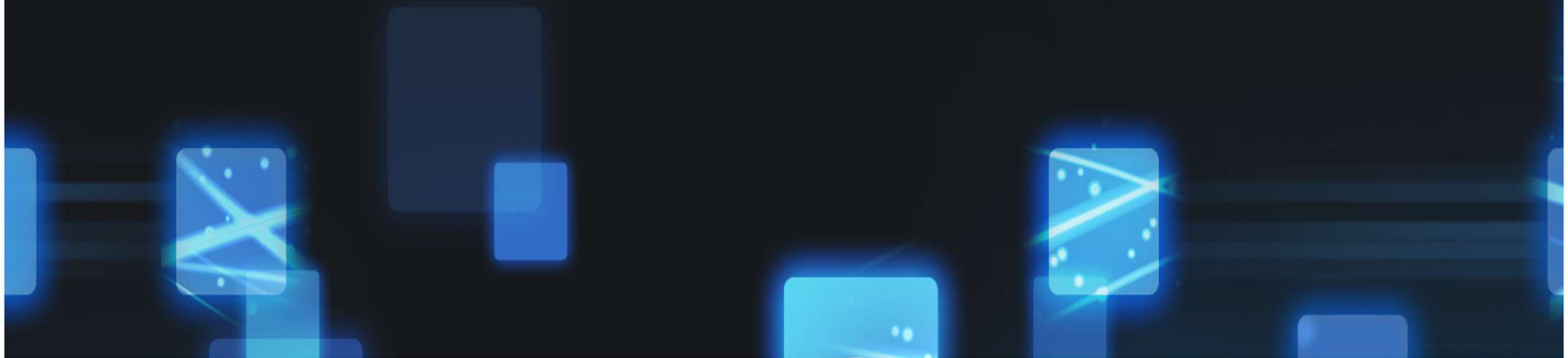


Introduction to Data Warehousing





Module Overview

- Overview of Data Warehousing
- Considerations for a Data Warehouse Solution



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

A successful business needs to be able to adapt— the following problems make that difficult:

1. Business data is spread across many systems
2. Data can be inconsistent, duplicated, and contradictory
3. Fundamental questions can't be easily answered



What Is a Data Warehouse?

A centralized store of business data for reporting and analysis that typically:

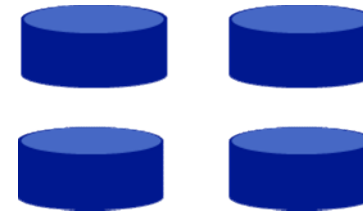
- Contains large volumes of historical data
- Is optimized for querying, as opposed to inserting or updating data
- Is incrementally loaded with new business data at regular intervals
- Provides the basis for enterprise BI solutions

