ATHENA
IT SOLUTIONS

Bridging the gap between
big data and big decisions.

INFO 7290:
Data Warehousing
& Business Intelligence

Fall 2015

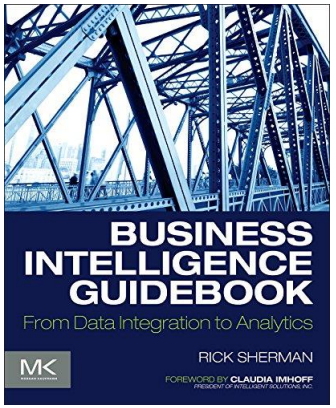
Week 1 – BI & DW Introduction

Northeastern University

Rick Sherman
Athena IT Solutions
ri.sherman@northeastern.edu

Topics – Week 1

- Course Overview
- Lecture:
  - ✓ Ch 1 Introduction
  - ✓ Ch 4 Architecture Introduction
  - ✓ Ch 8 Foundational Data Modeling – Brief
  - ✓ Ch 17 People, Process & Politics – Brief

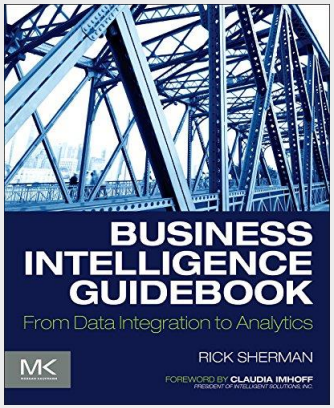


- Workshop:
  - ✓ Development Setup
  - ✓ Review DB schema
  - ✓ Queries

Microsoft SQL Server

Chapter 1:
The Business Demand for Data, Information and Analytics

- Data Deluge
- Information Demand
- BI Investments
- Information Gap
- BI Adoption Rate



Business Demand for Data, Information and Analytics

The Data Deluge

- Creating ever increasing amounts of data
  - ✓ Society





  - ✓ Business
    - Business-to-Business (B2B)
    - Business-to-Consumer (B2C)
- Businesses historically focused on enterprise application data
  - ✓ Managed it
  - ✓ Exchanged it with others that managed it

Business Demand for Data, Information and Analytics

The Analytics Need

- Business demand for analytics
  - ✓ Increase sales, manage costs & increase profits
  - ✓ Interact with customers, partners & suppliers
  - ✓ Respond to competitive pressures
  - ✓ Comply with government & industry regulations
  - ✓ Examine economic trends
- Need spans industries & enterprise size
  - ✓ Innovator & Early Adopters
  - ✓ Early & Late Majorities
- Analytics demand & awareness

The New York Times

Business Day

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH

Search | Global | DealBook

OFF THE SHELF

When Data Guys Triumph

By CADE MASSEY and BOB TEDESCHI

Published: October 1, 2011

JOSHUA MILBERG has plenty of business cred: an M.B.A. from Yale, experience in the mayor's office in Chicago, a job as a vice president for an energy consulting firm.

Enlarge This Image

Melinda Sue Gordon/Columbia Pictures

"Moneyball," with Brad Pitt. The 2003 book still makes business waves.

But all of that, Mr. Milberg says, matters less than his reputation as "the data guy" — someone who can offer insights through statistical analysis. And for that, he and a growing number of young executives can credit none other than "Moneyball: The Art of Winning an Unfair Game," by Michael Lewis.

Harvard Business Review

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

THE WALL STREET JOURNAL.

WSJ.com

August 28, 2013, 1:56 PM ET

Universities Go in Big for Big Data

ATHENA IT SOLUTIONS

Big Data

- Big Data & Data Scientist Hype
  - ✓ Everything is big data & big data analytics
    - Data big & small
    - Business analyst to Rocket (Data) Scientist
  - ✓ Technology is too tightly associated with it
- Technology myths
  - ✓ Technology solves all
  - ✓ Silver bullet solution
  - ✓ The One
    - Technology
    - Architecture
    - Vendor

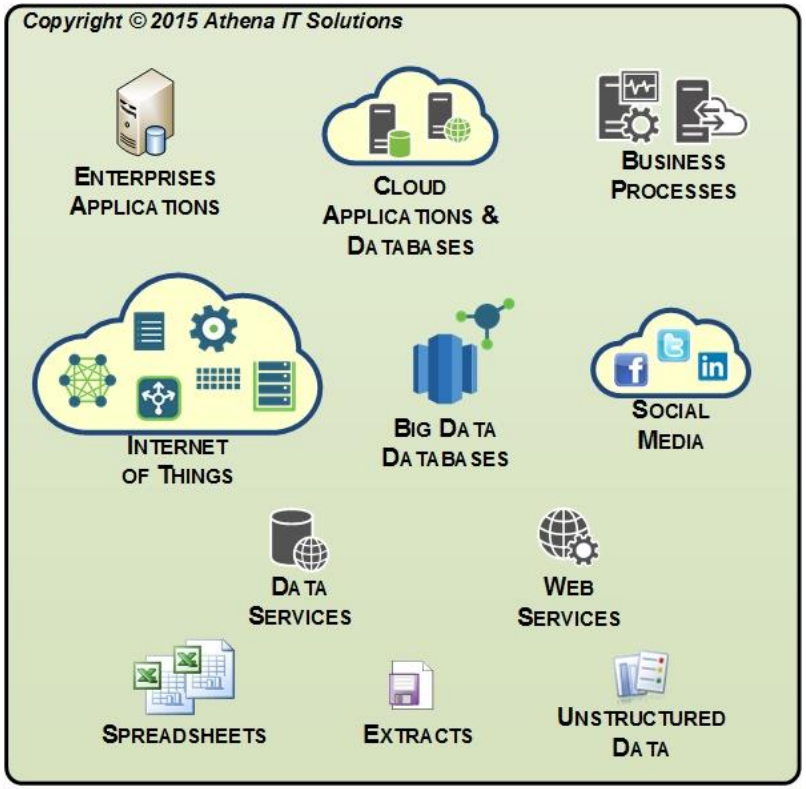
Big Data

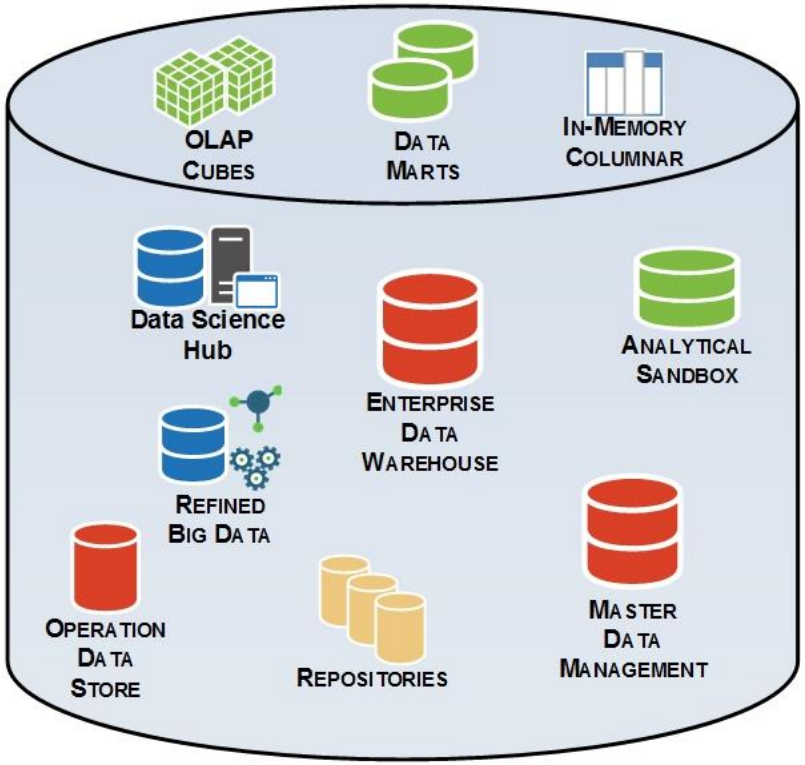
- Classic Definition: 3 V's
  - ✓ Volume - amount of data
  - ✓ Velocity - real-time
  - ✓ Variety - sources & types of sources

- Evolution more than revolution

- Expand Definition to 5 V's
  - ✓ Veracity - correctness & accuracy
  - ✓ Value - business value (ROI) of actions from insights

Differences in origin & use of data





Data Sources

- Data Capture
- Transactional or Operational

Integrated Data

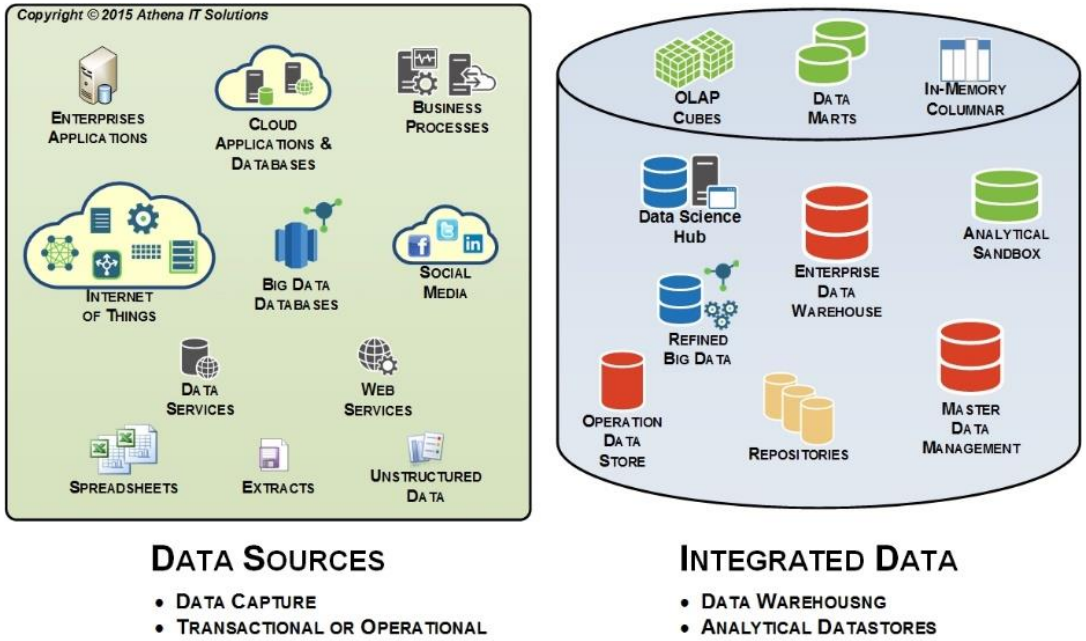
- Data Warehousng
- Analytical Datastores

Business Demand for Data, Information and Analytics

Data Variety

Key Differences from Past:

- Unstructured Data
- External Sources
- Unmanaged



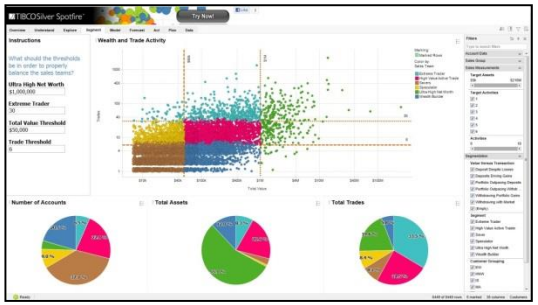
Considerations:

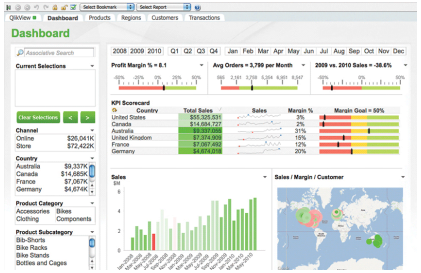
- Various technologies & architectures available for data
- Data use, purpose, value & cost drive implementation choices
- IoT more likely structured than unstructured; also domain-specific
- Many valuable unstructured data sources are low volume

Business Demand for Data, Information and Analytics

BI Trends

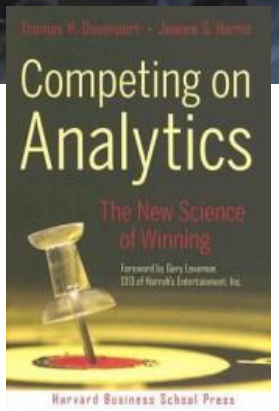
- Data Tsunami
  - ✓ 3 V's - Volume, Variety & Velocity
  - ✓ Big Data & "Small" Data
- Analytics Complexity Growing
- Business Needs & Adopting Advanced Analytics
  - ✓ Predictive Analytics
  - ✓ Data Visualization
  - ✓ Data Discovery
- Business Needs Pervasive BI
  - ✓ Self-Service BI





BI Investments

BI a Top Priority



- Leveraging data and analytics is the most important technology initiative for 2014, with 72-percent of CIOs surveyed stating that it's a critical or high priority.
  - *"2014 State of the CIO Survey." CIO magazine January 2014*
- BI and analytics will remain a top focus for CIOs through 2017 and benefits of fact-based decision-making are clear to business managers in a broad range of disciplines, including: marketing, sales, supply chain management, manufacturing, engineering, risk management, finance and HR.
  - *"Gartner Predicts Business Intelligence and Analytics Will Remain Top Focus for CIOs Through 2017." Press Release 16 December 2013.*
- "According to Gartner's annual survey of CIO technology priorities, BI and analytics has once again been named the top priority for 2012, a position it has held in three of the last five years."
  - *Gartner Research, Feb 2012*
- BI & analytics is a key IT investment over next 5 years for 83% of SMB firms
  - *IBM survey, 622 midmarket CIOs, 2Q11*

ATHENA IT SOLUTIONS

- BI, Analytics, Data Integration & Databases

- IDC projects worldwide business analytics market will grow from $37.7B in 2013 to $59.2B in 2018, a 9.4% CAGR in forecast period.
  - IDC Worldwide Business Analytics Software 2014–2018 Forecast and 2013 Vendor Shares
- BI & analytics market, combining software & services, estimated $79bn in 2012 growing at 16% to reach $143.3bn in 2016
  - The Business Intelligence Software & Services Market, 2012-2016 market study, Pringle & Company

Business Intelligence and Analytics Software by Segment, Worldwide, 2012-2013 (Millions of Dollars)

Subsegment	2013 Revenue	2013 Market Share (%)	2012 Revenue	2012-2013 Growth (%)
Analytic Applications and Performance Management	2,001	13.9	1,890	5.8
BI Platforms	8,550	59.5	7,857	8.8
CPM Suites	2,735	19.0	2,602	5.1
Advanced Analytics	1,082	7.5	962	12.5
Total	**14,368**	**100.0**	**13,311**	**7.9**

