Delivering a Relational Data Warehouse

Week 1 – Introducing the Data Warehouse

Module 01
The Business Case for a Data Warehouse

Microsoft

Module Outline
01 | The Business Case for a Data Warehouse

Topic
► Business Intelligence
► **Demo:** Delivering Business Intelligence
► Data Warehouse Goals

Module Outline
01 | The Business Case for a Data Warehouse

Topic
Business Intelligence
Demo: Delivering Business Intelligence
Data Warehouse Goals

Business Intelligence

- Today, Business Intelligence (BI) is a well-understood term
- According to Gartner, BI is defined as:

> A broad category of applications and technologies for gathering, storing, analyzing, sharing and providing access to data to help enterprise users make better business decisions

---

Business Intelligence
(Continued)

Or, put more simply…

> It is the application of knowledge derived from analyzing an organization's data to effect a more positive outcome

Or, put even more simply…

> It transforms data into knowledge to support decisions

Business Intelligence
(Continued)

- BI is used by decision makers to:
  - Understand the health of the organization
  - Collaborate on a shared view of:
    - Data
    - Business logic, and
    - Business drivers
  - Reduce the time to decision

Business Intelligence
(Continued)

- Its goal is often to:
  - Impact the bottom line by measuring specific operations
  - Enhance competitive advantage
- BI is no longer a luxury afforded by a larger companies—it is considered an essential part of the IT portfolio
- It encompasses a broad spectrum of technologies and requires skilled professionals to design, develop and manage

Business Intelligence
Addressing Common Business Questions

- What reseller sales have been made, and where?
- How are the salespeople performing?
- Which customers are likely to buy from us?
- Which products do our customers buy together?
- What is the sentiment of our new product?

Business Intelligence
Common Delivery Challenges

- Organizations have large volumes of related data stored in a variety of data systems, often in different formats, and possibly residing in different locations (on-premises, or cloud)
- Data systems may not:
  - Be easily queried
  - Be optimized for analytical queries
  - Contain all the data required, by design, or they store limited history
  - Manage historical context
  - Be available or accessible

Business Intelligence
Common Delivery Challenges (Continued)

- Employees may not have sufficient skills, tools, or permissions to query data systems
  - Commonly, employees do not have access to operational data systems, due to valid concerns that analytic queries can negatively impact on performance
- Systems may not have consistent definitions

Business Intelligence
Common Decision Maker Requirements

- Decision makers need:
  - Data catalogs to discover and understand data assets
  - Reliable, secure access to data
  - Flexibility in the ways they access data
  - Low latency query results
  - Tools and training to:
    - View, and interact with data assets and reports
    - Produce their own data models, reports and dashboards (Self-Service BI)

Business Intelligence
Delivery Scenarios

- Operational Reporting
  - Provides improved access to data from operational data systems

- Business Process or Activity Management
  - Provides improved analysis and reporting capabilities for specific business processes or activities

- Data Mart
  - Provides improved tools and access to business users of an application to enhance its value by improving decision making
  - Delivers integrated reporting and analytics

Business Intelligence
Scenarios (Continued)

- Enterprise Data Warehousing
  - Provides comprehensive integration of critical information across the enterprise
  - Breaks down the barriers between applications

Module Outline
01 | The Business Case for a Data Warehouse

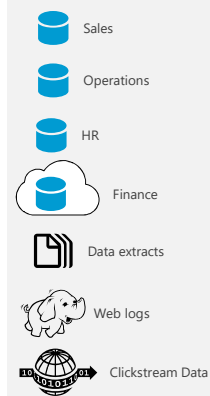| Topic |
| --- |
| Business Intelligence |
| **Demo:** Delivering Business Intelligence |
| Data Warehouse Goals |
| |
| |
| |

Demo

Delivering Business Intelligence

Demo objectives:

1. Describe common data challenges
2. Describe the purpose of the data warehouse
3. Describe the data warehouse ecosystem

