

Data Warehouse

Martin Clement

Head of Analytics

Martin.Clement@atvantage.com

ATVANTAGE



Data Warehouse Architecture & Organization

ATVANTAGE



TIMETOACT
GROUP

Session Topics

- **Data Warehousing Architecture and best practices**
 - ▶ Key Design Considerations
 - ▶ Data Architecture, Data Integration Architecture, Application Architecture
 - ▶ Hardware and Software Infrastructure for a data warehouse
- **Business value to build a DW and BI solution**
 - ▶ BI Maturity model
 - ▶ Aligning DW and BI to strategic business goals
 - ▶ Key industry business drivers for Data Warehousing
- **Data Governance**
 - ▶ Metadata Management, Performance Management, Capacity Planning
- **Data Warehouse Lifecycle Management**
 - ▶ BI Program Management

A general definition of architecture

Architecture: “*The architecture of a system describes its overall static structure and dynamic behavior. It models the system’s elements (which for IT systems are software, hardware, and its human users), the externally manifested properties of those elements, and the static and dynamic relationships among them.*”

Architecture can be at any scale – and the scale is defined by the “characteristic dimension” of the architecture’s elements (4 of 4)

- Paraphrasing the definition: “**Structure and behavior of the elements of a system**”

- ▶ What is a system?
 - ▶ What is an element?



- It's all a question of **scale**:

- ▶ The elements of a hi-fi are its tuner, amplifier, speakers, cables, and remote controls.
 - ▶ The elements of an amplifier are its case, the power supply, and the circuit board.
 - ▶ The elements of a circuit board are its resistors, fuses, and integrated circuits.
 - ▶ The elements of an integrated circuit are transistors and other components.



- Whatever the scale of the system we're focused on, we say:

- ▶ Above that scale are the **requirements** of the world in which the system sits
 - ▶ At that scale is the **architecture** between the elements of the system
 - ▶ Below that scale is the **design** of the “insides” of the elements of the system



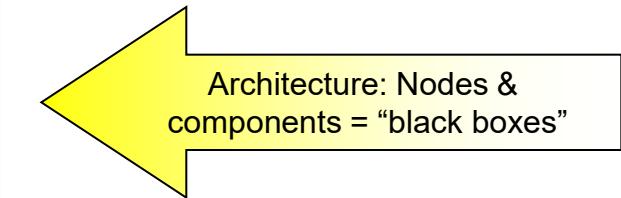
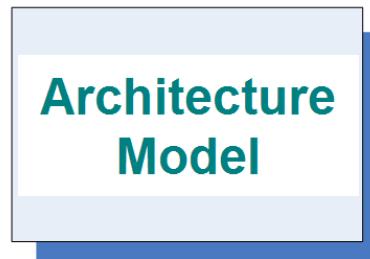
- So, in our world of IT:

- ▶ Our IT system's requirements are modeled as interactions (use cases, etc.) with actors and other IT systems
 - ▶ Our IT system's architecture's elements are components and nodes, etc.
 - ▶ Our IT system's design or designs model the insides of components, nodes, messages, and connections

Overall, architecture and design both focus on the content and the context of the target system, ensuring that it *does the right things* and *does things the right way*

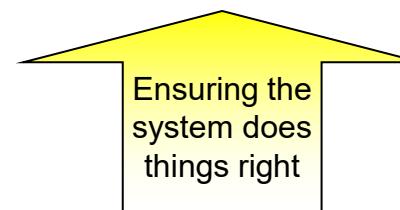
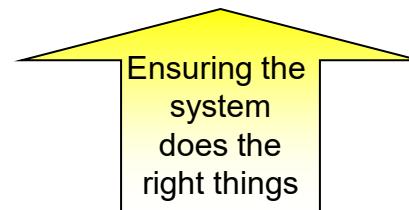
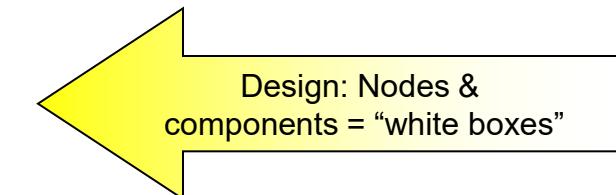
Architecture

Creating the Structure
(Architects)



Design

Detailing the Content
(Designers)



Main components of a data warehouse solution

Methodology and Best Practices

Application Architecture
(Reports, Queries, