

## 21<sup>st</sup> Century Data Modeling for Effective Analytics



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

<https://www.linkedin.com/in/annemariesmith>

# EWSolutions' Background



FBI  
Meritorious  
Service Award



**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.



2007 Excellence in Information Integrity Award Nomination



Excellence and Thought Leadership



2018: Most Innovative in Data Management Solutions



World Class Solutions Award Data Management



Best Business Intelligence Application Information Integration  
Client: Department of Defense



**For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training, email us at [info@EWSolutions.com](mailto:info@EWSolutions.com)**

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

For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training email us at [Info@EWSolutions.com](mailto:Info@EWSolutions.com)

- ❑ 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
- ❑ 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
- ❑ Has taught at several institutions of higher learning including LaSalle University
- ❑ Earned a Ph.D. in Management Information Systems and holds several certifications
- ❑ <https://www.linkedin.com/in/annemariesmith>

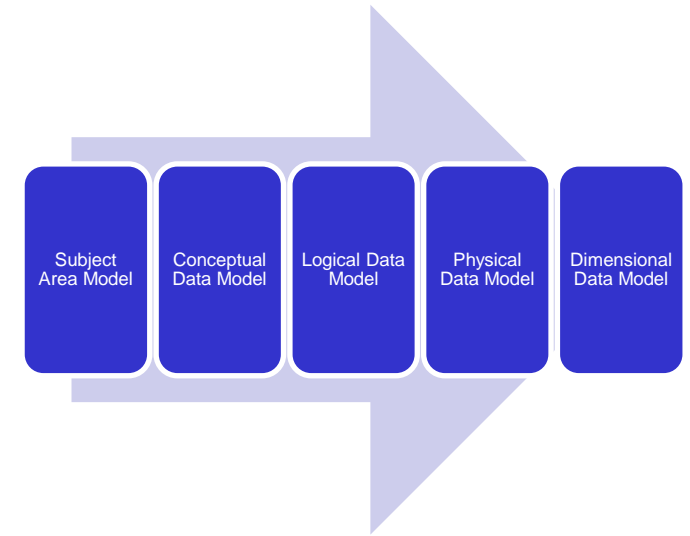


- ❑ Data Modeling and Data Architecture
- ❑ Data Modeling for Analytics – Best Practices
- ❑ Data Modeling at Scale
- ❑ Future of Data Modeling – Data Modeling for 21<sup>st</sup> Century



# Data Modeling and Data Architecture

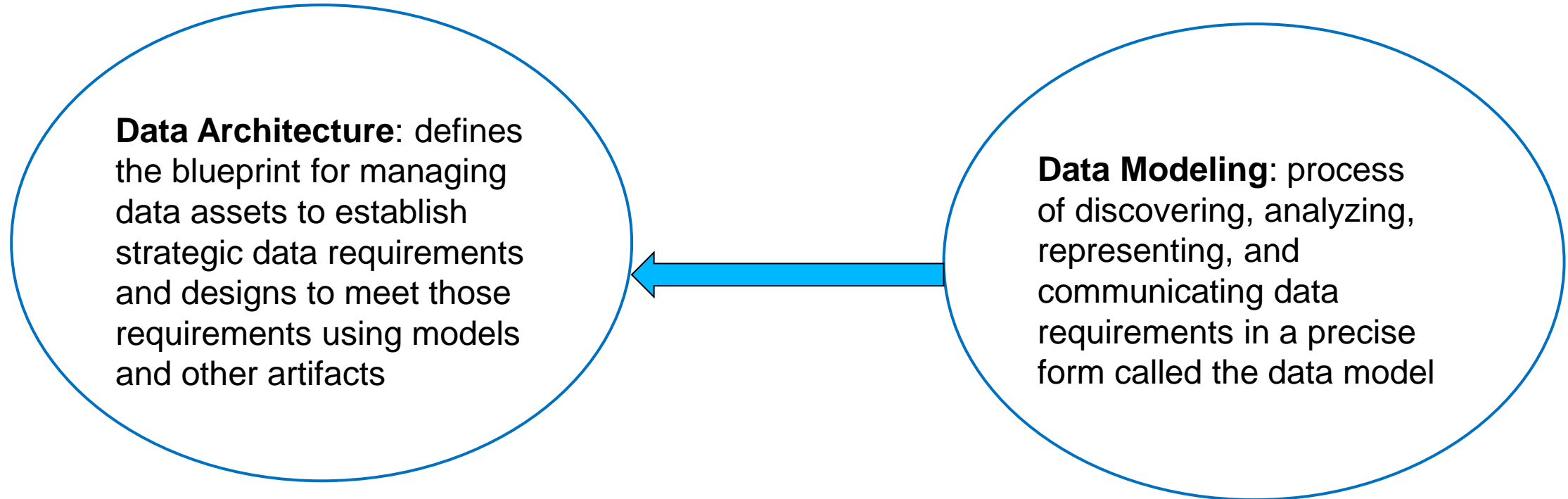
- ❑ **Data modeling:** process used to define and analyze data requirements that support the business processes for information systems
  - Process involves professional data modelers working closely with business stakeholders and potential users of the information system
- ❑ Using proper data modeling 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





