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 Architecture

Principles of BI Design

The Data Warehousing Institute 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: February 2013

TABLE OF CONTENTS

COURSE OBJECTIVES

To learn:

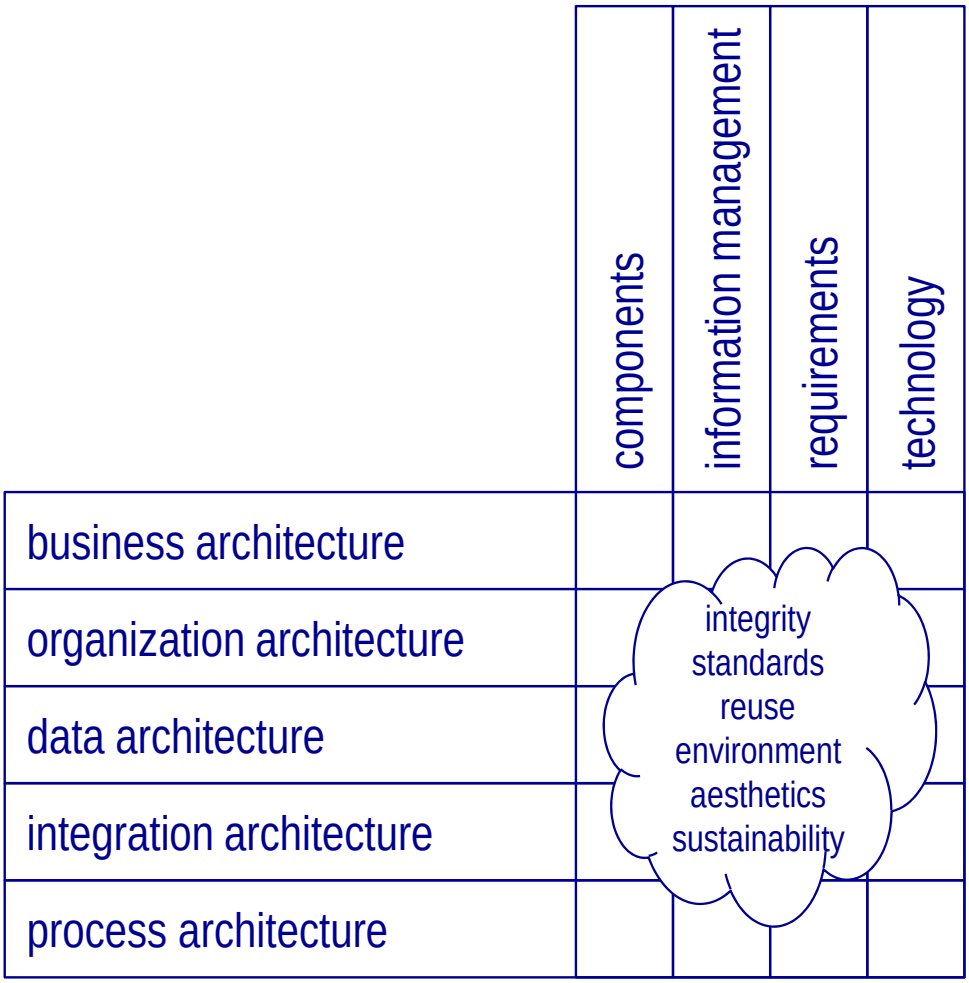
- ✓ ***The full scope of architectural objectives – structural integrity, standardization, reusability, environmental fit, aesthetics, and sustainability***
- ✓ ***A framework to ensure architectural completeness – business, organization, data, integration, and process views***
- ✓ ***A framework to organize BI components – access, analysis, presentation, storage, integration, and data source tiers***
- ✓ ***A framework to organize the information management stack – data, integration, rules, tools, teams, reports, analysis, and application***
- ✓ ***A framework to organize architectural requirements – functional, data, operations, environment, and structural requirements***
- ✓ ***A framework to organize technology requirements – data access, data manipulation, data analysis, reporting, visualization, security, portability, and accessibility***
- ✓ ***Technology trends and BI architecture – cloud, SaaS, open source, appliances, advanced visualization, etc.***
- ✓ ***Organizational options for best fit of BI into your culture – conglomerate, cooperative, and centralized***
- ✓ ***Data integration options in BI architecture – bus, hub-and-spoke, hybrid, federation, and virtualization***

Module 1

BI Architecture Concepts

Dimensions of BI Architecture

Layers and Views



Dimensions of BI Architecture

Layers and Views

BI ARCHITECTURE STACK

BI architecture can be viewed as a five-layer stack with business at the top of the stack and process at the bottom. Connecting from business to process involves intermediate levels of organization, data, and integration.

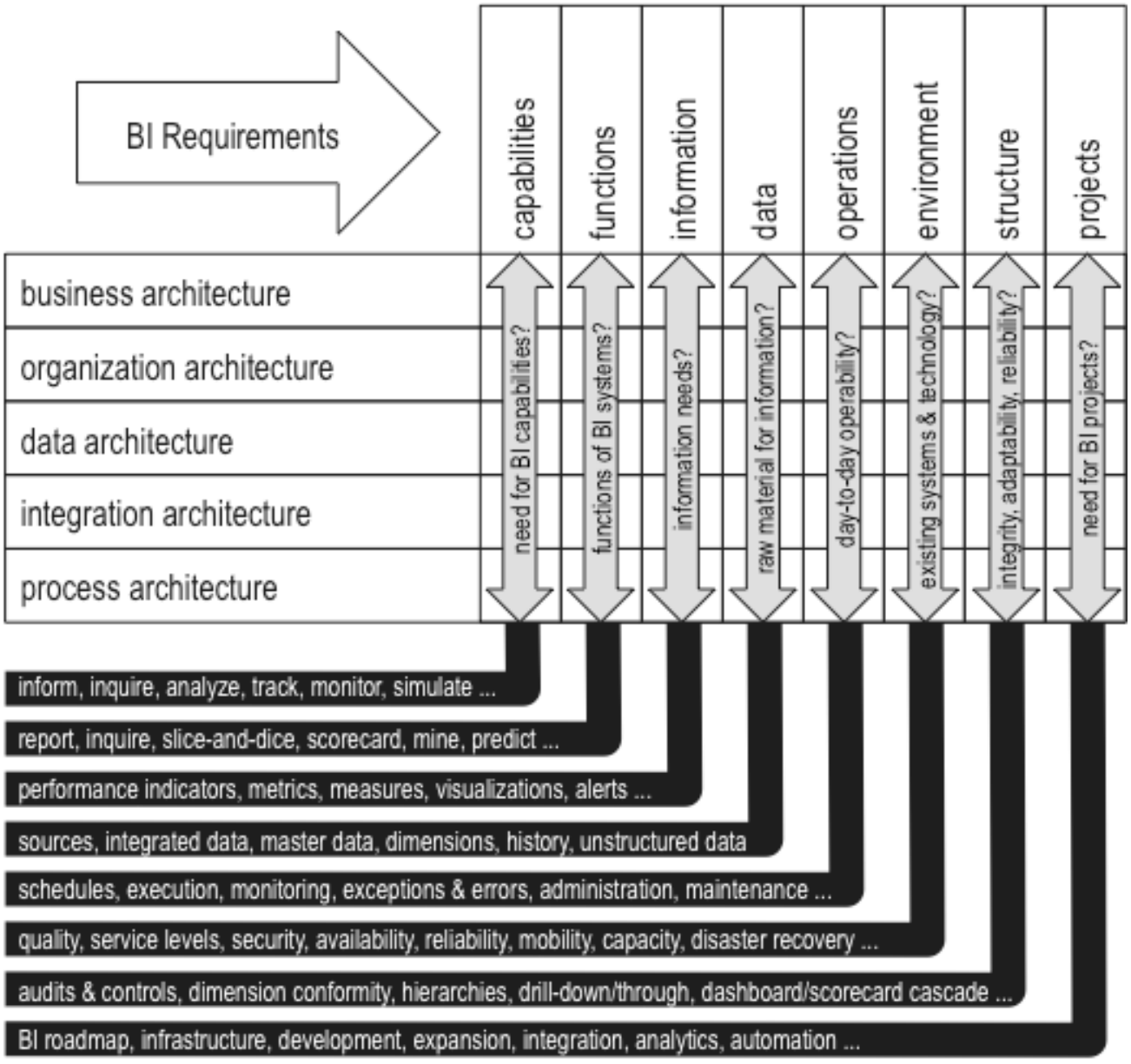
BI ARCHITECTURE PERSPECTIVES

Four slices of BI architecture – components, information management, requirements, and technology – are are viewed orthogonally. They're not presented as layers in the stack because they cut across all five layers. These vertical slices each provide different perspectives into the business, organization, data, integration, and process aspects of BI. At each intersection, the architect is responsible to:

- Provide frameworks, structures, and guidelines to assure integrity of BI systems.
- Define standards to ensure regulatory compliance and to assure consistency and cohesion among all parts of BI systems.
- Provide standards and frameworks that maximize opportunity for reuse of BI components.
- Maintain a map of the BI environment that enables smooth integration of new increments.
- Define standards of access, presentation, visualization, interface, and interaction to assure esthetically pleasing BI systems with consistency of look and feel throughout the BI environment.
- Develop architectural constructs from a forward-looking point of view with an eye toward long-term sustainability.