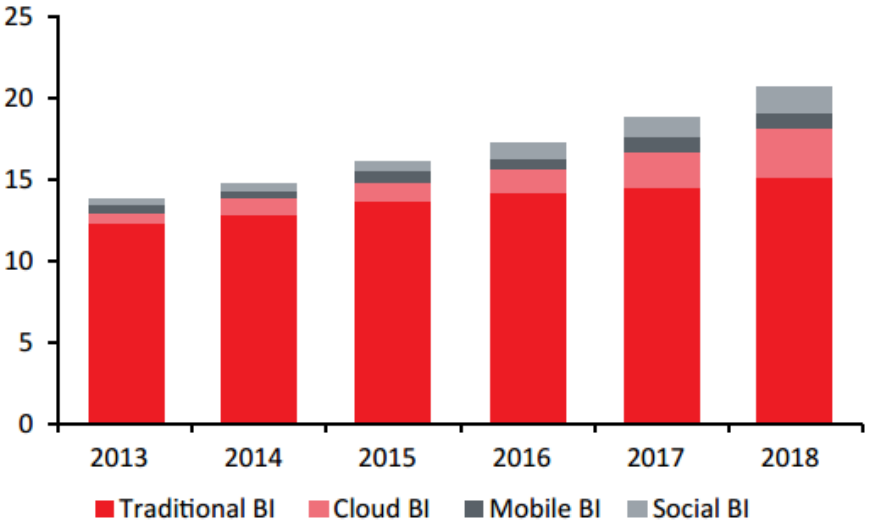
Source: Gartner (April 2014)

BI Investments
BI Market

- BI, Analytics, Data Integration & Databases

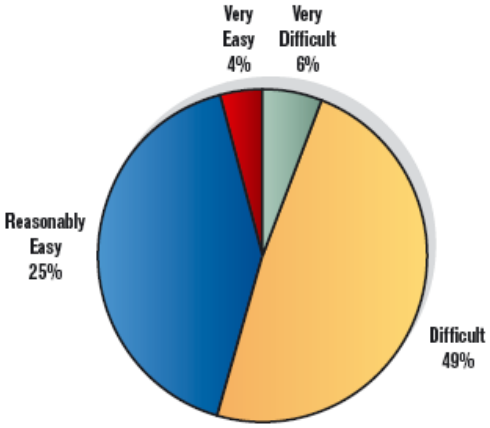
- "The global business intelligence market is projected to reach $20.81 billion in 2018, up from $13.98 billion in 2013, representing a Compound Annual Growth Rate (CAGR) of 8.28% Among all regions, North America is the largest, capturing 49% of the global BI market."
  - ✓ Redwood Capital, Gartner Research April, 2014

CHART 1: GLOBAL INTELLIGENCE MARKET SIZE, BY TECHNOLOGIES, 2013-2018 ($ BILLION)

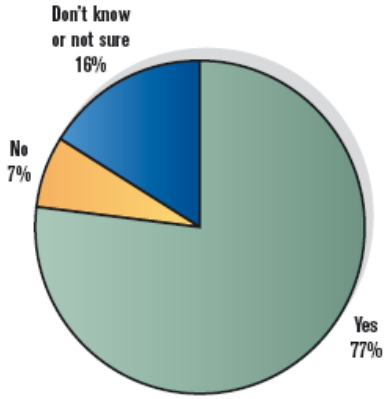


Sources: Gartner, Redwood Capital

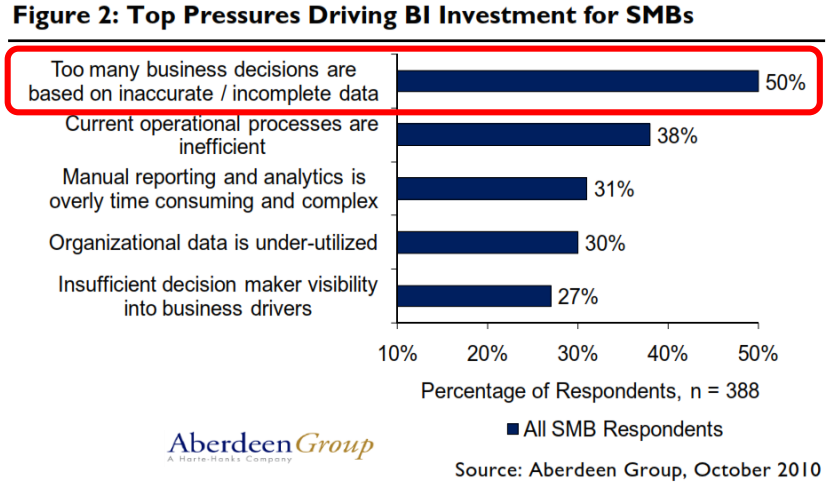
Information Gap
Business still not getting information

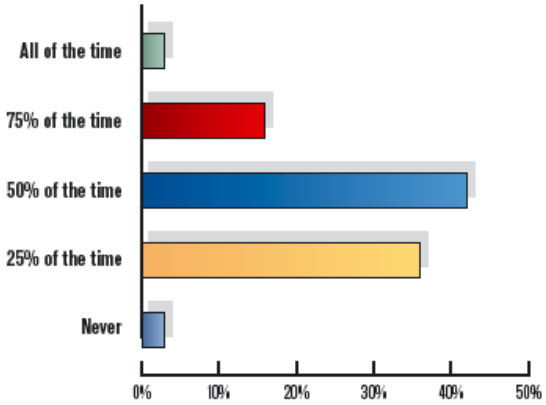


How difficult to get relevant corporate information to make business decisions



Aware of bad decisions managers have made due to insufficient information





Gut Feel Decisions

"Managers Say the Majority of Information Obtained for Their Work Is Useless, Accenture Survey Finds"

- Middle managers spend more than a quarter of their time searching for information necessary to their jobs, & when they do find it, it is often wrong
- Nearly three out of five respondents (59 percent) - poor information distribution results in missing information that might be valuable to their jobs almost every day
- 42 percent accidentally use the wrong information at least once a week
- 53 percent said less than half of the information they receive is valuable

Survey: 1,000+ middle managers of large companies

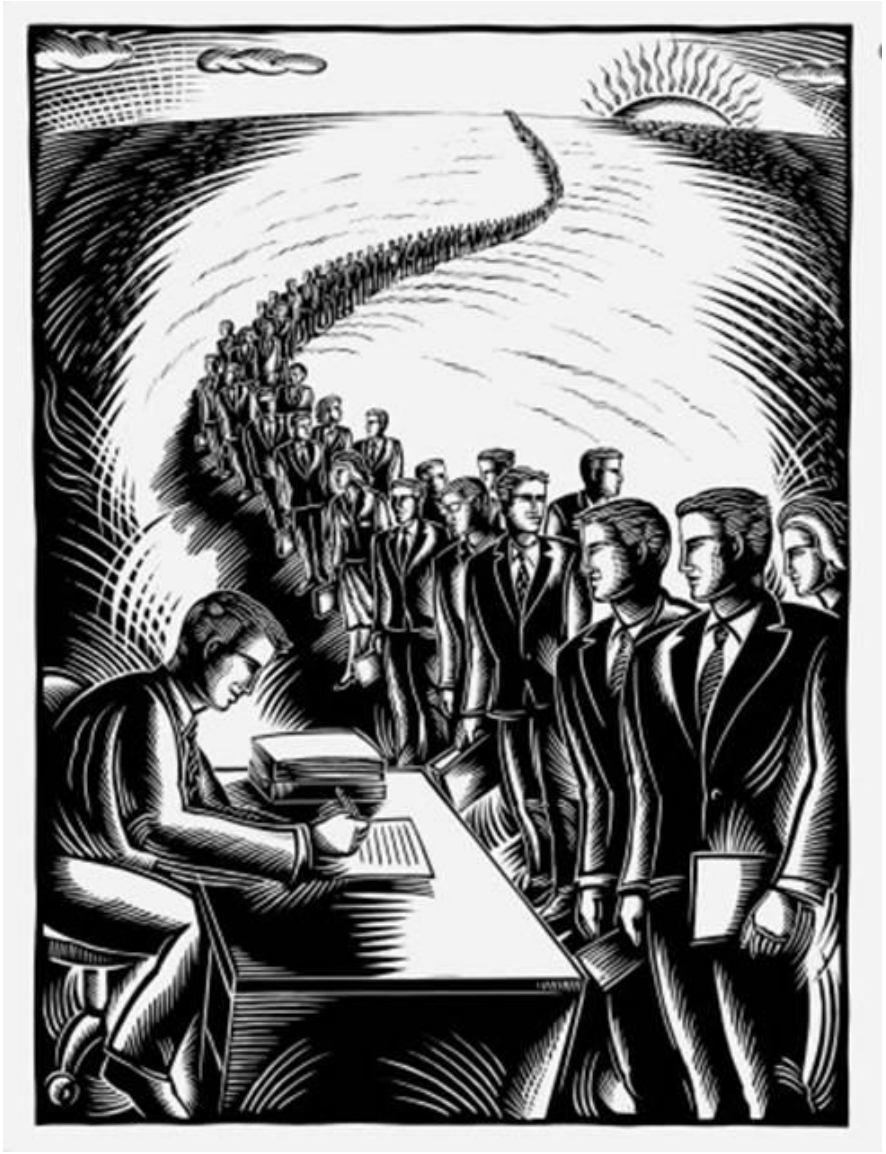
BusinessWeek

Survey: 675 US & European business executives & managers

Source: BusinessWeek Research Services

ATHENA IT SOLUTIONS

1st Wave - ERP Applications & Silos – ERP Applications



- In the beginning…
  - ✓ Green Bar Reports
  - ✓ Legacy applications
- Enterprise applications
  - ✓ Purchased & replaced legacy applications driven by Y2K
  - ✓ Business processes improved & TCO (total cost of ownership) lowered
- Data Rich, Information Poor
  - ✓ Reporting & analysis from enterprise applications fall far short of expectations
  - ✓ Rise of data shadow systems

Information Gap

1st Wave for Enterprise Information – ERP Applications

- Enterprise Resource Planning (ERP)
  - ✓ Replace disparate legacy applications
  - ✓ Operational application consolidation
  - ✓ 1990s boom time linked with Y2K
- ERP Modules related to Business Processes
  - ✓ Finance/Accounting
  - ✓ Human Resources
  - ✓ Manufacturing
  - ✓ Supply Chain Management (SCM)
  - ✓ Customer Relationship Management (CRM)
    - Salesforce automation (SFA)
    - Call Center
    - Many others

ORACLE

Microsoft

Microsoft Dynamics

SAP

INFOR

salesforce.com

NETSUITE

amdocs

sage

LAWSON

EPICOR

ATHENA IT SOLUTIONS

Current State of Reporting & Analysis

- Reporting from operational systems
  - ✓ Reporting bundled with application
  - ✓ Concentrated on specific business process & data
  - ✓ Numbers do not match across enterprise, i.e. data silos

- BI scenarios
  - ✓ Limited in scope & functionality
  - ✓ Not pervasive
  - ✓ Data & analytical silos
  - ✓ Much labor needed to create, deliver & expand

- Significant reporting built by business
  - ✓ Spreadsheets
  - ✓ Data Shadow Systems or Spreadmarts

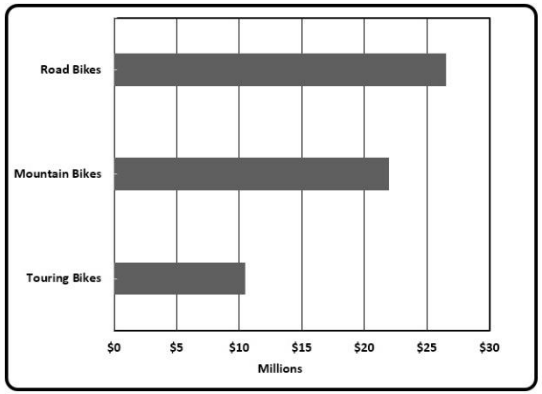
Current State of Reporting & Analysis

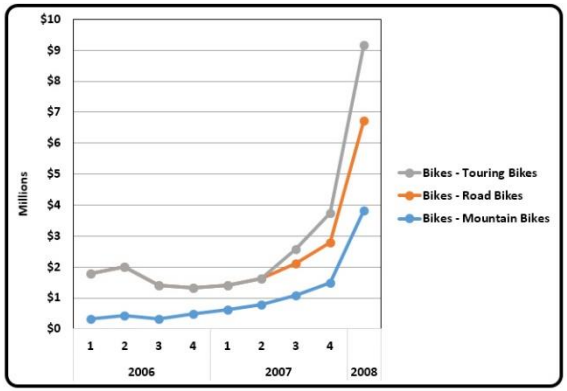
- Spreadsheets

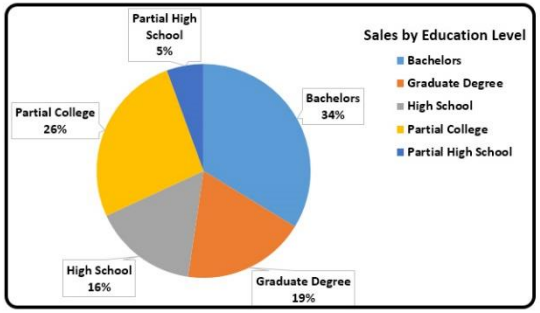
  ✓ Primary tool used by business to present & analyze data

