Module 1

Introduction to Data Warehousing
• Overview of Data Warehousing
• Considerations for a Data Warehouse Solution

Repositorio de todos los datos de la empresa
Lesson 1: Overview of Data Warehousing

• The Business Problem
• What Is a Data Warehouse?
• Data Warehouse Architectures
• Components of a Data Warehousing Solution
• Data Warehousing Projects
• Data Warehousing Project Roles
• SQL Server As a Data Warehousing Platform
The Business Problem

- Key business data is distributed across multiple systems
- Finding the information required for business decision-making is time-consuming and error-prone
- Fundamental business questions are hard to answer
What Is a Data Warehouse?

- A centralized store of business data for reporting and analysis
- Typically, a data warehouse:
  - Contains large volumes of historical data
  - Is optimized for querying data (as opposed to inserting or updating)
  - Is incrementally loaded with new business data at regular intervals
  - Provides the basis for enterprise BI solutions
Data Warehouse Architectures

- **Centralized** Data Warehouse
- Departmental **Data Mart**
- Hub and Spoke
Components of a Data Warehousing Solution

Data Warehouse

Master Data Management

Data Cleansing

ETL

Data Models

Reporting and Analysis

Data Sources

En el operacional buscamos un dato
En el datawarehouse contabilizamos datos
1. Start by identifying the business questions that the data warehousing solution must answer
2. Determine the data that is required to answer these questions
3. Identify data sources for the required data
4. Assess the value of each question to key business objectives versus the feasibility of answering it from the available data

- For large enterprise-level projects, an incremental approach can be effective:
  - Break the project down into multiple sub-projects
  - Each sub-project deals with a particular subject area in the data warehouse
Data Warehousing Project Roles

- Project manager
- Solution architect
- Data modeler
- Database administrator
- Infrastructure specialist
- ETL developer
- Business users/analyst
- Testers
- Data stewards
SQL Server As a Data Warehousing Platform

• Core Data Warehousing
  • SQL Server Database Engine
  • SQL Server Integration Services
  • SQL Server Master Data Services
  • SQL Server Data Quality Services

• Enterprise BI
  • SQL Server Analysis Services
  • SQL Server Reporting Services
  • Microsoft SharePoint Server
  • Microsoft Office

• Self-Service BI and Big Data Analysis
  • Excel Add-ins (PowerPivot, Power Query, Power View, Power Map)
  • Microsoft Office 365 Power BI
  • Windows Azure HDInsight
Lesson 2: Considerations for a Data Warehouse Solution

- Data Warehouse Database and Storage
- Data Sources
- Extract, Transform, and Load Processes
- Data Quality and Master Data Management
Data Warehouse Database and Storage

- Database Schema
- Hardware
- High Availability and Disaster Recovery
- Security
Data Sources

- Data Source Connection Types
- Credentials and Permissions
- Data Formats
- Data Acquisition Windows
Extract, Transform, and Load Processes

• Staging:
  • What data must be staged?
  • Staging data format

• Required transformations:
  • Transformations during extraction versus data flow transformations

• Incremental ETL:
  • Identifying data changes for extraction
  • Inserting or updating when loading
Data Quality and Master Data Management

• Data quality:
  • Cleansing data:
    • Validating data values
    • Ensuring data consistency
    • Identifying missing values
  • Deduplicating data

• Master data management:
  • Ensuring consistent business entity definitions across multiple systems
  • Applying business rules to ensure data validity
Lab: Exploring a Data Warehousing Solution

- Exercise 1: Exploring Data Sources
- Exercise 2: Exploring an ETL Process
- Exercise 3: Exploring a Data Warehouse

Logon Information
Virtual Machine: 20463C-MIA-SQL
User name: ADVENTUREWORKS\Student
Password: Pa$$w0rd

Estimated Time: 30 minutes
Lab Scenario

The labs in this course are based on Adventure Works Cycles, a fictional company that manufactures and sells bicycles and accessories to customers worldwide. Adventure Works sells direct through an e-commerce website, as well as an international reseller network. Throughout this course, you will develop a data warehousing solution for Adventure Works Cycles, including a data warehouse, an ETL process to extract data from source systems and populate the data warehouse, a data quality solution, and a master data management solution.
The lab for this module provides a high-level overview of the solution you will create later. Use this lab to become familiar with the various elements of the data warehousing solution that you will learn to build in subsequent modules. Don’t worry if you do not understand the specific details of how each component of the solution has been built, as you will explore these in greater depth later in the course.
Module Review and Takeaways

• Review Question(s)