Applied A.I. Solutions

Foundations of Data Management
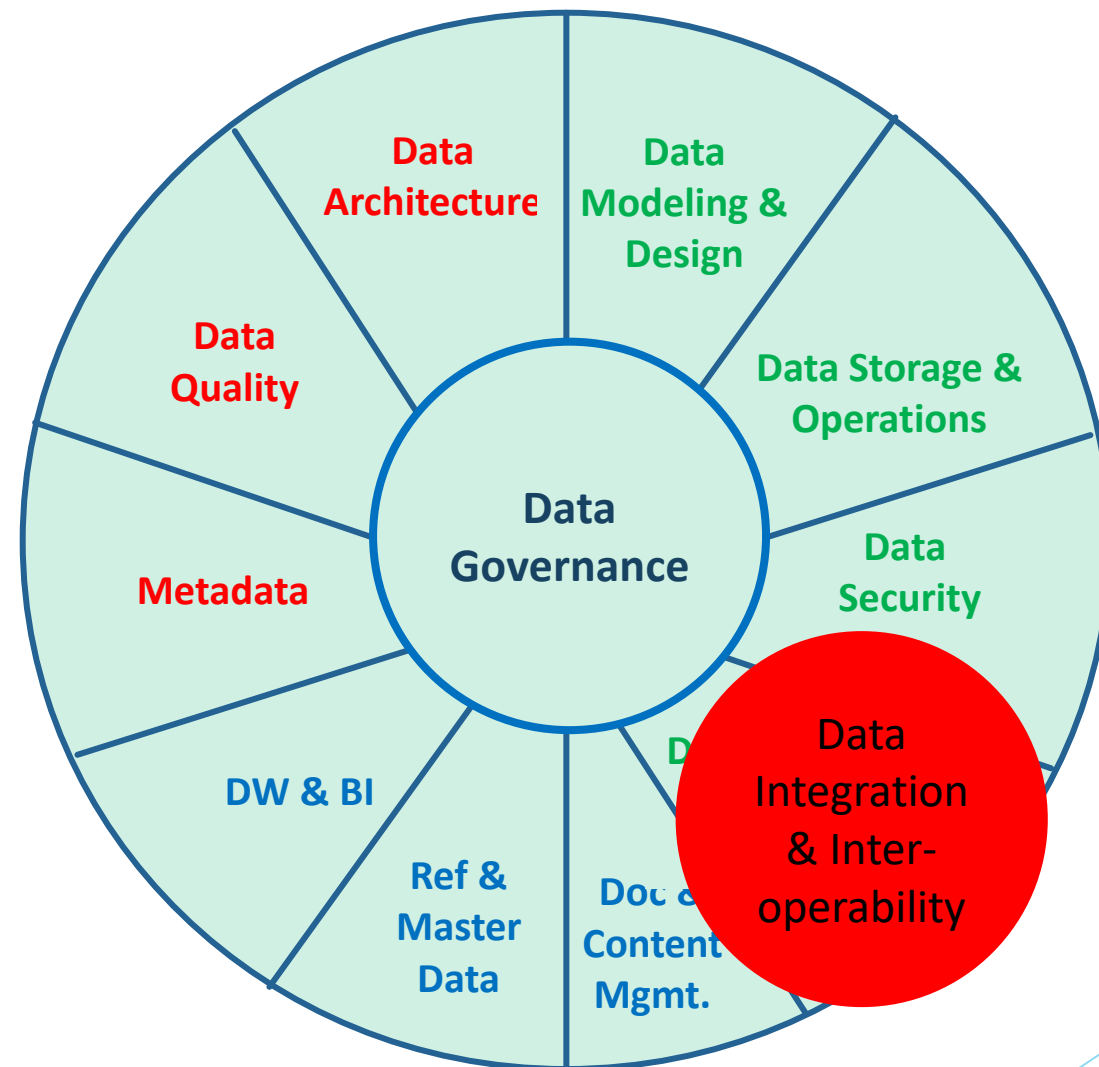
Professor
Daniel Vitaver-Bronstein, B.Sc., EMBA

daniel.vitaver-bronstein@georgebrown.ca

DATA INTEGRATION & INTEROPERABILITY

The DAMA Wheel

1. INTRODUCTION

- Integration <mark>consolidates data into consistent</mark> forms, physical or virtual

- Data Interoperability is the ability <mark>for multiple systems to communicate</mark>