Architectural Frameworks

A Requirements Framework



Architectural Frameworks

A Requirements Framework

DRIVING THE RIGHT PROJECTS

BI architecture is connected with requirements in two ways:

- Architecture has requirements. BI architecture, for example, has requirements to support needed business capabilities such as the ability to inform people and processes and to montior critical business performance indicators.

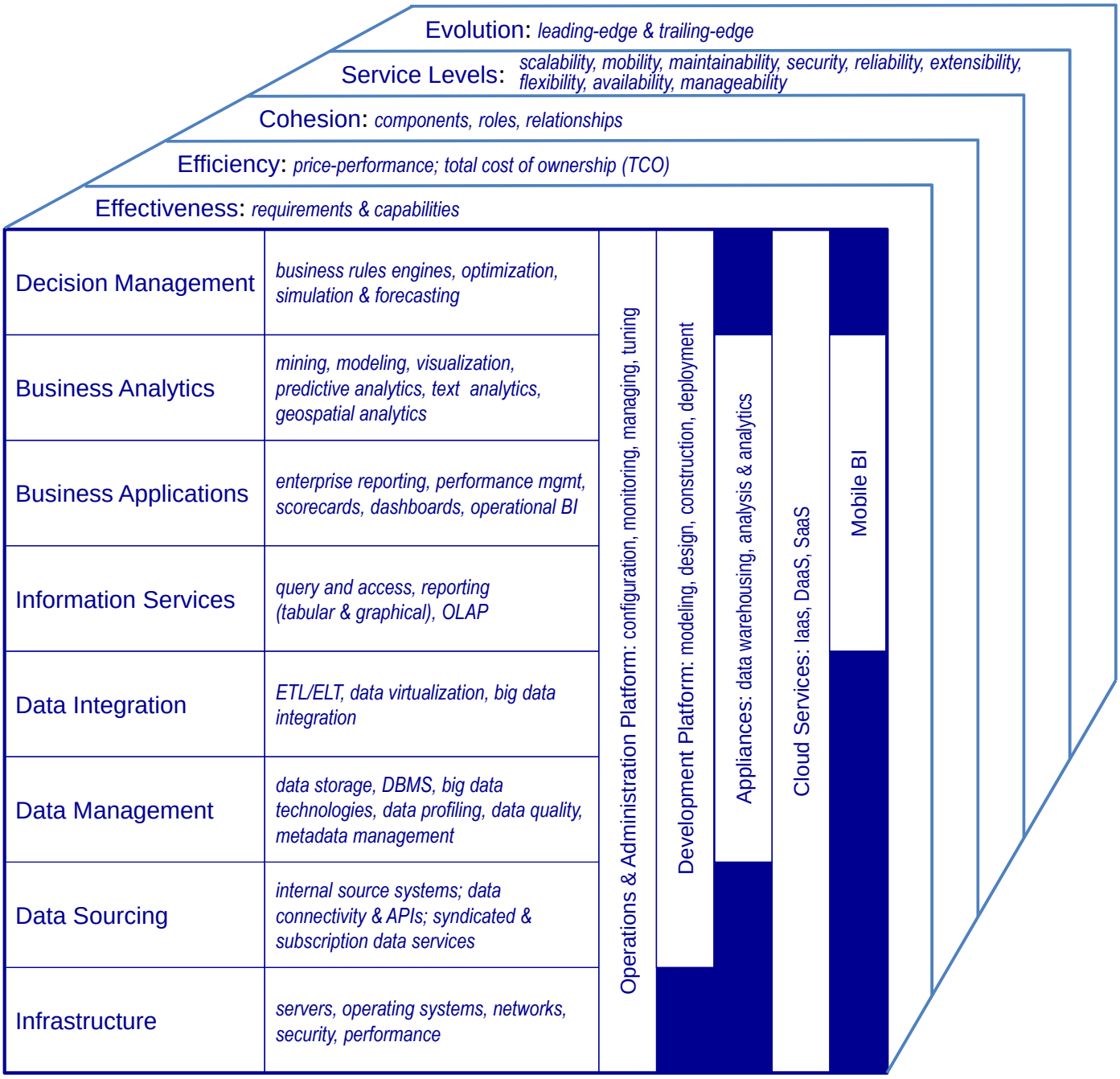
- Architecture frames project requirements. Knowing business-driven architectural needs helps to identify needed projects. Language, constructs, standards, and constraints expressed by architecture help to define requirements of individual projects.

The framework shown on the facing page is a tool to drive the right kinds of projects through BI architecture.

Each of the categories – capabilities, functions, information, data, etc. – represents an aspect of BI requirements that is driven to varying degrees by each of the five layers. Questions for each category are shown within the matrix. Examples for each category are shown below the matrix. Consider the questions from two perspectives: What project needs do they help to identify? What standards, frameworks, and guidelines should architecture provide to help shape individual project requirements?

Architectural Frameworks

A Technology Framework



Architectural Frameworks

A Technology Framework

THE TECHNOLOGY STACK

The diagram on the facing page illustrates the BI technology stack. Working from the bottom of the diagram to the top follows a progression that begins with foundation technology and leads to value through business capabilities.

- **Infrastructure** includes all of the hardware and foundation software that is needed.
- **Data sourcing** includes the systems and databases from which data is obtained.
- **Data management** includes data storage and database management systems, big data technologies such as NoSQL, and data profiling and quality tools.
- **Data integration** includes ETL and variations, data virtualization tools, and technologies for big data integration
- **Information services** technologies range from query languages to GUI-based query and reporting tools and OLAP technologies.
- **Business applications** technologies enable enterprise reporting, performance management systems, dashboards, scorecards, and operational BI with real-time feedback
- **Business analytics** technologies are used for data mining, analytic modeling, data visualization, and advanced analytics
- **Decision management** technologies are needed to implement decision management systems.

USABLE TECHNOLOGY

Every technology must have specific features and functions that make it usable. Required features and services include:

- Operations and administration platforms
- A development platform

Desired features and services may include:

- Appliances for data warehousing and/or analytics
- Cloud services for any combination of information as a service (IaaS), data as a service (DaaS), and software as a service (SaaS)
- Mobile BI

