Previews of TDWI course books offer an opportunity to see the quality of our material and help you to select the courses that best fit your needs. The previews cannot be printed.

TDWI strives to provide course books that are content-rich and that serve as useful reference documents after a class has ended.

This preview shows selected pages that are representative of the entire course book; pages are not consecutive. The page numbers shown at the bottom of each page indicate their actual position in the course book. All table-of-contents pages are included to illustrate all of the topics covered by the course.

This page intentionally left blank.

TDWI Business Intelligence and Analytics Architecture

TDWI takes pride in the educational soundness and technical accuracy of all of our courses. Please send us your comments—we'd like to hear from you. Address your feedback to:

info@tdwi.org

Publication Date: January 2018

TABLE OF CONTENTS

COURSE OBJECTIVES

You will learn:

- ✓ ***The full scope of architectural objectives—structural integrity, standardization, reusability, environmental fit, aesthetics, and sustainability***
- ✓ ***A framework to ensure overall architectural completeness and success—business purpose, organization, data and integration, process, and technology platforms***
- ✓ ***A framework for business architecture—performance, stakeholders, processes, rules, and information***
- ✓ ***A framework for organization architecture—people, purpose, process, and structure***
- ✓ ***A framework for data and integration architecture—collection, storage, operational data integration, data warehousing, big data integration, distribution/access/applications, and data modeling/metadata management***
- ✓ ***A framework for process architecture—methodologies, data governance, data modeling/metadata management, data flow, business processes, and operations/support***
- ✓ ***A framework for technology architecture—servers, data sourcing, databases, storage, data integration, business analytics, and data management***

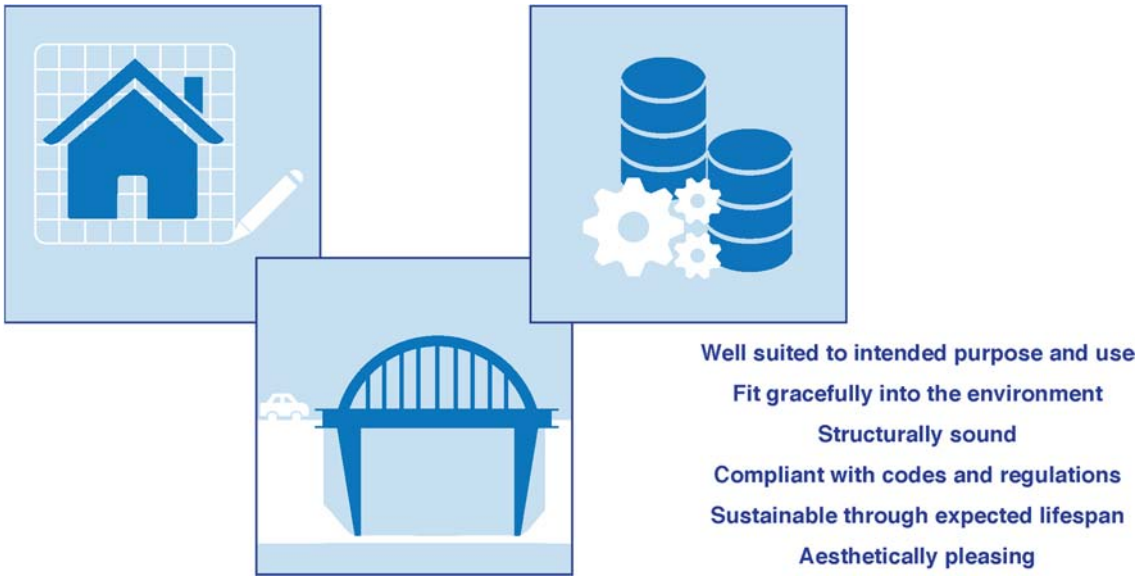
tdwi
Transforming Data With Intelligence™

Module 1

Business Intelligence and Analytics Architecture Concepts

Architecture Defined

What Is Architecture?



Architecture Defined

What Is Architecture?

DEFINITION

Architecture defines the roles, structure, relationships, and rules by which a collection of components constitute a cohesive whole—the glue that bonds individual parts into a system. Architecture is an early-stage design activity that precedes detailed design, specification, and construction.

ROLES OF ARCHITECTURE

Effective architecture ensures that the things we build:

- Are suited to the purposes for which they are intended
- Fit gracefully into their environment
- Are structurally sound
- Comply with codes, regulations, and standards
- Are sustainable through their expected lifespan
- Are aesthetically pleasing

These principles hold true for architecture of many things—buildings, bridges, information systems, and more.

BI and Analytics Defined

Business Intelligence

> "the ability of an organization to reason, plan, predict, solve problems, understand, innovate, and learn in ways that increase organizational knowledge, inform decision processes, enable effective actions, and help to establish and achieve business goals"
>
> David Wells, BI Consultant

BI and Analytics Defined

Business Intelligence

BUSINESS AND INTELLIGENCE

Dave Wells published a definition of BI focused on business capabilities, answering the question:

What does it mean to be an intelligent business?

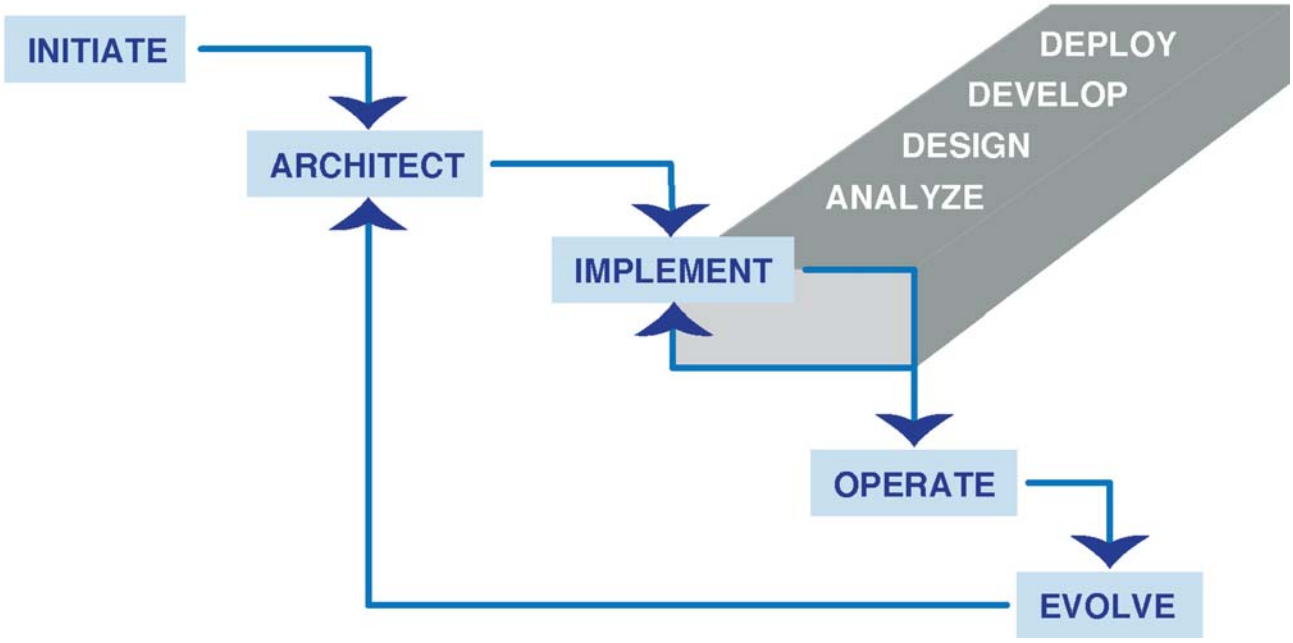
This definition describes business intelligence as "the ability of an organization to reason, plan, predict, solve problems, understand, innovate and learn ..." The definition also describes the purpose of BI, answering the question:

Why do we need business intelligence?

"... in ways that increase organizational knowledge, inform decision processes, enable effective actions, and help to establish and achieve business goals."

Architecture in the BI and Analytics Lifecycle

Iterative Implementation



Architecture in the BI and Analytics Lifecycle

Iterative Implementation

BI AND ANALYTICS PROGRAMS

BI and analytics solutions are developed as part of programs that balance business needs with available resources and materials.

Implementations take place within a larger lifecycle framework that evolves over time. This diagram depicts typical phases of the BI and analytics development lifecycle. Note the position of the architecture stage.

INITIATION

Every BI and analytics program begins with activities that establish the charter, sponsorship, initial funding, and expectations.

ARCHITECTURE

Definition of architecture typically follows initiation. Architecture identifies the components of solutions—data, information, technology, process, project, and organization—and establishes standards and guidelines for implementation and operation.

IMPLEMENTATION

Implementation projects are the development activities of BI and data warehousing (DW) programs. Typically executed iteratively as a series of small, short duration projects, this step delivers the databases, data warehouses, data transformation processes, data and information services, analytics applications, and technology deployments necessary to make BI a reality.

OPERATION

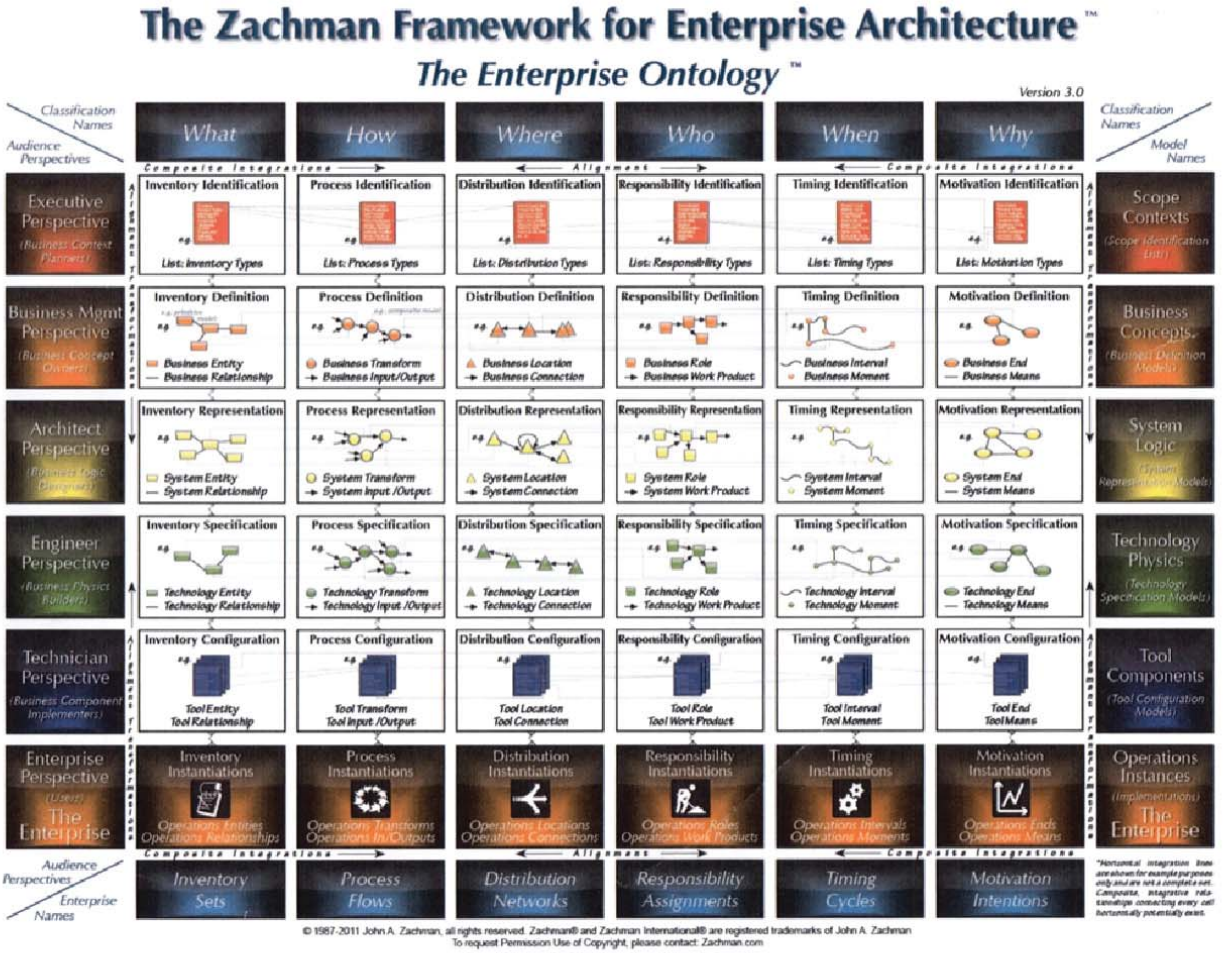
Operation encompasses administration of and delivery of information services through BI/DW solutions. Day-to-day operation of the data warehouse and delivery of information and analytics services are abundant with opportunities to recognize and even anticipate change.

EVOLUTION

Throughout the operational life of the BI program it is necessary to continuously align with changing business requirements. Recognizing need for change drives evolution, closing the lifecycle loop by returning to the architecture phase.

Architecture Frameworks

Zachman Framework



Architecture Frameworks

Zachman Framework

FRAMEWORK INCEPTION

Multiple levels of abstraction are needed to manage complexity, understand and document from multiple perspectives, communicate effectively, and provide a natural progression from need to solution. The Zachman Framework for Enterprise Architecture was developed by observing the way work was planned and performed in disciplines, such as building, that have existed for centuries. Zachman reasoned that the same fundamental concepts exist within the information environment and that applying them can provide systems and other products with the same lasting power and reliability as buildings and quality machinery. This framework was first described in "A Framework for Information Systems Architecture," IBM Los Angeles Scientific Center Report, John A. Zachman, 1986. For additional information, visit www.zachman.com.

FRAMEWORK ROLE

The Zachman Framework is an ontology; it is not a methodology. It identifies and classifies discrete components that, as a set, describe an enterprise. It is analogous to the periodic table in chemistry. It provides a predictable structure capable of producing repeatable results. It is the basis for architecture.

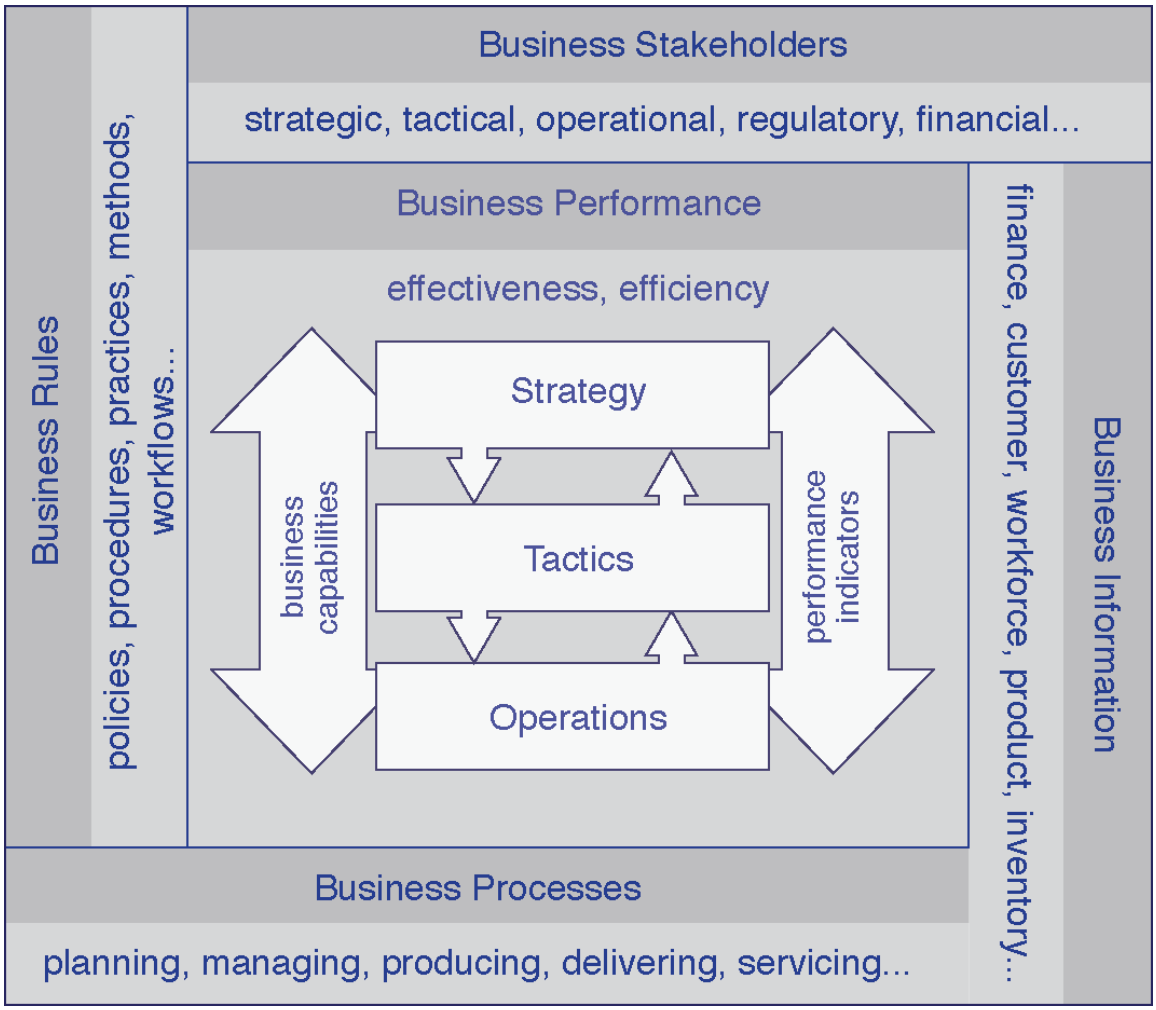
tdwi
Transforming Data With Intelligence™

Module 2

Architecting Business Capabilities

Business Architecture Concepts

Framework for Business Architecture



Business Architecture Concepts

Framework for Business Architecture

THE TOP OF THE STACK

Business architecture is the essential starting point for BI and analytics architecture. Without the right business focus, it makes little difference how well you organize or structure data, integration, and processing. When a BI program gives more attention to dashboards, scorecards, OLAP, reports, and data warehouses than to finance, R&D, marketing, operations, and customer support it is difficult to create real business value.

