Delivering a Relational Data Warehouse

Week 1 – Introducing the Data Warehouse

Module 02

Planning for a
Data Warehouse Implementation

Microsoft

Module Outline
02 | Planning for a Data Warehouse Implementation

Topic
► Data Warehouse Planning
► Gathering Requirements
► Data Warehouse Architectures
► **Demo:** Exploring the AdventureWorks Databases

Module Outline
02 | Planning for a Data Warehouse Implementation

Topic
Data Warehouse Planning
Gathering Requirements
Data Warehouse Architectures
Demo: Exploring the AdventureWorks Databases

Data Warehouse Planning

- The successful implementation of a data warehouse depends on good planning
  - Specifically, to manage the integration of numerous associated tasks and components
- A well recognized example of a mature approach to design, develop and deploy a data warehouse is the **Kimball Business Dimensional Lifecycle**

Data Warehouse Planning
Kimball Business Dimensional Lifecycle
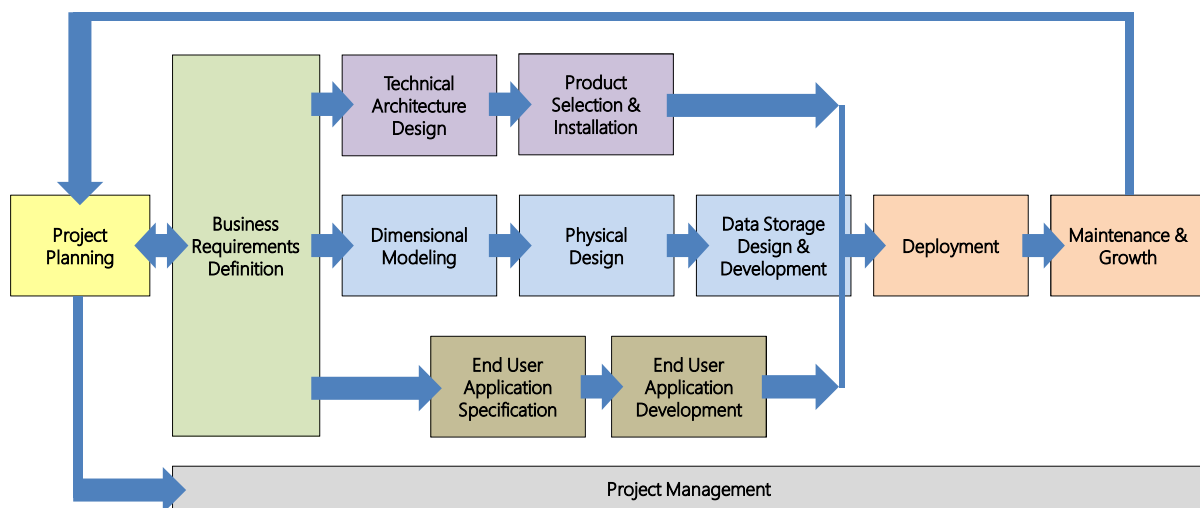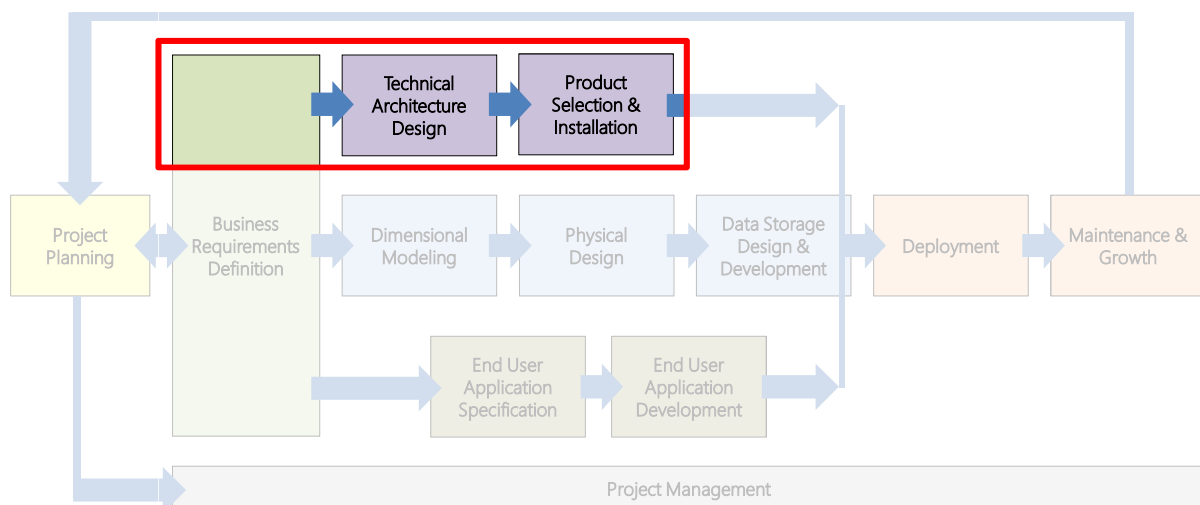

