# What Is In Your Data Warehouse?



peter.aiken@anythingawesome.com +1.804.382.5957

## Peter Aiken, Ph.D.

- I've been doing this a long time
- My work is recognized as useful
- Associate Professor of IS (vcu.edu)
- Institute for Defense Analyses (ida.org)
- DAMA International (dama.org)
- MIT CDO Society (iscdo.org)
- Anything Awesome (anythingawesome.com)
- Experienced w/ 500+ data management practices worldwide
- Multi-year immersions
  - US DoD (DISA/Army/Marines/DLA)
  - Nokia
  - Deutsche Bank
  - Wells Fargo
  - Walmart
  - HUD …
- 12 books and dozens of articles

**$1,500,000,000.00 USD**

# What Is In Your Data Warehousing Operations?





# What Do Your Data Warehousing Operations Consist Of?





# Of What Do Your Data Warehousing Operations Consist?



peter.aiken@anythingawesome.com +1.804.382.5957

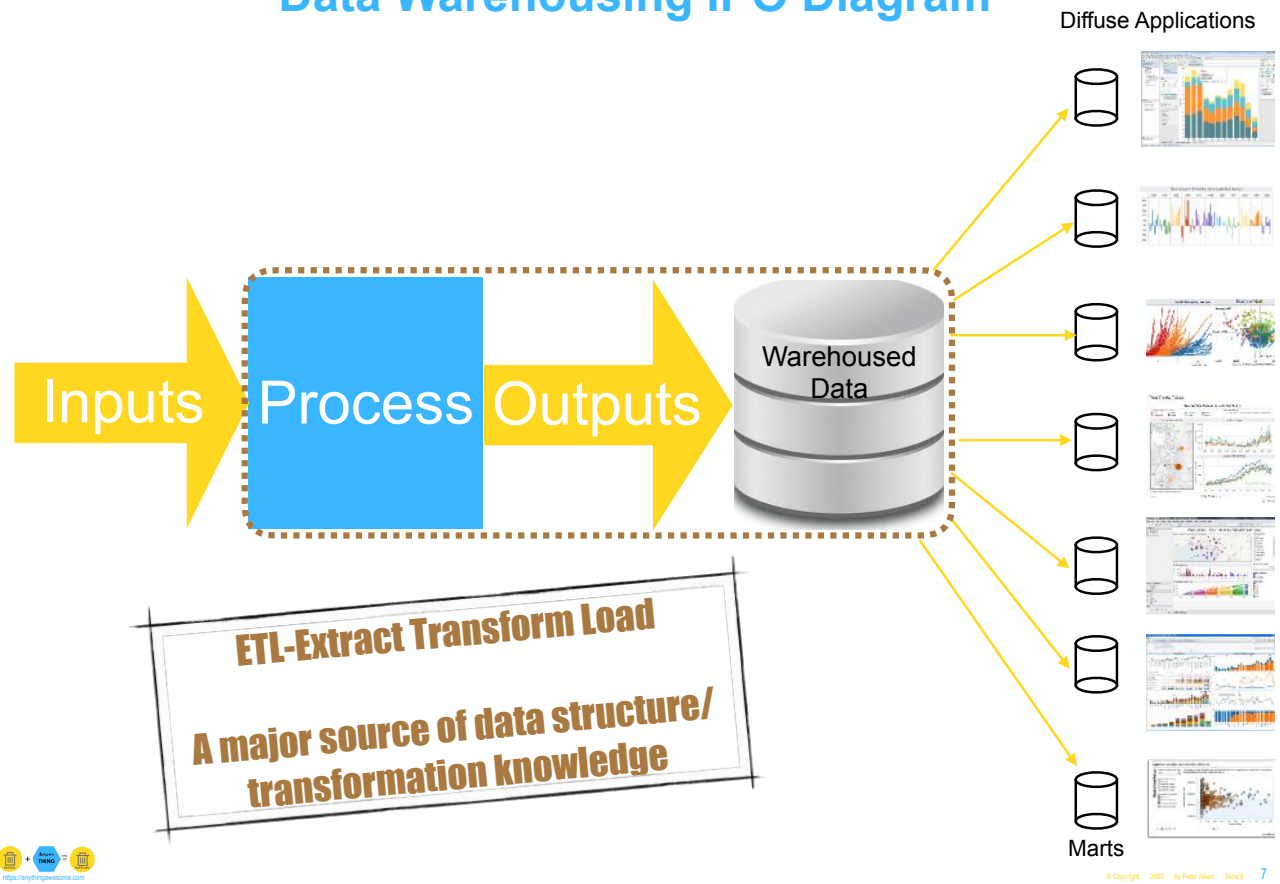
 

## Program Overview

- Definitions
  - Data Warehousing
  - DM BoK guidance
  - Legacy to digital conversion (cloud)
- Integration
  - Emphasize engineering talent
  - Incorporate leading not bleeding edge
  - Requires an adaptive rather than a prescriptive approach
- Preparation
  - Emphasis on storytelling first and visualization second
  - Analytics is both ubiquitous and not well understood
  - Keep improvements practically focused by strategy
- Best Practices
  - Cannot use what is not understood-understand what you have
  - PDCA (plan, do, check, act)
  - Iteratively refine/cull with respect to strategic direction
- Take Aways/References/Q&A


Of What Do Your Data Warehousing Operations Consist?

# Data Warehousing IPO Diagram






# Data Management Body of Knowledge (DM BoK V2)


**11 Practice Areas**

from *The DAMA Guide to the Data Management Body of Knowledge* 2E © 2017 by DAMA International

# Data Warehousing & Business Intelligence Management

**Definition**: Planning, implementation, and control processes to provide decision support data and support knowledge workers engaged in reporting, query and analysis.

**Goals**:
1. To support and enable effective business analysis and decision making by knowledge workers.
2. To build and maintain the environment / infrastructure to support business intelligence activity, specifically leveraging all the other data management functions to cost effectively deliver consistent integrated data for all BI activity.

**Inputs**:
- Business Drivers
- BI Data and Access Requirements
- Data Quality Requirements
- Data Security Requirements
- Data Architecture
- Technical Architecture
- Data Modeling Standards and Guidelines
- Transactional Data
- Master and Reference Data
- Industry and External Data

**Suppliers**:
- Executives and Managers
- Subject Matter Experts
- Data Governance Council
- Information Consumers (Internal and External)
- Data Producers
- Data Architects and Analysts

**Activities:**
1. **Understand Business Intelligence Information Needs (P)**
2. **Define and Maintain the DW / BI Architecture (P)**
3. **Implement Data Warehouses and Data Marts (D)**
4. **Implement BI Tools and User Interfaces (D)**
5. **Process Data for Business Intelligence (O)**
6. **Monitor and Tune Data Warehousing Processes (C)**
7. **Monitor and Tune BI Activity and Performance (C)**

**Participants**:
- Business Executives and Managers
- DM Execs and Other IT Mgmt
- BI Program Manager
- SMEs and Other Info Consumers
- Data Stewards
- Project Managers
- Data Architects and Analysts
- Data Integration (ETL) Specialists
- BI Specialists
- Database Administrators
- Data Security Administrators
- Data Quality Analysts

**Tools:**
- Database Management Systems
- Data Profiling Tools
- Data Integration Tools
- Data Cleansing Tools
- Business Intelligence Tools
- Analytic Applications
- Data Modeling Tools
- Performance Management Tools
- Meta-data Repository
- Data Quality Tools
- Data Security Tools

**Primary Deliverables**:
- DW/BI Architecture
- Data Warehouses
- Data Marts and OLAP Cubes
- Dashboards and Scorecards
- Analytic Applications
- File Extracts (for Data Mining/Stat. Tools)
- BI Tools and User Environments
- Data Quality Feedback Mechanism/Loop

**Consumers:**
- Knowledge Workers
- Managers and Executives
- External Customers and Systems
- Internal Customers and Systems
- Data Professionals
- Other IT Professionals

**Metrics:**
- Usage Metrics
- Customer/User Satisfaction
- Subject Area Coverage %s
- Response/Performance Metrics

**Activities: (P) – Planning (C) – Control (D) – Development (O) - Operational**



From *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

# Defining Data Warehousing, BI/Analytics

- Data Warehousing
  - A technology solution supporting … business capabilities such as: query, analysis, reporting and development of these capabilities
  - Analysis of information not previously integrated
  - Another, often new, set of organizational capabilities

- Business Intelligence (aka. decision support)
  - Dates at least to 1958
  - Support better business decision making
  - Technologies, applications and practices for the collection, integration, analysis, and presentation of business information
  - Understanding historical patterns in data to improve future performance
  - Use of mathematics in business
- Analytics (aka.) enterprise decision management, marketing analytics, predictive science, strategy science, credit risk analysis. fraud analytics - often based on computational modeling
- Reframing the question …
  - From: what data warehouse should we build?
  - To: how can data warehouse-based integration address challenges?

# Indiana Jones: Raiders Of The Lost Ark



# Organizing the Wheat Separated from the Chaff

- Better organized data increases in value
- Poor data management practices are costing organizations money/time/effort
- 80% of organizational data is **ROT**
  - **R**edundant
  - **O**bsolete
  - **T**rivial
- The question is which data to eliminate?
  - Most enterprise data is never analyzed

## Health Care Provider Data Warehouse

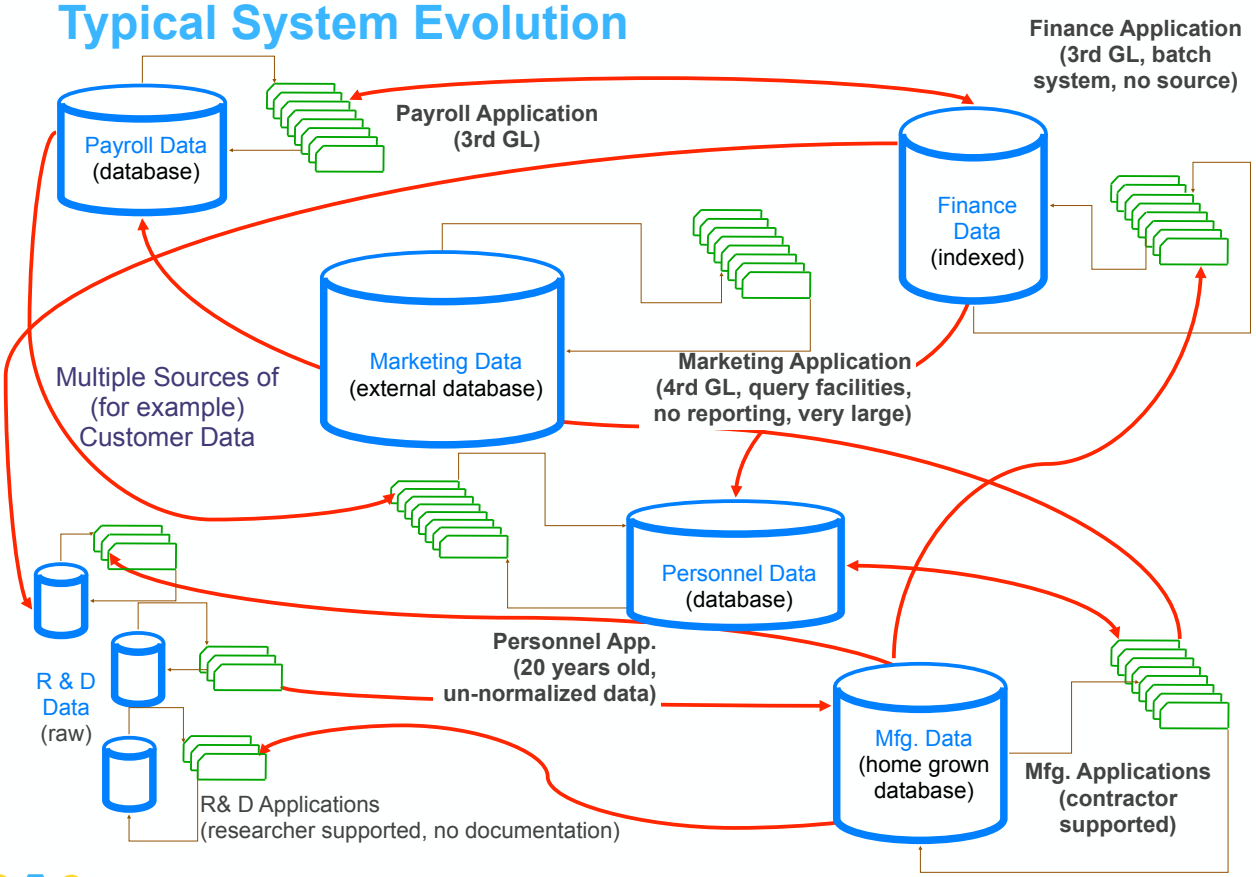
- 1.8 million members
- 1.4 million providers
- 800,000 providers no key
- 29% prov_ssn ≠ 9 digits
- 2.2% prov_number = 9 digits (required)
- 1 User
- $30 million


| | | |
|---|---|---|
| ,335 | $ | 4,812 |
| ,280 | $ | 55,445 |
| ,130 | $ | (118,059) |
| ,016 | $ | (18,016) |
| ,238 | $ | (70,786) |
| 4.39 | $ | (12.90) |
| per Visit | | |

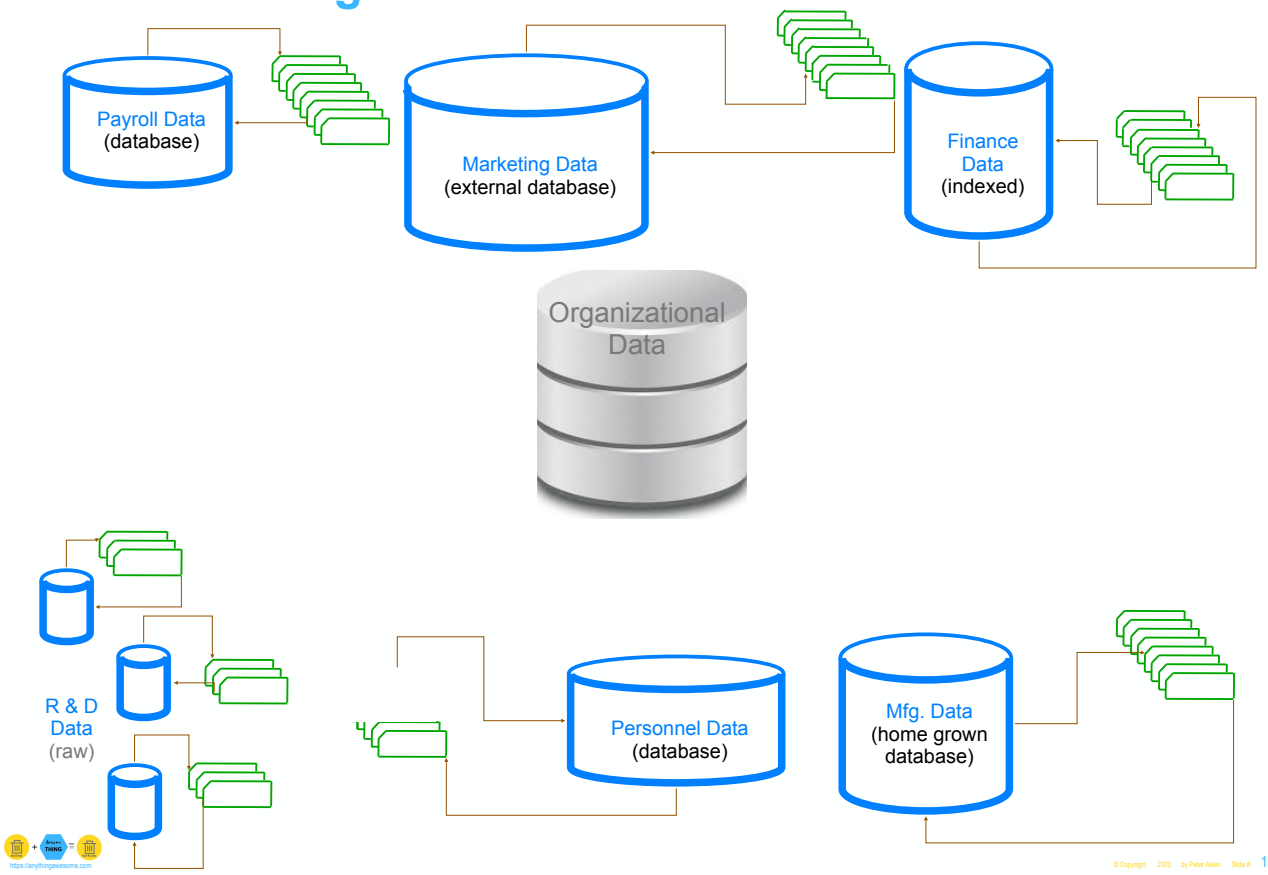
"A roomful of MBAs can accomplish this analysis faster!"