BUSINESS PERFORMANCE

Business performance is the core of business architecture. Business analytics and intelligence delivers value when it makes substantial contributions to business effectiveness and efficiency. Three primary ways that BI contributes are:

- Enabling new capabilities for business planning, management, and execution
- Creating feedback loops between strategy and tactics and between tactics and operations
- Formalizing the definition, measurement, and tracking of key business performance indicators

BUSINESS STAKEHOLDERS

The human connection is an important part of BI and analytics. Identifying and classifying stakeholders is an architectural construct. Multiple classifications such as by role (strategic, tactical, operational) and by interest (legal, regulatory, financial) are useful to fully understand the scope and relationships of stakeholders.

BUSINESS INFORMATION

Information is an obvious BI and analytics architecture topic. A primary purpose of both BI and analytics is to deliver information that makes a difference. Business architecture identifies business information subjects and relationships among them.

BUSINESS PROCESSES

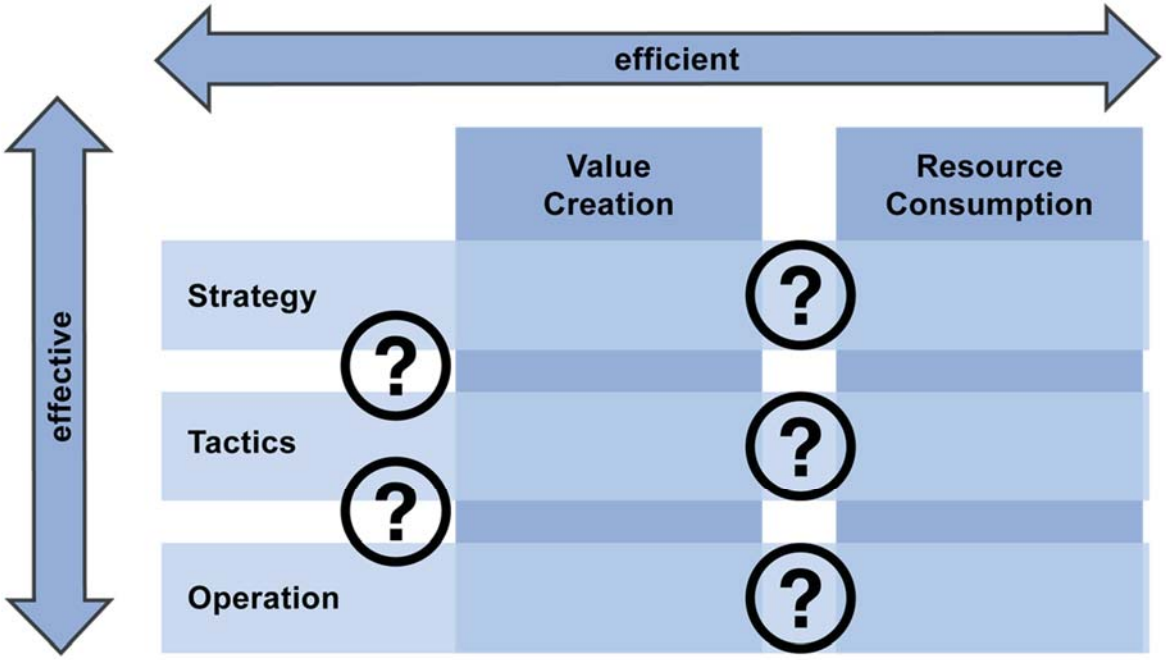
The processes of the business are among stakeholder interests and responsibilities, and they are also sources for and consumers of information. No business architecture is complete without a process view.

BUSINESS RULES

The policies, procedures, practices, methods, and workflows of business express rules and constraints that must be reflected through all of the organizational, integration, data, and process components of BI and analytics. Business architecture includes the structures and relationships to incorporate business rules into BI and analytics systems.

Business Performance

Definition and Concepts



Business Performance

Definition and Concepts

BUSINESS AT THREE LEVELS

Strategy defines a company's future position as a set of goals and the methods by which it plans to achieve those goals. Strategy is typically a long-term (three- to five-year) view of a business position with financial and market/competitive goals.

Tactics are the means by which strategy is carried out—the work of implementing strategy. Tactics encompass a set of planned activities to progress from one milestone to the next in pursuit of strategic goals.

Business operations are the day-to-day activities and the end-to-end business processes that carry out the work of the business. Operations includes creating, marketing, selling, and delivering products and services as well as managing the resources—human, financial, and physical—that are necessary to conduct business.

BUSINESS EFFECTIVENESS

Business effectiveness is the degree to which strategy is realized as action in business operations. Two translations occur that directly influence effectiveness: translating strategy to tactics, and translating tactics to operations. At both of these points performance measures are used to learn how well strategy is translated into action, and to provide feedback up the management chain.

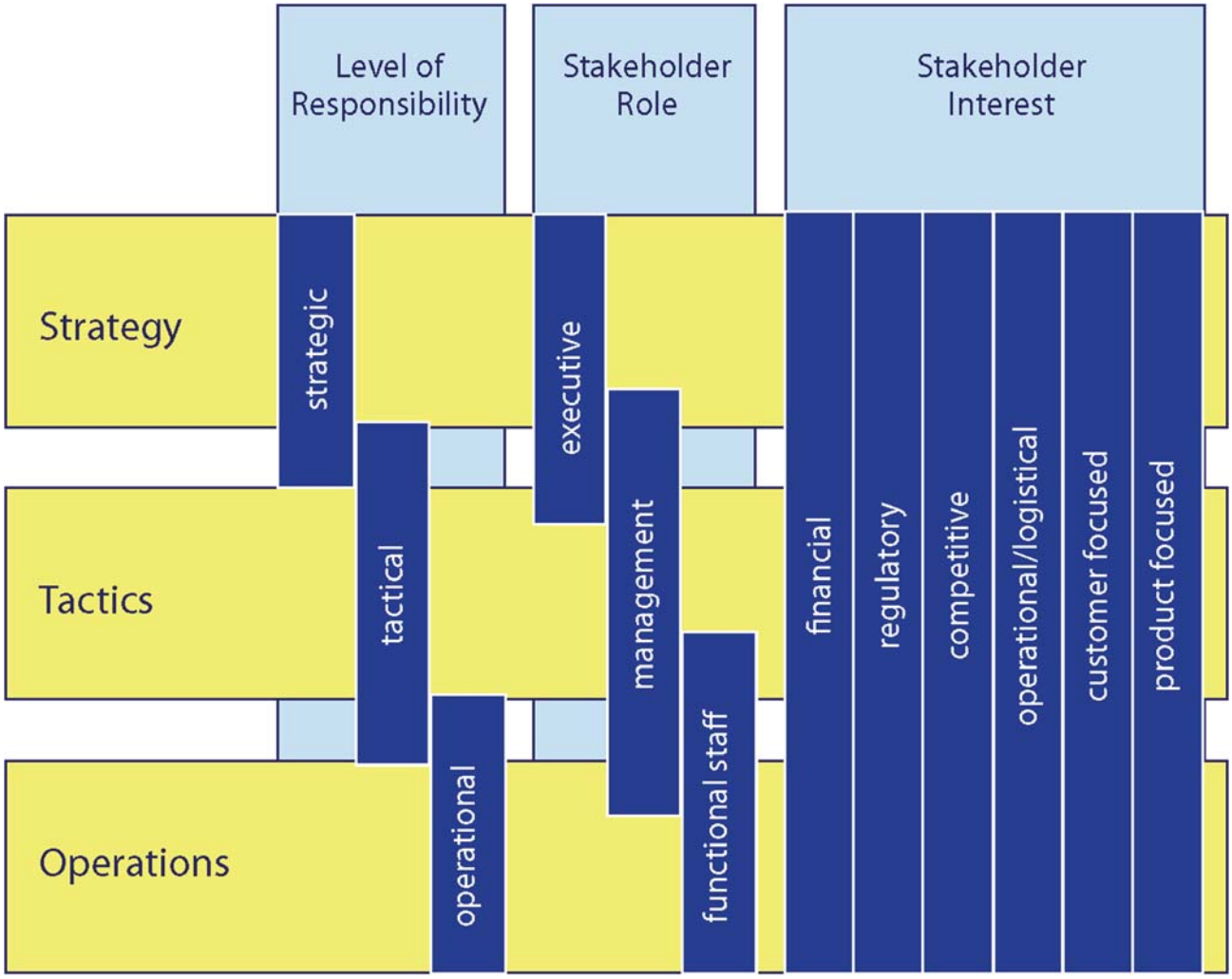
BUSINESS EFFICIENCY

Every business process or activity can be viewed from two perspectives: the value created and the resources consumed. This holds true at all three levels of business activity—strategic, tactical, and operational.

Business efficiency is the balance of value created and resources consumed. Improvements in efficiency are achieved by increasing value, by reducing resource consumption, or by a combination of the two.

Business Stakeholders

Responsibilities, Roles, and Interest



Business Stakeholders

Responsibilities, Roles, and Interest

INTERESTS AND EXPECTATIONS

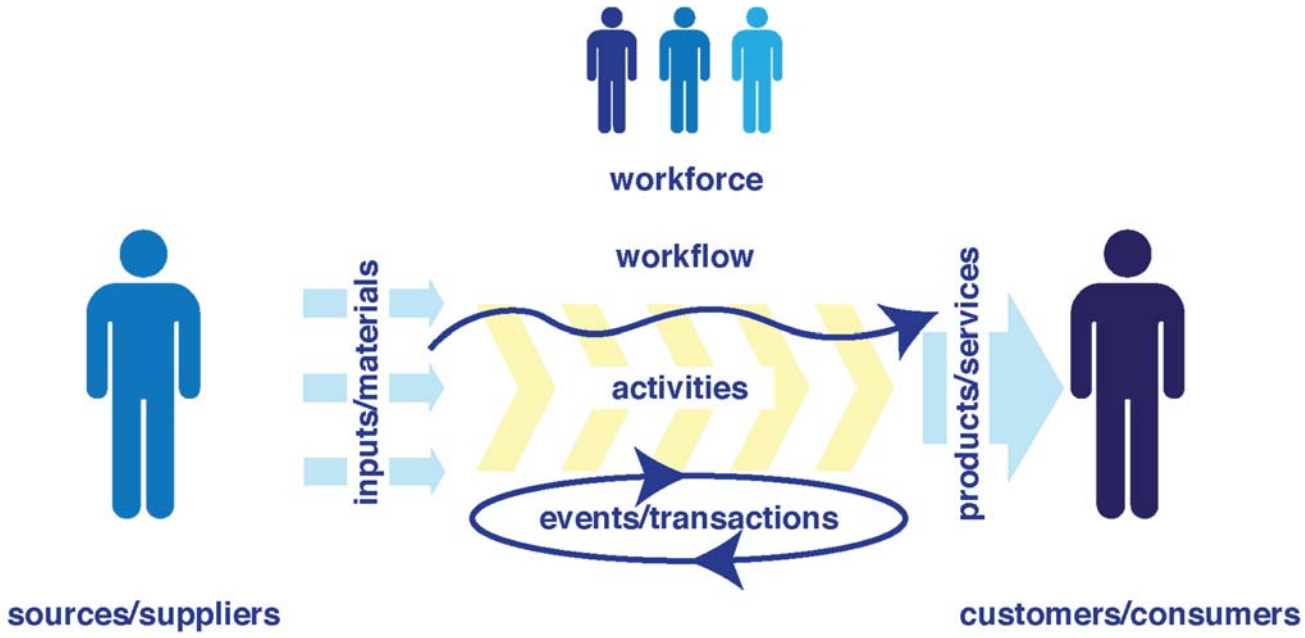
Business intelligence and analytics stakeholders are the people who use BI and analytics systems and services to help them do their work more effectively. Stakeholders can be classified in many ways including:

- Level of responsibility—strategic, tactical, operational
- Stakeholder role—executive, management, functional staff
- Stakeholder interest—financial, regulatory, competitive, customers, products, etc.

Each of these classifications is helpful to identify the full range of BI and analytics stakeholders and to understand their interests and expectations.

Business Processes

Process Concepts



	Planning	Managing	Producing	Delivering	Servicing
Products					
Services					
Revenue					
Customers					
Employees					
Partnerships					
Assets					

Business Processes

Process Concepts

DEFINITION

Simply defined, a business process is a set of linked activities that are performed with the purpose of providing a specific product to a known customer. Tom Davenport's definition adds depth, stating that a business process is "a structured, measured set of activities designed to produce a specific output for a particular customer or market. It implies a strong emphasis on *how* work is done within an organization, in contrast to a product focus's emphasis on *what*. A process is thus a specific ordering of work activities across time and space … the structure by which an organization does what is necessary to produce value for its customers."[1]

PURPOSE

The customers receiving value from a business process may be external customers such as those who purchase products, or they may be internal customers—people and departments who receive goods or services from other people and departments.

Each process typically focuses on one of several core business activities, including planning, managing, producing, delivering, and servicing, and on the application of that activity to core business entities such as products, services, and customers. Every business has many business processes. The matrix here depicts 35 core processes using a limited list of only seven business entities.

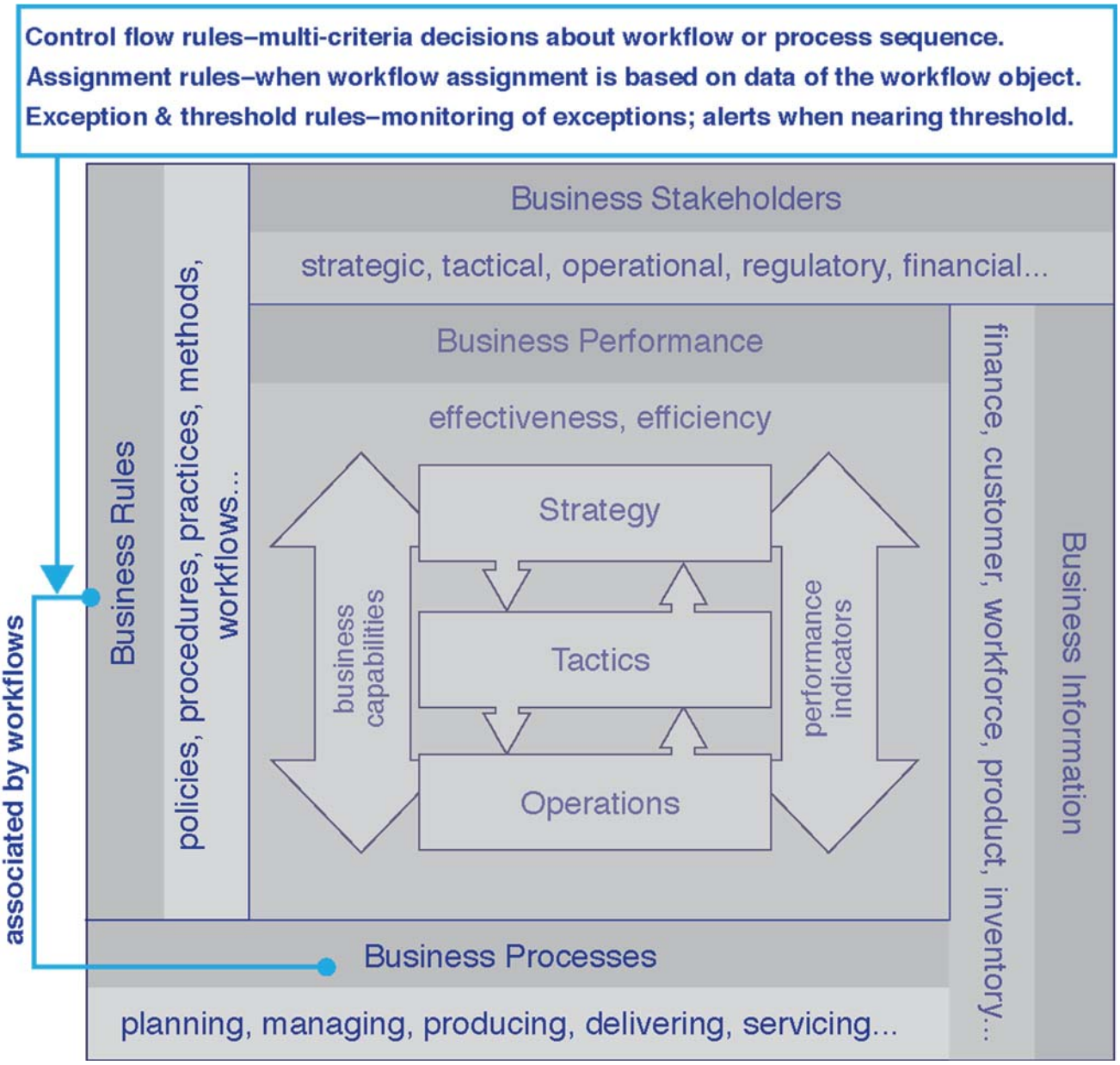
	Planning	Managing	Producing	Delivering	Servicing
Products					
Services					
Revenue					
Customers					
Employees					
Partnerships					
Assets					

Certainly, every business has more than seven core entities, and the list extends further with the need to include business concepts that are not typically considered as "entities"—compliance, risk, opportunity, data, information, knowledge, loyalty, collaboration, etc.

[1] *Process Innovation: Reengineering work through information technology*, Davenport.

Business Rules

Assertions and Constraints



Business Rules

Assertions and Constraints

RULES AND PROCESSES

A business rule is an assertion that defines constraints on an aspect of conducting business. Many of the rules in business are process oriented, expressing constraints on the activities and events within a process. Three common kinds of process rules are:

- Control flow rules where multiple decision criteria determine workflow or process sequence. In mortgage lending, for example: *when loan-to-value ratio is less than X percent and monthly payment is less than Y percent of monthly income, then no underwriter review is required.*
- Assignment rules where workflow activity assignment is based on data of the workflow object. *When the mortgage loan amount exceeds the limits established for conventional conforming loans, then jumbo loan approval steps must be performed.*
- Exception and threshold rules when approaching or exceeding defined limits or normal conditions. *When loan-to-value ratio is greater than X percent calculate the prepaid interest (points) that must be paid on the loan.*

MORE RULES

Not all business rules are process rules of the three types described above. Other common kinds of business rules include data constraints, access rules, regulatory compliance rules, legal mandates, and more.

