

ACS-4904 Data Warehousing

Ron McFadyen

Chapter 1 : Analytic databases and dimensional design

Figure 1-1

Operational Systems
Vs
Analytic Systems

	Operational System	Analytic System
Purpose	Execution of a business process	Measurement of a business process
Primary Interaction Style	Insert, Update, Query, Delete	Query
Scope of Interaction	Individual transaction	Aggregated transactions
Query Patterns	Predictable and stable	Unpredictable and changing
Temporal Focus	Current	Current and historic
Design Optimization	Update concurrency	High-performance query
Design Principle	Entity-relationship (ER) design in third normal form (3NF)	Dimensional design (Star Schema or Cube)
Also Known As	Transaction System On Line Transaction Processing (OLTP) System Source System	Data Warehouse System Data Mart

Figure 1-1 Operational systems vs. analytic systems

Dimensional Design

- Facilitate the analysis of a business process
- Design focuses on how the process is measured
- Design based on 2 concepts:
 - measurements
 - context (facts and dimensions)

Figure 1-2. A sample report - Any report can be examined in terms of its metrics and the context used for describing the metrics.

Order Report Western Region January 2009 (cont'd)					
Category	Product	SKU	Quantity Sold	Cost	Order Dollars
Packaging	Box - Large	011-4822	700	\$ 950.53	\$ 1,100.00
	Box - Medium	011-4899	1,250	\$ 1,001.84	\$ 1,380.00
	Box - Small	011-5744	1,200	\$ 1,200.72	\$ 1,330.00
	Clasp Letter	011-1729	400	\$ 352.82	\$ 356.00
	Envelope #10	021-0011	2,000	\$ 2,017.46	\$ 2,080.00
	Envelope Bubble	021-0012	1,200	\$ 866.51	\$ 1,212.00
	<i>All Packaging</i>			\$ 6,389.88	\$ 7,458.00
Pens	Gel Pen Black	017-1999	5000	\$ 116.39	\$ 120.00
	Gel Pen Blue	017-2444	2990	\$ 600.88	\$ 624.91
	Silver Pen	017-3001	50	\$ 128.46	\$ 130.00
	<i>All Pens</i>			\$ 845.73	\$ 874.91
Grand Total				\$207,229.42	\$214,896.91
Page 10 of 10					

Figure 1-2 Sample page from an orders report

Figure 1-3. A separation of attributes into facts and dimensions determined by considering how the ordering process is to be evaluated

Facts	Dimensions
Order Dollars	Product
Cost Dollars	Product Description
Quantity Ordered	SKU
	Brand Code
	Brand
	Brand Manager
	Category Code
	Category
	Order Date
	Month of Order
	Quarter of Order
	Fiscal Period of Order
	Year of Order
	Salesperson
	Salesperson ID
	Territory
	Territory Code
	Territory Manager
	Region
	Region Code
	Region VP
	Customer
	Customer ID
	Headquarters State of Customer
	Billing Address
	Billing City
	Billing State
	Billing Zip
	SIC Code of Customer
	Industry Name of Customer

