Business School
UNIVERSITY OF COLORADO DENVER

Information Systems Program

Module 3
Data Warehouse Design Practices and Methodologies

Lesson 5: Mini Case for Data Warehouse Design

Lesson Objectives

- Practice with data warehouse design problems
- Prepare for data warehouse design assignment
- Gain insights about analyzing data sources

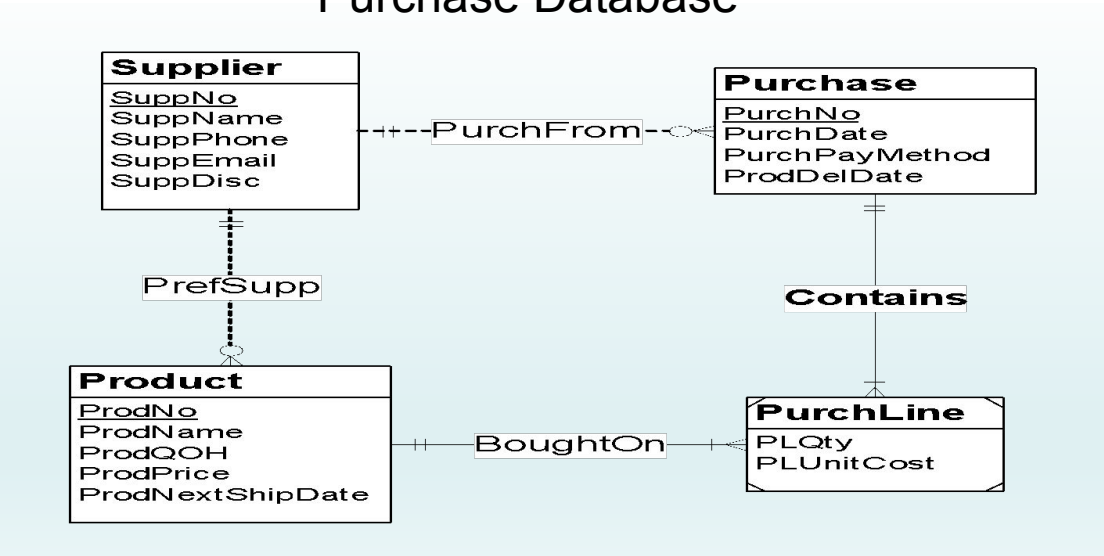
Mini Case on Data Warehouse Design

- Apply and integrate skills from module 3 lessons
- Acquire new skills
- Data source specifications, business needs, and sample data

- Design Requirements
  - Specify dimensions and measures
  - Determine grain
  - Create table design
  - Identify summarizability problems and suggest resolutions
  - Map data sources and populate tables

Data Sources

Purchase Database



Purchases Spreadsheet for Custom Products

ProdCode	ProdDesc	Supp	Qty	Stock	Unit Price	PurDate	Amount
CPC1	Souvenir 1	Omart	20	1	$2.00	13-Feb-2014	$40.00
CPC2	Souvenir 2	Smart	10	2	$3.50	14-Feb-2014	$35.00
CPC3	Souvenir 3	Pmart	20	0	$1.50	11-Feb-2014	$30.00

Business Intelligence Needs

- Track inventory over time by product and supplier
- Calculate inventory measures over time using quantity on hand and value
- Report on additions to inventory (purchases)
- No reporting on deletions to inventory (orders)

Important Design Decisions

- Grain determination and relative size calculations
- Simplification
- Mappings from source data to populate data warehouse tables

Grain Size Calculations

- **Fact table size**
  - **Use sizes of dimensions and sparsity cardinality estimate**
  - **Fill Ratio: 1 - Sparsity**
  - **Fact Table Size: Product of dimension sizes times fill ratio**
- **Sparsity**
  - **Match fact table to source tables**
  - **Use sizes of dimensions and source table**
  - **Fill Ratio: Source table size divided by product of dimension table sizes**
  - **Sparsity: 1 – Fill Ratio**

Mappings from Source Data

Associations

- Source column matching
- Conversions

Additions

- Generated PK values
- Default values
- Derived values

Data Warehouse Design Assignment

- Similar to design exercise
- Artifacts
  - Dimensional design with dimensions and members
  - ERD integrating data sources
  - Grain analysis
  - Summarizability problems and resolutions
  - Mapping from data sources
  - Population of DW tables using sample data from data sources

Summary

- Mini case study to help apply and integrate concepts and skills
- Case study requirements and data sources
- Concept extensions
  - Grain size
  - Mapping source data to data warehouse

Grain Size Determination

- Determine sparsity
  - Given dimension cardinalities and source table cardinality
  - Associate fact table to tables of data source
  - 1 minus source table cardinality divided by product of dimension cardinalities
- Determine fact table size
  - Given dimension cardinalities and sparsity estimate
  - Product of dimension cardinalities
  - Reduce by sparsity