Business Information

Information Uses

Business Information

Information Uses

PURPOSES OF INFORMATION ASSETS

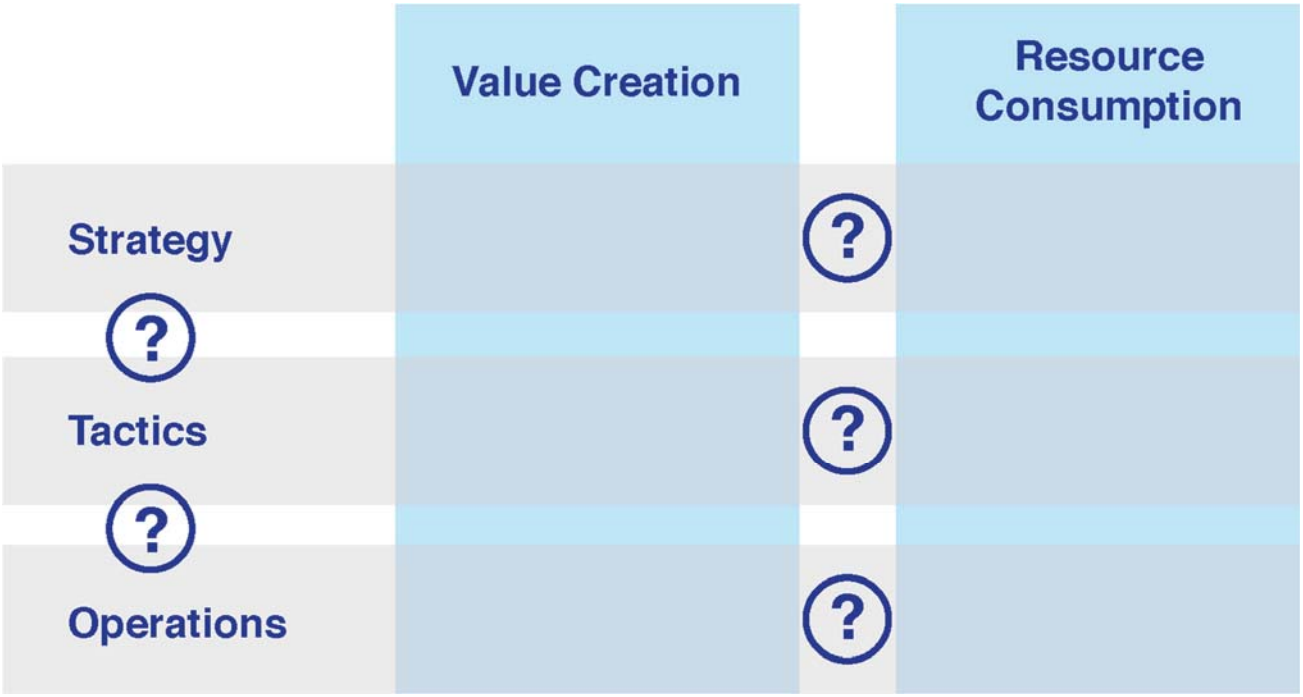
Most agree today that information is an asset and managing that asset is a fundamental purpose of data governance. It is important that the language, standards, and guidelines of BI and analytics architecture enable governance. Managing data and information assets is challenging because they are unique among corporate resources—very different from financial and human resources in one important respect. Money and staffing are consumable resources. You can't spend the same dollar or staff-hour twice. Data and information are reusable resources that can be applied multiple times for many different purposes:

- To keep records of business transactions
- To produce reports about business processes and performance
- To maintain audit trails of business events and activities
- To measure business results and performance
- To analyze business behaviors
- To inform decision-making processes
- To predict future conditions and outcomes
- To gain insight through discovery and learning
- To automate routine and repeatable operational decisions

Sometimes multiple purposes for the same data will create conflicts. Careful attention in BI architecture minimizes the impact of conflicting uses.

How Business Architecture Fits In

Align Investments with Business Goals



How Business Architecture Fits In

Align Investments with Business Goals

ALIGNMENT OF INVESTMENTS

Business intelligence and analytics architecture involves investment in all levels of architecture. Architecture enables business analytics and intelligence capabilities—capabilities that align with the business goals of the organization.

Business goals are the drivers of architectural choices. Strategic, tactical, and operational goals will influence all architectural components and need to be aligned to ensure attainment of the organizational strategy.

STRATEGIC GOALS

Strategy defines a company's future position as a set of goals and the methods by which it plans to achieve those goals. For business analytics and intelligence, strategic goals might translate into new capabilities needed to address business drivers. Capabilities such as prediction, simulation, or forecasting are examples.

TACTICAL GOALS

Tactics are planned activities to progress from one milestone to the next in pursuit of strategic goals. Strategic goals influence tactical goals. Tactical business analytics and intelligence goals translate into new projects that include the delivery systems, technology, or services to support new capabilities.

OPERATIONAL GOALS

Operational goals are the day-to-day activities and business processes that carry out the work of the business. Tactical goals influence operational goals. New operational goals might include new or changed business processes. New systems, for example, can influence changes to existing business processes or create different activities to be performed.

Module 3

Architecting Organizations

Organizational Architecture Concepts

Framework for Organizational Architecture

People	Purpose
roles skills competencies capabilities	goals results responsibilities accountabilities
Process	**Structure**
decision execution communication information	reporting relationships dependencies networks

Organizational Architecture Concepts

Framework for Organizational Architecture

BI, ANALYTICS, AND THE FABRIC OF BUSINESS

Organizational structure is a term used to highlight the way a company thinks about hierarchy, assigns tasks to personnel, and ensures its workforce works collaboratively to achieve a common goal.

Organizational architecture is primarily about people. It is the essential architectural perspective that connects people with business intelligence and analytics and weaves both into the fabric of the business. Process, data, and integration struggle to create value if people don't see them as valuable and understand how they are connected with their jobs.

Organization of teams may be formal or informal. BI and analytics projects often incorporate informal reporting structures that do not map to the official organization chart.

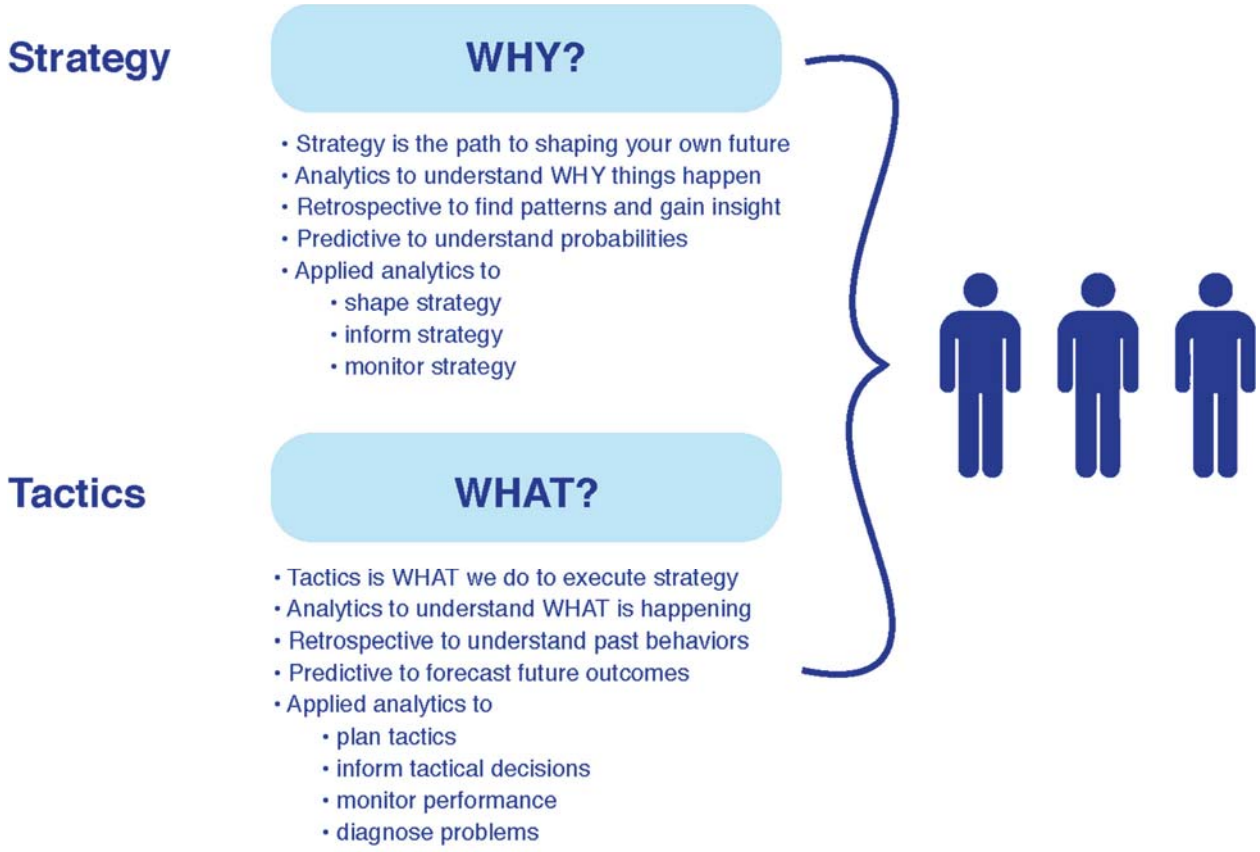
ORGANIZATIONAL FRAMEWORK

Organizational architecture consists of four categories of components:

- People—the "who" of organizations, encompassing roles, skills, competencies, and capabilities
- Purpose—the "why," including goals, results, responsibilities, and accountabilities
- Process—the "how," including decision, execution, communication, and information components
- Structure—relationships and dependencies among people

People

Aligning with the Business



People

Aligning with the Business

PLANS AND THE ORGANIZATION

Strategy execution is the primary focus of organizational leaders. As we have already reviewed, strategy defines the organization's desired future position and tactics are the planned activities that support the attainment of strategic goals. Strategic and tactical plans will determine how the organization will use its major resources.

BI and analytics have a role in determining "why things happen" for strategic plans and "what is happening" for tactical plans. Understanding the "why" and "what" also influences how the BI and analytics organization should be structured.

STRATEGIC PLANS

Strategy is the path to shaping the future. BI and analytics are used to understand why things happen. As part of the strategic plan, organizations need capabilities to initially shape the strategy, inform the strategy, and monitor if the strategy is working. To do this certain business analytics capabilities are needed such as retrospective and predictive analysis or performance management using BI products such as scorecards or dashboards.

TACTICAL PLANS

The tactical plan is what we do to execute the strategy. As part of the tactical plan, BI and analytics are applied to plan tactics, inform tactical decisions, monitor performance, and diagnose problems.

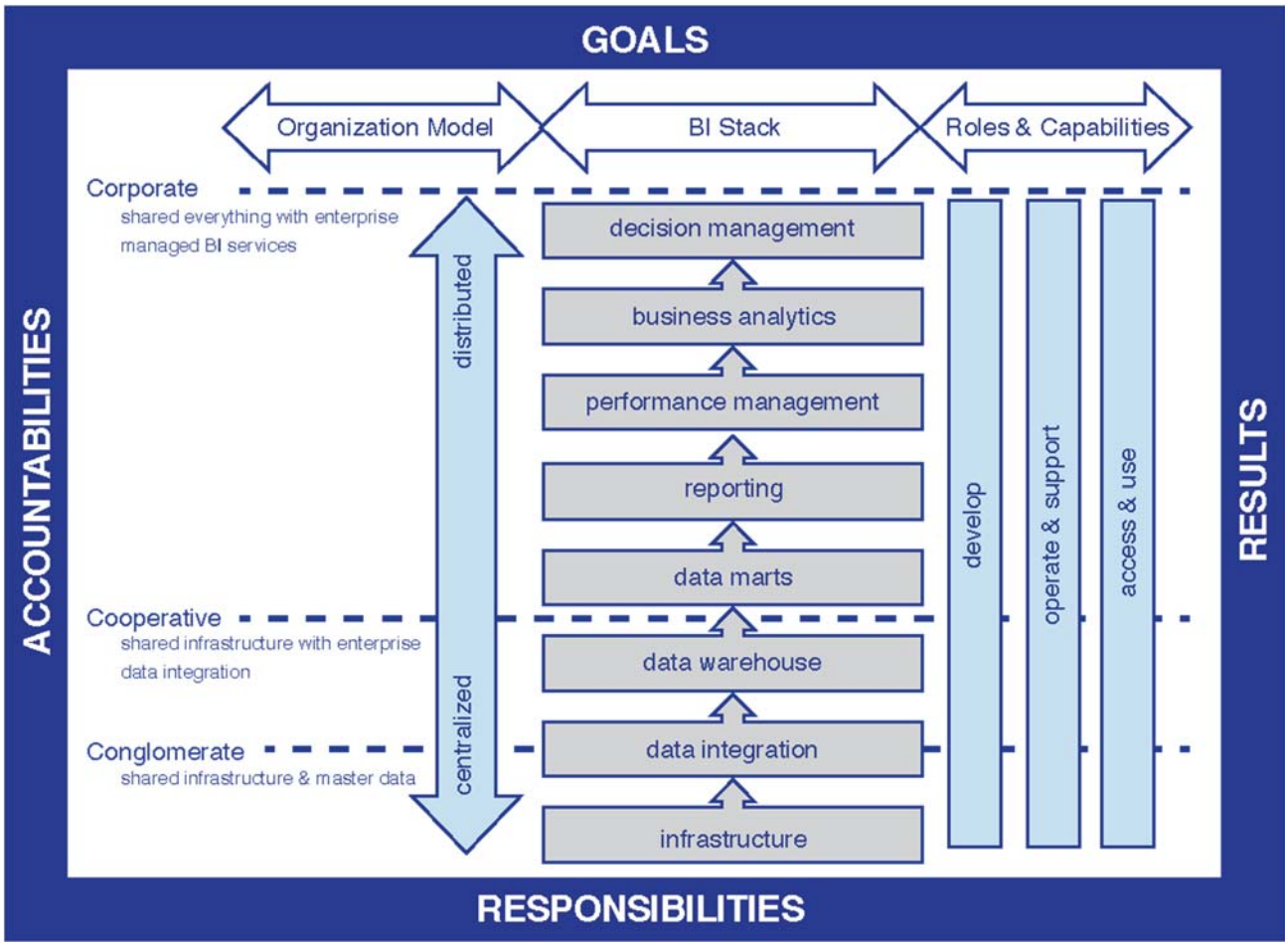
ORGANIZATIONAL STRUCTURE QUESTIONS

In order to determine the best organizational structure, consider the strategic and tactical plans and the goals to be accomplished:

- What are the business drivers and priorities associated with each goal and when do they need to be attained?
- What are the results to be achieved?
- Who has the right skills, competencies, and capabilities to support the strategic and tactical plans?
- How will decisions be made?
- Who is accountable and responsible?
- What should the dependencies and relationships be to execute the plan?

Purpose

From Goals to Results



Purpose

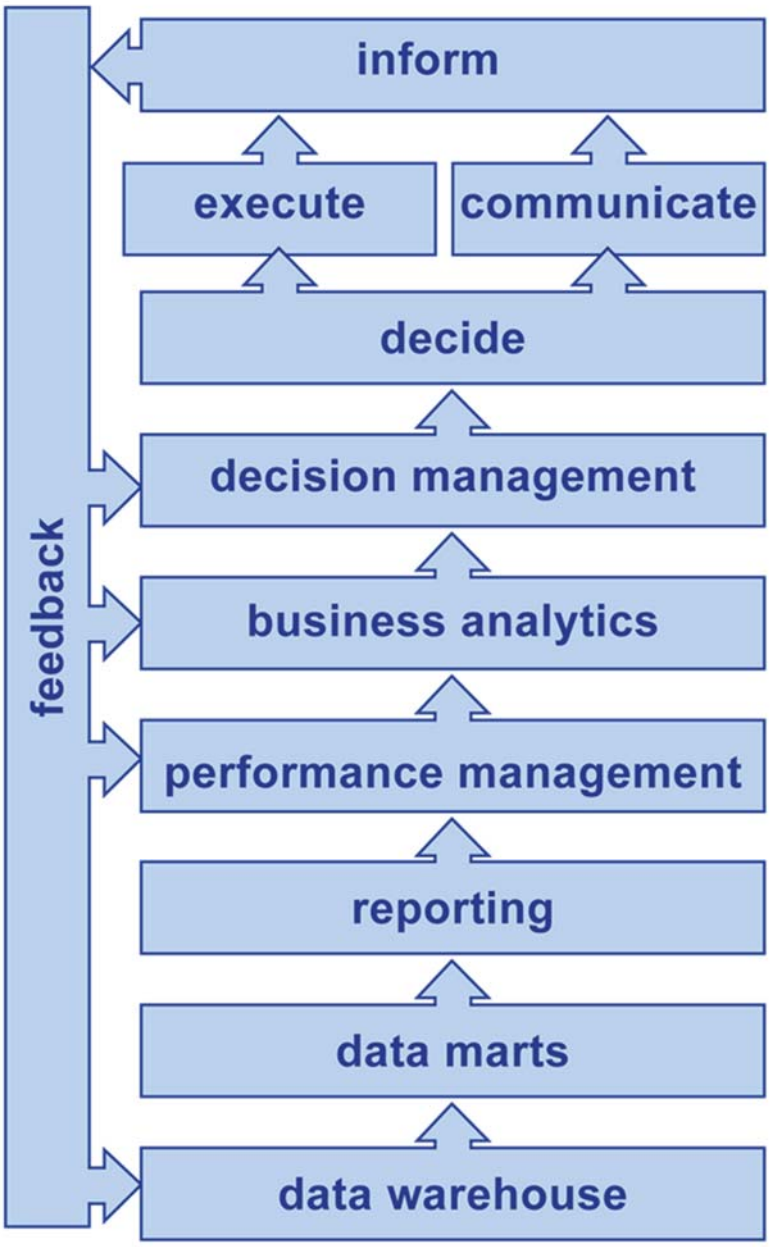
From Goals to Results

ORGANIZATIONAL BI AND ANALYTICS GOALS

Roles and responsibilities are strongly associated with the overall goals for BI and analytics and the goals for each organization's contribution. BI and analytics fits neatly into the business culture when responsibilities are well aligned with goals. The mirror of responsibility is accountability. Aligning responsibility with accountability and then linking accountability with results further strengthens your ability to weave BI into business culture, practices, and processes.

