21st Century Data Modeling for Effective Analytics

Anne Marie Smith, Ph.D.
Vice President of Education
Principal Consultant
EWSolutions, Inc.

https://www.linkedin.com/in/annemariesmith
EWSolutions’ Background

EWSolutions is a Chicago-headquartered strategic partner and full life-cycle systems integrator providing both award winning strategic consulting and full-service implementation services. This combination affords our client partners a full range of services for any size enterprise information management, metadata management, data governance and data warehouse/business intelligence initiative. Our notable client partner projects have been featured in the Chicago Tribune, Federal Computer Weekly, Journal of the American Medical Informatics Association (JAMIA), Crain’s Chicago Business, FBI, The Doings and won the 2004 Intelligent Enterprise’s RealWare award, 2007 Excellence in Information Integrity Award nomination, DM Review’s 2005 World Class Solutions award, 2016 CIO Review 20 Most Promising Enterprise Architecture providers and 2018 Innovation award.

For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training, email us at info@EWSolutions.com
Over 150 Client Partners Since 1997

AFLAC
AgVantis
AON (Hewitt Associates)
Arizona Supreme Court
Bank of America (Countrywide)
Bank of Montreal
BankUnited
Basic American Foods
Becton, Dickinson and Company
Blue Cross Blue Shield Association
Blue Cross Blue Shield companies
Booz Allen Hamilton
Branch Banking & Trust (BB&T)
British Petroleum (BP)
California DMV
California State Fund
Canadian Institute for Health Information (CIHI)
Canadian National Railway
Capella University
Cigna
College Board
Comcast
Comtech Cable Systems
Contech Systems, Inc.
CostCo
D.A. Davidson
Defense Logistics Agency (DLA)
Delta Dental
Department of Defense (DoD)
Discover
Driehaus Capital Management
Eli Lilly and Company
Environment Protection Agency (EPA)
Erwin
Farmers Insurance Group
Federal Aviation Administration
Federal Bureau of Investigation (FBI)
FedEx
Fidelity Information Services
FICO (Fair Issac)
Ford Motor Company
GlaxoSmithKline
Harbor Funds
Harris Bank (BMO)
The Hartford
Harvard Pilgrim HealthCare
Health Care Services Corporation
HP (Hewlett-Packard)
IDMA (Insurance Data Mgt. Institute)
Information Resources Inc.
International Paper
IQVIA (Quintiles)
Janus Mutual Funds
Johnson Controls
Key Bank
LiquidNet
Loyola Medical Center
Manulife Financial (John Hancock Fin.)
Mayo Clinic
McDonalds
Micron
Microsoft
MoneyGram
NASA
National City Bank (PNC Financial)
Nationwide
Neighborhood Health Plan
NORC
Physicians Mutual Insurance
Pillsbury
Quartz Benefits
Rowan College of South Jersey (RCSJ)
Sallie Mae
Saudi Arabia Minister of the Interior
Saudi Telecom Company (STC)
Schneider National
Secretary of Defense/Logistics
Singapore Defense Science & Technology Agency (DSTA)
Social Security Administration
South Orange County Community College
Spherion
Standard Bank of South Africa
SunTrust Bank
Surescripts
Target Corporation
Texas CPA
The Regence Group
Thomson Multimedia (RCA)
Thrivent Financial
United Healthcare (UHC)
United Health Group
United Nations (ICAO)
United States Air Force
United States Federal Courts
United States Department of State
United States Fish and Wildlife Service
United States Navy
United States Navy (BUPERS)
United States Transportation Command
University of Michigan
University of Wisconsin Health
USAA
US Cellular
Vera Bradley
Waste Management
Wells Fargo
Wisconsin Department of Transportation
Workers’ Compensation Insurance Rating Bureau of California (WCIRB)
Zurich Cantonional Bank

For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training email us at Info@EWSolutions.com

Over 155 Successful Client Partner Implementations
Contact us at INFO@EWSolutions.com to become our next success

© 1997 to present, Enterprise Warehousing Solutions, Inc. (EWSolutions) – 3
Anne Marie Smith, Ph.D.

- Internationally recognized expert and speaker in the fields of enterprise information management, data architecture and data modeling, metadata management and data warehousing

- One of the top industry experts in data governance and information management strategy and planning. Received DAMA International Professional Achievement Award - 2015

- Over 20 years experience in delivering solutions in enterprise information management for numerous companies / clients across varied industries

- Published almost 100 articles that have been featured in some of the industry’s most prestigious magazines and journals

- Contributing author and reviewer to the DAMA Data Management Body of Knowledge (DAMA-DMBOK©)

- Has taught at several institutions of higher learning including LaSalle University

- Earned a Ph.D. in Management Information Systems and holds CDMP© certification

- https://www.linkedin.com/in/annemariesmith
Agenda

- Data Modeling and Data Architecture
- Data Modeling for Analytics – Best Practices
- Data Modeling at Scale
- Future of Data Modeling – Data Modeling for 21st Century
Data Modeling and Data Architecture
Data modeling: process used to define and analyze data requirements that support the business processes for information systems

- Process involves professional data modelers working closely with business stakeholders and potential users of the information system

Using proper data modeling techniques and methodologies allow an organization to model data in a standard, consistent, predictable manner to manage data as a resource

- Model should reflect concepts of the organization and how they relate to one another and to the business processes of the application (operational or analytical)
- Data models should be part of the organization’s business documentation
- Data literate organizations develop and maintain data models for all applications and at all levels
Data Architecture Definition

- **Data architecture** describes the structure of an organization's logical and physical data assets and data management resources
  - Includes models, policies, rules, and standards that govern the collection, storage, arrangement, integration, and use of data in organization

- The goal of data architecture is to translate business needs into data and system requirements and to manage data and its flow through the enterprise and all applications

- Data literate organizations use data architecture to manage data and process flow throughout organization
Data Architecture vs. Data Modeling

Data Architecture: defines the blueprint for managing data assets to establish strategic data requirements and designs to meet those requirements using models and other artifacts.

Data Modeling: process of discovering, analyzing, representing, and communicating data requirements in a precise form called the data model.

Data architecture and data modeling are essential components of effective 21st century systems, data management capabilities, and user empowerment.
Data Management Framework

• **Data Governance**: Planning, supervision and control over data management and use.
• **Data Architecture Management**: Defining the blueprint for managing data assets.
• **Data Modeling and Design**: Analysis, design, implementation, testing, deployment, maintenance.
• **Data Storage and Operations**: Providing support from data acquisition to purging.
• **Data / Information Security Management**: Insuring privacy, confidentiality and appropriate access.
• **Data Quality Management**: Defining, monitoring and improving data quality.
• **Reference and Master Data Management**: Managing golden versions and replicas of data.
• **Data Warehousing and Business Intelligence** Management: Enabling reporting and analysis.
• **Unstructured Data Management**: Managing data found outside of databases.
• **Metadata Management**: Integrating, controlling and providing Metadata.
• **Data Integration**: Enabling transfer of data across applications and systems
Data Modeling Goals, Benefits and Phases
Goals of Data Modeling

- Ensures that all data objects required by the database are accurately represented
  - Omission of data will lead to creation of faulty reports and produce incorrect results for operations and decisions

- A data model helps design the database at the conceptual, physical and logical levels
  - Data model structure helps to define the relational tables, primary and foreign keys and stored procedures to support a robust database
  - Designing the data model can identify missing and redundant data early in the development process
Benefits and Value of Data Modeling

- **Higher quality data**
  - Data model diagrams display requirements and business rules; can anticipate large-scale data corruption opportunities in advance
  - Data models define rules that monitor data quality, which reduces the chance of errors
  - Data model development includes definition of concepts and attributes, establishing terms and meaning throughout organization

- **Increased internal communication about data and data processes**
  - Enables business to define how data is generated and moved throughout applications based on business needs, establishes organization-supported business definitions for terms (attributes) and concepts

- **Reduced development and maintenance costs**
  - Data modeling surfaces errors and inconsistencies early so they are far easier and cheaper to correct

- **Improved performance**
  - An organized database operates more efficiently; building a database from a well-designed data model stops endless searching across database; enables faster results
Data Modeling Phases

-q Four types of data models
  - Conceptual
  - Logical
  - Physical
  - Dimensional – for analytics, data marts, business intelligence

-q Every data modeling effort should begin with a conceptual data model, expanded to a logical data model, then into the physical data model
  - Starting at the physical data model level prevents discovery of business requirements and hinders ability to scale the database appropriately
Conceptual Data Model

- Identifies business concepts to support the organization’s activities and requirements
  - Represents data concepts and their relationships, concepts and relationships defined for organizational use

- Conceptual data models are designed and developed for a business audience
  - Business requirements should drive all database and application design, starting at conceptual level

- Developed independently of hardware specifications like data storage capacity, location or software specifications like DBMS-specific content - no technology requirements
Conceptual Data Model Example
Logical Data Model

- Describes data requirements in more detail than the conceptual level data model
  - Designed based on the conceptual data model
  - Identifies business requirements, including attribute definitions to support business case
  - Definitions should be documented and accessible across organization

- Designed and developed independently from the DBMS – includes business terms, business relationships

- Each entity has associated attributes, and each attribute has one location in the logical data model
  - Data attributes include datatypes and usual length, derived from business requirements
Logical Data Model Example
Physical Data Model