## Typical System Evolution

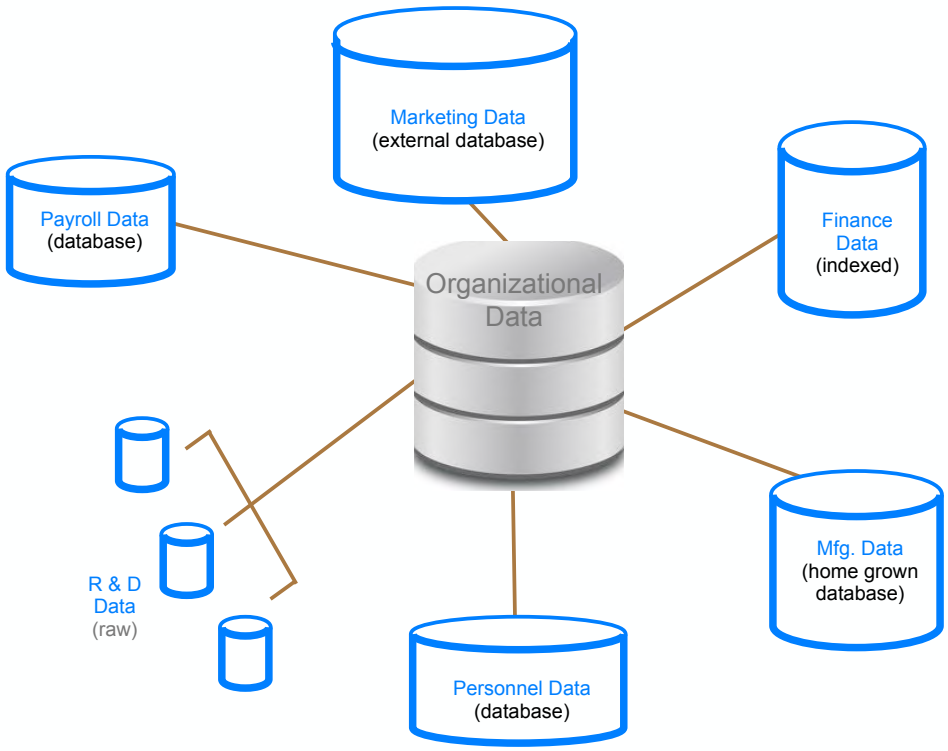


# ... Then Integrate

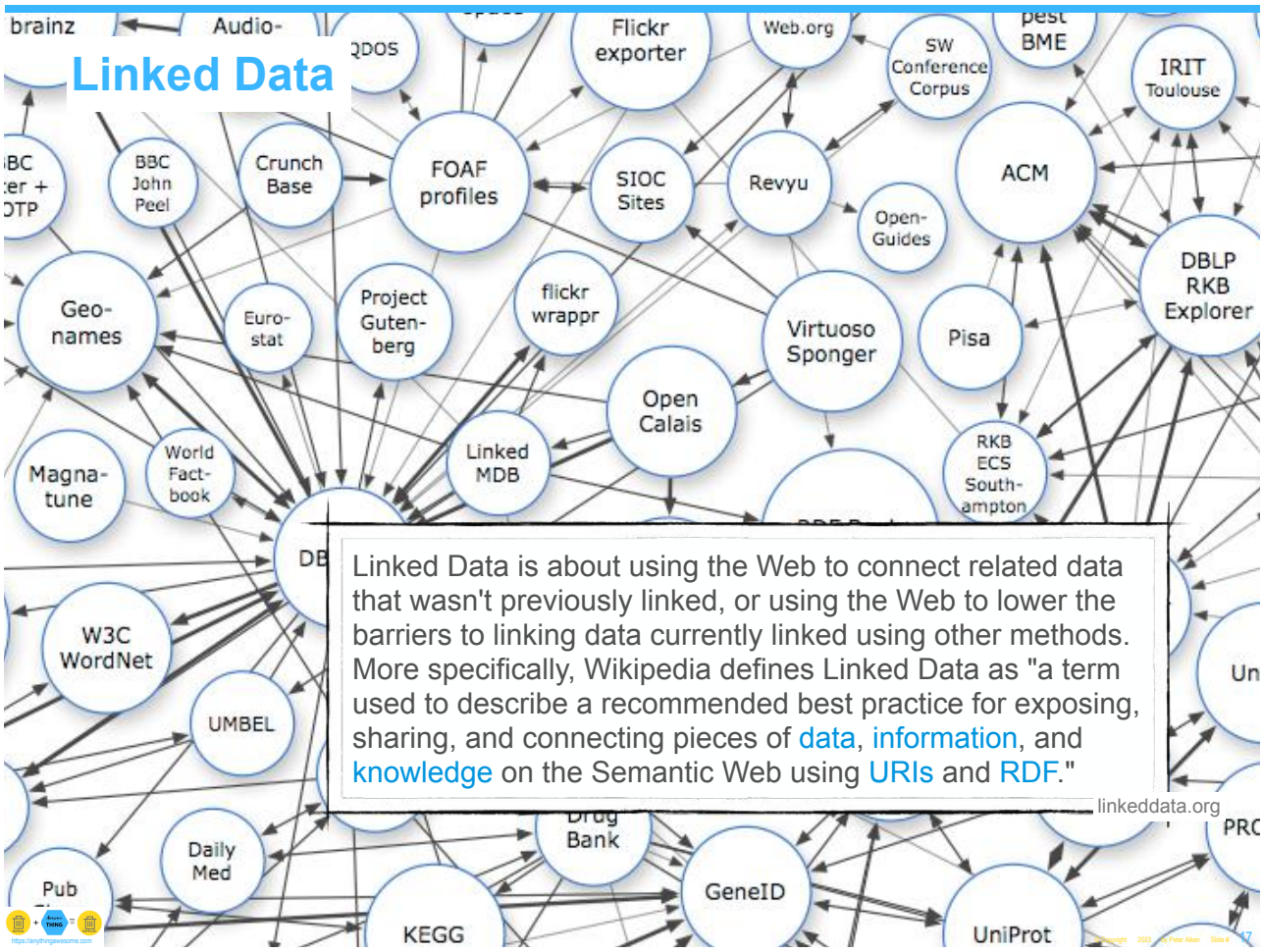


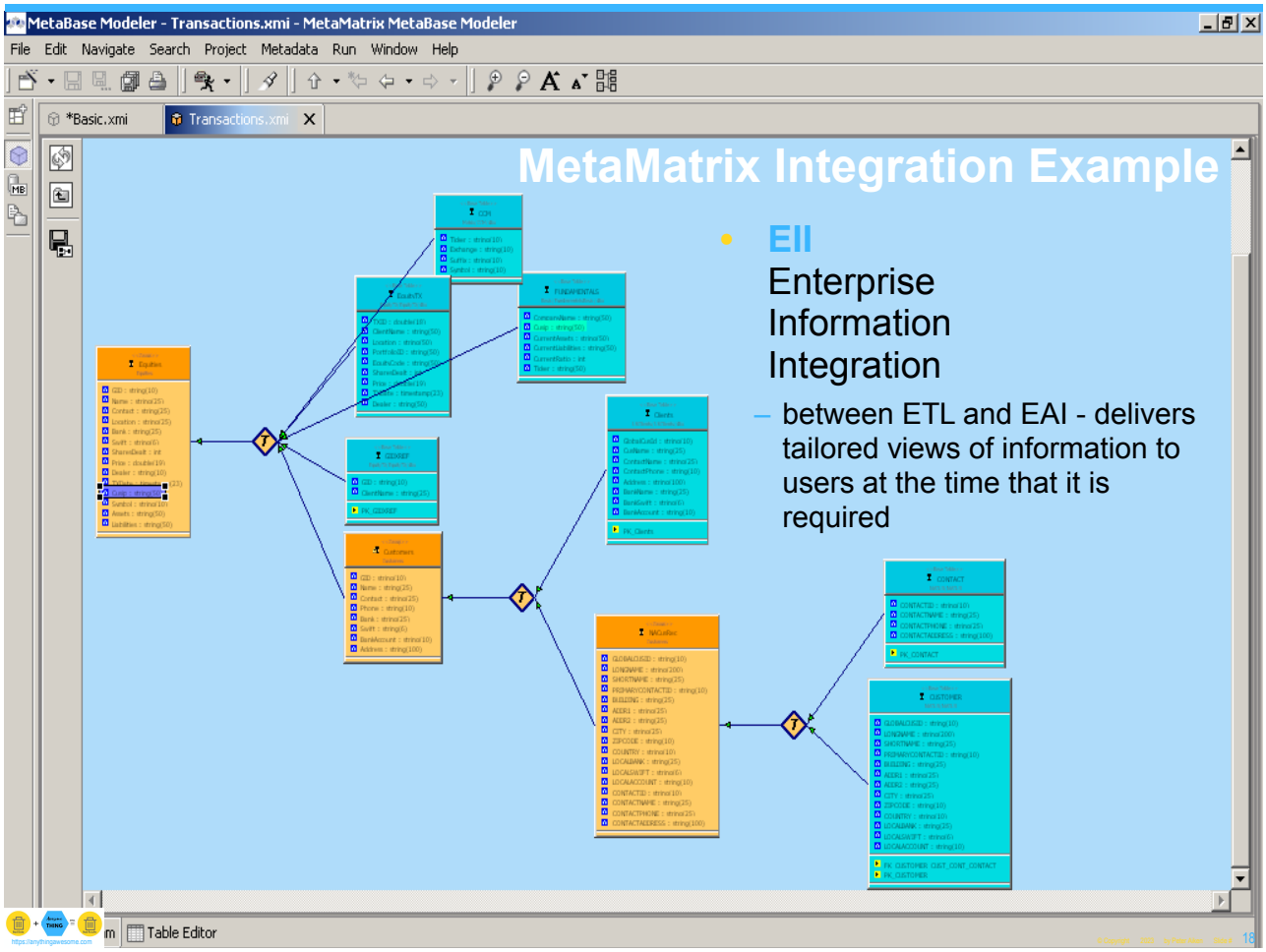

 

# ... Then Re-architect

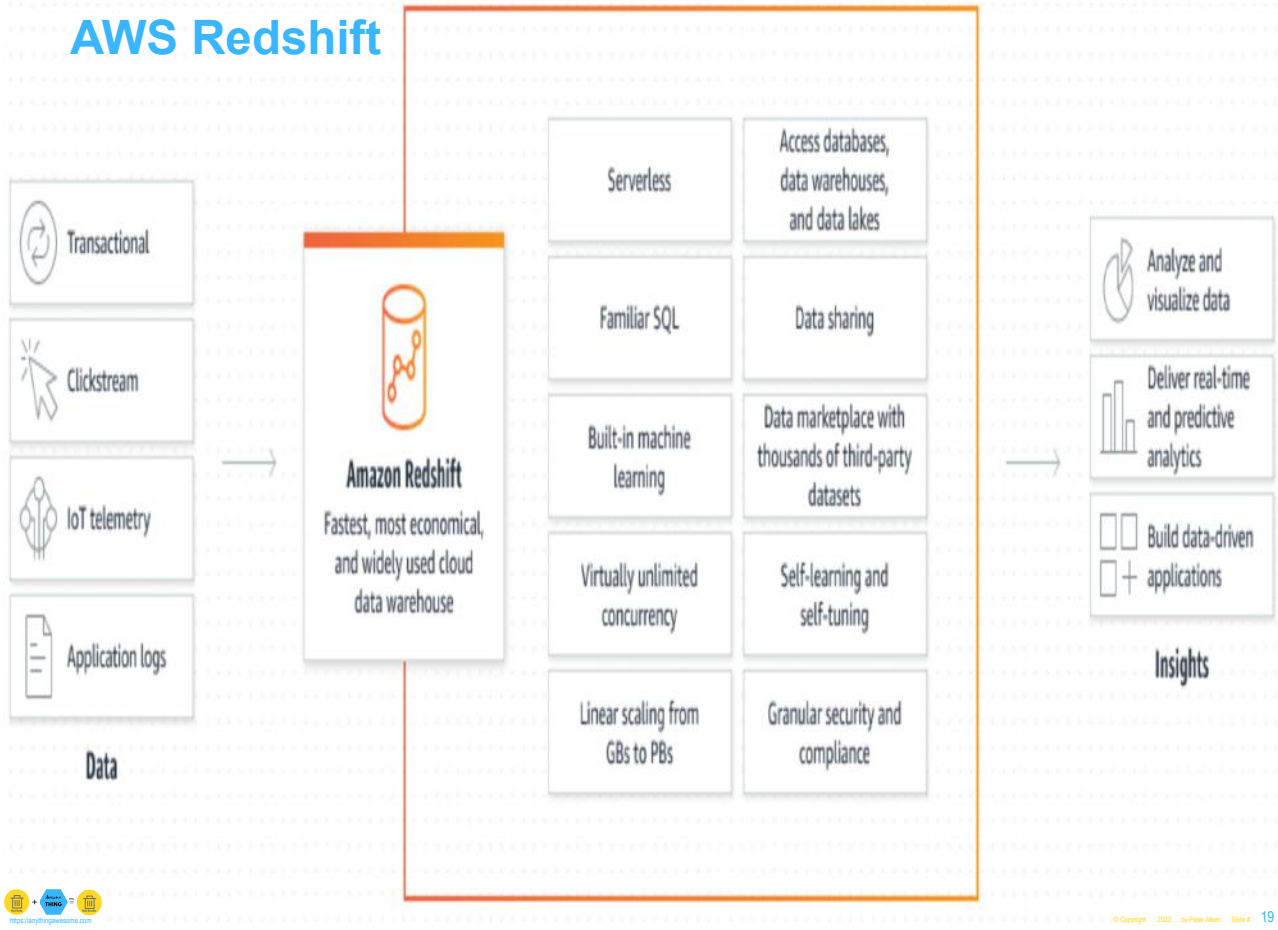


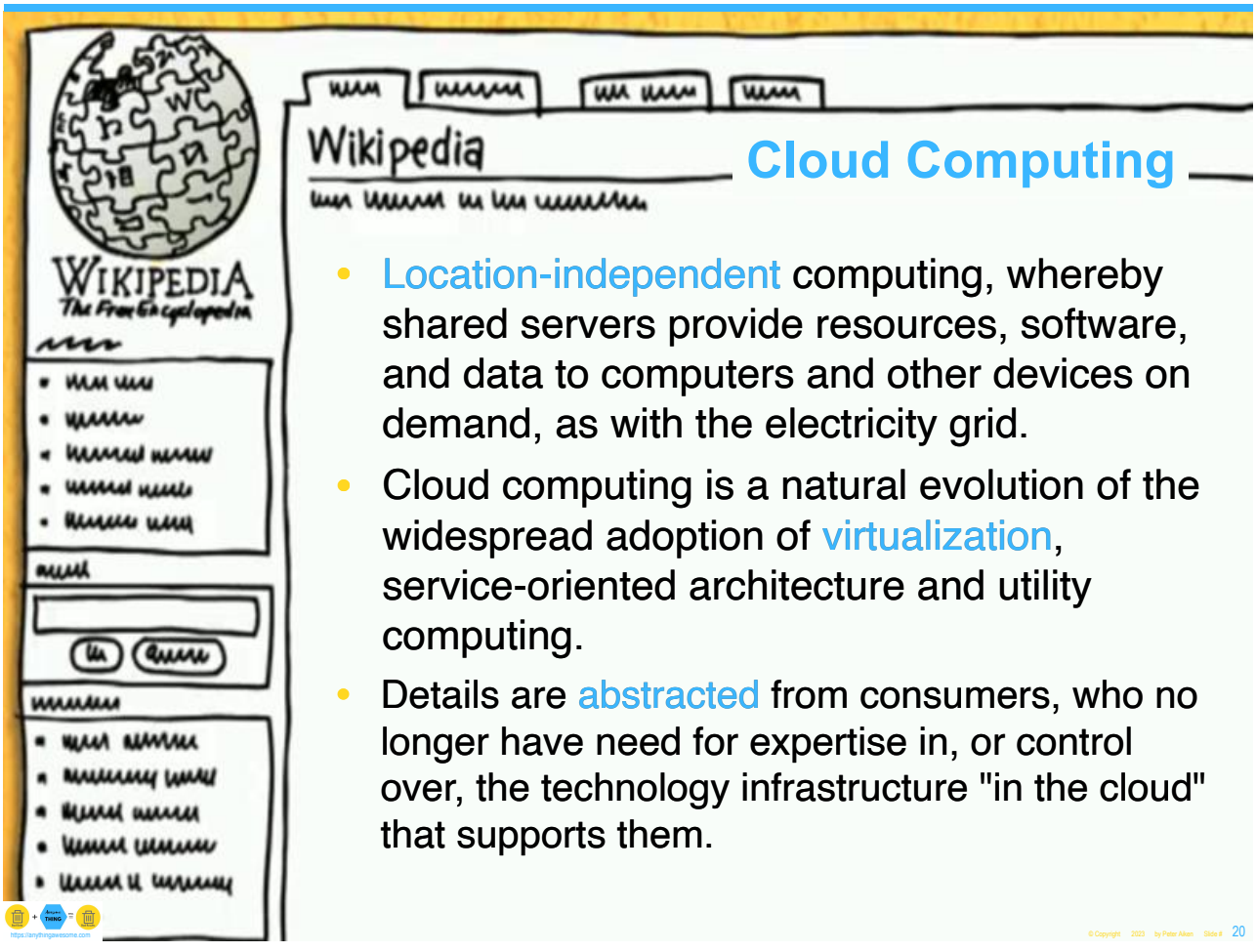

Linked Data
brainz
Audio-
Flickr exporter
Web.org
SW Conference Corpus
IRIT Toulouse
BBC John Peel
Crunch Base
FOAF profiles
SIOC Sites
Revyu
ACM
Open-Guides
DBLP RKB Explorer
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Virtuoso Sponger
Pisa
Open Calais
Linked MDB
RKB ECS South-ampton
Magna-tune
World Fact-book
W3C WordNet
UMBEL
Daily Med
Drug Bank
GeneID
KEGG
UniProt
Pub
Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF."
linkeddata.org


MetaBase Modeler - Transactions.xmi - MetaMatrix MetaBase Modeler
File Edit Navigate Search Project Metadata Run Window Help
*Basic.xmi
Transactions.xmi
MetaMatrix Integration Example
EII
Enterprise Information Integration
between ETL and EAI - delivers tailored views of information to users at the time that it is required
Table Editor

# AWS Redshift



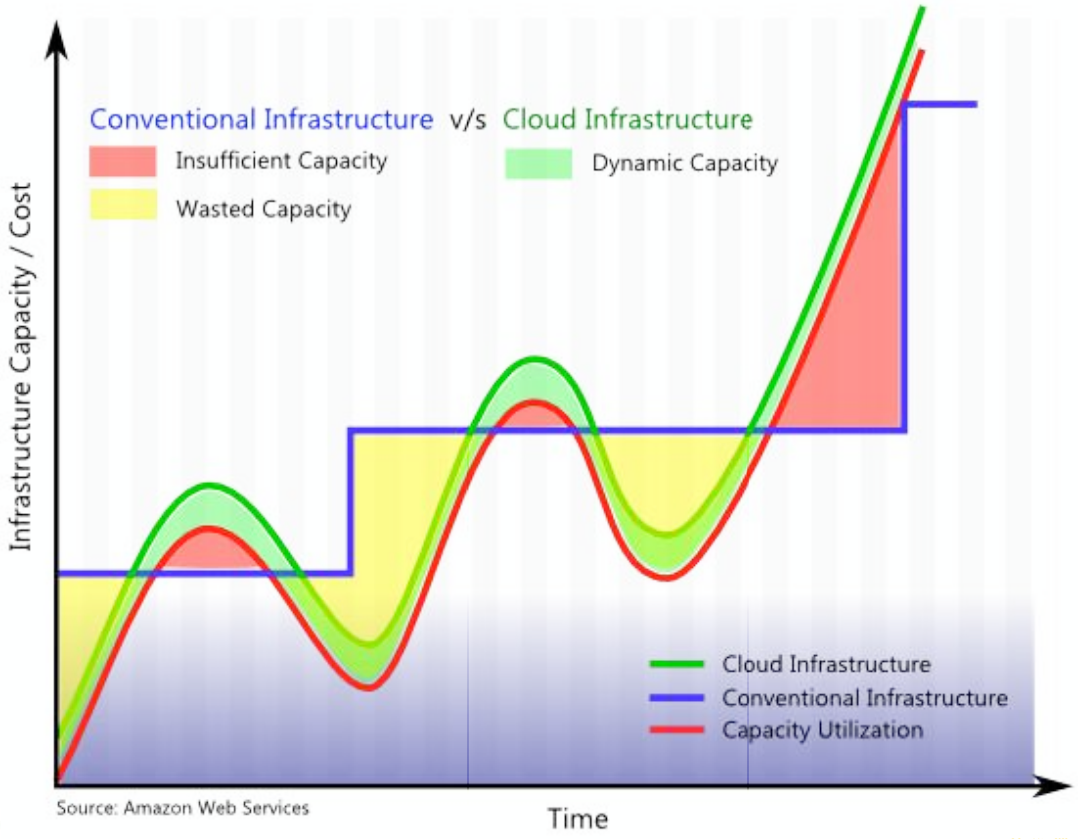



# Cloud Computing

- Location-independent computing, whereby shared servers provide resources, software, and data to computers and other devices on demand, as with the electricity grid.
- Cloud computing is a natural evolution of the widespread adoption of virtualization, service-oriented architecture and utility computing.
- Details are abstracted from consumers, who no longer have need for expertise in, or control over, the technology infrastructure "in the cloud" that supports them.