- <mark>DII solutions</mark> enable DM functions such as:

  - data migration and conversion

  - data consolidation into hubs or marts

  - integration of vendor packages

DII Management Framework

Definition

Managing the movement and consolidation of data within and between data stores, applications and organizations

Goals

1. Provide data <mark>securely, regulatory</mark> complaint, in the format, <mark>timeframe</mark> needed

2. Lower cost, complexity, developing <mark>shared</mark> models and interfaces

3. Identify events, automatically <mark>trigger alerts and actions</mark>

4. Support BI, Analytics, MDM, and other efficiency efforts

Business   Drivers

Principles:

1. Take an <mark>enterprise perspective</mark> in design to ensure future scalability and extensibility

2. Balance local data needs with enterprise data needs

3. Ensure business accountability

Inputs

- Business Goals and strategies

- Data needs and standards

- Regulatory compliance and security requirements

- Data, process, application, and technical architectures

- Data semantics

- Source Data

Activities

1. Plan & Analyze
2. Design DII Solutions
3. Develop DII solutions
4. Implement and Monitor

Deliverables

1. DII Architecture

2. Data exchange specifications

3. Data access agreements

4. Data services

5. Complex event processing thresholds and alerts

Suppliers

- Data Producers
- IT Steering Committee
- Executives and Managers
- SME

Participants

- Data Architects
- Business and Data Analysts
- Data Modelers
- Data Stewards
- ETL, service interface developers
- Project and Program managers

Consumers

- Information consumers
- knowledge workers
- Managers and Executives

Technical Drivers

Techniques
- Hub and Spoke integration
- ETL
- Enterprise Application integration (EAI)
- Service Oriented Architecture (SOA)

Tools
- Data Transformation engine
- Data virtualization server
- Enterprise service bus
- Data, Process modeling tool
- Metadata repository

Metrics
- Data volumes, speed delivery, data latency
- time to market for enhancement
- solution costs, complexity
- Value delivered

Drivers

- Managing data movement <mark>efficiently</mark>

- Managing associated <mark>costs</mark> of moving data using multiple technologies