Data Warehouse Architectures

Central Data Warehouse



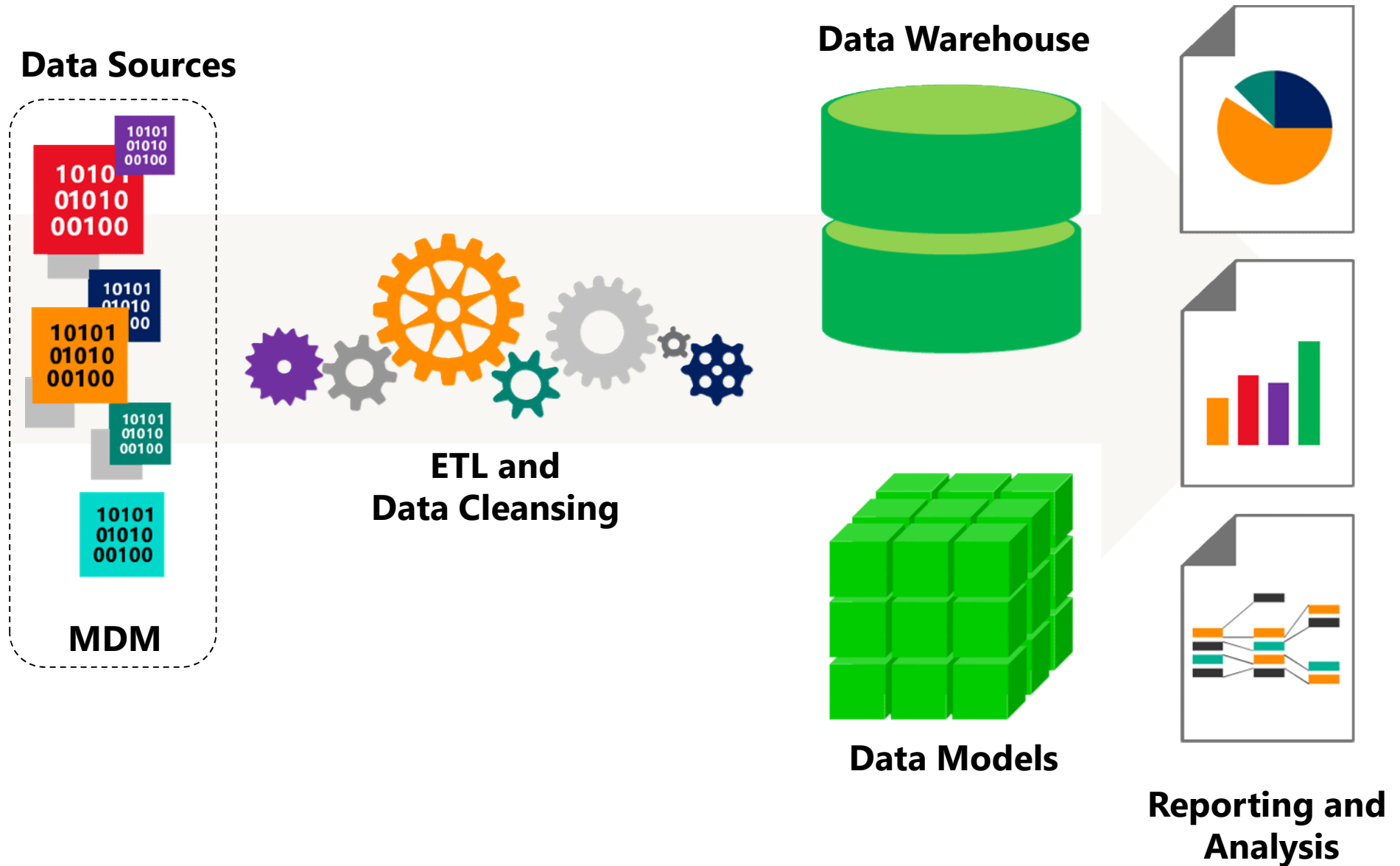
Departmental Data Marts



Hub-and-Spoke



Components of a Data Warehousing Solution





Data Warehousing Projects

1. What are the business questions that need to be answered?
2. What data is required to answer them?
3. Where is this data and how easy is it to obtain?
4. Knowing the above, assess the importance of each question against the ability to answer it from existing data

Data Warehousing Project Roles



Project Manager	Solution Architect	Data Modeler	Database Administrator	Infrastructure Specialist
ETL Developer	Business Users	Business Analyst	Testers	Data Stewards



SQL Server As a Data Warehousing Platform

Core Data Warehousing

- Database Engine
- Integration Services
- Master Data Services
- Data Quality Services

Business Intelligence

- Analysis Services
- Reporting Services



Lesson 2: Considerations for a Data Warehouse Solution

- Data Warehouse Database and Storage
- Columnstore Indexes
- 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

Columnstore Indexes

Row-based index

- Can be clustered and nonclustered
- Improve performance on row level queries, inserts and updates
- Best used in OLTP databases
- All data in a row is processed

Column-based index

- Can be clustered and nonclustered
- Improve performance on queries that scan a table, aggregation and analytical queries
- Best used in data warehouses
- Only columns needed are processed

A clustered columnstore index can be combined with multiple nonclustered row indexes to have the benefits of both types



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