SUSTAINABLE TECHNOLOGY

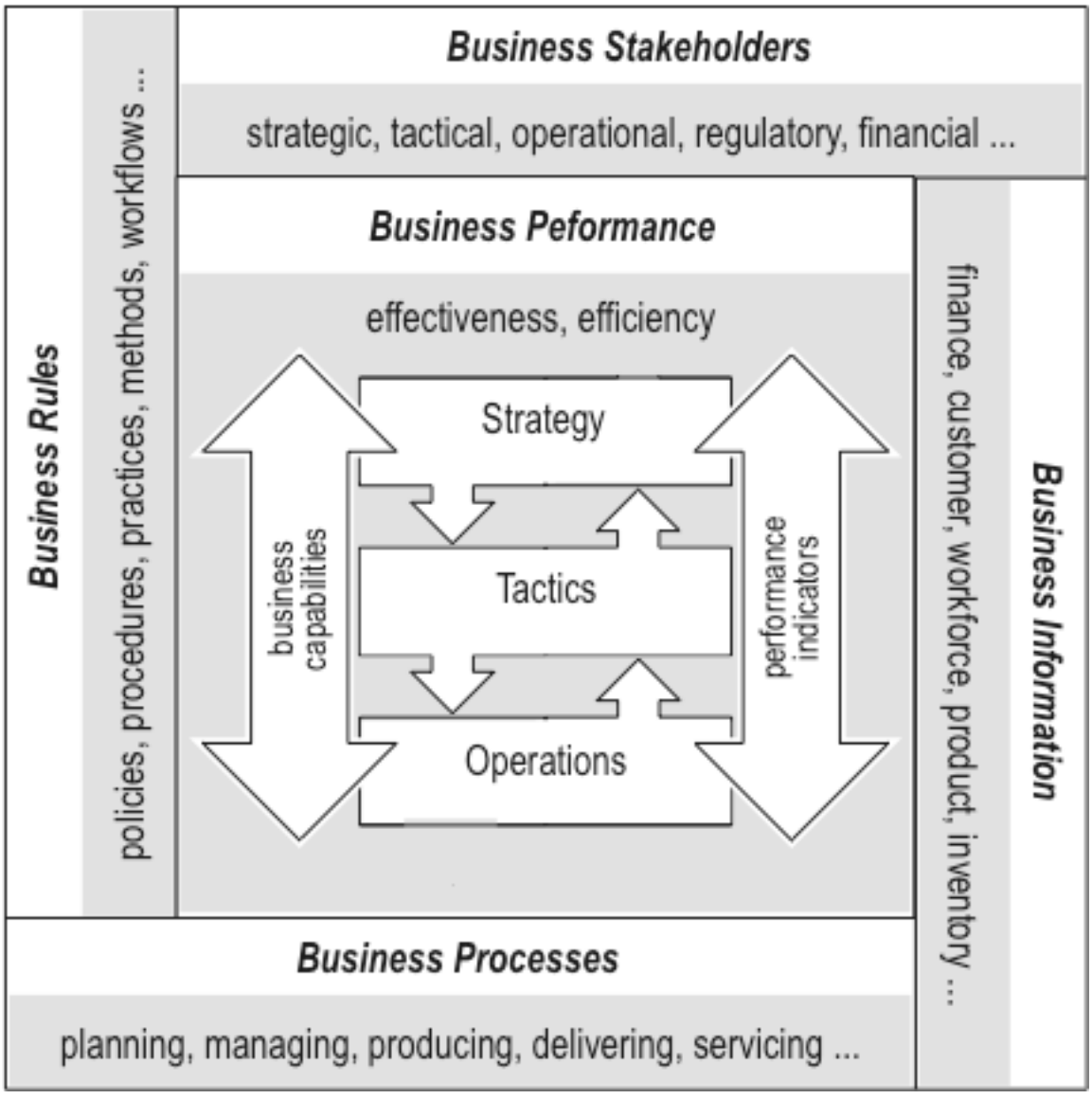
Technology architecture is needed to ensure that the technology stack will adapt to change and can be sustained throughout the lifespan of enterprise BI systems. Architectural considerations and responsibilities include effectiveness, efficiency, cohesion, service levels, and evolution.

Module 2

Business Architecture

Business Architecture Concepts

What and Why



Business Architecture Concepts

What and Why

THE TOP OF THE STACK

Business architecture is the essential starting point for BI 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. BI delivers value when it makes substantial contributions to business effectivness 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. Identifying and classifying stakeholders is an architectural construct. Multiple classifications such as by role (strategic, tactical, operational) and by interest (legal, regulatory, financial) is useful to fully understand the scope and relationships of stakeholders.

BUSINESS INFORMATION

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

BUSINESS PROCESSES

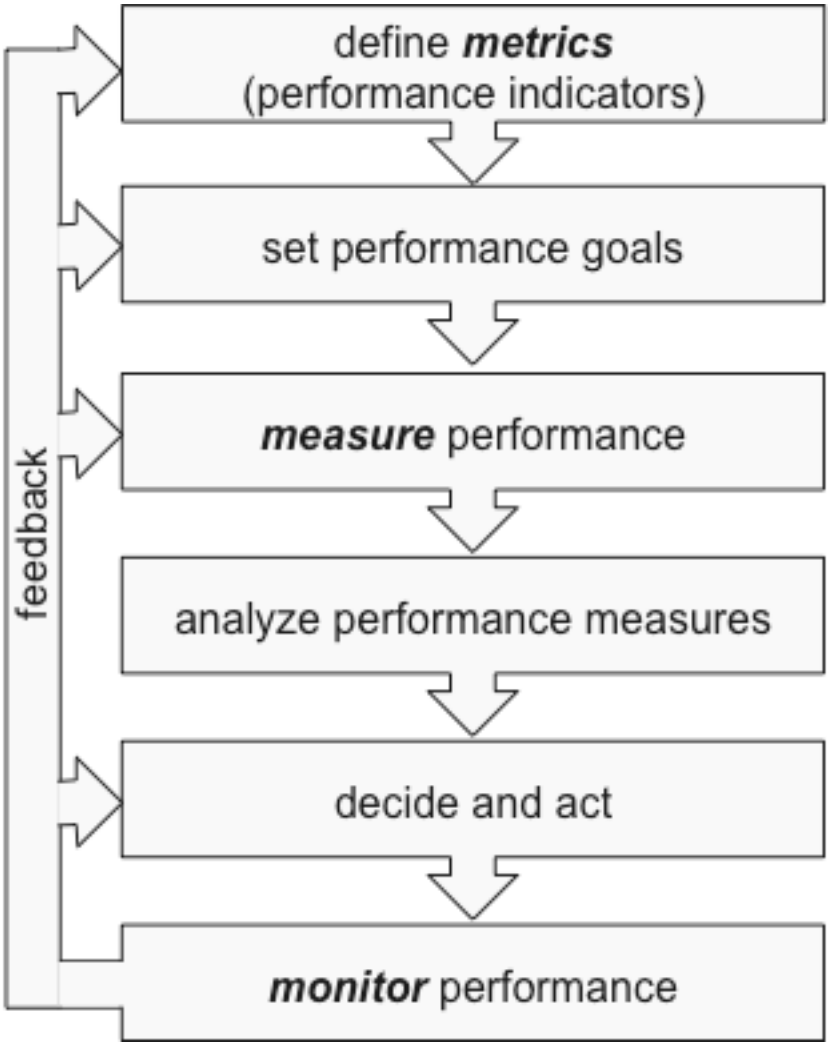
The processes of the business are among stakeholder interests and responsibilities, and are also sources for and consumers of information. No business architecture is complete without a process view.

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. Business architecture includes the structures and relationships to incorporate business rules into BI systems.

Business Performance

Performance Mangement



Business Performance

Performance Management

PERFORMANCE MANAGEMENT DEFINED

Gary Cokins defines performance management concisely as "the translation of plans into results – execution. It is the process of managing an organization's strategy."[1]

Wayne Eckerson says performance management "consists of a series of processes and applications designed to optimize the execution of business strategy… a framework that takes the long-standing task of *measuring performance* to the next level, that of *managing performance*."[2]

Frank Buytendijk states, "Performance management tries to capture an organization's business model. As it becomes clear how various business domains affect the business results, performance management provides insight into who drives results and how results are driven."[3]

While each of these definitions offers a slightly different perspective, they support some common themes about performance management:

- Performance management focuses on execution of strategy.
- Measurement is only a part of performance management.
- Analysis, especially cause-and-effect analysis – is a key part of performance management.
- The right actions and results are the goals of performance management.