Essential Concepts

1. <mark>Extract, Transform, and Load</mark>

2. <mark>Latency</mark>

3. Replication

4. Archiving

5. Enterprise Message Format / Canonical Model

6. Interaction Models

7. DII Architecture Concepts

8. Data Exchange Standards

1. ETL vs ELT
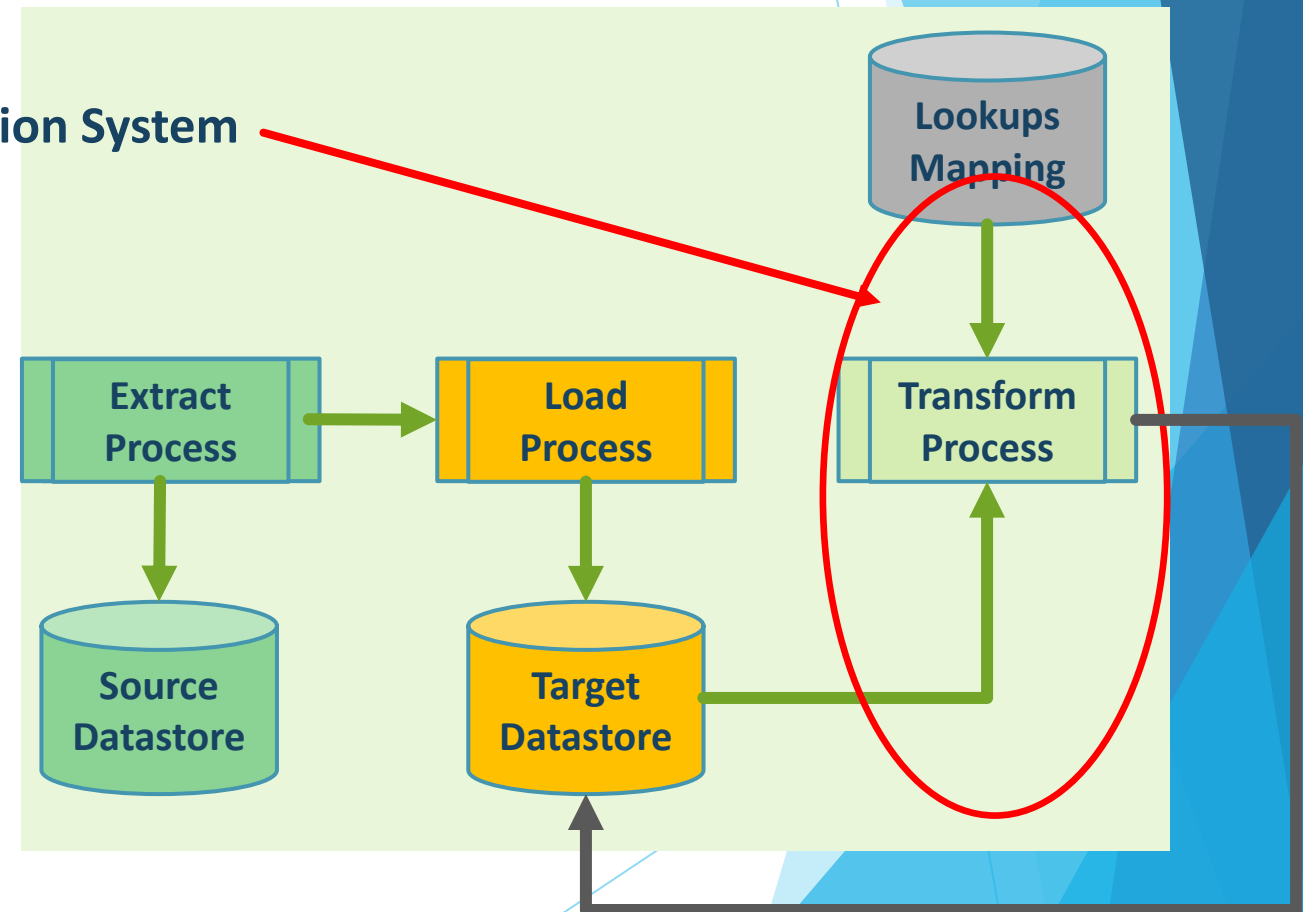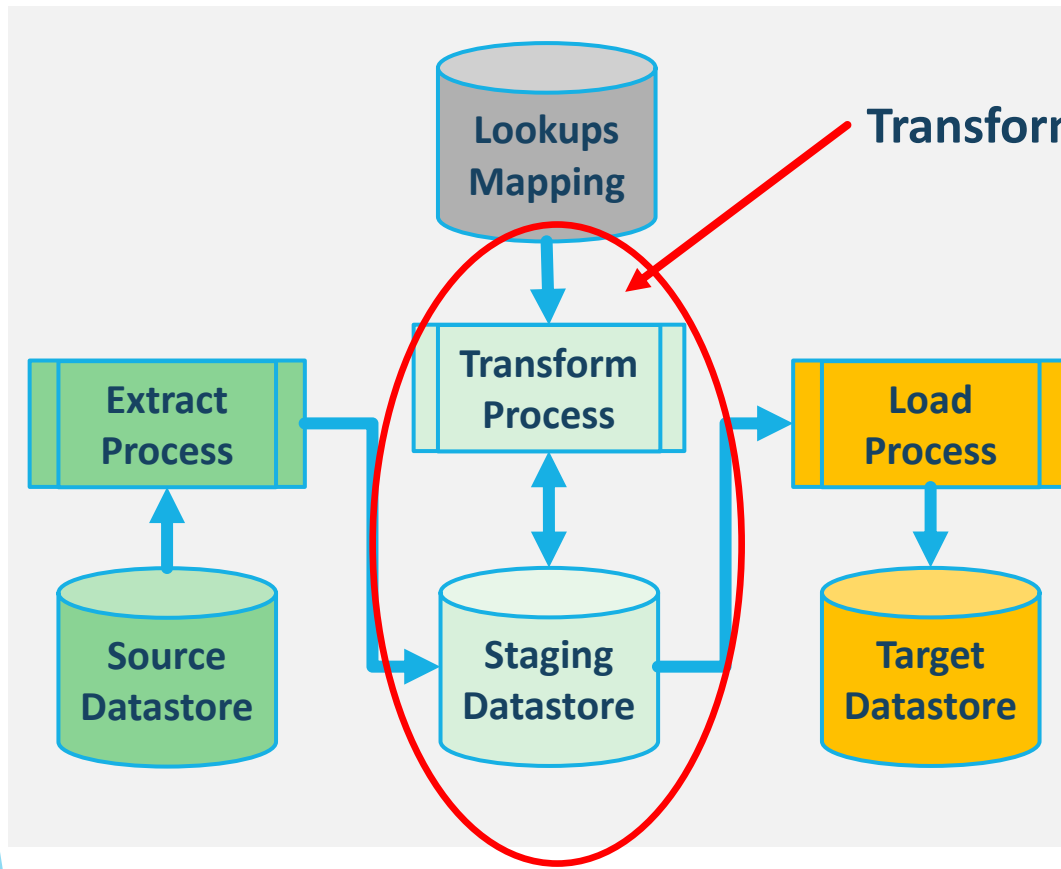
High

2. Latency

   o Batch

   o Change Data Capture

   o Near-real-time and Event-driven

   o Asynchronous

   o Real-time, synchronous

   o Low latency or Streaming

Low

3. Replication

   • Applications ==maintain exact copies== of datasets in multiple

   locations

4. Archiving

   • Storage or ==archives of data that is used infrequently== or not

   actively used

5. Enterprise Message Format / Canonical Mode

- Common model used by an organization or data exchange group to ==standardize== the format in which data is shared

6. Interaction Models

- **Point-to-Point**: systems pass data directly to each other

- **Hub-and-Spoke**: consolidates shared data (physically or virtually) in a central **hub** that many applications can use

- **Publish** – Subscribe model involves systems pushing data out (publish), and other systems pulling data in (subscribe)

7. DII Architecture Concepts

    a) Application Coupling

    b) Orchestrating and Process Controls

    c) Enterprise Application Integration

    d) Enterprise Services Bus (ESB)

    e) Service-Oriented Architecture (SOA)

7. DII Architecture Concepts – cont'd

   f) <mark>Complex Event Processing (CEP)</mark>

   g) Data Federation and Virtualization

   h) Data-as-a-Service (DaaS)

   i) Cloud-based Integration

8. Data Exchange Standards (ISO, XML)

2. ACTIVITIES

a) Plan and Analyze

- o Define Data Integration and Lifecycle <mark>Requirements</mark>

- o Perform Data Discovery

- o Document Data Lineage

- o <mark>Profile Data</mark>

- o Collect Business Rules

b) Design Data Integration Solutions

- o techniques and technologies

- o inventory of the data structures involved

- o orchestration, frequency of data flow