Process

Action and Information



Process

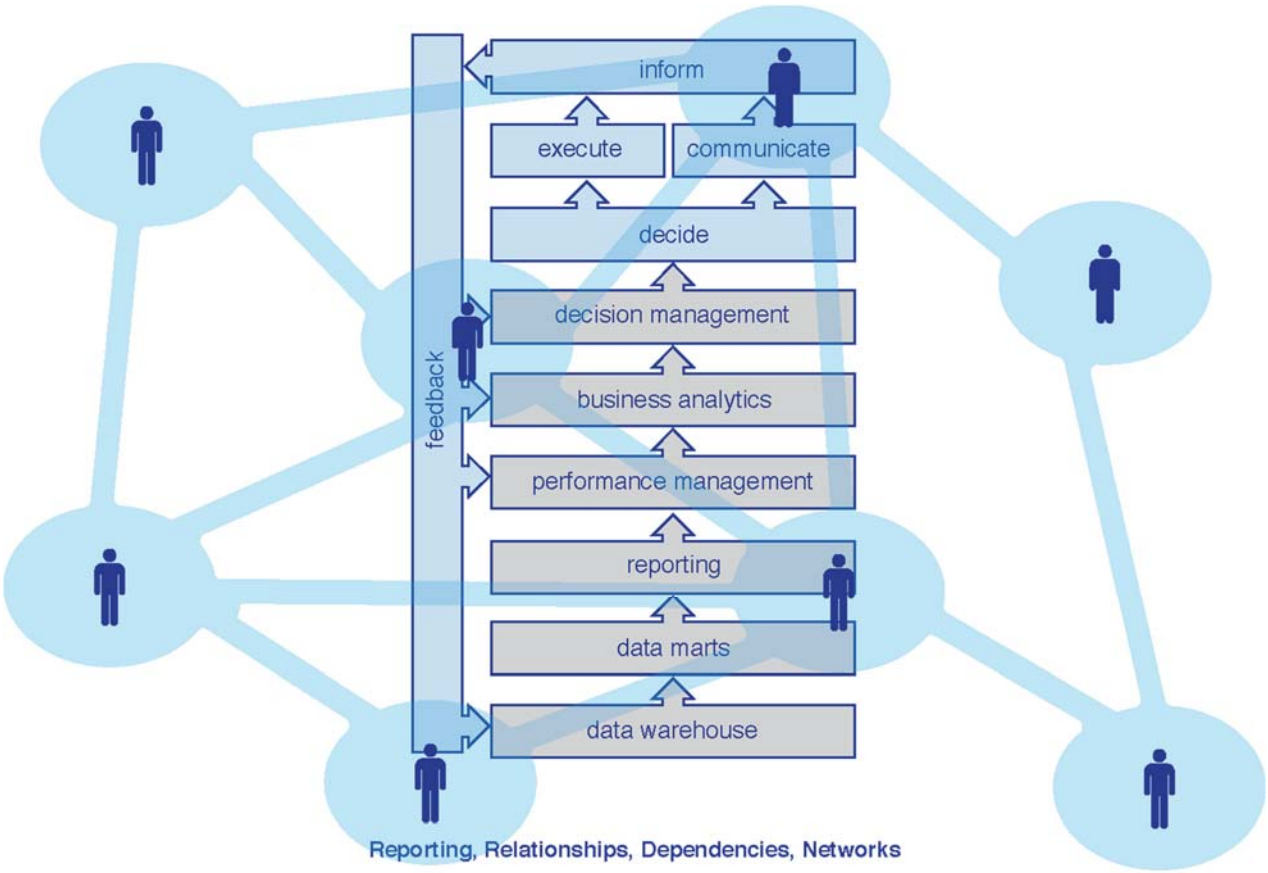
Action and Information

FEEDBACK LOOPS

The process component of organizational architecture has a strong affinity with business processes in business architecture and with the broader category of process architecture. In organizational architecture, the goal is to frame those processes that use BI and analytics services and depend on capabilities. Here the concept of decision support goes beyond decision making to encompass executing and communicating decisions and informing people and processes to create feedback loops. Feedback can advance capabilities in performance management, business analytics, and decision management. Feedback may also be reflected as new data leading to new information opportunities.

Structure

Connecting People



Structure

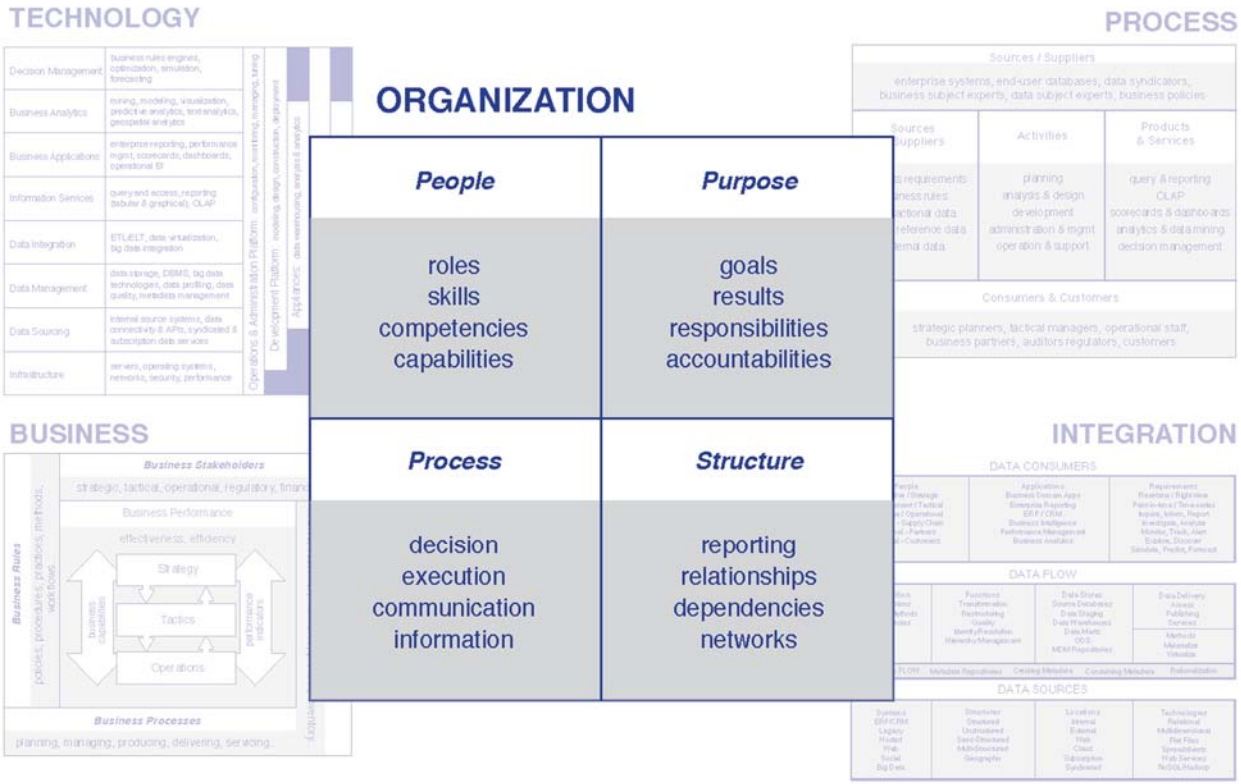
Connecting People

HIERARCHIES, NETWORKS, ETC.

BI and analytics at its best drives conversation and collaboration among people. Thus the architect's view of the organization must include an understanding of relationships, both formal and informal, among people. Knowing how people relate, communicate, and depend on others is important for the architectural objective of seamlessly fitting into the environment. BI and analytics is least disruptive culturally and organizationally when it readily adapts to fixed organization reporting hierarchies, adaptive matrix-based teams, and unofficial but very real collaboration and communication networks.

How Organizational Architecture Fits In

Organizational Fit



How Organizational Architecture Fits In

Organizational Fit

ORGANIZATIONAL FIT

The human dimension is one of the main ways that the five architectural layers are interconnected. People in the organizational view are sometimes suppliers in process architectures, sometimes business stakeholders in business architecture, and sometimes data consumers in integration architecture. Supplier, stakeholder, and consumer are roles in which people act. Organizational architecture looks at people independently of their roles and inclusive of all of their roles.

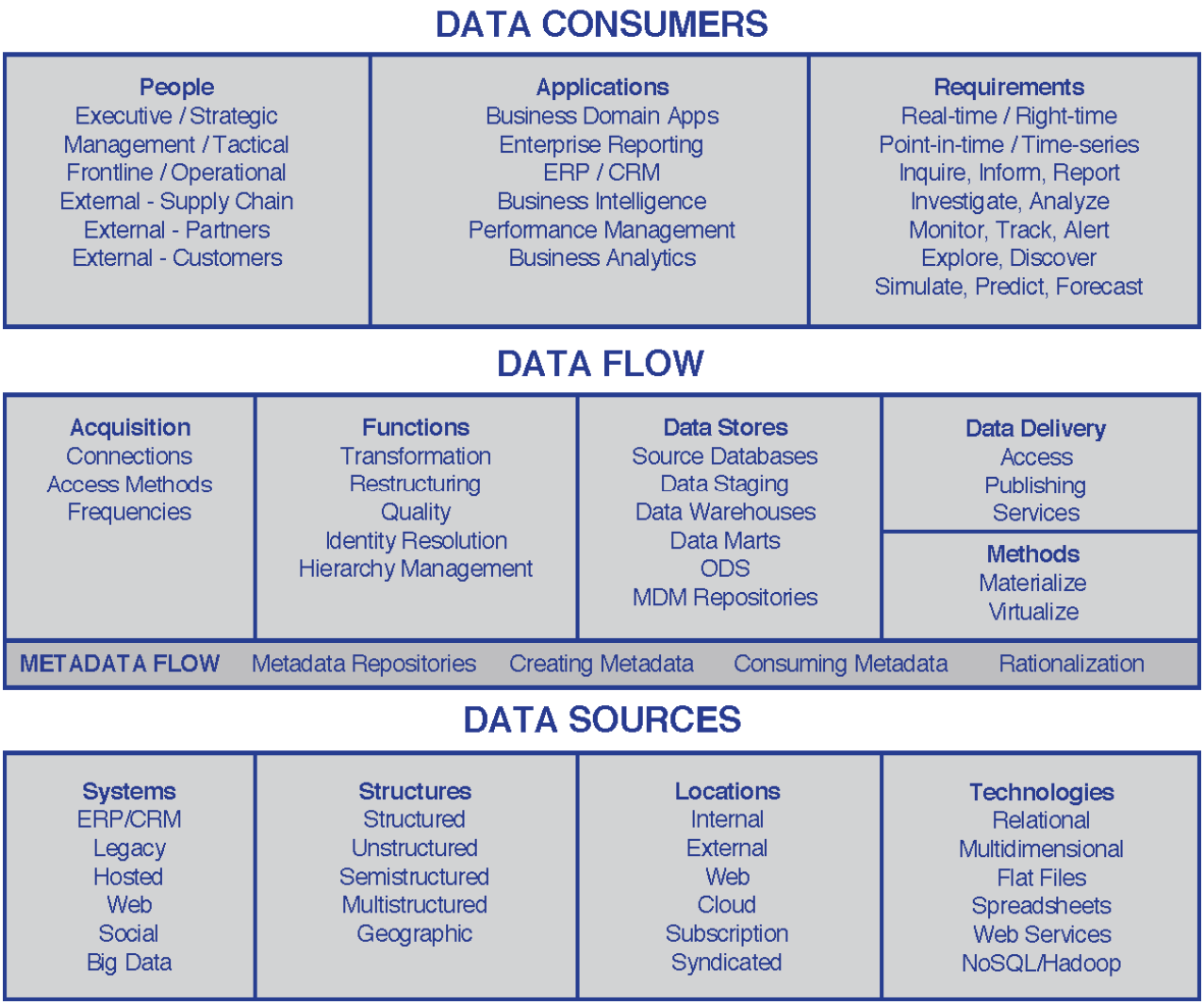
tdwi
Transforming Data With Intelligence™

Module 4

Architecting Data and Integration

Data and Integration Architecture Concepts

Framework for Data and Integration Architecture



Data and Integration Architecture Concepts

Framework for Data and Integration Architecture

DATA AT THE CORE

Data is fundamental to business analytics and intelligence. Without data, none of the BI and analytics services and products are possible. Without good data, it is impossible to deliver high-quality BI and analytics products and services. Data architecture components are interwoven throughout integration architecture, which includes data models, data consumers, data flow, and data sources.

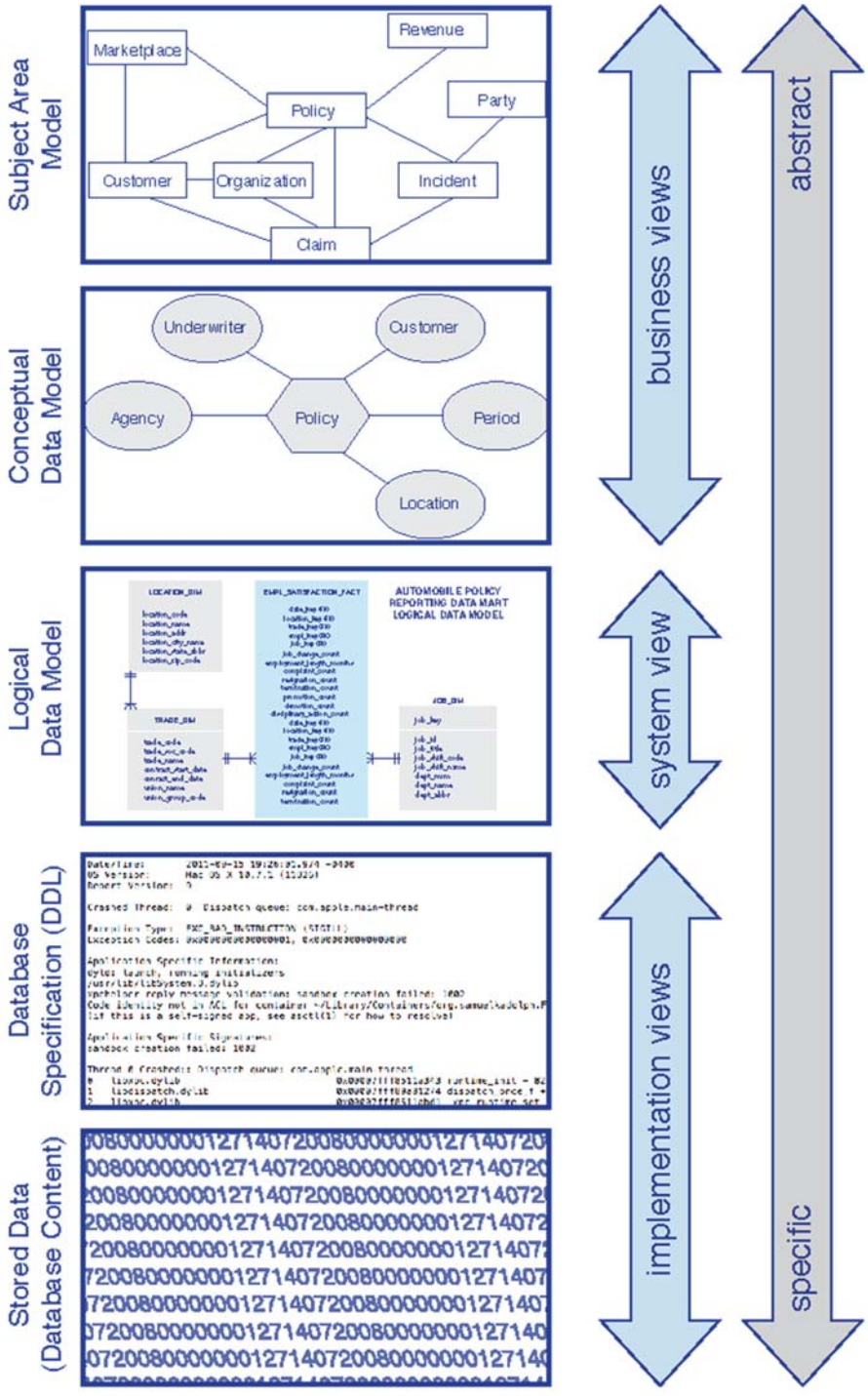
RATIONALIZING DISPARATE DATA

Data integration focuses on all the components used to rationalize disparate data. Integration architecture includes:

- Data sources—systems, structures, locations, and technologies of source data
- Data flow—acquisition, functions for data transformation and manipulation, data stores, integration methods, delivery methods, and metadata management
- Data consumers—both people and applications, with special attention to requirements

Describing Data

Data Abstraction



Describing Data

Data Abstraction

DATA MODELING

Data resources are most frequently described using data models. A data model is a representation of information used by the enterprise. Data can be represented at multiple levels of abstraction, each providing a different perspective and understanding of the data. The highest level of abstraction is a business context view with both external (outside looking in) and internal (looking from within) perspectives. This level corresponds with business activities and their associated information needs. Subsequent levels transform needs into solutions, culminating with a physical specification at the lowest level.

REMOVING DETAIL AND SPECIFICS

Data abstraction is the process of removing specifics and details from views of data. Abstraction is the opposite of specification, which increases the level of detail and specifics. Multiple levels of abstraction are used to manage the connection of data consumers with data sources:

- Business views remove the specifics of application systems and database implementation from data models and descriptions. Data is described only by its business meaning. Data models at this level are often described as conceptual, canonical, or semantic models.
- System views remove the details of database implementation from data models and descriptions. Data is described by business meaning and associated with keys and relationships that support access and navigation. Data models at this level are often called logical models.
- Implementation views capture the details of database implementation and the specifics of keys and indexing. Data is typically described in systems context rather than business meaning. Data models at this level are known as physical models.