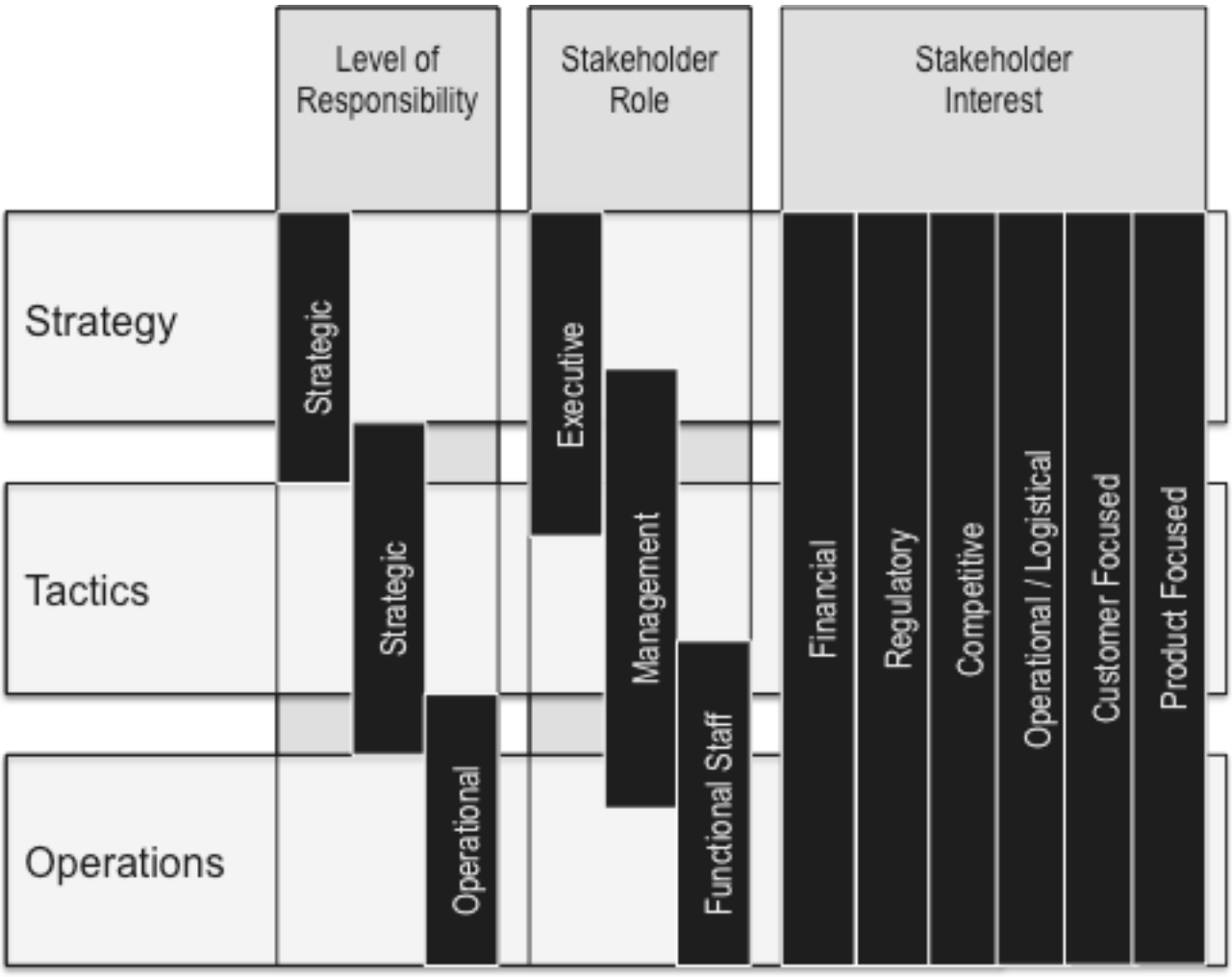
THREE Ms OF PERFORMANCE MANAGEMENT

Metrics, measures, and monitoring are the core elements that make performance management work. Metrics are the quantitative criteria that are used to manage performance. Measures are point-in-time supporting data for metrics. Monitoring is the act of tracking measures against metrics-based goals.

[1] *Performance Management,* pp. 9, Cokins.
[2] *Performance Dashboards,* pp. 11, Eckerson.
[3] *Performance Leadership,* pp. 17, Buytendijk.

Business Stakeholders

Responsibilities, Roles, and Interests



Business Stakeholders

Responsibilities, Roles, and Interests

INTERESTS AND EXPECTATIONS

Business Intelligence stakeholders are the people who use BI systems 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 stakeholders and to understand their interests and expectations.

Module 3

Organization Architecture

Organization Architecture Concepts

What and Why

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

Organization Architecture Concepts

What and Why

BI IN THE FABRIC OF BUSINESS

Organization architecture is primarily about people. It is the essential architectural perspective to connect people with BI and to weave BI 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.

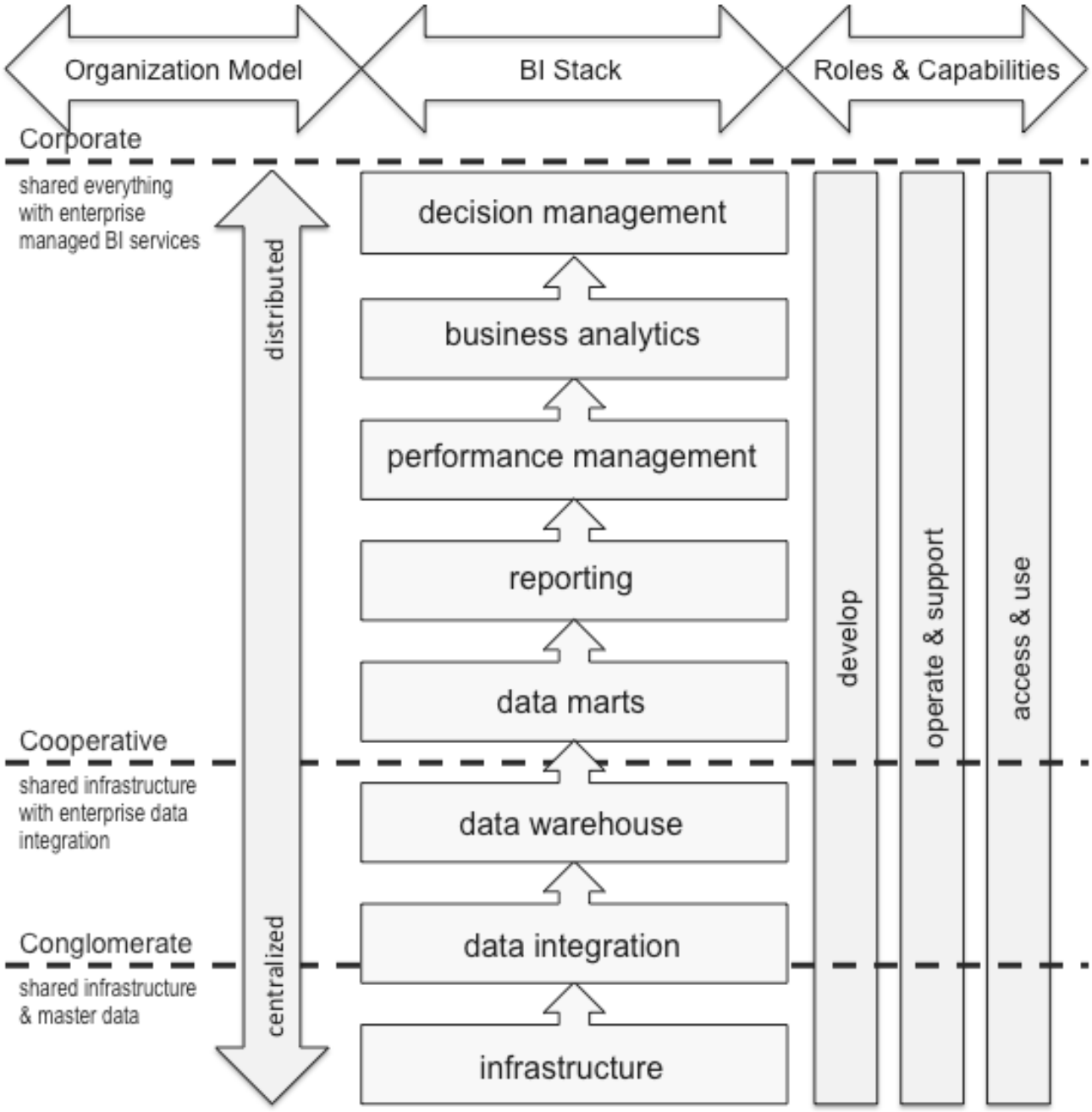
BI ORGANIZATION FRAMEWORK

Organization architecture is structured as four categories of components:

- People describes the "who" of BI organization, encompassing roles, skills, competencies, and capabilities.
- Purpose describes the "why" with components of goals, results, responsibilities, and accountabilities.
- Process describes the "how" with decision, execution, communication, and information components.
- Structure describes relationships and dependencies among people.

People

Roles and Capabilities



People

Roles and Capabilities

ORGANIZATIONAL VARIATIONS

One of the most important architectural views of people is found in mapping people to the BI stack, specifically considering the point where enterprise management and shared BI shift to end-user autonomy and business unit independence. Wayne Eckerson[1] describes three models:

- **Conglomerate** where infrastructure and metadata are shared and all of the upper layers of the stack are independently managed.
- **Cooperative** with shared infrastructure and data integration extending to autonomous access to and business applications of the data.
- **Corporate** where all layers of the stack are shared and centrally managed as enterprise resources.

ROLES AND CAPABILITIES

Another important view of people identifies which organizations or groups are responsible for development, operations, and usage activities at each level of the BI stack. Identifying the organizations and responsibilities enables guidelines and standards for the skills, capabilities, and competencies needed to fulfill those responsibilities.

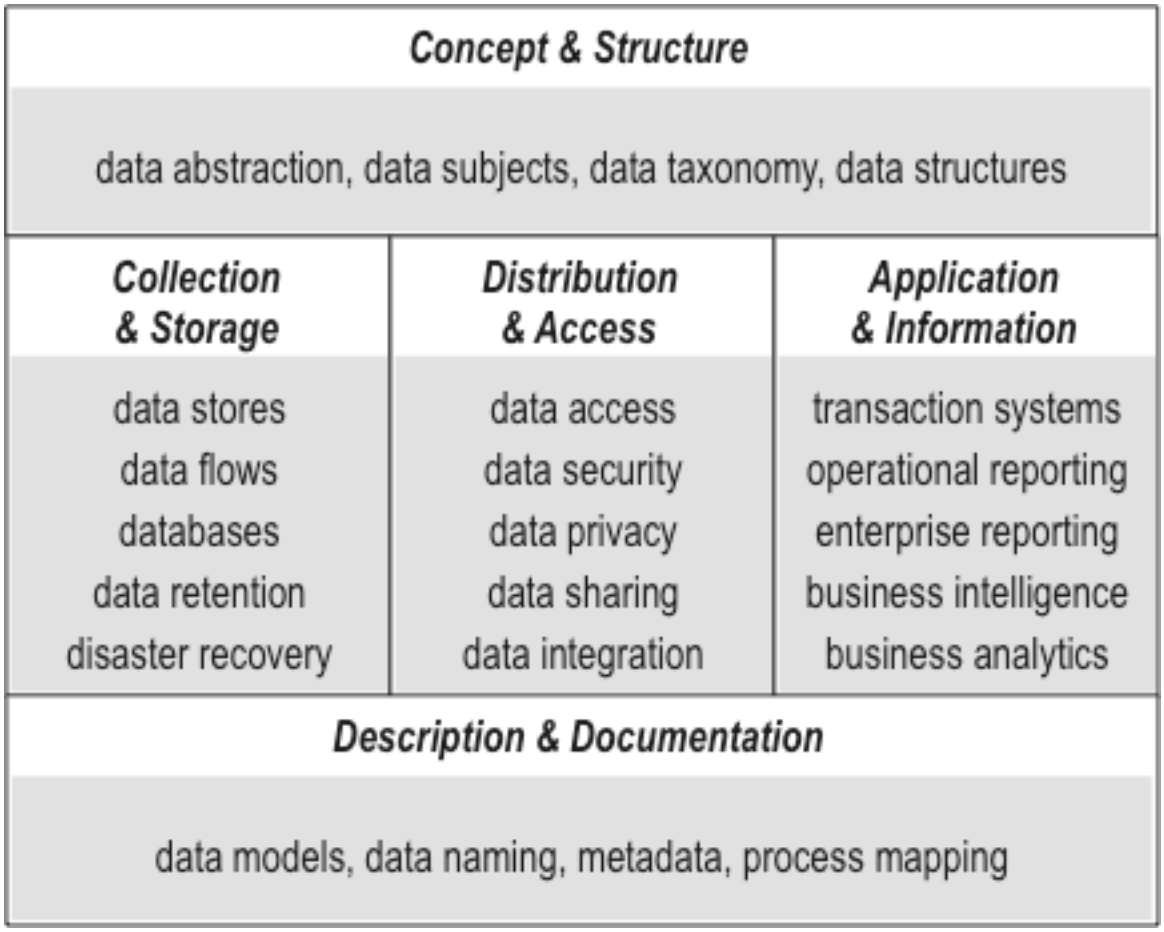
[1] *Business Intelligence Architecture: The Dividing Line* (http://www.b-eye-network.com/view/15523), Eckerson.

Module 4

Data Architecture

Data Architecture Concepts

What and Why



Data Architecture Concepts

What and Why

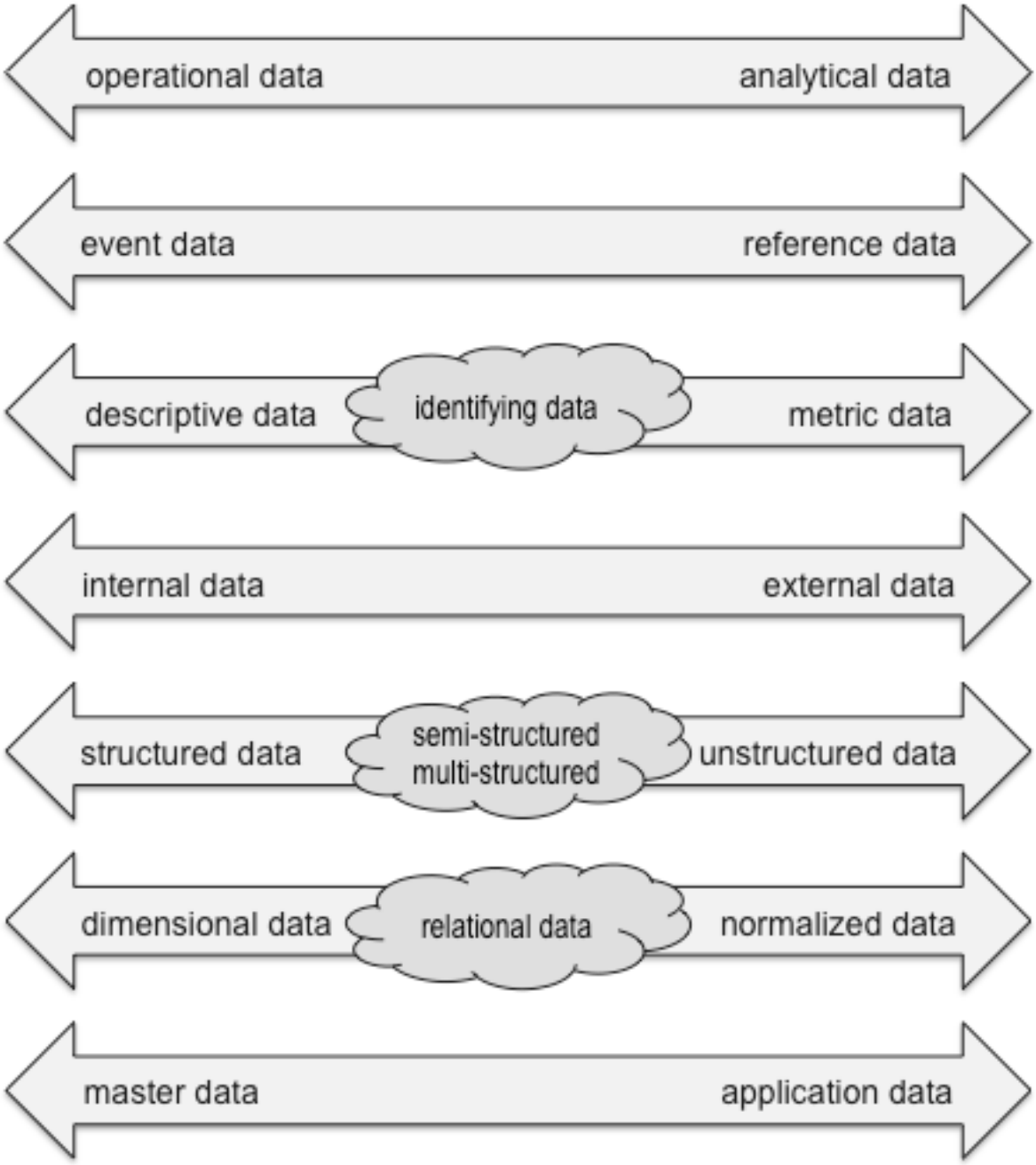
CRITICAL COMPONENTS

Data is fundamental to business intelligence. Without data none of the BI services and products are possible. Lacking good data it is impossible to deliver high-quality BI products and services. Thus data architecture is a critical part of BI architecture. Data architecture components are grouped as:

- Concept and structure components describing how data is organized and how it is viewed.
- Collection and storage components that support ability to obtain and retain data.
- Distribution and access components enabling connection of data with people who need data.
- Application and information components describing the variety of ways that data is used in business activities.
- Description and documentation components for sustainable understanding of data meaning, structure, and processing.

Concept and Structure

Data Taxonomy



Concept and Structure

Data Taxonomy

CLASSIFYING DATA

Event data is data whose values are determined by a business occurrence. For example, when a customer places an order the quantity is determined by the business event of placing a single specific order. Reference data provides the context of business events. Reference data values are not determined by the same business events for which they provide context.

Descriptive data records the non-quantitative or non-measurement properties of things. Metric data records quantifiable business facts that may be used as business measures. Identifying data is a subset of descriptive data that is used to distinguish between unique occurrences of business entities.

Internal data generally refers to data that is collected by the enterprise, managed by its systems, and stored within its databases. External data is not collected, managed, and owned by the enterprise. It is purchased or otherwise obtained from sources such as data syndication and subscription services.

Structured data can be described using the relational data model. Unstructured data does not have a defined data model or doesn't fit the relational model. Semi-structured data is a form of structured data where the expression of structure is embedded in the data (e.g., XML). Multi-structured data contains a variety of data formats and types in a single database or data store.

Dimensional and normalized data are both variations of relational data. Dimensional data is organized logically as a collection of related business measures associated with dimensions that are used to group and filter the measures. Normalized data follows a set of rules to remove dependency and redundancy and is organized logically as entities, relationships among the entities, and attributes of entities.

Master data is the non-transactional reference data of an enterprise – customers, products, parts, services, suppliers, accounts, etc. – that supports transactional activities and that is used and shared by many groups and processes throughout the organization. Application data is generally transactional, unshared or lightly shared through interfaces, and unique to the applications by which it is collected and used.

Module 5

Integration Architecture

Integration Architecture Concepts

What and Why

Data Consumers

People	Applications	Requirements
executive / strategic	business domain apps	real-time / right-time
management / tactical	enterprise reporting	point-in-time / time-series
front line / operational	ERP / CRM	inquire, inform, report
external – supply chain	business intellignece	investigate, analyze
external – partners	performance management	monitor, track, alert
external – customers	business analytics	explore, discover
		simulate, predict, forecast

Data Flow

Acquisiton	Functions	Data Stores	Data Delivery
connections	transformation	source databases	access
access methods	restructuring	data staging	publishing
frequencies	quality	data warehouses	services
	identity resolution	data marts	**Methods**
	hierarchy management	ODS	materialize
		MDM repositiories	virtualize

Metadata Flow

metadata repositories | creating metadata | consuming metadata | rationalization

Data Sources

Systems	Structures	Locations	Technologies
ERP / CRM	structured	internal	relational
legacy	unstructured	external	multidimensional
hosted	semi-structured	web	flat files
web	multi-structured	cloud	spreadsheets
social	geographic	subscription	web services
big data		syndicated	NoSQL / Hadoop

Integration Architecture Concepts

What and Why

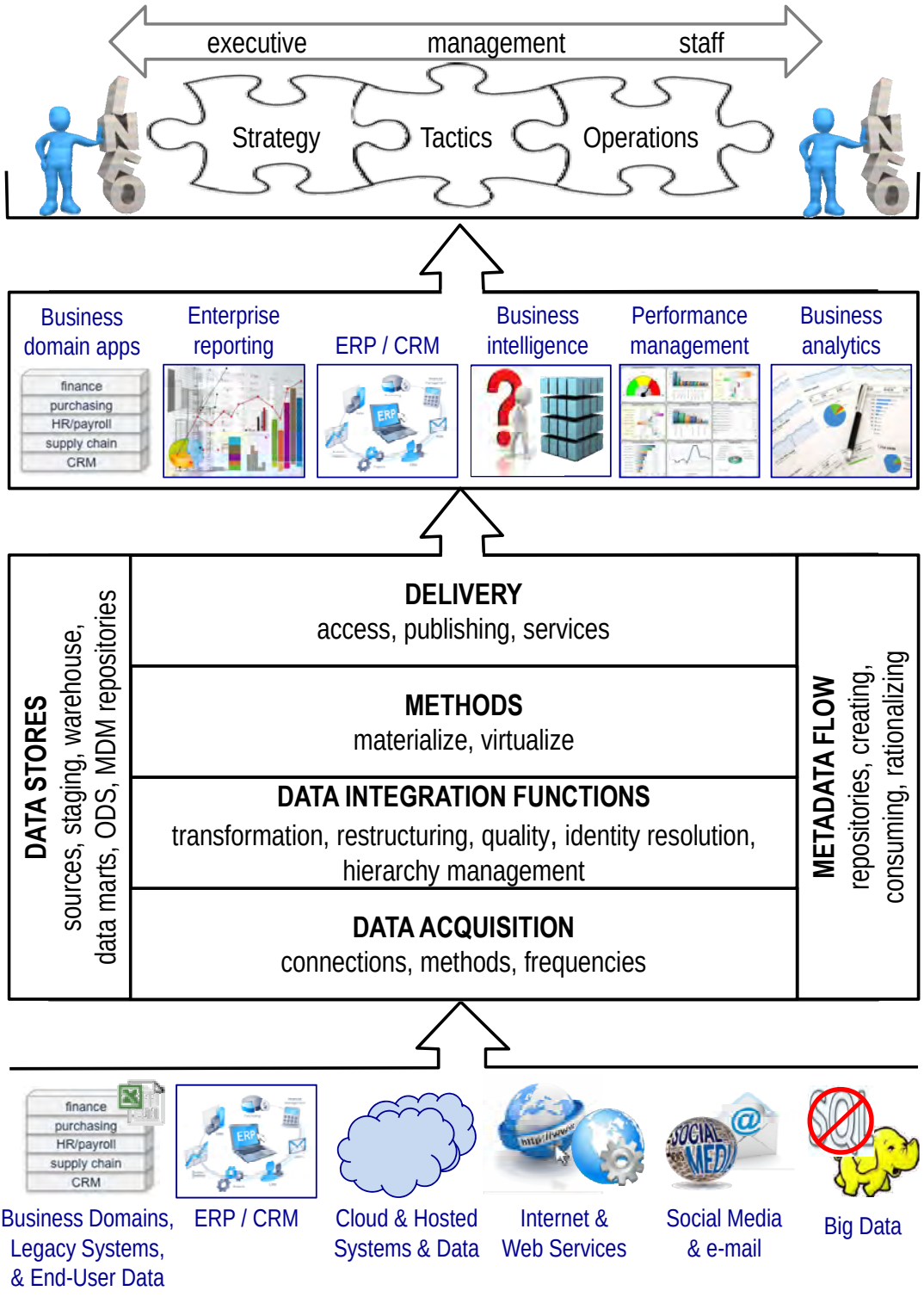
RATIONALIZING DISPARATE DATA

Data integration is a part of data distribution, with integration architecture a subset of data architecture. Yet integration is sufficiently complex with a large enough set of components that it merits positioning as a distinct layer of BI architecture. Integration architecture comprises:

- Data sources including systems, structures, locations, and technologies of source data
- Data flow including 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

Data Sources

Source Systems and Databases



Data Sources

Source Systems and Databases

SYSTEM PERSPECTIVE

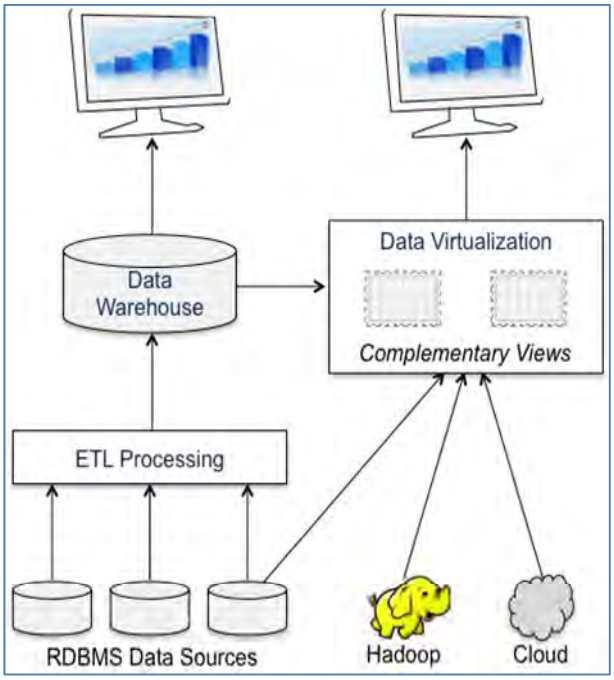
Data sources are clearly one of the critical components of data integration architecture. It is necessary to understand those sources from more than a simple database point of view. The architecture should include components for all of the various types of systems from which you acquire (or will acquire) data. 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
- E-mail – an increasingly 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

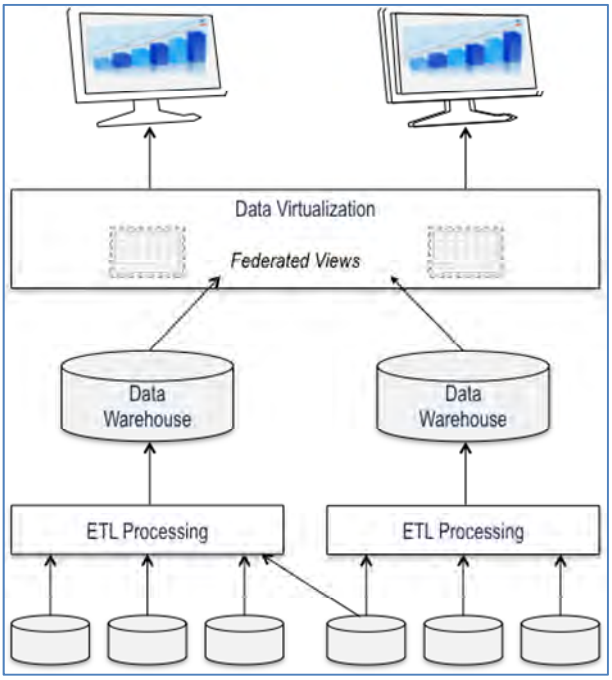
Data Warehouse Architecture

Extending the Data Warehouse with Virtualization

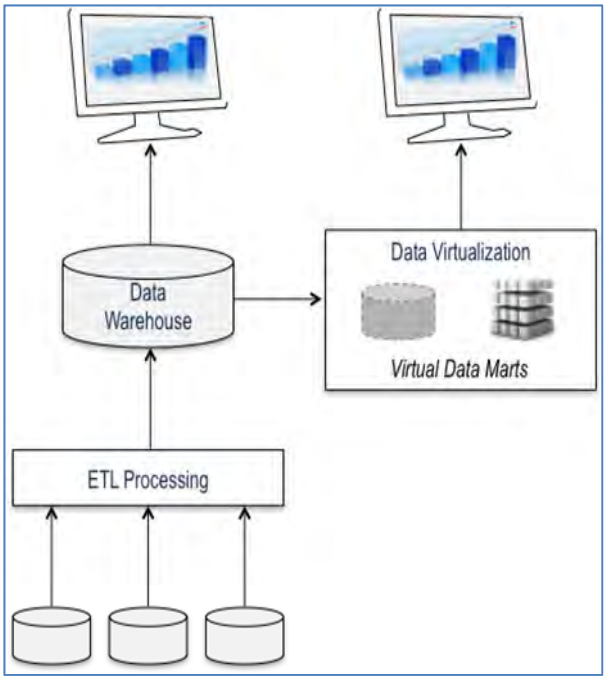
Augment



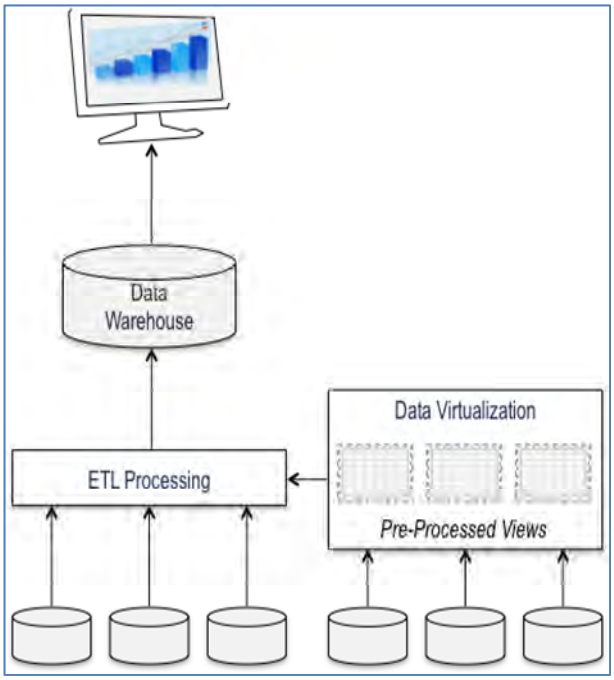
Federate



Hub and Virtual Spoke



Complement ETL



Data Warehouse Architecture

Extending the Data Warehouse with Virtualization

AUGMENTATION

Data virtualization can be applied to augment an existing data warehouse with virtual views to meet new information needs. Unstructured data, cloud data, and real-time data integration can be implemented without disruptive changes to the core data model and ETL processes. Speed of delivery and speed of data are accelerated with virtualization. Leveraging new and existing data sources more rapidly advances business agility. Unstructured data is integrated with structured data, and new reporting and analytic applications can be implemented more quickly.

FEDERATION

Many organizations have multiple data warehouses for a variety of reasons – mergers and acquisitions, independent departmental initiatives for data integration, purchased or hosted applications with data warehouse components, etc. Whatever the causes, the result is new silos of data without full enterprise integration. The challenge is to deliver an integrated enterprise view from many different data warehouses without disrupting the independent operations of each warehouse. With virtualization, each individual warehouse operates independently, serving the users and purposes for which it is designed while simultaneously participating in federated views that support an enterprise-wide perspective.

HUB AND VIRTUAL SPOKE

As demand for information accelerates, the demand for new data marts also grows. Data virtualization enables the creation of virtual data marts – "virtual spokes" that can be deployed quickly, without the increased development effort and production workload of more ETL processing. Virtualization enables the concept of disposable data marts and creation of new data marts easily – a particularly powerful technique in highly volatile business and systems environments.

