TDWI Business Intelligence and Analytics Foundations

Charting the Course to Data-Driven Success

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.

TABLE OF CONTENTS

COURSE OBJECTIVES

You will learn:

- ✓ ***Meaningful and actionable definitions of business intelligence and analytics***
- ✓ ***Roadmap development techniques for your BI and analytics program and projects***
- ✓ ***A capability framework that links BI and analytics program initiatives to business goals***
- ✓ ***Best practices for BI and analytics services, including performance management, analytics, OLAP, reporting, visualization, and self-service***
- ✓ ***Types of data consumed by BI and analytics applications, including enterprise, external, and big data sources***
- ✓ ***Architecture, implementation, and operational practices for data integration services***
- ✓ ***Data management practices including data governance, data quality management, data profiling, and data cleansing***
- ✓ ***Technologies needed to support BI and analytics***

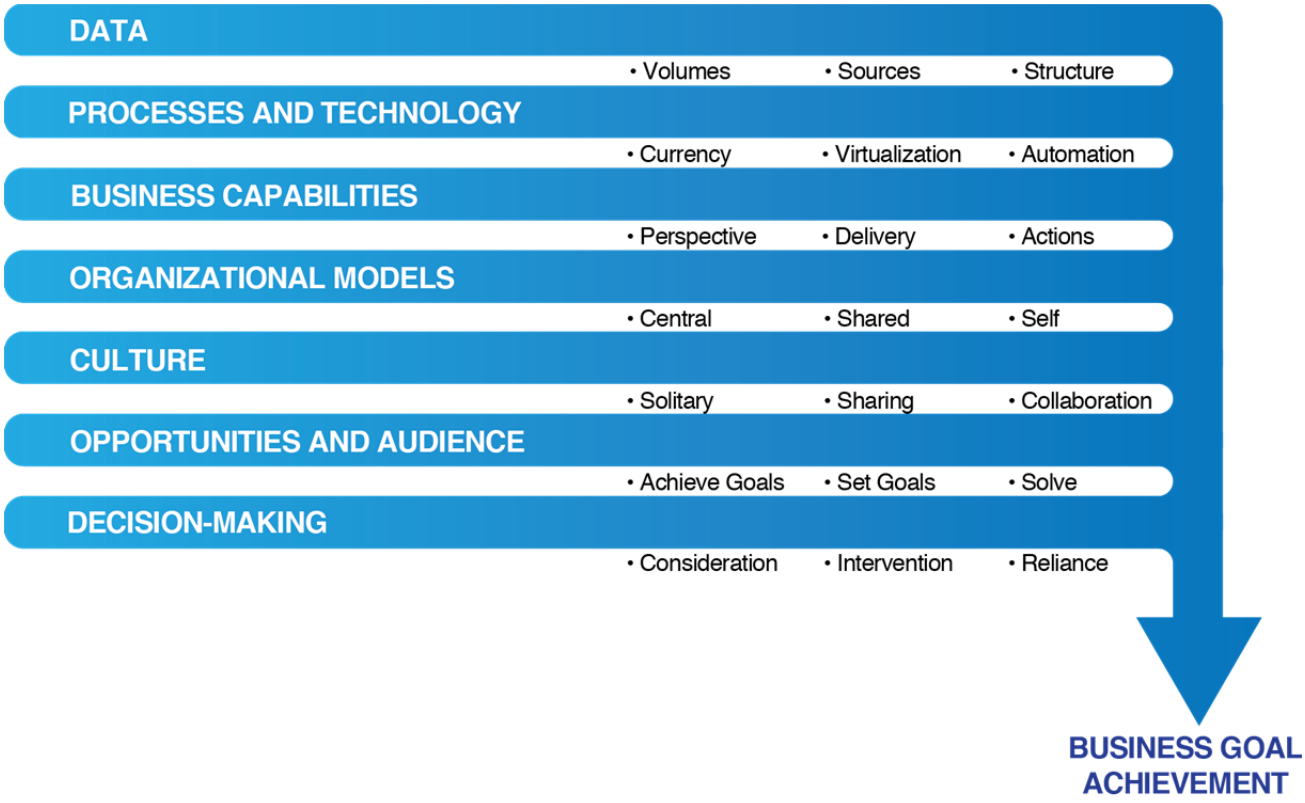
Transforming Data With Intelligence™

Module 1

Introduction to BI and Analytics

Definitions

Evolution of BI and Analytics



Definitions

Evolution of BI and Analytics

CHANGING LANDSCAPE

The graphic on the facing page depicts key elements of the definitions of BI and business analytics. In the past decade, there have been major changes in each of these areas.

DATA

Data volumes have increased significantly over time, from data warehouses measured in megabytes and gigabytes to ones that are measured in terabytes and petabytes. Exabyte-sized environments are in the foreseeable future.

Initially, data was obtained from internal operational systems or external sources and was highly structured in nature. As the ecosystem evolved, additional sources (e.g., machine-generated data, social media, and Internet applications) have been incorporated, encompassing other data structures that may be stable or in motion. These are described further in *Module 4: Data Integration*.

PROCESSES AND TECHNOLOGY

As demands have increased for more data, near real-time access, and additional functionality, supporting processes and technology have evolved.

Initial environments depended on manual coding of the data movement algorithms. ETL tools were then introduced. These have evolved into more comprehensive data integration and aggregation tools that employ other technologies, including data virtualization. Tools to automate major parts of the development process have also emerged. Tools and technology are described further in *Module 6: BI and Analytics Technology*.