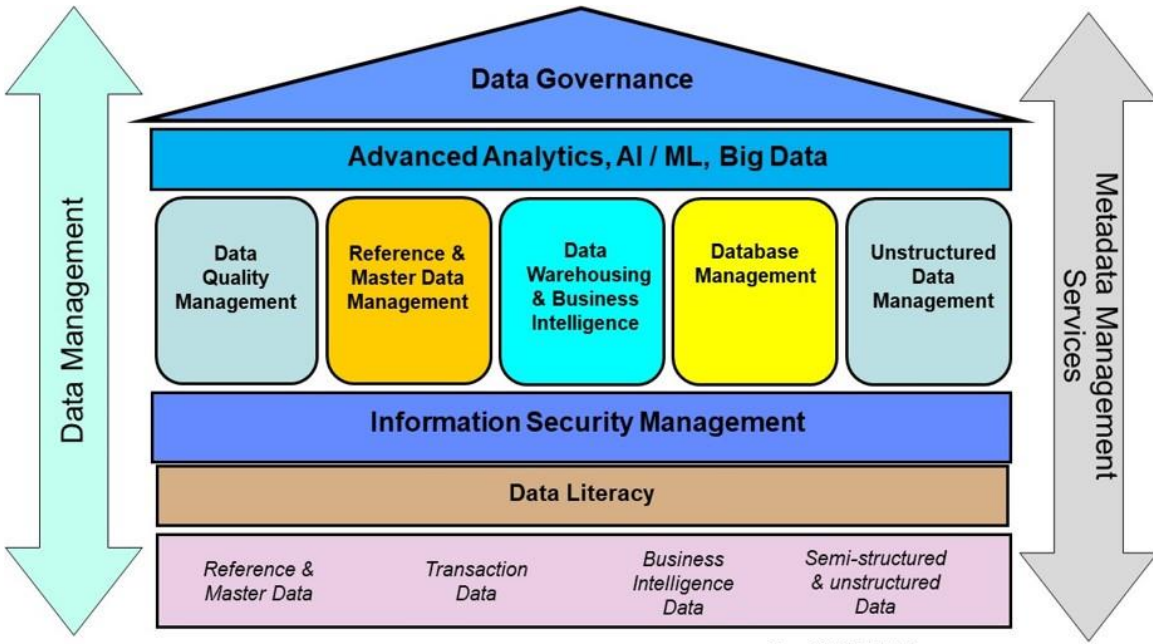
- ❑ **Data architecture** describes the structure of an organization's logical and physical data assets and data management resources
  - Includes models, policies, rules, and standards that govern the collection, storage, arrangement, integration, and use of data in organization
- ❑ The goal of data 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 and data modeling are essential components of effective 21<sup>st</sup> century systems, data management capabilities, and user empowerment***

