Design Approach for a Data Sharing Environment

Presented by
Gene Boomer
CNO Financial
Discussion topics include:

- Why Data Sharing
  - Impetus for establishing such an environment

- The Artifacts
  - Design approach and logical models

- Effect of Big Data
  - Evolving the architecture

- Wrap up and Questions
  - Final thoughts
The Usual Issue

• Sharing data meant building point-to-point (P2P) interfaces

• Main issues with P2P
  – Minimal to no reuse opportunity
  – Data redundancies
  – Data integrity
  – Business rule integration
  – Does not scale
  – Hindrance to growth

• Indicative of a low Information maturity level
Conceptual Solution

- Design a central repository of Trusted Sources, we call ours CNO Unified Repository (CURe)
- Assure scalable design approach for storing and moving data among trusted sources
- Establish guiding tenets
  - No “boil the ocean”
  - No data cleansing
  - No “system” codes
  - No direct access to Trusted Sources
- Sponsorship is important, engaged business is critical
The Artifacts

- Adoption of the DM-BOK framework (we will adopt DM-BOK2)
- Data at rest – establish common data model patterns for hub subject areas
- Data in movement – establish common and reusable integration patterns to move content into and out of data sharing hub
Subject Area Design

Subject Area design highly abstracted:
- Subject Object
- Object Attribute
- Attribute Value

Surrogate Key generated per each new distinct object

Cross-reference via hashed business key(s) along with supplying Admin System

Lineage provided at the object and attribute grain
Subject Area Linking

- The Relationship design pattern is highly abstracted similar to Subject Areas:
  - Base Subject Object
  - Related Subject Object
  - Related Subject Area
  - Relationship Attribute
  - Attribute Value

- Relationship categorized by relationship type

- Base and related subject object can be from the same Subject Area (e.g. Rider related to base Policy within Account SA)

- Lineage provided at the relationship and attribute grain
Architecture Approach

- **Enterprise Data Sharing approach utilized:**
  - What: effective data sharing environment
  - How: establish a solid reference architecture
  - Where: at the enterprise level in line with corporate tenets
  - Who: sponsored by Business areas, joint governance by Business and IT, implementation and maintenance by EDM
  - When: iteratively built upon business plan projects over time
  - Why: because it is the correct and best approach to help the enterprise grow
Data Process Flow

- Process flow followed as an approach of data governance
- Business involvement with the process although official data stewardship is lacking at enterprise level
- Help address semantic issues across enterprise and within business lines
- Potentially treat flow like an agile sprint.
- Goal to share content quickly
- Determined that same flow can be used for report process
• Data Architecture approach put forth:
  – Source content from administrative systems and other applications that contain data to be shared across the enterprise
  – Extract content to an enterprise level landing zone
  – Content is transformed into a “common contributor format” based on defined canonical / semantic model and ingested into CURe
  – Content is subscribed from CURe based on request
  – Content can be consumed into other data stores or fed back to admin systems
Big Data Implication

- Implement landing zone as a data lake within Big Data environment:
  - Use BD environment to capture raw content from admin systems
  - Utilize advantage of BD environment with regards to volume, velocity and variety

- Two main use cases supported by this approach:
  - Operational efficiency by utilizing BD environment for staging / archiving content
  - Ability to perform advanced / predictive analytics on content
  - Plan is to support both use cases

What is a data lake?
A repository for large quantities and varieties of data, both structured and unstructured.

- Data generalists/programmers can tap the stream data for real-time analytics.
- The data lake accepts input from various sources and can preserve both the original data fidelity and the lineage of data transformations. Data models emerge with usage over time rather than being imposed up front.
- Data scientists use the lake for discovery and ideation.
- The lake can serve as a staging area for the data warehouse, the location of more carefully “treated” data for reporting and analysis in batch mode.

Data lakes take advantage of commodity cluster computing techniques for massively scalable, low-cost storage of data files in any format.

www.pwc.com/techforecast
Final Thoughts

- Have to move beyond P2P protocol
- Establishment of common data model and data integration patterns was essential
- Information maturity progresses in the right direction over time
- Utilize data sharing approach in support of our vision statement

Make Data a strategic asset to the Enterprise. Information derived from our Data assets supports CNO’s vision. Information needs of all the stakeholders in the Enterprise will be met in terms of information availability, security, and quality.

We accomplish this by focusing on:
- Integrity: obtaining highest degree of data quality and information confidence
- Customer-Focus: meeting and exceeding the information needs of all customers whether internal or external
- Excellence: exceptional information utilization and distribution process
- Teamwork: utilizing the right resource with the right role at the right time
Questions

Thank you for your time and interest

Contact Information:
Gene Boomer
Director – Data Strategy
CNO Financial Group, Inc.
gene.boomer@cnoinc.com