COMPLEMENTING ETL

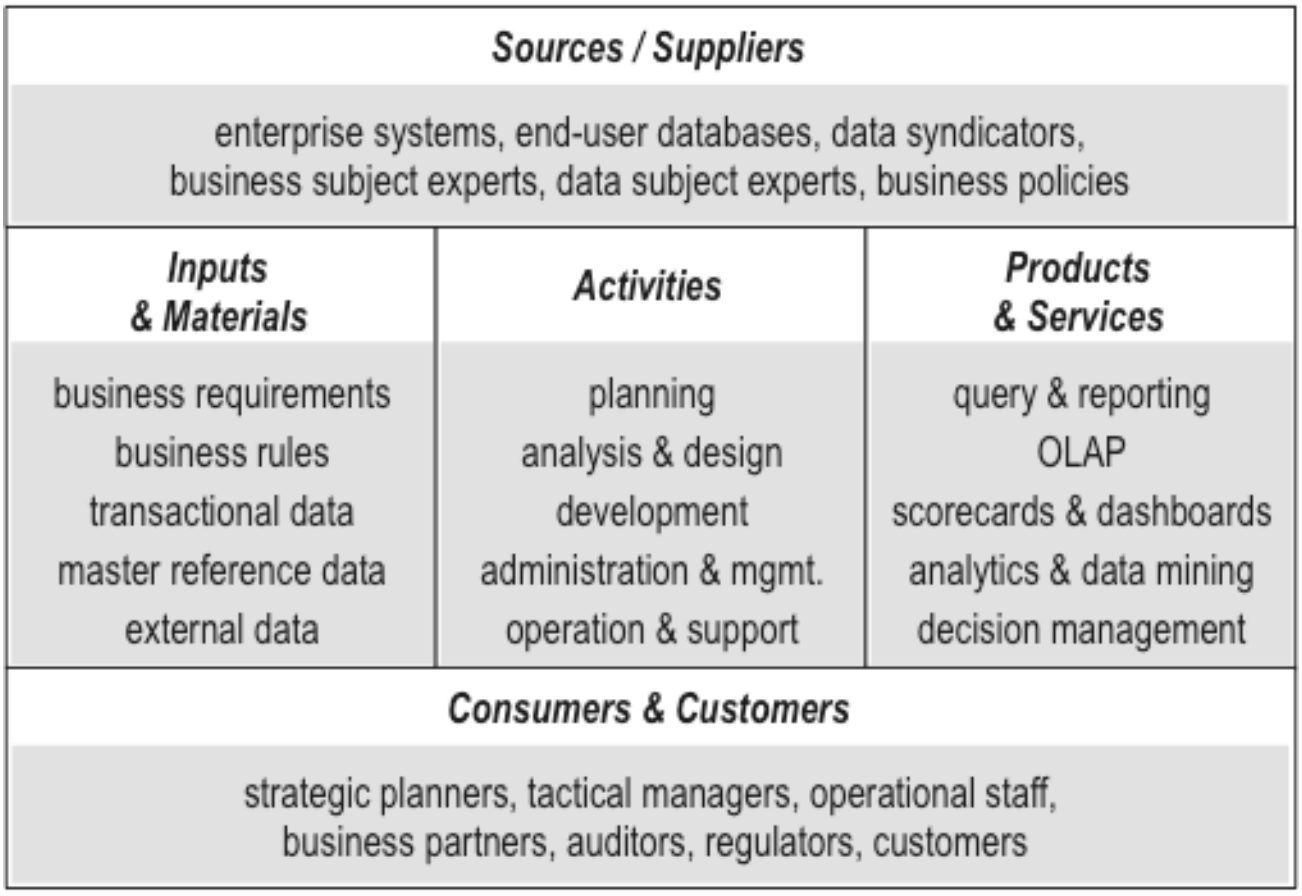
Older data warehouses acquire, process, and load data through ETL processing, using data sources that are typically structured relational data. With technology advances, many new data sources are compatible with traditional ETL processing. Common examples of data source and ETL incompatibility include ERP-embedded databases and web services data. Modifying existing ETL processes to access these sources is complex and risky. Data virtualization is an effective way to remove or reduce these incompatibilities. Using a virtualization tool, you can pre-process problem data sources, creating views that are readily accessible by your ETL technology. Changes to existing ETL processes and the risks inherent in those changes are substantially reduced.

Module 6

Process Architecture

Process Architecture Concepts

What and Why



Process Architecture Concepts

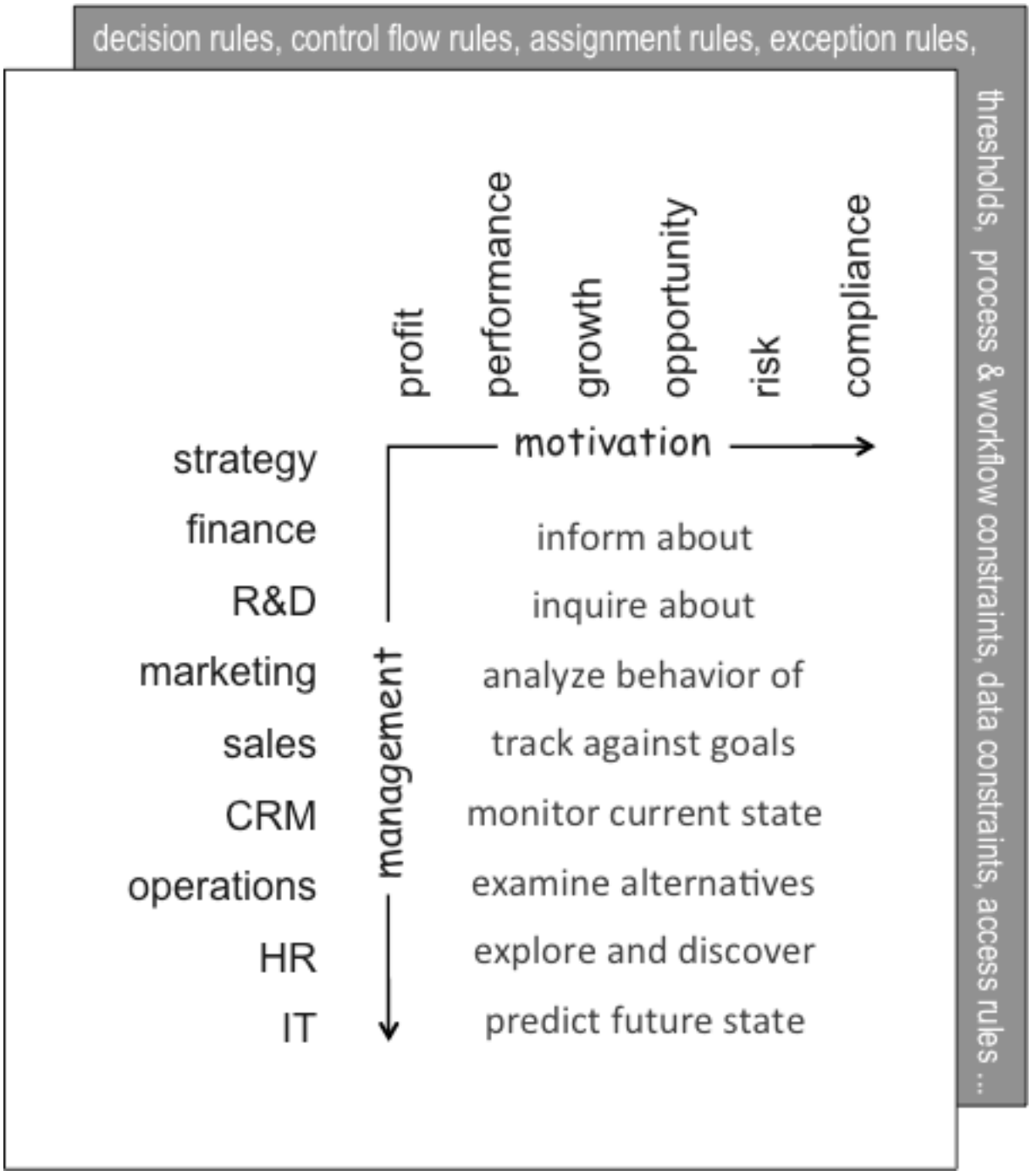
What and Why

APPLYING THE PROCESS MODEL TO BI

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

Inputs and Materials

Business Rules



Inputs and Materials

Business Rules

MANY KINDS OF BUSINESS RULES

In *Module 2: Business Architecture* we briefly discussed business rules as part of business architecture, looking primarily at rules that govern workflow for business processes – the control flow, assignment, and exception rules. Additional rule types that are architecturally important include:

- Decision rules expressing the logic and criteria by which routine and recurring decisions are classified (informed, assisted, advised, automated) and the actions that are taken with each decision type.
- Thresholds describing not-to-exceed conditions that are the basis of automated monitoring and alerts.
- Process and workflow constraints such as the rules that constrain a hiring process based on collective bargaining labor agreements.
- Data access and privacy rules such as patient data constraints imposed by HIPAA in the healthcare industry.

FINDING AND ORGANZING BUSINESS RULES

The variety, abundance, and complexity of business rules make it impractical to attempt to identify and catalog all of them. A pragmatic architectural approach provides the structure and guidelines to find, classify, and catalog rules on a project-by-project basis and as part of the requirements gathering process. Seeking rules at the intersection of management and motivation is a practical process. Consider, for example, the management domain of human resources (HR).

- At the intersection of HR and compliance, seek the rules of compliance with the Occupational Safety and Health Act (OSHA), with Equal Employment Opportunity (EEO) regulations, with labor union contract stipulations, etc.
- At the intersection of HR and performance, seek the policies that establish rules for performance review criteria, bonus programs, etc.
- At the intersection of HR and profit, look at the policies, contracts, and legal stipulations that govern pay rates and cost of employee benefits.

The same technique can be applied for any combination of management and motive. It provides a direct and focused way to find the business rules that matter at the time that they matter.