  ✓ Final destination of data

  ✓ Only "BI" tool with pervasive business use





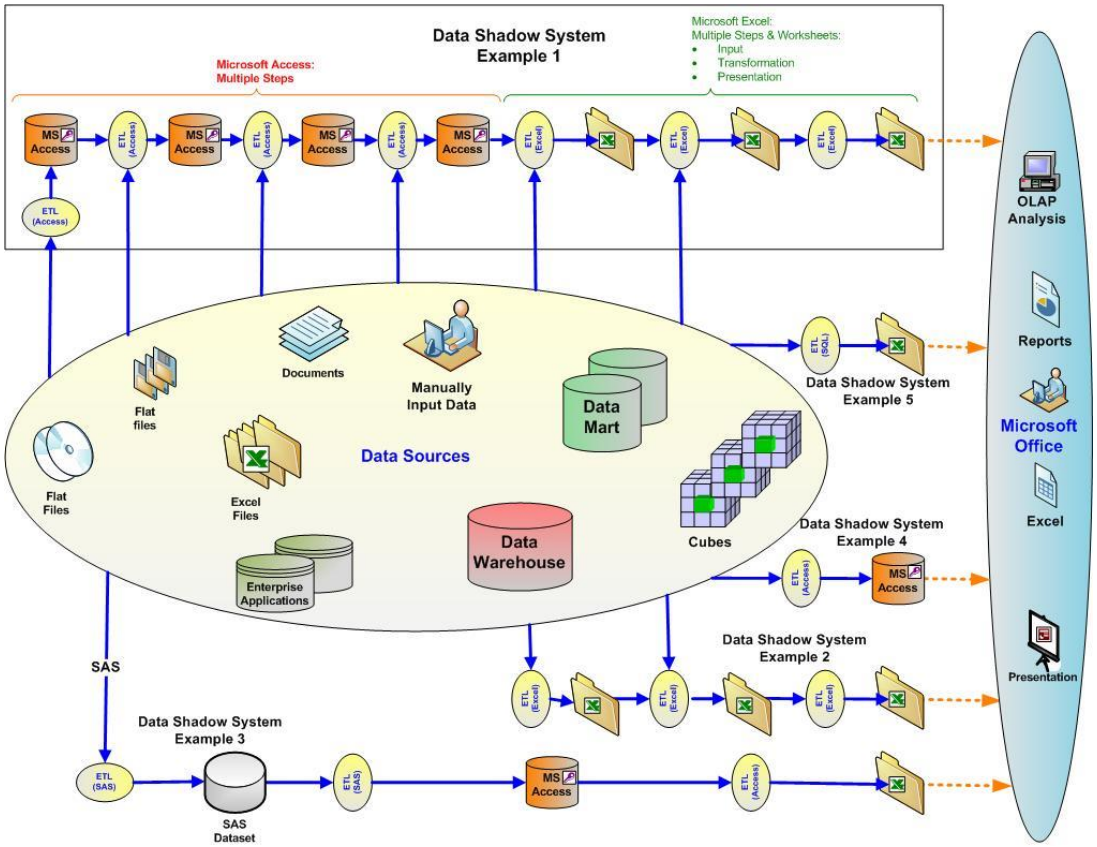


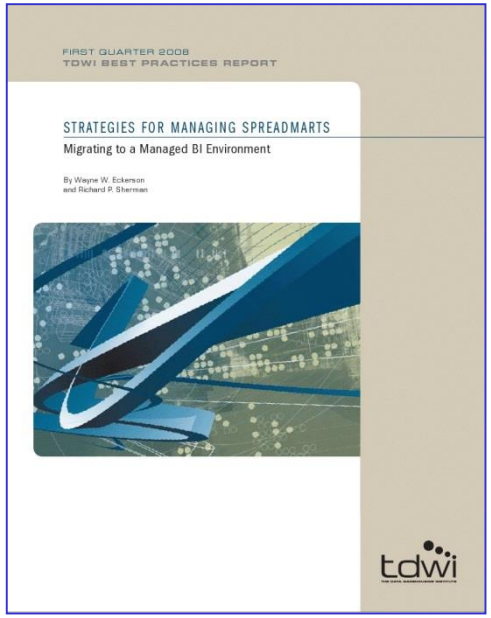
Sales	Bikes			Bikes Total	Grand Total
	Mountain Bikes	Road Bikes	Touring Bikes		
2006	$1,562,457	$4,967,887		$6,530,344	$6,530,344
1	$314,949	$1,476,749		$1,791,698	$1,791,698
2	$440,199	$1,573,813		$2,014,012	$2,014,012
3	$329,530	$1,067,304		$1,396,834	$1,396,834
4	$477,779	$850,020		$1,327,799	$1,327,799
2007	$3,989,638	$3,952,029	$1,417,435	$9,359,103	$9,359,103
1	$626,185	$787,346		$1,413,530	$1,413,530
2	$780,916	$843,055		$1,623,971	$1,623,971
3	$1,081,343	$1,028,987	$459,349	$2,569,678	$2,569,678
4	$1,501,195	$1,292,642	$958,086	$3,751,923	$3,751,923
2008	$3,814,691	$2,920,268	$2,427,366	$9,162,325	$9,162,325
Grand Total	$9,366,786	$11,840,184	$3,844,801	$25,051,771	$25,051,771

Information Gap
Current State of Reporting & Analysis

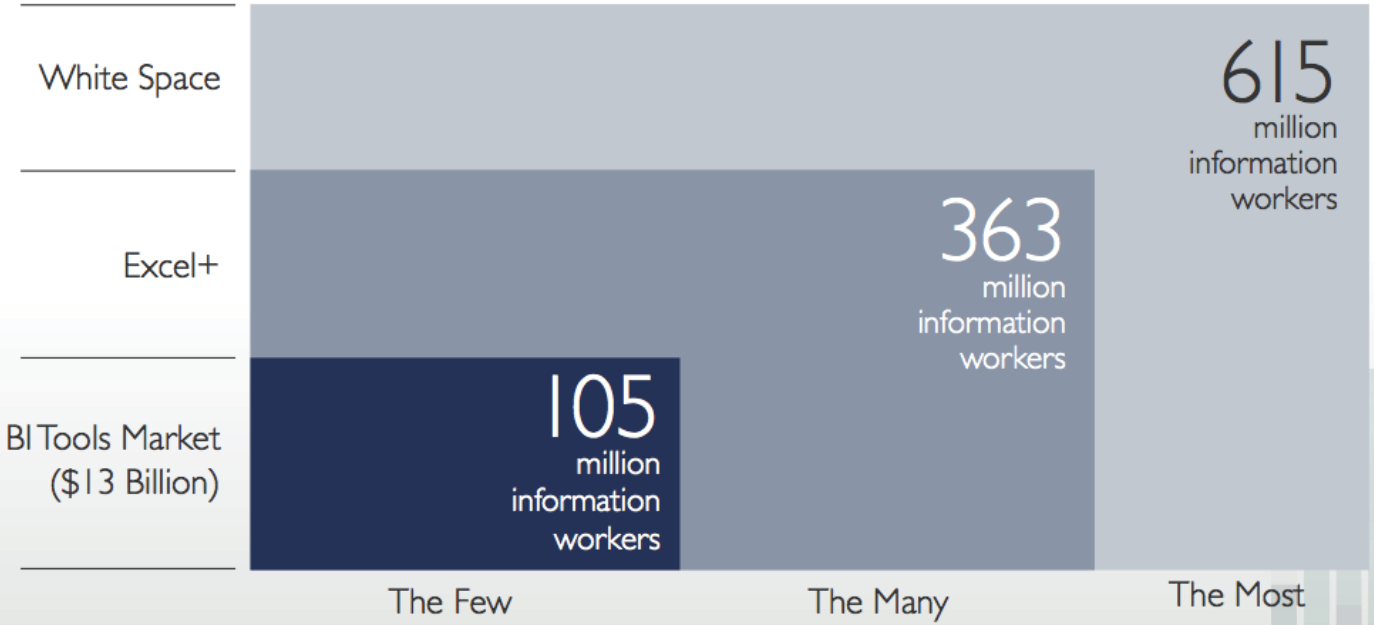
…but Spreadsheets become Data Shadow Systems or Spreadmarts!

- Data superglue





Market Opportunity



Source: 2012 BI Tools Market size per IDC; total information workers per Forrester and Forrsights. According to Forrester, there are expected to be 615mm information workers in 2013. According to Forrsights, as of August 2012, 17% (105mm workers) use BI tools and 59% (363mm workers) use spreadsheets.

Q2'14

BI Adoption Rate
BI Project Results

- 70% to 80% of corporate BI projects fail
  - ✓ IT mistake
    - Viewing BI as an engineering problems & tool focus
    - If we build it, they will come
  - ✓ Business mistake
    - Thinking BI gives answers rather than enabling analysis
    - We just need a dashboard
- Not technical but business failure
  - ✓ Not technical failure with almost all creating databases filled with data
  - ✓ Expectations and Usage Shortfall!!!
- Business - BI, data & expectations are moving targets
  - ✓ "BI has been in the top 10 issues for CIOs for the past ten years. It is a moving target and it changes every year as tools become more mature. BI as a competency has become an expectation for knowledge workers as well as executive managers. In other words the finishing line keeps moving forward"
    - Patrick Meehan, president & research director, Gartner's CIO Research group, Jan 10, 2011

BI Adoption Rate
BI Inhibitors

- IT Adoption
  - ✓ Complexity
  - ✓ Costs
  - ✓ Skills Shortage
  - ✓ Hype
  - ✓ Awareness Gap

- Business Usage
  - ✓ Complexity
  - ✓ Time to Implement
  - ✓ Data Silos
  - ✓ Costs
  - ✓ Awareness Gap

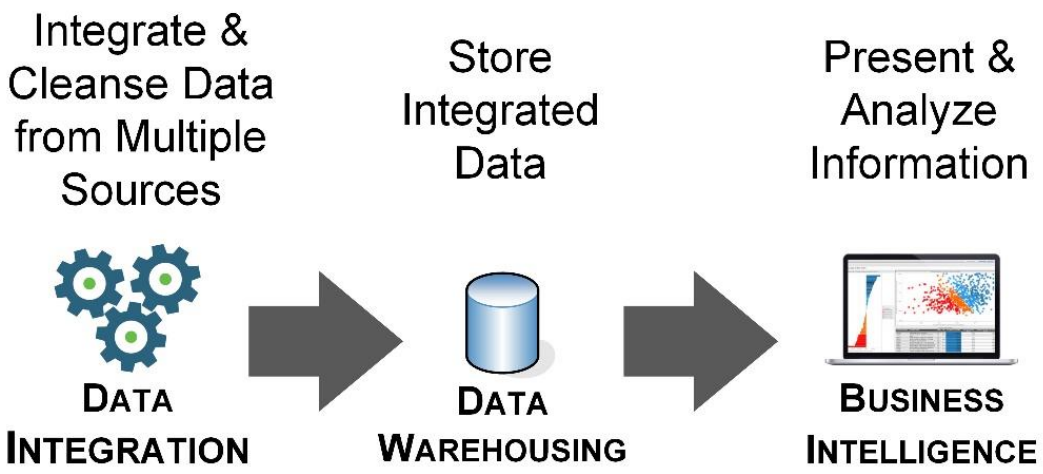
Why BI?

Why not just access the data where it is

Operational Data	Analytical Data
Structured for efficiently processing and managing business transactions and interactions	Structured for business people to understand and analyze
Live in the here and now. Record the business event as-is	Support the past, present and future. Tracks changes in dimensions – products, customers, businesses, geopolitical, account structures and organizational hierarchies – so that information can be examined as-is, as-was and as-will-be.
Typically contains a relatively short time span	Historical
Data is spread out over many source systems, making it hard to bring together and analyze	Very enterprise, no matter how large or small, must perform data integration

- Operation systems – outside of scope:
  - ✓ There are many business algorithms used to transform data to information outside of operations systems. Business groups transform the data into the context they need to perform their work.
  - ✓ Enterprise-wide and group-specific key performance indicators (KPIs) that need to be derived outside of operational systems.

Data Is Not Necessarily Information

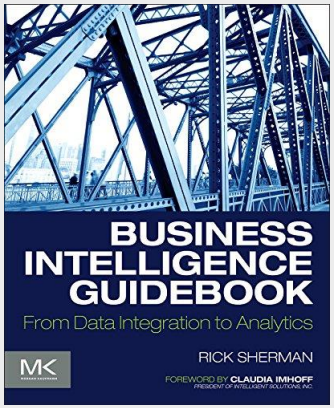


Information 5 C's

- Clean
- Consistent
- Conformed
- Current
- Comprehensive

Chapter 4:
Architecture Framework

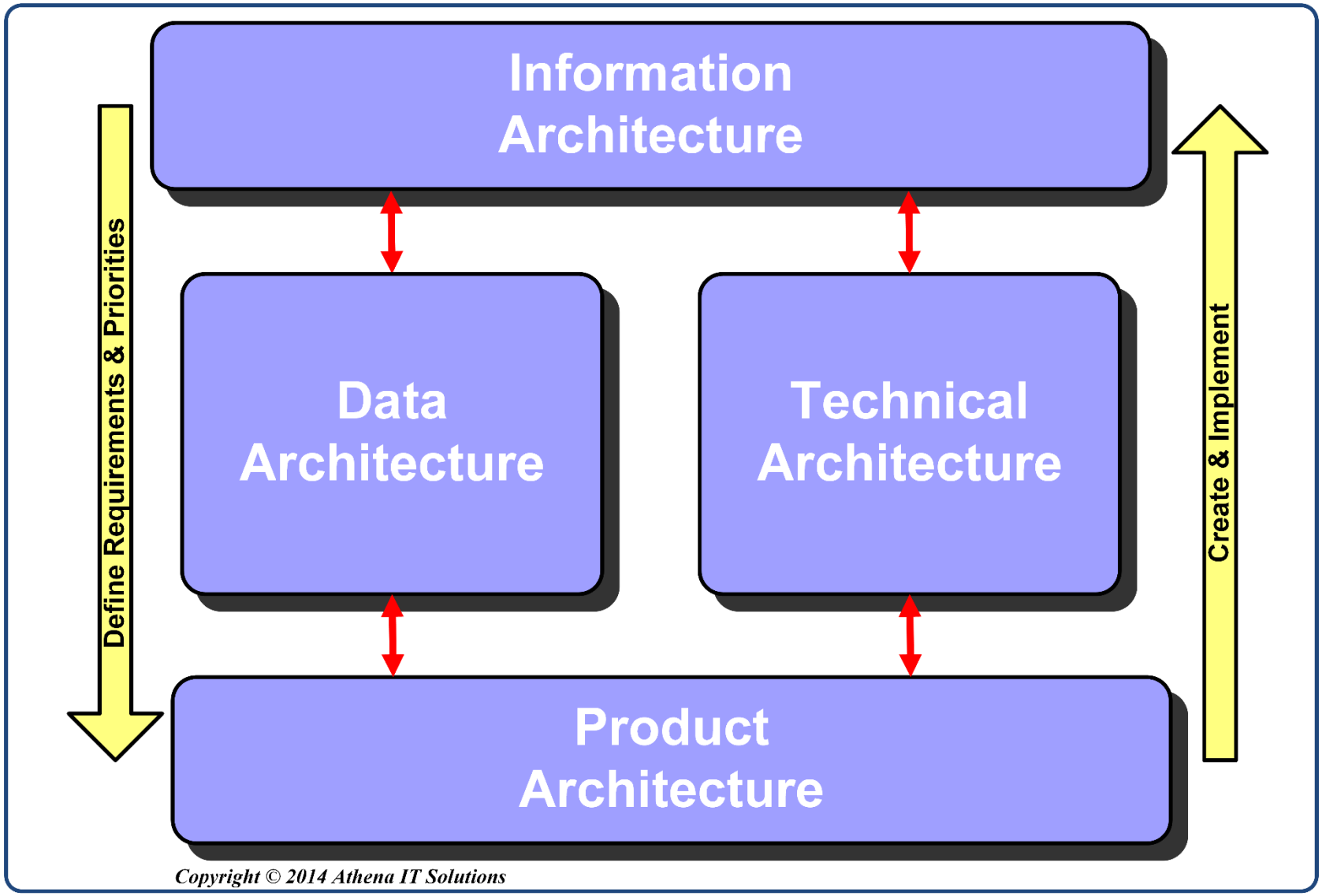
- 4 Architectures
- Information Architecture
- Data Architecture
- Technical & Product Architectures
- Definitions



Data, Integration & Application Silos Litter Enterprises

- Business Applications
  - ✓ Business Transactions
  - ✓ Business Functions
  - ✓ Business Processes
- Cloud applications
  - ✓ Support business processes
  - ✓ Create silos
- BI
  - ✓ Application-specific
  - ✓ DW
  - ✓ BI-specific independent
  - ✓ Report-specific
- Integration
  - ✓ DW feeds
  - ✓ BI feeds – often manual
  - ✓ Application integration

Architecture Framework:
Building Blocks



What, Who, Where & Why

QUESTION	DESCRIPTION
WHAT	• What business processes or functions are going to be supported
	• What types of analytics will be needed
	• What types of decisions are affected
WHO	• Who will have access - employees, customers, prospects, suppliers or other stakeholders
WHERE	• Where is the data now
	• Where will it be integrated
	• Where will it be consumed in analytical application
WHY	• Why will the BI solution(s) be built, i.e. what are the business and technical requirements