ARCHITECTURE

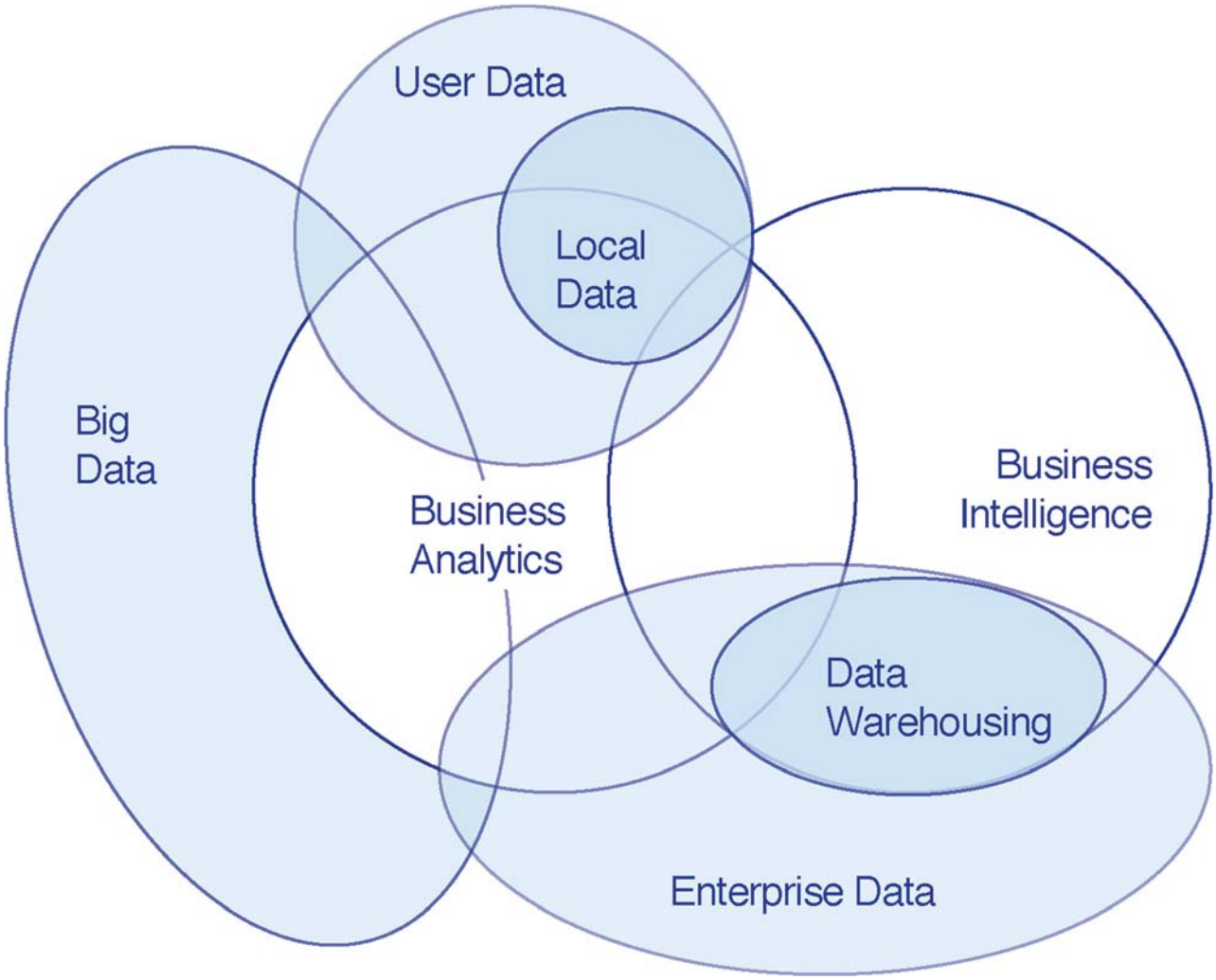
Data architecture prescribes modeling techniques and their use:

- Types of data stores and their intended use
- Modeling techniques and their role in the development lifecycle
- Data modeling standards for describing information resources

Data architecture supports business architecture and generates requirements for integration architecture.

Collection and Storage

Diversity of Sources



Collection and Storage

Diversity of Sources

INGESTING AND MANAGING DATA

Today's enterprise information architecture comprises multiple categories of information. Enterprise information assets include user data, local data, and various big data sources.

Supporting BI and analytics typically involves the establishment of additional repositories for data. Data stores in the BI and analytics environment may include:

- Data staging
- Data lakes
- Analytics sandboxes
- Data warehouses
- Data marts
- Master data repositories

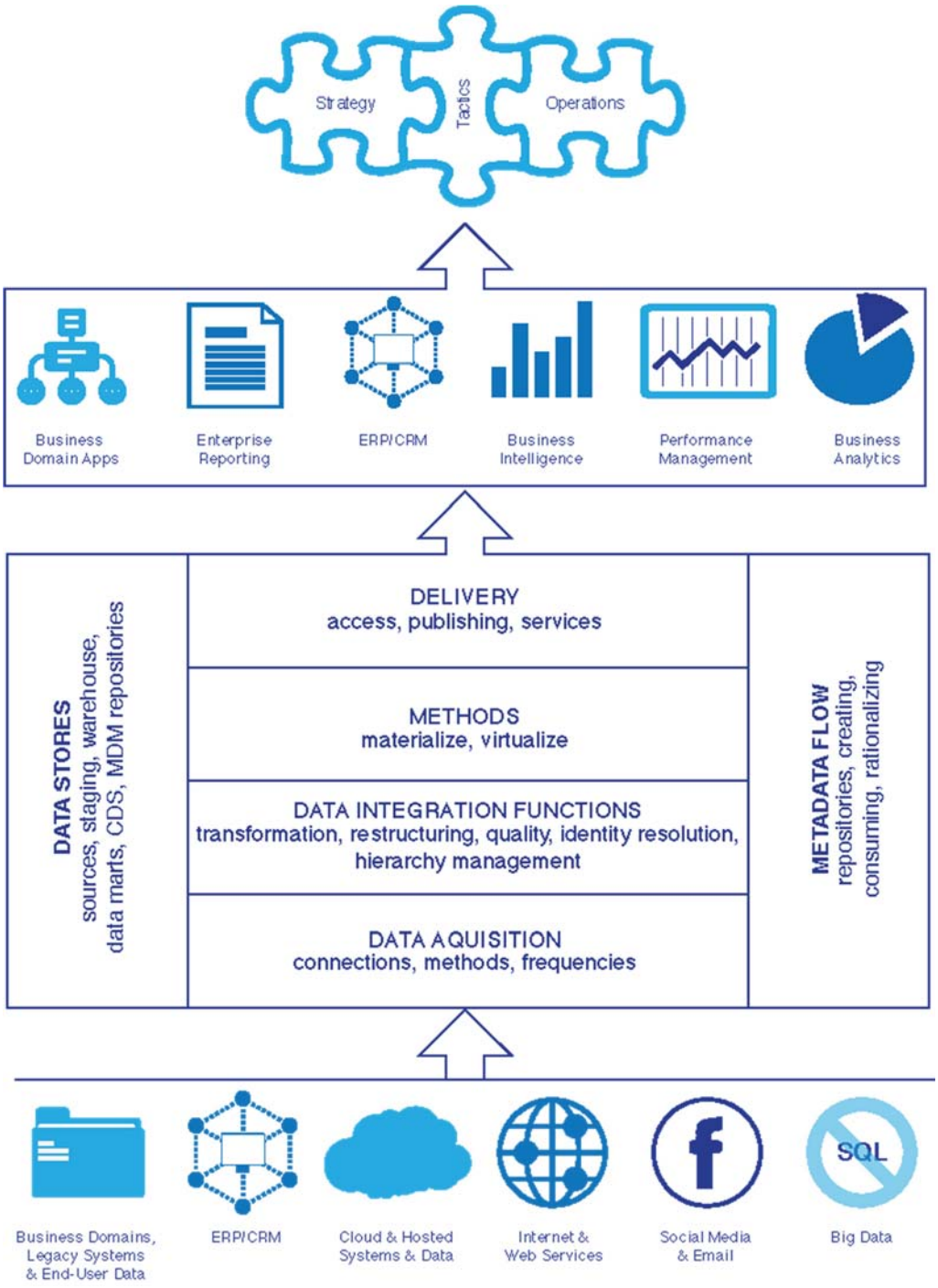
Additional repositories may be necessary to perform utility functions, such as data standardization, quality assessment, and data cleansing.

IMPACT OF BIG DATA

Big data has expanded the number and types of data sources that can be used to enrich the analytics process. It expands analysis opportunities by including data from Web searches, online shopping, email, text messaging, social media activity, machine-to-machine communications, sensor data, and much more. Big data is not a new or standalone silo of data.

Data Integration

Integrating Information



Data Integration

Integrating Information

MULTIPLE SOURCES

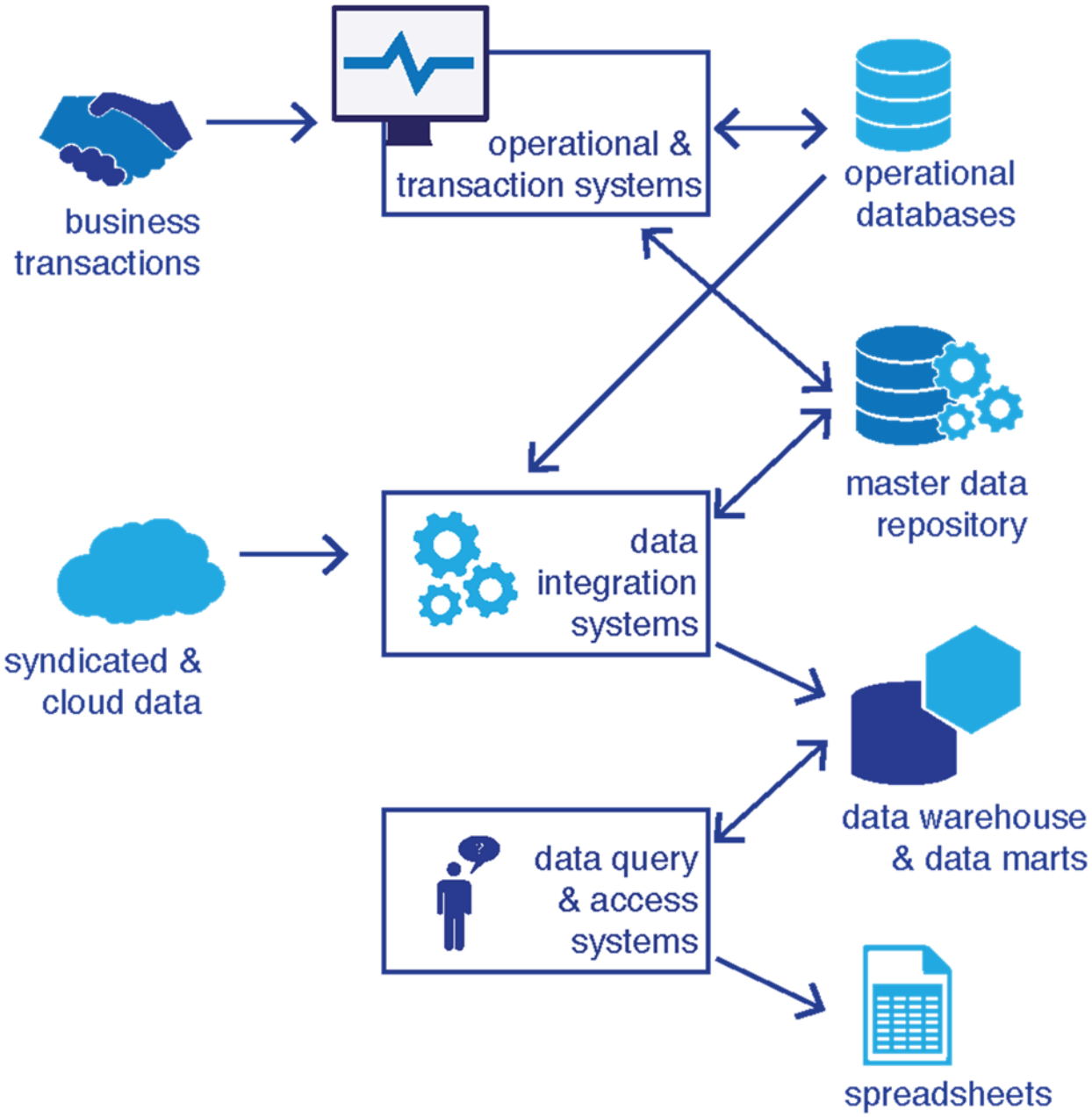
Data sources are one of the critical components of data and integration architecture. It is necessary to understand those sources from more than a simple database point of view. Multiple sources are common in any single organization. Data sources do not come integrated, but disintegrated, and the purpose of the architecture is to facilitate integration. Common types of systems include:

- Business domain systems—the non-integrated operational and transactional systems of various business areas such as finance, purchasing, human resources, and payroll
- Legacy systems—all of the older systems still operating in the information systems environment
- End-user data—spreadsheets and end-user databases
- Hosted systems—cloud, application service provider (ASP), and outsourced
- Web systems—Internet, intranet, and Web services data opportunities
- Social media systems—Twitter, Facebook, and other sites from which you can gain much customer and brand insight
- Email—an important data source especially for sentiment analysis and regulatory compliance monitoring
- Big data systems—the increasing volume and variety of data that is available and accessible today

Some of these data sources are external, meaning outside the control of the organization, including hosted systems, social media systems, and Web systems.

Access and Delivery

Retrieving Data



Access and Delivery

Retrieving Data

DATA DISTRIBUTION

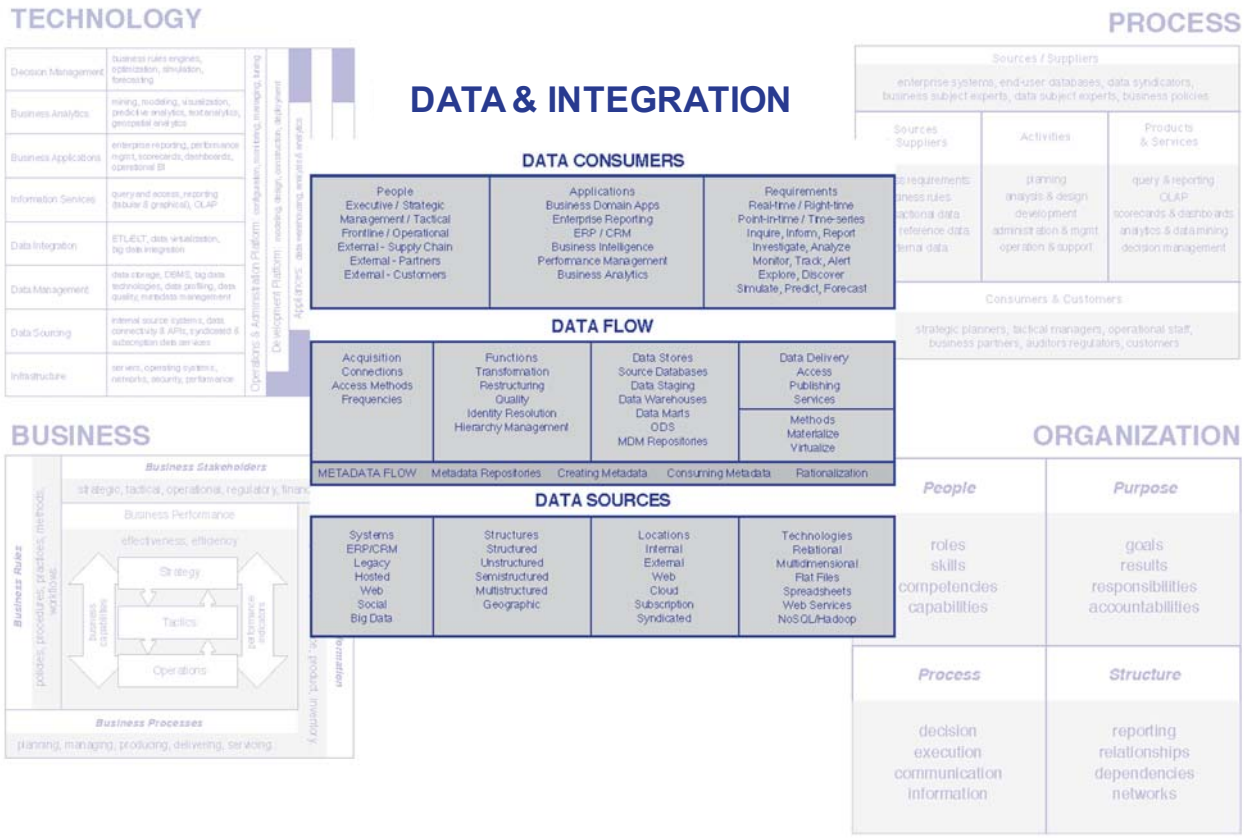
Data distribution prepares data for consumer access. Distribution often includes data integration and data cleansing activities and uses data stores such as warehouses and data marts. Ultimately the goal is to prepare data for access by separating it from the operational environment and representing it as schema or views that are sharable, understandable, and optimized for access.

CONSUMER ACCESS

Consumer access encompasses the activities of retrieving data, both by people and by applications, for the variety of BI uses that range from simple query to analytics and decision management. Architecture must strike the right balance between ease of access and the constraints of data privacy and security.

How Data & Integration Architecture Fits In

Integrated Information Fit



How Data & Integration Architecture Fits In

Integrated Information Fit

INTEGRATED INFORMATION FIT

Information integration is critical to the rationalization of data sources within the enterprise information architecture. Having siloed or disparate data decreases the value of data and information as an asset. Reviewing the different layers of architecture, the shared nature of the BI and analytics architecture becomes apparent. Information is used by multiple and different business processes.

