Previews of TDWI course books are provided as an opportunity to see the quality of our material and help you to select the courses that best fit your needs. The previews can not 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. The pages shown are not consecutive. The page numbers as they appear in the actual course material are shown at the bottom of each page. All table-of-contents pages are included to illustrate all of the topics covered by a course.

TDWI Data Warehousing Architectures

Choosing the Right Data Warehousing Approach

TABLE OF CONTENTS

Module 1

Introductory Concepts

Context and Scope

Definition of Data Warehousing

A data warehouse is a subject-oriented, integrated, non-volatile, time-variant collection of data organized to support management needs.

W. H. Inmon, Database Newsletter, July/August 1992

I look at Information Warehousing as something that provides two real business benefits: data integration and data access. It removes much unnecessary and unwanted data and processing from the classic operational environment

Susan Osterfelt, Executive Systems Journal, January 1993

The process whereby organizations extract value from their information assets through the use of special stores called data warehouses

Ramon Barquin, Planning & Designing the Data Warehouse, 1997

The Data Warehouse is nothing more than the union of all the constituent data marts.

Ralph Kimball, et al, , The Data Warehouse Life Cycle Toolkit , 1998

A data warehouse is a data structure that is optimized for distribution. It collects and stores integrated sets of historical data from multiple operational systems and feeds them to one or more data marts. It may also provide end-user access to support enterprise views of data."

TDWI

Context and Scope

Definition of Data Warehousing

CONSENSUS DEFINITIONS

Multiple, and sometimes conflicting, definitions of data warehousing terms do exist (many of the differences will be discussed later in this course). Still, there is some consensus of definitions—or at least intent—among the varied definitions. Common themes are: integrated, subject-oriented, non-volatile, time-variant, accessible, meets business information needs, and is a process of turning data into information.

INTEGRATED

Warehousing provides a single comprehensive source of information for and about the business. Answering a business question does not require accessing multiple sources across a variety of technology platforms with potentially inconsistent data.

SUBJECT-ORIENTED

Data and information is organized and presented as business subjects aligned with information needs, not as computer files designed for transactional processing needs.

TIME-VARIANT

The warehouse contains a history of the business, as well as relatively current business information. Structures and intervals are kept consistent across time, allowing time-specific analytics such as trend analysis.

NON-VOLATILE

The warehouse provides stable information. Business data, once written to the warehouse, is not overwritten. The body of data grows through regular addition of new data in a way that maintains accurate historical records.

ACCESSIBLE

The primary purpose of a data warehouse is to provide readily accessible information to business people. The data is organized for easy access.

MEETS BUSINESS INFORMATION NEEDS

Warehousing provides an organized data resource, against which a variety of standard tools can be applied by business knowledge workers to manipulate, analyze, and generate answers to business questions.

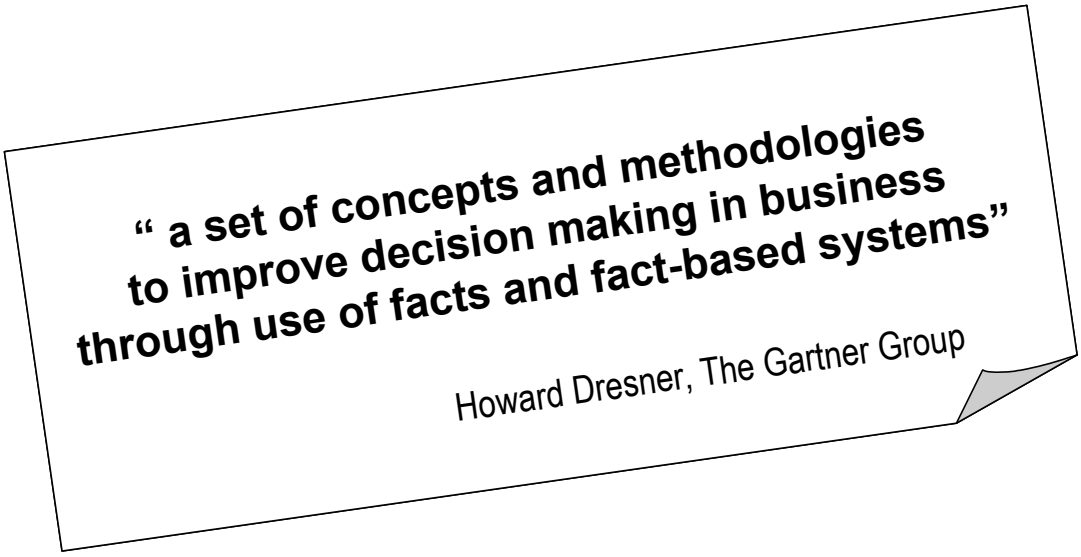
DATA-TO-INFORMATION PROCESS

David Loshin, in *Business Intelligence: The Savvy Manager's Guide*, describes the data-to-information process as: "…The process of determining what data are to be collected and managed and in what context." Data is a collection of facts from which conclusions can be drawn. Through the process of interpretation by people or systems, data takes on meaning and becomes information. The warehouse is the primary source used specifically for analyzing business data and developing informational views of direct value to business people.

Context and Scope

Definition of Business Intelligence

Business Intelligence



- **The goal is improved decision making.**
- **Concepts & methodologies – Not just technology.**
- **Facts & fact-based systems are necessary to implement.**

Context and Scope

Definition of Business Intelligence

BUSINESS INTELLIGENCE

Howard Dresner (Gartner Group) first defined business intelligence as shown on the facing page. As BI became a mainstream term, additional definitions have emerged. Among them is the following definition:

"BI is neither a product nor a system. It is an architecture and a collection of integrated operational as well as decision-support applications and databases that provide the business community easy access to business data." (Larissa T. Moss and Shaku Arte, *Business Intelligence Roadmap,* Pearson Education, 2003)

TDWI defines business intelligence as: "The processes, technologies, and tools needed to turn data into information, information into knowledge, and knowledge into plans that drive profitable business actions. Business intelligence encompasses data warehousing, business analytic tools, and content knowledge management."

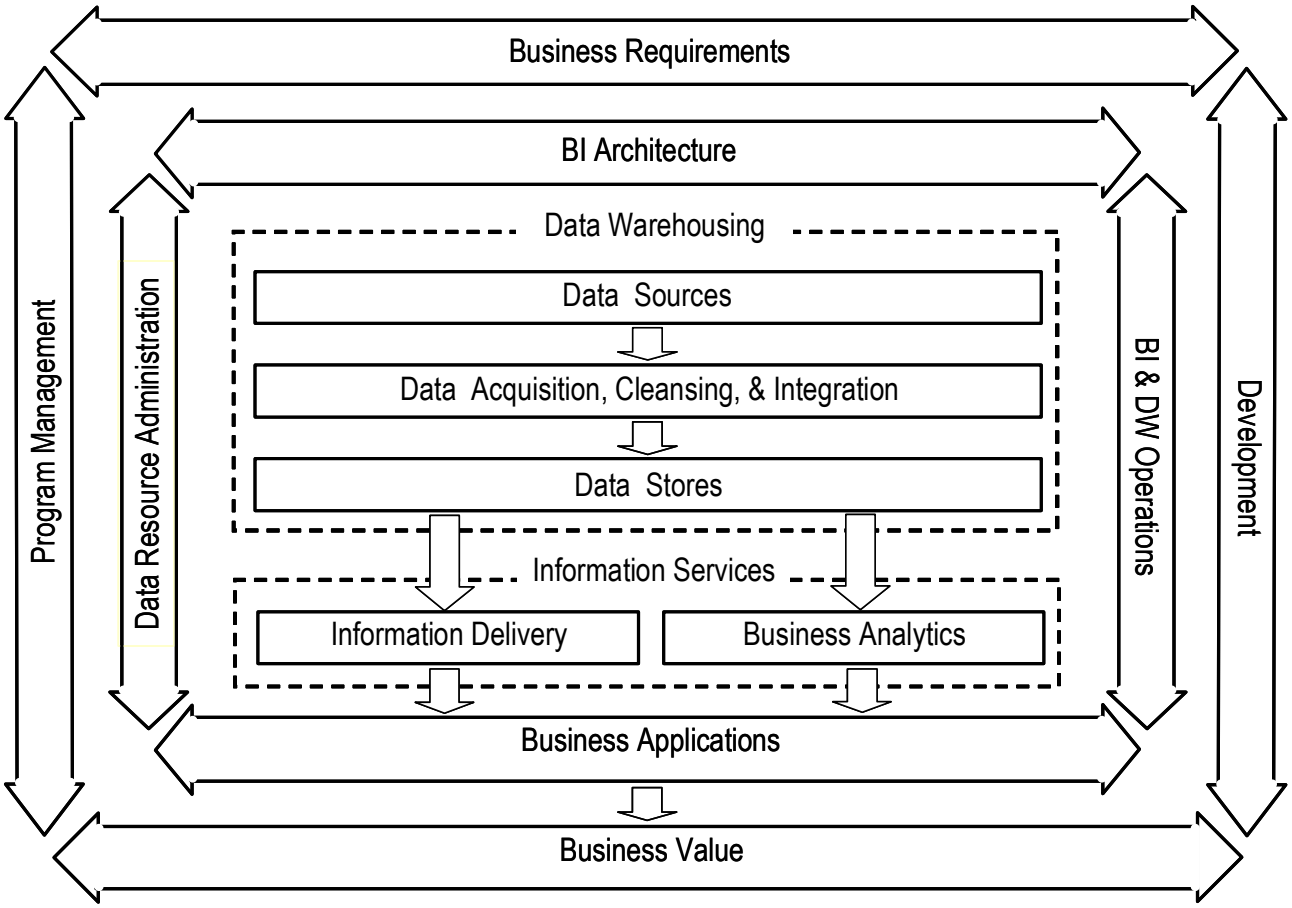
Business Intelligence provides the ability to transform data into usable, actionable information for business purposes. BI requires:

- Collections of quality data and metadata important to the business
- The application of analytic tools, techniques, and processes
- The knowledge and skills to use business analysis to identify/create business information
- The organizational skills and motivation to develop a BI program and apply the results back into the business

The foundation that enables BI is the enterprise architecture—business, data, and technical. A well-implemented data warehousing program provides much of that foundation.