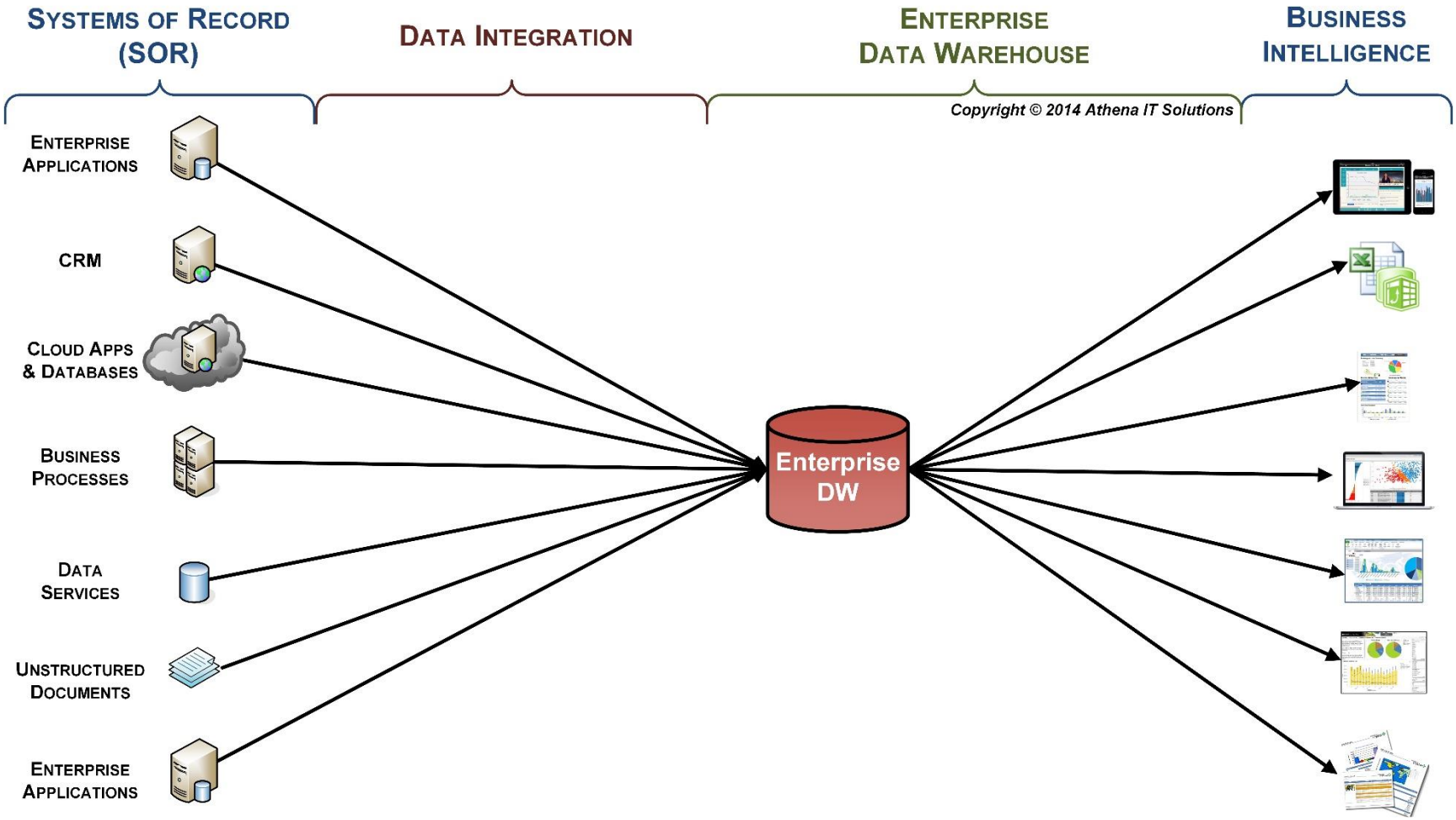
Product Evaluation Trap

Product Twilight Zone

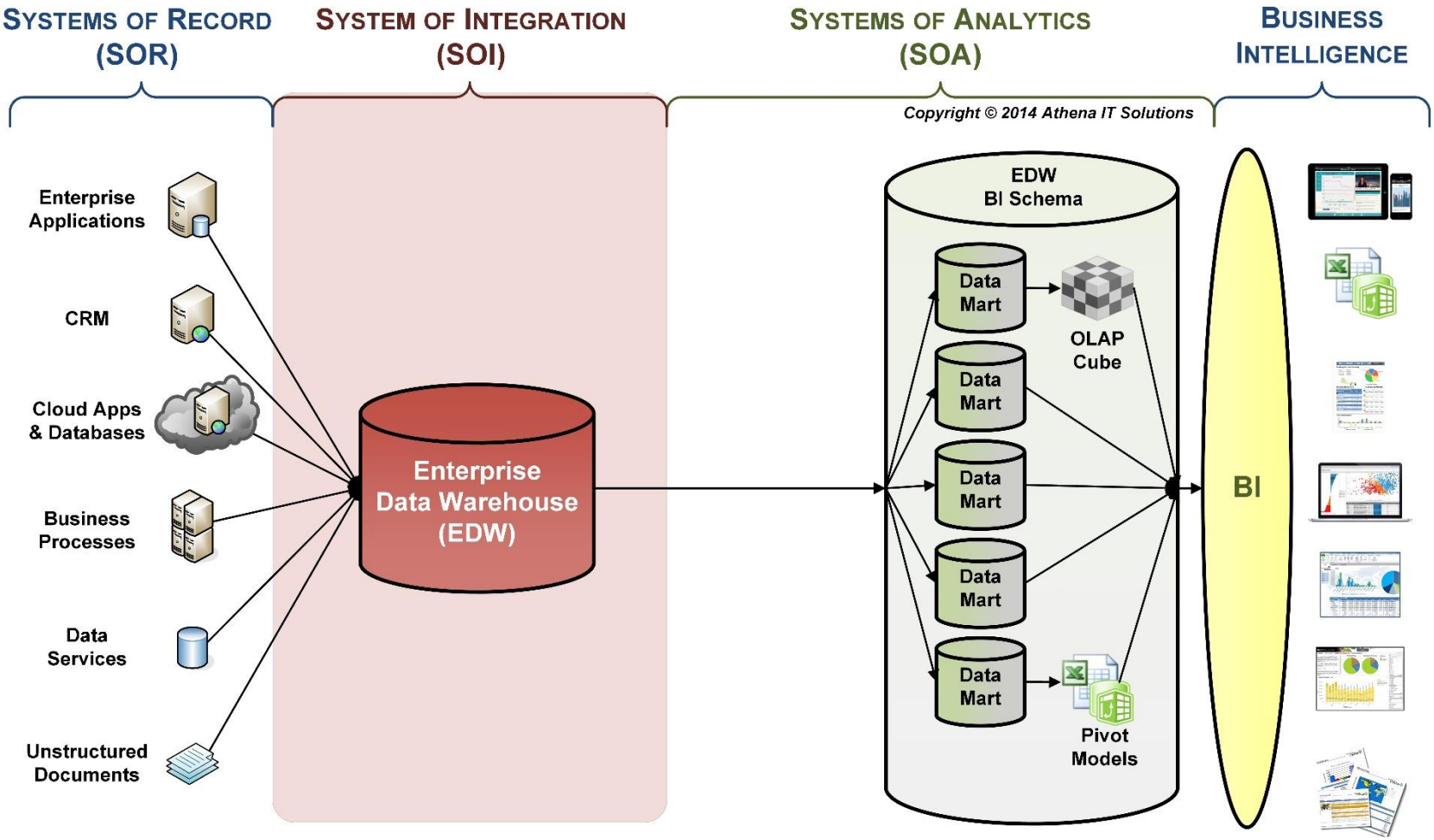
1) Evaluate and select “best” products
2) Implement BI solution without an information architecture
3) BI project is late and costs more than planned
4) BI solution fails to meet expectations and active adoption
5) Blame current products used
6) Evaluate and select a new “best” product
7) Go back to step 2

- Typical Enterprise
  - ✓ 6+ BI Tools
  - ✓ ETL to DW
  - ✓ >1 App Integration
  - ✓ Excel for BI
  - ✓ Manual coding for BI

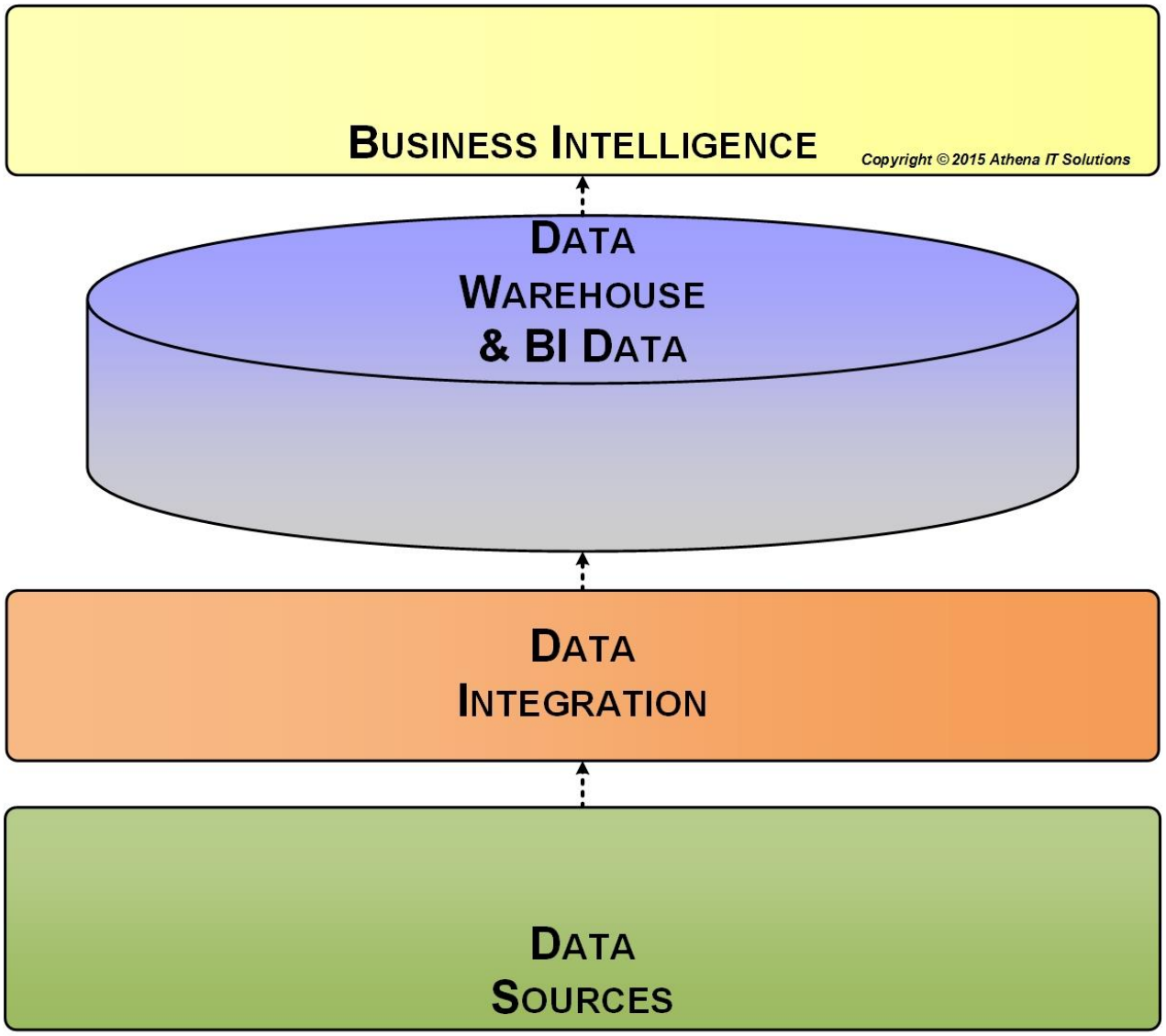
SOR - Inconsistency, Data Quality & Variety



