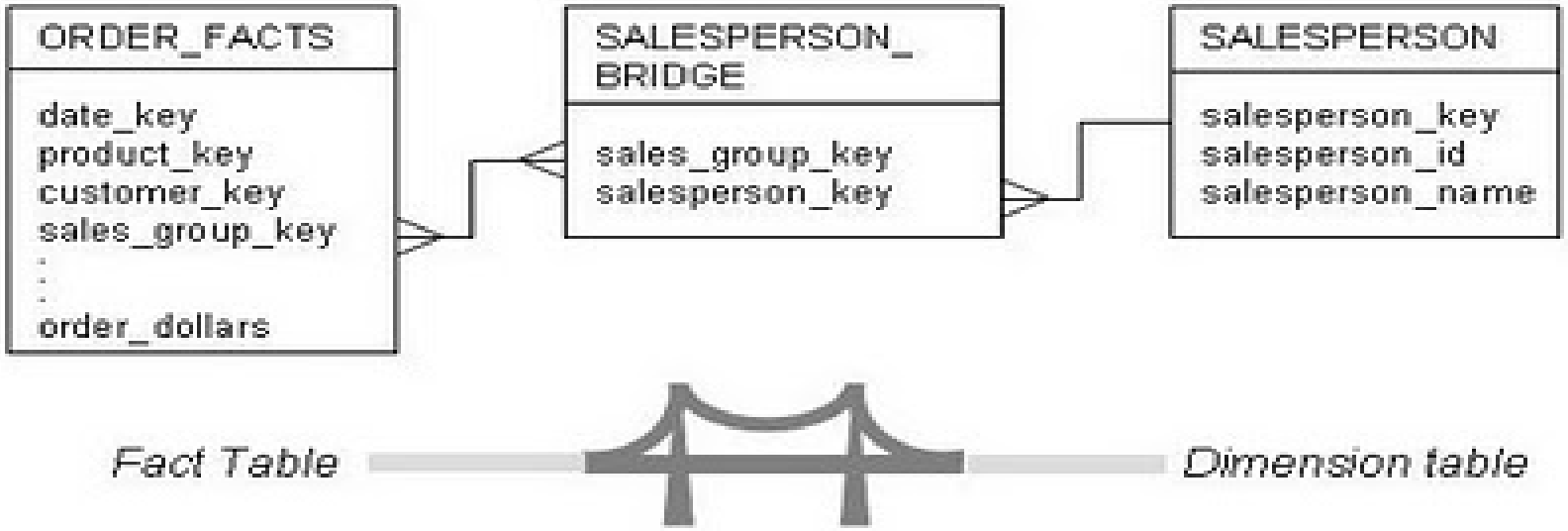
Handling Hierarchies via Bridge Table



person_id	supervisor_id	name	salary_per_anno
1		Sarah	100000
2	1	Sally	80000
3	2	Kim	65000
4	2	Jane	63000
5	1	Diana	75000
6	5	Greg	64000
7	5	Mark	63000
8	5	Nick	61000
9	3	John	57000
10	3	Richard	55000
11	3	Ken	54000
12	4	Chris	53000
13	4	Ashley	55000

person_id	supervisor_id	distance
…	…	…
9	9	0
9	3	1
9	2	2
9	1	3

Many Valued Dimensions with Bridge Table (when more than one sales persons in an order)



3 Fundamental Fact Table Grains

- Transaction
  - One row per transaction/line of transaction
  - Rows are inserted into fact tables only when a transaction activity occurs

3 Fundamental Fact Table Grains

- Periodic snapshot
  - At predetermined intervals snapshots of the same level of details are taken and stacked consecutively in the fact table
  - Example: most financial reports, bank account value
  - Complements detailed transaction facts but not substitutes them
  - Share the same conformed dimensions but have less dimensions