## Cloud Scalability






## Cisco's Ladder to the Cloud

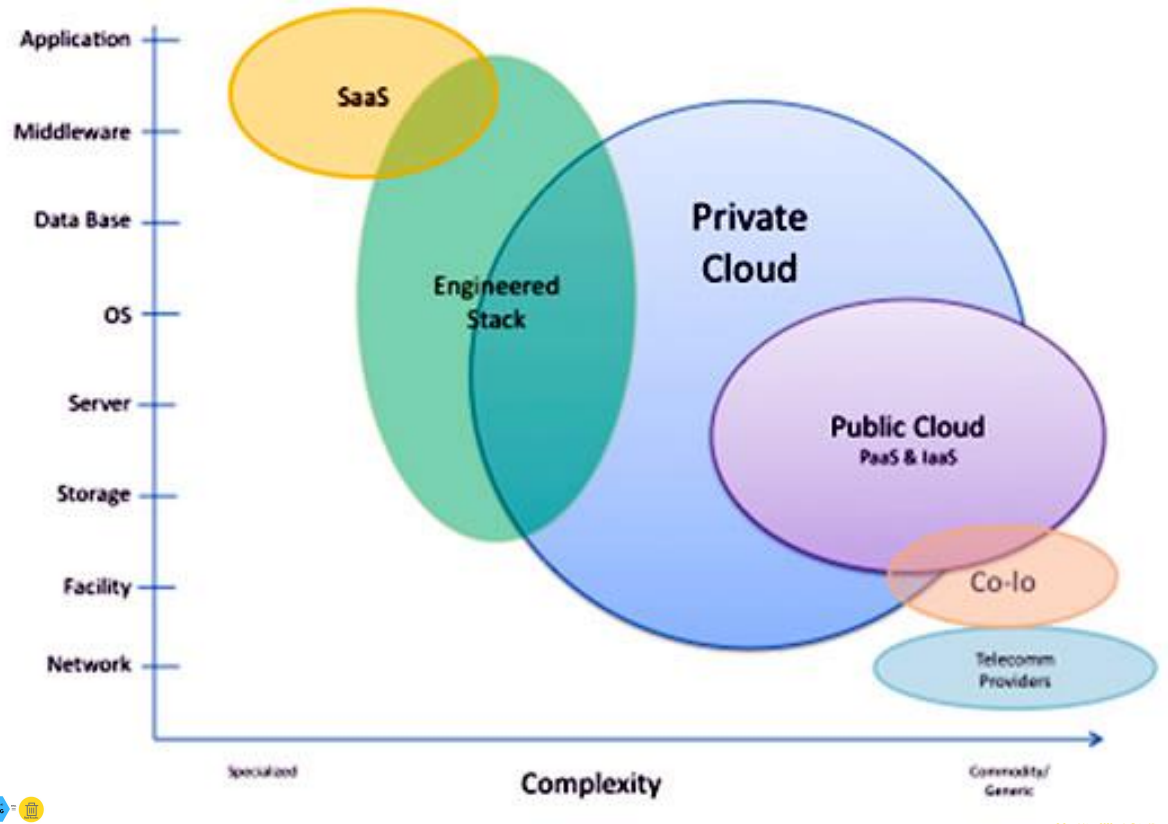
**Public Cloud**
No capital expense, extreme scalability

**Private Cloud**
Automated provisioning, chargeback

**Virtual Data Center**
Less downtime, lab management, automatic resource allocation

**Consolidated Data Center**
Fewer machines, faster provisioning, virtual disk recovery

## Cloud Options






## Data in the cloud should have three attributes that data outside the cloud/warehouse should not have. It should be:

Cleaner

Smaller

Sharable-er

SHAREABLE

# Making ~~Warehousing~~ Successful

Problems with forklifting

1. no basis for decisions made
2. no inclusion of architecture/ engineering concepts
3. no idea that these concepts are missing from the process
4. 80% of organizational data is ROT

Transform

Less

Cleaner

More shareable ... data

**Data Branding**



# Fixing Data in the Cloud is Like Using A Glovebox

# Program Overview

- Definitions
  - Data Warehousing
  - DM BoK guidance
  - Legacy to digital conversion (cloud)
- Integration
  - Emphasize engineering talent
  - Incorporate leading not bleeding edge
  - Requires an adaptive rather than a prescriptive approach
- Preparation
  - Emphasis on storytelling first and visualization second
  - Analytics is both ubiquitous and not well understood
  - Keep improvements practically focused by strategy
- Best Practices
  - Cannot use what is not understood-understand what you have
  - PDCA (plan, do, check, act)
  - Iteratively refine/cull with respect to strategic direction
- Take Aways/References/Q&A

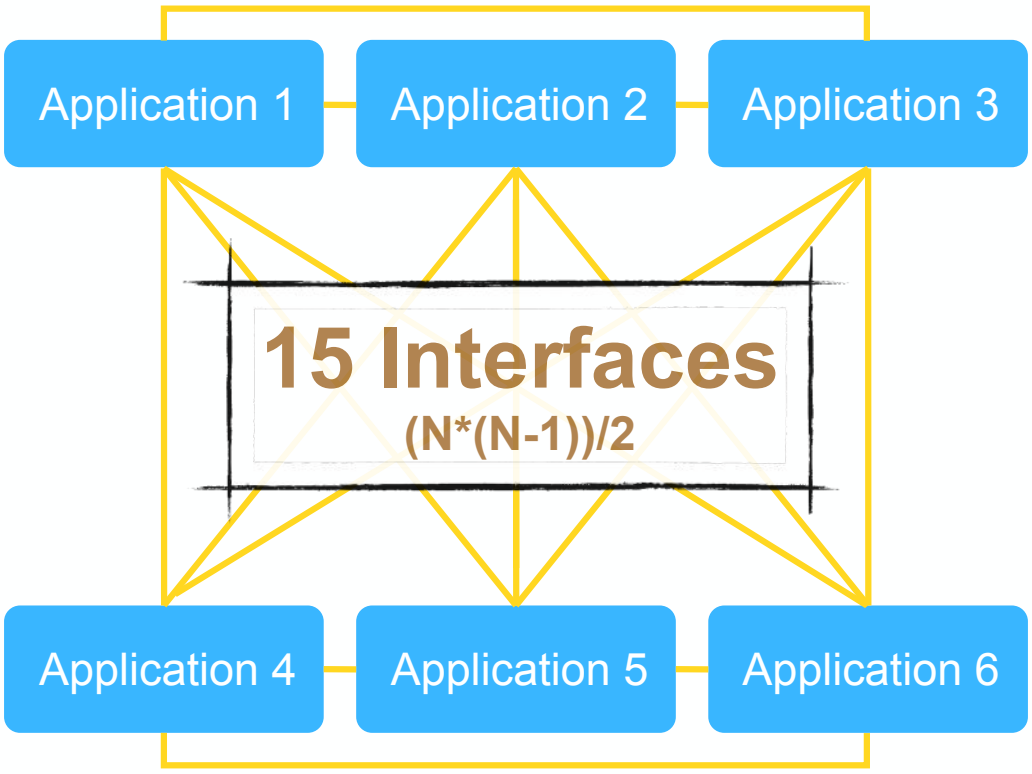

**Of What Do Your Data Warehousing Operations Consist?**




# Two Basic Warehouse Purposes

- Integration
  - of disparate data sources for purpose of potential subsequent analyses
  - Most organization data is never analyzed
  - Same type of inputs as output
  - Downstream knowledge is incorporated upstream
- Preparation
  - Seen as the last mile (preparation) of the data before presentation as part of decision making activities
  - Closed ended activities
  - Final possible application of programmatic quality measures


## How many interfaces are required to solve this integration problem?



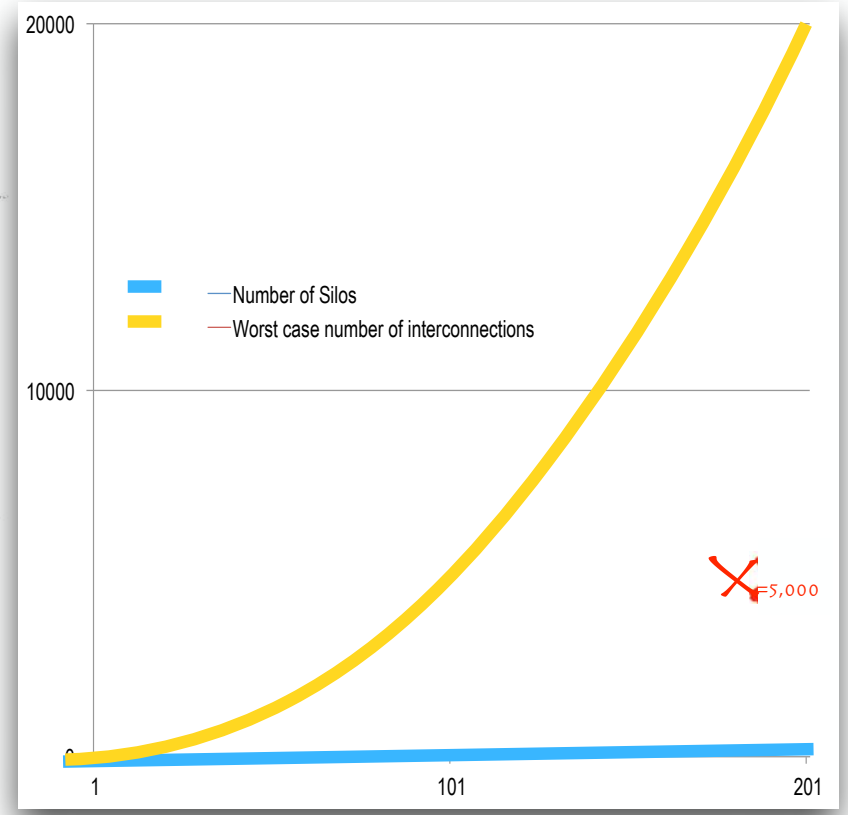

RBC: 200 applications - 4900 batch interfaces



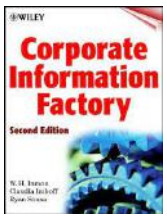
## The rapidly increasing cost of complexity

**Silos / Interconnections**

- 6 / 15
- 60 / 1,770
- 600 / 179,700
- 200 / 19,900
- 200 / 5,000 (actual)



# Corporate Information Factory Architecture



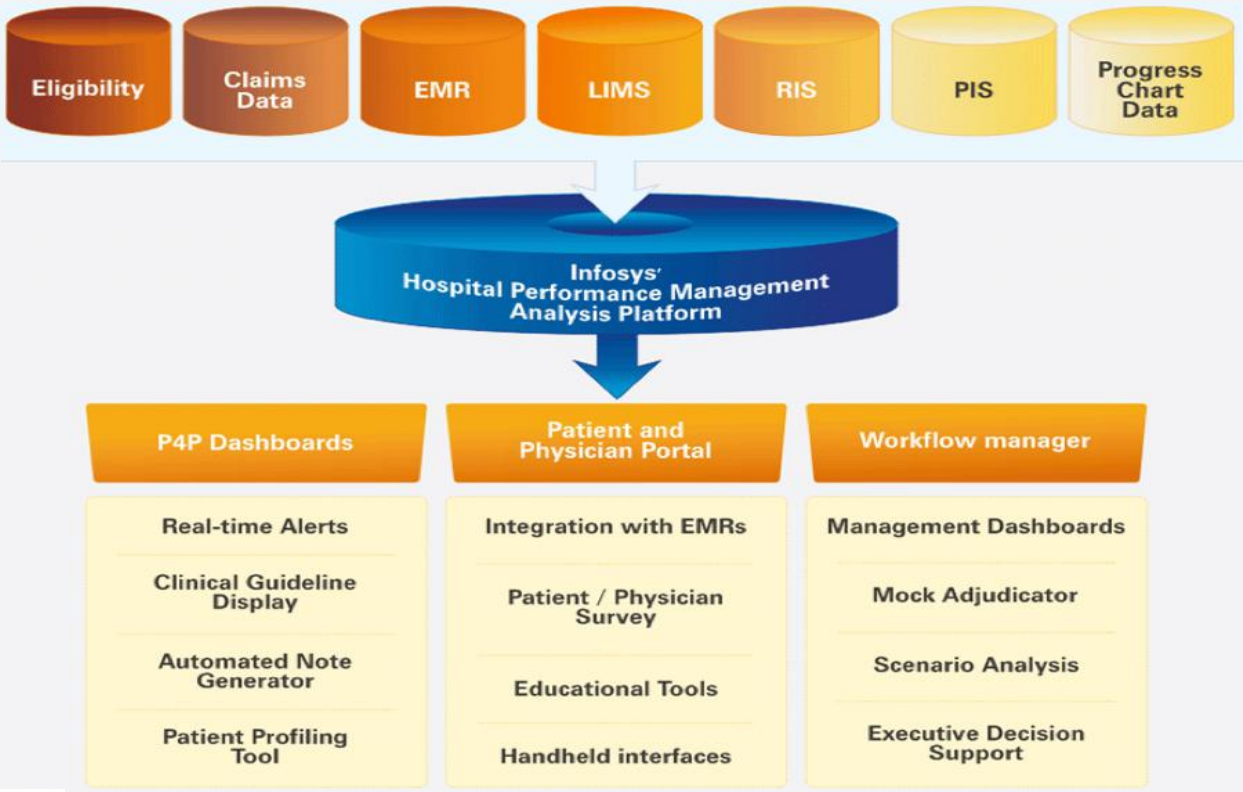




from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

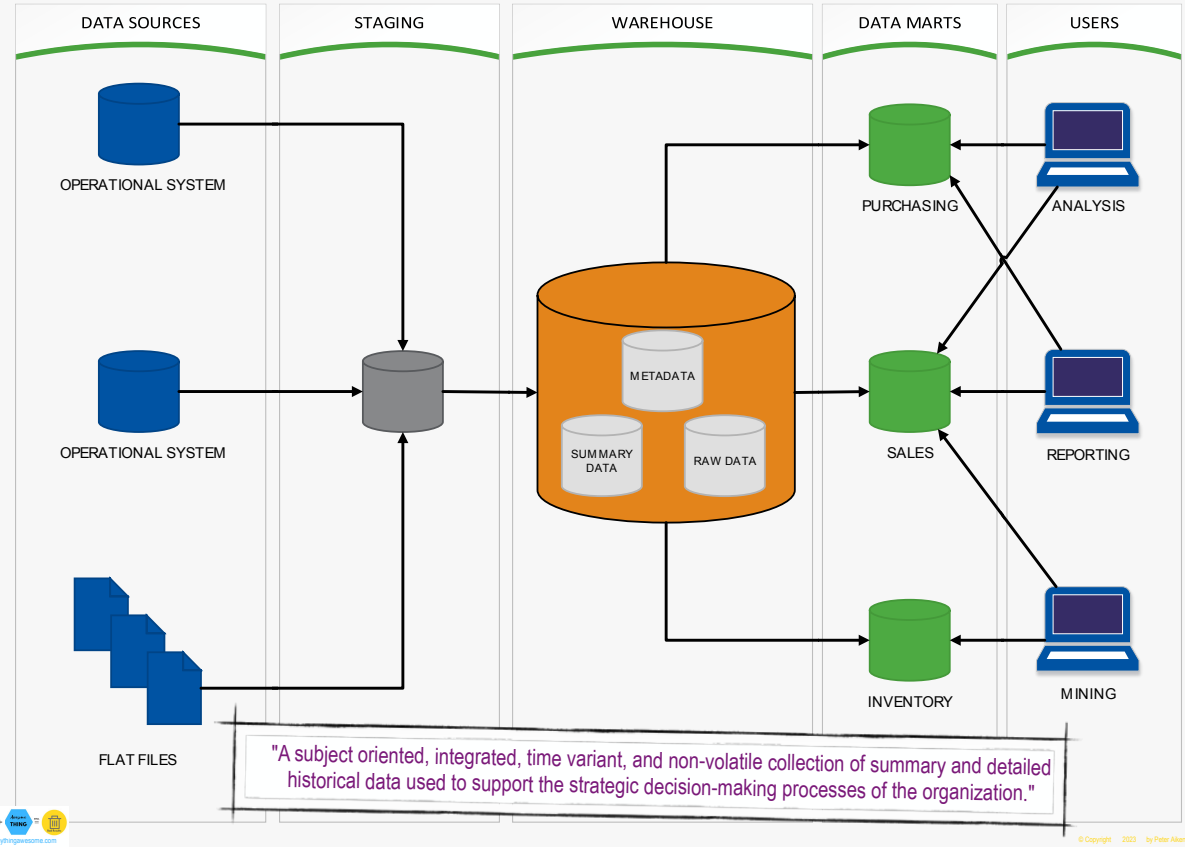


# Warehousing applied to a specific challenge




*http://www.infosys.com/industries/healthcare/industryofferings/Pages/healthcare-data-warehousing.aspx*

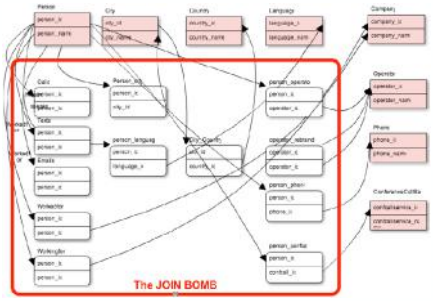
# Inmon Implementation/3NF



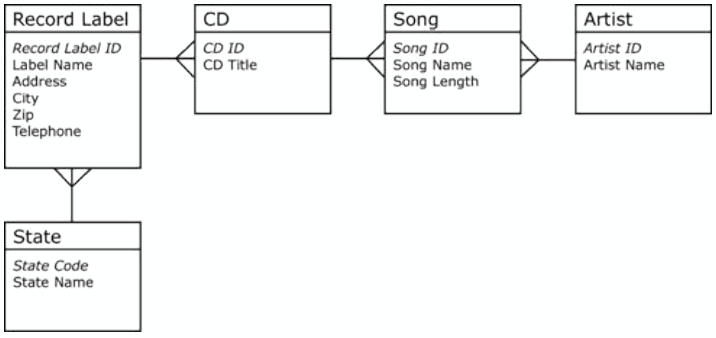



# Third Normal Form

- Each attribute in the relationship is a fact about a key
  - Highly normalized structure




- Not much true expertise in IT
  - Little understanding of where functionality should reside
  - Concepts taught unevenly



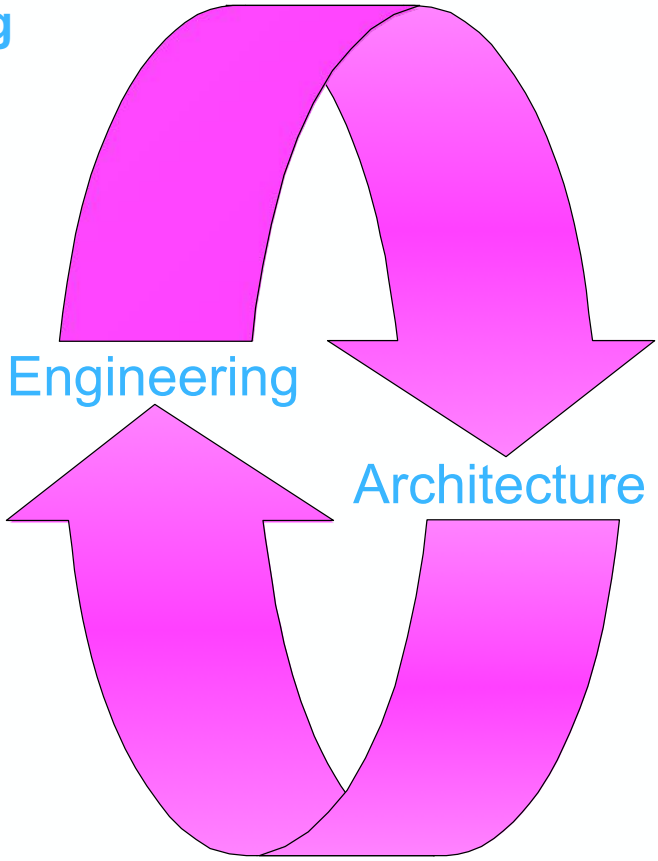
# Third Normal Form: Pros and Cons

- Pros
  - Easily understood by business and end users
  - Reduced data redundancy
  - Enforced referential integrity
  - Indexed attributes/flexible querying
- Cons
  - Joins can be expensive
  - Does not scale

# Engineering/Architecting Relationship

- Architecting is used to create and build systems too complex to be treated by engineering analysis alone
  - Require technical details as the exception
- Engineers develop the technical designs
  - Engineering/Crafts-persons deliver components supervised by:
    - Manufacturer
    - Building Contractor



# Attracting and Retaining Engineering Talent

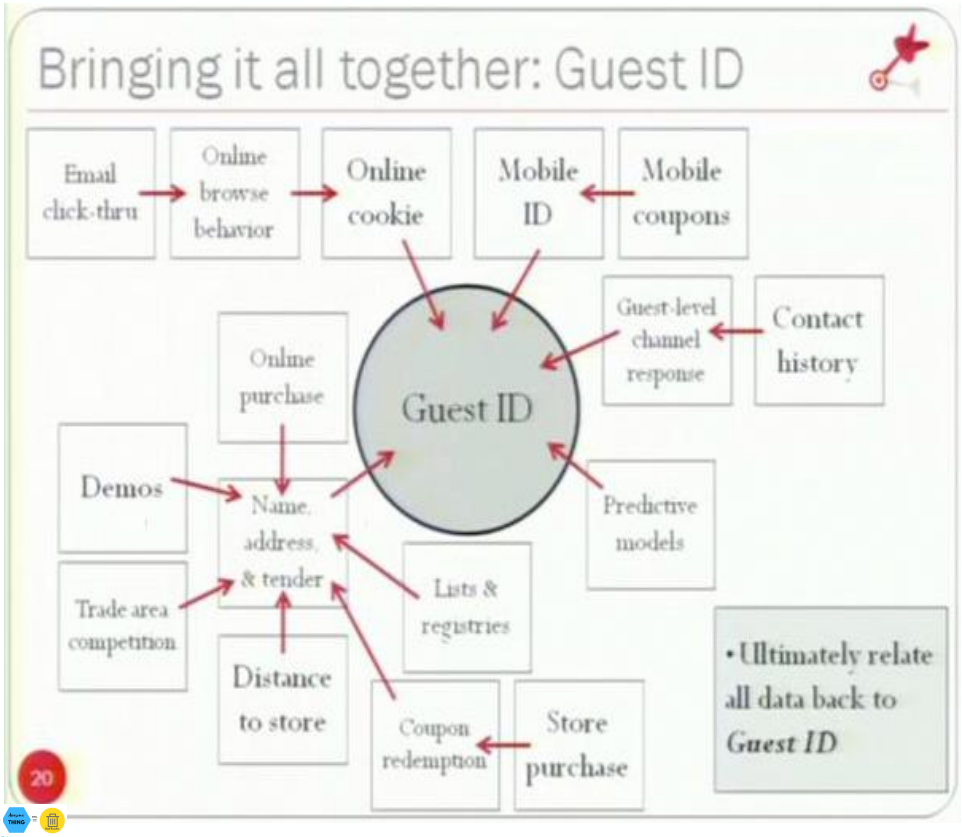
## Emphasize Engineering Talent

- Solution design is not based on semantic understanding
- Focus is often critical due to speed issues
- AI/ML focus on learning how the system performs and how to improve it

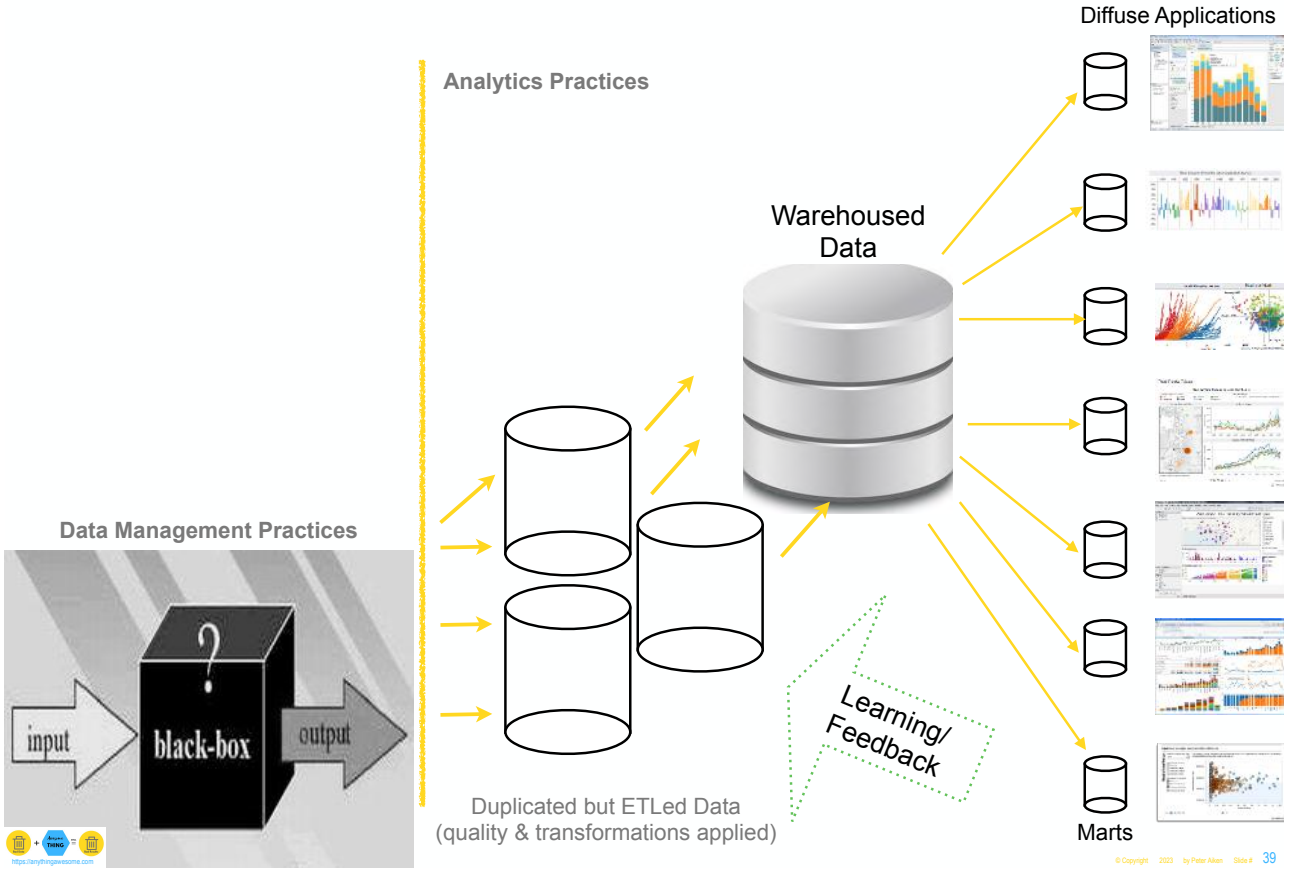


http://rmportal.performedia.com/node/1373 and http://www.predictiveanalyticsworld.com/patimes/target-really-predict-teens-pregnancy-inside-story/ http://rmportal.performedia.com/rm/paw10/gallery_01#1373

## Target Isn't Just Predicting Pregnancies




# Where To Focus Limited Resources?






# Mission

- *Advance change in America by ensuring equitable access to nutritious food for all in partnership with food banks, policymakers, supporters, and the communities we serve*
- Great understanding and skill executing rapid response to local needs

- What else could be done with their data?
- Rerouting busses = eliminating food deserts for thousands of residents


# Program Overview

- Definitions
  - Data Warehousing
  - DM BoK guidance
  - Legacy to digital conversion (cloud)
- Integration
  - Emphasize engineering talent
  - Incorporate leading not bleeding edge
  - Requires an adaptive rather than a prescriptive approach
- Preparation
  - Emphasis on storytelling first and visualization second
  - Analytics is both ubiquitous and not well understood
  - Keep improvements practically focused by strategy
- Best Practices
  - Cannot use what is not understood-understand what you have
  - PDCA (plan, do, check, act)
  - Iteratively refine/cull with respect to strategic direction
- Take Aways/References/Q&A


**Of What Do Your Data Warehousing Operations Consist?**



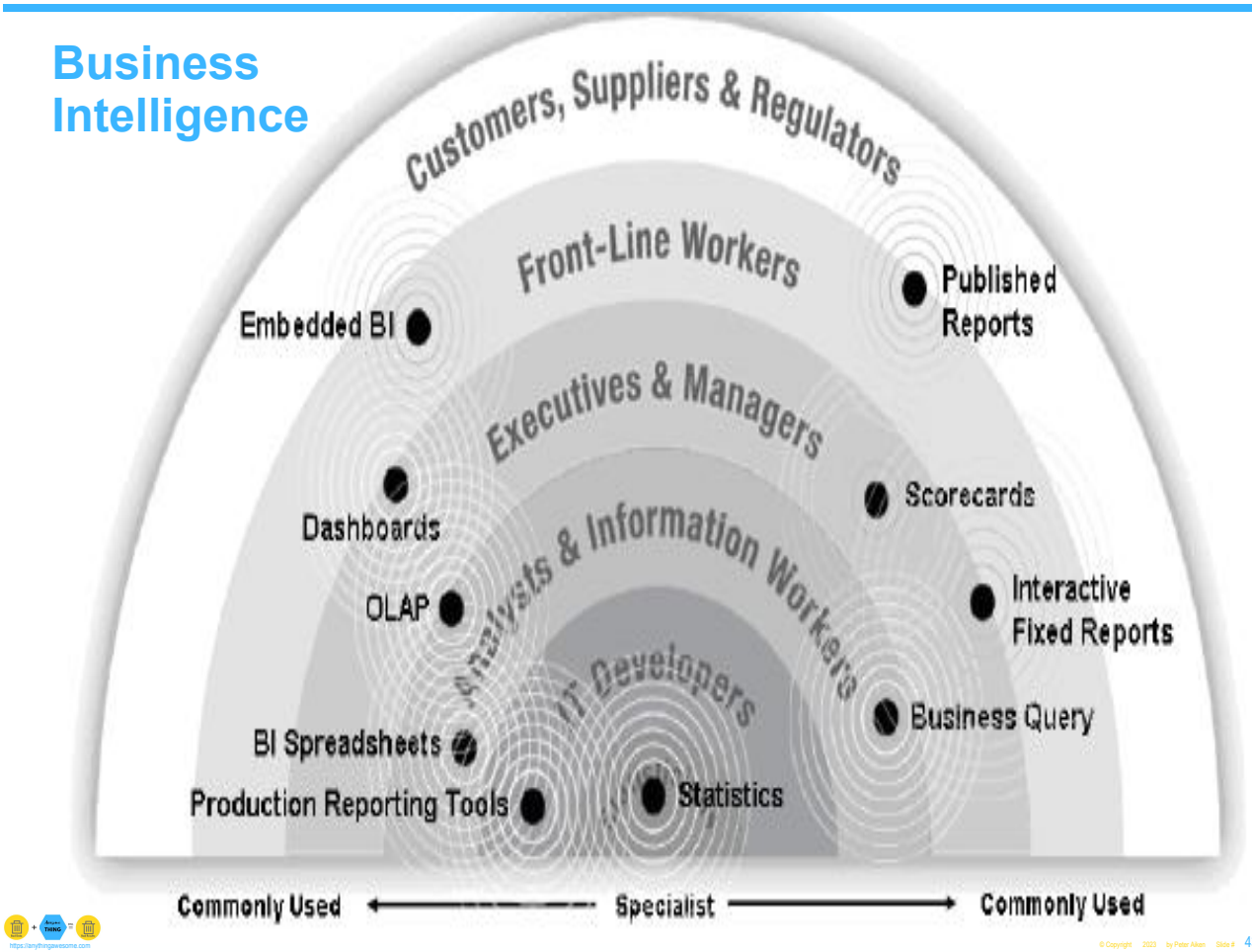

# Program Overview

- Definitions
  - Data Warehousing
  - DM BoK guidance
  - Legacy to digital conversion (cloud)
- Integration
  - Emphasize engineering talent
  - Incorporate leading not bleeding edge
  - Requires an adaptive rather than a prescriptive approach
- Preparation
  - Emphasis on storytelling first and visualization second
  - Analytics is both ubiquitous and not well understood
  - Keep improvements practically focused by strategy
- Best Practices
  - Cannot use what is not understood-understand what you have
  - PDCA (plan, do, check, act)
  - Iteratively refine/cull with respect to strategic direction
- Take Aways/References/Q&A

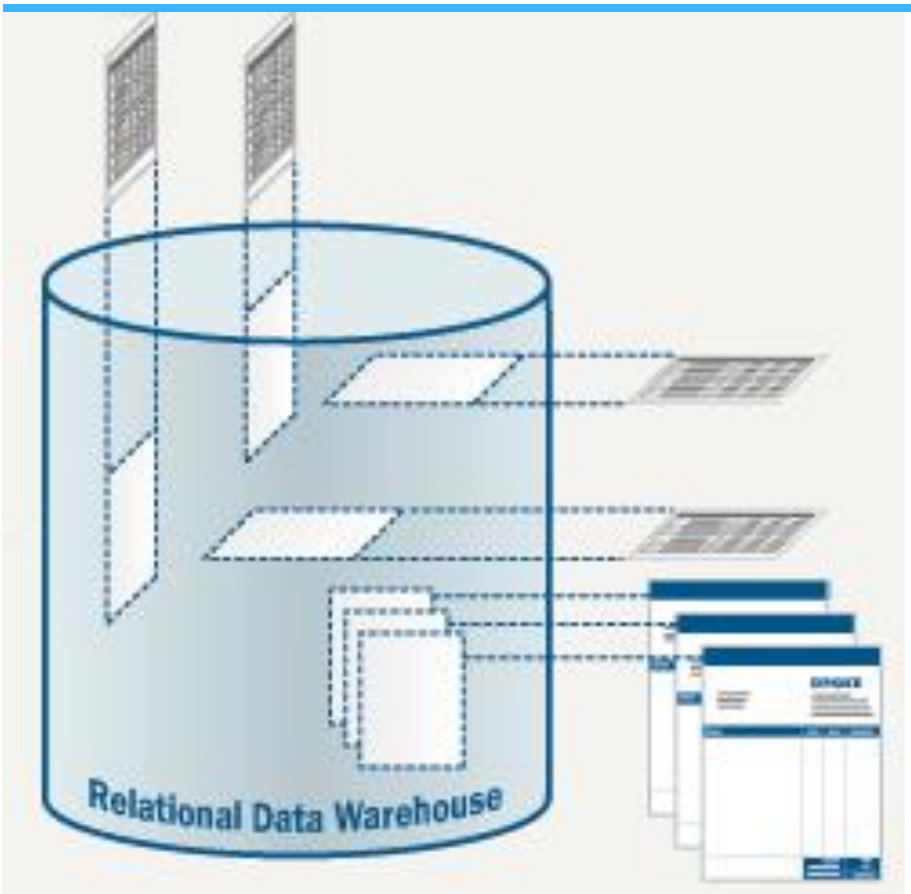

**Of What Do Your Data Warehousing Operations Consist?**


© Copyright 2023 by Peter Aiken Slide # 42

# Business Intelligence



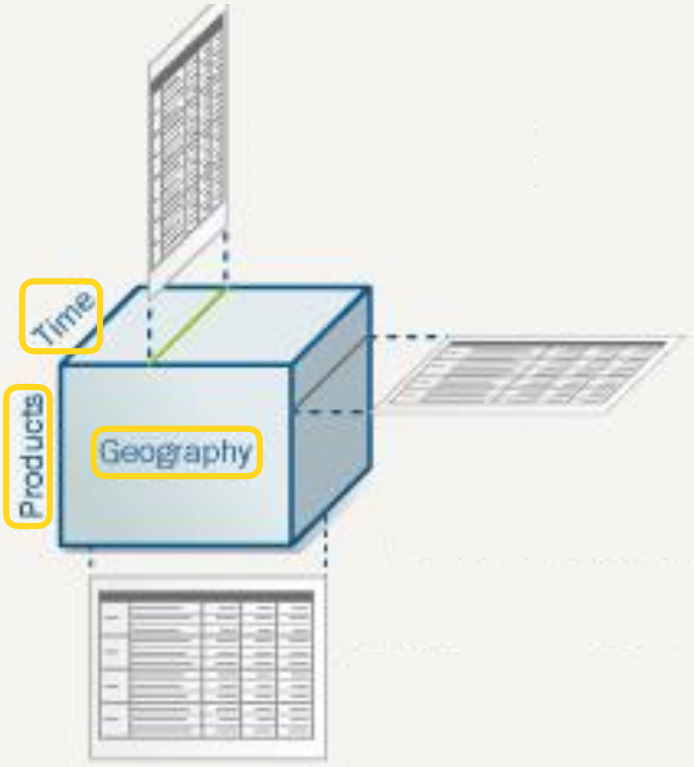



## Basics

- Users can "drill" anywhere
- Entire collection "cube" is accessible
- Summaries to transaction-level detail

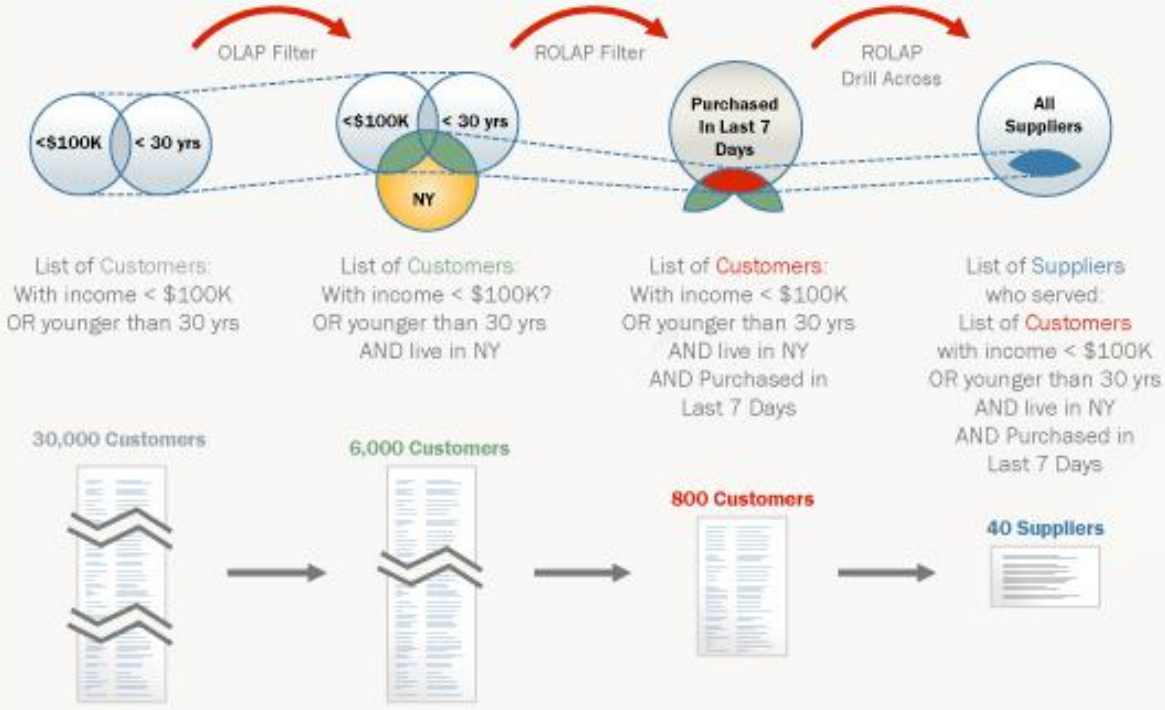
## Sample questions …



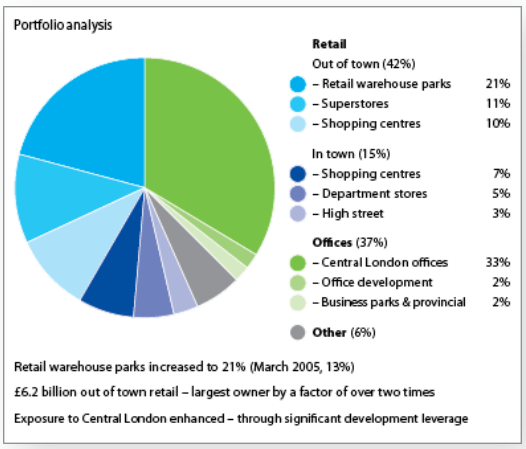

- Emphasis on the "cube"
- 'N' dimensions
- Permits different users to "slice and dice" subsets of data
- Viewing from different perspectives



## Example: Set Analysis



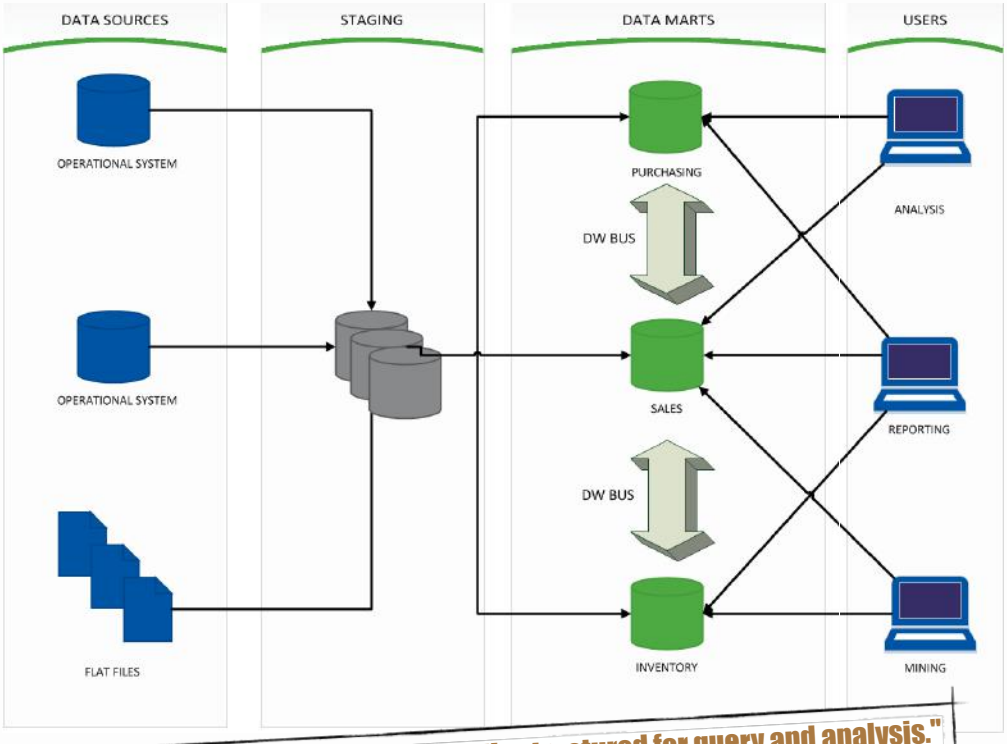
# Portfolio Analysis




- Bank accounts are of varying value and risk
- Cube by
  - Social status
  - Geographical location
  - Net value, etc.
- Strategy: balance return on the loan with risk of default
- How to evaluate the portfolio as a whole?
  - Least risk loan may be to the very wealthy, but there are a very limited number
  - Many poor customers, but greater risk
- Solution may combine types of analyses
  - When to lend?
  - Interest rate charged?

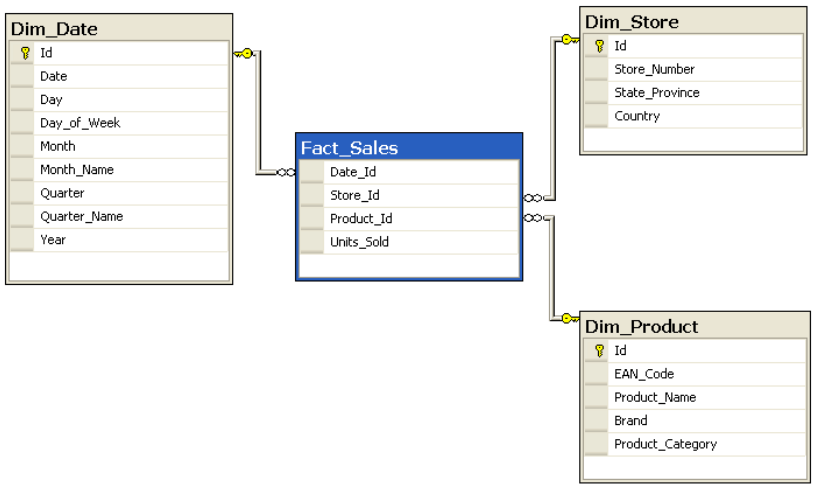


# Kimball Implementation/Dimensional




"A copy of transaction data specifically structured for query and analysis."

# Star Schema




- Comprised of “fact tables” that contain quantitative data, and any number of adjoining “dimension” tables
- Optimized for business reporting
- Use Cases
  - OLAP (Online Analytic Processing)
  - BI



# Star Schema Pros and Cons

- Pros
  - Simple Design
  - Fast Queries
  - Most major DBMS are optimized for Star Schema Designs
- Cons
  - Questions must be built into the design
  - Data marts are often centralized on one fact table

# the architecture of talent

## Emphasize Data Architecture Talent

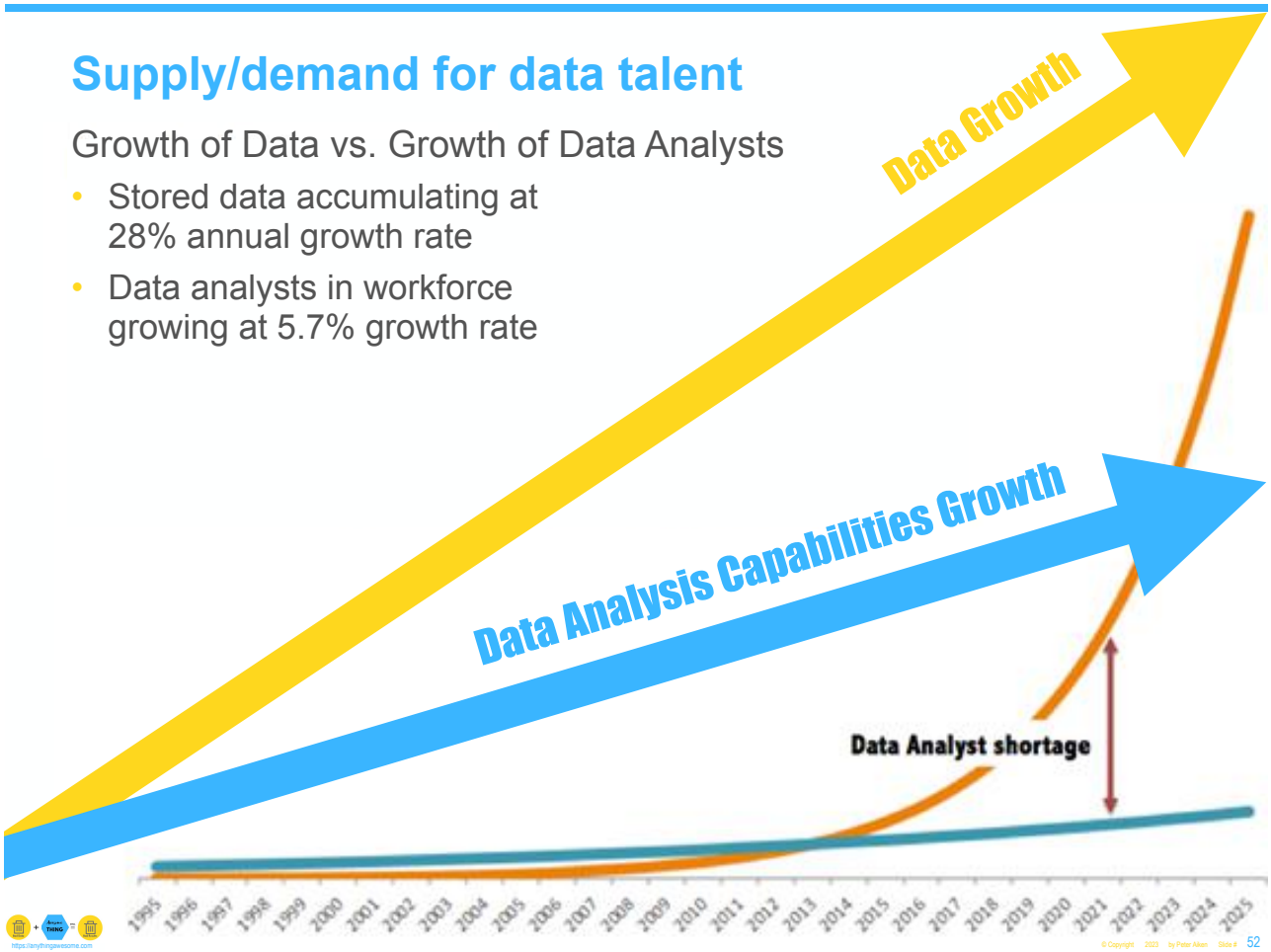
- Solution design is based on semantic understanding
- Broad category ranging from CRM to analytics
- AI/ML focus on data exploration



## Supply/demand for data talent

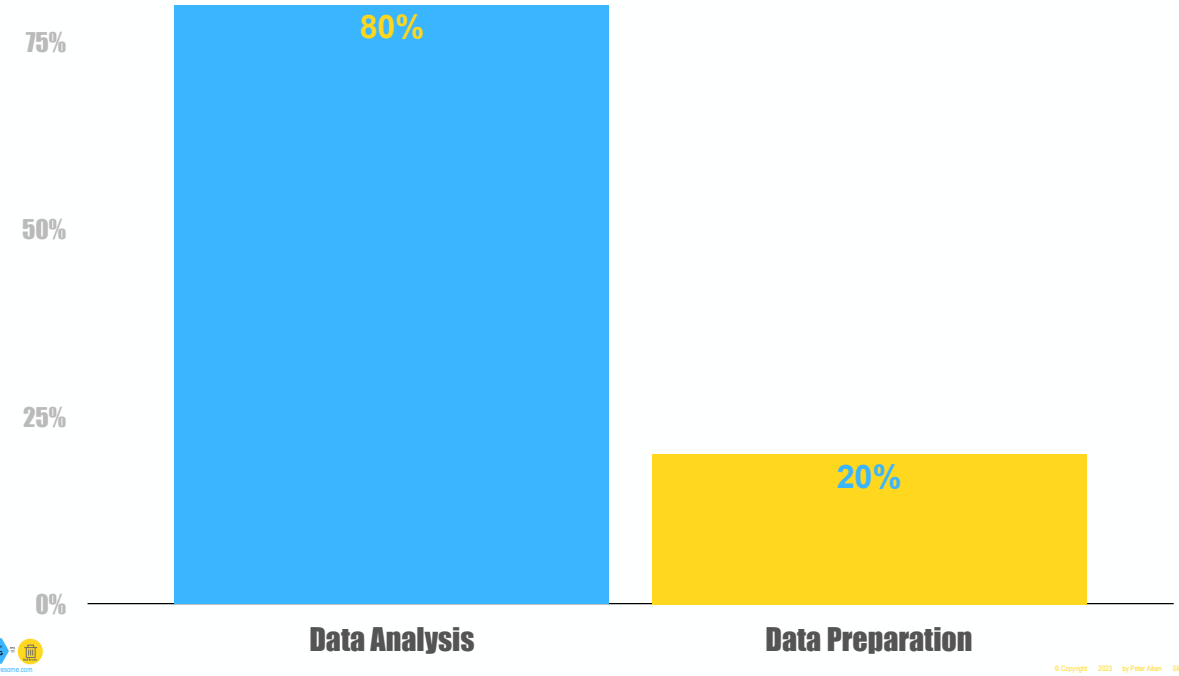
Growth of Data vs. Growth of Data Analysts

- Stored data accumulating at 28% annual growth rate
- Data analysts in workforce growing at 5.7% growth rate



# Everyone wants to do better data analysis …

- Some data preparation is inevitable
- What would a 'good' ratio be?
- "Everyone knows"






Forbes

2020 American Airlines market value ~ $6b
AAdvantage valued between $19.5-$31.5b
2020 United market value ~ $9b
MileagePlus ~ $22b

EDITORS' PICK | Jul 15, 2020, 09:00am EDT

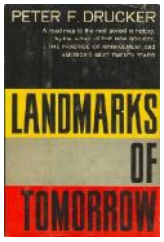
## How Airlines Make Billions From Monetizing Frequent Flyer Programs

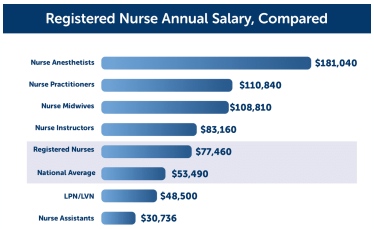
Forbes Advisor

JT Genter Advisor Contributor
Advisor Contributor Group
Personal Finance

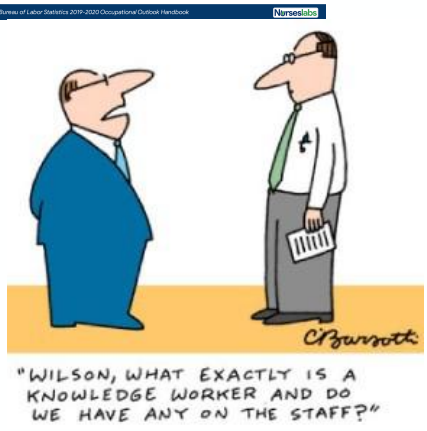
https://www.forbes.com/sites/advisor/2020/07/15/how-airlines-make-billions-from-monetizing-frequent-flyer-programs/?sh=66da87a614e9

# Simple Business Case

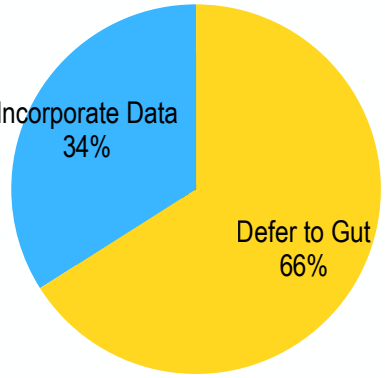

- "Knowledge workers"
  - **The Landmarks of Tomorrow** (1957) by Peter Drucker
  - Examples include programmers, physicians, pharmacists, architects, engineers, scientists, design thinkers, public accountants, lawyers, and academics, and any other white-collar workers, whose line of work requires one to "think for a living".
  - Think about inputs and make decisions based on the data and their processes



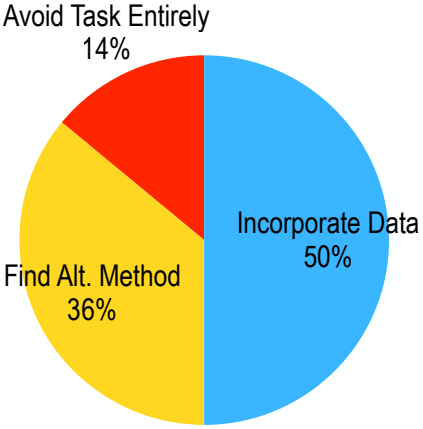



http://TheDataLiteracyProject.org

# When asked to incorporate data


- Data appreciation isn't translating into employee adoption
  - 48% frequently make gut decisions
  - 66% for C-suite executives
- Lack of data skills is limiting workplace productivity
  - 36% said they would find an alternative method to complete the task without using data
  - 14 percent would avoid the task entirely



# Data Literacy Levels

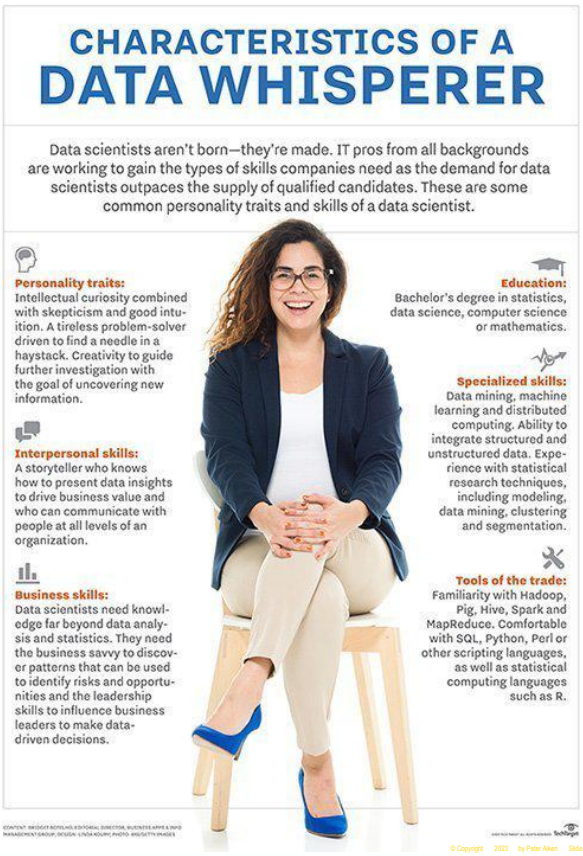
**Everything here is required of data scientists and cyber professionals**

| | | Ethical Perspective | Behavioral Focus |
|---|---|---|---|
| Level 3 | Knowledge worker (1,000,000,000) | Avoid Fooling or Manipulating | Organizational |
| Level 2 | Adult data spreader | Avoid Being Fooled or Manipulated | Individual |
| Level 1 | Mobile data spreader | | |



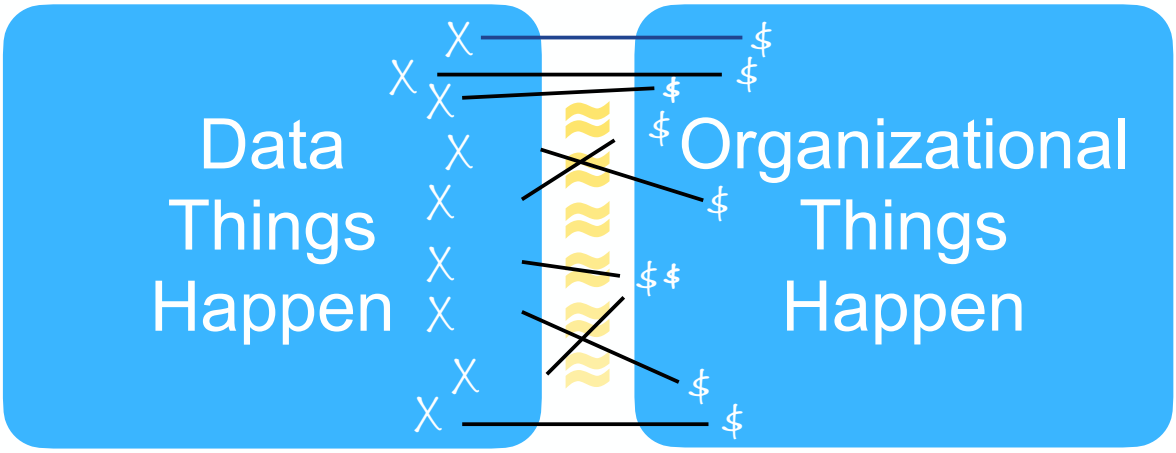

# Level 3–KW–Citizen Data Knowledge Areas

- Elevator story
- Data stewardship
- Demonstrating value
- Currency
- Fiduciary responsibilities
- Shared fate




– https://www.techtarget.com/searchenterpriseai/definition/data-scientist

## More work required to incorporate greater focus





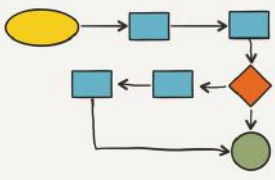

## What is Strategy?


**strat·e·gy**
/ˈstratəjē/

*noun*

1. a plan of action or policy designed to achieve a major or overall aim.
   "time to develop a coherent economic strategy"
   *synonyms:* master plan, grand design, game plan, plan (of action), action plan, policy, program; More


- Current use derived from military
  - **a pattern in a stream of decisions**
    [Henry Mintzberg]



## Former Walmart Business Strategy

# Every Day Low Price

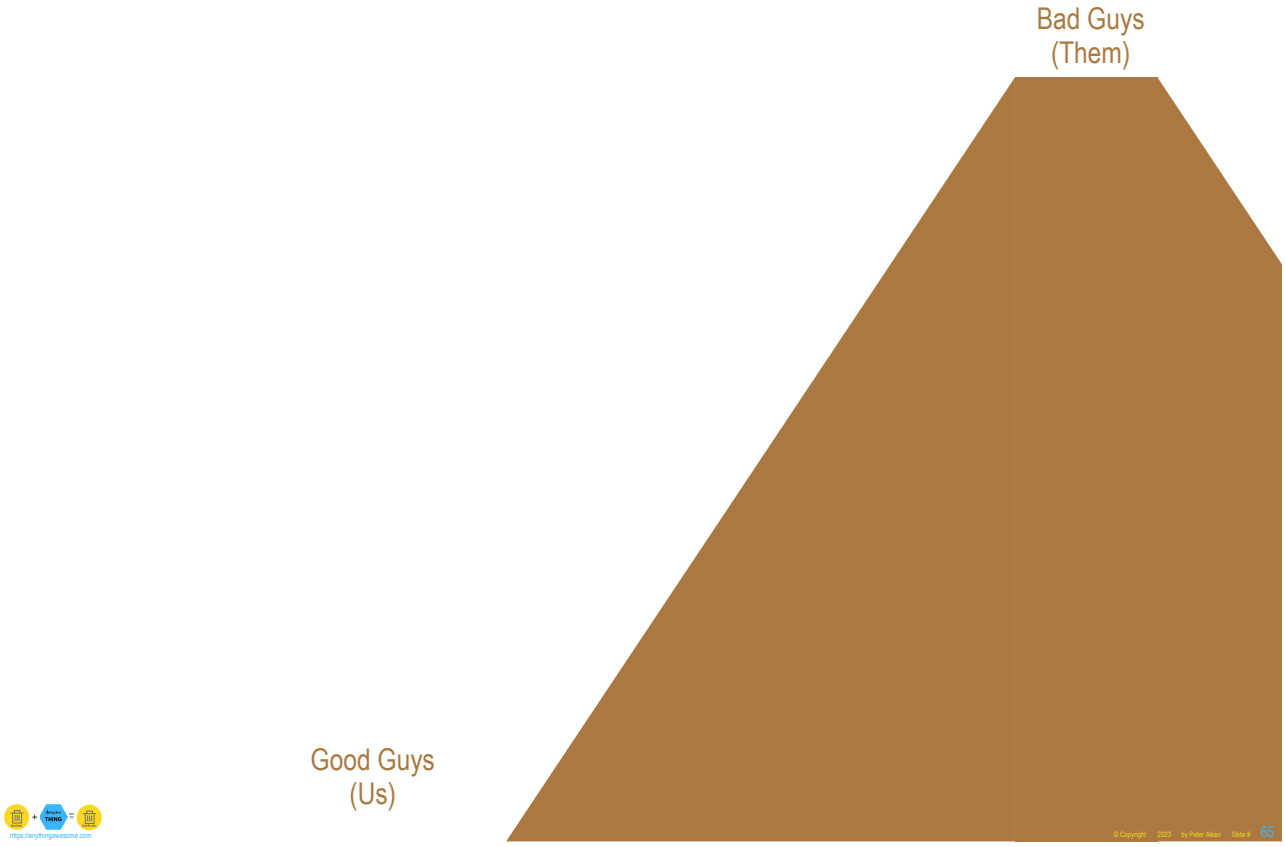


# Strategy Example 3




# Strategy Example 3

## Strategy Example 3



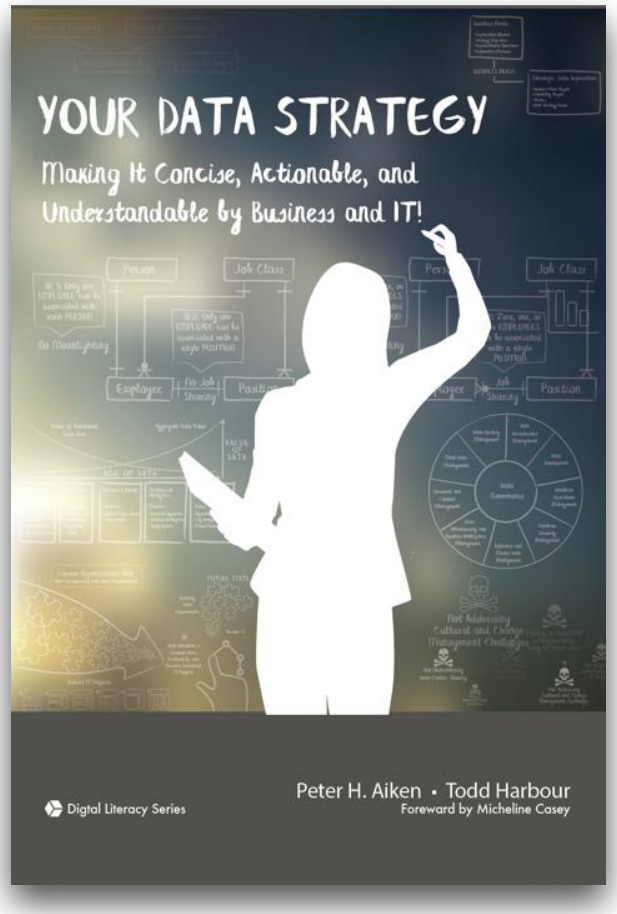
## Strategy Guides Workgroup Activities

# A pattern in a stream of decisions

## Your Data Strategy

- Highest level data guidance available ...
- Focusing data activities on business-goal achievement ...
- Providing guidance when faced with a stream of decisions or uncertainties
- Data strategy most usefully articulates how data can be best used to support organizational strategy
- This usually involves a balance of remediation and proactive measures






## Program Overview

- Definitions
  - Data Warehousing
  - DM BoK guidance
  - Legacy to digital conversion (cloud)
- Integration
  - Emphasize engineering talent
  - Incorporate leading not bleeding edge
  - Requires an adaptive rather than a prescriptive approach
- Preparation
  - Emphasis on storytelling first and visualization second
  - Analytics is both ubiquitous and not well understood
  - Keep improvements practically focused by strategy
- Best Practices
  - Cannot use what is not understood-understand what you have
  - PDCA (plan, do, check, act)
  - Iteratively refine/cull with respect to strategic direction
- Take Aways/References/Q&A


**Of What Do Your Data Warehousing Operations Consist?**

# Program Overview

- Definitions
  - Data Warehousing
  - DM BoK guidance
  - Legacy to digital conversion (cloud)
- Integration
  - Emphasize engineering talent
  - Incorporate leading not bleeding edge
  - Requires an adaptive rather than a prescriptive approach
- Preparation
  - Emphasis on storytelling first and visualization second
  - Analytics is both ubiquitous and not well understood
  - Keep improvements practically focused by strategy
- Best Practices
  - Cannot use what is not understood-understand what you have
  - PDCA (plan, do, check, act)
  - Iteratively refine/cull with respect to strategic direction
- Take Aways/References/Q&A


**Of What Do Your Data Warehousing Operations Consist?**




# Reframing the question

From: How shall we build this data warehouse?

- (or worse) … What should go into this warehouse?

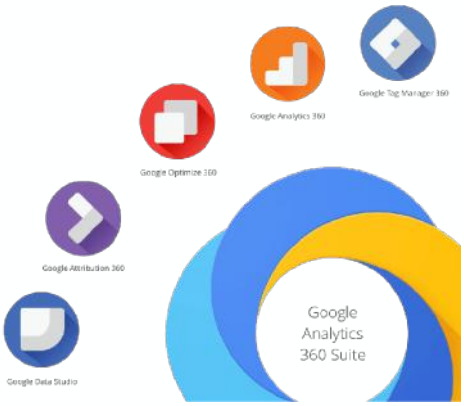
To: How can warehousing capabilities solve this business challenge?

- (better still) … How can warehousing capabilities solve this class of business challenges?

Other examples

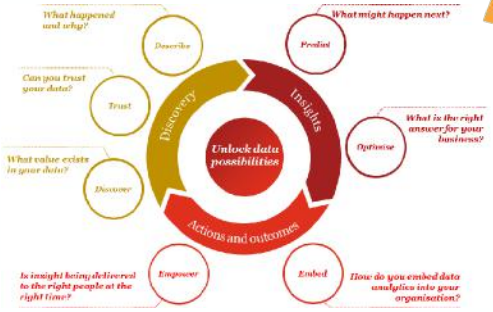
- Are you ready for warehousing?
  - Foundational practices
  - Project deliverables
- Will you get it right the first time?
  - Is the business environment constantly evolving?
  - Do you have an agreed upon enterprise-wide vocabulary?
- Is your data warehouse intended to be the enterprise audit-able system of record?
  - Extract, transform and load requirements
  - Data transformation requirements
  - How fast do you need results?
  - Performance of inserts vs reads

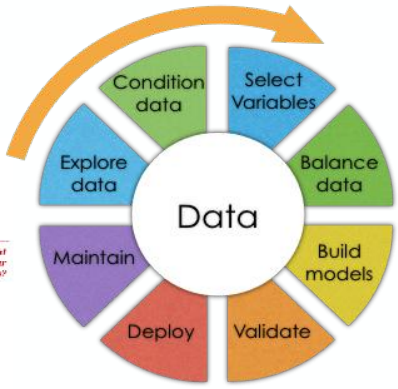






# Analytics



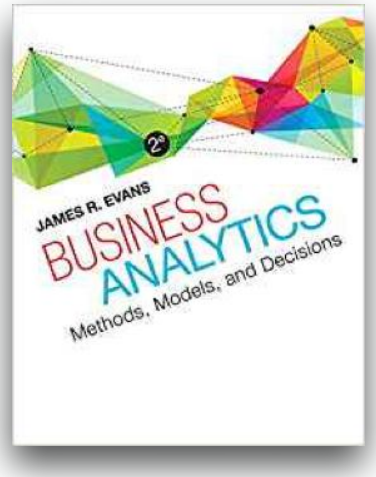








## Business Analytics, or simply analytics, is:

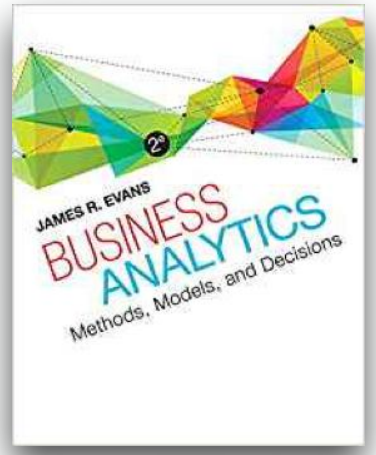
- is the use of data, information technology, statistical analysis, quantitative methods, and mathematical or computer-based models to help managers gain improved insight about their business operations and make better, fact-based decisions.
- is “a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving.”
- supported by various tools such as Microsoft Excel and various Excel add-ins, commercial statistical software package such as SAS or Minitab and more complete business intelligence suites that integrate data with analytical software




*Business Analytics: Methods, Models, and Decisions*, Second Edition by James R. Evans, page 4.

# Analytics

- is the use of data, information technology, statistical analysis, quantitative methods, and mathematical or computer-based models to help managers gain improved insight about their business operations and make better, fact-based decisions.
- is “a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving.”
- supported by various tools such as Microsoft Excel and various Excel add-ins, commercial statistical software package such as SAS or Minitab and more complete business intelligence suites that integrate data with analytical software

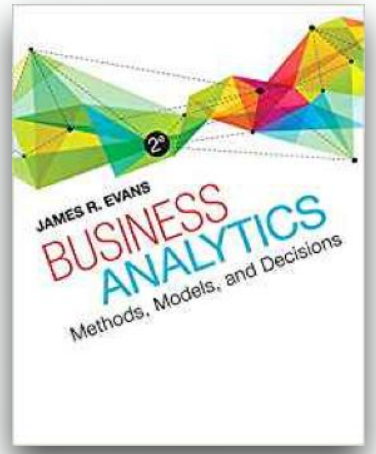



*Business Analytics: Methods, Models, and Decisions*, Second Edition by James R. Evans, page 4.



# Analytics

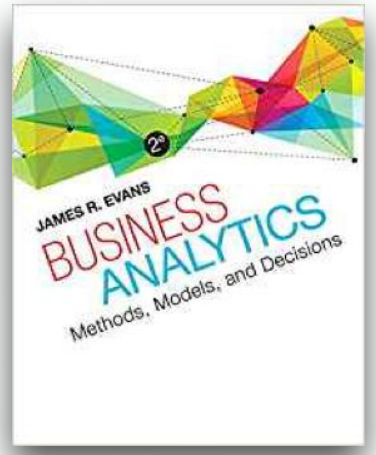
- is the use of data, hardware/software, statistical analysis, quantitative methods, and mathematical or computer-based models to help managers gain improved insight about their business operations and make better, fact-based decisions.
- is “a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving.”
- supported by various tools such as Microsoft Excel and various Excel add-ins, commercial statistical software package such as SAS or Minitab and more complete business intelligence suites that integrate data with analytical software




*Business Analytics: Methods, Models, and Decisions*, Second Edition by James R. Evans, page 4.

# Analytics

- is the use of data, hardware/software, quantitative methods, and mathematical or computer-based models to help managers gain improved insight about their business operations and make better, fact-based decisions.
- is “a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving.”
- supported by various tools such as Microsoft Excel and various Excel add-ins, commercial statistical software package such as SAS or Minitab and more complete business intelligence suites that integrate data with analytical software

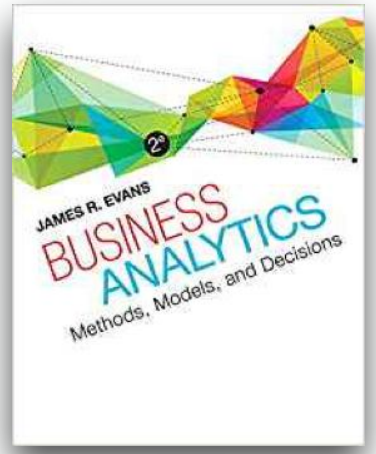



*Business Analytics: Methods, Models, and Decisions*, Second Edition by James R. Evans, page 4.



# Analytics

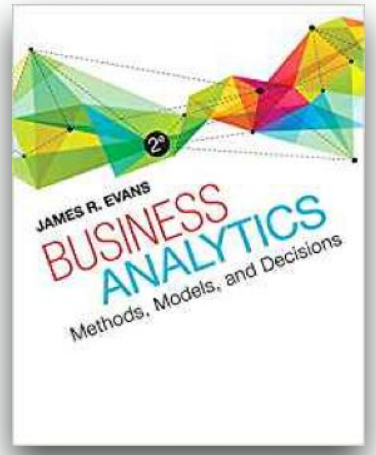
- is the use of data, hardware/software, quantitative methods, and models to help managers gain improved insight about their business operations and make better, fact-based decisions.
- is “a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving.”
- supported by various tools such as Microsoft Excel and various Excel add-ins, commercial statistical software package such as SAS or Minitab and more complete business intelligence suites that integrate data with analytical software




*Business Analytics: Methods, Models, and Decisions*, Second Edition by James R. Evans, page 4.

# Analytics

- is the use of data, hardware/software, quantitative methods, and models to help managers make better decisions.
- is "a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving."
- supported by various tools such as Microsoft Excel and various Excel add-ins, commercial statistical software package such as SAS or Minitab and more complete business intelligence suites that integrate data with analytical software

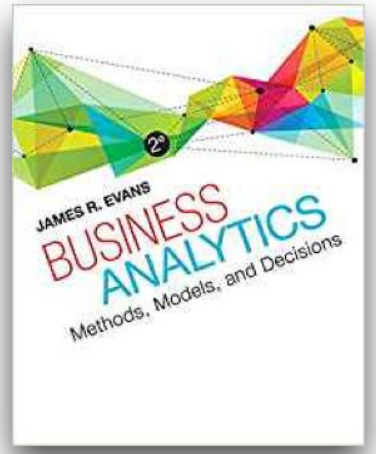



*Business Analytics: Methods, Models, and Decisions*, Second Edition by James R. Evans, page 4.



# Analytics

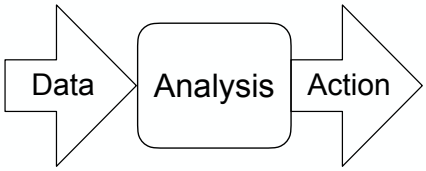
- is the use of data, hardware/software, quantitative methods, and models to help managers make better decisions.
- the process of transforming data into actions through analysis to solve problems
- supported by various tools such as Microsoft Excel and various Excel add-ins, commercial statistical software package such as SAS or Minitab and more complete business intelligence suites that integrate data with analytical software




*Business Analytics: Methods, Models, and Decisions*, Second Edition by James R. Evans, page 4.

## Analytics (31 words almost ⅓ of original)

- is the use of data, hardware/software, quantitative methods, and models to help managers make better decisions- the process of transforming data into actions through analysis to solve problems using **tools**

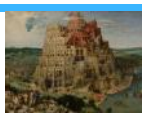




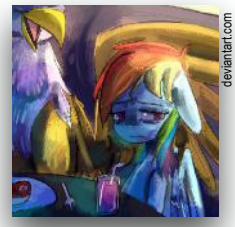



# Data Analysis

# Understanding = Interoperability



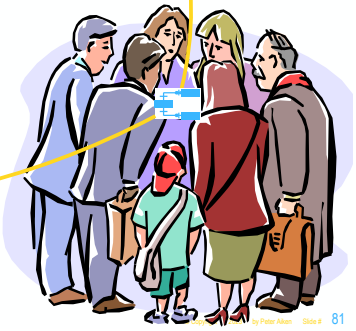
- All organizations have architectures
  - Business
  - Process
  - Systems
  - Security
  - Technical
  - Data/Information




- Some are better understood and documented (and therefore more useful to the organization)
- 'Understanding an architecture'
  - Documented and articulated as a (digital) blueprint illustrating the commonalities and interconnections among the architectural components
- Ideally the understanding is shared by
  - Business
  - Technical
  - Systems

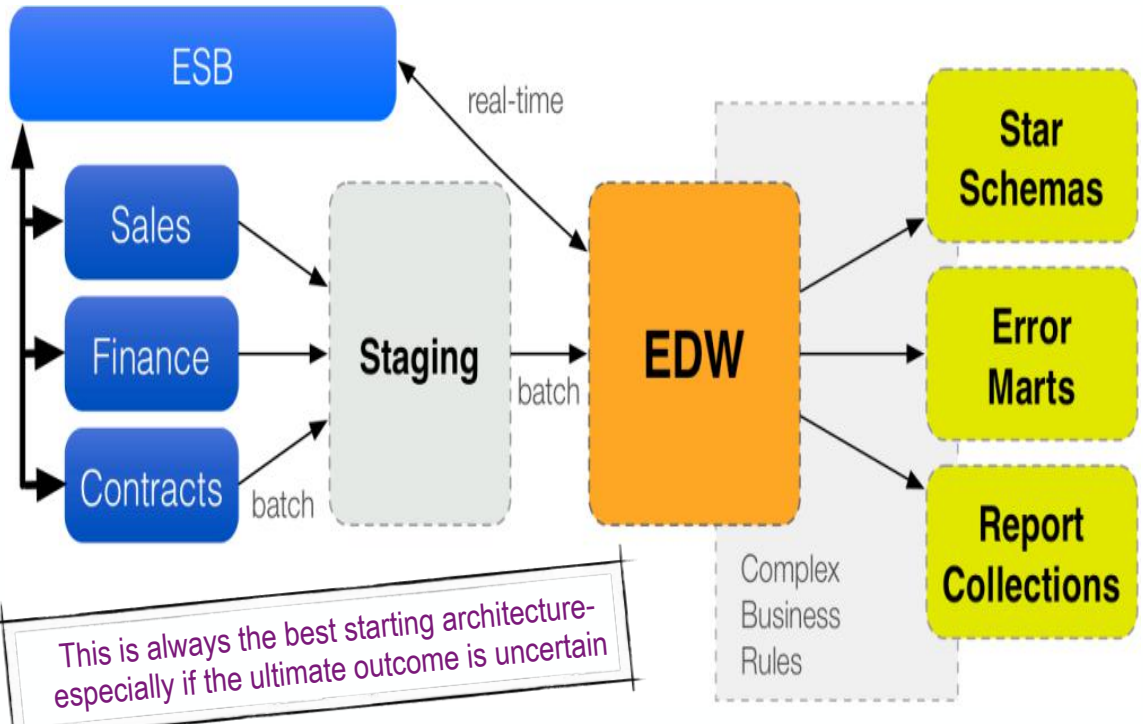
Common vocabulary expressing integrated requirements ensuring that data assets are stored, arranged, managed, and used in systems in support of organizational strategy

Trusted Catalog



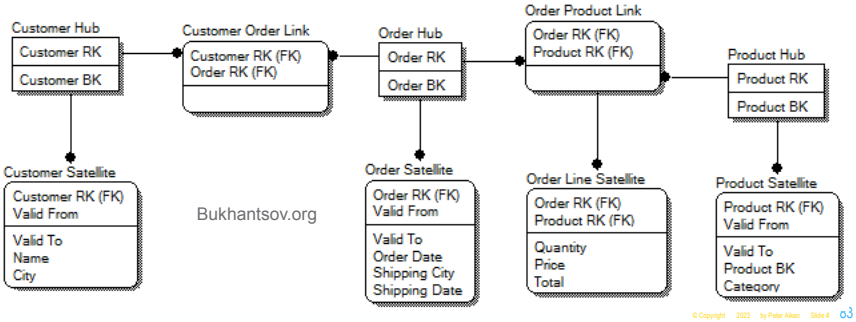


# Data Vault Implementation



# Data Vault

- Designed to facilitate long-term historical storage, focusing on ease of implementation
- Retains data lineage information (source/date)
- “All the data, all the time” - hybrid approach of Inmon and Kimball.
- Comprised of
  - Hubs (which contain a list of business keys that do not change)
  - Links (Associations/transactions between hubs)
  - Satellites (descriptive attributes associated with hubs and links)






# Data Vault Pros and Cons

- Pros
  - Simple integration
  - Houses immense amounts of data with excellent performance
  - Full data lineage captured
- Cons
  - Complication is pushed to the “back end”
  - Can be difficult to setup for many data workers
  - No widespread support for ETL tools yet

# Comparison

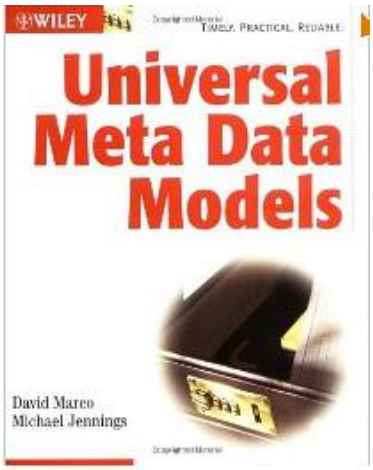
Initial Design Starting Point

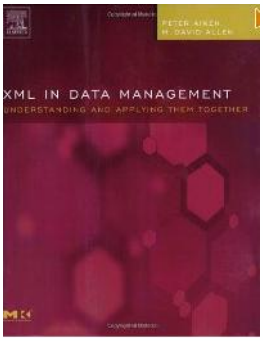
| | 3NF | Dimensional | Vault |
|---|---|---|---|
| Scalability | ☑ | ☑ | ☑ |
| Flexibility | ☒ | ☒ | ☑ |
| Reengineering | ☒ | ☒ | ☑ |
| Auditability | ☐ | ☐ | ☑ |
| Business Interpretable | ☑ | ☑ | ☒ |
| Presentation Layer | ☒ | ☑ | ☒ |
| Performance | ☒ | ☑ | ☑ |
| Support | ☑ | ☑ | ☐ |

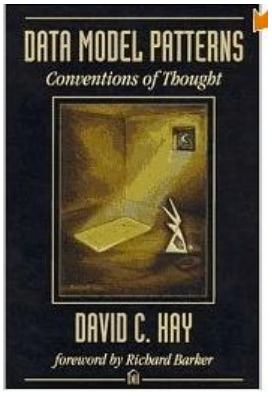


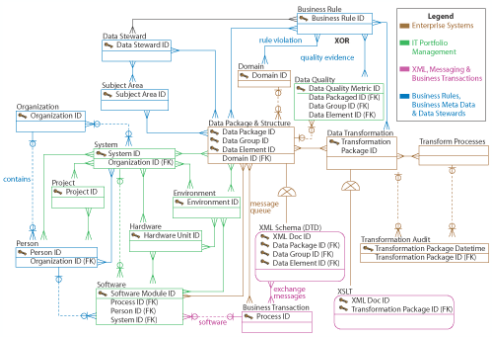


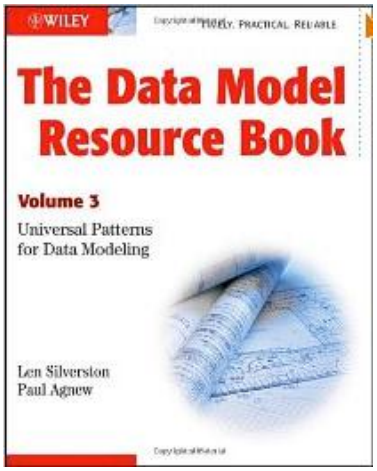
# Meta Data Models



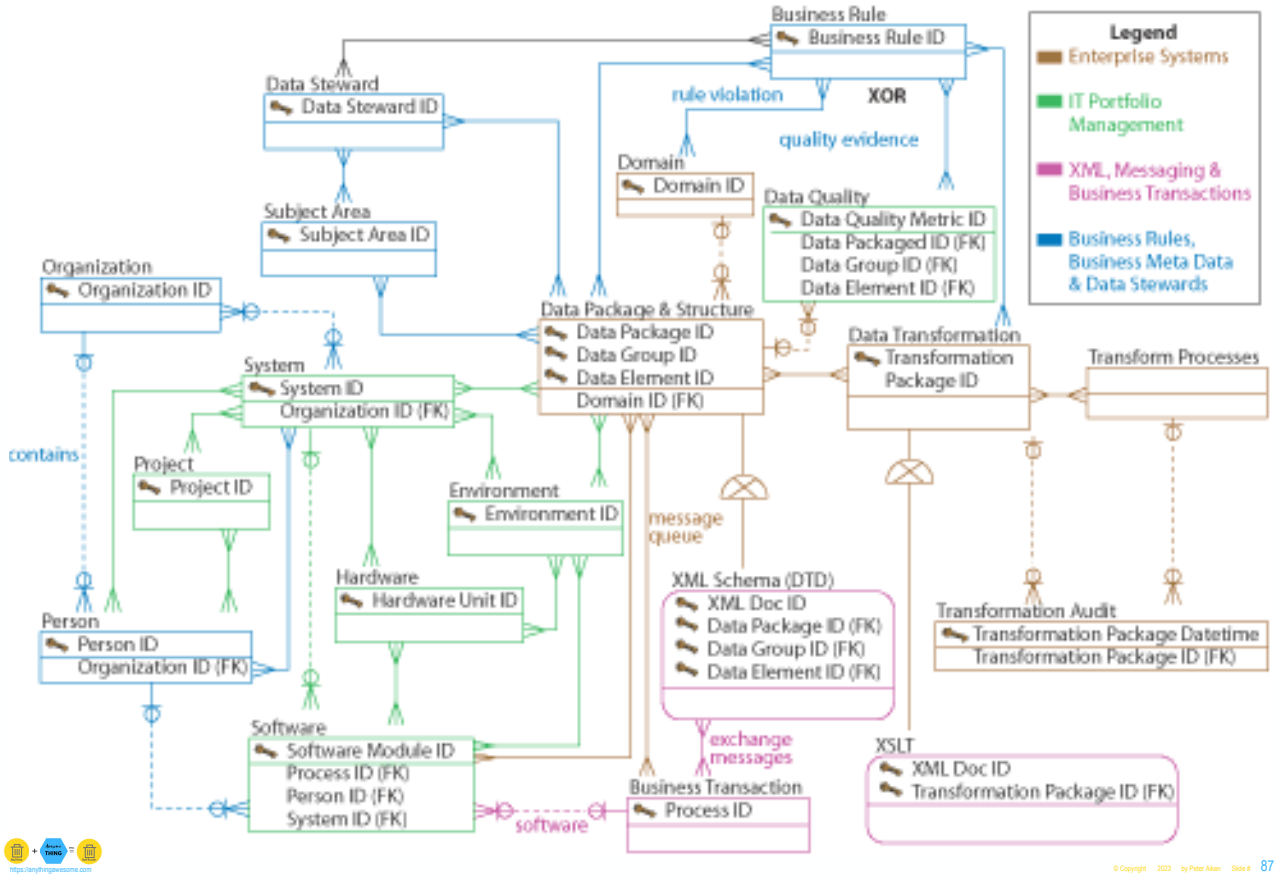




Source:http://dmreview.com/article_sub.cfm?articleID=1000941

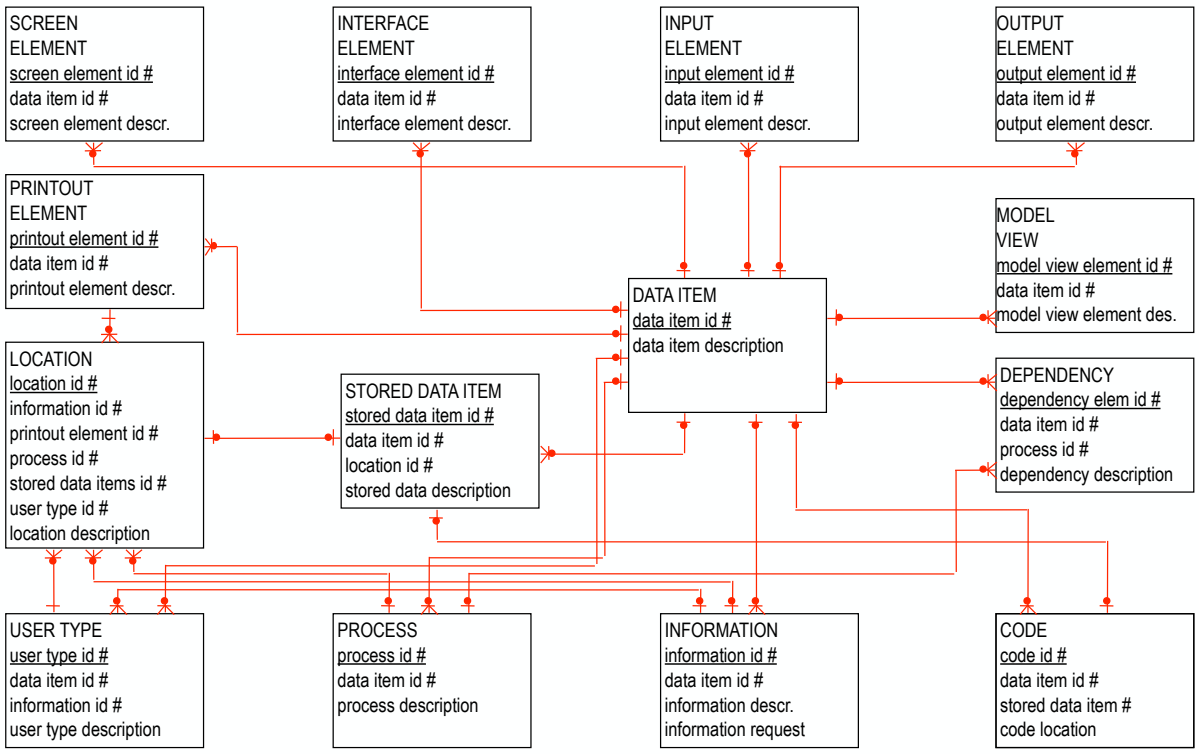
# Marco & Jennings's Metadata Model



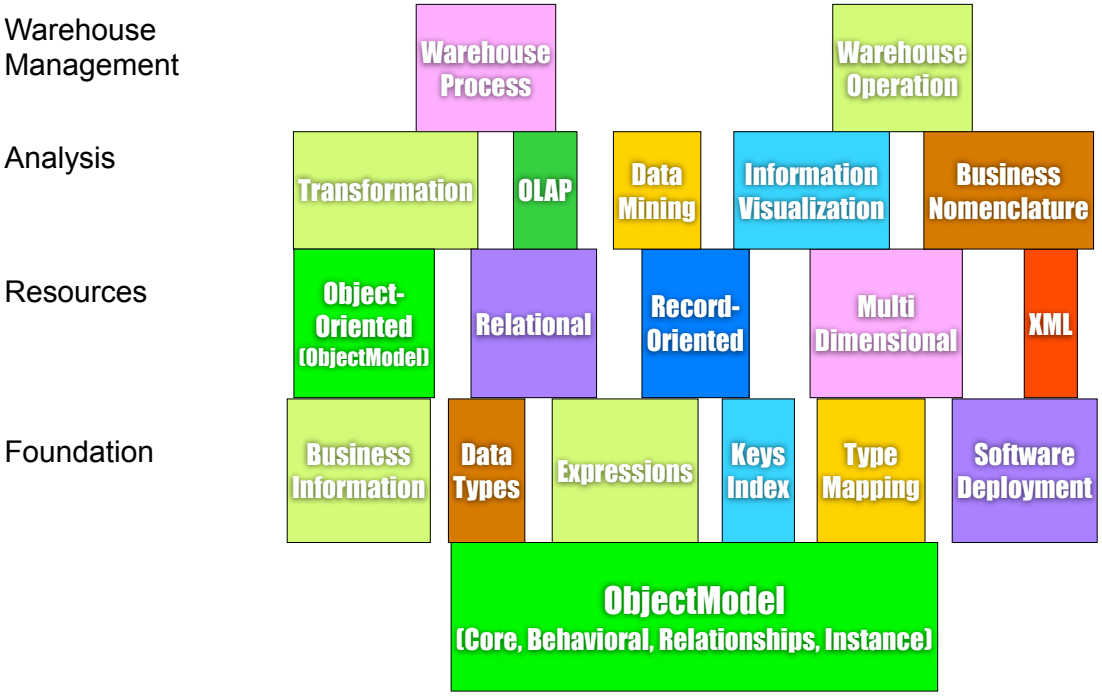



# ETL Metadata Data Model

https://anythingawesome.com/reverseengineeringofdata.html



# Overview of CWM Metamodel




http://www.omg.org/technology/documents/modeling_spec_catalog.htm



# Program Overview

- Definitions
  - Data Warehousing
  - DM BoK (and iterative use)
  - Legacy (to digital conversion)
- Integration
  - Emphasize engineering talent
  - Incorporate leading not bleeding edge
  - Requires an adaptive rather than a prescriptive approach
- Preparation
  - Emphasis on storytelling first and visualization second
  - Analytics is both ubiquitous and not well understood
  - Keep improvements practically focused by strategy
- Best Practices
  - Cannot use what is not understood-understand what you have
  - PDCA
  - Cull with respect to strategic direction
- Take Aways/References/Q&A


**Of What Do Your Data Warehousing Operations Consist?**

# (16) Causes of Data Warehousing Failure

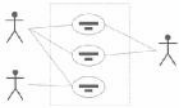
- The project is over budget
- Slipped schedule
- Unimplemented functions and capabilities
- Unhappy users
- Unacceptable performance
- Poor availability
- Inability to expand
- Poor quality data/reports
- Too complicated for users
- Project not cost justified
- Poor quality data
- Many more values of gender code than (M/F)
- Incorrectly structured data
- Provides correct answer to wrong question
- Bad warehouse design
- Overly complex





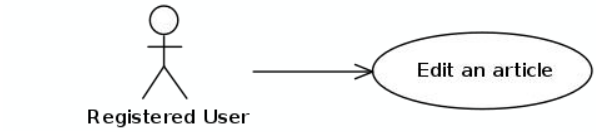
# Warehouse Requirements are Largely Use Case Driven

**Use Case**



- A usage scenario for a piece of software; often used in the plural to suggest situations where a piece of software may be useful.
- A potential scenario in which a system receives an external request (such as user input) and responds to it.

https://en.wikipedia.org/wiki/Use_case




- Difficult to holistically evaluate without integrated glossary
- Unable to capture non-functional requirements

Trusted Catalog/Glossary/Dictionary/Encyclopedia

- The plan for implementing non-functional requirements is detailed in the system architecture, because they are usually architecturally significant requirements.
- The average data warehouse is rebuilt 7 times before it is considered useful

# Upcoming Events

Time: 19:00 UTC (2:00 PM NYC) | Presented by: Peter Aiken, PhD


## Data Management Best Practices
*12 December 2023*

DGIQ Data Governance & Information Quality CONFERENCE

## Data Strategy Best Practices
*9 January 2024*


## Data Modeling Fundamentals
*12 February 2024*

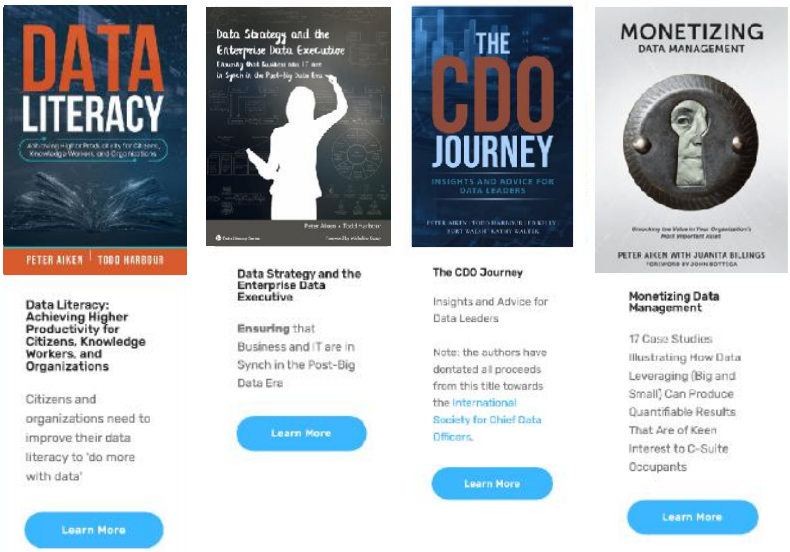
Brought to you by:

DATAVERSITY

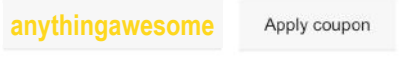
Note: In this .pdf, clicking any webinar title opens the registration link



# Event Pricing




- 20% off directly from the publisher on select titles
- My Book Store @ http://anythingawesome.com
- Enter the code "anythingawesome" at the Technics bookstore checkout where it says to "Apply Coupon"

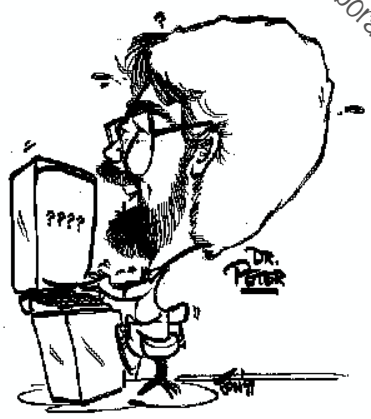


Critical Design Review?

Executive Data Literacy Training?

Collaboration?

Mentoring?


Independent Verification & Validation

Peter.Aiken@AnythingAwesome.com +1.804.382.5957

Reverse Engineering Expertise?

Hiring Assistance?

# Thank You!

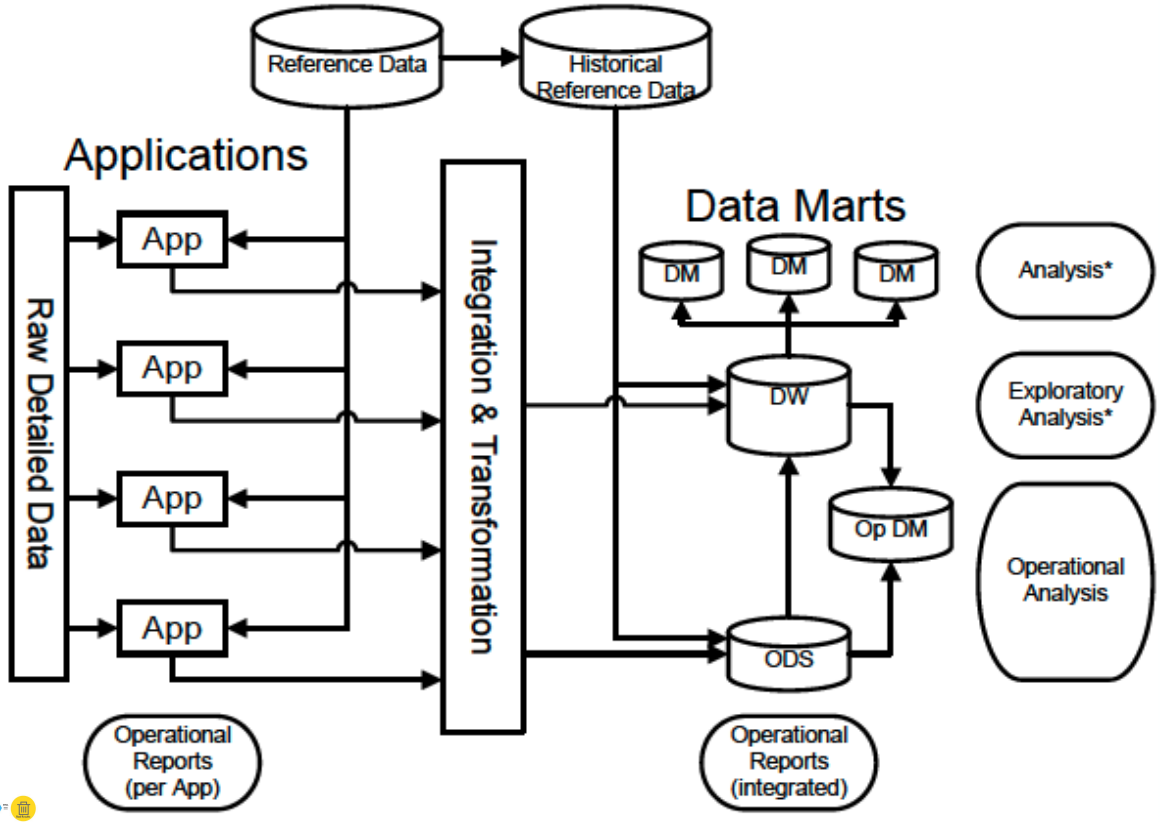
Use your data more strategically?

Tool/automation evaluation?

Book a call with Peter to discuss anything - https://anythingawesome.com/OfficeHours.html



## Corporate Information Factory Architecture



# Corporate Information Factory Architecture

| | |
|---|---|
| Raw Detailed data | Operational / Transactional Application data of the enterprise. The raw detailed data provides the source data to be integrated into the Operational Data Store (ODS) and DW components. They can also be in database or other storage or file format. |
| Integration and Transformation | This layer of the architecture is where the un-integrated data from the various application sources stores is combined / integrated and transformed into the corporate representation in the DW. |
| Reference Data | Reference data was a precursor to what is currently referred to as Master Data Management. The purpose was to allow common storage and access for important and frequently used common data. Focus and shared understanding on data upstream of the Data Warehouse simplifies the integration task in the DW. |
| Historical Reference Data | When current valued reference data is necessary for transactional applications, and at the same time it is critical to have accurate integration and presentation of historical data, it is necessary to capture the reference data that was in place at any point in time. For more discussion on reference data, see Chapter 8 Master and Reference Data Management. |

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International



from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

# Corporate Information Factory Architecture

| | |
|---|---|
| Operational Data Store (ODS) | The focus of data integration is meeting operating and classically operational reporting needs that require data from multiple operational systems. The main distinguishing data characteristics of an ODS compared to a DW include current-valued vs. DW historical data and volatile vs. DW non-volatile data.<br><br>Note: ODS is an optional portion of the overall CIF architecture, dependent upon specific operational needs, and acknowledged as a component that many businesses omit. |
| Operational Data Mart (Oper-Mart) | A data mart focuses on tactical decision support. Distinguishing characteristics include current-valued vs. DW historical data, tactical vs. DW strategic analysis, and sourcing of data from an ODS rather than just the DW. The Oper-Mart was a later addition to the CIF architecture. |
| Data Warehouse (DW) | The DW is a large, comprehensive corporate resource, whose primary purpose is to provide a single integration point for corporate data in order to serve management decision, and strategic analysis and planning.<br><br>The data flows into a DW from the application systems and ODS ,and flows out to the data marts, usually in one direction only. Data that needs correction is rejected, corrected at its source, and re-fed through the system. |
| Data Marts (DM) | The purpose of the data marts is to provide for DSS / information processing and access that is customized and tailored for the needs of a particular department or common analytic need. |

# Corporate Information Factory Architecture

| | | |
|---|---|---|
| Applications | Isolated Operational Reports | Limited to data within one application instance |
| ODS | Integrated Operational Reports | Reports requiring data from multiple source systems. Typically, they have more operational than analytical orientation, with little historical data. |
| DW | Exploratory Analysis | The complete set of corporate data allows for discovery of new relationships and information. Many BI data mining tools work with flat-file extracts from the DW, which can also offload the processing burden from the DW. |
| Oper-Mart | Tactical Analytics | Analytic reporting based on current-values with a tactical focus. Dimensional data modeling techniques employed. |
| Data Mart | Analytics – classical management decision support, and Strategic Analytics | Inmon's early focus was on "departmental analysis", which was experientially true for real-world organizational issues, such as political and funding expediency. Later work expanded concepts to common-analytic needs crossing departmental boundaries. |

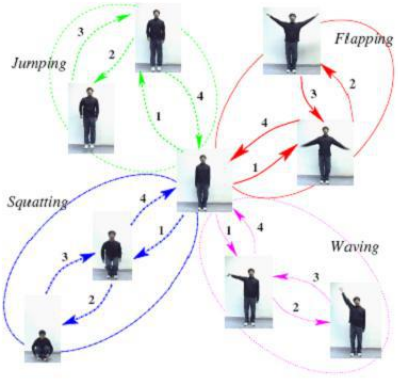




# Goals and Principles

1. To support and enable effective business analysis and decision making by knowledgeable workers
2. To build and maintain the environment/infrastructure to support business intelligence activities, specifically leveraging all the other data management functions to cost effectively deliver consistent integrated data for all BI activities

## Activities




- Understand BI information needs
- Define and maintain the DW/BI architecture
- Process data for BI
- Implement data warehouse/data marts
- Implement BI tools and user interfaces
- Monitor and tune DW processes
- Monitor and tune BI activities and performance

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International



## Primary Deliverables

- DW/BI Architecture
- Data warehouses, marts, cubes etc.
- Dashboards-scorecards
- Analytic applications
- Files extracts (for data mining, etc.)
- BI tools and user environments
- Data quality feedback mechanism/loop

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

# Roles and Responsibilities

- **Suppliers**
  - Executives/managers
  - Subject Matter Experts
  - Data governance council
  - Information consumers
  - Data producers
  - Data architects/analysts
- **Participants**
  - Executives/managers
  - Data Stewards
  - Subject Matter Experts
  - Data Architects
  - Data Analysts
  - Application Architects
  - Data Governance Council
  - Data Providers
  - Other BI Professionals
- **Consumers**
  - Application Users
  - BI and Reporting Users
  - Application Developers and Architects
  - Data integration Developers and Architects
  - BI Vendors and Architects
  - Vendors, Customers and Partners


from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International



from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

# Technology

- ETL
- Change Management Tools
- Data Modeling Tools
- Data Profiling Tools
- Data Cleansing Tools
- Data Integration Tools
- Reference Data Management Applications
- Master Data Management Applications
- Process Modeling Tools
- Meta-data Repositories
- Business Process and Rule Engines

# Guiding Principles


- Obtain executive commitment and support.
- Secure business SMEs.
- Be business focused and driven. Let the business drive the prioritization.
- Demonstrate data quality is essential.
- Provide incremental value.
- Transparency and self service.
- One size does not fit all: Find the right tools and products for each of your segments.
- Think and architect globally, act and build locally.
- Collaborate with and integrate all other data initiatives, especially those for data governance, data quality and metadata.
- Start with the end in mind.
- Summarize and optimize last, not first.

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International



# References

Adamson, Christopher. Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance. John Wiley & Sons, 2006. ISBN 0-471-77709-9. 345 pages.

Adamson, Christopher and Michael Venerable. Data Warehouse Design Solutions. John Wiley & Sons, 1998. ISBN 0-471-25195-X. 544 pages.

Adelman, Sid and Larissa T. Moss. Data Warehouse Project Management. Addison-Wesley Professional, 2000. ISBN 0-201-61635-1. 448 pages.

Adelman, Sid and others. Impossible Data Warehouse Situations: Solutions from the Experts. Addison-Wesley, 2002. ISBN 0-201-76033-9. 432 pages.

Brackett, Michael. The Data Warehouse Challenge: Taming Data Chaos. New York: John Wiley & Sons, 1996. ISBN 0-471-12744-2. 579 pages.

Caserta, Joe and Ralph Kimball. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data. John Wiley & Sons, 2004. ISBN 0-764-56757-8. 525 pages.

# References


Correy, Michael J. and Michael Abby. Oracle Data Warehousing: A Practical Guide to Successful Data Warehouse Analysis, Build and Roll-Out. TATA McGraw-Hill, 1997. ISBN 0-074-63069-5.

Covey, Stephen R. The 7 Habits of Highly Effective People. Free Press, 2004. ISBN 0743269519. 384 Pages.

Dyche, Jill. E-Data: Turning Data Into Information With Data Warehousing. Addison-Wesley, 2000. ISBN 0-201-65780-5. 384 pages.

Gill, Harjinder S. and Prekash C. Rao. The Official Guide To Data Warehousing. Que, 1996. ISBN 0-789-70714-4. 382 pages.

Hackney, Douglas. Understanding and Implementing Successful Data Marts. Addison Wesley, 1997. ISBN 0-201-18380-3. 464 pages.

Imhoff, Claudia, Nicholas Galemmo and Jonathan G. Geiger. Mastering Data Warehouse Design: Relational and Dimensional Techniques. John Wiley & Sons, 2003. ISBN 0-471-32421-3. 456 pages.

Imhoff, Claudia, Lisa Loftis and Jonathan G. Geiger. Building the Customer-Centric Enterprise: Data Warehousing Techniques for Supporting Customer Relationship Management. John Wiley & Sons, 2001. ISBN 0-471-31981-3. 512 pages.

Inmon, W. H. Building the Data Warehouse, 4th Edition. John Wiley & Sons, 2005. ISBN 0-764-59944-5. 543 pages.

Inmon, W. H. Building the Operational Data Store, 2nd edition. John Wiley & Sons, 1999. ISBN 0-471-32888-X. 336 pages.

Inmon, W. H., Claudia Imhoff and Ryan Sousa. The Corporate Information Factory, 2nd edition. John Wiley & Sons, 2000. ISBN 0-471-39961-2. 400 pages.

Inmon, W. H. and Richard D. Hackathorn. Using the Data Warehouse. Wiley-QED, 1994. ISBN 0-471-05966-8. 305 pages.

Inmon, William H., John A. Zachman and Jonathan G. Geiger. Data Stores, Data Warehousing and the Zachman Framework. McGraw-Hill, 1997. ISBN 0-070-31429-2. 358 pages.

Kimball, Ralph and Margy Ross. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2nd edition. New York: John Wiley & Sons, 2002. ISBN 0-471-20024-7. 464 pages.

Kimball, Ralph, Laura Reeves, Margy Ross and Warren Thornwaite. The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses. John Wiley & Sons, 1998. ISBN 0-471-25547-5. 800 pages.

Kimball, Ralph and Richard Merz. The Data Webhouse Toolkit: Building the Web-Enabled Data Warehouse. John Wiley & Sons, 2000. ISBN 0-471-37680-9. 416 pages.

Mattison, Rob. Web Warehousing & Knowledge Management. McGraw Hill, 1999. ISBN 0-070-41103-4. 576 pages.

Morris, Henry. *Analytic Applications and Business Performance Management*. DM Review Magazine, March, 1999. www.dmreview.com. Note: www.dmreview.com is now www.information-management.com.

Moss, Larissa T. and Shaku Atre. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley, 2003. ISBN 0-201-78420-3. 576 pages.

Poe, Vidette, Patricia Klauer and Stephen Brobst. Building A Data Warehouse for Decision Support, 2nd edition. Prentice-Hall, 1997. ISBN 0-137-69639-6. 285 pages.

Ponniah, Paulraj. Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals. John Wiley & Sons – Interscience, 2001. ISBN 0-471-41254-6. 528 pages.

Westerman, Paul. Data Warehousing: Using the Wal-Mart Model. Morgan Kaufman, 2000. ISBN 155860684X. 297 pages.


### 9.5.2 Business Intelligence


Biere, Mike. Business Intelligence for the Enterprise. IBM Press, 2003. ISBN 0-131-41303-1. 240 pages.

Eckerson, Wayne W. Performance Dashboards: MEasuring, Monitoring, and Managing Your Business. Wiley, 2005. ISBN-10: 0471724173. 320 pages.

Bischoff, Joyce and Ted Alexander. Data Warehouse: Practical Advice from the Experts. Prentice-Hall, 1997. ISBN 0-135-77370-9. 428 pages.

Howson, Cindi. "The Business Intelligence Market". http://www.biscorecard.com/. Requires annual subscription to this website.

Malik, Shadan. Enterprise Dashboards: Design and Best Practices for IT. Wiley, 2005. ISBN 0471738069. 240 pages.

Moss, Larissa T., and Shaku Atre. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley, 2003. ISBN 0-201-78420-3. 576 pages.

Vitt, Elizabeth, Michael Luckevich and Stacia Misner. **Business Intelligence**. Microsoft Press, 2008. ISBN 073562660X. 220 pages.


### 9.5.3 Data Mining


Cabena, Peter, Hadjnian, Stadler, Verhees and Zanasi. Discovering Data Mining: From Concept to Implementation. Prentice Hall, 1997. ISBN-10: 0137439806




# Additional References

- http://www.information-management.com/infodirect/20050909/1036703-1.html
- http://www.agiledata.org/essays/dataWarehousingBestPractices.html
- http://www.cio.com/article/150450/Five_Key_Business_Intelligence_Trends_You_Need_to_Know?page=2&taxonomyId=3002
- http://www.computerworld.com/s/article/9228736/Business_Intelligence_and_analytics_Conquering_Big_Data?taxonomyId=9
- http://www.enterpriseirregulars.com/5706/the-top-10-trends-for-2010-in-analytics-business-intelligence-and-performance-management/
- http://www.itbusinessedge.com/cm/blogs/vizard/taking-the-analytics-pressure-off-the-data-warehouse/?cs=50698
- http://www.informationweek.com/news/software/bi/240001922

# References & Recommended Reading

### 11.4.1 General Reading


Brathwaite, Ken S. Analysis, Design, and Implementation of Data Dictionaries. McGraw-Hill Inc., 1988. ISBN 0-07-007248-5. 214 pages.

Collier, Ken. Executive Report, Business Intelligence Advisory Service, *Finding the Value in Metadata Management* (Vol. 4, No. 1), 2004. Available only to Cutter Consortium Clients, http://www.cutter.com/bia/fulltext/reports/2004/01/index.html.

Hay, David C. Data Model Patterns: A Metadata Map. Morgan Kaufmann, 2006. ISBN 0-120-88798-3. 432 pages.

Hillmann, Diane I. and Elaine L. Westbrooks, editors. Metadata in Practice. American Library Association, 2004. ISBN 0-838-90882-9. 285 pages.

Inmon, William H., Bonnie O'Neil and Lowell Fryman. Business Metadata: Capturing Enterprise Knowledge. 2008. Morgan Kaufmann ISBN 978-0-12-373726-7. 314 pages.

Marco, David, Building and Managing the Meta Data Repository: A Full Life-Cycle Guide. John Wiley & Sons, 2000. ISBN 0-471-35523-2. 416 pages.

Marco, David and Michael Jennings. Universal Meta Data Models. John Wiley & Sons, 2004. ISBN 0-471-08177-9. 478 pages.

Poole, John, Dan Change, Douglas Tolbert and David Mellor. Common Warehouse Metamodel: An Introduction to the Standard for Data Warehouse Integration. John Wiley & Sons, 2001. ISBN 0-471-20052-2. 208 pages.




# References, cont'd


Poole, John, Dan Change, Douglas Tolbert and David Mellor. Common Warehouse Metamodel Developer's Guide. John Wiley & Sons, 2003. ISBN 0-471-20243-6. 704 pages.

Ross, Ronald. Data Dictionaries And Data Administration: Concepts and Practices for Data Resource Management. New York: AMACOM Books, 1981. ISN 0-814-45596-4. 454 pages.

Tannenbaum, Adrienne. Implementing a Corporate Repository, John Wiley & Sons, 1994. ISBN 0-471-58537-8. 441 pages.

Tannenbaum, Adrienne. Metadata Solutions: Using Metamodels, Repositories, XML, And Enterprise Portals to Generate Information on Demand. Addison Wesley, 2001. ISBN 0-201-71976-2. 528 pages.

Wertz, Charles J. The Data Dictionary: Concepts and Uses, 2nd edition. John Wiley & Sons, 1993. ISBN 0-471-60308-2. 390 pages.

# References, cont'd


### 11.4.2 Meta-data in Library Science

Baca, Murtha, editor. Introduction to Metadata: Pathways to Digital Information. Getty Information Institute, 2000. ISBN 0-892-36533-1. 48 pages.

Hillmann, Diane I., and Elaine L. Westbrooks. Metadata in Practice. American Library Association, 2004. ISBN 0-838-90882-9. 285 pages.

Karpuk, Deborah. METADATA: From Resource Discovery to Knowledge Management. Libraries Unlimited, 2007. ISBN 1-591-58070-6. 275 pages.

Liu, Jia. Metadata and Its Applications in the Digital Library. Libraries Unlimited, 2007. ISBN 1-291-58306-6. 250 pages.

### 11.4.3 Geospatial Meta-data Standards

http://www.fgdc.gov/metadata/geospatial-metadata-standards.




# References, cont'd


### 11.4.4 ISO Meta-data Standards

ISO Standards Handbook 10, Data Processing—Vocabulary, 1982.

ISO 704:1987, Principles and methods of terminology.

ISO 1087, Terminology—Vocabulary.

ISO 2382-4:1987, Information processing systems—Vocabulary part 4.

ISO/IEC 10241:1992, International Terminology Standards—Preparation and layout.

FCD 11179-2, Information technology—Specification and standardization of data elements - Part 2: Classification for data elements.

ISO/IEC 11179-3:1994, Information technology—Specification and standardization of data elements - Part 3: Basic attributes of data elements.

ISO/IEC 11179-4:1995, Information technology—Specification and standardization of data elements - Part 4: Rules and guidelines for the formulation of data definitions.

ISO/IEC 11179-5:1995, Information technology—Specification and standardization of data elements - Part 5: Naming and identification principles for data elements.

ISO/IEC 11179-6:1997, Information technology—Specification and standardization of data elements - Part 6: Registration of data elements.

# More Analytical Applications

### Typical Analytical Applications for Internal Processes

**Activity-based costing (ABC).** The first step in activity-based management is to allocate costs accurately to aspects of the business such as customers, processes, or distribution channels; models incorporating activities, materials, resources, and product-offering components then allow optimization based on cost and prediction of capacity needs.

**Bayesian inference (e.g., to predict revenues).** A numerical estimate of the degree of belief in a hypothesis before and after evidence has been observed.

**Biosimulation (e.g., in pharmaceutical "in silico" research).** Manipulation of biological parameters using mathematics and/or rule bases to model how cells or other living entities react to chemical or other interventions.

**Combinatorial optimization (e.g., for optimizing a product portfolio).** The efficient allocation of limited resources to yield the best solution to particular objectives when the values of some or all of the variables (e.g., a given number of people) must be integers (because people can't be split into fractions) and there are many possible combinations. Also called *integer programming*.



# More Analytical Applications (continued)

**Constraint analysis (e.g., for product configuration).** The use of one or more constraint satisfaction algorithms to specify the set of feasible solutions. Constraints are programmed in rules or procedures that produce solutions to particular configuration and design problems using one or more constraint satisfaction algorithms.

**Experimental design (e.g., for Web site analysis).** In the simplest type of experiment, participants are randomly assigned to two groups that are equivalent to each other. One group (the program or treatment group) gets the program and the other group (the comparison or control group) does not. If the program results in statistically significant differences in the outcome variable, it is assumed to have the hypothesized effect.

**Future-value analysis.** The decomposition of market capitalization into current value (extrapolation of existing monetary returns) and future value, or expectations of future growth.

**Monte Carlo simulation (e.g., for R&D project valuation).** A computerized technique used to assess the probability of certain outcomes or risks by mathematically modeling a hypothetical event over multiple trials and comparing the outcome with predefined probability distributions.

# More Analytical Applications (continued)

**Multiple regression analysis (e.g., to determine how nonfinancial factors affect financial performance).** A statistical technique whereby the influence of a set of independent variables on a single dependent variable is determined.

**Neural network analysis (e.g., to predict the onset of disease).** Systems modeled on the structure and operation of the brain, in which the state of the system is modified by training until it can discriminate between the classes of inputs; used on large databases. Typically, a neural network is initially "trained," or fed large amounts of data and rules about data relationships—for example, "A grandfather is older than a person's father."

**Textual analysis (e.g., to assess intangible capabilities).** Analysis of the frequency, semantic relationships, and relative importance of particular terms, phrases, and documents in online text.

**Yield analysis (e.g., in semiconductor manufacturing).** Employing basic statistics (mean, median, standard deviation, etc.) to understand yield volume and quality, and to compare one batch of items with another—often displayed visually.