Context and Scope

Business Intelligence Framework



Context and Scope

Business Intelligence Framework

A STRUCTURE FOR BI COMPONENTS

The BI components framework identifies the parts of a BI program and the relationships among them. The framework consists of three layers:

Layer	Description
Business Layer	The components needed for BI to fit seamlessly into business organizations, processes, and activities.
Administration & Operation Layer	The components that connect technical components with business components.
Implementation Layer	The technical components needed to capture data, turn data into information, and deliver that information to the business.

BUSINESS LAYER

The business layer is made up of the following:

Component	Description
Business Requirements	The reasons to implement BI, and the kinds of results needed including information needs, essential business metrics, etc.
Business Value	The benefits anticipated from or achieved by BI, including such things as increased revenue, improved profit margins, risks mitigated or avoided, reduced costs, etc.
Program Management	The ongoing activities of managing the BI program for maximum business value, including establishing enterprise structures and standards, synchronizing multiple and parallel projects, realigning with changing business needs, etc.
Development	The project activities that create and deploy BI and DW products including methodology, project decomposition, project success measures, etc.

ADMINISTRATION & OPERATION LAYER

The administration & operation layer is composed of the following:

Component	Description
BI Architecture	Frameworks, standards, and conventions that describe BI environment components and the relationships among them including business, data, technology, organizational, and project architectures.
Business Applications	Business processes and procedures that access and/or receive information and employ that information to achieve business results.
Data Resource Administration	Policies, procedures, and processes for data governance including data owner, steward, and custodial responsibilities.
BI & DW Operations	Execution, monitoring, and maintaining acceptable quality, availability, and performance of the DW and BI functions and services.

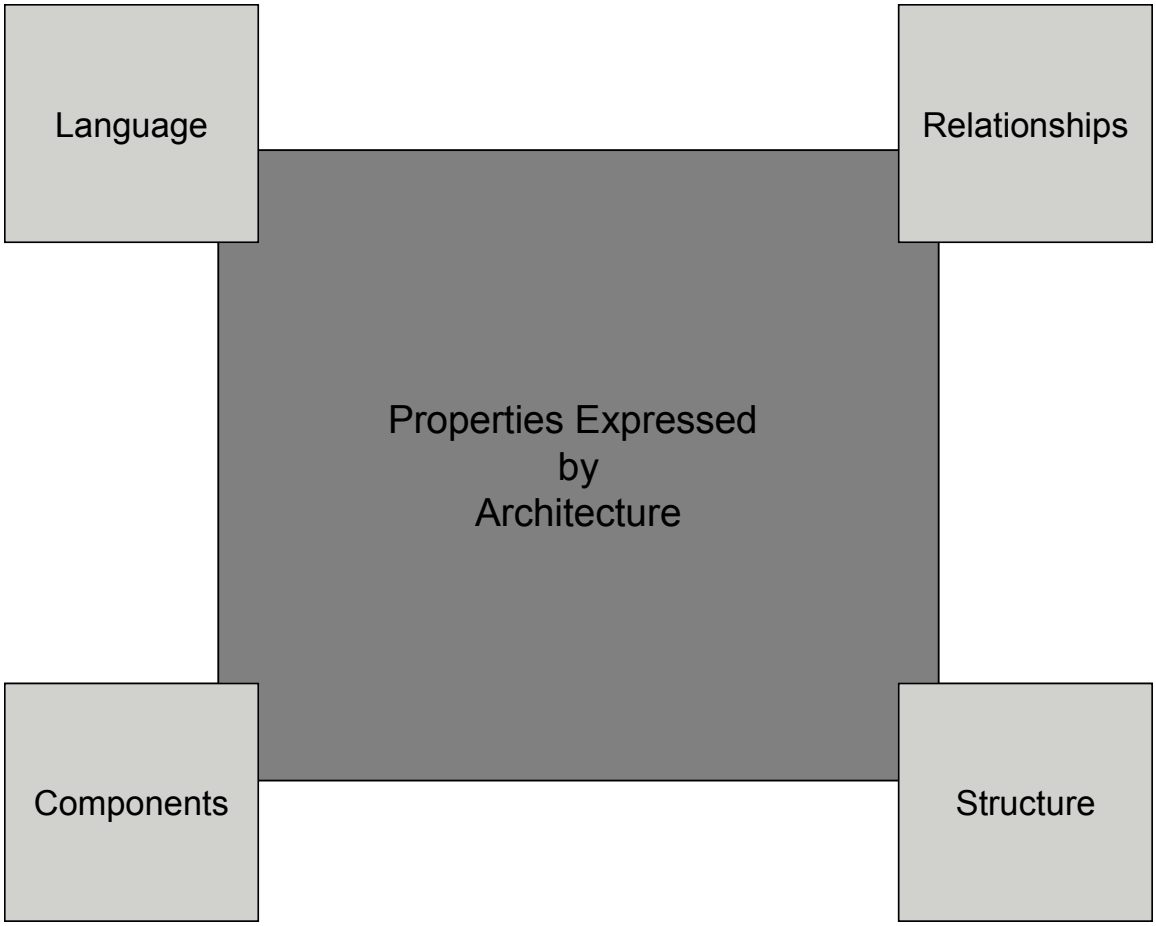
IMPLEMENTATION LAYER

The implementation layer consists of:

Component	Description
Data Warehousing	Systems, processes, and procedures to integrate data and prepare it to become information.
Information Services	Systems, processes, and procedures that turn data into information and deliver that information to the business.

Architecture and Methodology

Definition of Architecture



Architecture and Methodology

Definition of Architecture

DEFINITIONS

Architecture describes, at an abstract level, the major components of something to be constructed, the roles and functions of those components, and the relationships among them. Information systems architecture describes the components of information systems (data, function, network, etc.) and the roles and relationships of those components. Data warehousing architecture is a subset of information systems architecture specifically oriented toward systems whose purpose is to receive disparate data and deliver integrated information.

Many data warehousing architectures exist today, with differences in the components, structure, relationships, and language. Architectural needs across multiple organizations are as varied as their information needs. The ideal architecture for one organization may be especially ill-suited to another. The purpose of this course is to examine several representative architectures and provide guidelines to determine which is best suited (or which parts of several may be combined) to meet your organization's unique needs.

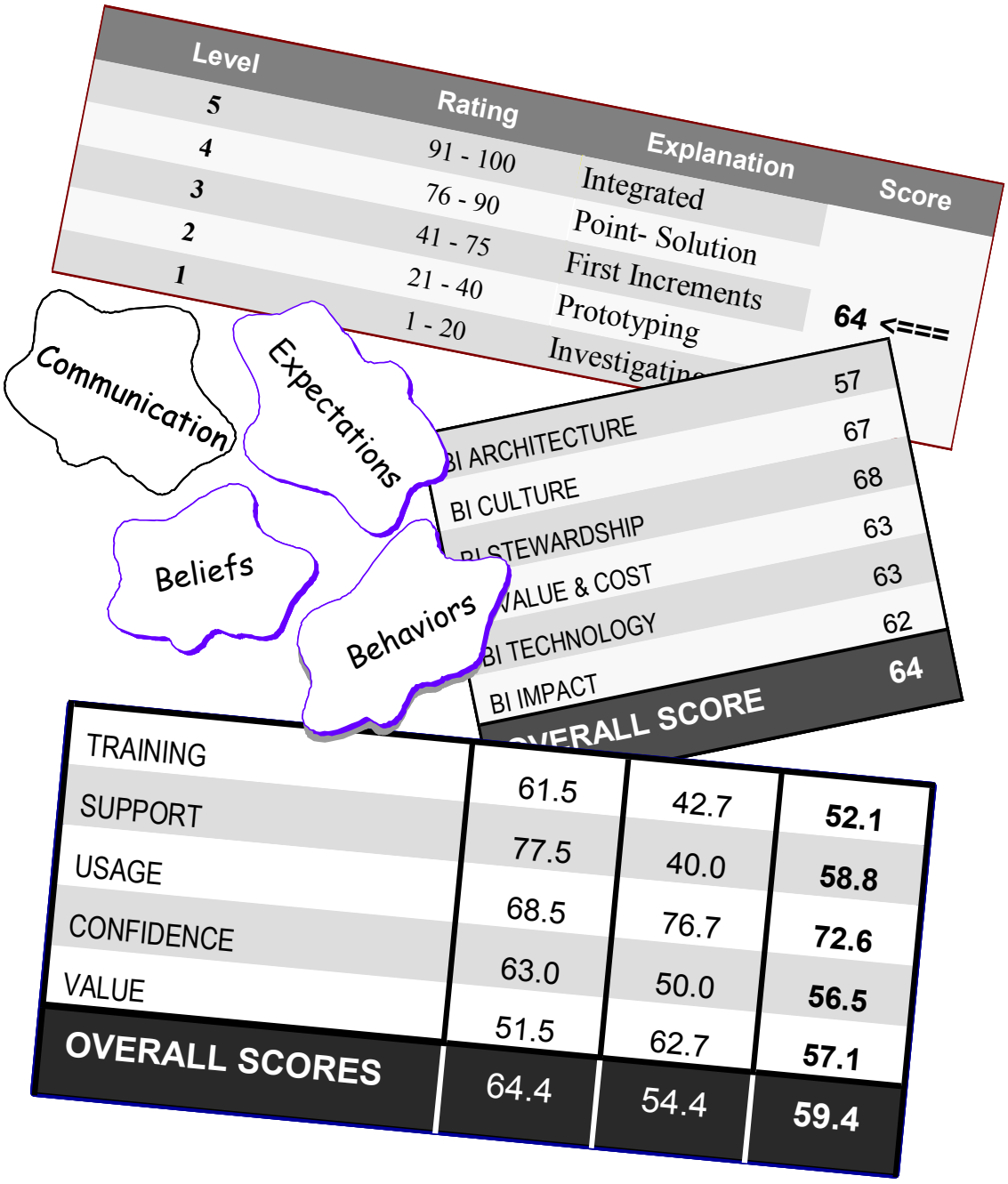
Module 2

Architecture Assessment

This page intentionally left blank.