# Data Management Framework



Since 2010 © EWolutions  
All Rights Reserved

- **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

- ❑ Ensures that all data objects required by the business process are accurately represented
  - Omission (or repetition) of data will lead to creation of faulty reports and produce incorrect results for operations and decisions
- ❑ A data model helps **identify and organize data** needed at the conceptual, logical and physical levels
  - Developing the series of data models can identify missing and redundant data early in the development process (conceptual and logical levels)
  - Data model structure helps to define the relational tables, primary and foreign keys and stored procedures to support a robust database

- ❑ **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
  - Supports data governance processes through identification of relevant data stewards and clear data definitions
- ❑ **Reduced development and maintenance costs**
  - Data modeling identifies 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 databases; enables faster results

- ❑ Five types of data models:
  - Subject Area
  - Conceptual
  - Logical
  - Physical
  - Dimensional – for analytics, data marts, business intelligence
- ❑ Once the organization has developed a robust subject area model for the entire organization, all data modeling efforts should begin with a conceptual data model based on the subject area model. The conceptual data model is 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

- ❑ Each subject area is a **high-level classification** that represents a group of concepts pertaining to a major topic of interest to an organization
- ❑ Subject areas can represent generic business concepts (customer, product, employee and finance), as well as industry specific areas
- ❑ SAM is not intended to represent each subject area as a “silo”, it is a view of the business
- ❑ There can be some discussion about where to place a concept, since each concept belongs in one and only one subject area
  - Resolve conflicts or identify concepts that can relate to multiple subject areas through conceptual model’s relationships



# Subject Area Model Example

## Finance

- Regulation
- Reporting
- Financial Accounting
- Accounts Payable
- Accounts Receivable
- Asset
- Treasury

## Product

- Finished Good
- Manufacturing Service
- Manufacturing Equipment
- Product Taxonomy
- Research & Development
- Design
- Testing
- Bill of Materials
- Build Plan

## Location

- Site
- Land
- Building
- Office
- Factory
- Port
- Airfield
- Warehouse
- Geographical area

## Customer

- Individual
- Company
- Pricing
- Sale
- Sales

**SAMPLE**

## Legal

- Policy
- Contract
- Compliance and Ethics
- Dispute
- Data Security and Privacy
- Government Affairs

## Material

- Raw Material
- Part
- Work in Process (WiP)
- Inventory

## Supplier

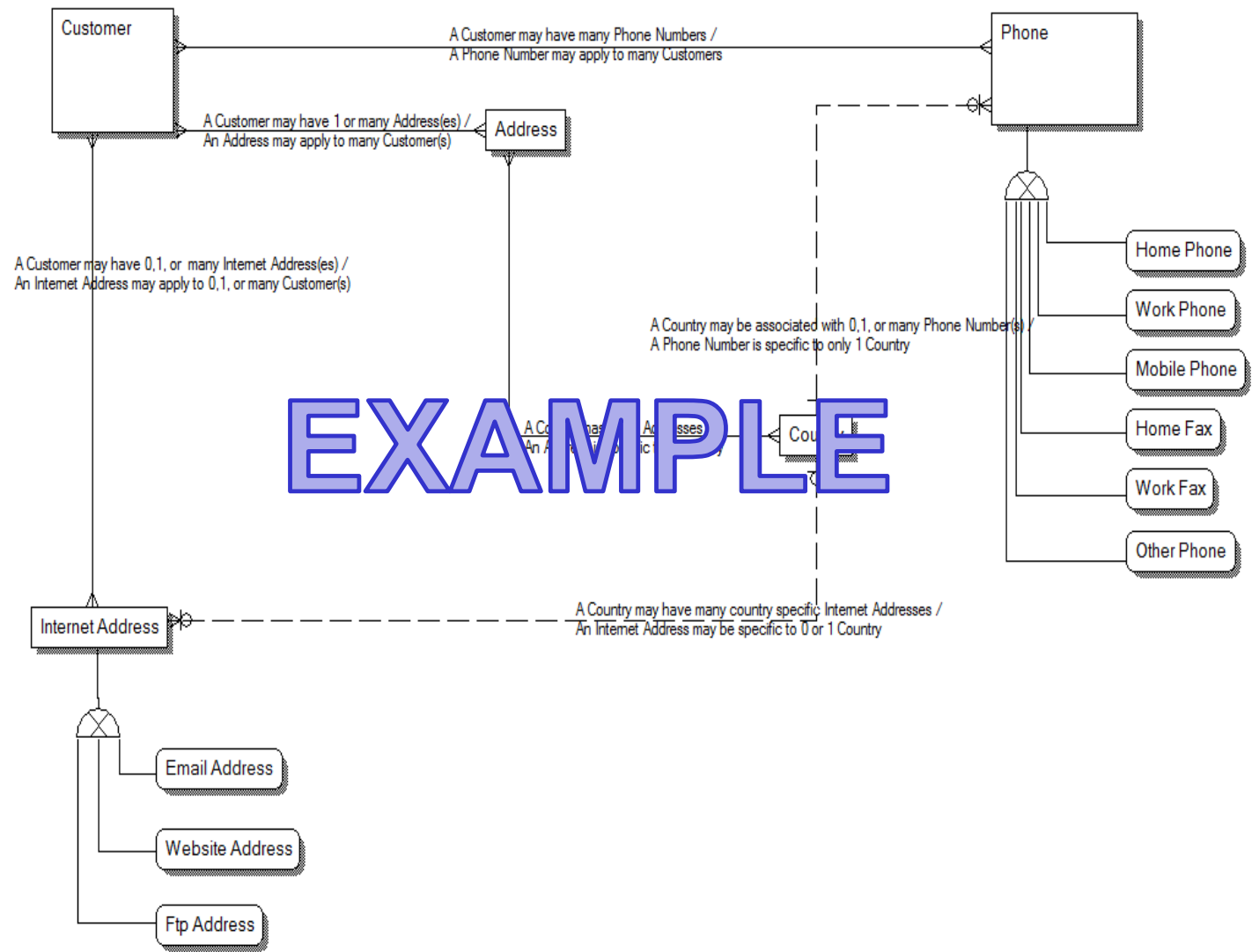
- Supplier Type
- Supplier Organization
- Commitment
- Performance
- Pricing
- Specification

## Party

- Employee
- Partner
- Tenant
- Third Party
- Government Agency

- ❑ Identifies business concepts to support the organization's activities and requirements
  - **Represents data concepts** and their relationships, defined for organizational use
  - Based on the organization's subject area model
- ❑ Conceptual data models are designed and developed for a business audience
  - Business requirements should drive all database and application design, starting at conceptual level – helps avoid redundant applications or conflicting databases
- ❑ Developed independently of hardware specifications like data storage capacity, location, or software specifications like DBMS-specific content - no technology requirements

# Conceptual Data Model Example

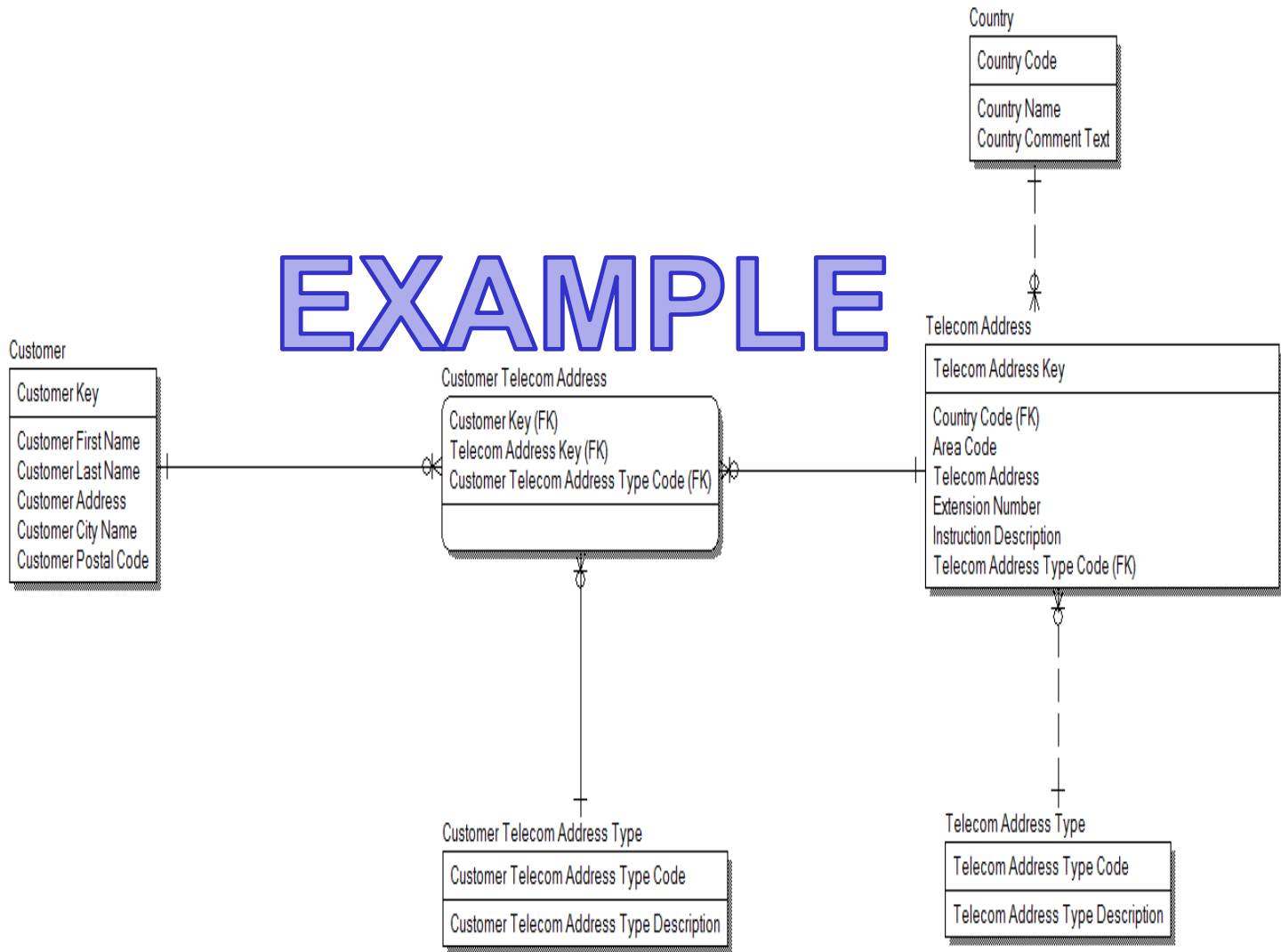


EXAMPLE

- ❑ 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 requirements
  - 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

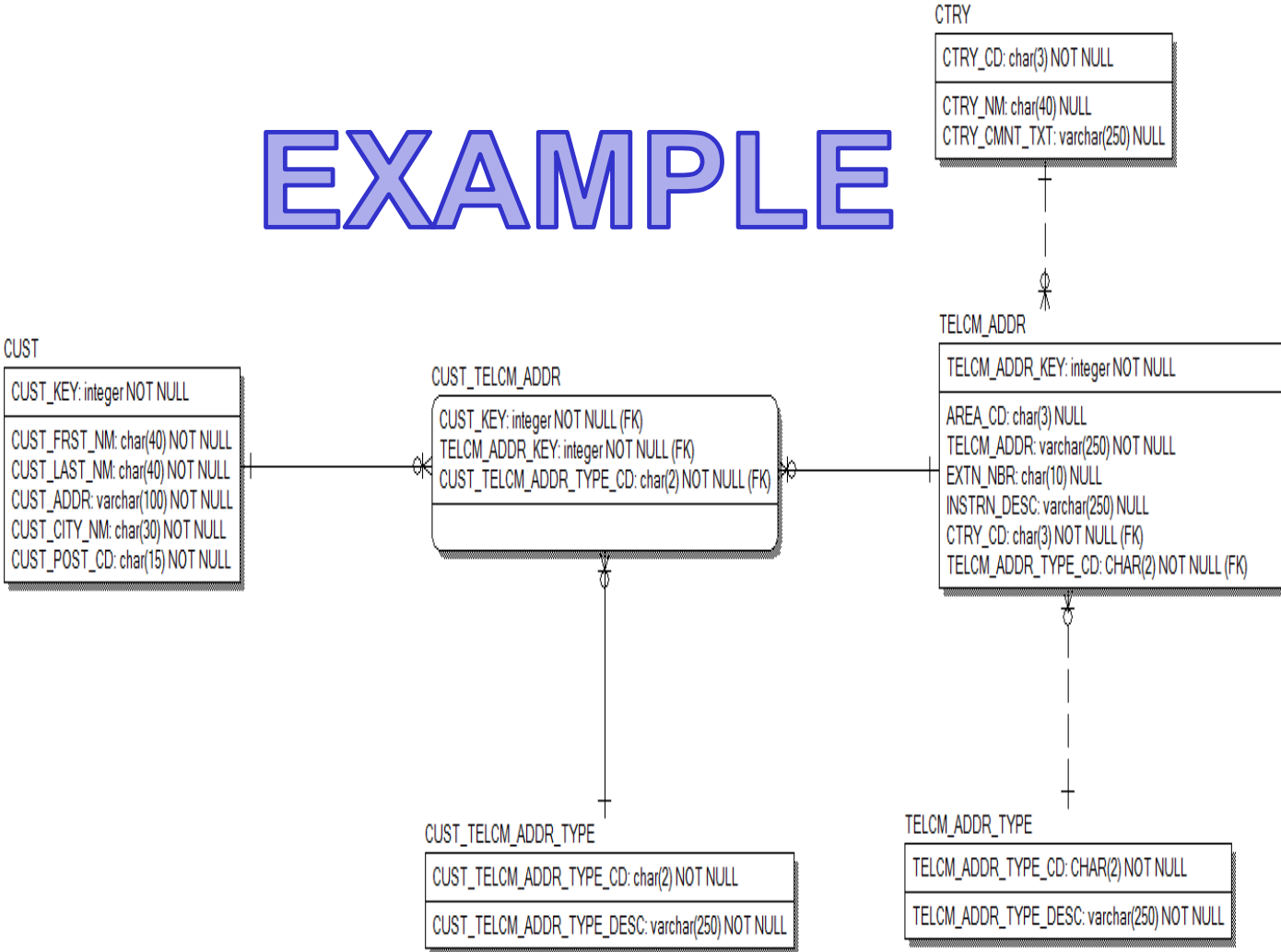
# Logical Data Model Example

EXAMPLE



- ❑ 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

## EXAMPLE

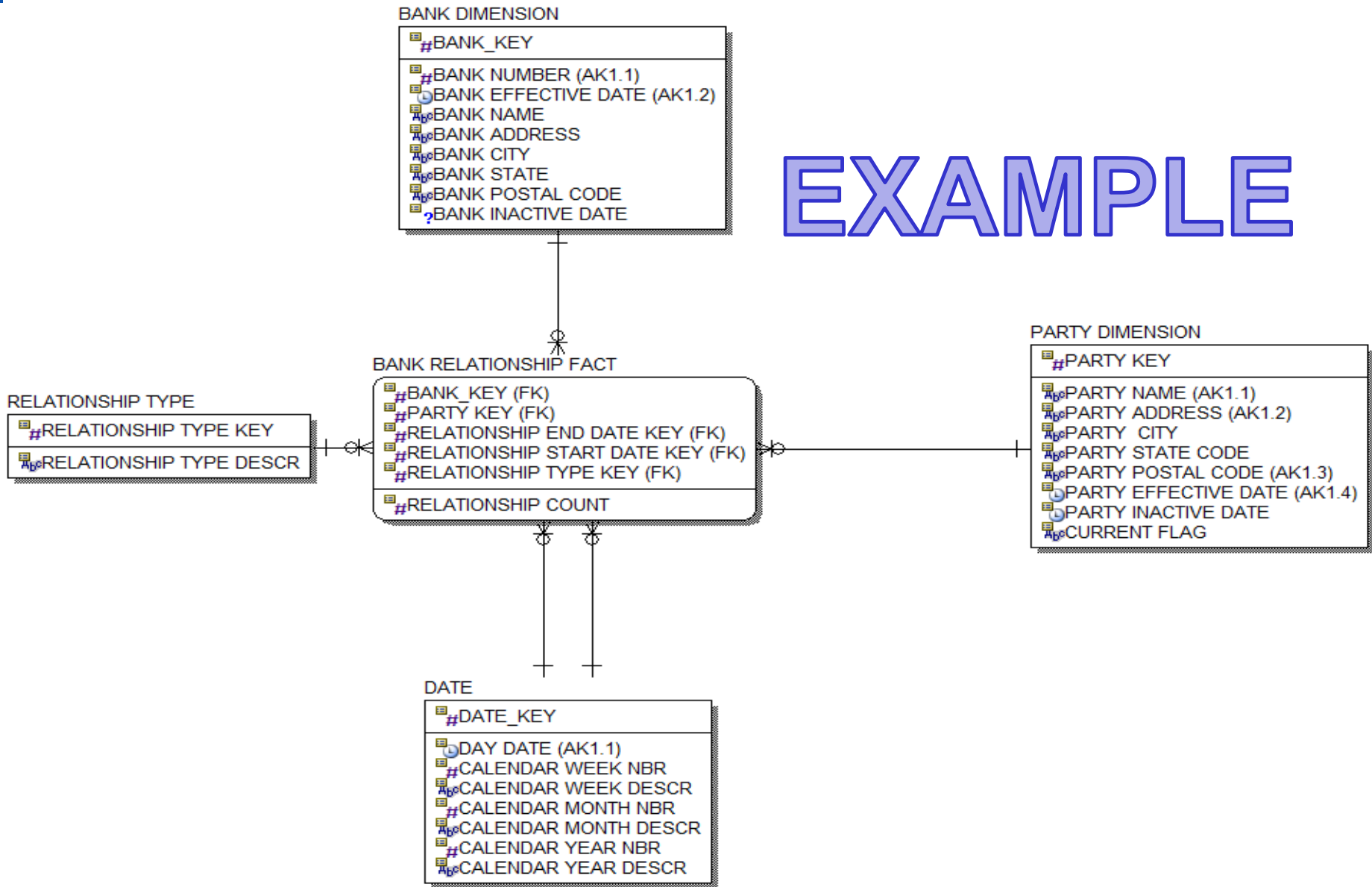




- ❑ Describes data in a format optimized for data storage in a data warehouse or data mart, used for reporting and analytics
- ❑ Optimizes the database for faster retrieval of data, using “fact” and “dimension” tables
  - **Facts** are the measurements/metrics or facts from a business process. Example: 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

# Dimensional Model Example

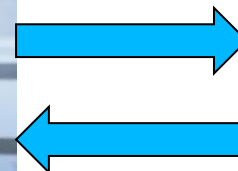
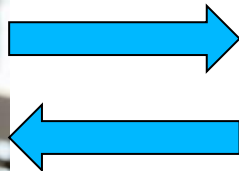
# EXAMPLE



# 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 and related content

- ❑ 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 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

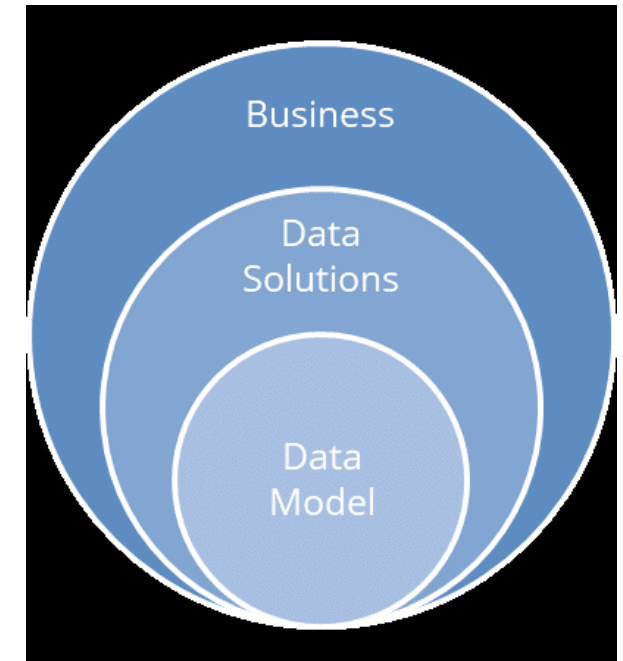
- ❑ Initial investment in **proper enterprise data architecture and accurate data models** is essential to scaling any application (operational, analytical)
  - Application/database scale may start small, can grow properly only with clear requirements and accurate data models
- ❑ Overly complex architecture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

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

- ❑ 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 21<sup>st</sup> century organizations

# Future of Data Modeling

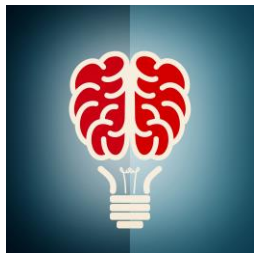
- ❑ High-quality data foundation is important for success in digital technologies and business competitive advantage
  - High data quality based on underlying structure that connects an organization's data assets from business requirements
  - Business case for all development should include all stages of data models
  
- ❑ Starting with an **effective data model set is essential** for enabling 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



- ❑ Identify business goals for data with each initiative – do not focus only on processes
  - Develop a standard approach to enterprise data architecture that includes all stages of data models and related documentation
  - Develop good business use cases for data modeling to demonstrate effectiveness and value of the technique
  - 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)
- ❑ Start with the organization's subject area model to identify the relevant subjects as focus
  - 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 across 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



# 21<sup>st</sup> Century Data Modeling Skills



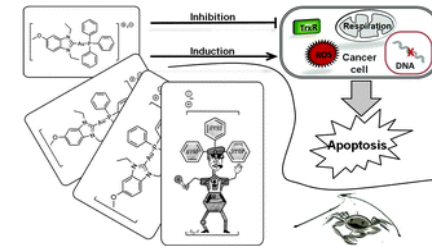
Conceptual and logical thinking and design



Critical thinking and analysis



Effective communications across business and IT



Abstract thinking and detailed analysis capabilities



Documentation skills

Field Name	Data Type	Description	Dep
ALBUM	Text	The title of this pressing	
ARTIST	Text	The name of the recording artist or group	
MEDIA	Text	CD, LP, EP, CA	
PUB DATE	Number	The date this pressing was published (YYYY)	
RECORD	Text	The recording number of this pressing	
DISCNUM	Text	Unique ID of this pressing	

Field Properties	
General	Lookup
Field Size	255
Format	
Input Mask	
Caption	
Default Value	
Validation Rule	
Validation Text	You must enter a unique serial number
Required	Yes
Allow Zero Length	No
Indexed	Yes (No Duplicates)

Data representation and definition



Data modeling tool competency



Adaptability and teamwork



**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



## Data Management Consulting Services

Over 155 Client Partners

100% Data Management Success since 1997



**DATAMANAGEMENT  
EDUCATION.COM**

**Online, On-Demand Courses**

# EW Solutions™

### World-Class Training

More than 40 courses, taught by industry experts. Over 150 satisfied client partners

**Data  
Warehouse  
Data Models**



**Big Data  
Meta Model**



**DataManagementU.com**

*Free knowledge portal*

Anne Marie Smith, Ph.D.  
VP of Education, Chief Methodologist

EWSolutions, Inc.  
[AMSmith@ewsolutions.com](mailto:AMSmith@ewsolutions.com)  
[856-494-3722 \(O\)](tel:856-494-3722)  
<https://www.linkedin.com/in/annemariesmith>

***EWSolutions***

1 Tiffany Point, Suite #100  
Bloomington, IL 60108-2960

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

<http://www.EWSolutions.com>

Enterprise Warehousing Solutions, Inc.  
© All Rights Reserved