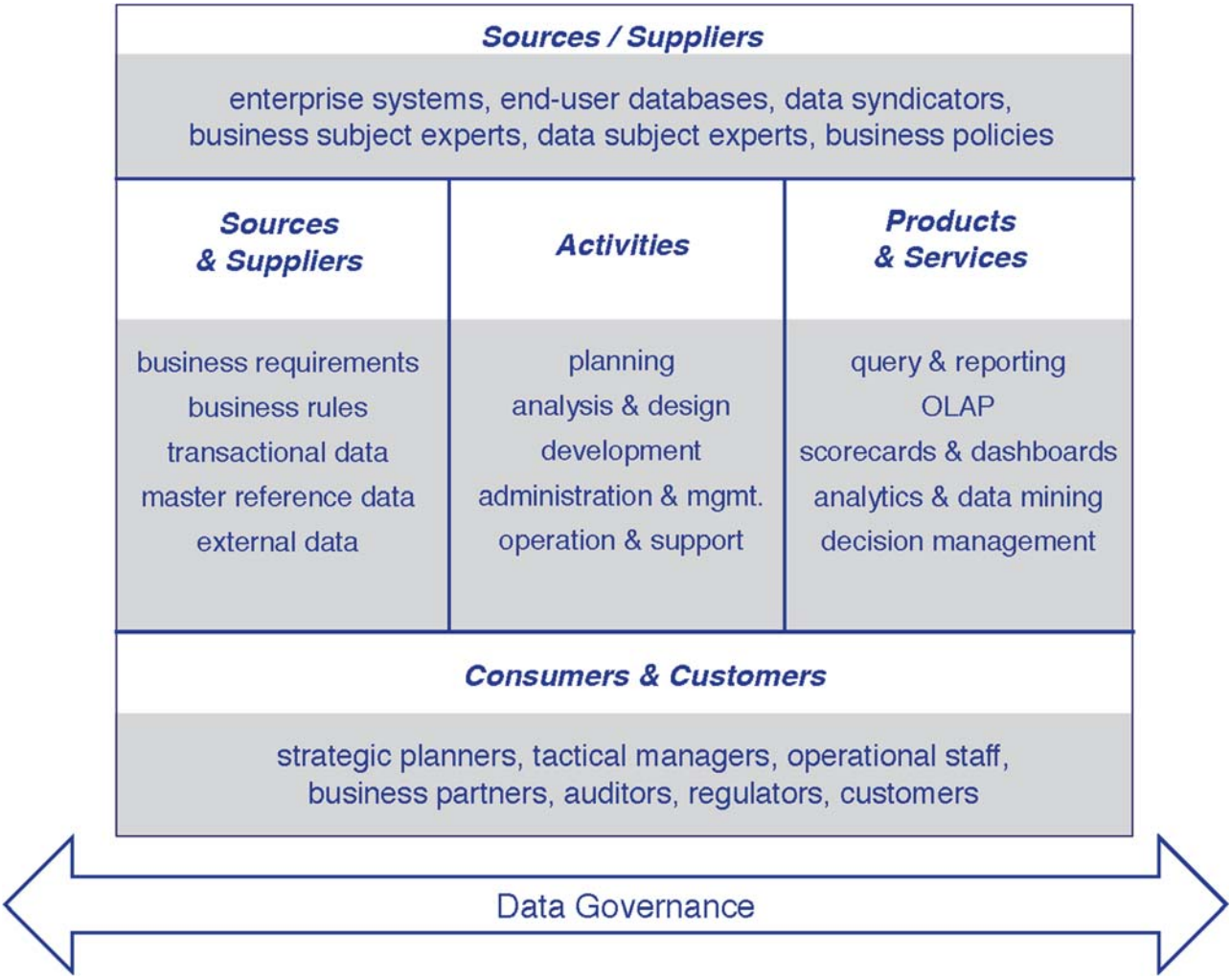
tdwi
Transforming Data With Intelligence™

Module 5

Architecting Process

Process Architecture Concepts

Framework for Process Architecture



Process Architecture Concepts

Framework for Process Architecture

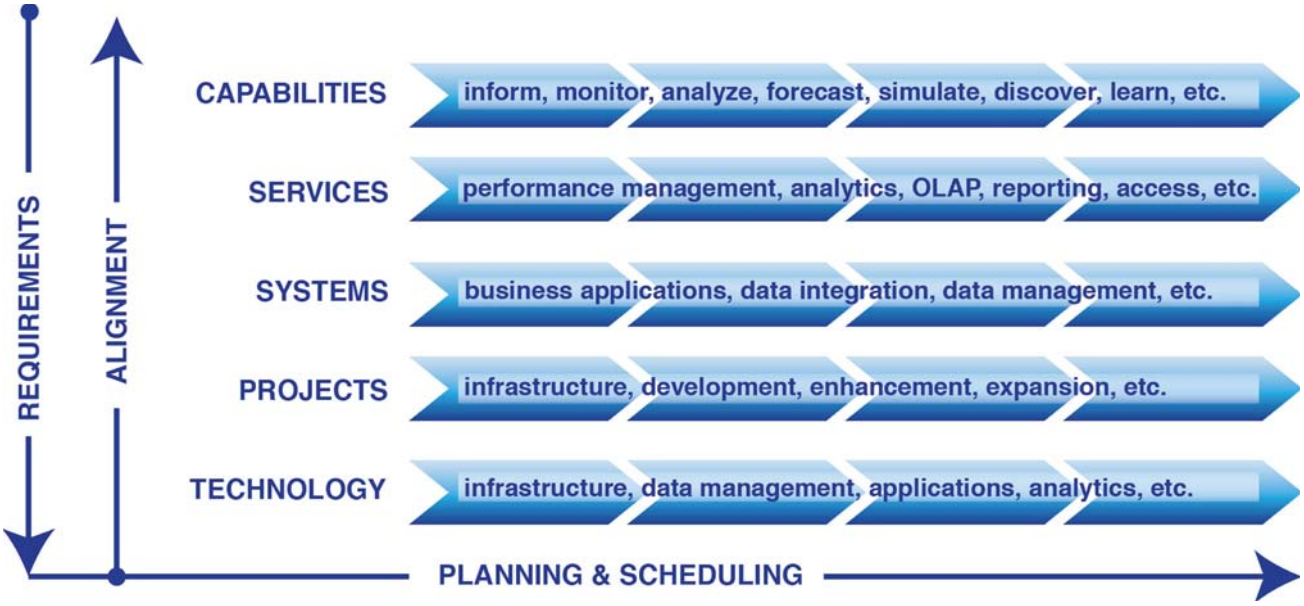
PROCESS IN BI AND ANALYTICS

BI and analytics processes have all the same components as described earlier for business processes. They produce BI and analytics products for BI and analytics customers. They need materials; they acquire materials from data sources and perform value-added activities to create products from materials. BI and analytics processes are business processes; therefore, the standard model of business processes is the architectural basis of BI and analytics processes.

Many of the components of the process architecture have been addressed in other modules of this course—sources, consumers, inputs, and products. This module will focus on the activity components of the process architecture and on data governance.

Planning

Priorities and Resources



Planning

Priorities and Resources

GOVERNANCE OF THE BI AND ANALYTICS PROGRAM

The BI and analytics lifecycle begins with planning. Planning processes determine what will be built, in what sequence, and with what resources. The process that establishes priorities and resources is the cornerstone of the process architecture. It governs the program management function.

As we saw earlier in this course, BI and analytics solutions are not developed as part of a "big bang" project with enterprise focus. Rather, they are developed in smaller pieces through increments. The project roadmap, an output of the planning process, defines and schedules the increments.

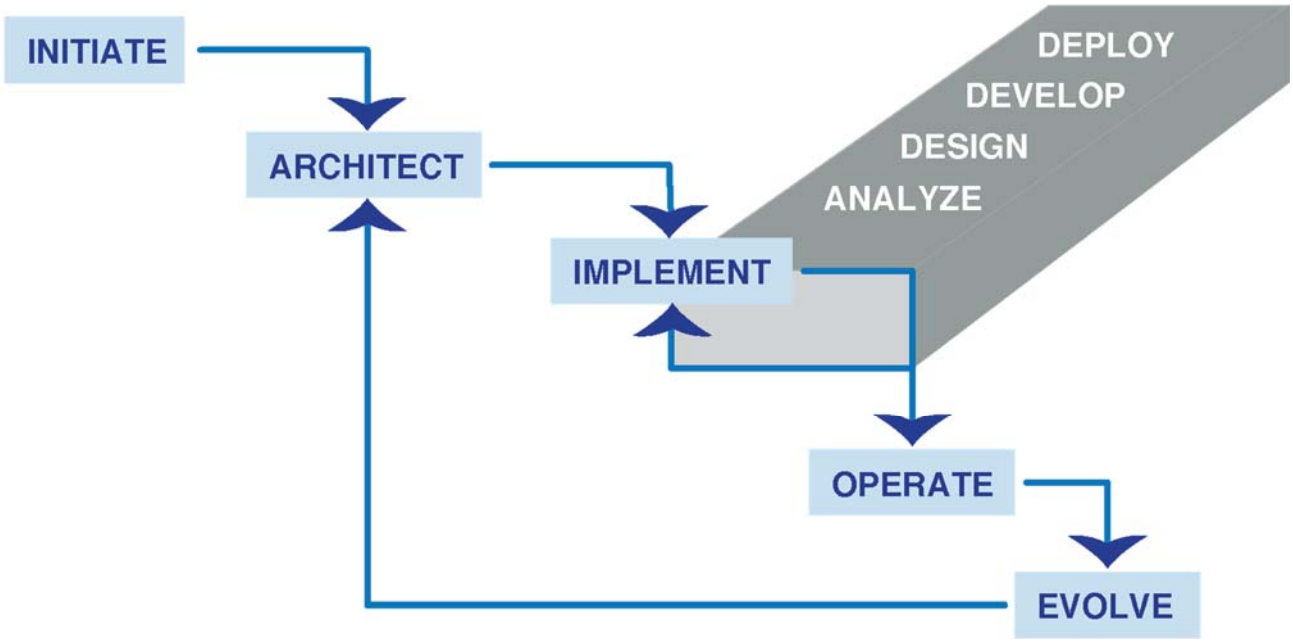
Small organizations may establish a budget and leave it to the program manager to determine priorities for development. On the opposite extreme, large organizations often feature multiple governing bodies, such as executive committees and steering committees, which follow a procedure to establish priorities and resources.

Process architecture also dictates the outputs of the planning process. Minimally, these include:

- A program roadmap
- Development schedules
- Defined scope for each project
- Budgets
- Resources for development, maintenance, and support
- A process for revising and evolving the program management function

Development

Repeatable Processes



Development

Repeatable Processes

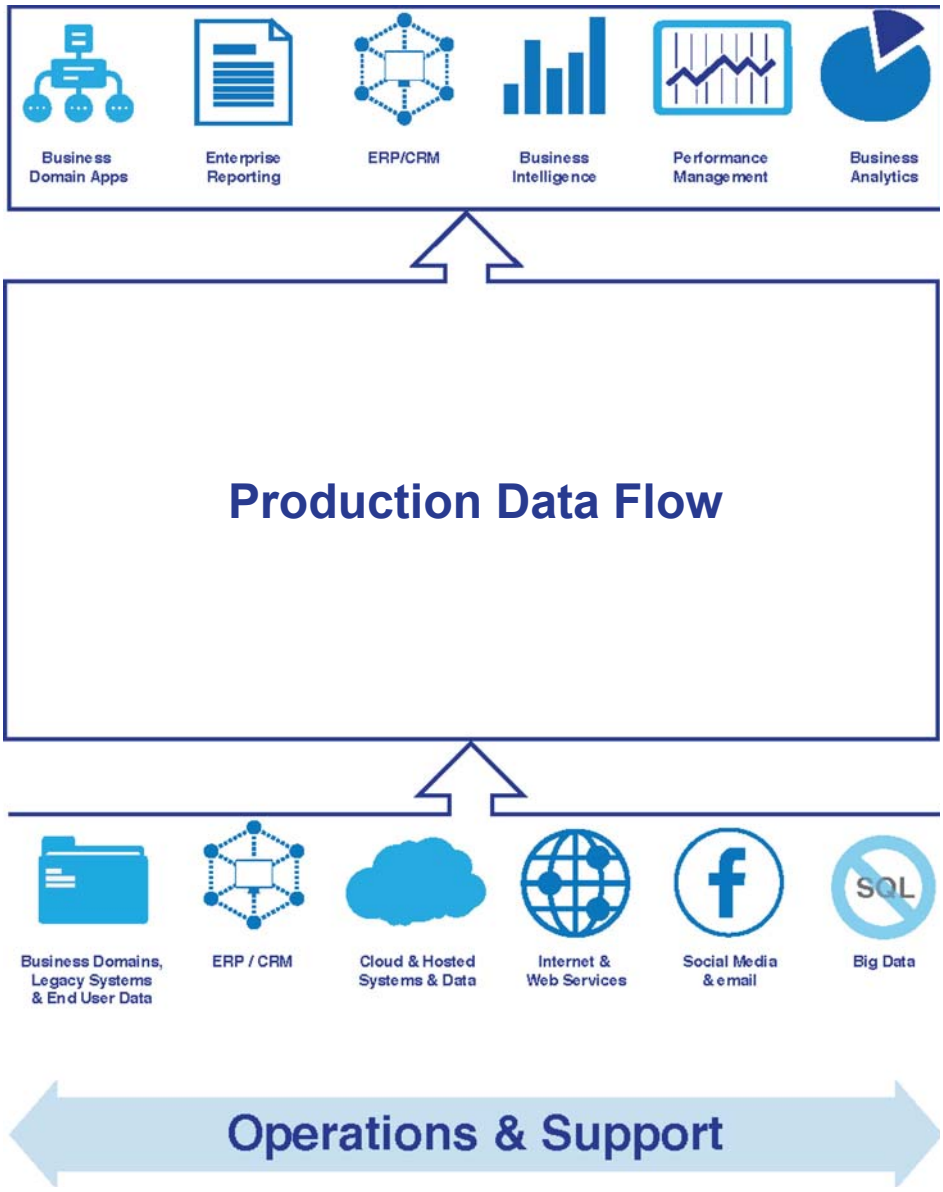
ITERATIVE DEVELOPMENT

The process architecture defines the development approaches that will be used for development projects. As we will see in the next pages, different approaches may be called for based on the type and scope of the project. They may range from tightly scoped agile development efforts to larger and more broadly scoped development efforts. Process architecture describes what types of approach are appropriate in the organization and how they are matched to the type and scope of the project.

Regardless of approach, the goal is iterative development. Solution development is iterative in two senses. First, each solution developed will follow a repeatable process, enabling predictable planning, resource allocation, management, and maintenance. Second, many project approaches follow an internal process of iteration and refinement. This provides valuable business feedback, leading to improved quality in the final solution.

Operations

Production Data Management



Operations

Production Data Management

DATA INTEGRATION OPERATIONS

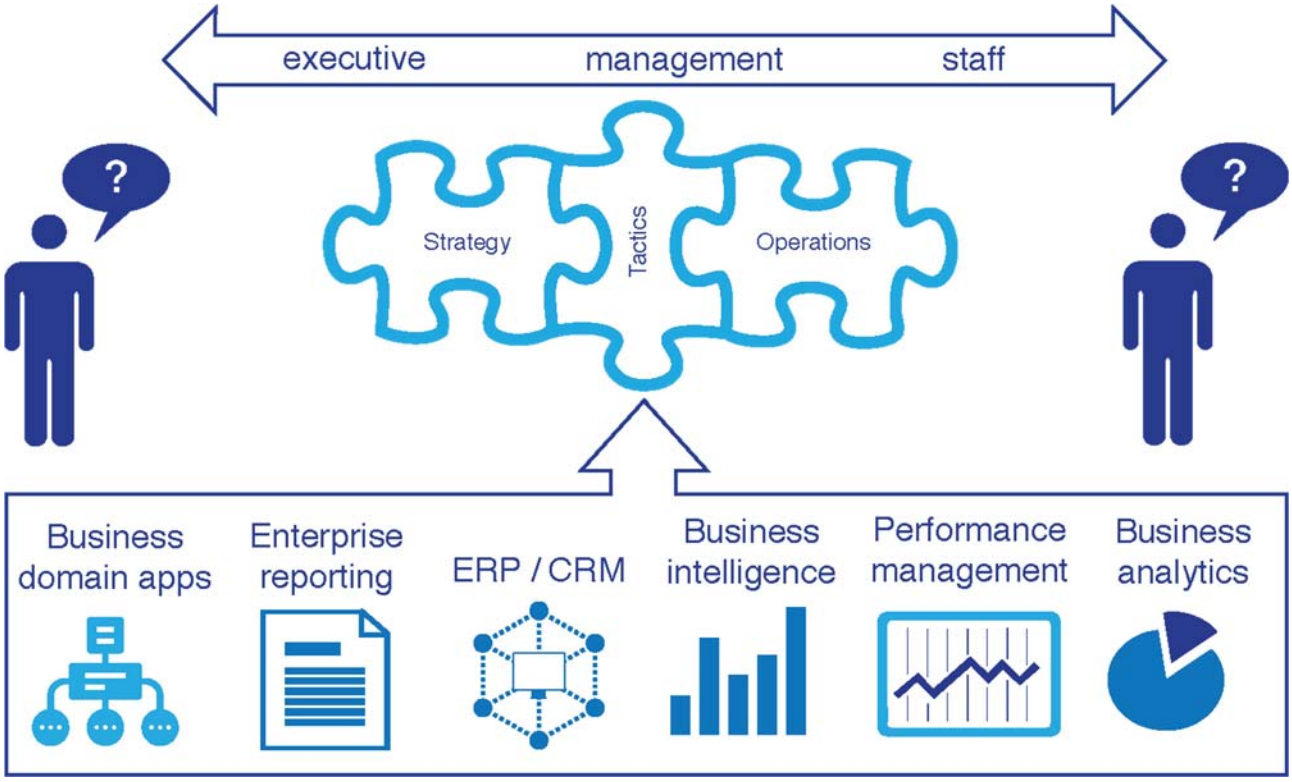
BI and analytics products and services depend on timely and automated data integration processes to stay valuable. Data integration processes include the acquisition, transformation, cleansing, and loading of data, but these processes do not happen without operational support.

Day-to-day support, system maintenance, and system monitoring are all important to integration systems as much as they are to any other system. Integration systems also require operational support not common to all systems in areas such as periodic audits, growth management, and capacity planning.

Proactive change management is another concern in supporting data integration. The data sources used in integration will change. The requirements, priorities, expectations, and needs of the business will change. Technology will change. Managing change from multiple sources while operating, maintaining, and adding functionality to a system is a challenge. Formal change management processes will help to meet that challenge.

Products and Services

Consumers and Applications



Products and Services

Consumers and Applications

PEOPLE PERSPECTIVE

Process architecture must recognize the variety of consumers and their data needs. Consumers include both applications and people. People who consume data may include:

- Executives
- Managers
- Frontline staff
- Supply chain partners
- Business partners
- Customers

APPLICATIONS PERSPECTIVE

Much of data consumption is by application systems—not directly by people. Data integration architecture needs to recognize and support the variety of application consumers that you have or expect to have. Common consuming applications include:

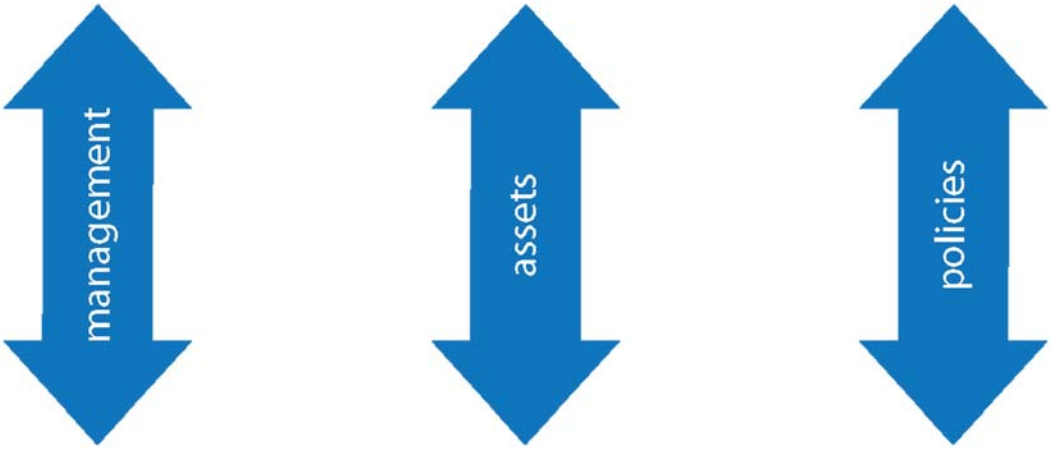
- Business domain applications—systems used to manage finance, human resources, manufacturing, product inventory, service delivery, sales, marketing, and other business functions and processes
- Enterprise reporting systems—for compliance reporting, financial reporting, and other reporting functions that depend on integrated cross-functional and cross-application data
- ERP and CRM systems—vendor-supplied systems that need to synchronize overlapping data with external applications
- Business intelligence systems—providing integrated data for decision support, integrated reporting, OLAP, and other common BI functions
- Performance management systems—supporting metrics-driven business management with measurement data and performance scorecards and dashboards
- Business analytics systems—supporting decision analysis, discovery analysis, and predictive analytics

Data Governance

Data Governance Defined

Data governance is an emerging, cross functional management program that treats data as an enterprise asset: A collection of corporate policies, standards, processes, people and technology essential to managing critical data to a set of goals.