Delivering Business Intelligence
01 – Organizations accumulate data in many data sources and formats
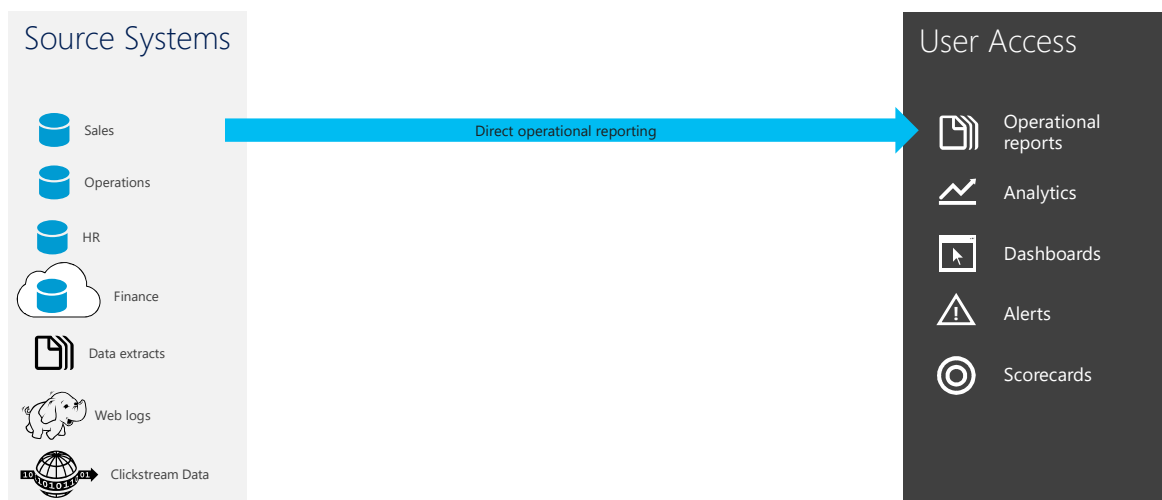
Source Systems

Sales

Operations

HR

Finance

Data extracts

Web logs

Clickstream Data

Delivering Business Intelligence
02 – Users need access to the data

Source Systems

- Sales
- Operations
- HR
- Finance
- Data extracts
- Web logs
- Clickstream Data

User Access

Delivering Business Intelligence
02 – Users may access data sources directly

Source Systems

- Sales
- Operations
- HR
- Finance
- Data extracts
- Web logs
- Clickstream Data

Direct operational reporting →

User Access

- Operational reports
- Analytics
- Dashboards
- Alerts
- Scorecards

Delivering Business Intelligence
03 – Data sources can be mirrored/replicated to reduce contention

Source Systems

- Sales
- Operations
- HR
- Finance
- Data extracts
- Web logs
- Clickstream Data

Direct operational reporting →

User Access

- Operational reports
- Analytics
- Dashboards
- Alerts
- Scorecards
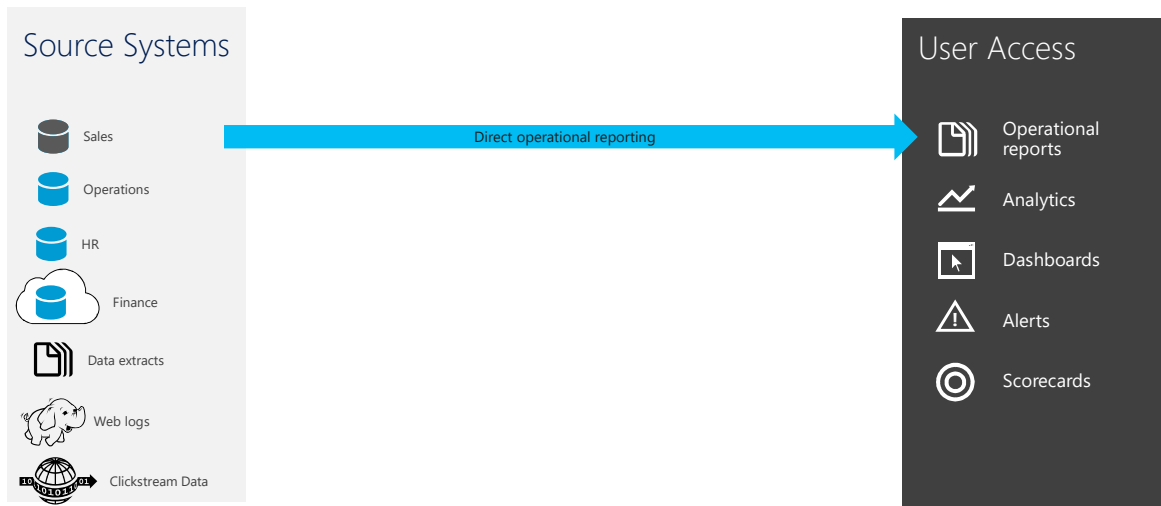
Delivering Business Intelligence
05 – Or, operational data sources can be prepared for analytics