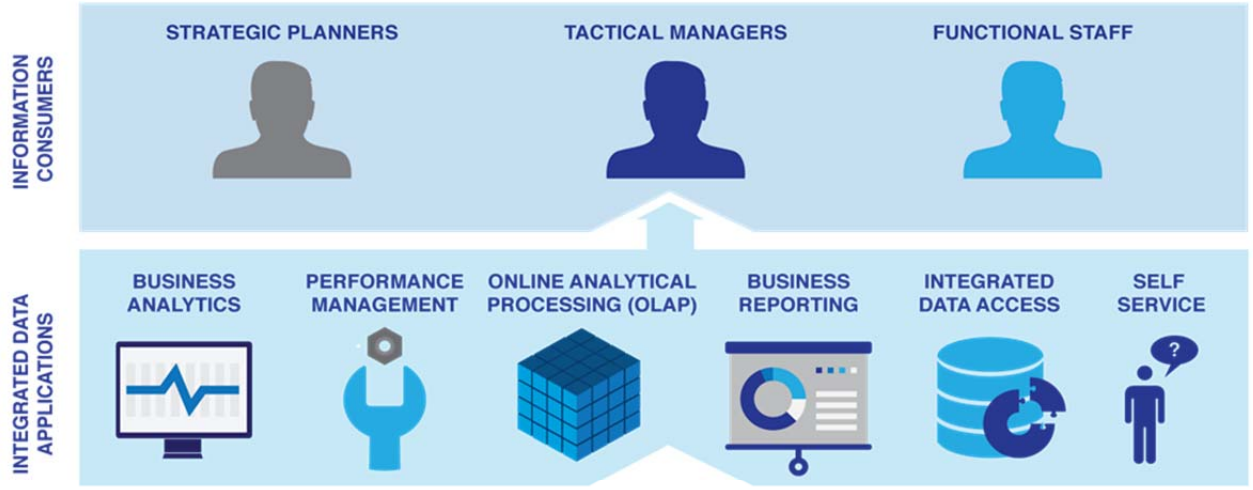
A word of caution is needed. There is no silver bullet. As you consider emerging technologies, be sure to carefully evaluate the offerings and your organization's readiness to adopt them and to deal with potentially being on the leading edge.

BUSINESS CAPABILITIES

Early business intelligence solutions delivered information mostly in the form of queries and reports with some simple graphics. Today, the landscape has evolved to include dashboards and scorecards, analytic models and simulations, and data visualizations. These topics are discussed in *Module 2: Performance Management and Analytics, and Module 3: OLAP and Other Information Services.*

Components

People and Applications



Mistakes to Avoid

When Validating Direction

TEN MISTAKES TO AVOID

Ten Mistakes to Avoid When Validating Your Business Intelligence or Data Warehousing Direction by Jonathan Geiger.

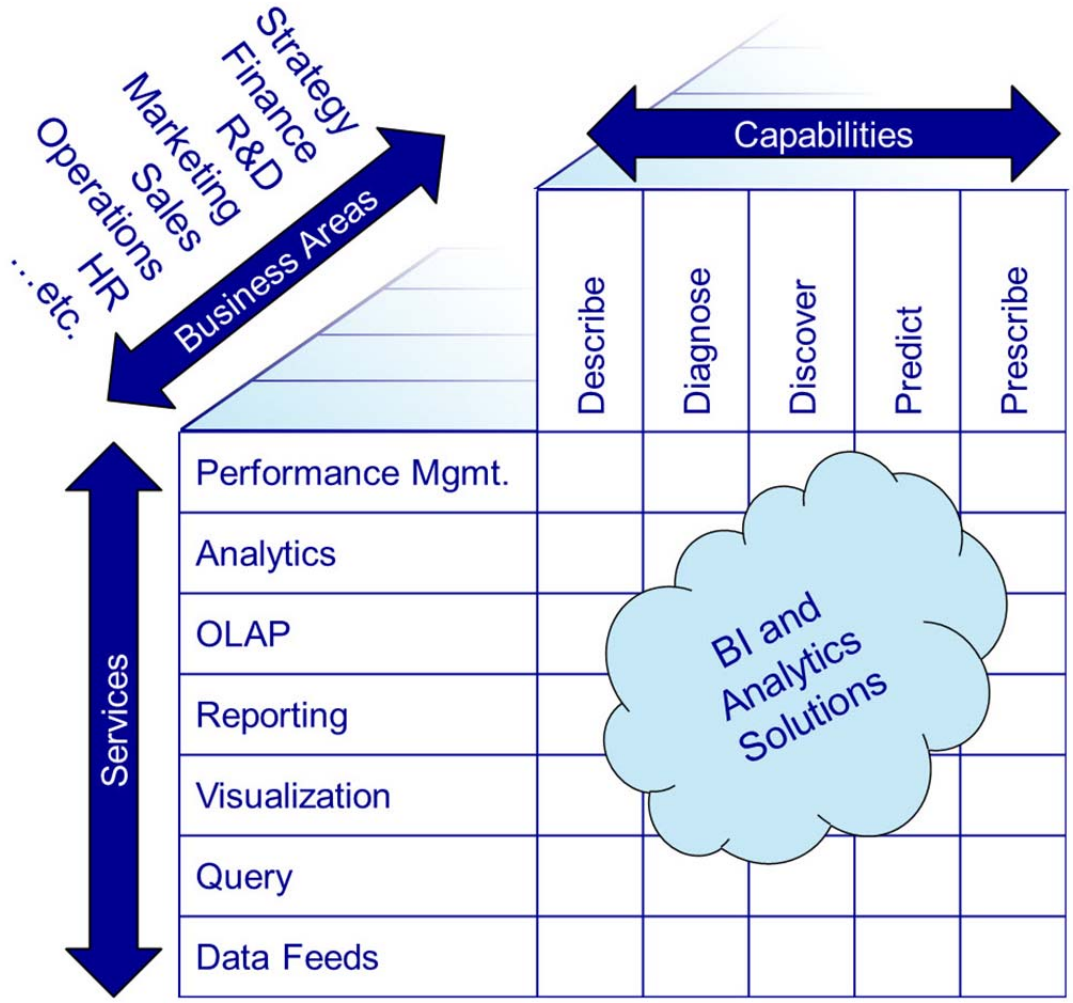
It is often said that "If it ain't broke, don't fix it," but how do you know if your program isn't broken? Even if it meets some business needs and provides business users with data to support analyses and decision making, there are always opportunities for improvement. Companies should periodically review where they've been with their program and more important, where they're heading. However, exploring your direction without a well-thought-out approach and ignoring critical areas may lead to inaccurate conclusions about the environment and suboptimal improvements. Mistakes to avoid are shown on the facing page.

