Applied A.I. Solutions

Foundations of Data Management

Professor
Daniel Vitaver-Bronstein, B.Sc., EMBA

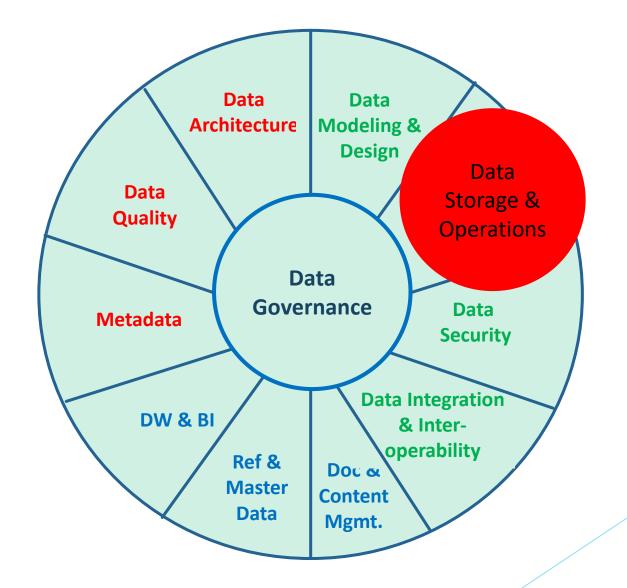daniel.vitaver-bronstein@georgebrown.ca

DATA STORAGE & OPERATIONS

The DAMA Wheel

1. INTRODUCTION

- DS&O includes the design, implementation and support of stored data, to maximize its value through its lifecycle, from creation/acquisition to disposal

  o <mark>Database</mark> support

  o Data <mark>technology</mark> support

Data Storage and Operations Management Framework

Definition

The design, implementation, and support of stored data to maximize its value

Goals

1. Manage availability of data throughout the data lifecycle

2. Ensure the integrity of data assets

3. Manage performance of data transactions

Business   Drivers

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

DS&O Programs are guided by the following principles:

1. Identify and act on automation opportunities

2. Build with reuse in mind

3. Understand and appropriately apply best practices

4. Utilize database standards to support requirements

5. Set expectations for the DBA role in project work

Inputs

- Data Architecture

- Data Requirements

- Data Models

- SLA

Activities

1. Manage Database Technology

2. Manage Database Operations

Deliverables

1. Database Technology Evaluation Criteria

2. Database Environments

3. Migrated/Replicated/Versioned Data

4. Business Continuity Plan

5. Database Performance

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Participants
- Data Architect
- Data Modeler
- Software Developer
- Application Testing Team

Suppliers
- DBA
- Data Architect

Consumers
- Data Modeler
- Software Developer
- Application Testing Team
- Infrastructure Operations

Technical ⬆ **Drivers**

Techniques
- Change Implementation Path
- Physical Naming Standards
- Data Lifecycle Management
- Script Usage for all Changes

Tools
- Data Modeling Tools
- Database Monitoring Tools
- Database Management Tools
- Developer Support Tools

Metrics
- Data Storage Metrics
- Performance Metrics
- Operations Metrics
- Services Metrics

Drivers

- DS & Op activities are instrumental to organizations that rely on data

- <mark>Business Continuity is the main driver</mark>

- Reliable Data Storage Infrastructure for IT Operations minimize the risk of disruption

Essential Concepts

- Database Terminology

- Database Lifecycle

  Management

- <mark>DBAs</mark>

- <mark>Database Architecture Types</mark>

- <mark>Database Processing Types</mark>

- Data Storage Media

- Database Environments

- Database Organization

- Specialized Databases

- Common Database

  Processes

- Database Terms

  - o Database

  - o Instance

  - o Schema

  - o Node

  - o Database abstraction

- Database Lifecycle Management

  o DBA is the ==custodian== of all database changes and controlled

    environments

  o DLM Includes implementing policies and procedures for

    - ==data acquisition==
    - ==data migration==
    - ==data retention==
    - ==data expiration==
    - ==data disposition==
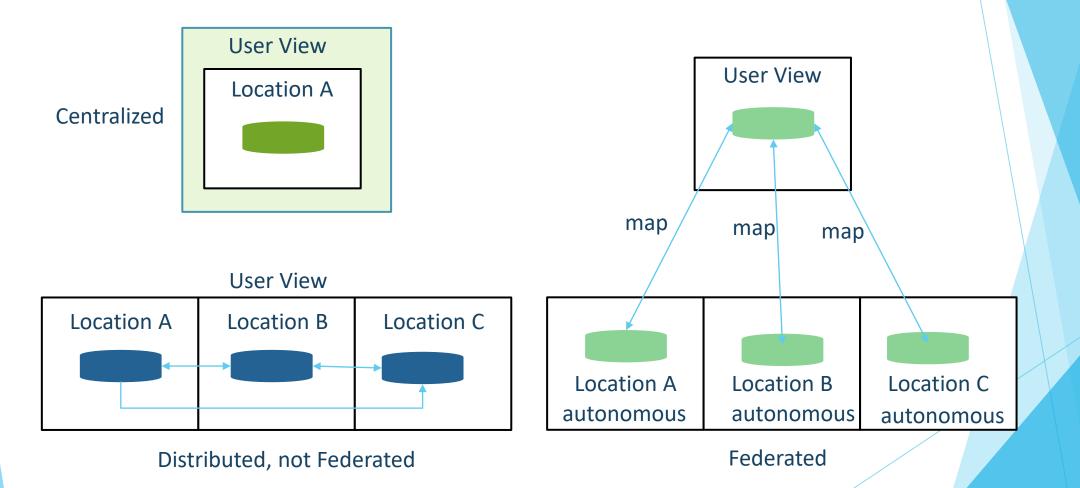   - ==data change request management==