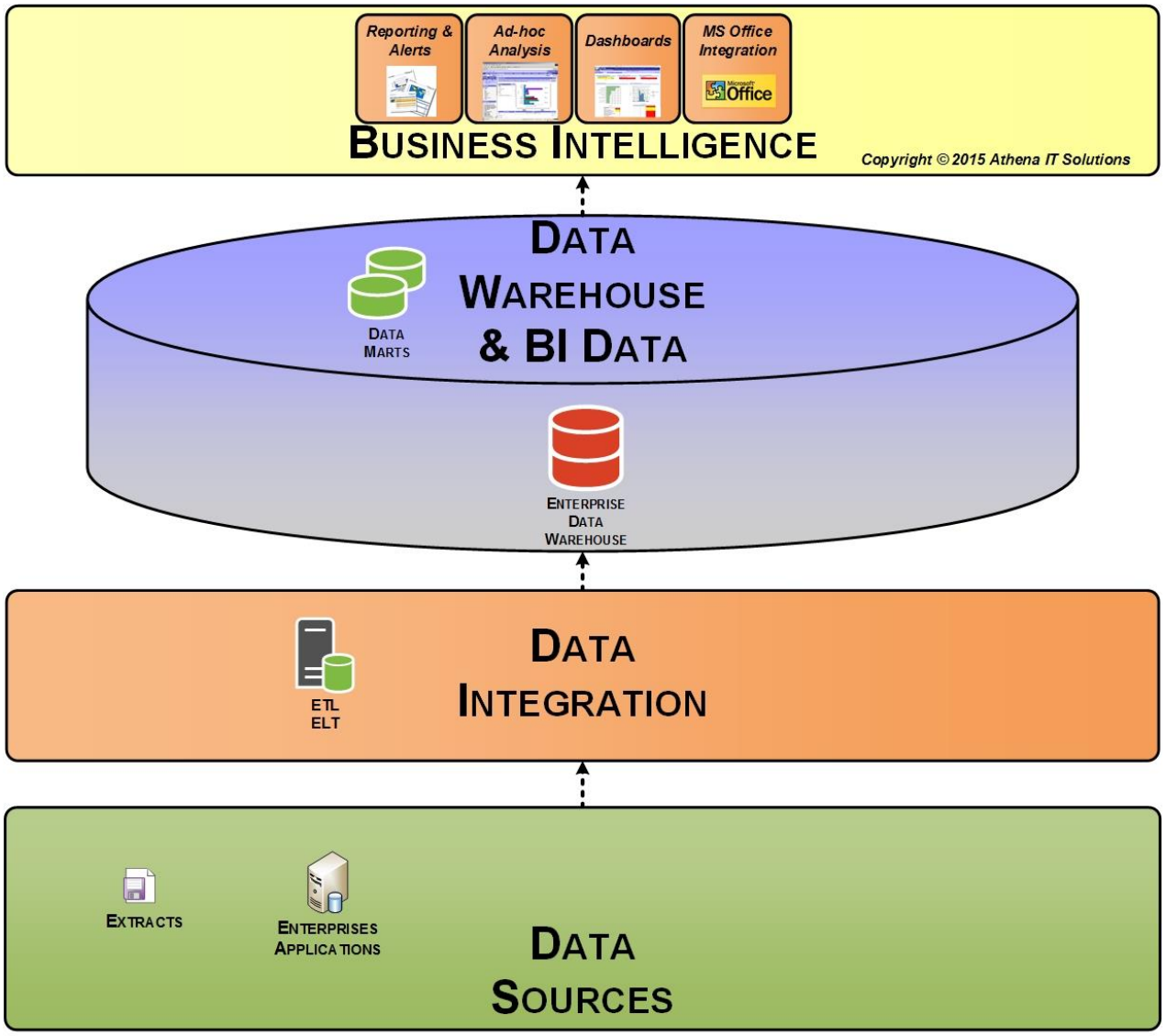
Specialization



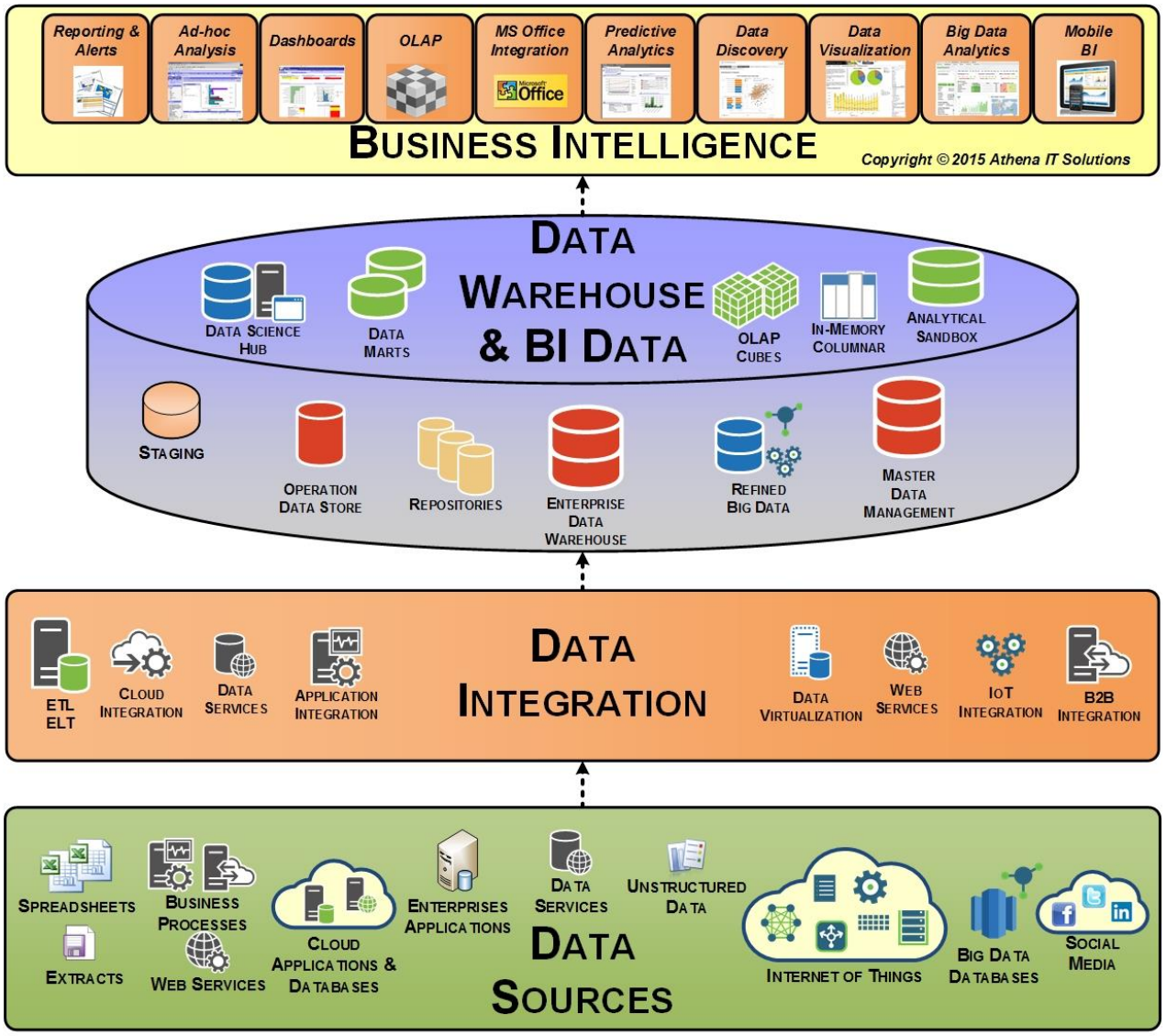
BI TA - Four Major Functional Layers



Technology Architecture
4 Layers - Traditional

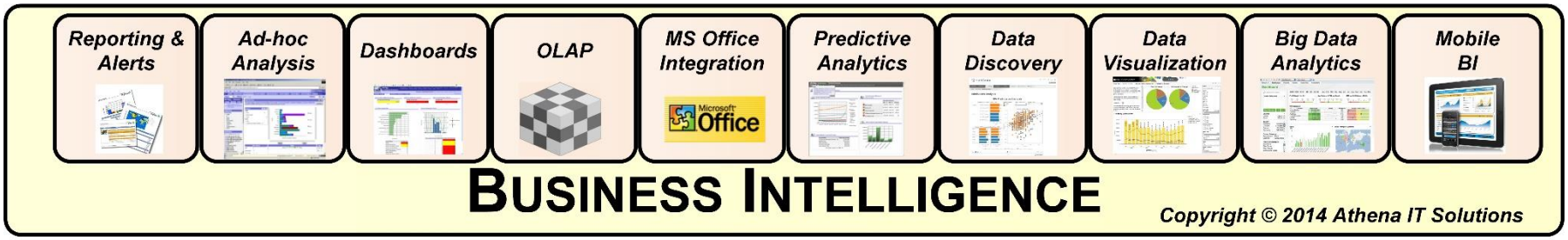


Technology Architecture
4 Layers - Expanded



BI Technical Architecture Categories

- BI Analytical Styles

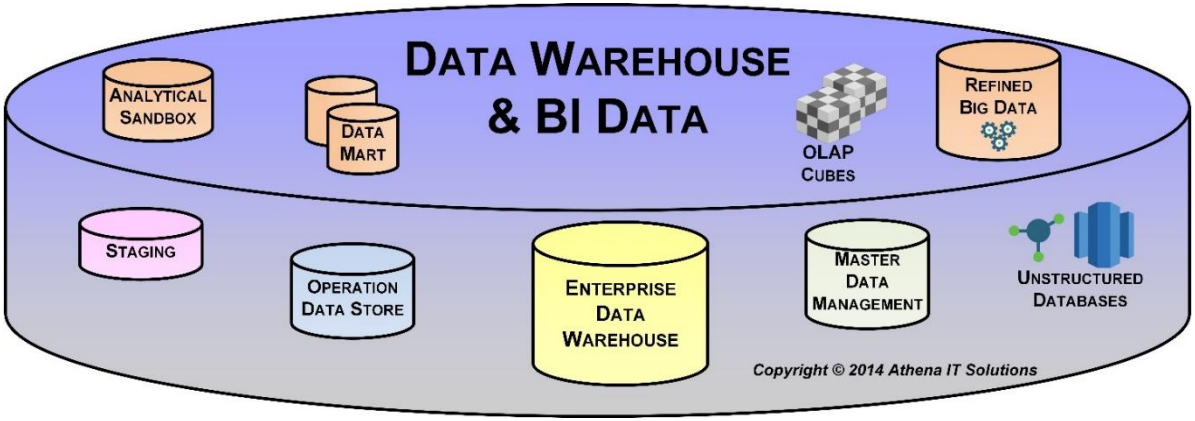


When designing the BI function of the technical architecture, a BI team needs to:

- Examine the business and data requirements
- Explore with business people what types of analytical processing they plan to perform
- Assess the analytical skills of the business users
- Select the BI functionality, i.e. type of capabilities and styles, needed

BI Technical Architecture Categories

- Data Warehouse and BI Data Stores

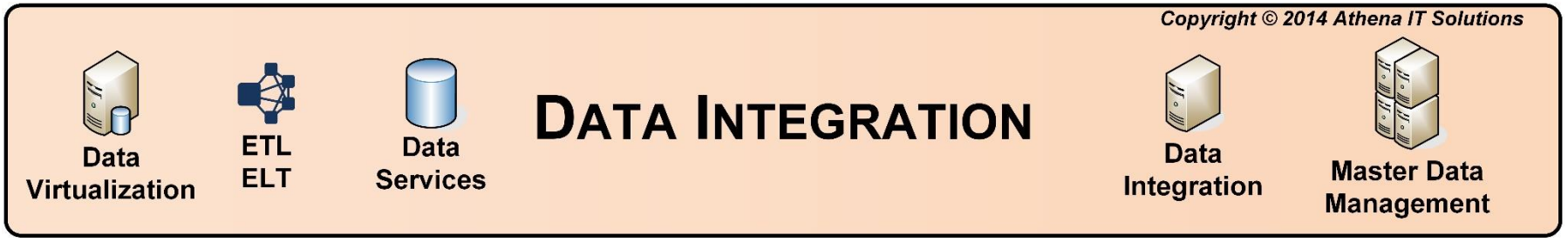


Options beyond relational databases:

- OLAP databases
- Columnar
- Massively Parallel Processing (MPP) databases
- Data virtualization
- In-database analytics
- In-memory analytics
- Cloud-based BI, DW or data integration
- NOSQL databases

BI Technical Architecture Categories

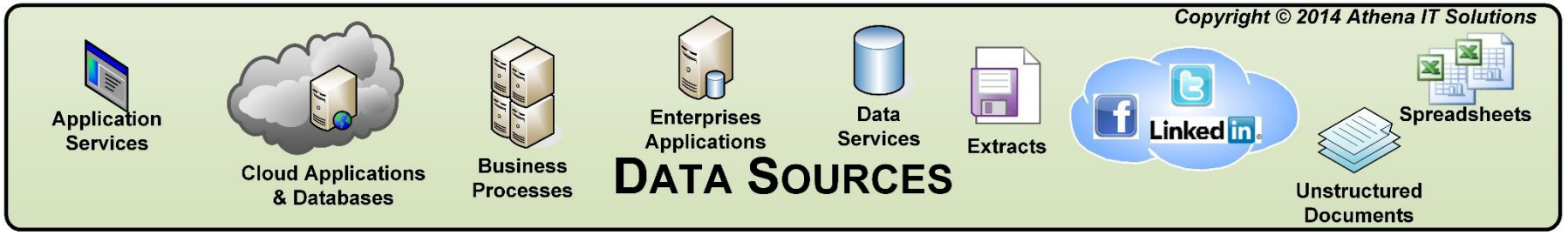
- Data Integration Layer



- Data integration has grown to encompass many different technologies and capabilities beyond ETL (Extract, Load & Transform)
- Different integration use cases
  - ✓ Classic ETL as just one use case of data integration.
- Different integration technologies
- Rise of integration suites
  - ✓ Many-to-many data movement or integration
  - ✓ Leverage appropriate technology based on need

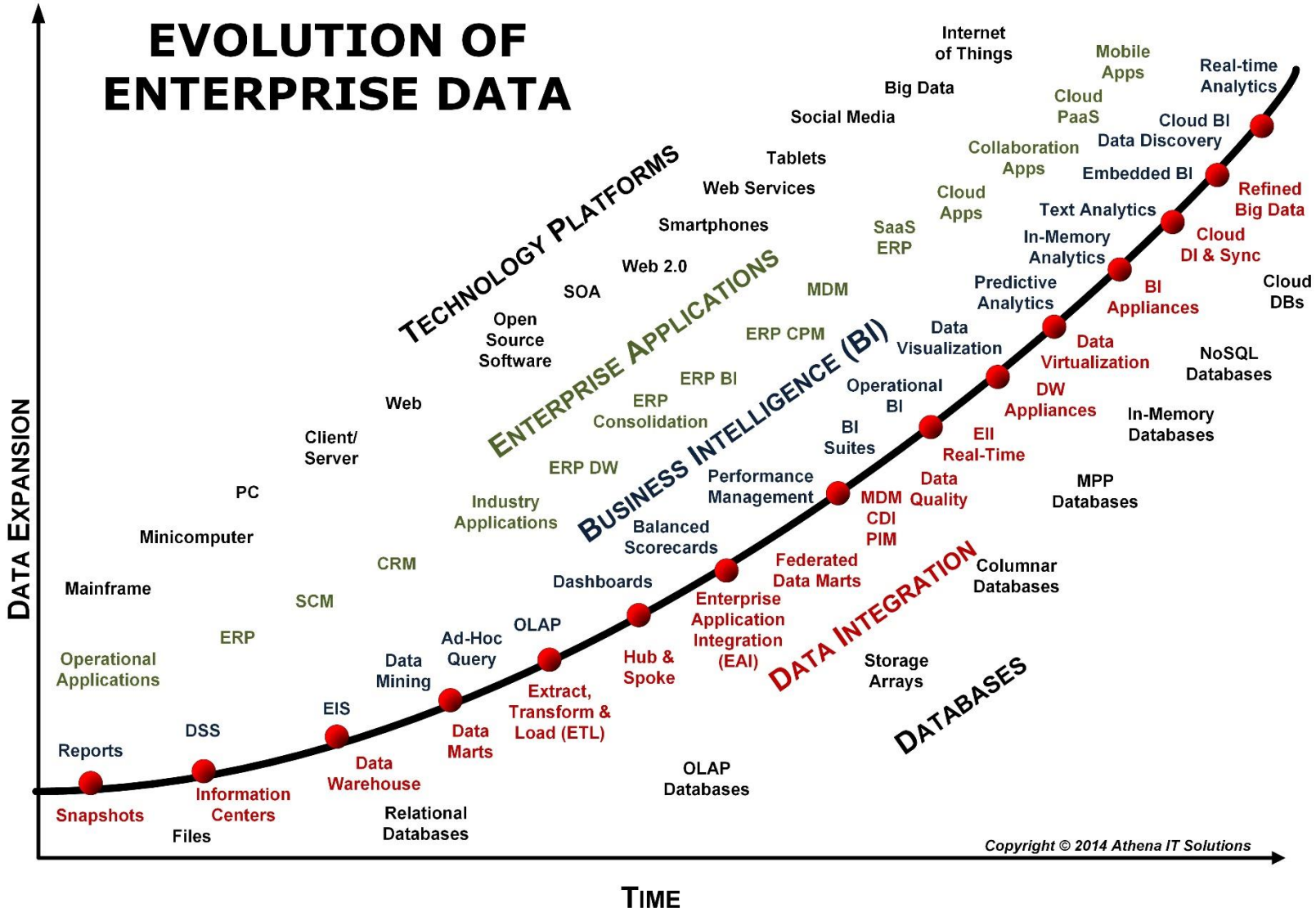
BI Technical Architecture Categories

- Data Sources Layer



- Data created in front-office and back-office systems, as well as data that is sourced or exchanged externally with customers, prospects, suppliers, partners and other stakeholders.
- Structured, unstructured and semi-structured data that may need to be integrated on a real-time basis.
- Data is now created in all types of business and people activities, thus greatly expanding the data volumes that need to be integrated.