Module 2

Business Metrics and Analytics

Business Capabilities

Capabilities Through Services



Business Capabilities

Capabilities Through Services

DELIVERING VALUE

There is not a one-to-one mapping between a business capability and a business intelligence solution.

For example, OLAP services may provide descriptive or diagnostic capabilities; performance management services may provide descriptive or prescriptive capabilities.

SERVICES FOCUS

As the illustration suggests, business applications are solutions built for specific business functions. They provide capabilities through one or more services.

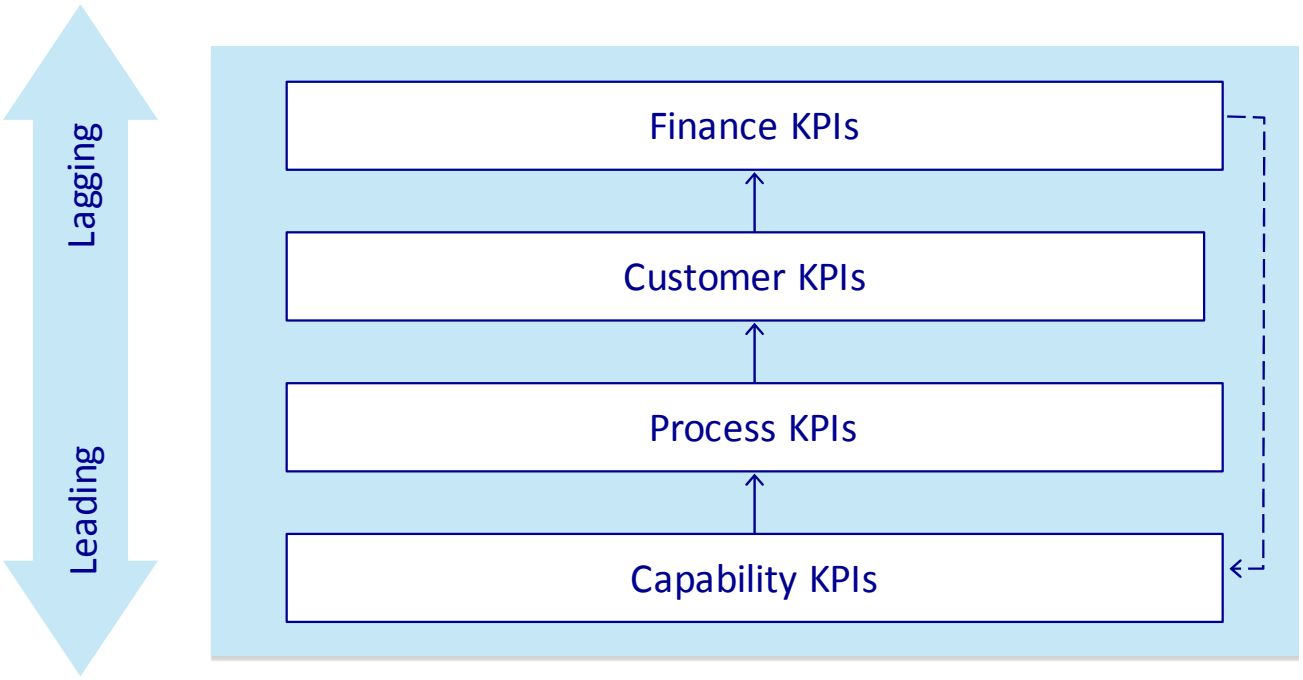
Services include:

- *Performance management services* identify and track key performance indicators and associated goals, surfacing them through dashboards and scorecards.
- *Analytics services* leverage analysis, statistics, and data mining to generate insights that are deeper than what can be revealed by surface-level inspection.
- *OLAP services* provide the ability to view historical data from various perspectives and levels of aggregation. Often referred to as "slice and dice analysis."
- *Reporting services* provide information needed on a regular and predictable basis at strategic, tactical, and operational levels of the business.
- *Visualization and storytelling services* facilitate communication of BI and analytics insights, helping channel them into actions and business impact.
- *Query services* allow businesspeople to begin their own inquiries into data resources through ad hoc and managed access environments.
- *Self-service BI and analytics services* enable businesspeople to bring their own data resources into an analytics environment.

The remainder of this module will explore the first two services in this list. The rest are explored in *Module 3: OLAP and Other Information Services.*

Performance Management

Key Performance Indicators



Performance Management

Key Performance Indicators

KPI DEFINED

Key performance indicators are "financial and non-financial metrics used to help an organization define and measure progress toward organizational goals." [1] This definition contains several key concepts:

- KPIs include both financial and non-financial metrics.
- KPIs help to define organizational goals.
- KPIs help to measure progress toward goals.

A metric refers to a measurement of business activity. However, in a performance management system, we want to do more than just measure business activity; we want to measure performance aligned with business strategy.

THE BALANCED SCORECARD

KPIs are the cornerstone of the Balanced Scorecard (BSC) introduced by Robert Kaplan and David Norton (Harvard Business School) in 1992. Since that time it has evolved substantially and become the de facto standard for strategic business scorecards.

The BSC approach is founded on these principles:

- *Financial performance* metrics are lagging indicators
- *Customer outcomes* drive financial performance
- *Internal process excellence* drives customer performance
- *Organizational capability* (learning and growth) drives process performance

These four categories can be tailored to a business as a whole, to a function or department, or even to an individual employee.

In each category, a set of metrics are selected that align with business goals and correspond with the activities and focus of the target audience. In combination with goals, these metrics can be monitored and presented using dashboards and scorecards, enabling businesspeople to take actions to have positive business impact.

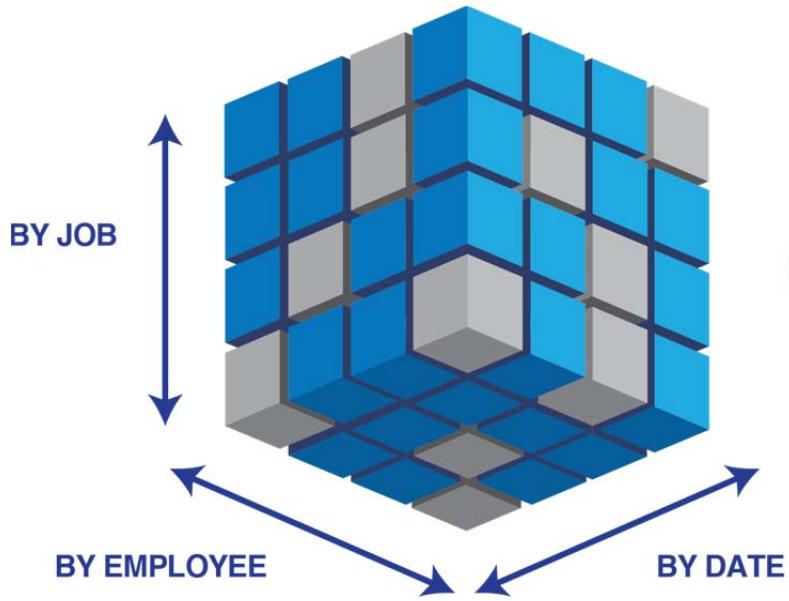
[1] *Business Dashboards: A Visual Catalog for Design and Deployment*, pp. 24, Rasmussen, Chen, and Bansal

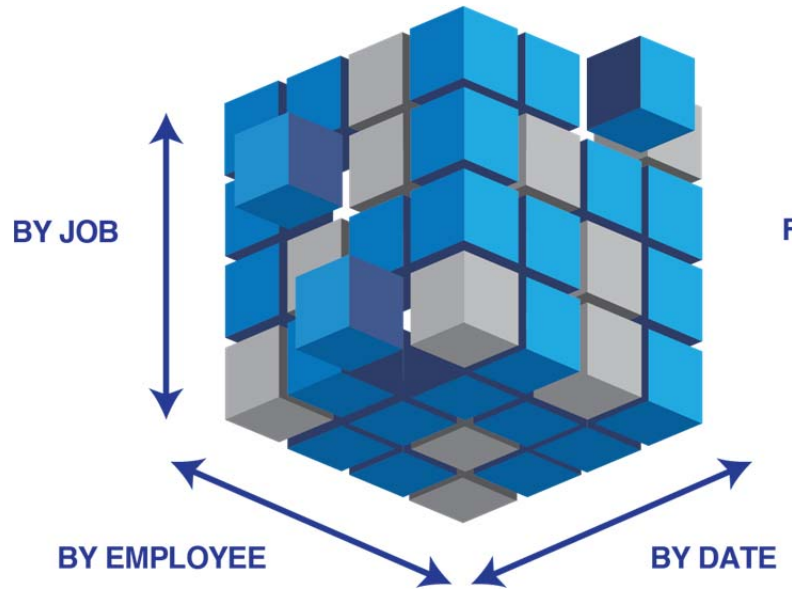
Module 3

OLAP and Other Information Services

OLAP Services

Online Analytical Processing





OLAP Services

Online Analytical Processing

OLAP DEFINED

OLAP is a data access and manipulation technology that allows users to selectively access data and view it from different perspectives. OLAP tools are used to analyze different dimensions of multidimensional data. The data typically represents business measures that can be aggregated along dimension hierarchies, selected and limited by elements of dimensions, and viewed at the intersections of multiple dimensions.

OLAP is among the most common business analysis tools currently in use. OLAP is powerful; the tools are relatively easy to learn and use, and they are a widely accepted and integral part of business analysis.

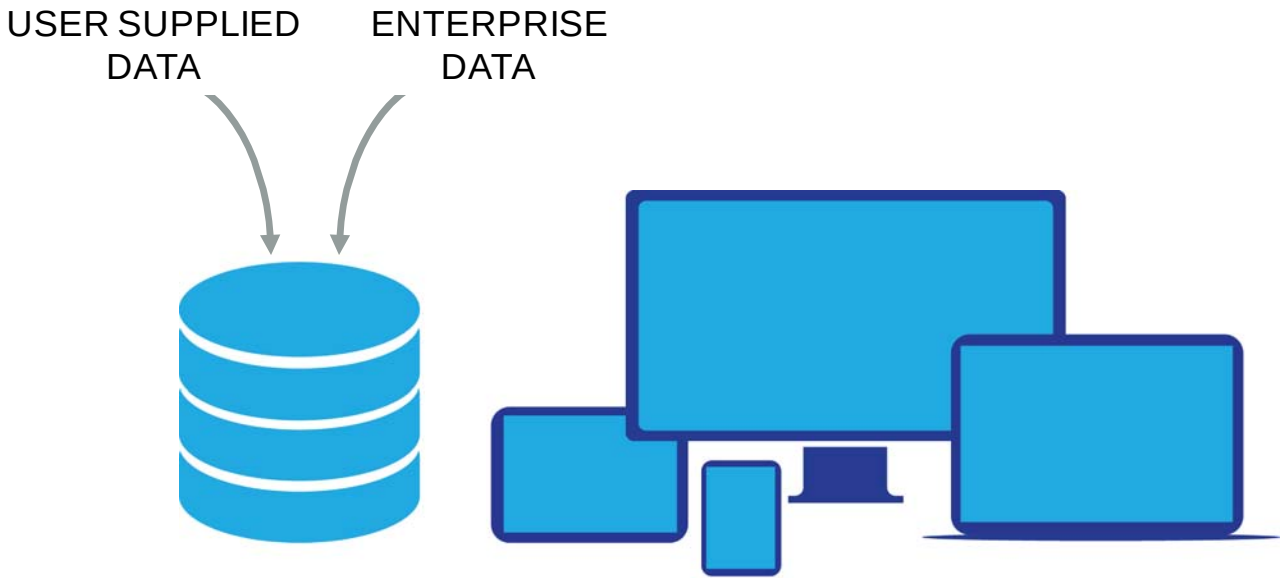
THE OLAP INTERFACE

Selecting which dimensions are in view and on which axis is achieved using a pivot-table-like interface. The OLAP cube (or star schema) with this interface supports several operations:

- *Slice*—Slicing selects a subset of the data by limiting data to a single value for one dimension. Slicing produces a two-dimensional array.
- *Dice*—Dicing selects a subset of the data by limiting the values that are active in two or more dimensions. Dicing produces a three-dimensional array that is in effect a smaller cube.
- *Drill Down*—Drilling down moves from summary to greater detail, resulting in finer grain for the data that is in view. Drill is an important analysis function as most analysis begins by looking at summary data then drilling to detail where questions arise.
- *Roll Up*—Rolling up summarizes data to show totals at varying levels along a hierarchical dimension.
- *Pivot*—Pivot rotates a cube in space to change which dimensions are positioned on the x, y, and z axes of a three-dimensional view.

Self-Service

Evolving Service Models



- ✓ DATA IMPORT
- ✓ PROFILING
- ✓ DATA BLENDING
- ✓ EXPLORATION

Self-Service

Evolving Service Models

BRING YOUR OWN DATA

Self-service is one of the latest trends in BI and analytics. A self-service environment allows someone to import their own data sets into the data management environment and then perform their own data exploration activities. These may include:

- Data *import (*or *ingestion)*, often supported by a *data lake* (to be discussed in Module 4)
- Data *profiling* (to be discussed in Module 5)
- Data *blending* (or *integration*) with other enterprise data (to be discussed in Module 4)
- Data exploration
- Visualization and reporting

OTHER SERVICE MODELS

Self-service is one extreme of a spectrum of service models.

Central services are the "we build it for you" model that works well for standard reports and routinely published information. In the central services model, standards, processes, and technology are prescribed. A single centralized team is responsible for development, deployment, and management of information services. This model works well when goals are exceptional consistency, strong governance, rapid delivery, and managed costs. The central services model may be challenged to scale up to meet high demand for services.

Shared services is the "we build the Legos" model where a central team builds and publishes reusable data components that are accessed, configured, and assembled by distributed teams to meet their local needs. With published interfaces it is practical for local data to be appended to or integrated with central data. The shared services model defines processes, standardizes architecture, and maintains a centralized team for shared work, but much project and process work occurs in individual project teams and distributed business units. The blend of centralized and decentralized resources achieves good efficiency of resource utilization.

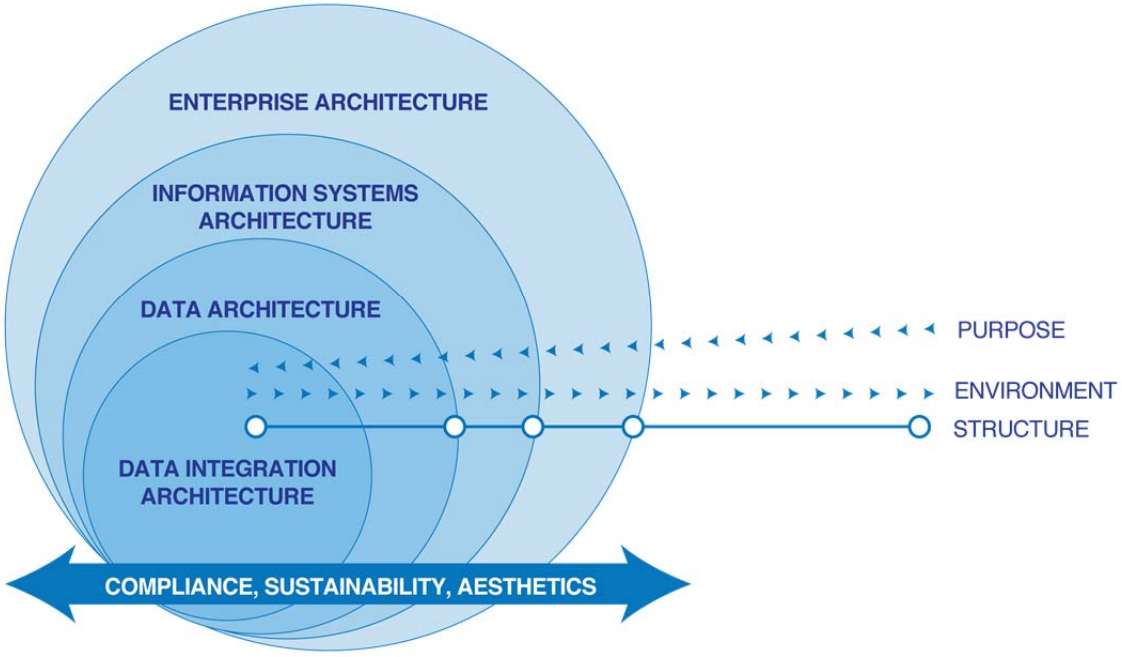
Hybrid services. As a practical matter, many organizations evolve to a mix-and-match hybrid of service models. Good guidelines and clear understanding of the criteria by which projects and service models are matched is important to ensure appropriate use of each level.

Module 4

Data Integration

Data Integration Architecture

Integration and Data



Data Integration Architecture

Integration and Data

DATA ARCHITECTURE

Data architecture defines the roles, structure, relationships, and rules to manage the data assets of an enterprise. Data architecture is a subset of information systems architecture, which is in turn a subset of enterprise architecture.

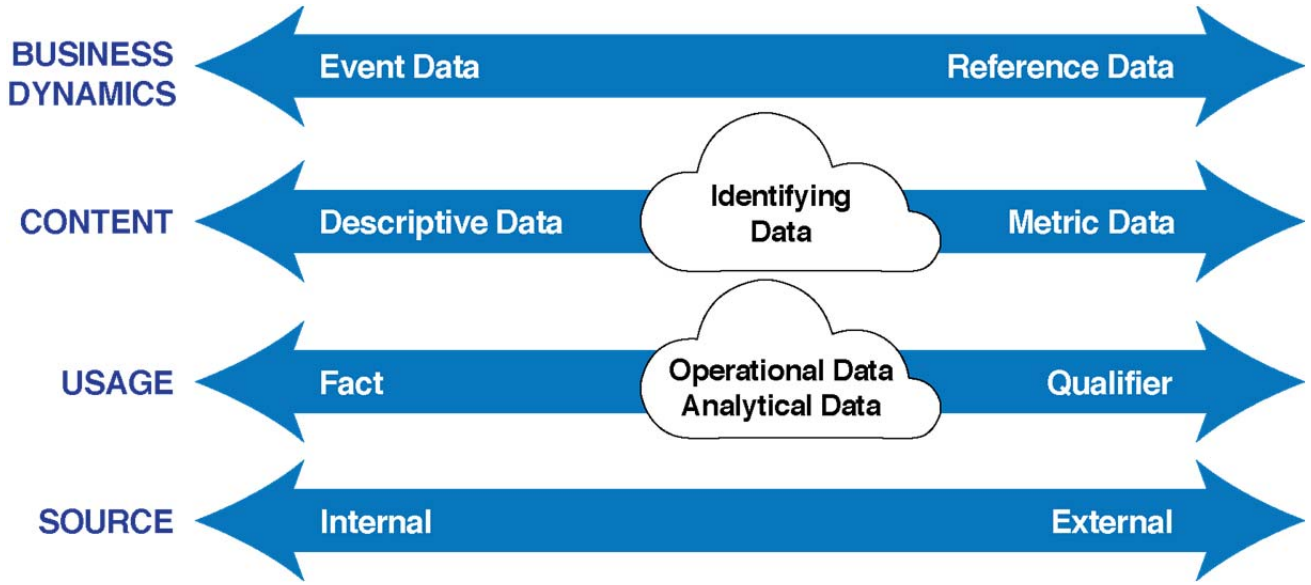
DATA INTEGRATION ARCHITECTURE

Data integration architecture is a subset of data architecture. The three supersets in which it is contained partially define the purpose of data integration and comprise the data integration environment. Architectural structure may be independent but compatible at each level. Compliance, sustainability, and aesthetics are consistent themes across all levels.

Data integration architecture defines the roles, structure, relationships, and rules to aggregate a collection of data integration components into a data integration system.

Data Types and Sources

Data Properties



Data Types and Sources

Data Properties

PROPERTIES

Business intelligence and analytics work with many different types of data, and it is important to understand the types. A single data item generally has characteristics in more than one category. Not all categories are comprised of mutually exclusive data types. One set of data types can be viewed based on the data properties.

BUSINESS DYNAMICS

Business dynamics refers to how the data contributes to recording business activities. *Event Data* is data whose values are determined by a business occurrence. *Reference Data* is data whose values provide the context for business events—these values are not determined by the events for which they provide context.

CONTENT

Content classifies the nature of domain values of a data item. *Descriptive Data* is data that records the non-quantitative or non-measurement properties of things. *Identifying Data* is a subset of descriptive data that distinguishes among unique occurrences of an item. *Metric Data* is data that records the quantifiable facts that may be used as business measures.

BUSINESS USAGE

Business usage indicates how data is applied in specific business scenarios. A *fact* is a discrete item of business information of interest to a business person seeking information. A *qualifier* is a criterion by which information is accessed, sorted, grouped, summarized, and presented by a business person seeking information. *Operational Data* is data used in day-to-day business activities. *Analytical Data* is data used to analyze and understand what is happening and why it is happening.