Source Systems

- Sales
- Operations
- HR
- Finance
- Data extracts
- Web logs
- Clickstream Data

Data Storage and Aggregation

Direct operational reporting →

ETL → Details Operational Data Store

User Access

- Operational reports
- Analytics
- Dashboards
- Alerts
- Scorecards
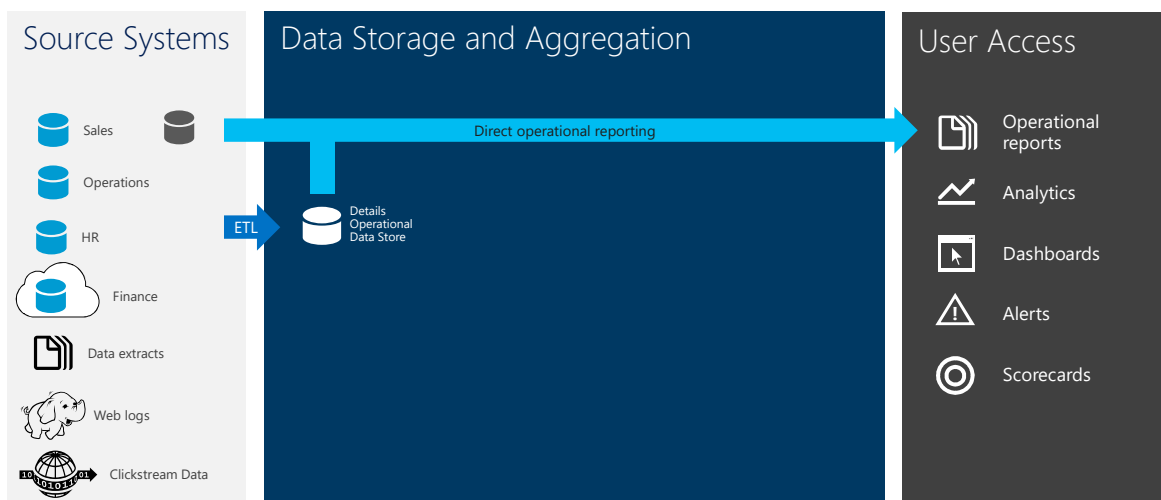
Delivering Business Intelligence
06 – The data warehouse manages data for reporting and analytics

Source Systems

- Sales
- Operations
- HR
- Finance
- Data extracts
- Web logs
- Clickstream Data

Data Storage and Aggregation

Direct operational reporting

ETL

Details
Operational
Data Store

Enterprise Data Warehouse

- Sales Data Mart
- Operations Data Mart
- HR Data Mart
- Finance Data Mart
- Planning Data Mart

User Access

- Operational reports
- Analytics
- Dashboards
- Alerts
- Scorecards
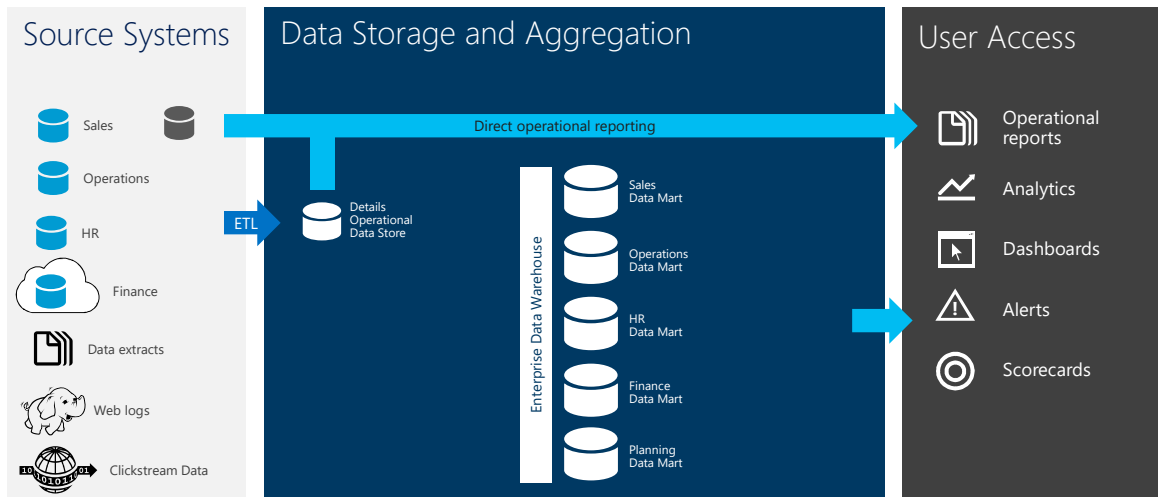
Delivering Business Intelligence
08 – Staging areas may simplify the loading of the data warehouse

Source Systems

- Sales
- Operations
- HR
- Finance
- Data extracts
- Web logs
- Clickstream Data

Data Storage and Aggregation

Direct operational reporting

ETL

Details
Operational
Data Store

ETL

Staging

Enterprise Data Warehouse

- Sales Data Mart
- Operations Data Mart
- HR Data Mart
- Finance Data Mart
- Planning Data Mart

User Access

- Operational reports
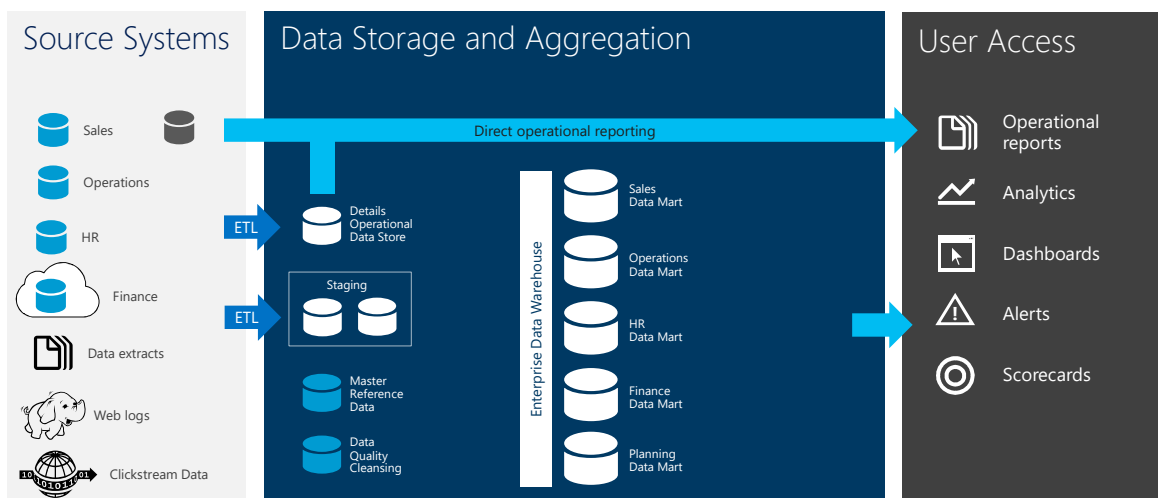- Analytics
- Dashboards
- Alerts
- Scorecards

Delivering Business Intelligence
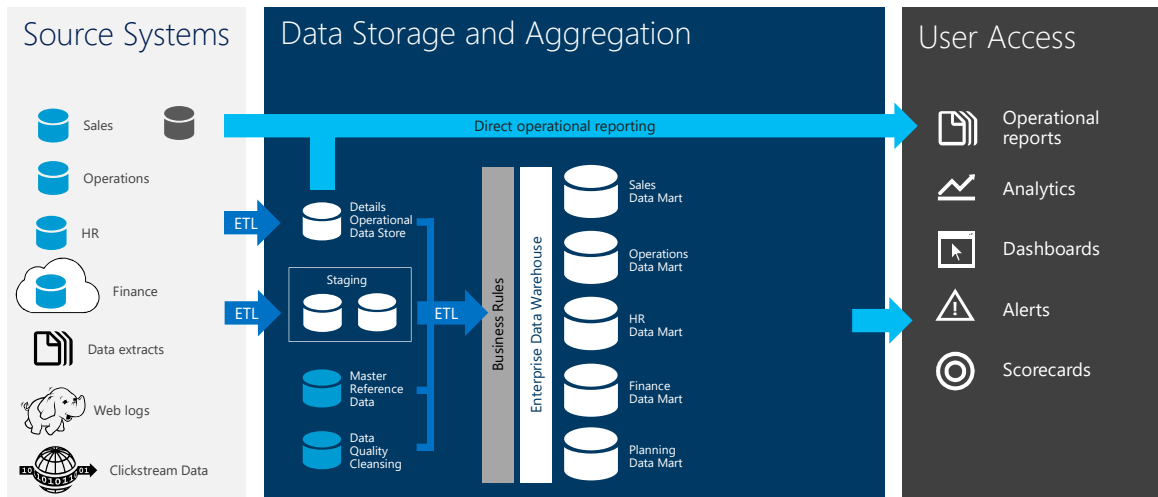09 – Master data systems can manage consistent definitions
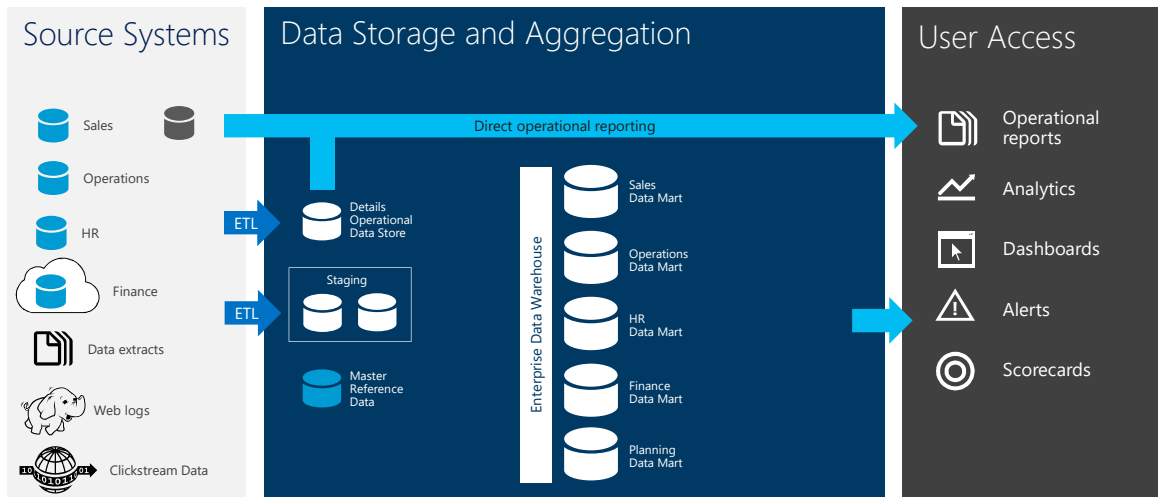
Delivering Business Intelligence
10 – Data quality systems can help detect and cleanse dirty data
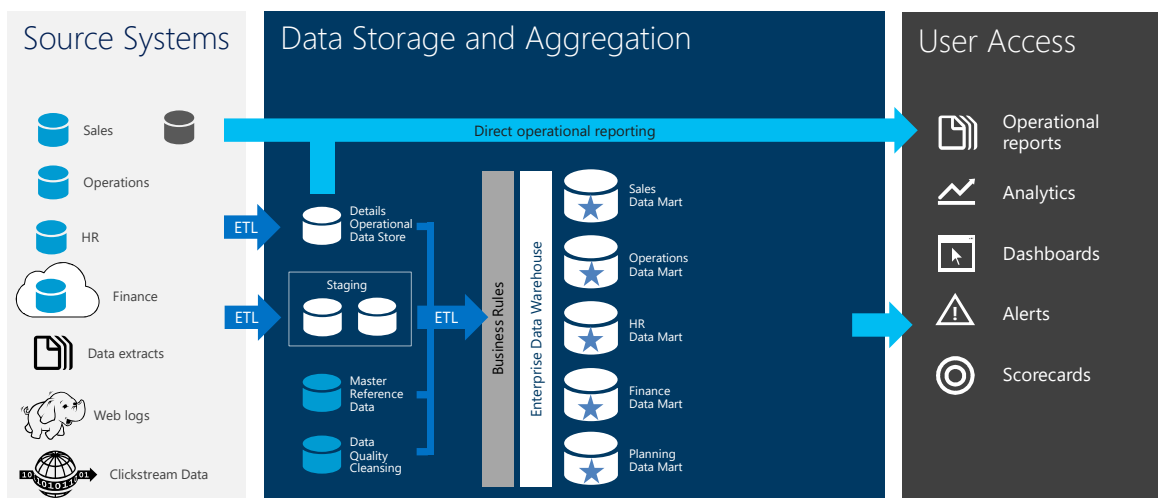
Delivering Business Intelligence
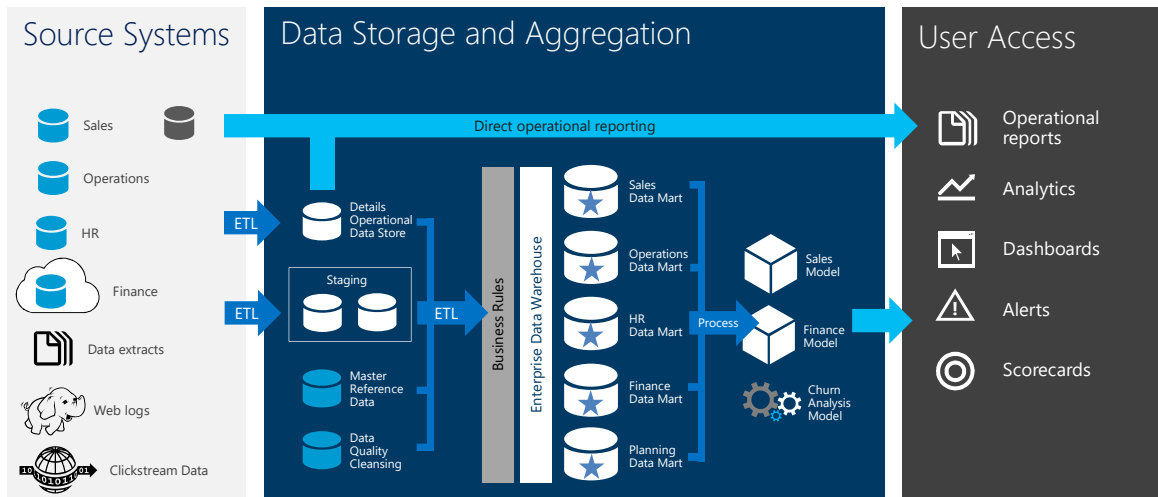## 11 – Only complete, credible, clean, consistent data is loaded