Three Fundamental Fact Grains

- Periodic snapshot Fact Tables

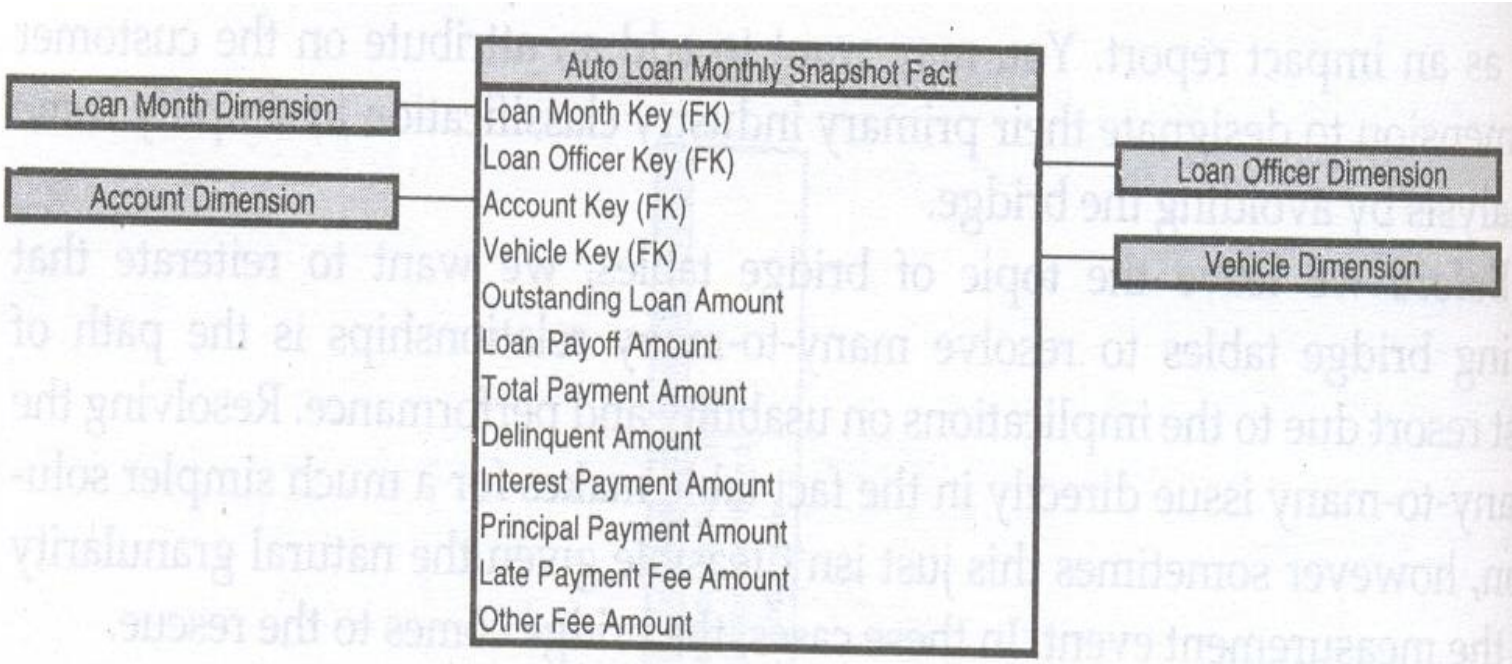


Figure 6-18 Periodic snapshot grained fact table for monthly loans.

3 Fundamental Fact Table Grains

- Accumulating snapshot
  - Less frequently used
  - Have multiple date FK that correspond to each milestone in the workflow
  - Lots of N/A or Unknown fields when a row is originally inserted
    - Requires a special row in date dimension table as discussed earlier

Accumulating snapshot fact tables

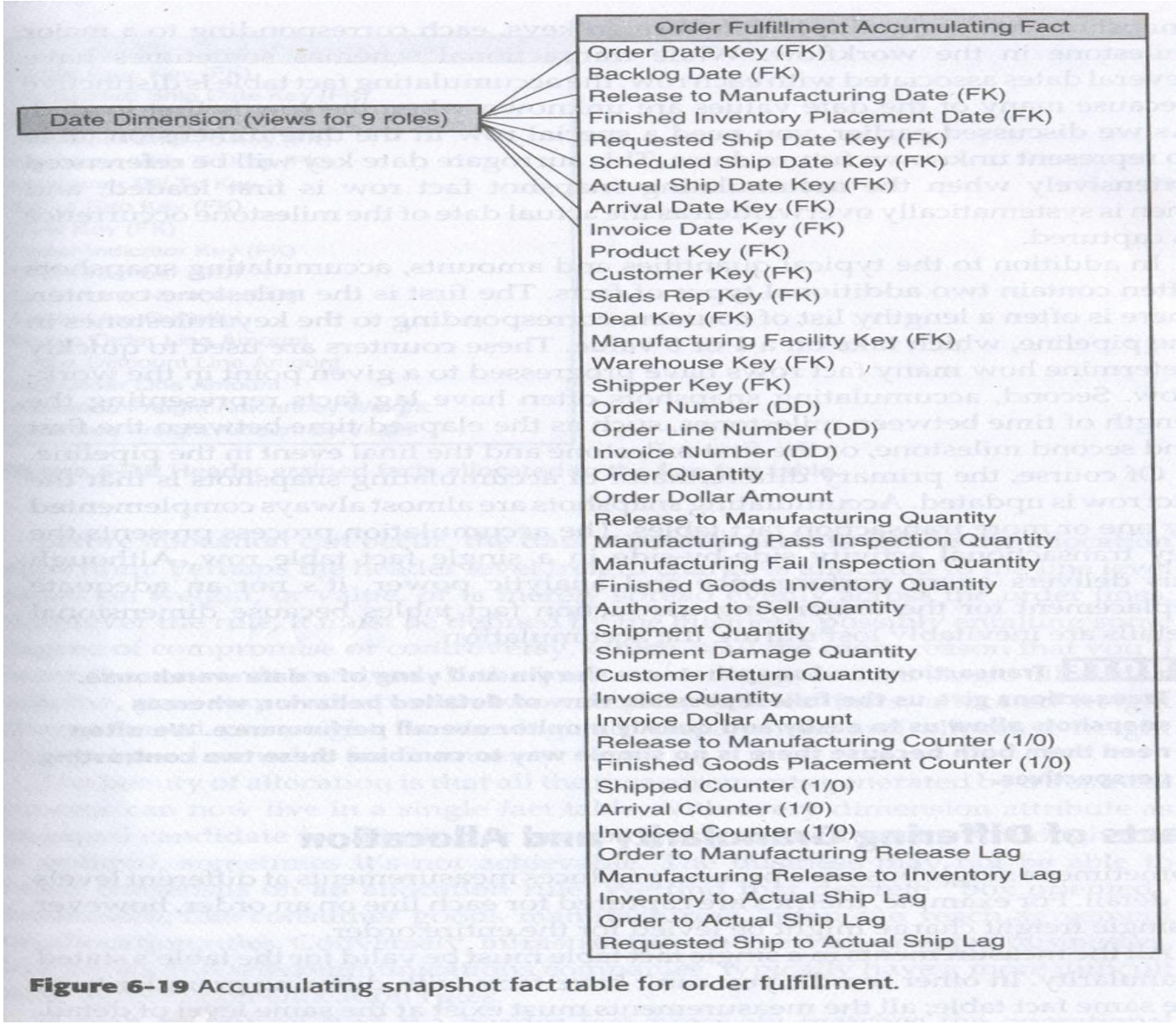
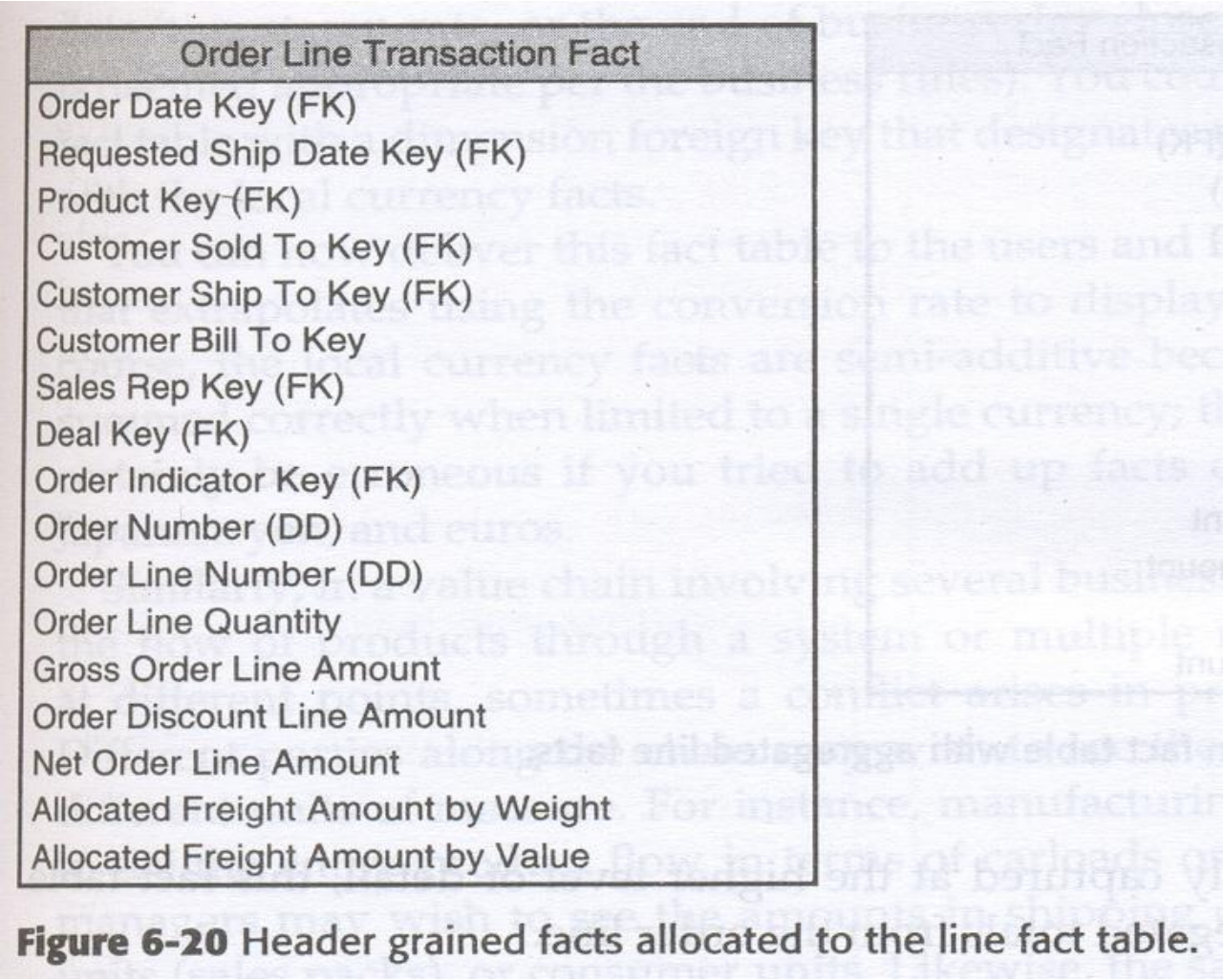


Figure 6-19 Accumulating snapshot fact table for order fulfillment.

Facts of Different Granularity

- A single fact table cannot have facts with different granularity
  - All measurements must be in the same level of details
  - Example:
    - Measurements are captured for each line order except for the shipping charge which is for the entire order
  - Solutions:
    - Allocating higher level facts to a lower granularity
    - Create two separate fact table

Facts of Differing granularity and Allocation
(Single freight might be levied on for entire order)



Order Line Transaction Fact
Order Date Key (FK)
Requested Ship Date Key (FK)
Product Key (FK)
Customer Sold To Key (FK)
Customer Ship To Key (FK)
Customer Bill To Key
Sales Rep Key (FK)
Deal Key (FK)
Order Indicator Key (FK)
Order Number (DD)
Order Line Number (DD)
Order Line Quantity
Gross Order Line Amount
Order Discount Line Amount
Net Order Line Amount
Allocated Freight Amount by Weight
Allocated Freight Amount by Value

Figure 6-20 Header grained facts allocated to the line fact table.

Order Header Transaction Fact
Order Date Key (FK)
Requested Ship Date Key (FK)
Customer Sold To Key (FK)
Customer Ship To Key (FK)
Customer Bill To Key
Sales Rep Key (FK)
Deal Key (FK)
Order Indicator Key (FK)
Order Number (DD)
Gross Order Header Amount
Order Discount Header Amount
Net Order Header Amount
Order Header Freight Amount

Figure 6-21 Header grain fact table with aggregated line facts.

facts that are naturally captured at the higher level of detail, this fact table could also store aggregated totals from the order line.

Multiple Currencies and Units of Measure

Sometimes fact measurements need to be expressed in multiple currencies or units of measure. In both cases, packaging the measurements and conversion factors in the same fact table row provides the safest guarantee that the factors will be used correctly.

In the case of currency conversion, the most common requirement is to provide the measurements in the local currency in which they were captured, as well as converted to the standardized international currency for the organization. As shown in Figure 6-22, you could store one set of measurements, along with the appropriate conversion rate fact (corresponding to the transaction

Physical structure:

Order Transaction Fact
Order Date Key (FK)
Product Key (FK)
Customer Ship To Key (FK)
Customer Bill To Key (FK)
Sales Rep Key (FK)
Deal Key (FK)
Local Currency Dimension Key (FK)
Order Number (DD)
Order Line Number (DD)
Order Quantity
Local Currency Gross Order Amount
Local Currency Order Discount Amount
Local Currency Net Order Amount
Currency Conversion Rate

Business view:

Order Transaction Fact
Order Date Key (FK)
Product Key (FK)
Customer Ship To Key (FK)
Customer Bill To Key (FK)
Sales Rep Key (FK)
Deal Key (FK)
Local Currency Dimension Key (FK)
Order Number (DD)
Order Line Number (DD)
Order Quantity
Local Currency Gross Order Amount
Local Currency Order Discount Amount
Local Currency Net Order Amount
Standard US Dollar Gross Order Amount
Standard US Dollar Order Discount Amount
Standard US Dollar Net Order Amount

Currency Dimension

Figure 6-22 Fact table supporting multiple currencies.

Multiple Currencies and Units of Measures

- Measurements are provided in a local currency
- Measurements are also converted to a standardized currency or conversion rates must be stored
- Similarly, in case of multiple units of measures, conversions to all different units of measure are provided

Physical structure:

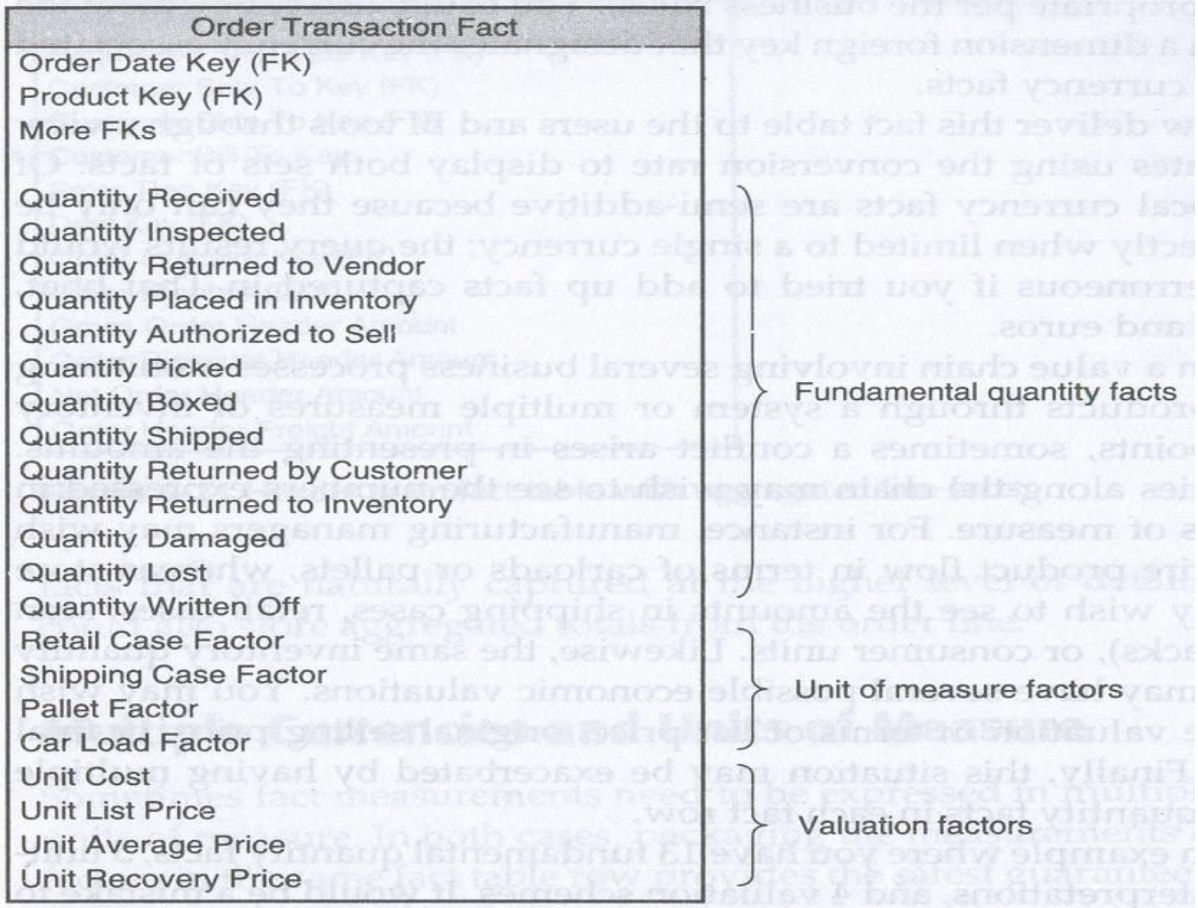
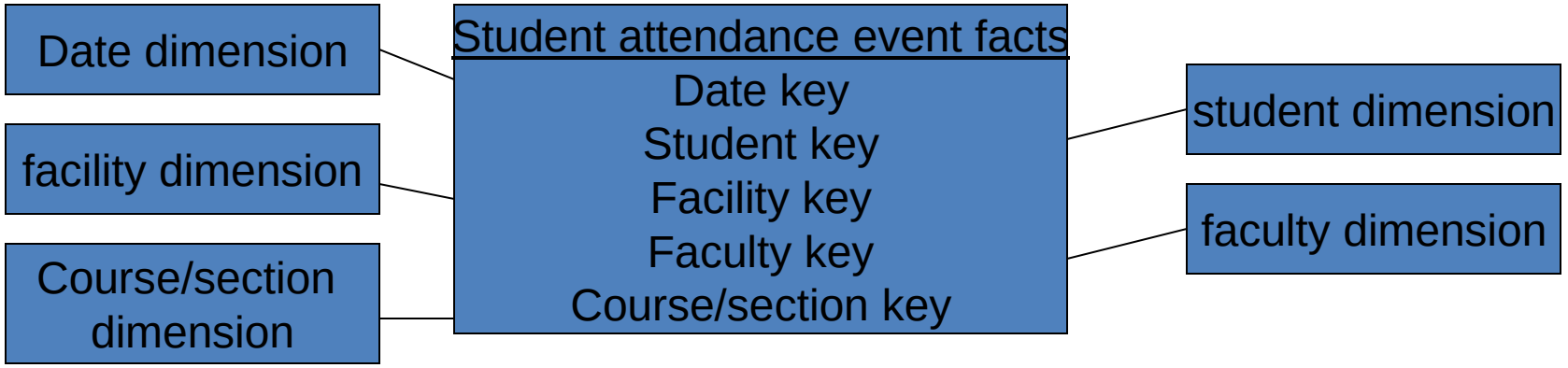


Figure 6-23 Fact table to support multiple units of measure.

Factless Fact Tables

- business processes that do not generate quantifiable measurements
- Example: student attendance



- Can be easily converted into traditional fact tables by adding an attribute Count, which is always equal to 1.
  - Helps to perform aggregations

Factless Fact Tables

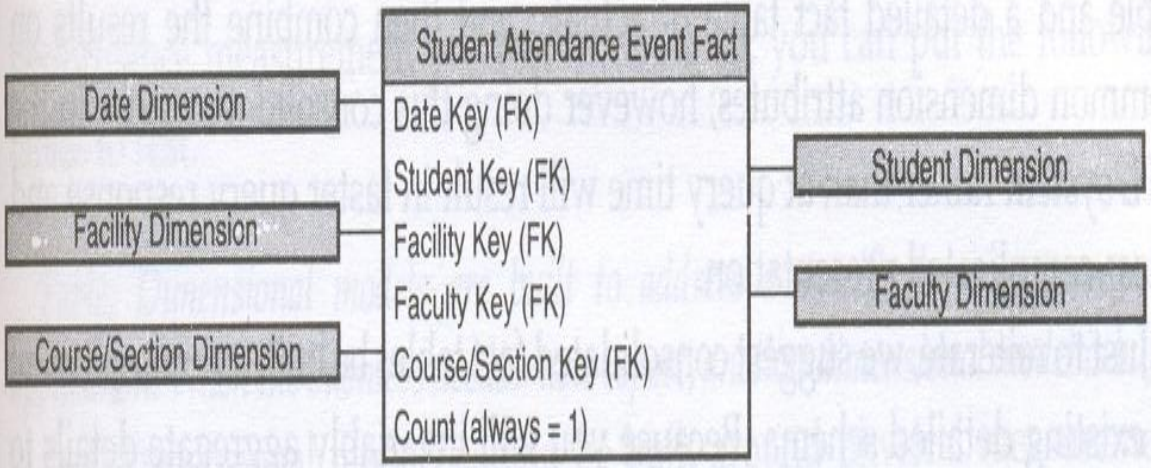


Figure 6-24 Factless fact table for student attendance events.