- Describes data to be used in an application with a specific database management system (DBMS)
  - Location, data storage or technology to be used in the project are included in physical data model content

- Columns (attributes) should have exact datatypes, lengths assigned and default values
  - Each table (entity) may contain columns that are repeated for performance or application reasons
  - Table and column names may be abbreviated based on modeling standards

- Primary and foreign keys (drawn from logical model), views, indexes, access profiles, and authorizations, etc. are defined at physical level
Physical Data Model Example
Dimensional Data Model

- Describes data in a format optimized for data storage in a data warehouse or data mart, used for reporting and analytics
- Optimizes the database for faster retrieval of data, using “fact” and “dimension” tables
  - Facts are the measurements/metrics or facts from a business process. Example: quarterly sales number
  - Dimensions provide context for a business process event; who, what, where of a fact. Example: for the fact quarterly sales number, dimensions include customer names, customer location, product name
- Structure used by BI query, reporting and analytical tools
  - Dimensional models have specific advantages for reporting and analysis, arranging data so it is easier to retrieve information and generate reports
Some Data Modeling Best Practices for Analytics
Data modeling is different from data analysis. Well-designed and implemented data models optimize data to support data analysis (operational or analytical).

Developing data models in all stages (conceptual, logical, physical, dimensional) enables improved understanding of data and supports effective data analysis:

- Developing a dimensional data model based on the logical data model can reduce time for development and improve usability of data.

Data definition, part of best practices in data modeling, is an essential component of effective analytics:

- Tables and attributes without definitions are counterintuitive.
- Data catalog, business glossary, data dictionary can support all forms of analytics.
Human Factors in Data Modeling

Business domain subject experts – provide requirements to data modeler / data architect, validates data model content, may include data analysts and data scientists

Data modeler / data architect – develops conceptual and logical data models from business requirements, supports physical and dimensional database creation

Database developer, database administrator, programmer – builds physical and dimensional data structures from data models
Data Models for Analytics

- Requirements for queries and reports will change, models must reflect changing needs with relevant data and analytical processes

- Optimizing performance is a primary goal of analytically-focused data models; techniques include:
  - De-normalizing data from logical data model to physical version can support improved access for queries and reports; denormalized database may not resemble normalized logical data model
  - Creating views based on specific and repeated business needs
  - Using pre-calculated results in appropriate tables, appending detail data to master record table, redundant data storage, etc.
  - Analytics and reporting are not data mining. Data mining uses data structures unique to a specific set of questions or a set of relationships under exploration
Data Modeling at Scale
Data Modeling at Scale

- Data modeling decisions affect performance and usability of data
  - Initially
  - Over time

- Effective data architecture and data modeling are key aspects of performance improvements, operations effectiveness, and user access to data for analytics
  - Comprehensive, pragmatic data model facilitates BI tool usage for business value

- Scalability of any application is based, in part, on the data architecture and the content of each data model
  - Cannot “tune” around a poorly designed data architecture and inaccurate data models
  - Business requirements must be documented and understood for effective data models regardless of application/database scale
  - Accurate and validated data definitions are essential for scalable data models and applications
Design for Scalability with Data Models

- Initial investment in proper enterprise data architecture and accurate data models is essential to scaling any application (operational, analytical)
  - Application/database scale may start small, can grow properly only with clear requirements and accurate data models

- Overly complex architecture and “clever” data models often do not scale well, and negatively affect performance and data usage
  - Clear, well-documented, and business-focused data models offer best opportunities to support scaling application and/or database as needed

- Design for practicality while following best practices and industry standards for effective enterprise and application data architecture with accurate data models
# Proven Strategies for Scalability with Data Models

<table>
<thead>
<tr>
<th>Define</th>
<th>Define and document data architecture and data modeling standards for the enterprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicate</td>
<td>Communicate standards to organization – clearly and consistently – expect them to be followed</td>
</tr>
<tr>
<td>Train</td>
<td>Train business analysts and application developers to work with data models to reduce communication barrier with SMEs and data architects / data modelers</td>
</tr>
<tr>
<td>Incorporate</td>
<td>Incorporate data modeling practices and techniques into all business subject areas to maintain enterprise view while supporting subject area’s requirements</td>
</tr>
<tr>
<td>Capture and store</td>
<td>Capture and store all relevant metadata centrally and consistently</td>
</tr>
<tr>
<td>Use</td>
<td>Use the data to test the data model’s accuracy – validate each model before proceeding to next stage</td>
</tr>
<tr>
<td>Insist on</td>
<td>Insist on conceptual and logical data modeling (before physical data model design) for all business areas throughout the lifecycle to support actual business requirements</td>
</tr>
<tr>
<td>Follow</td>
<td>Follow release management process consistently to manage and document data model changes</td>
</tr>
</tbody>
</table>
Challenges with Scaling Traditional Data Models

- Data variety, especially using big data and analytics, can offer challenges to data modeling practices
  - Data models traditionally focused on only structured, relational data
  - Big data and analytics requirements can (and should be) part of a data model

- Organizations think that time-to-delivery value is a concern for developing the data model forms
  - Data modeling does not have to be a lengthy and tedious effort – skill and best practices
  - Most organizations focus on development and struggle to see value in staged data modeling
    - Conceptual
    - Logical
    - Physical
    - Dimensional

- Many traditional data model processes are cumbersome and can be streamlined to reduce challenges
  - Building conceptual and logical data models before physical can identify data requirements from business perspective to support effective database and application access
  - Documenting data definitions is essential and can be done during data model development

- Perception of data modeling value and lack of data modeling skills are major challenges in 21st century organizations
Future of Data Modeling
Fundamental Pillar of 21st Century Businesses

- High-quality data foundation is important for success in digital technologies and business competitive advantage
  - High data quality based on underlying structure that connects an organization’s data assets from business requirements
  - Business case for all development should include all stages of data models

- Starting with an effective data model set is essential for high-quality data that can transform an organization
  - Advanced analytics
  - Self-service business intelligence and reporting capabilities
  - Artificial intelligence and machine learning
  - Other digital business applications
Data Modeling Success Tips

- Identify business goals for data with each initiative – do not focus only on processes
  - Develop a standard approach to enterprise data architecture that includes the three stages of data models
  - Develop good business use cases for data modeling to demonstrate effectiveness and value
  - Focus on business needs and data assets from project inception

- Select the right data modeling approach, without sacrificing the essential model stages (conceptual, logical, physical, dimensional)
  - Base approach on the business use case and underlying technology
  - Design each model to the appropriate level of detail for the business requirements

- **Data model must fully address the requirements, but not over-engineer them**

- Seamless and open communication between all teams and stakeholders are essential for successful data modeling
  - Ensure all relevant data elements are incorporated and have the same meaning and interpretation across the organization, with supporting documentation

- Remember a data model is a living artifact that requires updating and maintenance
  - Creating a formal process for updating all models is critical, including updating model documentation
21st Century Data Modeling Skills

Conceptual and logical thinking and design

Critical thinking and analysis

Effective communications across business and IT

Abstract thinking and detailed analysis capabilities

Documentation skills

Data representation and definition

Data modeling tool competency

Adaptability and teamwork

Over 155 Successful Client Partner Implementations
Contact us at INFO@EWSolutions.com to become our next success
Future Data Modeling Trends and Techniques

**Agile Data Modeling:** “just-in-time data modeling” using “best practices with a minimally sufficient design” and “the right data model for specific situations” for a mix of unstructured data, relational data, master data, and dimensional data.

**Graph Databases:** Graph databases provide easy, quick, visuals of business cases, and they need data models.

**Business Intelligence and Reporting:** Data modelers will continue to diagram and document relational databases, focused on specific business and technical use cases.

**Cloud-based Computing:** Data modeling for cloud applications focuses on scalability, speed, and reduced cost. Cloud computing will encourage “larger applications that circulate more data through their hardware”, and require well-documented data models and effective metadata management.

**Wider Focus:** data models must address and document multi-level concerns: business level, solution level, and implementation challenges.
In the Final Analysis
Data modeling is an essential component for any application / system – operational, analytical, etc.

- Data modeling is a component of an organization’s enterprise data architecture

Adopting data modeling best practices can offer a competitive advantage for organizations that value data as an organizational asset

- Data modeling techniques should be practiced consistently throughout organization

There are many aspects to effective data modeling, including data management, metadata management, data architecture, etc.

Building and sustaining an effective data modeling competency requires commitment across the organization, from business and IT

Continuous learning is a crucial component of data modeling competency
Questions & Answers

Data Management Consulting Services
Over 155 Client Partners
100% Data Management Success since 1997

World-Class Training
More than 40 courses, taught by industry experts. Over 150 satisfied client partners

Over 155 Successful Client Partner Implementations
Contact us at INFO@EWSolutions.com to become our next success
Anne Marie Smith, Ph.D.
VP of Education, Chief Methodologist

EWSolutions, Inc.
AMSmith@ewsolutions.com
856-494-3722 (O)
https://www.linkedin.com/in/annemariesmith

Contact us at INFO@EWSolutions.com to become our next success

EWSolutions
1 Tiffany Point, Suite #100
Bloomingdale, IL 60108-2960

Correspondance & Processing
PO Box 70
Hinsdale, IL 60522-0070

http://www.EWSolutions.com
Enterprise Warehousing Solutions, Inc.
© 2021 All Rights Reserved