Delivering Business Intelligence
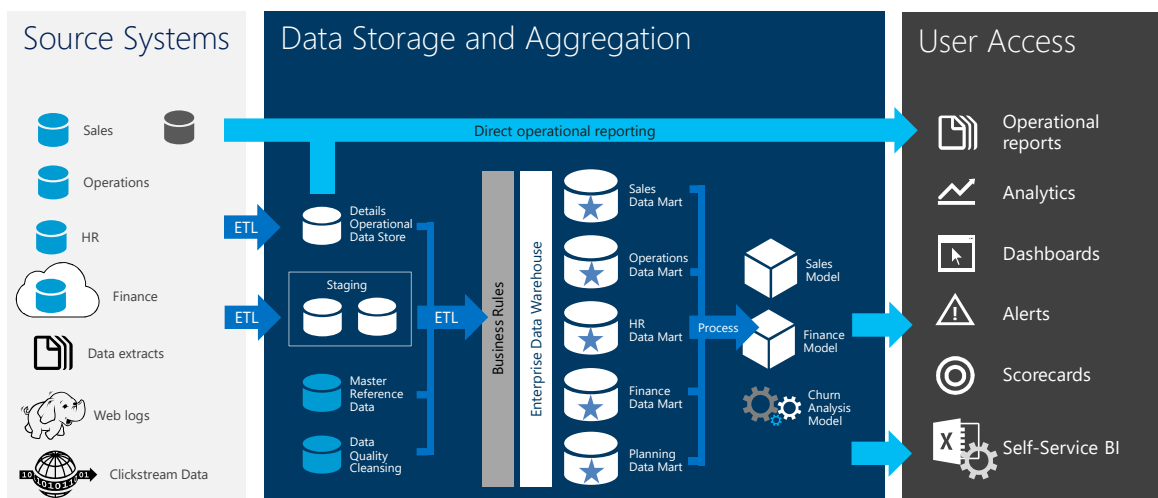## 12 – Data marts are designed with schemas optimized for analytics