Figure 1-3 Facts and dimensions for the orders process

Figure 1-4. One possible logical dimensional design

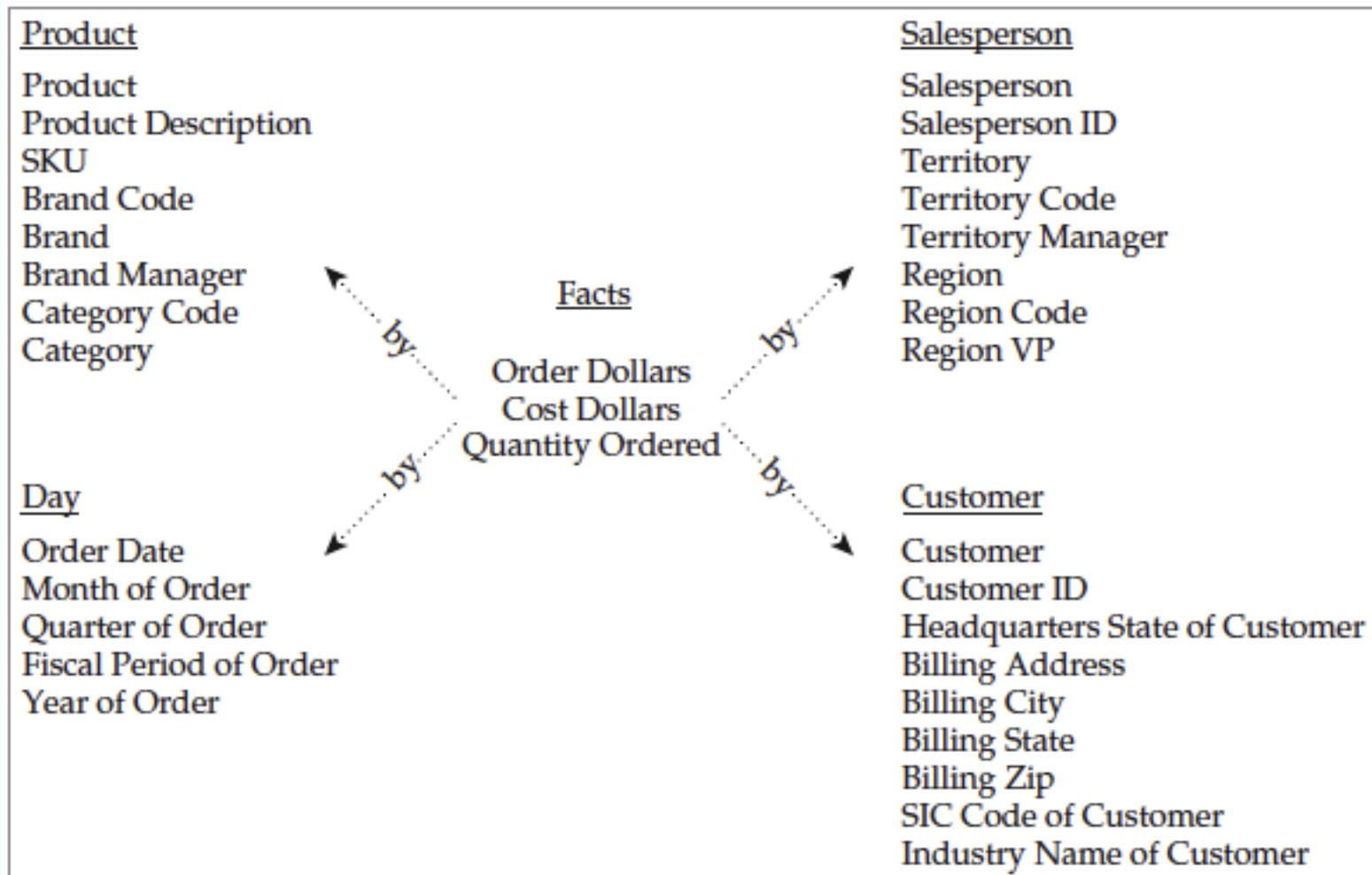


Figure 1-4 Sorting out facts and dimensions for the orders process

Figure 1-5. Physical view showing foreign keys. Not normalized

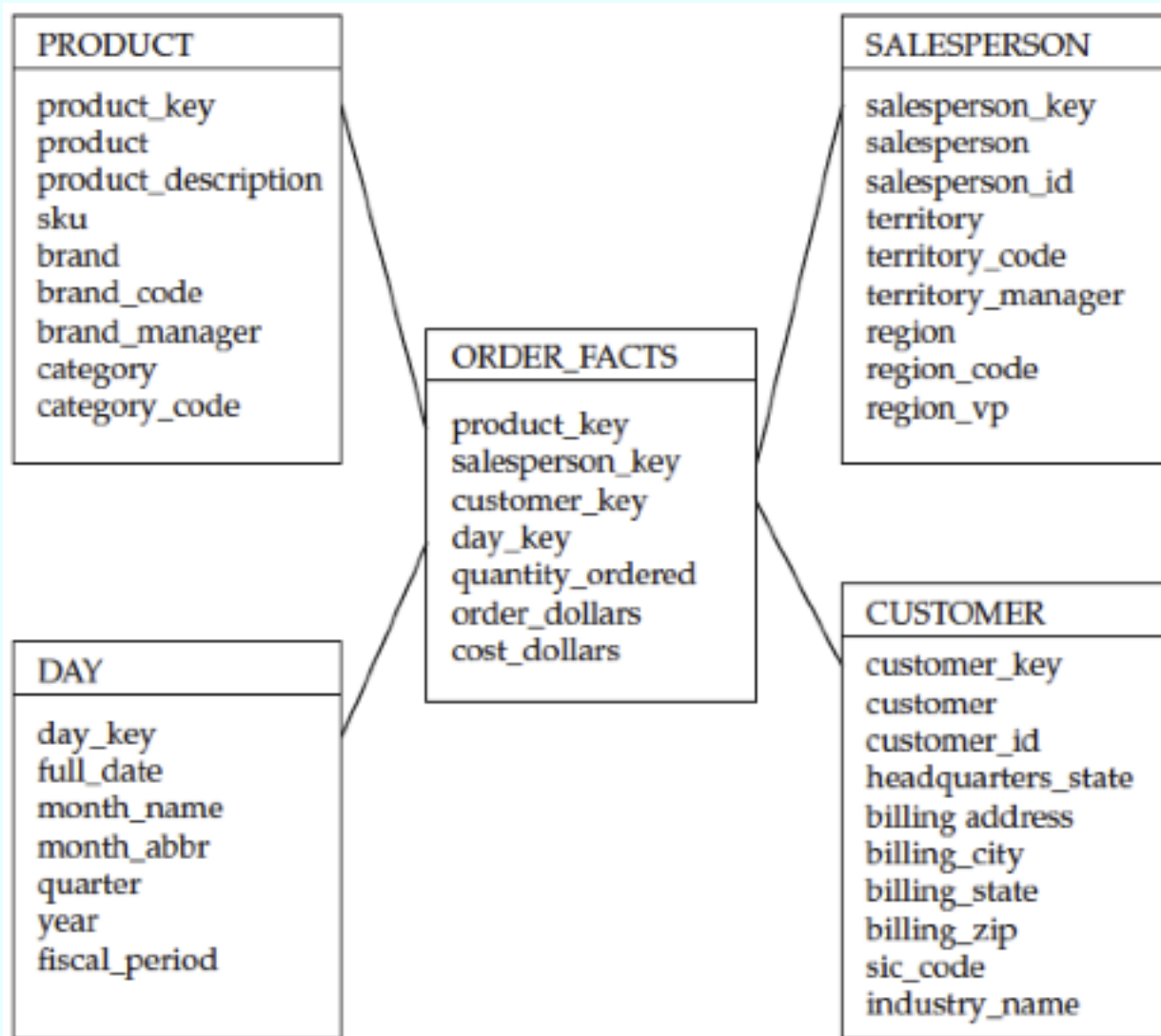
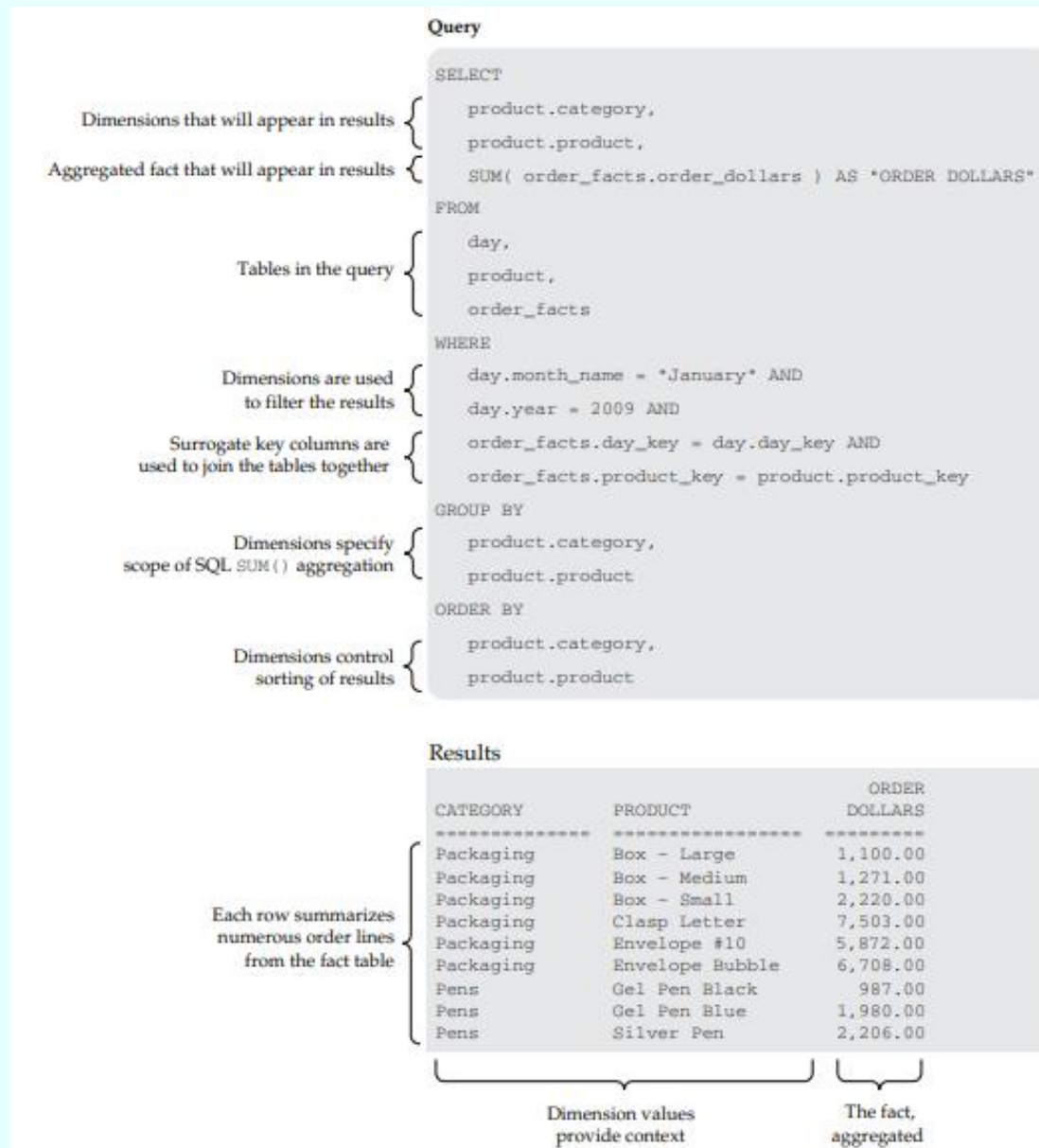


Figure 1-5 A simple star schema for the orders process

Figure 1-6. Order dollars by product, category



If dimensions are normalized then the design is referred to as a *snowflaked* design and the additional tables are referred to as *outriggers*.

Keys

Primary Keys

Natural keys

Surrogate keys

PK for Fact tables

Sometimes the catenation of FKs

Sometimes a subset of the FKs is used - or even some other attributes may be involved

Primary Keys:

Each dimension has a surrogate key that serves as the PK

PKs are assigned by ETL processes

An IT organization should have naming standards

... the text indicates each surrogate key using the `_key` suffix

Natural keys:

The identifiers that serve as PKs in the operational system

E.g. `customer_id`, `sku`, `employeeID`, `studentNumber`, ...

Surrogate keys are used in a warehouse to track changes in dimensions

- A natural key alone cannot serve as a unique identifier if we allow for multiple rows in a dimension to refer to the same entity.
- Ch 8 More slow change techniques

PK for Fact tables

Sometimes the concatenation of FKs serves as the fact table's PK - Figure 1-5

Sometimes a subset of the FKs is used - or even some other attributes may be involved