SOURCE

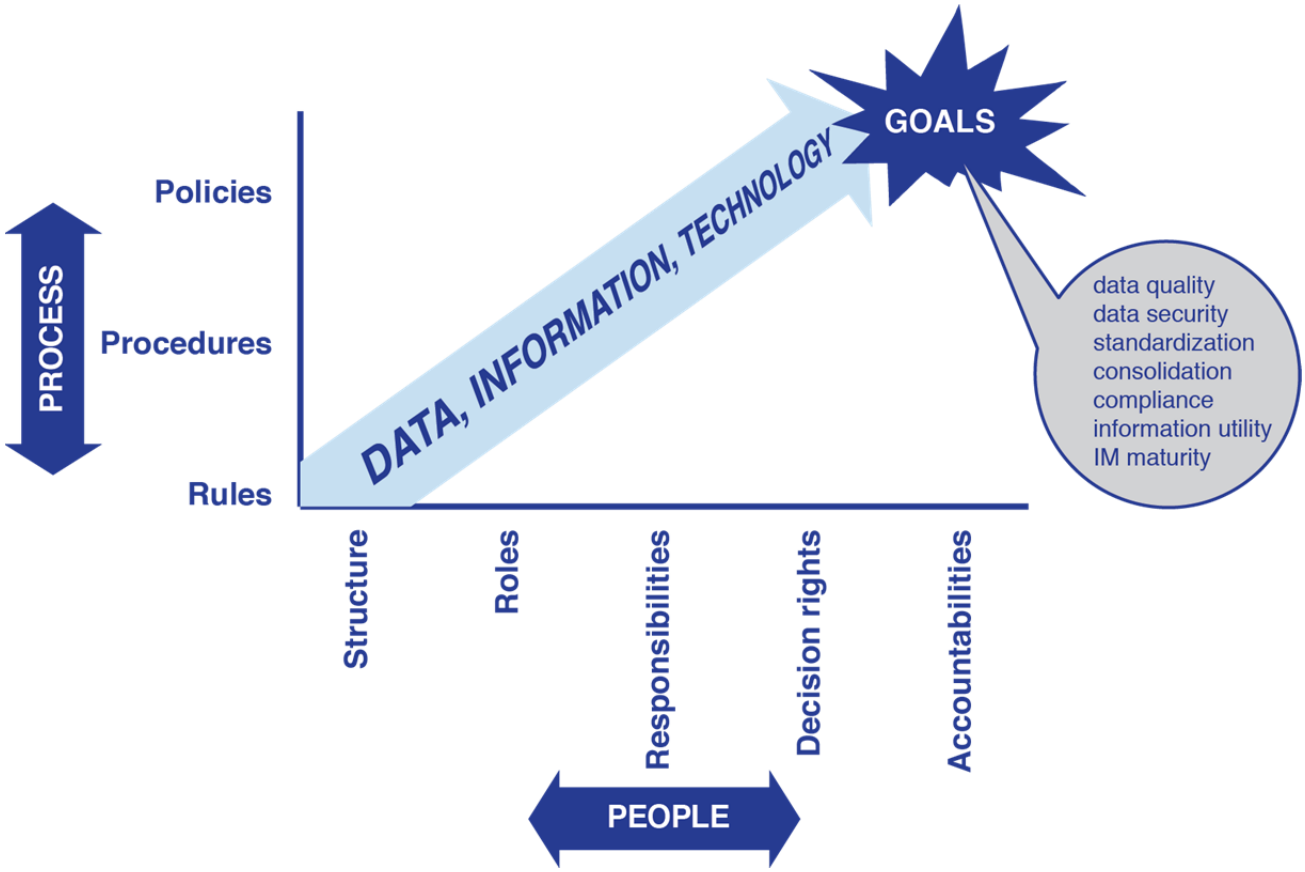
Source indicates the origin of the data. *Internal Data* is data that is collected by the enterprise, managed by its systems, and stored within its databases, regardless of whether or not it is stored on site. It includes data in transactional databases as well as data in decision support databases, intranet content, email repositories, spreadsheets, etc. *External Data* is data that is not collected, managed, or owned by the enterprise and is acquired from sources such as data syndication and subscription services, including postal service databases, social media feeds, etc., to enrich internal data.

Module 5

Data Management

Data Governance

Data Governance Concepts



Data Governance

Data Governance Concepts

GOVERNANCE DIMENSIONS

Data governance is a program of managing information assets to achieve defined information management goals. Governance establishes the processes that are needed and designates the responsibilities of people to achieve the goals.

The process dimension of data governance includes policies, procedures, and rules. The people dimension of data governance includes organizational structure, roles, responsibilities, decision rights, and accountabilities.

These dimensions create a management framework within which data and information are managed and technologies are employed to achieve specific information management goals.

GOVERNANCE GOALS

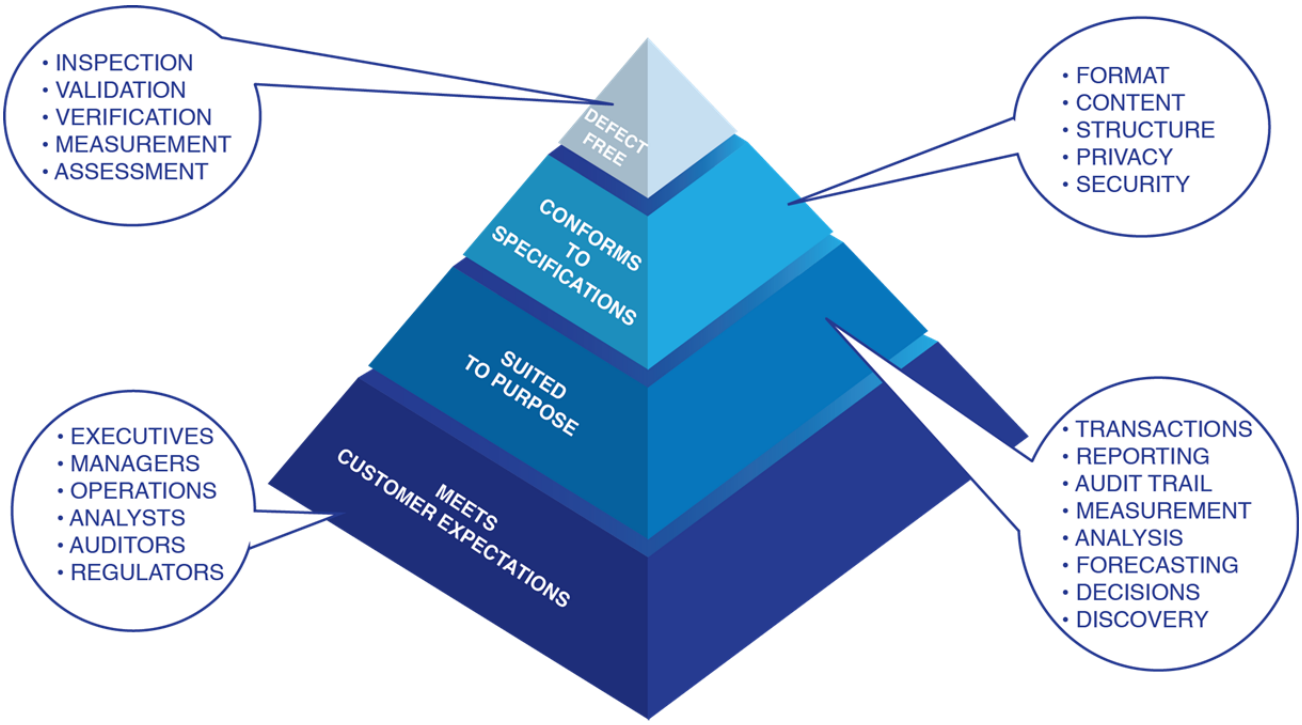
Goals are the driving force of data governance—the reasons to govern data and the foundations upon which governance processes are built. Common goals include such things as:

- data quality
- data security
- data standardization
- data consolidation
- regulatory compliance
- information utility
- information management maturity

As with any program, data governance goals are not static. They change over time as the business evolves and the governance program matures.

Data Quality

Data Quality Concepts



Data Quality

Data Quality Concepts

QUALITY DEFINITIONS

The Merriam-Webster dictionary defines quality as "degree of excellence." The important point here is that quality is not an absolute, but something that exists in degrees. One common definition describes high quality as *defect free*. This interpretation comes from the community of quality practitioners who base their practice on the principle of zero defects. They define quality as *conformance to specifications* and defects as variance from specifications. Another widely used definition states that quality is *suitability to purpose*—a thing is of high quality when it is well suited to its intended purpose, and it is of poor quality when badly suited to its purpose. The principles of Total Quality Management (TQM) define quality as consistently *meeting customer expectations*. This principle promotes the idea that quality doesn't reside within a product; it can only be judged in relation to the expectations of the customer using the product.

DATA AND DEFECTS

Defect-free data requires identification of the things that are data defects (more about this later), after which you can manage data by inspecting it to find defects, by validating and verifying data as free of defects, and by measuring defects as part of data quality assessment.

DATA AND SPECIFICATIONS

Conformance to specifications requires formal data specifications, which may address any or all of data format, content, and structure as well as usage-oriented specifications such as those for data privacy and security. Data quality management will test data against specifications.

DATA AND PURPOSE

Suitability to purpose must consider all purposes for which data is used, ranging from business transactions and operational reporting to BI and analytics. Expect the quality criteria to vary widely among the different uses. Variations in quality criteria increase the level of difficulty in data quality management, but attention to them makes quality management efforts more effective and far-reaching.

DATA AND EXPECTATIONS

Data quality as meeting customer expectations must consider the wide range of data and information consumers. Expect wide variation in the expectations through the range of consumers, both internal and external. The quality management implications of varied expectations are much like those for varied purpose—greater complexity and greater impact.

Module 6

BI and Analytics Technology

The Technology Stack

Technology Layers

Decision Management	business rules engines, optimization, simulation, forecasting
Business Analytics	mining, modeling, visualization, predictive analytics, text analytics, geospatial analytics
Business Applications	enterprise reporting, performance mgmt, scorecards, dashboards, operational BI
Information Services	query and access, reporting (tabular & graphical), OLAP
Data Integration	ETL/ELT, data virtualization, big data integration
Data Management	data storage, DBMS, big data technologies, data profiling, data quality, metadata management
Data Sourcing	internal source systems, data connectivity & APIs, syndicated & subscription data services
Infrastructure	servers, operating systems, networks, security, performance

The Technology Stack

Technology Layers

FROM TECHNICAL FOUNDATION TO BUSINESS VALUE

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

- *Infrastructure* includes all of the hardware and foundation software needed to enable and implement higher-level layers in the stack—the servers, operating systems, networks, security management, and performance optimization.
- *Data sourcing* includes the systems and databases from which data is obtained as well as the languages, protocols, services, and connectors that provide access to data sources.
- *Data management* includes data storage and database management systems, big data technologies such as NoSQL, data profiling tools, data quality and cleansing tools, and metadata management systems and repositories.
- *Data integration* includes ETL and variations such as ELT (extract, load, then transform) for data consolidation, data virtualization tools for federation and unstructured data integration, 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, analytics modeling, data visualization, and advanced analytics methods such as text analysis and spatial analysis.
- *Decision management technologies* are needed to implement decision management systems. They include business rules engines and tools to support simulation, forecasting, and process optimization.