BI Evolution



Architecture Framework:
Terminology Overload and Hype

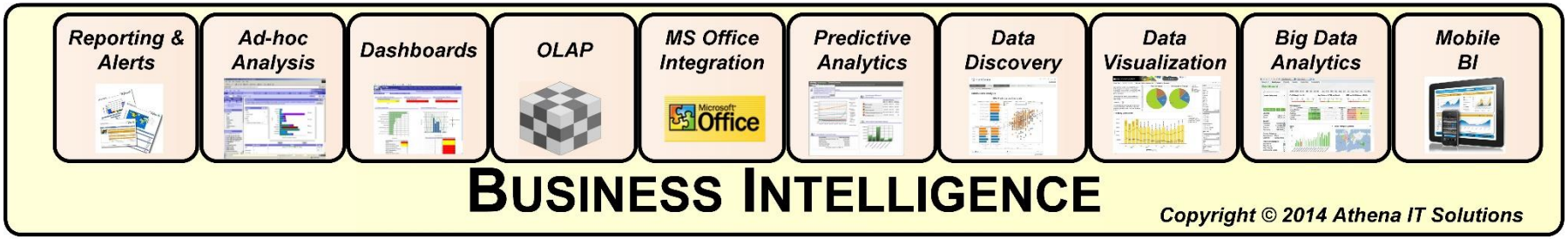
Definitions:
Business Intelligence (BI)

- Business intelligence (BI)
  - ✓ Transform data into business information & insights
  - ✓ Enables access and delivery of information to business users
  - ✓ Information presentation layer

ATHENA IT SOLUTIONS

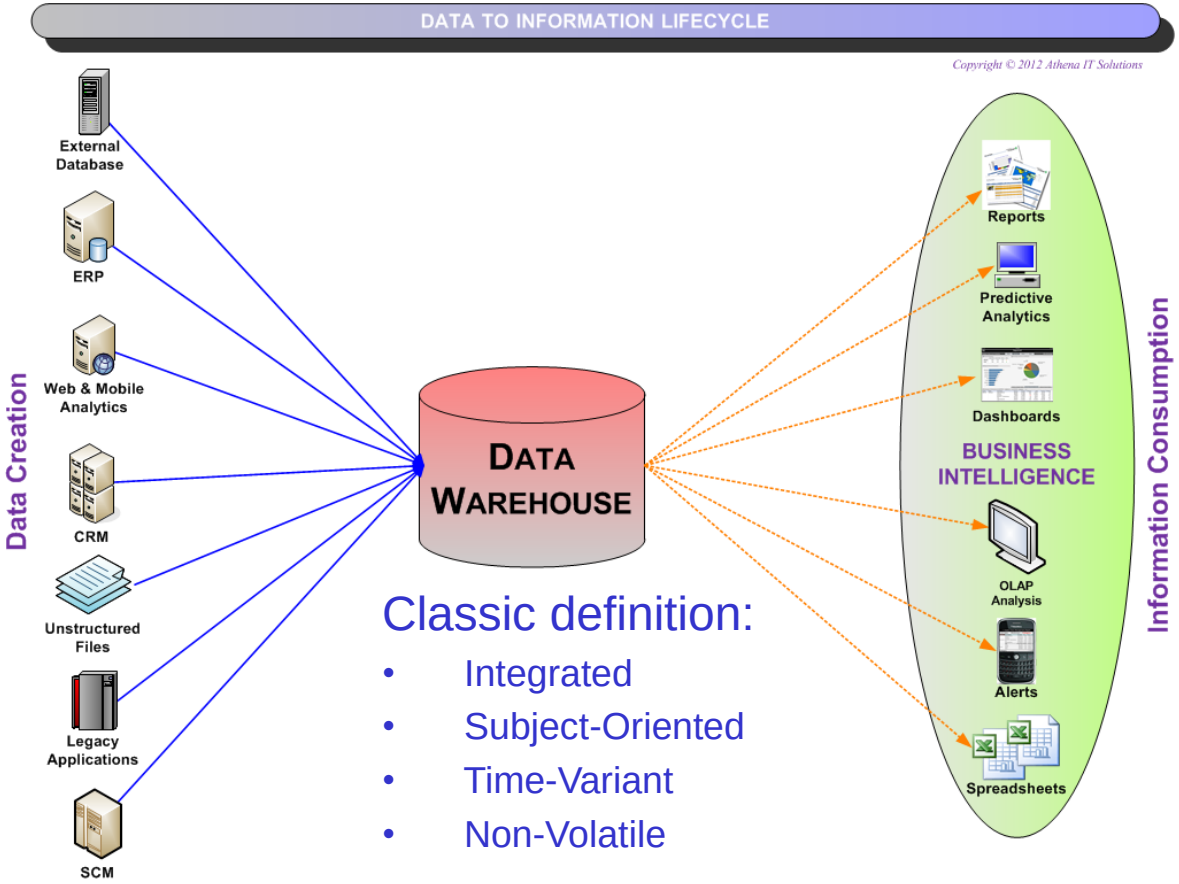
Definitions:
Business Intelligence (BI)

- Business intelligence (BI) Suites
  - ✓ Multiple BI presentation styles, tools & technologies packaged together as single product
  - ✓ Best of Breed vs All-in-One Product



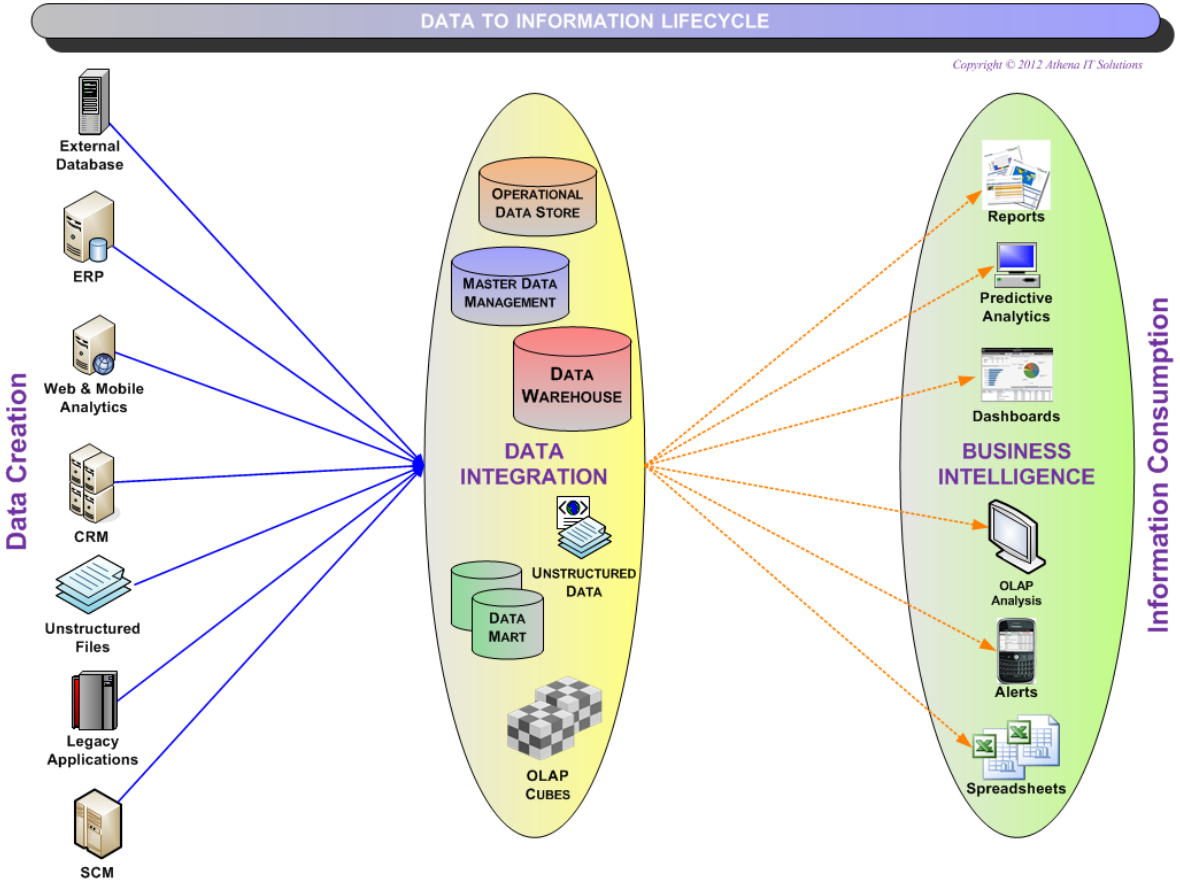
Definitions:
Data Warehousing (DW)

- Data Warehouse (DW)
  - ✓ In the beginning a place to store data for reporting, decision support systems (DSS) & BI in a centralized database



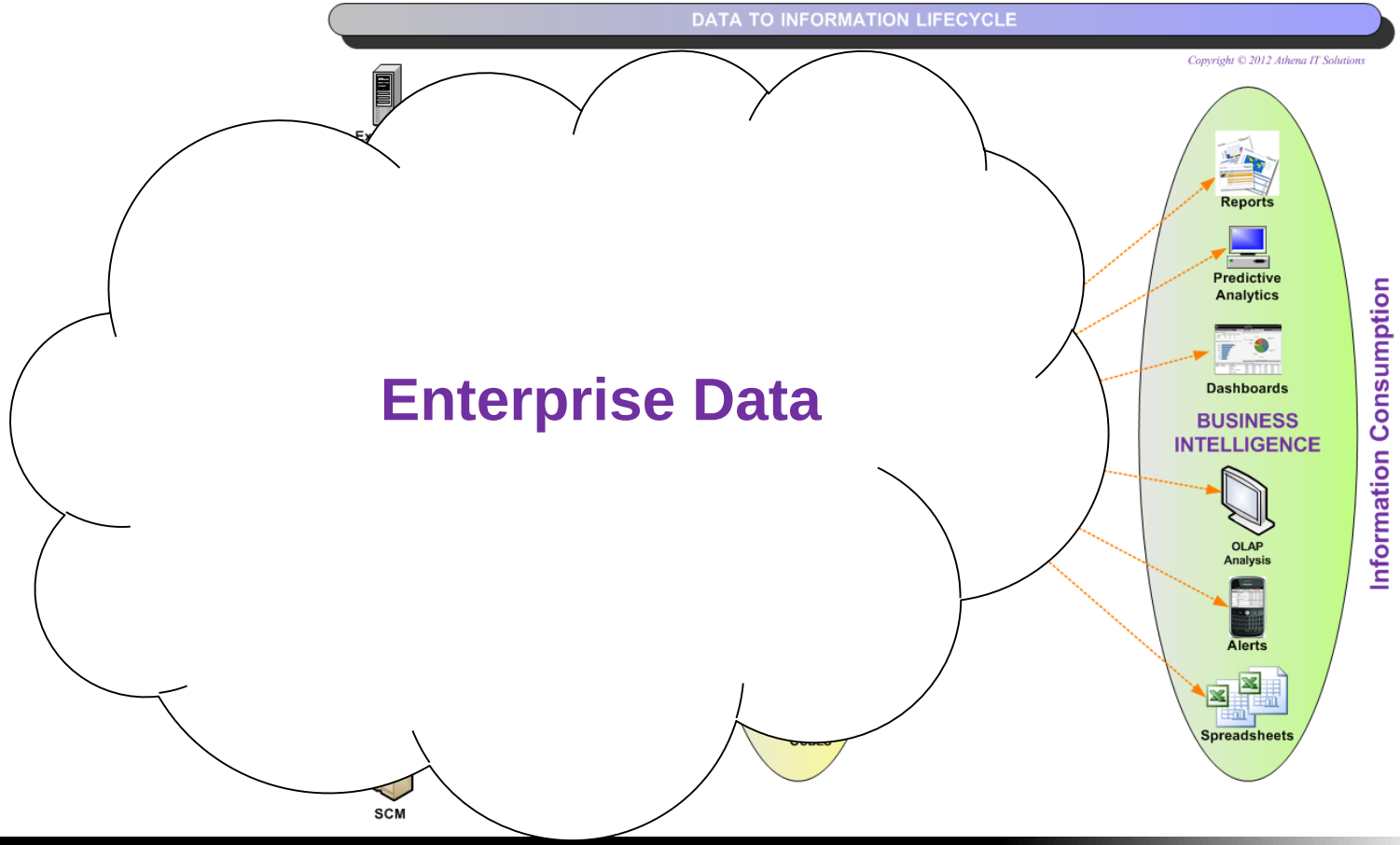
Definitions:
Data Warehousing (DW)

- Data Warehouse versus Data Warehousing
  - ✓ Various types of data stores are added to an overall information workflow
  - ✓ Data stores used for specific functions



Definitions:
Data Warehousing (DW)

- Business does not care about definitions or details
  - ✓ Wants comprehensive, consistent, conformed, clean & current data (5C's)
  - ✓ Will use best data available to them for decision-making

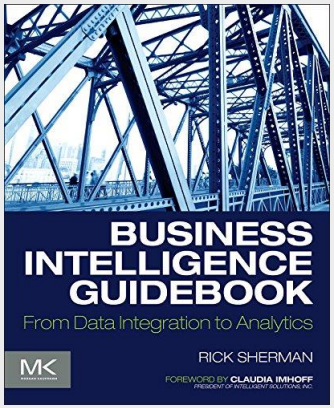


Architecture Framework:
Summary of Architecture Action Plan

ARCHITECTURE	DELIVERABLES
DATA	• Define what data is needed to meet business user needs. • Examine the completeness and correctness of source systems that are needed to obtain data. • Identify the data facts and dimensions. • Define the logical data models. • Establish preliminary aggregation plan.
INFORMATION	• Define the framework for the transformation of data into information from the source systems to information used by the business users. • Recommend the data stages necessary for data transform and information access. • Develop source-to-target data mapping for each data stage. • Review data quality procedures and reconciliation techniques. • Define the physical data models.
TECHNOLOGY	• Define technical functionality used to build a data warehousing and business intelligence environment. • Identify available technologies available and review tradeoffs associated between any overlapping or competing technologies. • Review current technical environment and company's strategic technical directions. • Recommend technologies to be used to meet your business requirements and implementation plan.
PRODUCT	• List product categories needed to implement the technology architecture. • Review tradeoffs between overlapping or competing product categories. • Outline implementation of product architecture in stages. • Identify short list of products in each of these categories. • Recommend products and implementation schedule.

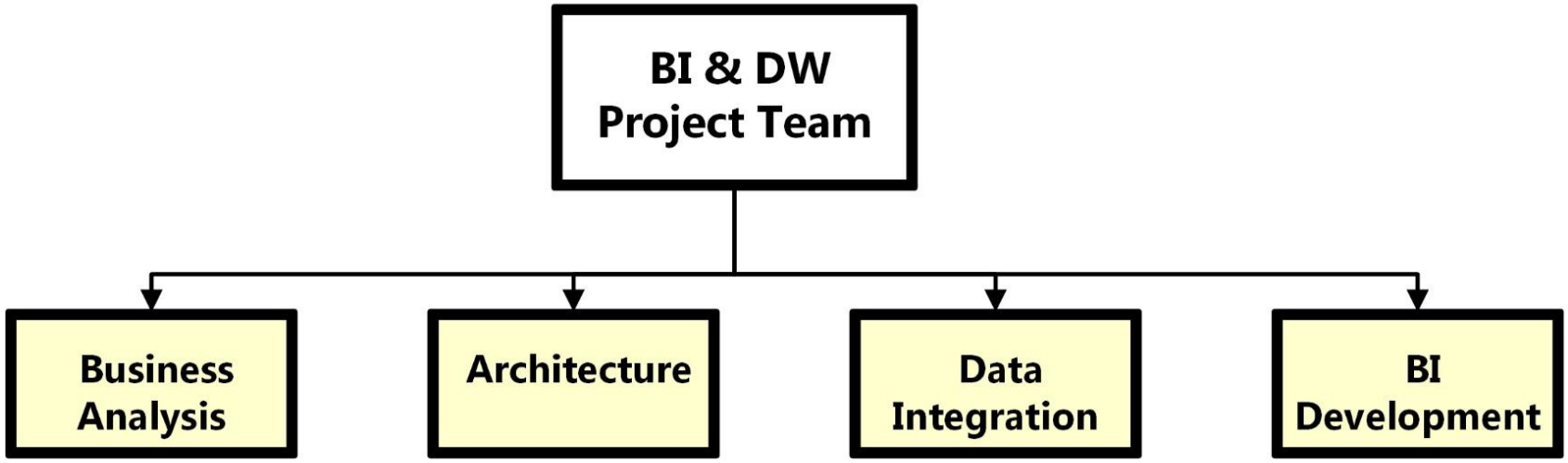
Chapter 19:
Foundational Data Modeling

- Introduction
- ER Modeling
- Normalization

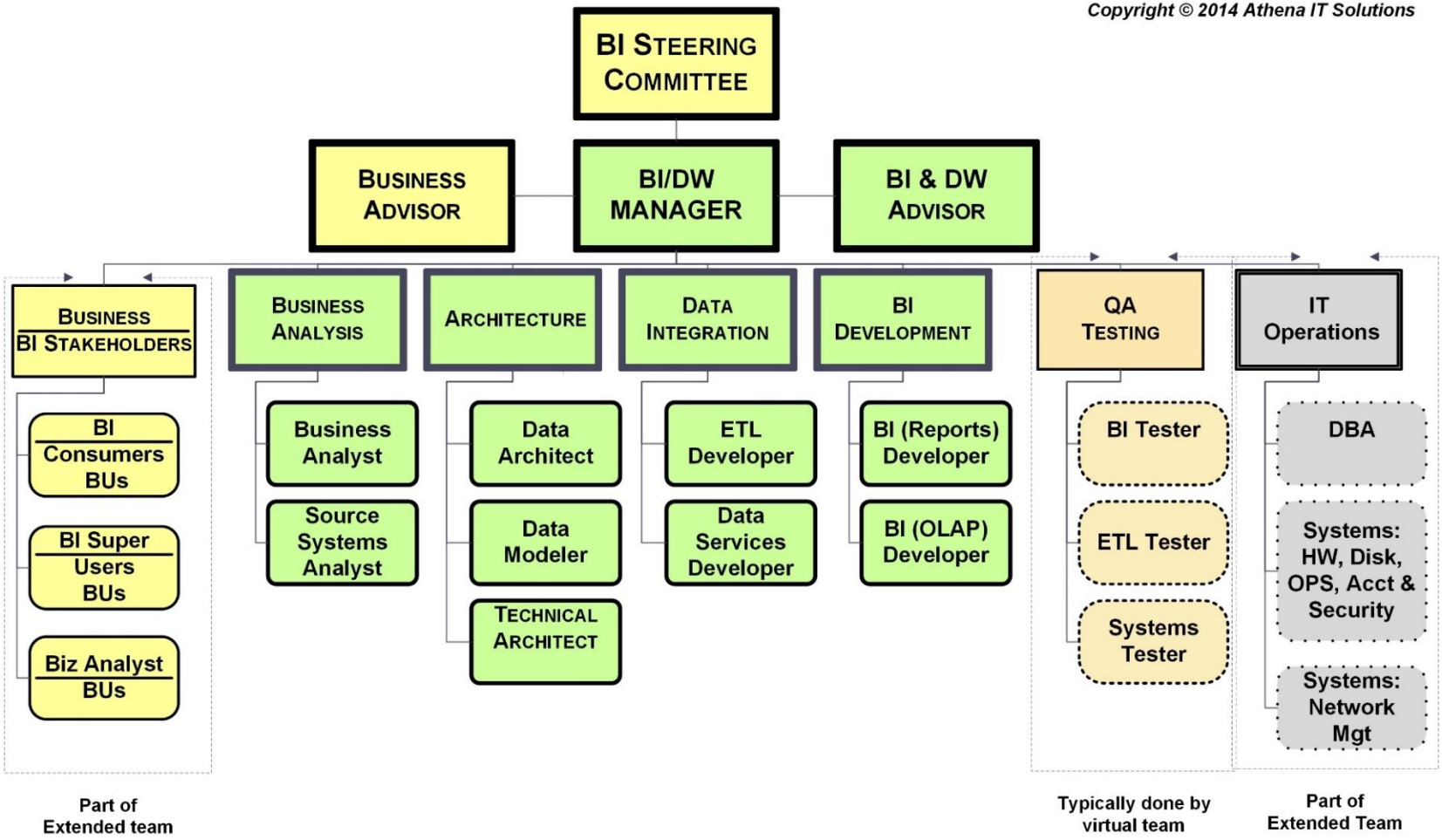


BI Project Team

- BI/DW Team by Function

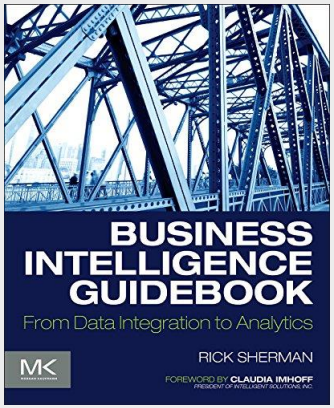


BI Organization – Roles (Not Necessarily Individual Job Positions)



Chapter 9:
Foundational Data Modeling

- Introduction
- ER Modeling
- Normalization



Data Modeling:
Different Types of Models

Foundational Data Modeling:
Introduction

Data Model is a specification of data structures & business rules representing business requirements.

Data Modeling is a structured approach used to identify the data components of an information system's specifications.

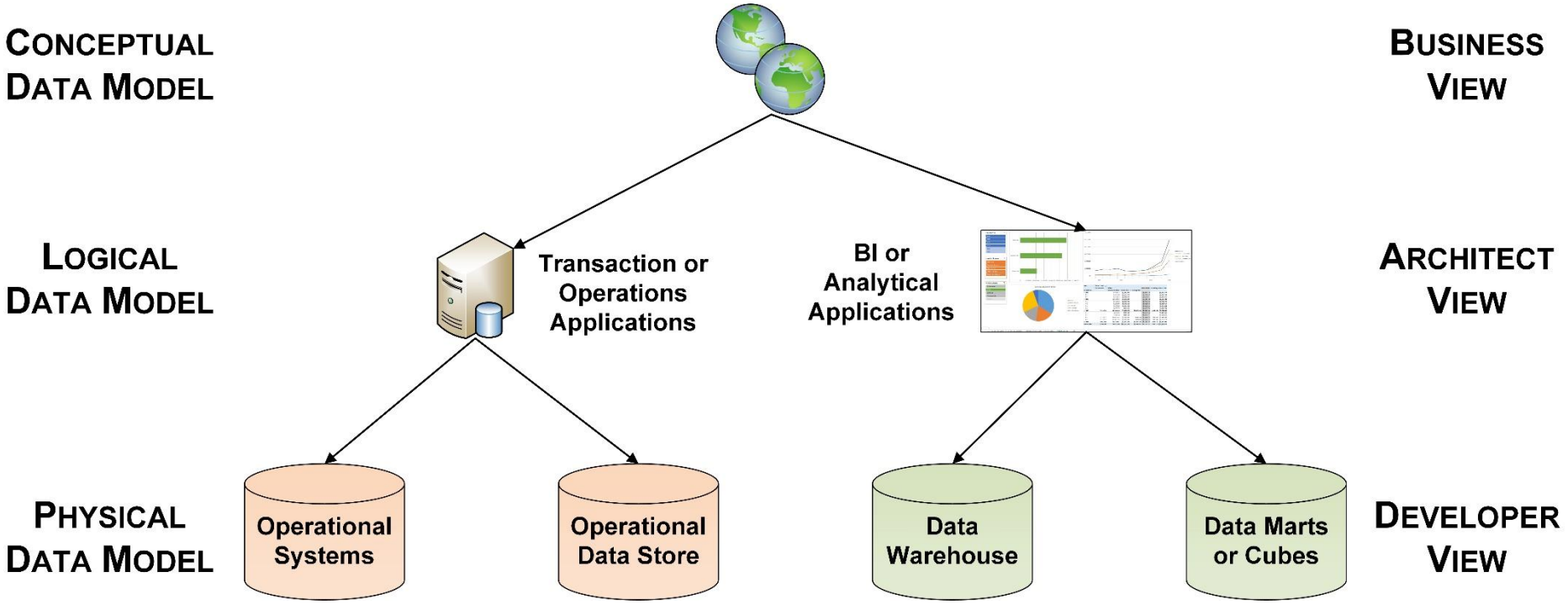
Purpose:

- A method to communicate using a visual representation the information that is needed, collected and used by an organization
- Data design specification for an IT application

Other types of modeling:

- Business Process, Functional & Object Modeling

Foundational Data Modeling:
Three Levels of Data Models



Foundational Data Modeling:
Conceptual Data Model

Conceptual Data Model (CDM) is a structured business view of the data required to support current business processes, business events, and related performance measurements.

- Single integrated data view identifying the structure of business functions rather than data processing flow or its physical characteristics

- Characteristics
  - ✓ Represents overall logical structure of the data
  - ✓ Independent of software or physical data storage structure
  - ✓ Often contains objects not implemented in physical
  - ✓ Represents data needed to run an enterprise or a business activity

Foundational Data Modeling:
Logical Data Model

Logical Data Model (LDM) builds upon the business requirements and includes a further level of detail that supports both the business and system requirements.

- Business rules are incorporated into LDM & it loses some of the 'generalities' from enterprise CDM
- Characteristics
  - ✓ Independent of specific software and data storage structure
  - ✓ Includes more specific entities and attributes
  - ✓ Includes business rules and relationships
  - ✓ Includes foreign keys, alternate keys, and inversion entries
- The **logical data model** is the next layer down, and is the one we're most involved in when designing the BI application.
- It helps us understand the details of the data, but not how they are implemented.

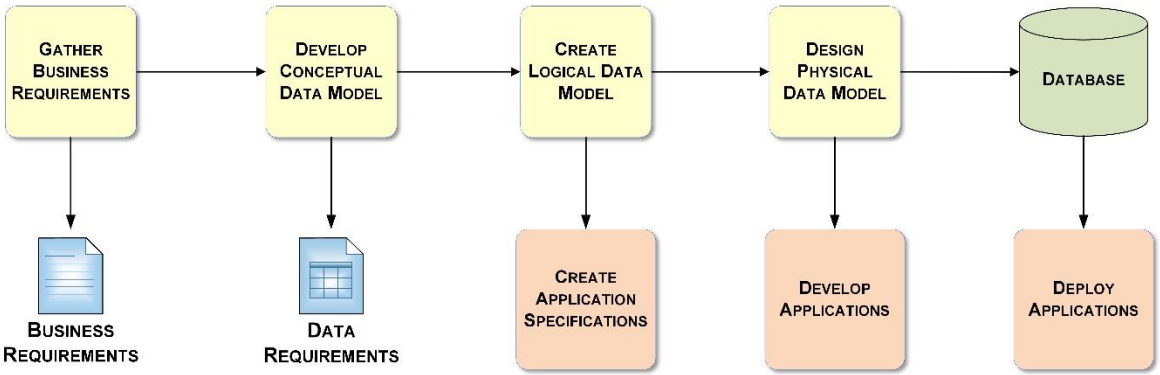
Foundational Data Modeling:
Physical Data Model

Physical Data Models (PDM) are specific to the software and performance constraints of the database management system used in the implementation.

- Both software and data storage structures are considered and the model is often modified to meet performance or physical constraints
- Characteristics
  - ✓ Dependent on specific software and data storage structure
  - ✓ Includes tables and columns
  - ✓ Includes physical database objects (triggers, stored procedures, tablespaces, indexes, partitions, materialized views, etc.)
  - ✓ Includes referential integrity rules that restrict relationships between tables
- Physical will vary based on database technology used for BI while Logical should remain the same

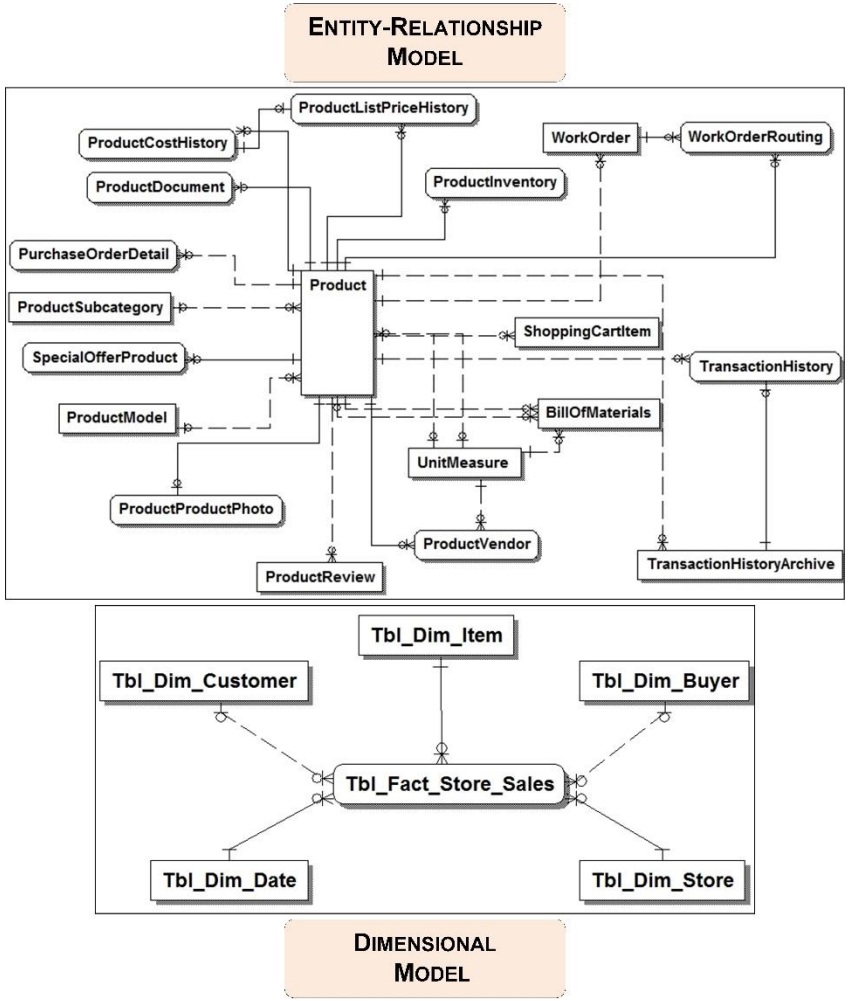
Foundational Data Modeling:
Workflow

- Gather business requirements
  - ✓ Analyze the data needed by the business requirements
  - ✓ Identify data relationships
- Create the various data models needed
  - ✓ Conceptual
  - ✓ Logical
  - ✓ Physical
- Support the application development
  - ✓ Create application specifications
  - ✓ Develop applications
  - ✓ Deploy applications



Foundational Data Modeling:
Where are data models used

- Use Cases:
  - ✓ Transactional processing or operational systems
  - ✓ BI applications
- Relational databases
  - ✓ Transactional systems
  - ✓ DW portion of BI Architecture
- Modeling Approaches:
  - ✓ ER Modeling
  - ✓ Dimensional Modeling