Source: The Data Warehouse Lifecycle Toolkit

Data Warehouse Planning
Business Requirements

- The greatest risk to the success of a data warehouse project is lack of trust, which leads to the lack of adoption
- Data warehouse designers must have a sound understanding of the business, and the end user requirements
- This understanding then leads to parallel tracks focused on:
  - Technology
  - Data
  - End user access

Data Warehouse Planning
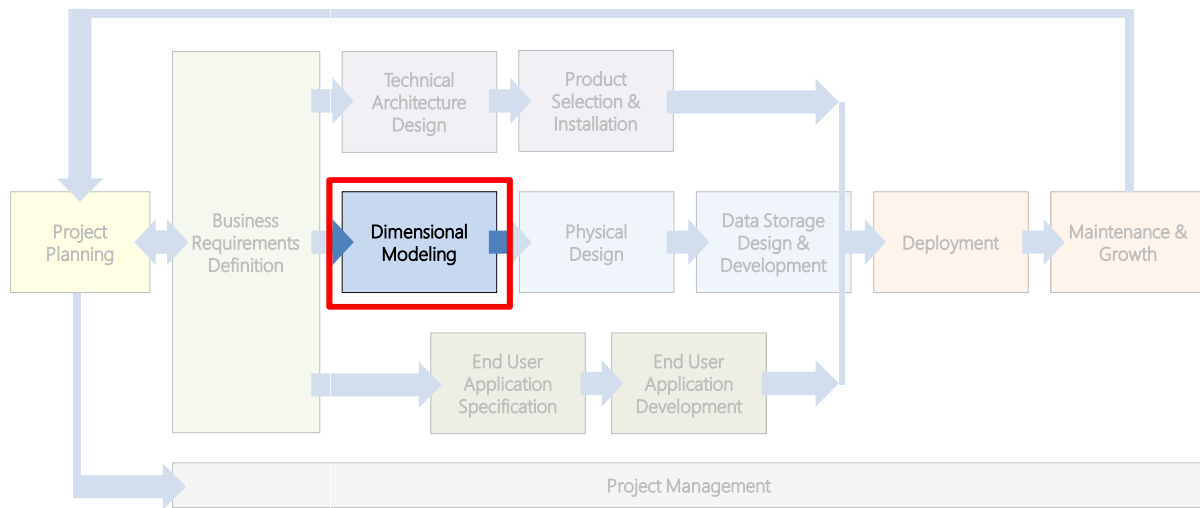Course Flow ► Week 1 ► Introducing the Data Warehouse

Project Planning	Business Requirements Definition	Technical Architecture Design	Product Selection & Installation

Source: The Data Warehouse Lifecycle Toolkit
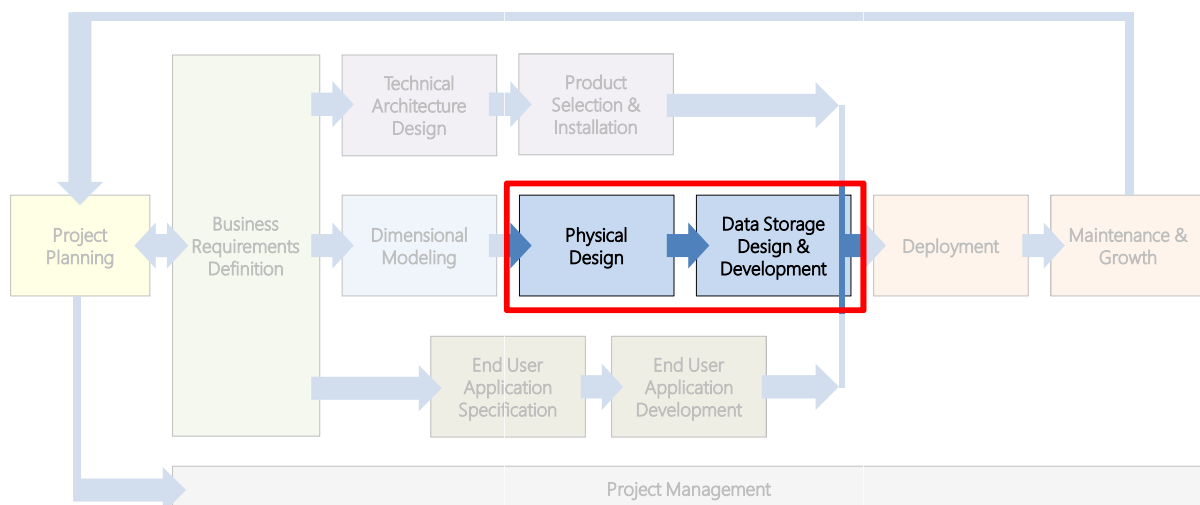
Data Warehouse Planning