- o regulatory and security concerns

- o operating concerns

b) Design Data Integration Solutions – cont'd

- o Design Data Integration Architecture

  - – select interaction model

  - – design data services or exchange pattern

- o Model Data Hubs, Interfaces, Messages, Data Services

- o Map data Sources to Targets

- o Design Data Orchestration

c) **Develop Data Integration Solutions**

o Develop Data Services

o Develop Data Flows

o Develop Data Migration Approach

o Develop a Publication Approach

o Develop Complex Event Processing Flows

o Maintain DII Metadata

d) **Implement and Monitor**

3. TOOLS

- Data Transformation Engine / ETL Too[l]
- Data Virtualization Server
- Enterprise Service Bus
- Business Rules Engine
- Data and Process Modeling Tools
- Data Profiling Tools
- Metadata Repository

Data Integration Tools Reviews 2021 | Gartner Peer Insights

https://www.gartner.com/doc/reprints?id=1-1ZHS1JTN&ct=200717&st=sb

Magic Quadrant

Figure 1. Magic Quadrant for Data Integration Tools

Source: Gartner (August 2020)

4. TECHNIQUES

- The main goals are to keep the applications
    - o keep applications coupled loosely
    - o limit the number of interfaces
    - o use hub-and-spoke approach
    - o create standard (or canonical) interfaces

5. IMPLEMENTATION GUIDELINES

- o Readiness Assessment / Risk Assessment
- o Organizational and Cultural Change

6. GOVERNANCE

- <mark>Data Sharing Agreements</mark>
- <mark>DII and Data Lineage</mark>
- <mark>Data Integration Metrics</mark>

Data Availability
- availability of data requested

Data Volume and Speed
- volumes of data transported
- volumes of data transformed
- volumes of data analyzed
- speed of transmission
- latency between data update and availability
- latency between event and triggered action
- time to availability of new data sources

Solution Costs and Complexity
- cost of developing solutions
- cost of managing solutions
- ease of acquiring new data
- complexity of solutions and operations
- number of systems using data integration solutions