- DBAs

  o Critical role in both Data Storage & Operations and Data Security

  o Supports development, test, QA, database environments

  o DABs, is a Data Team member. Roles:

    - Production DBA

    - Application DBA

    - Procedural and Development DBAs

    - NSA - Network Storage Administrators

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA
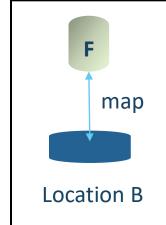
Database Architecture Types

Centralized

Centralized box:
- User View
  - Location A

User View

Distributed, not Federated

- Location A
- Location B
- Location C

User View

map   map   map

- Location A autonomous
- Location B autonomous
- Location C autonomous

Federated
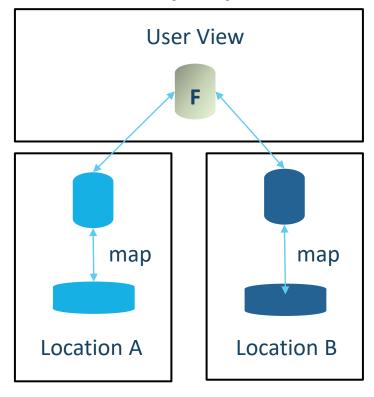
Tightly Coupled

Federated

Loosely Coupled

User View

User View

F

map

map

Location A

Location B

F

map

map

Location A

Location B

- Virtualization / Cloud Platforms

  o Provides computation, software, data access, and storage services <mark>independently</mark> from the physical location and system configuration

  o Methods:
    - Virtual machine image
    - Database-as-a-service (DaaS)
    - Managed database hosting on the cloud

o Database Processing Types

1. ==ACID==
   o Atomicity
   o Consistency
   o Isolation
   o Durability

2. ==BASE==
   o Basically Available
   o Soft State
   o Eventual Consistency

3. ==CAP== or Brewer's Theorem
   o Consistency
   o Availability
   o Partition Tolerance (partial failure tolerance)

Item	ACID	BASE
Casting (data structure)	Schema must exist	Dynamic
	Table structure exists	Adjust on the fly
	Columns data typed	Store dissimilar data
Consistency	Strong consistency available	Strong, eventual, or none
Processing Focus	Transactional	Key-value stores
	Row/column	Wide-column stores
History	1970s application storage	2000s unstructured storage
Scaling	Product dependent	Automatically spreads data across servers
Origin	Mixture	Open-source
Transaction	Yes	Possible

- Database Environments

  o Pre-production Environments

    1. Development

    2. Testing

    3. Sandboxes or Experimental

  o Production

- Database Organization

  1. Hierarchical (tree schema)

  2. ==Relational== (schema on write)
     - ==Multidimensional==
     - Temporal (valid time, transaction time, decision time)

  3. Non-relational (schema on Read)
     - Spatial
     - Column-oriented

In- memory

002,Desk,500

10,5,25,1,100

ID	Item	Price
0001	Chair	100
0002	Desk	500
0003	Table	250
0004	Cabinet	150
0005	Marker	10

row-oriented

ID	Time	I/O	Errors
0001	8:00:01	10	1
0002	8:01:25	5	0
0003	8:02:15	25	2
0004	8:05:00	1	1
0005	8:15:00	100	4

column-oriented

- Common Database Processes

  o Capacity and growth projections

  o Change Data Capture (CDC)

  o Replication (mirroring, change log shipping)

  o Data Recovery

  o Archiving, Purging and Retention

2. ACTIVITIES

a) Database Technology Support

   Select and maintain the software that stores and manages

   the data

b) Database Operations Support

   Specific to the data and process that the software manages

a) **Database Technology Support**

　　1.　Understand DB Technology Characteristics

　　2.　Evaluate DB Technology

　　3.　Manage and Monitor DB Technology

b) Database Operations Support (Manage Databases)

1. Understand requirements

2. Plan for Business Continuity

3. Develop Database Instances

4. Manage Database Performance

5. Manage Test Data Sets

6. Manage Data Migrations

3. TOOLS

1. Data Modeling Tools

2. Database Monitoring Tools

3. Database Management Tools

4. Developer Support Tools

4. TECHNIQUES

1. Test in Lower Environment

2. Physical Naming Standards (ISO/IEC 11179) –

   Metadata registries (MDR)

3. Script Usage for changes

5. IMPLEMENTATION GUIDELINES

- Readiness Assessment / Risk Assessment

  o Data Loss

  o Technology readiness

- Organizational and Cultural Change

  o Enhanced communications

6. GOVERNANCE

- Metrics

- Information Asset Tracking

- Data Audits and Data Validation