Maria Villar & Theresa Kushner



Data governance is the organization and implementation of policies, procedures, structure, roles, and responsibilities which outline and enforce rules of engagement, decision rights, and accountabilities for the effective management of information assets.

John Ladley & Danette McGilvray

Data Governance

Data Governance Defined

DATA GOVERNANCE

Maria Villar and Theresa Kushner define data governance as a "management program that treats data as an enterprise asset: a collection of corporate policies, standards, processes, people, and technology…"[1]

John Ladley and Danette McGilvray define data governance as "policies, procedures, structure, roles and responsibilities which outline and enforce rules of engagement, decision rights, and accountabilities for the effective management of information assets."[2] Considering these definitions together provides a good sense of the core of data governance as an asset management practice with attention to data-related policies and processes.

PROCESS ARCHITECTURE

There are multiple definitions of data governance, but the key concept is its relationship to process architecture. The definitions below expose the role data governance plays in process architecture:

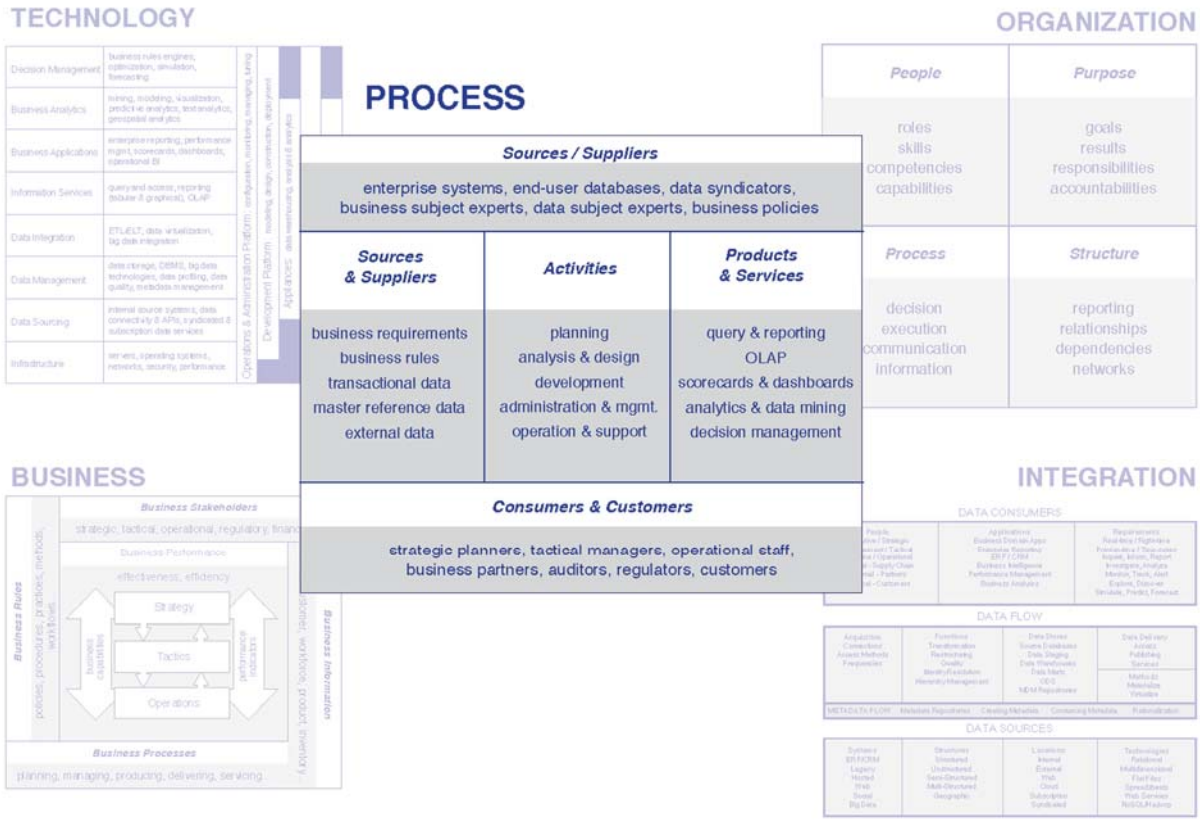
- Gwen Thomas defines it as "execution and enforcement of authority over the management of data and data-related processes."[3]
- Alex Berson and Larry Dubov define it as "a process focused on managing the quality, consistency, usability, security, and availability of information."[4]
- David Loshin describes data governance as "a program for defining information policies that relate to the constraints of the business …"[5]

Data governance influences how processes are defined and developed, which, in turn, influences the what and how of sources and suppliers, activities, products and services, and information consumers in the BI and analytics architecture.

1 Source: *Data Governance Fundamentals,* www.elearningcurve.com. Villar & Kushner are authors of *Managing Your Business Data*.
2 Source: *Executing Data Quality Projects* by McGilvray. Ladley is a well-known EIM consultant.
3 Source: *Data Governance Defined* by John Ladley (http://admin.dataversity.net/data-governance-defined).
4 Source: *Master Data Management and Customer Data Integration for a Global Enterprise* by Berson and Dubov.
5 Source: *Master Data Management* by Loshin.

How Process Architecture Fits In

Process Fit



How Process Architecture Fits In

Process Fit

PROCESS FIT

The process architecture enables and integrates multiple components. The integration of these components results in the ability to acquire material from sources and perform value-added activities to create products from materials, for example. Processes for BI and analytics development, use, and operations can be used across technology platforms and business groups.

This module did not dive deep into sources/suppliers, inputs/materials, and consumers as these are shared components across business, organization, integrated information, and technology architectural components.

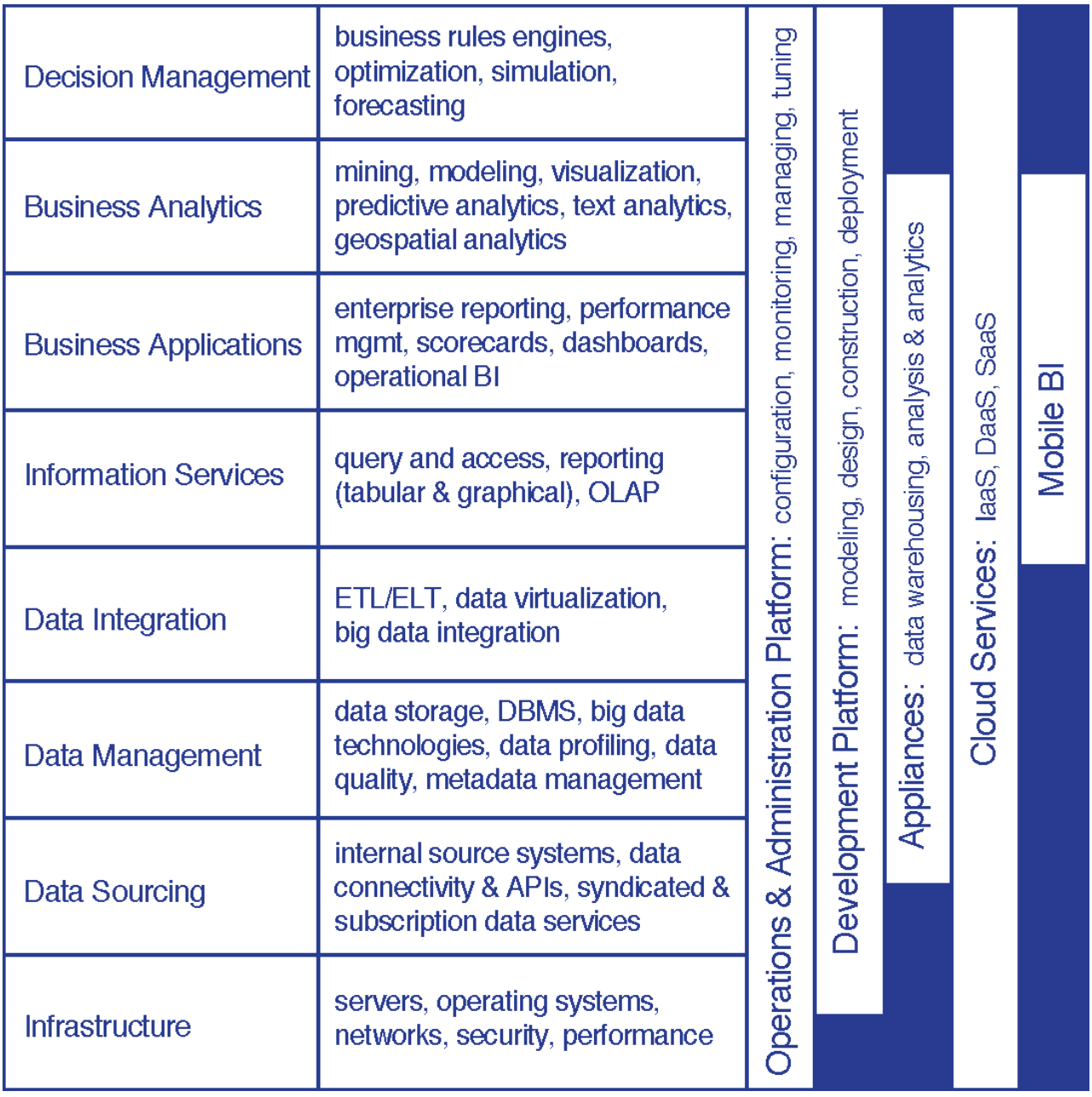
tdwi
Transforming Data With Intelligence™

Module 6

Architecting Technology Platforms

Technology Architecture Concepts

Business Intelligence and Analytics Technology Stack



Technology Architecture Concepts

Business Intelligence and Analytics Technology Stack

THE TECHNOLOGY STACK

The diagram on the facing page illustrates the BI and analytics technology stack.

Working from the bottom of the diagram to the top follows a progression from foundation technology to value through business capabilities.

Each of the technologies in the stack must provide specific features and functions that make it usable and support people working with the technology.

Required features and services include the following:

- Operations and administration platforms with functions to monitor, manage, tune, and configure systems and technology at every layer of the stack.
- A development platform or multiple platforms that include functions for modeling, design, construction, and deployment of applied technology systems spanning from data sourcing to decision management.

Desired features and services may include:

- Appliances for data warehousing and/or analytics that provide an integrated set of servers, storage, operating systems, and database management preconfigured for easy installation and high performance. Analytics appliances integrate software for analytics capabilities into the configuration.
- Cloud services to support hosting of infrastructure-as-a-service (IaaS) for servers and operating systems, data-as-a-service (DaaS) for storage and DBMS, or software-as-a-service (SaaS) for specific information services and business applications.
- Mobile BI with communications, compression, visualization, and other features needed to deliver information services and business applications to smartphones and tablets.

Platforms

Networks and Servers

Platforms

Networks and Servers

NETWORKS

Decisions about network design and capacity can directly impact how business analytics and intelligence perform. Network capacity is an enterprise resource the same as data and information. Networks connect all applications and repositories of data together, internally and externally. For instance, networks are a critical success factor in cloud models. Physically, data centers may be large distances away from each other, and the success of cloud models depends on having the right network capacity and redundancy to ensure service levels are met.

Business analytics and intelligence products may need to process large amounts of data from point to point, especially with the incorporation of big data sources. Without the right network design and capacity, business analytics and intelligence product and application quality cannot be maintained. High amounts of computing power will provide little benefit if data cannot quickly be moved from point to point. Firewalls are also a consideration in design and would be influenced by data security requirements.

Because networks are a shared enterprise resource, having an up-to-date network diagram including capacity and connection is necessary to support the network requirements of business analytics and intelligence products and applications.

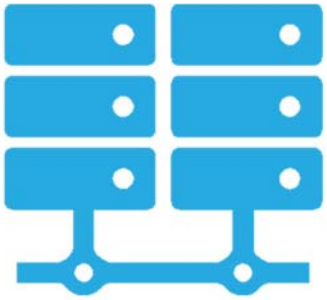
SERVERS

Different types of servers used in the business analytics and intelligence architecture are addressed later in the module, but server and network design go together to ensure performance. A fast network without the associated computing power will not meet operational service levels. An end-to-end network and server design is essential to architecting the correct infrastructure within the technical architecture.

Servers

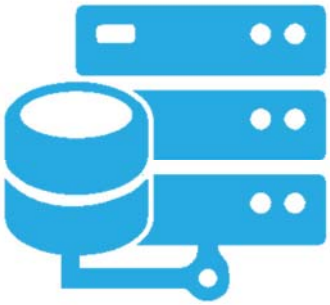
Many Options

Cloud



Appliance

Virtualized Server



Dedicated Server

Servers

Many Options

SERVERS

Servers are an integral part of any technology infrastructure. Servers provide the computing power for business analytics and intelligence processing. The amount of processing resources in the infrastructure should be architected based on operational service levels. Examples of operational service levels could include data availability for analysis or query response time. A business requirement to have real-time data, for instance, would directly influence the computing power needed.

Business requirements are not the only influence on server choice. The overall infrastructure strategy will also determine the organization's approach to architecting the infrastructure. Dedicated servers, virtual servers, data appliances, and cloud options exist.

DEDICATED

Dedicated servers refer to processing resources used for a single purpose. It is not uncommon to have a server dedicated for data integration purposes to ensure operational service levels. In some cases, dedicated servers may also refer to a physical server.

VIRTUALIZED

Virtual servers are used where flexibility is needed. There are several advantages to using virtual servers, such as scaling up and down quickly, restoration in event of a disaster, and environmental friendliness. Virtual servers can be easy to provision.

APPLIANCE

The appliance is a category of server that is simple to use and configured for optimal workloads. The appliance is an integrated server including storage, operating system, and DBMS.

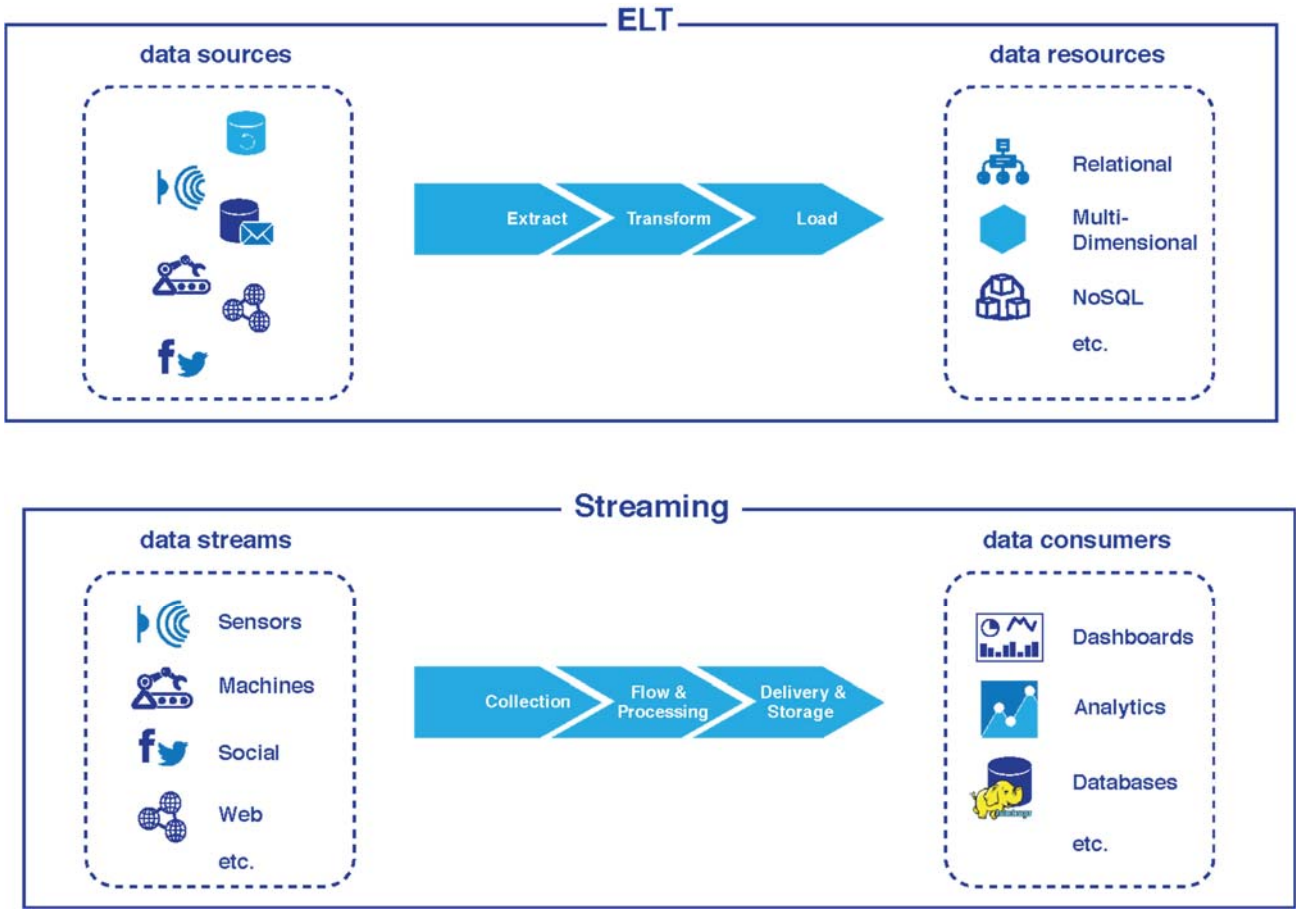
CLOUD SERVER

Cloud servers work similarly to physical servers but the functions they provide can be very different. When using a cloud server, virtual server space is rented versus purchasing physical servers. Cloud servers are often paid for by the hour or Cloud Compute Unit (CCU) based on capacity needed.

Cloud implementations can be public, which is offered to the general public, private, which is kept within an organization's network, and hybrid, which is a combination of public and private.