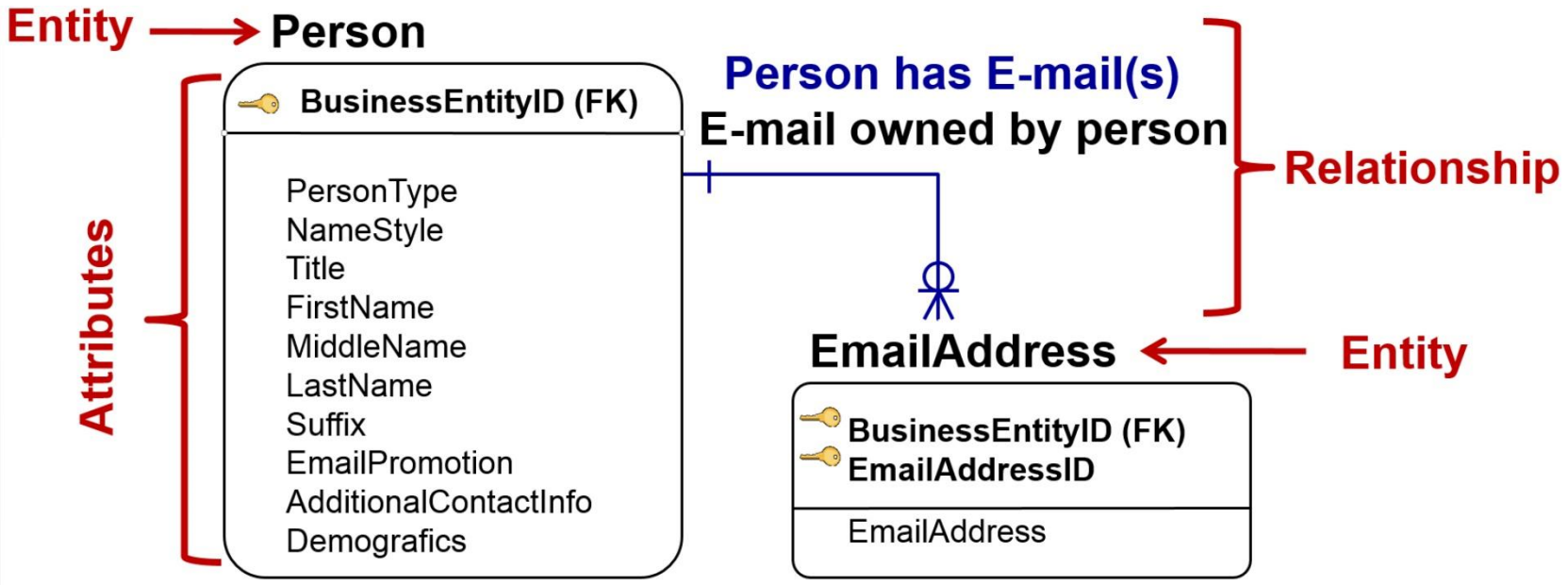


Foundational Data Modeling:
E-R Modeling

- Entity-Relationship (ER) modeling:
  - ✓ Logical data modeling technique
  - ✓ Also referred to as 3NF (3rd normal form) or normalized model
  - ✓ Sometimes referred to as relational model but that is incorrect
- Primarily used in transactional processing applications implemented using relational databases
  - ✓ Enterprise Resource Planning (ERP) systems often have thousands of tables
  - ✓ Used to eliminate data redundancy
  - ✓ Enable fast loading & updates

E-R Modeling
Building Blocks

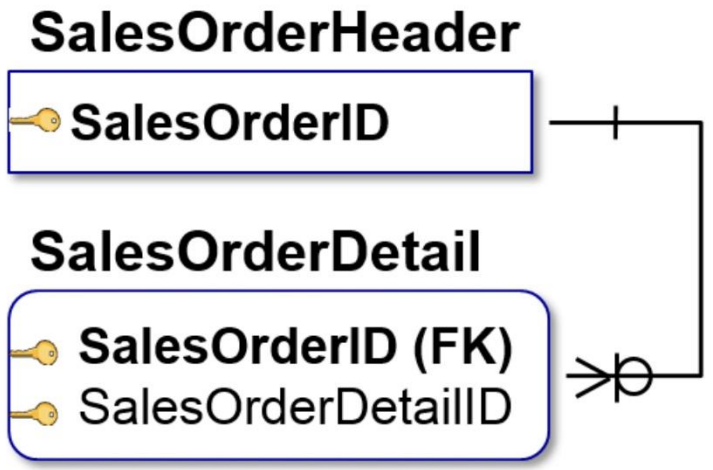
- **Entity** is a person, place, thing, event, or concept about which the business keeps data
- **Relationship** is a logical link between two entities that represents a business rule or constraint
- **Attribute** is a distinct characteristic of an entity for which data is maintained



Entity & Attribute Types

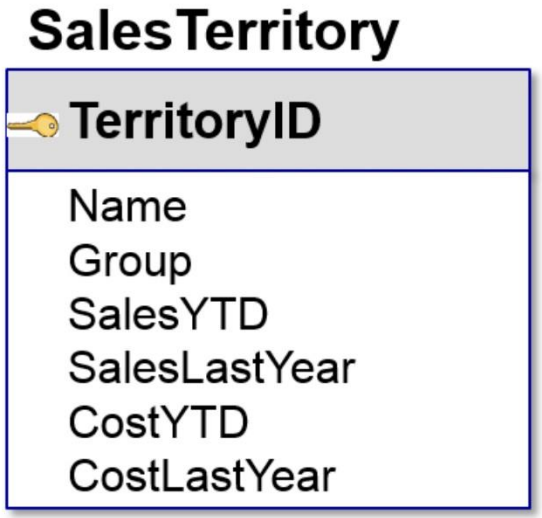
Two Types Of Entities:

- **Independent**
- **Dependent**

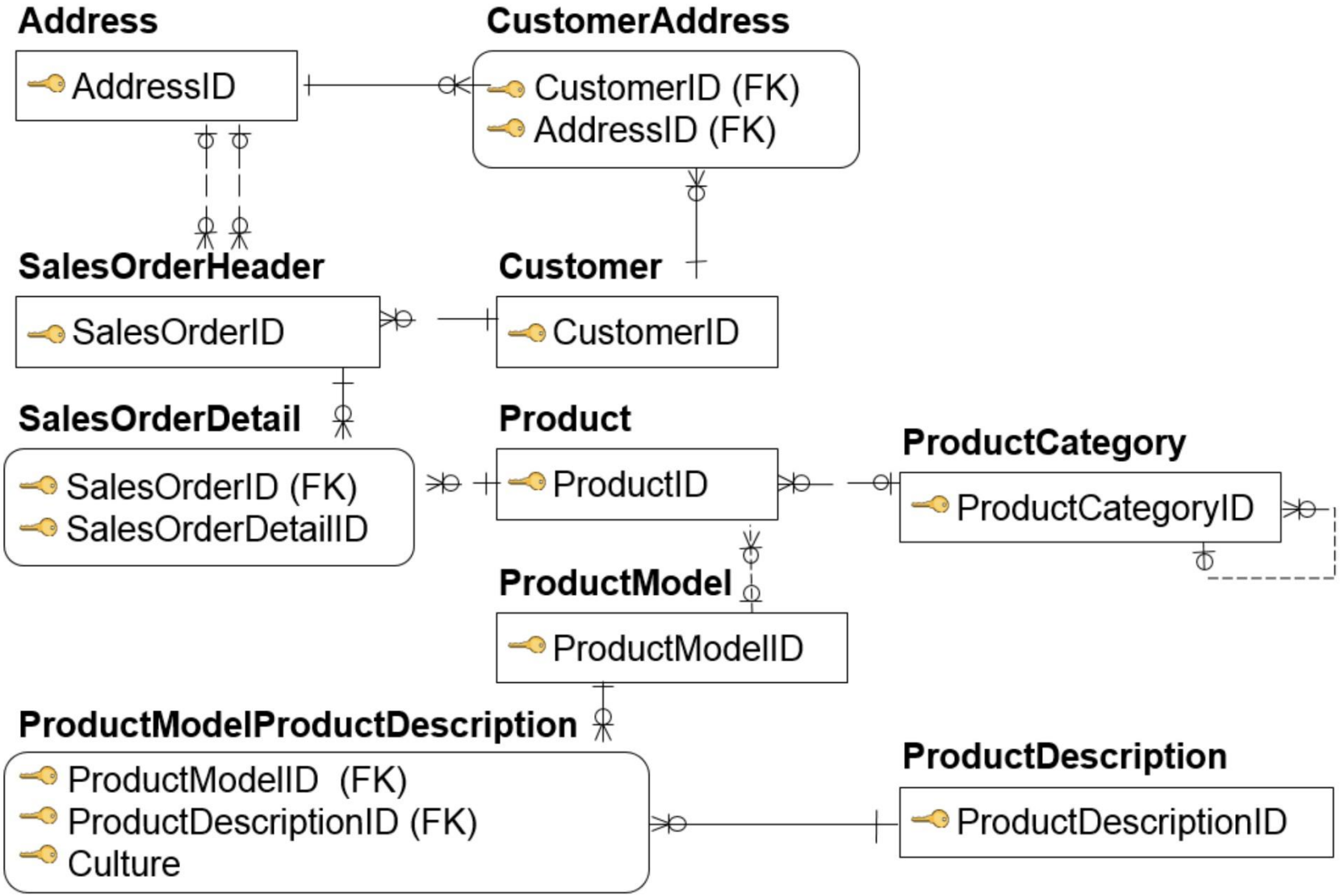


Two types of attributes:

- **Key**
- **Non-Key**



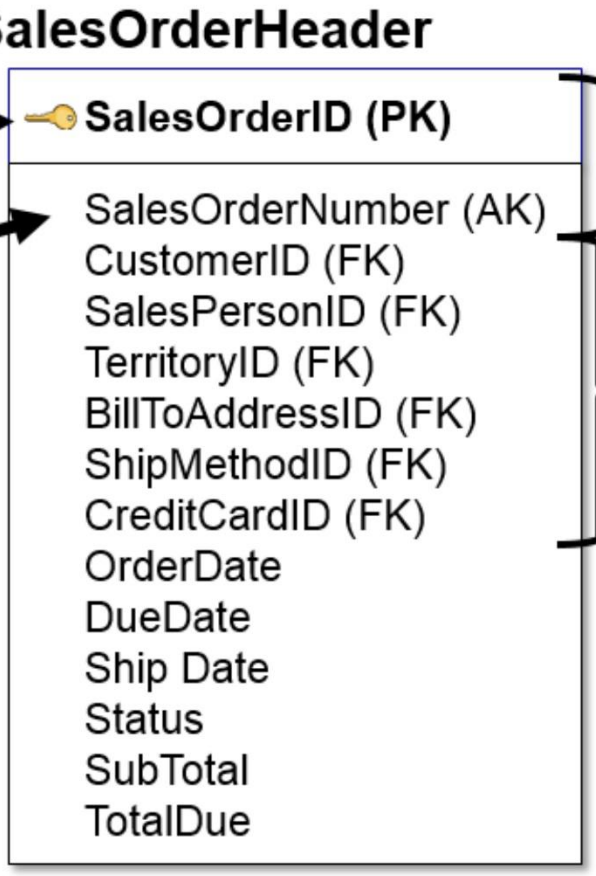
E-R Modeling:
ER Model Example



E-R Modeling:
Keys

PRIMARY KEY: An attribute or group of attributes that uniquely identifies an instance of the entity

ALTERNATE KEY: An attribute or set of attributes which uniquely identifies each instance, but is not chosen as the primary key



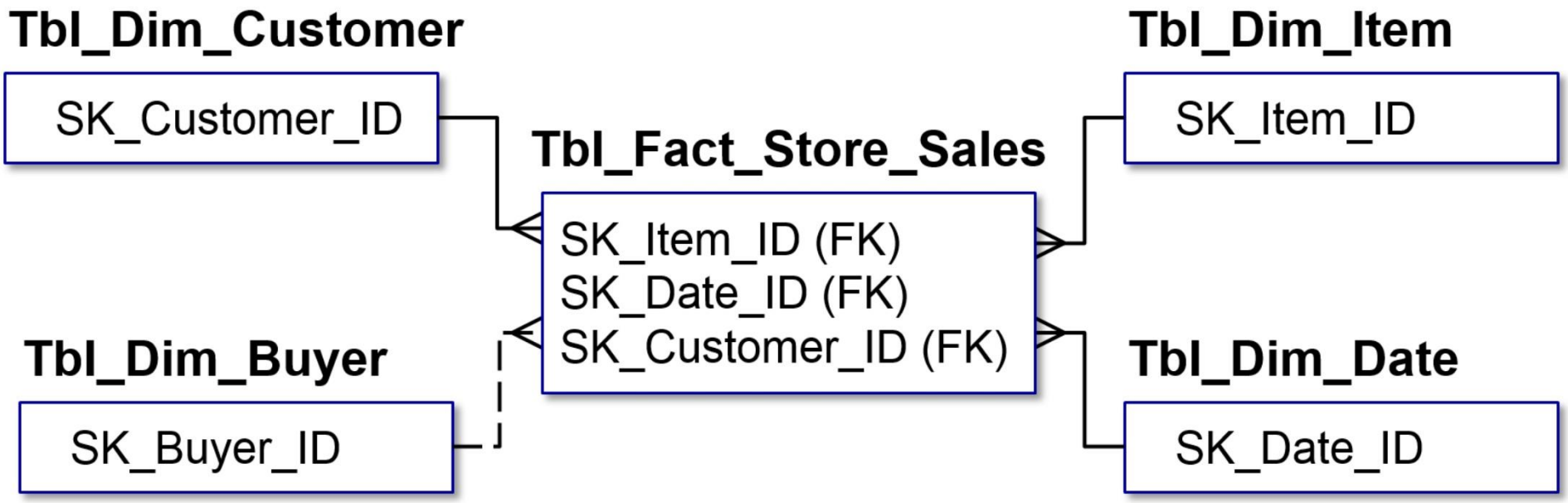
CANDIDATE KEYS: An attribute or group of attributes which serves to uniquely identify each instance of an entity

FOREIGN KEYS: A primary key of a parent entity that is contributed to a child entity across a relationship

E-R Modeling:
Referential Integrity

Referential Integrity rules determine actions taken when a parent or child row is inserted, updated or deleted

- None of these actions should violate relationships
- How is it enforced?
  - ✓ Through relational database constraints
  - ✓ In DW often enforced within ETL



Data Modeling:
Normalization

Normalization is a formal data modeling approach to examining and validating the model.

Pros:

- Ensures each attribute belongs to the entity to which it is assigned
- Redundant storage of information is minimized

Cons:

- Can adversely affect performance if rigorously enforced
- Can adversely affect deadlines if rigorously implemented

Levels of Normalization

- Dr. E. F. Codd identified ‘normal forms’ as the different states of a ‘normalized relational’ data model

- Levels
  - ✓ 1NF = No repeating groups
  - ✓ 3NF = No non-key interdependencies
  - ✓ 2NF = No partial key dependencies
  - ✓ 4NF = No independent multiple relationships
  - ✓ 5NF = No semantically related multiple relationships

E-R Modeling:
Why Normalize (3NF) ?

- Why normalize?
  - ✓ Data is easier to define
  - ✓ Data interdependencies are identified
  - ✓ Data ambiguities are resolved
  - ✓ Data model can be more flexible
  - ✓ Data model is easier to maintain

- Issues with normalization:
  - ✓ Structure can be very complex
    - ERP applications >10,000 tables
  - ✓ Difficult to understand
  - ✓ Performance can be an issue
    - Querying is especially tough

E-R Modeling:
ER Model Recap

Sales Order bought by a Customer with various Products in it: