Chapter 8

Data Integration and Interoperability

Preparation and Operational Activities in Data Interchange

➢ Preparation and operational activities associated with data interchange include:

- **Specification for dataset creation, design, build, and test of dataset creation, specification and design of operational process,** and **operation.**
- **All of the above** are necessary for the effective implementation of data interchange processes.

Reference: Comprehensive data interchange processes require proper specification, design, testing, and operational monitoring.

Critical Data Interchanges

➤ Critical data interchanges are involved in:

- **RRP submissions to a regulator, issue of contracts and ITTs for bidding, business-specific information published to the organization's web pages,** and **GIS datasets published to support safety checks before operational digging.**

- **All of the above** are examples of critical data exchanges necessary for regulatory compliance and business operations.

Reference: Critical interchanges are vital for maintaining operational integrity and regulatory compliance.

Discovering and Documenting Metadata About Physical Data Assets

➢ Discovering and documenting metadata about physical data assets provides information on how data is transformed as it moves between systems.

• This metadata is essential for tracking data lineage and ensuring accurate data movement.

Reference: Metadata discovery helps in understanding data transformations and movements across different systems.

Data Lineage Tool

- ➢ A data lineage tool enables a user to **track the data from source system to a target database, understanding its transformations.**
- This tool helps visualize how data moves through systems and undergoes changes..

Reference: Data lineage tools are key for maintaining transparency and ensuring data integrity across transformations.

Difficulty in Integrating Multiple Source Systems

➢ One of the difficulties in integrating multiple source systems is **determining valid links or equivalences between data elements.**

• It can be challenging to establish consistent relationships and mappings between disparate data sources.

Reference: Data element equivalency is critical for successful data integration and maintaining data quality.

Ensuring Secure Data Interchange

➢ To ensure that all data interchange scenarios are managed securely, it is essential to **comply with legal and regulatory or external obligations to publish data.**

• Security and compliance are critical for protecting data during interchange processes.

Reference: Compliance with regulations ensures data privacy and integrity during data exchanges.

Goal of Data Discovery in Data Integration

➢ The goal of data discovery in data integration is to **identify potential sources and perform a high-level assessment of Data Quality.**

• This ensures that integration efforts are based on reliable and high-quality data sources.

Reference: Data discovery is the first step in ensuring that integration processes are grounded in accurate data.

Definition of Change Data Capture

➢ **Change data capture** is a **Data Integration approach that updates a Data Warehouse with small changes from Operational Systems.**

- It captures and processes incremental changes to keep the warehouse up to date..

Reference: Change data capture is essential for maintaining real-time or near-real-time synchronization between systems.

Not Part of Data Integration and Interoperability

➢ **Archiving data** is **not** part of Data Integration and Interoperability.

- Data integration focuses on combining and making data interoperable, while archiving is about storing data for long-term retention.

Reference: Data archiving is a separate process focused on data preservation rather than integration.

Advantage of Point-to-Point Interaction Model

➢ The advantage of using the **point-to-point interaction model** instead of a hub-and-spoke model is **lower latency**.

- Point-to-point interactions provide direct communication between systems, reducing delays.

Reference: Point-to-point systems are efficient for real-time data exchanges where low latency is critical.

ETL Acronym

➤ **ETL** stands for **Extract, Transform, Load.**
• It is the standard process for moving data from source systems, transforming it, and loading it into a target system.

Reference: ETL is a core data integration process, particularly in data warehousing.

Impact of Synchronous Data Integration

➢ When integrating two data stores using batch or real-time synchronous approaches, the result is often an issue with **latency**.

• Synchronous processes can introduce delays as they wait for all systems to align before completing operations.

Reference: Latency in synchronous systems can affect performance, especially in real-time integrations.

Example of Incoming Formatted Dataset

➢ An example of an incoming formatted dataset is **received content formatted into Excel tables as a formatted file.**

• Data sent to organizations is often formatted in specific ways, such as Excel tables, for ease of use and integration.

Reference: Formatting ensures that data is structured and ready for analysis or integration upon receipt..

Tightly Coupled System Interface

➤ When systems are tightly coupled, the interface is typically a **synchronous interface**.

- Tightly coupled systems require real-time communication and data synchronization between systems.

Reference: Synchronous interfaces facilitate real-time data sharing but can introduce dependencies between systems.

Integration Approach with Higher Latency

➤ **Batch data integration** has a higher latency compared to event-driven approaches.

• Batch processes involve collecting and processing data in groups, leading to delays in data availability.

Reference: Batch integration is suitable for scenarios where immediate data availability is not required.

Example of Internal Data Interchange

➢ An example of **internal data interchange is data sent and received between people or systems within an organization, such as reports, documents, GIS datasets, drawings, models, photographs, or records.**

• This type of interchange focuses on sharing data internally rather than externally.

Reference: Internal data interchanges support collaboration and operational efficiency within organizations.

Service-Oriented Architecture (SOA) Implementation

➢ Implementing a **Services-Oriented Architecture (SOA)** will often use **an enterprise service bus.**

- An enterprise service bus facilitates communication between services in an SOA environment.

Reference: SOA helps integrate different services within an organization, enabling modularity and flexibility.

Not an Example of External Outgoing Data Interchange

➢ **Purchased prebuilt data** is **not** an example of external outgoing data interchange.

• External outgoing data interchange involves sending data from an organization to external entities, not purchasing data.

Reference: Outgoing interchanges typically involve sending reports, documents, or other data generated within the organization.

Benefit of Service-Oriented Architecture (SOA)

➢ One of the key benefits of SOA is that it **enables application independence and the ability to replace systems without significant changes to interfacing systems.**

- SOA allows for modularity and flexibility in system integration.

Reference: SOA is designed to reduce dependencies between applications, making systems easier to maintain and upgrade.

Purpose of ISO 8000

➤ The purpose of **ISO 8000** is to **enable the exchange of complex information in an application-neutral form.**

• This standard ensures that data can be shared and understood across different systems and platforms.

Reference: ISO 8000 focuses on improving data quality and ensuring compatibility across systems..

Example of Interoperability

➢ The ability of a photo app to share images with various social media applications is an example of **Interoperability**.

• Interoperability ensures that different systems can work together and exchange data seamlessly.

Reference: Interoperability is critical for enabling systems to interact and share data across platforms.

Essential for Successful Data Integration

➢ The most essential factor for the successful integration of data is **understanding data content and structure**.

• Without a clear understanding of the data, integration efforts are likely to fail.

Reference: Data content and structure provide the foundation for designing integration processes that work effectively.

Mapping Data from Source to Target

➢ Mapping requirements and rules for moving data from source to target enables **Transformation**.

• Transformation ensures that data is properly structured and formatted as it moves between systems.

Reference: Data transformation is a key step in ensuring that data is usable and consistent in the target system.

Goal of Data Discovery

➢ One of the goals of data discovery is **identifying potential sources of data for the Data Integration effort.**

• Data discovery helps organizations find relevant data sources that can be integrated into their systems.

Reference: Identifying data sources is a critical step in planning data integration efforts.

Common Interaction Models for Data Integration

➢ Three common interaction models for data integration are **point to point, hub and spoke, and publish and subscribe.**

• Each model has its advantages depending on the scale and requirements of the integration effort.

Reference: These models provide different ways of structuring data exchanges between systems.

Best Use Case for Point-to-Point Interfaces

➢ The best use case for point-to-point interfaces is **integrating two systems with data only needed by those systems.**

• Point-to-point integration is ideal for simple, direct data exchanges between a limited number of systems.

Reference: Point-to-point interfaces are efficient for small-scale integrations where complex architecture is unnecessary.

Thank You