Delivering Business Intelligence
13 – The data warehouse comprises additional data assets

Delivering Business Intelligence
14 – The data warehouse can enable governed Self-Service BI

Module Outline
01 | The Business Case for a Data Warehouse

Topic
Business Intelligence
Demo: Delivering Business Intelligence
Data Warehouse Goals

Data Warehouse Goals

- In essence, the data warehouse delivers data for the purposes of analytics and reporting
- Specifically the data warehouse:
  - Stores quality, conformed, and historically accurate data as a "single version of the truth"
  - Delivers understandable and navigable data with fast performance
  - Is designed for resilience to adapt to changes as requirements evolve
  - Ensures access of the right data to only the right people
  - Forms the foundation for decision making

Data Warehouse Goals
Data Warehouse Ecosystem

- The data warehouse ecosystem consists of components and services to deliver the data warehouse goals
  - Extract, Transform and Load systems
  - Operational Data Stores (ODS)
  - Staging systems
  - Master Data Management systems
  - Data Quality Management systems
  - Data models (OLAP)
  - Prediction models (data mining, or machine learning)
  - Data dictionaries

Data Warehouse Goals
Data Warehouse Ecosystem ► ETL Systems

- Extract, Transform and Load (ETL) systems are concerned with moving and transforming data
- Transformations can include data integration, aggregation, change detection, calculations, conforming or cleansing
- ETL development is often complex and development can be challenging
  – It is estimated that 60-80% of the data warehouse development effort is dedicated to the ETL process

Data Warehouse Goals
Data Warehouse Ecosystem ► Operational Data Stores

- An ODS delivers a subject-oriented integrated store
- It serves two possible purposes:
  – Integration point for operational systems, providing real-time source for critical details (balances, etc.)
  – To supply current and detailed data for decision support
- The ODS is updated frequently
- Contains all text and numbers required to describe low-level transactions

Data Warehouse Goals
Data Warehouse Ecosystem ► Staging Systems

- Staging systems support data loading and transformation requirements
- They can:
  - Minimize the impact on operational systems
  - Be used to assist with, and optimize, ETL processing
- They are never intended to be accessed by end users
- They can also provide ETL restartability, without the need to reload data from source systems

Data Warehouse Goals
Data Warehouse Ecosystem ► Master Data Management Systems

- Master Data Management (MDM) systems delivers a master data hub that provides access to authoritative, standardized, and validated versions of data
- They are concerned with the concept of the "Golden Record"
  - This concept represents the "single version of the truth", also encompassing all the data in every System of Record (SOR) within the organization
- They are commonly used to define and maintain key business dimension data, for example, a master list of products

Data Warehouse Goals
Data Warehouse Ecosystem ► Data Quality Management Systems

- Data Quality Management systems profile data to discover inconsistencies and other anomalies
- They can also perform data cleansing activities to improve the data quality
  - For example, repairing incorrect values, detecting duplicate records, removing outliers, etc.

Data Warehouse Goals
Data Warehouse Ecosystem ► Data Models

- Data models deliver intuitive browsing and high performance query results—even over large volumes of data
- They can:
  - Perform calculations difficult to achieve by using relational queries
  - Surface actions, like drill through
  - Be effectively secured for different roles
- Data models are more commonly known as cubes

Data Warehouse Goals
Data Warehouse Ecosystem ► Prediction Models

- Prediction models are trained from data warehouse data, to detect patterns, clusters, relationships and rules
- They deliver a model which can be used to explore data, or to perform predictions
- Common business scenarios include:
  - Churn analysis, customer segmentation, association rules, forecasting, anomaly detection, and fraud detection

Data Warehouse Goals
Data Warehouse Ecosystem ► Data Dictionaries

- Data warehouse users and developers can benefit from an authoritative data dictionary which document the data warehouse
- A dictionary typically includes:
  - Names and descriptions of tables and columns
  - Relationships between tables
  - Calculation logic descriptions
  - Owners, including contact details
- They are particularly useful to support Self-Service BI (SSBI)

Data Warehouse Goals
Current Trends

- Based on recent technological advancements, today we are witnessing a paradigm shift in data warehousing
  - Integration of big data analytics
  - Real-time insights
  - Complex data types
  - In-memory technologies
  - Seamless integration of data among cloud, on-premises, and hybrid environments

Microsoft