Data Sources

Batch, Event, Streaming



Data Sources

Batch, Event, Streaming

PROCESSING DATA

The data sourcing component is responsible for acquiring and formatting the necessary data to support analytics use cases. The data may exist in a wide range of source data locations, technologies, and formats. Ingestion must connect to all desired data sources, acquire the data, transform it to meet integration, usability, and quality requirements and then deliver it to a persistence component when data is to be retained and stored.

Ingestion activities can exist in three distinct forms—batch, on-demand, and streaming.

BATCH

Batch ingestion is based on familiar extract, transform, and load (ETL) processes and is triggered by a defined schedule. Data is acquired from sources, transformed as needed, and delivered to a target location.

EVENT

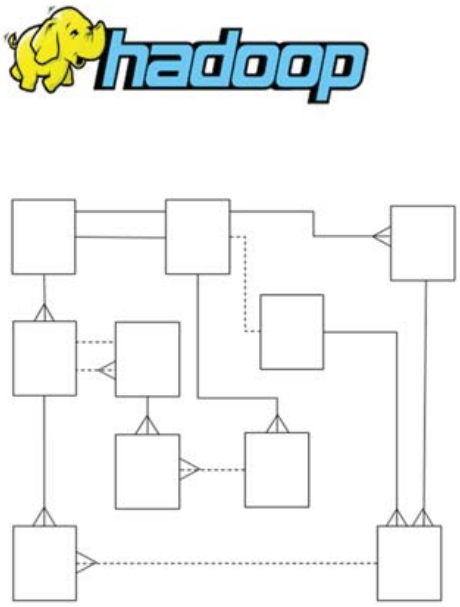
On-demand or event-triggered ingestion uses data virtualization technologies. Data is acquired from sources and transformed in motion as it moves from source to consumer. On-demand ingestion is triggered by a user query.

STREAMING

Streaming technologies acquire, transform, and deliver data in real time. Streaming is effective when working with continuously incoming event data such as sensor data, financial trading data, and social media data. Streaming data may be analyzed in real time, captured and stored for future analysis, or a combination of the two.

Databases and Storage

RDBMS, Hadoop, Nonrelational



Databases and Storage

RDBMS, Hadoop, Nonrelational

DATABASES

Databases are a part of the data management architectural component. Databases are no longer exclusively relational due to the volume, variety, and velocity of big data. Other storage options include Hadoop and nonrelational data stores.

RDBMS

A relational database management system (RDBMS) is a database management system (DBMS) that is based on the relational model (tables, rows, columns, and relationships). An RDBMS is best used with structured and transactional data.

HADOOP

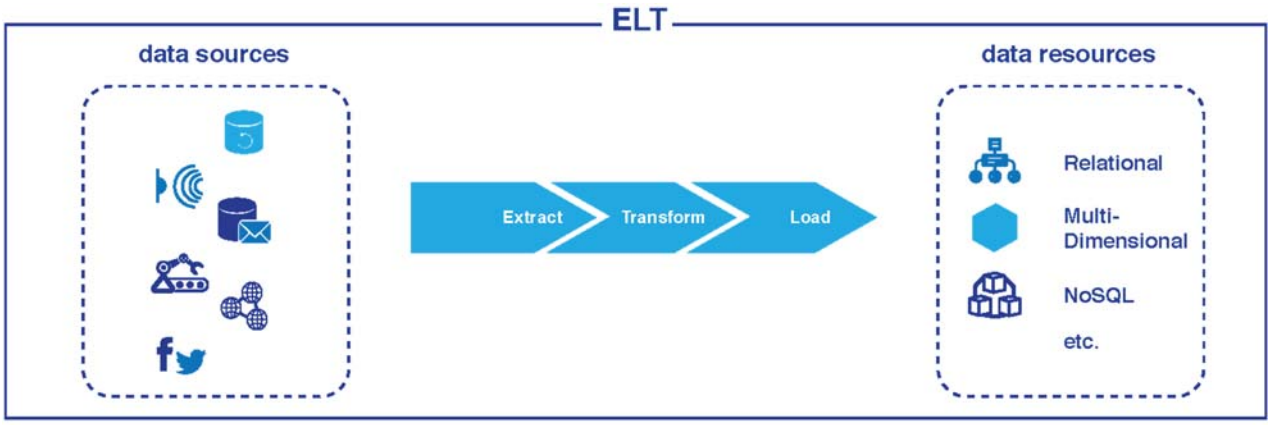
Hadoop is an open source software framework for distributed storage and processing of big data on computer clusters using commodity hardware. The framework used by Hadoop is called MapReduce. A MapReduce job usually splits the input data into independent chunks that are processed by the map tasks in a completely parallel manner. To implement the data storage layer, Hadoop uses a feature known as HDFS—the Hadoop Distributed File System. HDFS is not a file system in the traditional sense and isn't usually directly mounted for a user to view.

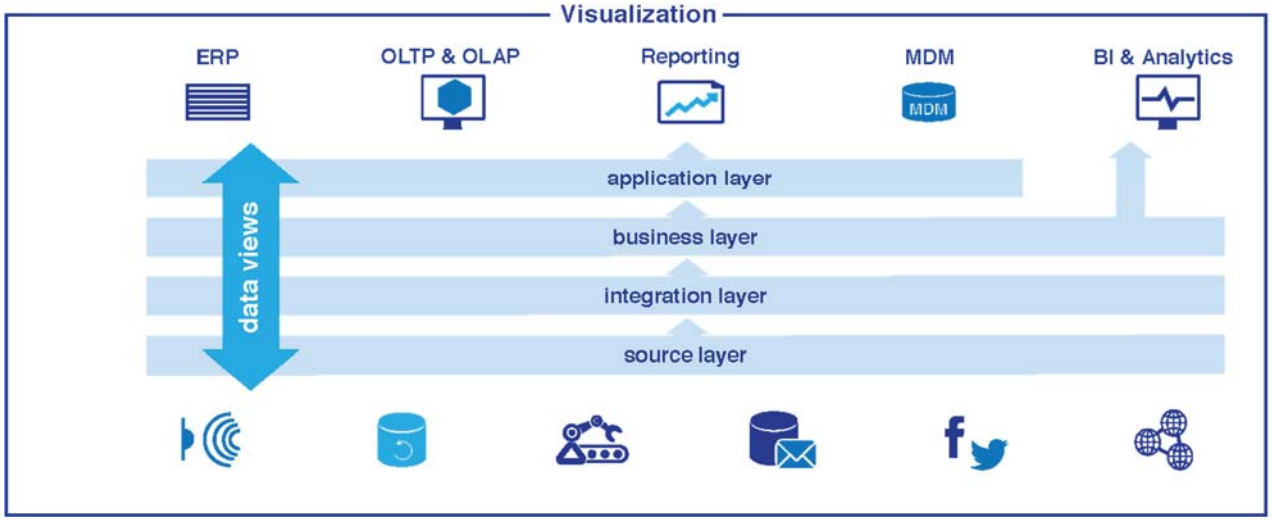
NONRELATIONAL

A nonrelational database is any database that does not follow the relational model. Nonrelational databases are also called NoSQL databases and have increased in popularity with the increased use of big data. Unlike relational models, which require a predefined schema, NoSQL databases can handle a flexible design that makes it easier to ingest data and handle change. NoSQL databases are designed to handle unstructured data and are typically open source.

Integration

Integration Tools





Integration

Integration Tools

INTEGRATION TOOLS

Technically, data integration can happen different ways. Acquisition, transformation, cleansing, and loading of data can occur in different sequences and using different technical approaches.

ETL AND ELT

Extract, transform, and load (ETL) functionality is considered core data integration technology. ETL tools provide data acquisition, change data capture, filtering, formatting, and cleansing capabilities. ETL tools typically depict ETL flows via a graphical user interface referred to as a mapping. ETL tools typically include administration tools for scheduling and mapping administration. ETL is often associated with batch or mini-batch load frequency. ELT is a version of ETL where the focus is on data acquisition and loading first. ELT tools are architected to utilize the power of the database and complete transformation processing within the database platform.

VIRTUALIZATION

Virtualization focuses on logically integrating data at the time of the query. Advantages of virtualization include not having to physically move data for integration. Transformation and cleansing rules would be applied in the integration layer. Virtualization is a flexible option for integration especially when data is accessed less frequently.

Business Analytics Technology

A Variety of Tools

Business Analytics Technology

A Variety of Tools

DISCOVERY AND ANALYTICS MODELING

Many different tools provide analytics capabilities. Even traditional BI tools provide functionality such as analytics modeling. The different categories of analytics tools include:

- Text analytics
- Statistical analysis
- Machine learning
- Social media analytics
- Real-time analytics
- Predictive analytics

Although categories of tools are listed under technology architecture such as business intelligence and business analytics, many of these categories are becoming blurred as functionality across tools tends to increase. Tools that have been traditionally known for business analytics such as SAS or SPSS are being joined in the landscape by new open source and other technology vendors. A quick look at the landscape on the prior page shows the number of technologies available today that can support various analytics capabilities.

Business Intelligence Technology

Variety of BI Tools

Ad Hoc Reporting
Enterprise Reporting

Scorecards

Dashboards

Business Intelligence Technology

Variety of BI Tools

MANY OPTIONS

Business intelligence tools provide a variety of functionality. Multiple BI products can be created with an integrated suite of tools. Core BI products include capabilities such as ad hoc queries, enterprise reports, OLAP, scorecards, and dashboards, as well as visualization. SAP, Microsoft, Oracle, Microstrategy, and IBM represent a vendor-centric option for BI tools. Many other vendors offer best-of-breed options such as Tableau and Qlik. These are not exhaustive lists.

Data Visualization

Presenting Compelling Data Stories

Data Visualization

Presenting Compelling Data Stories

COMMUNICATING WITH IMAGES

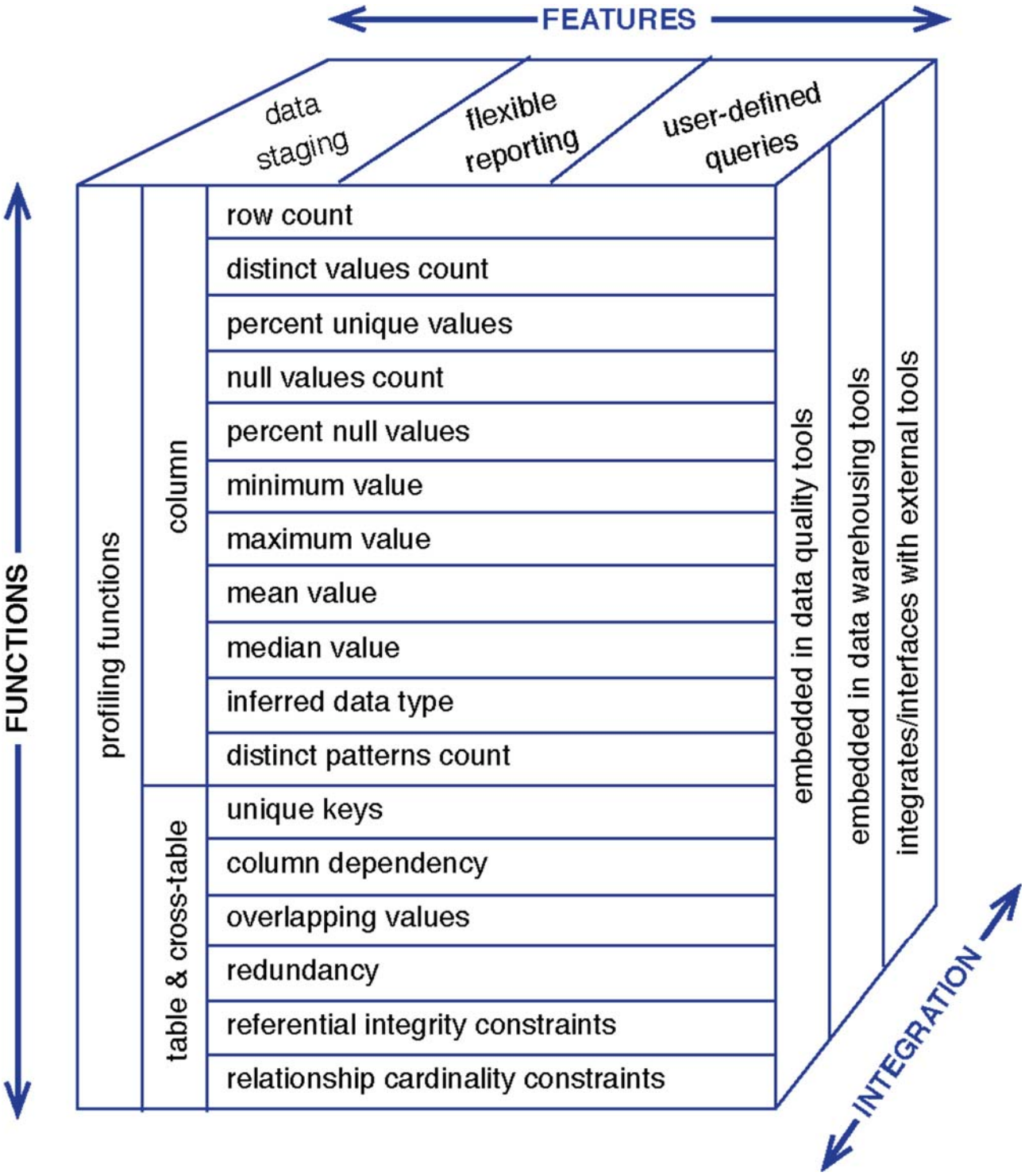
Data visualization is not confined to a single technology. Visualization capabilities exist in data mining tools, traditional BI tools, and Excel, for example. Visualization tools such as Tableau and Qlik have emerged to promote fast analytics and to accelerate the analytics development process. Visualization tools create charts and graphs as well as more complex visualizations such as infographics and data storytelling.

DATA STORYTELLING

Data stories group several related visualizations and connect them with narrative that expresses the analyst/storyteller's interpretation of their meaning. Storytelling makes analytics more real and personal than individual charts and graphs without interpretation. A good data story connects the data and the visuals with cause and effect, elicits personal responses, and drives conversation and interaction.

Data Management

Profiling, Quality, Metadata, Governance



Data Management

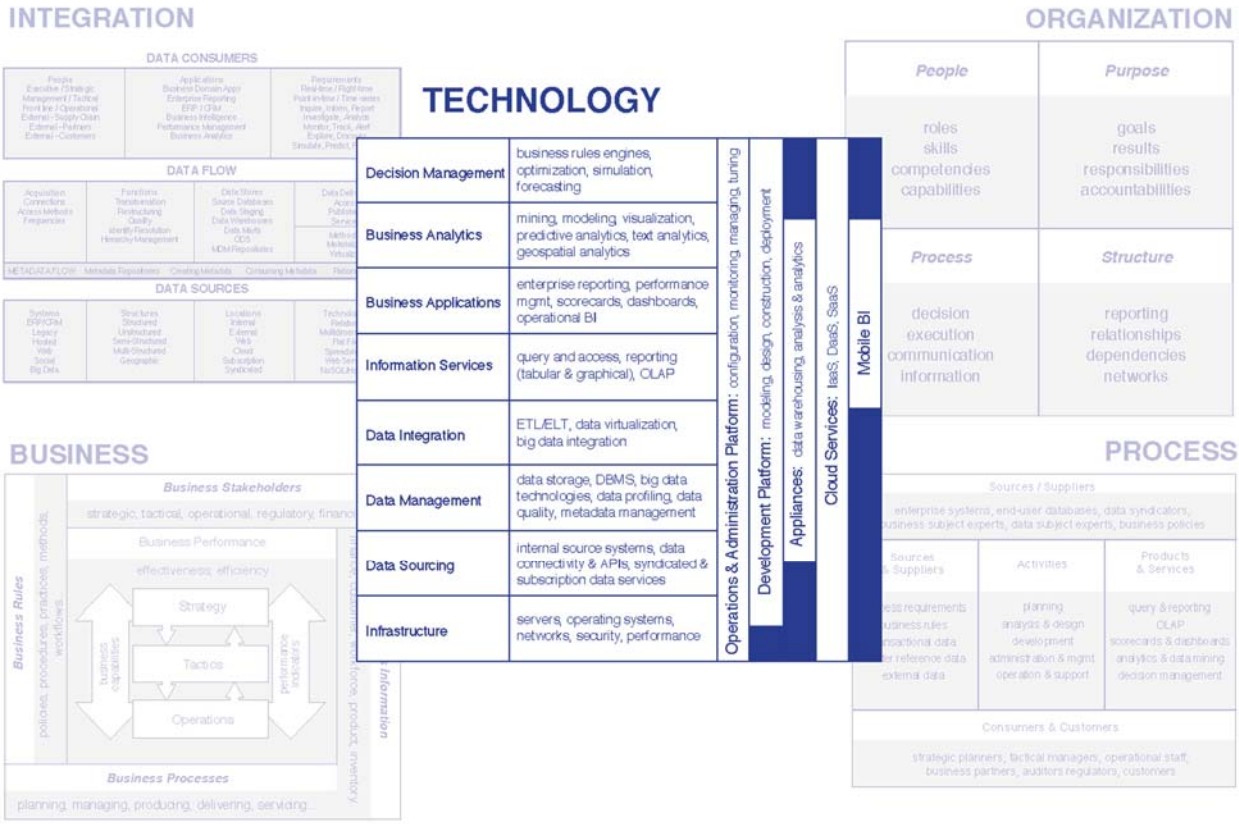
Profiling, Quality, Metadata, Governance

QUALITY AND GOVERNANCE TOOLS

As outlined in process architecture, data governance focuses on managing data as an asset. Functions such as data profiling and metadata management contribute to data quality management and often fall under the umbrella of data governance. Suites of data quality tools exist where profiling, data quality rule management, and metadata management are part of an integrated tool. These technologies may also be provided separately or as part of data integration technology. For instance, data profiling functions and data cleansing functions may also be found in ETL tools. This is yet another category of technology to consider in the technology architecture.

How Technology Architecture Fits In

Technology Fit



How Technology Architecture Fits In

Technology Fit

TECHNOLOGY AS A FOUNDATION

Technology offerings in business intelligence and analytics are constantly changing. Technology enables core capabilities that are necessary to be able to manage data as an asset. Servers, networks, databases, data governance, business intelligence, and analytics tools are necessary for data integration, data curation, data management, and data analysis.

Many stakeholders will use common technology architecture to get the most value from the organization's technology investment. For instance, an organization tends to have an integrated network to connect all employees and locations—without an integrated network the sharing of information and communication would not happen. The technology architecture to support the use of data as an asset is no different. Technology is shared and leveraged across an organization. This does not mean all technology is standard or enterprise, but it means that information does not become an asset unless technology is leveraged.

This page intentionally left blank.