Course Flow ► Week 2 ► Designing a Relational DW Schema

Source: The Data Warehouse Lifecycle Toolkit
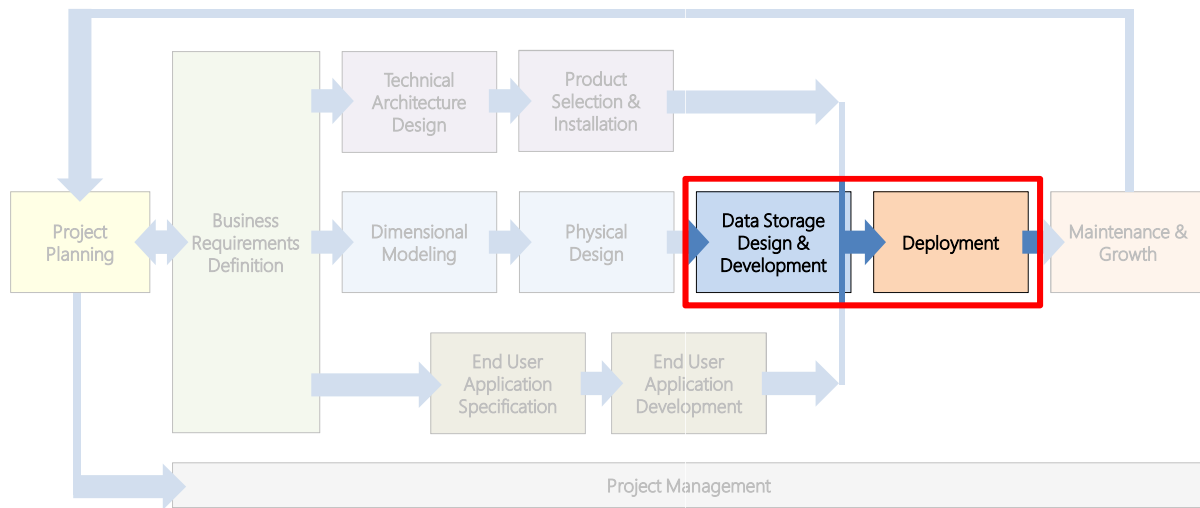
Data Warehouse Planning

Course Flow ► Week 3 ► Optimizing a DW for Scale and Performance

Source: The Data Warehouse Lifecycle Toolkit

Data Warehouse Planning
Course Flow ► Week 4 ► Loading and Maintaining a Data Warehouse

Source: The Data Warehouse Lifecycle Toolkit

Data Warehouse Planning
Course Flow ► Scope

- This course focuses only on the delivery of relational data warehouse solutions
  – It does not focus on OLAP data modeling, or end user applications, which are covered in different courses

- Several mature data warehouse architectures exist

- This course will refer to those published by **The Kimball Group**
  – A comparison of architectures will be covered in this module
  – This does not imply that other architectures are unsuitable, or cannot be delivered with Microsoft products and services

Data Warehouse Planning
Experience From the Field

- Assemble a team of skilled professionals, and manage the project like any other major undertaking
- Where possible, strive to future-proof the data warehouse design
  – Design for extensibility, as change is inevitable (new sources, new formats, new columns, new calculations, etc.)
  – Know that data volumes almost always grow
  – Recognize that data warehouse maintenance will be required

- Understand that the end users know best, though they often require guidance to elicit what they genuinely need

Data Warehouse Planning
Experience From the Field (Continued)

- Do not attempt to deliver and entire enterprise data warehouse in a single project
  – Plan to deliver single data marts by an agreed priority and timeline

- Demand top-level organizational support for the project, so to avoid internal politics denying access to data, overriding sensible design, and other hindrances
- Be prepared to re-think the data warehouse logical and physical designs based on newer technology or evolving requirements

Module Outline
02 | Planning for a Data Warehouse Implementation

Topic
Data Warehouse Planning
Gathering Requirements
Data Warehouse Architectures
Demo: Exploring the AdventureWorks Databases

Gathering Requirements

- Business requirements impact on every aspect of the data warehouse project
- Listening to the right people, and asking the right questions will help you to
  - Understand the business, and
  - Produce an accurate list of requirements and priorities
- This is usually best achieved by conducting interviews, which can be formal or informal

Gathering Requirements
Interviews

- Always start with the project sponsor, and ensure that the scope is clear, understood and achievable
  - Also enquire as to whether there have been other data warehouse projects—successful or not
  - If there have been other projects, be sure to learn about their outcomes
- Next, interview end users, sampling across the organization, and at different levels of the organization (executives to analysts)
  - Learn about their particular job challenges and objective, and understand how they make decisions
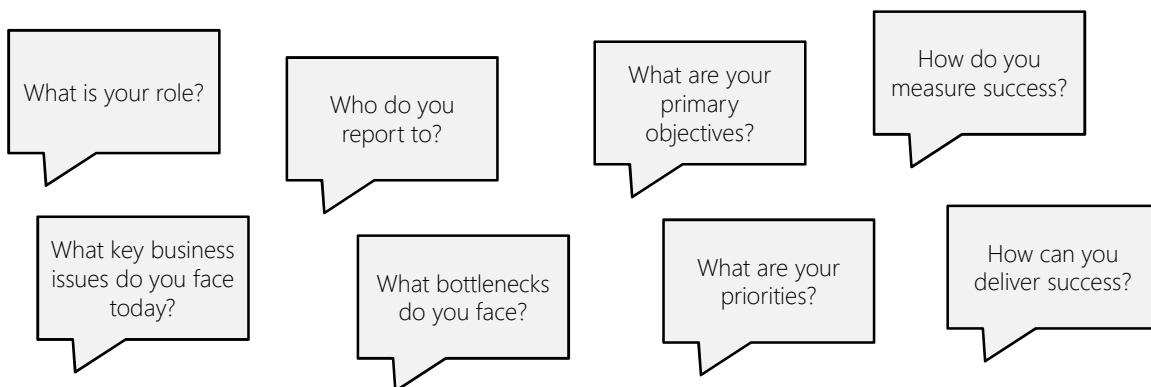
Gathering Requirements
Interviews (Continued)

- Once a consistent theme arises from the business, commence more formal discussions with IT professionals
  - This can include business analysts, DBAs, source system experts, data stewards, and software developers
- This should lead to an understanding of current systems, infrastructure, and data audits to perform a high-level mapping of data to requirements

Gathering Requirements
Interviews ► Questions

- Consider starting with interview questions to learn about their particular job challenges and objectives

What is your role?

Who do you report to?

What are your primary objectives?

How do you measure success?

What key business issues do you face today?

What bottlenecks do you face?

What are your priorities?

How can you deliver success?

Gathering Requirements
Interviews ► Questions (Continued)

- Next, ask questions to understand their data requirements…

Gathering Requirements
Interviews ► Questions (Continued)

What reports do you presently rely on?

Who produces these reports?

How often are these reports published?

How are these reports delivered to you?

Who else should have access to these reports?

What ad hoc analysis do you do, and who requests this?

How much history data do you require?

How do you report by time period?

Could these reports be improved?

How much data latency can you tolerate?

How do you categorize large sets of entities (e.g. products)?

How do you define business entities?

Who should not have access to these reports?

Gathering Requirements
Interviews ► Responses

- The interview responses, when collated, usually provide a good start to this phase of project, and consistent themes usually emerge
  - This often leads to another round of questions to refine questions based on others' responses
  - Usually by now, the word "by" is revealing potential dimensions (or levels), and a list of business measures (and calculations) is emerging

- When responses reveal the existence of data assets, obtain access to, or copies of, these assets

Gathering Requirements
Interviews ► Responses ► Data Assets

- Existing data assets can be in the form of:
  - Reports (automatically, or manually produced)
  - Spreadsheets
  - Data extracts
  - Access databases
  - Self-Service BI solutions

- Note that end users (non-IT) can produce surprisingly complex solutions to get their job done

Gathering Requirements
Interviews ► Responses ► Data Assets (Continued)

- Data assets are usually extremely useful in allowing you to determine:
  - Data source
  - Data ownership
  - Dimensionality
  - Grain
  - Business logic (e.g. Excel formulas)
  - Data presentation

Gathering Requirements
Recommended Practices

- Ask the right people the right questions
- Listen carefully
- Repeat asking questions until clear requirements have emerged
- Do not focus only on the top requests
  - Less important requests can have an equal bearing on design
  - Overlooking a minor request can result in major design rework
- Do not underestimate the importance of this project phase
- Do not underestimate the time required to fully complete it

Gathering Requirements
Recommended Practices (Continued)

- Collate a requirements document to drive the entire project design, and ensure that the business signs off on this
  - Commonly, not all requirements can be met
  - Assist the business to prioritize, and if time or budget is limited, assign requirements to later releases

- Identify requirements that could be better delivered as SSBI solutions
  - Accept that a data warehouse can never meet every requirement
  - Accept that existing SSBI solutions may not need to be replaced

Microsoft

Module Outline
02 | Planning for a Data Warehouse Implementation

Topic
Data Warehouse Planning
Gathering Requirements
Data Warehouse Architectures
Demo: Exploring the AdventureWorks Databases

Data Warehouse Architectures

- There are several data warehouse design architectures adopted today:
  - Kimball's Dimensional Data Warehouse
  - Inmon's Corporate Information Factory
  - Stand-alone data mart

- This course does not attempt to suggest that one architecture is more superior, nor does it discuss the pros and cons of each
  - All approaches have been applied to deliver successful implementations, and suitability is based on specific requirements

Data Warehouse Architectures
Dimensional Data Warehouse

- The **Dimensional Data Warehouse** is also know as:
  - Enterprise data warehouse
  - Bus architecture
  - Architected data marts
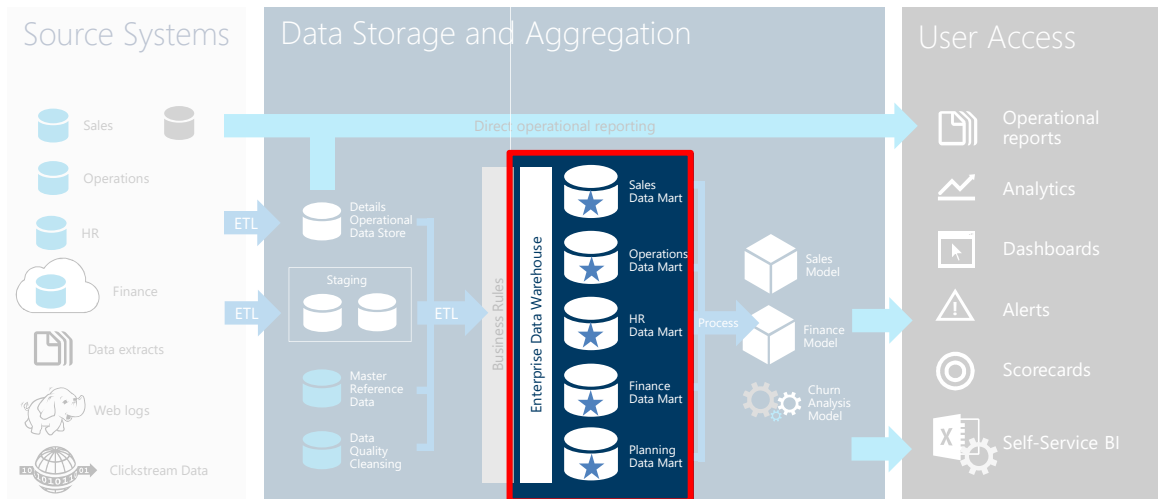 - Virtual data marts

Data Warehouse Architectures
Dimensional Data Warehouse (Continued)

- In essence:
  - The design is an integrated repository of atomic data
  - All data is organized dimensionally
  - It may be accessed by end users directly
  - Data marts are subject areas within the data warehouse, and are not required to be separate databases

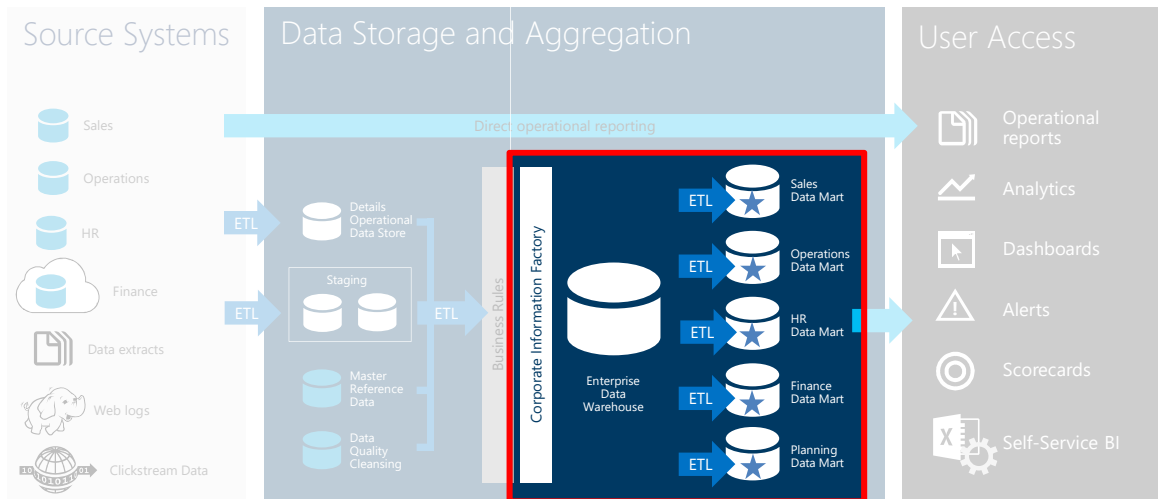# Data Warehouse Architectures
Diagram ► Dimensional Data Warehouse

Data Warehouse Architectures
Corporate Information Factory

- The **Corporate Information Factory** is also know as:
  – Atomic data warehouse
  – Enterprise data warehouse

- In essence:
  – The data warehouse is an integrated repository of atomic data, often normalized to 3NF
  – It is not accessed directly
  – Dimensional design is used to create data marts which reorganize data into subject areas

Data Warehouse Architectures
Diagram ► Corporate Information Factory

Data Warehouse Architectures
Stand-Alone Data Mart

- The **Stand-Alone Data Mart** is also know as:
  - Data mart
  - Silo
  - Stovepipe
  - Island

- It is a subject area implementation <u>without an enterprise context</u>
- It commonly adopts a dimensional design
- Perhaps useful for a POC, or first subject area exploration

Data Warehouse Architectures
Review of Enterprise Architectures

- The Dimensional Data Warehouse and
  Corporate Information Factory implementations
  designs are commonly adopted
  - They both have an enterprise focus, supporting analytic requirement across the business, and addressing requirements within subject areas

- They both implement dimensional format, but at different levels
  - Dimensional Data Warehouse at enterprise level
  - Corporate Information Factory at subject area level

- The theory in this course can be applied to either architecture

Microsoft

Module Outline
02 | Planning for a Data Warehouse Implementation

Topic
Data Warehouse Planning
Gathering Requirements
Data Warehouse Architectures
Demo: Exploring the AdventureWorks Databases

Demo

Exploring the AdventureWorks Databases

Demo objectives:

1. Introduce the AdventureWorks databases
2. Explore the **AdventureWorks2016** database
3. Explore the **AdventureWorks2016DW** database

Introducing AdventureWorks

- Many examples, and the demos and labs delivered with this course are based upon the AdventureWorks sample databases
- Company background:
  - The company represents a large, multinational manufacturing company which manufactures and sells bicycles to North American, European, and Pacific commercial markets
  - Products are sold through reseller and internet channels

Introducing AdventureWorks
AdventureWorks (Continued)

- The **AdventureWorks2016** database represents a collection of operational source systems
- The **AdventureWorksDW2016** database represents an enterprise data warehouse architecture, comprising subject areas for:
  - Call center
  - Sales (internet and reseller)
  - Finance, and
  - Product inventory

Introducing AdventureWorks
AdventureWorks (Continued)

- They are available as a free download from CodePlex
- These databases serve as helpful resources to help convey theory on design practices
  - Microsoft's online documentation uses the database to convey examples
- They are also convenient for supporting demos and labs ☺
- However, they do not necessarily reflect sound design practices

Microsoft