Consolidated Fact Tables

- Fact tables populated from different sources may potentially be consolidated into single one
  - Level of granularity must be the same
  - Measurements are listed side-by-side
  - Example: by combining forecast and actual sales amounts, a forecast/actual sales variance amount can be easily calculated and stored

Consolidated Fact Tables

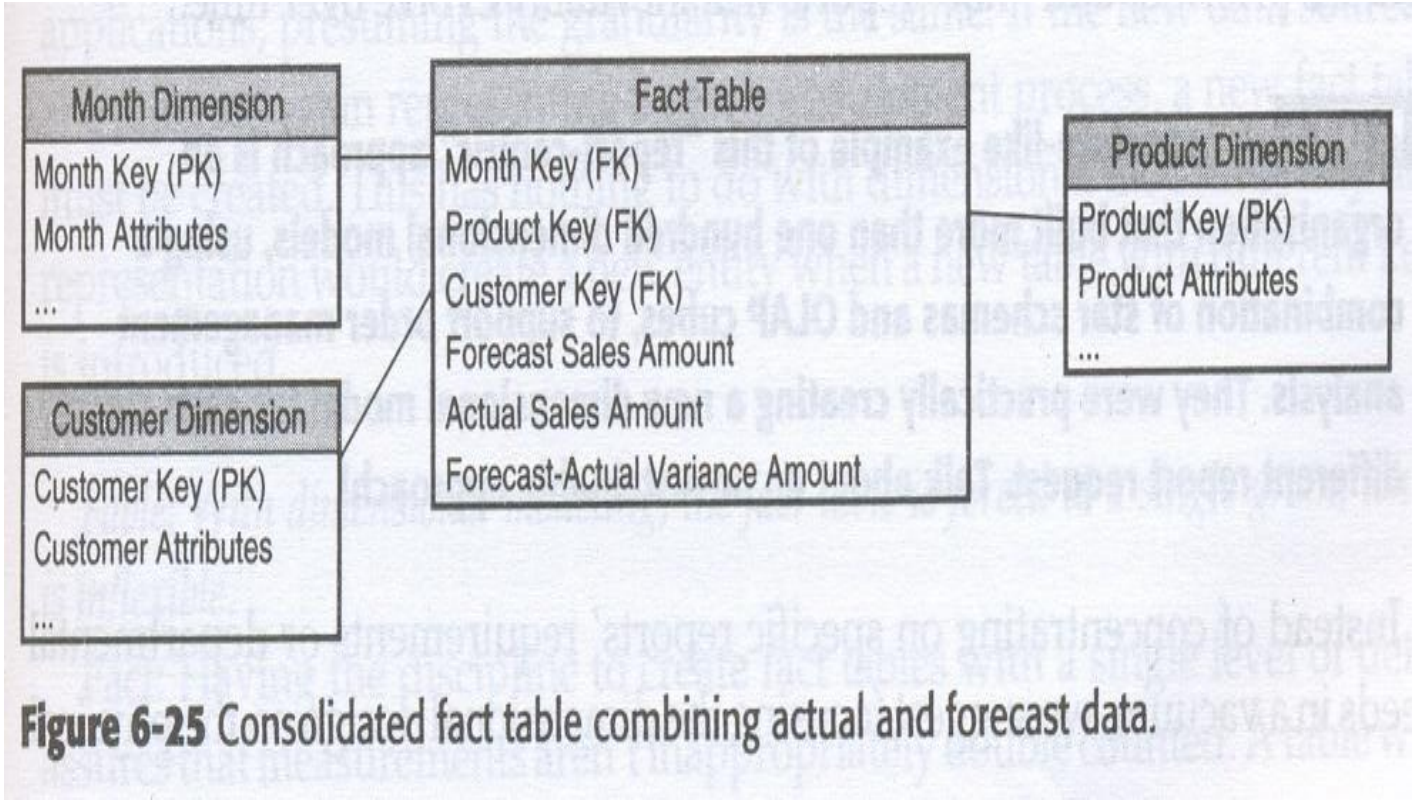


Figure 6-25 Consolidated fact table combining actual and forecast data.