Principles of Assessment

Introduction and Definition



Principles of Assessment

Introduction and Definition

INTRODUCTION TO ASSESSMENT

In general terms, an assessment is a structured process of evaluation. It is a means used to appraise and measure some relevant characteristics a "thing" we are interested in. These measured characteristics are used to help us objectively judge if the "thing" is suitable based on our expectations to help us meet our defined goals and objectives.

The assessment process includes the following steps.

1. define the "thing" that needs to be assessed
2. identify meaningful dimensions or characteristics of the "thing" that will help decision makers understand if and how it will serve its intended purpose
3. determine a measurement scale and technique to quantify the characteristics that can be compared to various reference points such as standards, targets, thresholds, etc
4. determine a set measurement levels or thresholds that provide context around the raw measurements to help decision makers interpret the raw measurement points into informed decisions

Assessment techniques applied to data warehousing Architectures allow us to objectively evaluate a variety of characteristics that define how a working system would behave if it conformed to the architecture under investigation. Assessment examines characteristics such as cost, value, quality, effort, alignment with best practices, and probability of success.

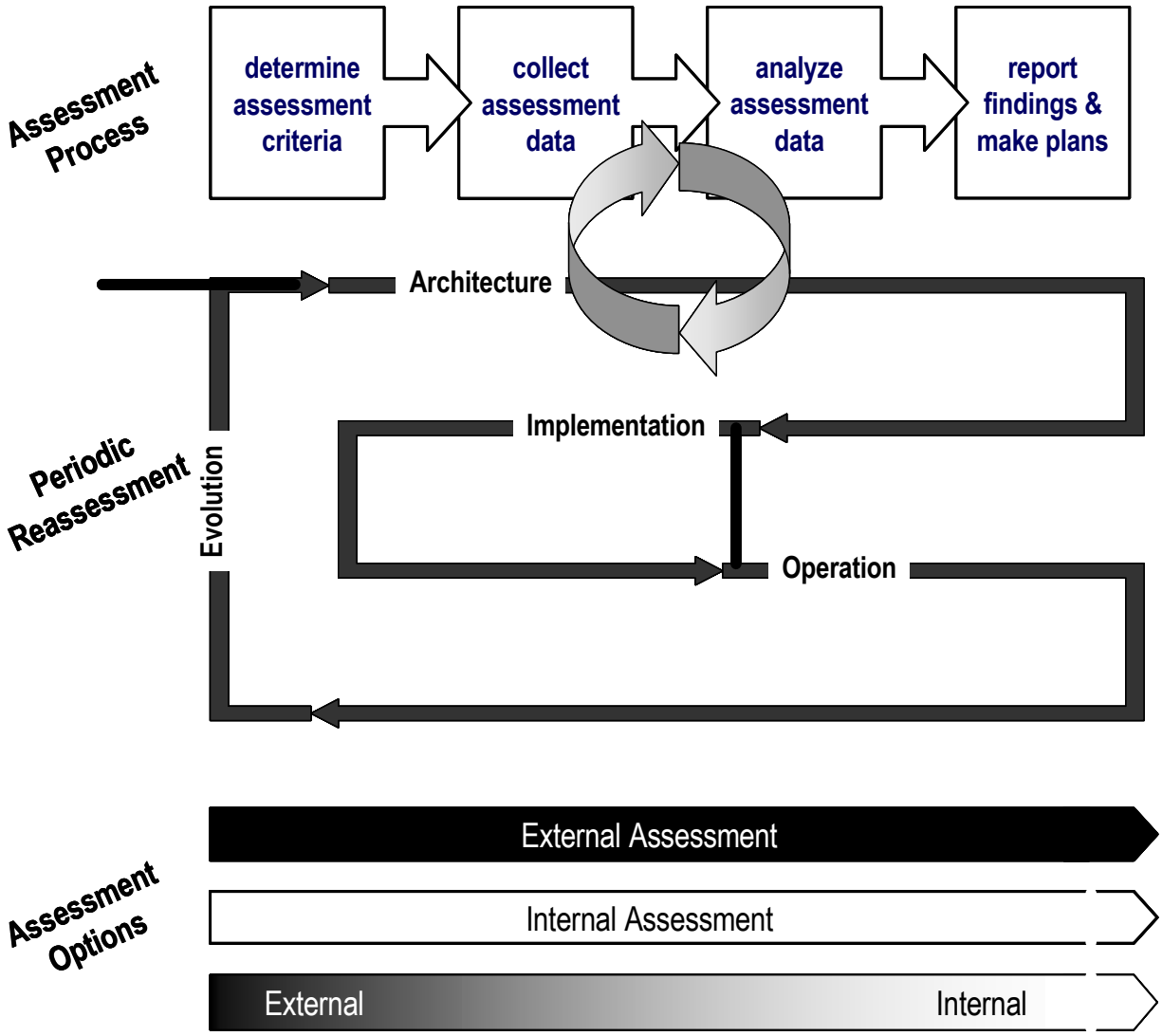
Measurement is a key element of effective assessment. Although many of the factors to be evaluated are somewhat subjective, quantifying those factors is a necessary step. Measures and ratings help to understand the architectural details relative to a range of possibilities. More importantly they offer a means to compare results of periodic assessments and to measure growth and change.

GOALS AND OBJECTIVES

It must be understood that an assessment is always conducted and interpreted within the context of goals and objectives. If goals and objectives are not clearly defined, it is not possible to complete an assessment. Setting of goals and objectives for the "thing" being assessed is critical to obtain meaningful results from the assessment. A major challenge encountered when conducting an assessment is obtaining agreement and alignment across multiple stakeholder groups about the correct set of goals and objectives.

Principles of Assessment

Process Activities



Principles of Assessment

Process Activities

EXTERNAL vs. INTERNAL ASSESSMENT

Two techniques – external assessment and self assessment – are common to evaluate readiness. These techniques may be used, individually or in combination, to meet particular assessment needs. **External Assessment** is a formal assessment of readiness by an experienced data warehouse consultant. This technique typically delivers a draft program charter and a customized plan for near-term projects. **Self Assessment** uses a survey method to measure perceptions. This technique identifies strengths, weaknesses, and areas of risk as an aid to planning and management. External assessment is useful at the start of large and complex programs or to revitalize a severely troubled program. For ongoing and repeatable assessment, self-assessment is generally more practical.

QUICK AND CONCISE

BI/DW assessment should be a relatively simple and straight-forward process that avoids complexity and minimizes effort. Complex, time-consuming, or difficult processes discourage assessment and diminish the benefits of regular review. A quick, concise, easily-executed, and tool-enabled process makes regular assessment practical.

FOUR STEPS

The four basic steps of assessment are:

Determine Assessment Criteria based on the scope and maturity of your program. Limit assessment complexity and effort by assessing only those things related to the current state of the program.

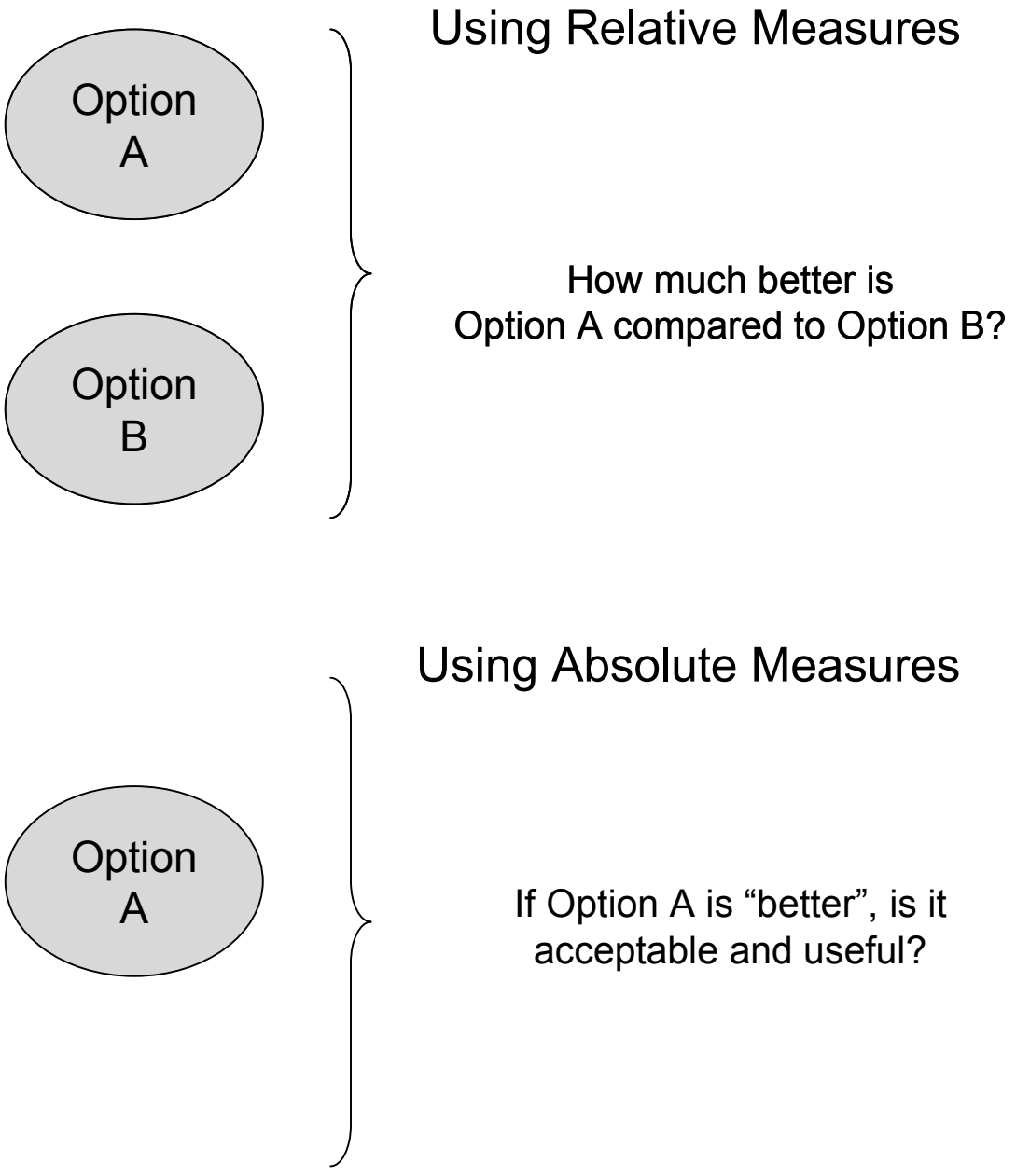
Collect Assessment Data using survey methods. Collecting a small amount of data is as effective as working with large volumes. Remember that the goal of assessment is to understand how BI/DW solutions align with business, cultural, organizational, and technological needs and constraints. Surveys of individual perceptions and beliefs are as useful as more concrete measures and result in small volumes of easily gathered data.

Analyze Assessment Data to answer three questions: (1) Where are the strengths? (2) Where are the weaknesses? (3) Where are the risks?

Report Findings and Make Plans. Continue the assessment by itemizing each of the strengths, weaknesses, and risks indicated by analysis. Then make plans – specific action items with assigned responsibility – to gain advantage from strengths limit the impact and duration of weakness and lessen the probability that risks become realities.

Measurement Concepts

Relative and Absolute Measures



Measurement Concepts

Relative and Absolute Measures

RELATIVE MEASURES

Relative measures belong to a category of measures used to help us understand how specific characteristics vary across two or more architectures being assessed and compared. When relative measures are being designed, the measurement technique does not concern itself with placing a given measure on a scale to imply that a given architecture has a certain rating. This type of measure is concerned with measuring whether a given characteristic is "better" in a given architecture than in another one. It should be recognized that this type of measure does not indicate if either architecture is "good" at something. It simply expresses the fact the architecture "A" is better than architecture "B".

Relative measures are useful to select a those candidate architectures that are "better" at something than others are.

ABSOLUTE MEASURES

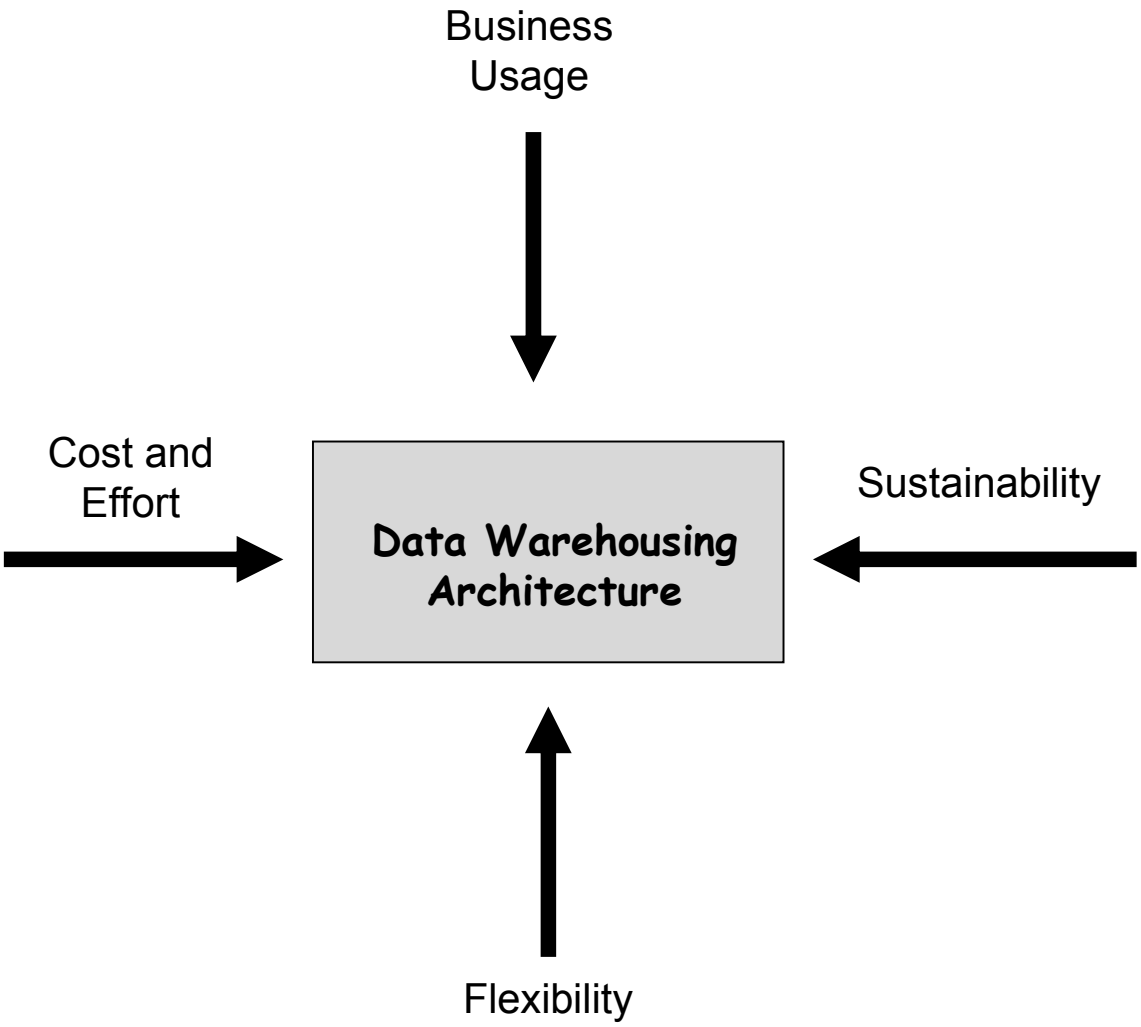
Absolute measures indicate how well a given architecture performs a specific function or implements a certain characteristic. This type of measure is used to understand if the given architecture is "good" or "poor" at something as opposed to just indicting that it is better or worse than other candidates.

Absolute measures are useful to eliminate any candidate architectures that do not meet some minimum level of performance.

Both types of measures are useful an assessment process.

Measurement Concepts

Dimensions of Architecture Performance



Dimensions of Architecture Performance

Measurement Concepts

Dimensions of Architecture Performance

DESCRIPTION

Factors are the characteristics of architecture that are considered to be important by the assessment team. Data warehousing architectures express many different characteristics and relationships between those characteristics. This course introduces and defines sixteen factors that describe various traits about different architectures. Not all factors are equally important to all assessments at all times. The assessment team must decide which factors are important and then conduct the assessment based on that selection.

The individual factors used for evaluation are grouped into the following dimensions.

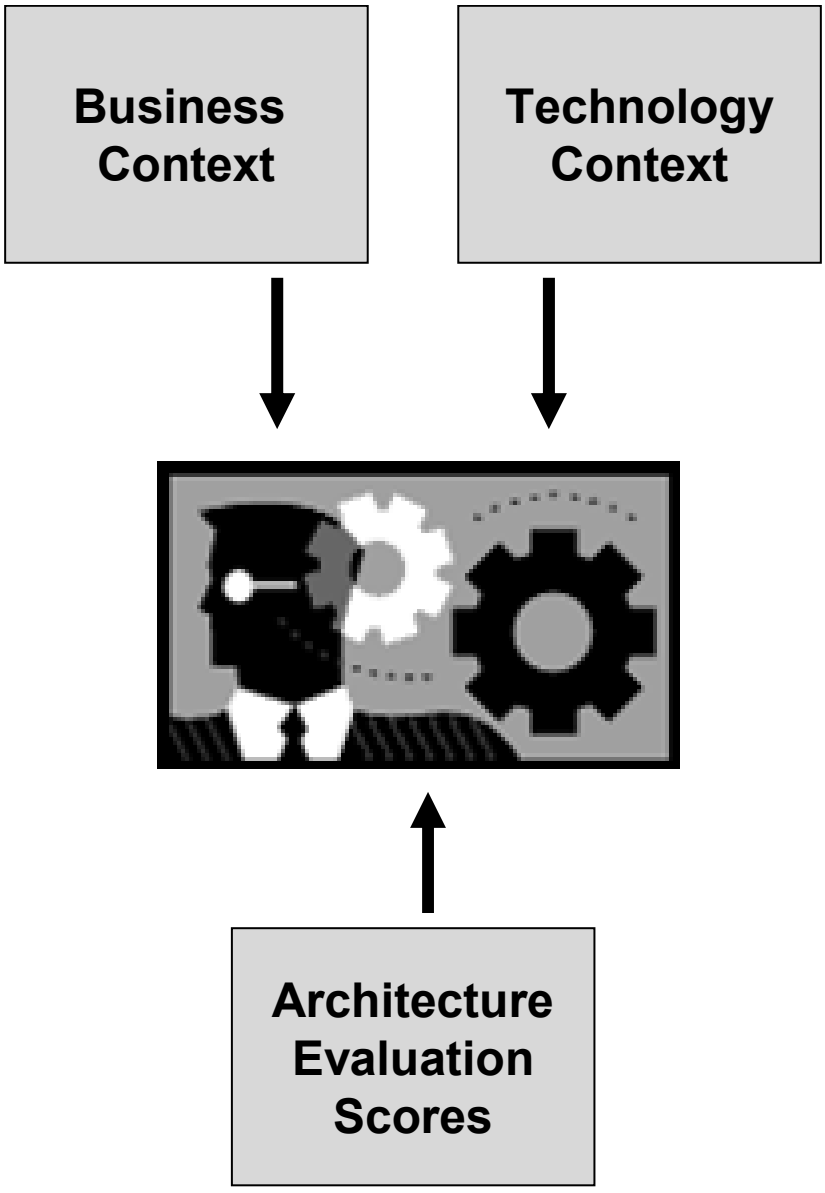
- Business Usage
- Cost and Effort
- Sustainability
- Flexibility

Before the factors can be selected to support a given assessment, it is important to understand their underlying concepts and what characteristics of the architecture that help measure.

The individual factors are defined in a later section of the course.

Analysis and Decision Making

Establishing Context



Analysis and Decision Making

Establishing Context

THE IMPERATIVE OF CONTEXT

Establishing the business and technology context for the assessment team is critical if it is expected to produce any kind of meaningful evaluation scores.

Context is also critical for data warehousing stakeholders and managers to interpret the scores of the assessment team. It should be recognized that the purpose of conducting an architectural assessment is to make decisions about either modifying an existing architecture or selecting a new architecture. These decisions are not possible without providing the right level of context around the architecture and the purpose it is trying to achieve.

There are different sets of information that make up a useful context. The information sets should be defined and provided to the assessment team and to the decision making team. Context information includes the following information.

1. What business outcomes and types of insights does the organization want to achieve from its investment in data warehousing?

2. What are the short term and long term objectives of the Data Warehousing environment as defined by the different groups of sponsors and stakeholders?

3. What timing, financial, resource and external constraints are in place that will dictate what architectural trade-offs have to be made to achieve the defined short term and long term objectives?

4. Who are the key stakeholders and sponsors of the data warehousing environment and what are their short term and longer term expectations regarding the environment?

5. What previous business and technology decisions have been made that will influence the data warehousing architecture?

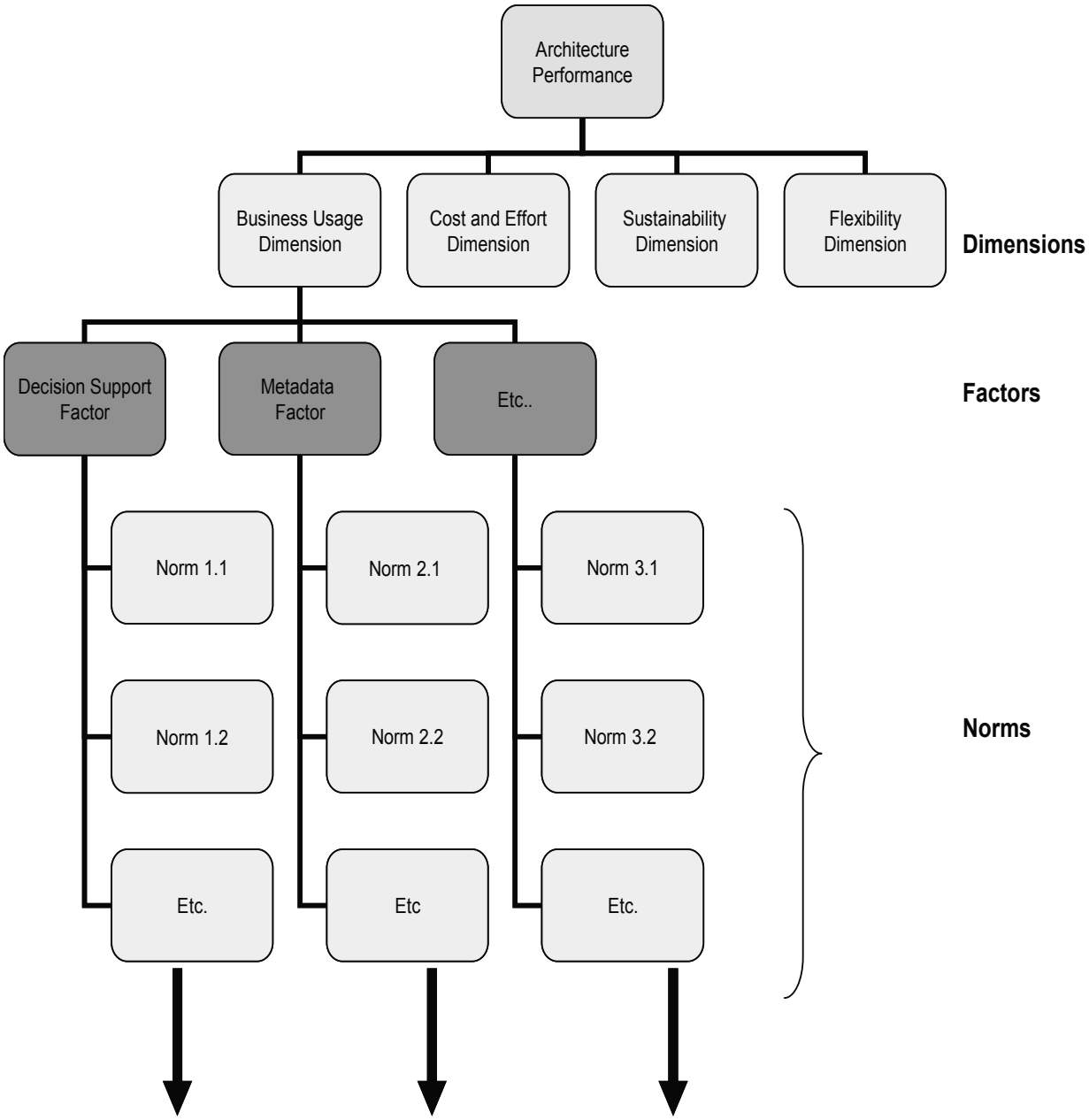
Module 3

Assessment Factors

This page intentionally left blank.

Introductory Concepts

Overview



Introductory Concepts

Overview

DESCRIPTION

The decision making processes used to evaluate and select an appropriate data warehousing architecture for an organization depend on having insight into how well different architectures are expected to perform according to different categories of performance.

A measurement technique designed to quantify the expected performance of a given architecture provides input to the decision making process. The performance measurements, combined with other information such as goals, priorities, expectations and constraints provide context to the raw measurements.

The information produced by the measurement technique will be organized into a dimensional structure to facilitate the analytics and decision making process.

DIMENSIONS AND FACTORS

The dimensions of architecture performance, introduced earlier are summarized below.

- Business usage – characteristics related to how information will be consumed and analyzed. Integration scope, quality, latency and granularity are included in this dimension.
- Cost and effort – characteristics related to the cost and effort to implement the features defined by architecture.
- Sustainability – characteristics related to how easy the implemented architecture can be maintained and supported over time.
- Flexibility – characteristics related to how well the implemented architecture can respond to changing business and technology conditions.

Architectural dimensions are used to group more detailed factors that further define architecture performance at a lower level. A total of sixteen factors are identified and used to measure architectural performance.

Introductory Concepts

Architectural Factors

✓	Decision Support	Business Usage
✓	Integration	
✓	Metadata Consistency	
✓	Unstructured Data	
✓	Data Latency	
✓	Data Quality	
✓	Granularity	
✓	Cost to Deploy	Cost & Effort
✓	Cost to Operate	
✓	Time to Deliver	
✓	Scalability	Sustainability
✓	Sustainability	
✓	Staffing and Skills	
✓	Persistence	
✓	Structure of Data	Flexibility
✓	Technology Dependency	

Introductory Concepts

Architectural Factors

BUSINESS USAGE DIMENSION

The following factors are lower level details making up the business usage architecture dimension.

- Decision support factor
- Integration factor
- Metadata consistency factor
- Unstructured data factor
- Data latency factor
- Data quality factor
- Granularity factor

COST AND EFFORT DIMENSION

The following factors are defined as lower level details within the cost and effort architecture dimension.

- Cost to deploy factor
- Cost to operate factor
- Time to deliver factor

SUSTAINABILTY DIMENSION

The following factors are defined as lower level details within the sustainability architecture dimension.

- Scalability factor
- Sustainability factor
- Staffing and skills factor
- Persistence factor

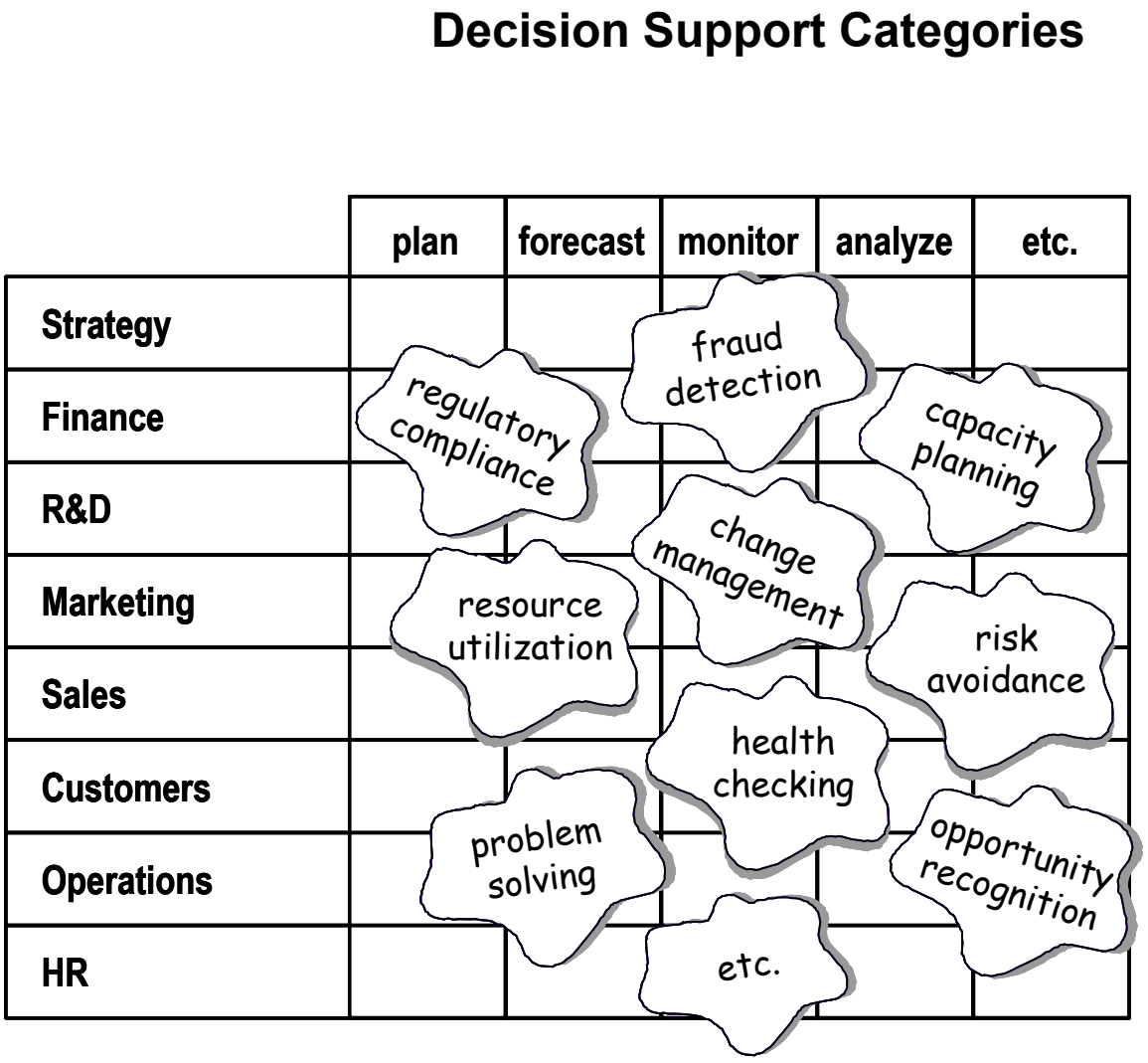
FLEXIBILITY DIMENSION

The following factors are defined as lower level details within the flexibility architecture dimension.

- Structure of data factor
- Technology dependency factor

Business Usage Dimension

Decision Support Factor



Business Usage Dimension

Decision Support Factor

INFORMATION NEEDS VARIABLES

Some organizations engage in data warehousing primarily to deliver information to business executives in support of strategic decision making. Others pursue warehousing to meet the needs of managers involved in day-to-day operation of the business – specifically needs for business information and analytics.

A key question to be answered when evaluating data warehousing architectures is:

Who are the customers and what are their needs & expectations?

SUPPORTING STRATEGIC NEEDS

Strategic decision support targets business executives to provide information about progress toward meeting business goals, effectiveness of business initiatives in accomplishing those goals, and discovery of opportunity for other initiatives. Strategic decision-making normally demands an enterprise perspective. If your expectations for data warehousing include strategic decision support, then the architecture must address enterprise integration to some degree.

SUPPORTING TACTICAL NEEDS

Tactical decision support targets managers of business units to provide information and support analytics related to business tactics and processes. Tactical information needs are more limited in scope than strategic needs – typically confined to a specific process or organization. Enterprise integration is less significant in supporting these kinds of expectations.

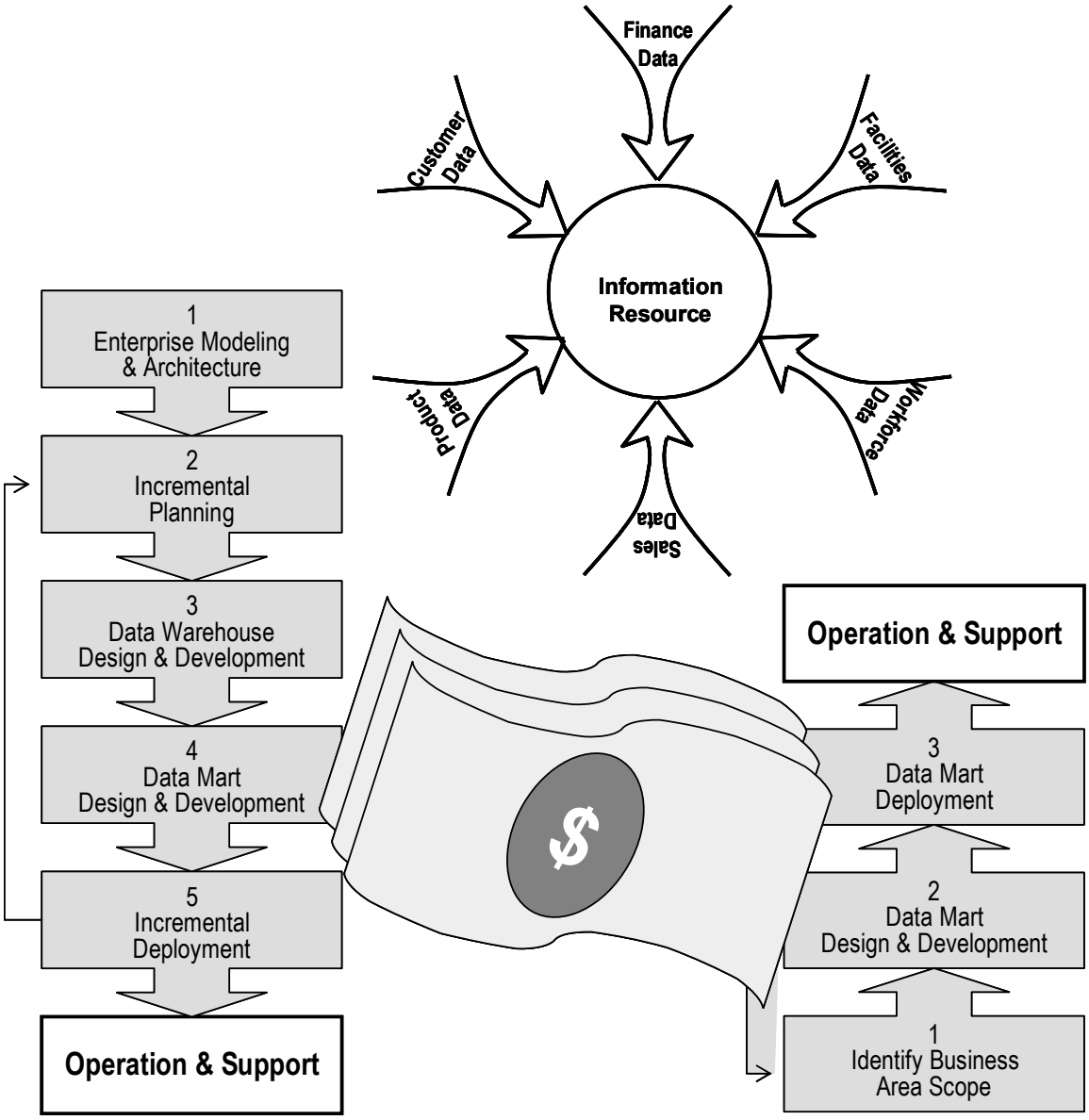
HOW TO ASSESS

Effectiveness of support for both strategic and tactical decision support is evaluated using the criteria described below:

architecture properties	methodology properties
• has structures & components to deliver the right kinds of information, in friendliest form, to executives responsible for business strategies • has structures & components to deliver the right kinds of information, in the most useful forms, to managers and knowledge workers responsible for business processes	• has specific activities & deliverables for enterprise focus in the business architecture • has specific activities & deliverables for business process alignment in the business architecture • has specific activities & deliverables to identify & express information needs of executives, managers & knowledge workers

Cost and Effort Dimension

Cost to Deploy Factor



Cost and Effort Dimension

Cost to Deploy Factor

DEPLOYMENT COST CONSTRAINTS

Cost of delivery is one of the most commonly considered critical success factors for data warehousing. Sponsor and business tolerance for costly start-up efforts is relatively low. Moderate costs and rapid return on investment are expectations of many warehousing initiatives.

When evaluating warehousing architectures it is important to know:

What cost constraints are imposed on the development effort?
What return on investment expectations have been established?

ARCHITECTURE AND DEPLOYMENT COSTS

Top-down approaches producing a warehouse with dependent data marts typically have higher development cycles than bottom-up methods, especially for early increments. As the body of data in the warehouse grows and is reused, development costs may decline.

Bottom-up approaches producing independent data marts usually have lower development costs than can top-down approaches. Development cost does not significantly change as the warehouse matures.

Bottom-up approaches producing data marts with a shared architecture normally have higher start-up costs than independent marts, as some expense is incurred to identify and define conformed dimensions. Once standards of conformity have been defined, development costs are similar to those for independent data marts.

HOW TO ASSESS

Adaptability to deployment cost expectations is evaluated using the criteria described below:

architecture properties	methodology properties
• business architecture that defines the scope of the warehousing program • limited startup time & projects with small scope • deployable with limited, moderate cost technology	• incremental & evolutionary deployment through a series of small projects • not dependent on one or a few costly products

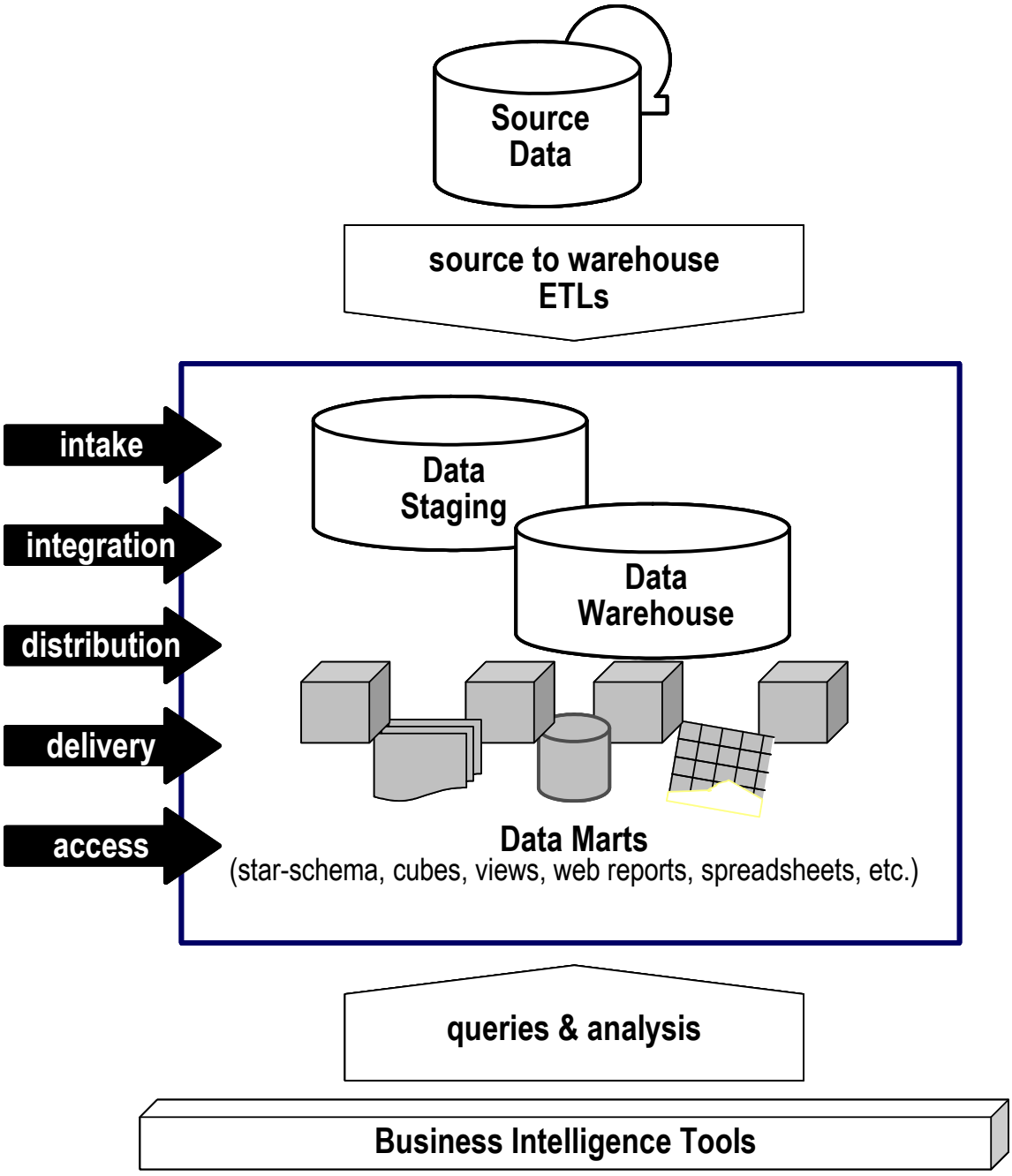
Module 4

Architecture Building Blocks

This page intentionally left blank.

Data Architecture Models

Functions and Roles of Data Stores



Data Architecture Models

Functions and Roles of Data Stores

THE ROLES Every data warehousing environment, regardless of architecture and flow of data, must provide for five roles to be complete. Different architectures assign these roles to data stores in various ways.

INTAKE Data stores with intake responsibility receive data into the warehousing environment. Data is acquired from multiple source systems of varying technologies, at different frequencies, and into numerous warehousing files and/or tables. Further, the data typically requires many and diverse transformations. Most data is extracted from operational systems whose data is most certainly not all clean, error-free, and/or complete. Data cleansing is commonly performed as part of the intake process to ensure completeness and correctness of data.

INTEGRATION Integration describes how the data fits together. The challenge for the warehousing architect is to design and implement consistent and interconnected data that provides readily accessible, meaningful business information. Integration occurs at many levels—"The key level, the attribute level, the definition level, the structural level, and so forth …" (*Data Warehouse Types*, www.billinmon.com) Additional data cleansing processes, beyond those performed at intake, may be required to achieve desired levels of data integration.

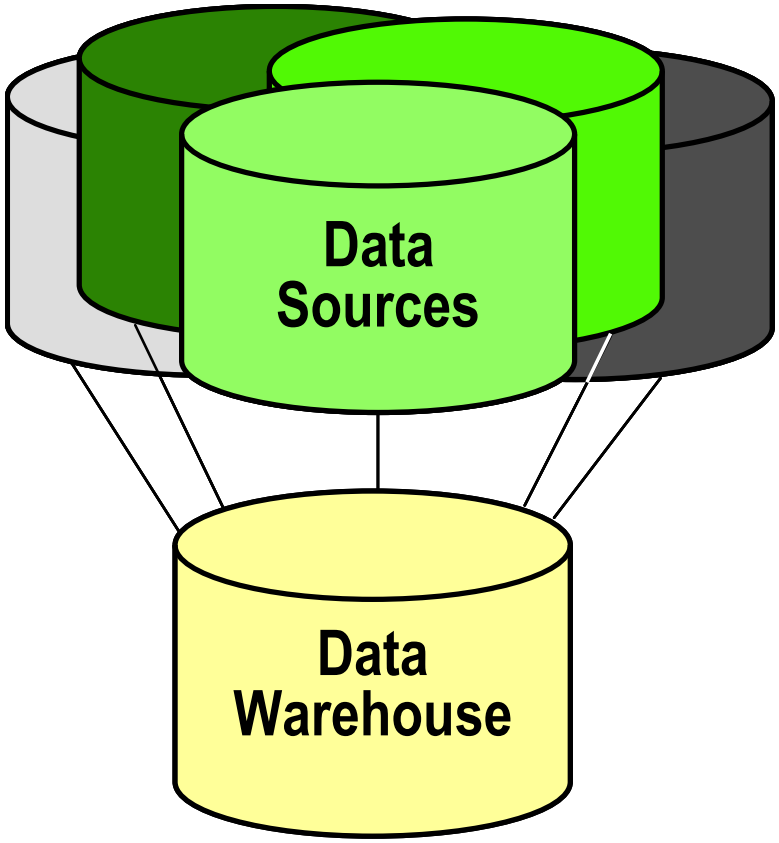
DISTRIBUTION Data stores with distribution responsibility serve as long-term information assets with broad scope. Distribution is the progression of consistent data from such a data store to those data stores designed to address specific business needs for decision support and analysis.

DELIVERY Data stores with delivery responsibility combine data as "in business context" information structures to present to business units who need it. Delivery is facilitated by a host of technologies and related tools—data marts, data views, multidimensional cubes, web reports, spreadsheets, queries, etc.

ACCESS Data stores with access responsibility are those that provide business retrieval of integrated data—typically the targets of a distribution process. Access-optimized data stores are biased toward ease of understanding and navigation by business users.

Data Architecture Models

Centralized Hub Structure



Centralized Hub Structure

Data Architecture Models

Centralized Hub Structure

DESCRIPTION

A common structure within the data architecture is the centralized hub. This structure is commonly a normalized relational design. Diverging from its strict third normal form logical view, it can also support de-normalized structures and derived values to optimize the delivery and access functions. Its primary roles are to achieve data integration using a common data structure and to adapt to longer term business changes because of its normalized and more flexible design.

SUPPORTING THE FIVE FUNCTIONS

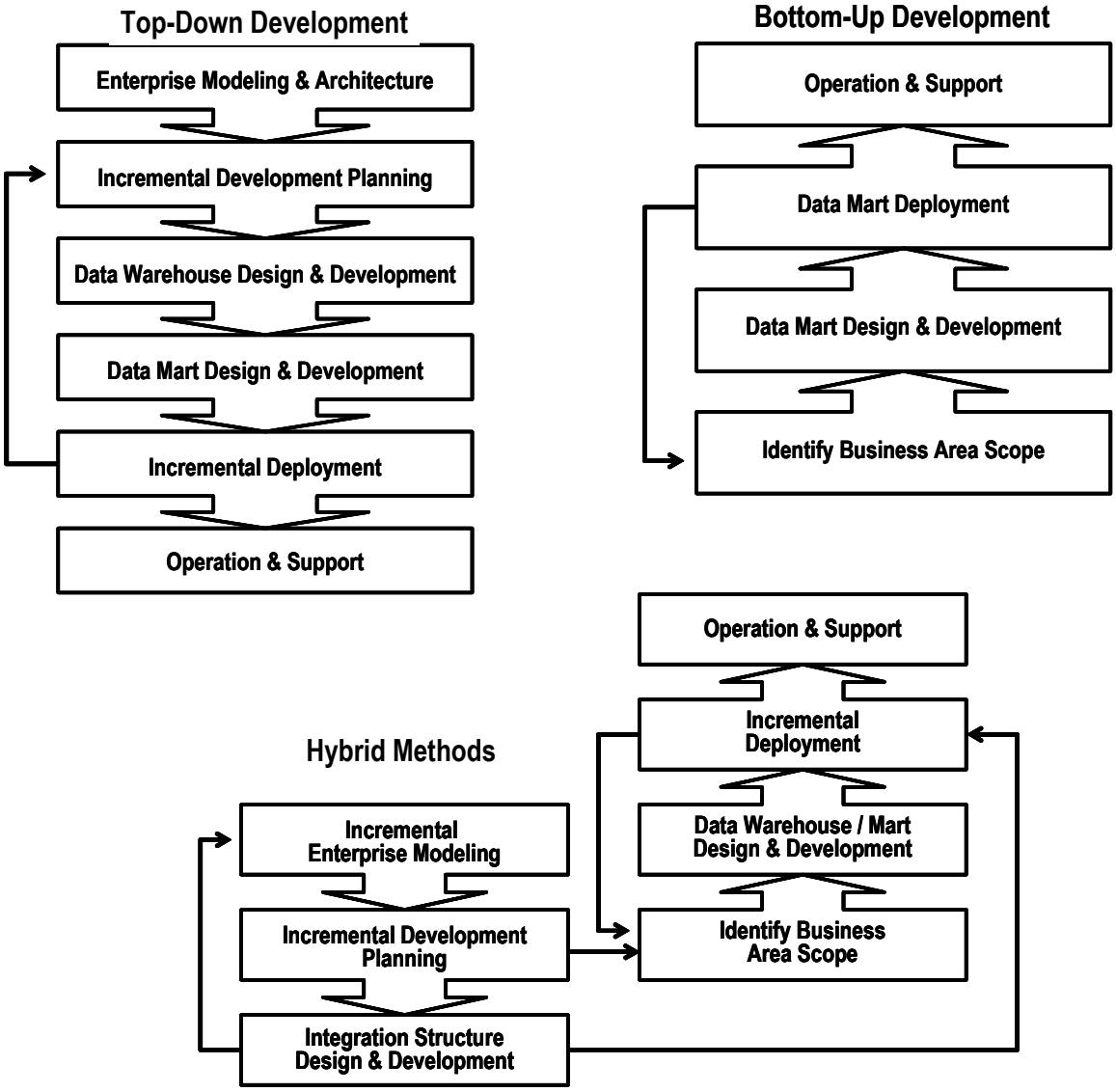
The centralized hub structure supports all five architecture functions, ie intake, integration, distribution, delivery and access through a single data structure. The challenge is to achieve an optimized level of performance across all five architectural functions. Intake and integration tend to work well in a normalized environment, while distribution, delivery and access tend to work better when the data structures are de-normalized. The challenge of the data warehouse architect is to achieve optimum performance across all functions with a single data structure.

The major motivations for implementing the centralized hub model are to achieve an integrated, subject oriented view of business data and to remain somewhat flexible and adapt to changing future business requirements.

This structure is attributed to the original work in data warehousing conducted by Bill Inmon.

Project Architecture Models

Methodology Structure



Project Architecture Models

Methodology Structure

TOP-DOWN

Top-down approaches are also commonly called enterprise approaches. Top-down data warehouse development begins at the enterprise, and typically emphasizes the data warehouse as a primary integrated information resource. Data warehouse structure is determined through enterprise modeling. Content is determined by a combination of business information needs and available source data. Top-down approaches are generally associated with longer start-up times due to the need for enterprise perspective.

BOTTOM-UP

Bottom-up approaches begin with business information needs for a single business unit or limited business domain. Bottom-up methods are most compatible with bus integration approaches, using conformity instead of an enterprise repository to achieve integration. Bottom-up development generally trades the strength of an integration hub for the benefits of quick start-up and rapid deployment.

BALANCING ENTERPRISE & BUSINESS UNIT FOCUS

Hybrid approaches combine some elements of bottom-up development with some elements from top-down methods. The objective of a hybrid approach is rapid development within an enterprise context. A typical hybrid approach quickly develops a skeletal enterprise model before beginning iterative development of data marts. The data warehouse is populated only as data is needed by data marts, and is sometimes constructed in a retrofit mode after data marts have been deployed. Metadata consistency and conformed dimensions are the initial integration tools, with the data warehouse being a secondary means of integration.

Module 5

Published Data Warehousing Approaches

This page intentionally left blank.

Corporate Information Factory

Introduction

The Corporate Information Factory (CIF)

Basic Principles

Program Management

Internal and External Data Sources

Enterprise Data Integration of Data in the DW and ODS

Delivery of Business Data through Data Marts

Meta Data Management throughout the Environment.

Corporate Information Factory

Introduction

OVERVIEW

The Corporate Information Factory (CIF) is a conceptual or logical architecture whose purpose is to deliver business intelligence and business management capabilities driven by data provided by business operations. The architecture describes and categorizes the information stores and processes used to operate and manage a robust business intelligence environment. The CIF components include producers of data and consumers of information.

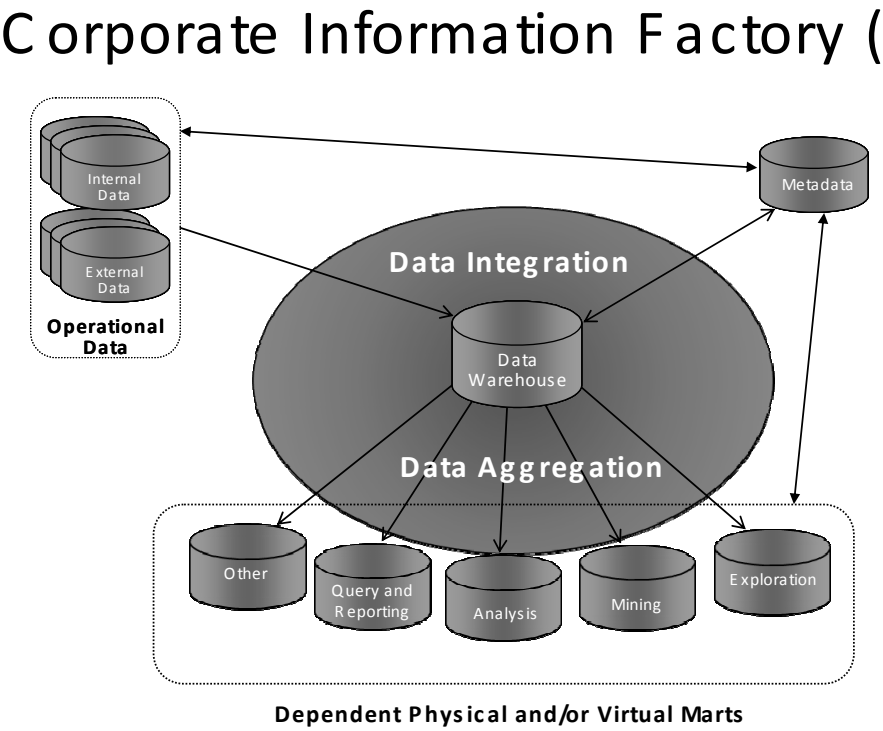
BASIC PRINCIPLES

The CIF continues to evolve and is based on the following basic principles.

- The BI environment is Program oriented and is managed using Program Management principles
- Data is sourced from internal operational systems and from external locations
- Data is stored as an enterprise view within the data warehouse and the operational data store
- Delivery of data to the business community is through a series of data marts that are tailored to the business need or directly from the operational data store
- Meta data management is included throughout the environment.

Corporate Information Factory

Data Architecture



Corporate Information Factory

Data Architecture

DATA STORES

The flow of data is predominantly from operational systems, both internal and external, where detailed data is extracted, transformed, cleansed, and formatted into the corporate standard for data analysis.

This integrated and consolidated granular data is loaded into the Data Warehouse or Operational Data Store (ODS) for more timely analytics. A staging area may optionally be implemented. It may be temporary or persistent depending on the complexity of the data integration processes. The data warehouse serves as the "hub" that acts as the data source for the dependent data marts or "spokes".

An advantage of the CIF originates from the design of the data warehouse. It is a somewhat de-normalized relational design and it has the ability to support multiple forms of analyses such as OLAP, statistics, ad hoc reporting, fixed reporting and data mining. It does not dictate what data structures should be used in the data marts. Data marts can be physically distinct from the data warehouse due to performance or technical reasons or they can be virtual structures such as universe layers or logical data base views of the data warehouse structure, set up to serve a particular analytic purpose.

METADATA

Metadata is captured from the operational systems and the data integration / data aggregation processes. Other metadata may be captured from or used by the data warehouse, ODS and data marts.

DATA INTEGRATION

Data Integration processes are typically performed by data consolidation capabilities of ETL technologies with assistance from data quality tools. Data Aggregation processes can be performed by the data consolidation and/or by the data federation capabilities of EII technologies. The results from the downstream analytics processes may be replicated back into the operational systems to support closed loop processing by using either data consolidation or data propagation capabilities of EAI technologies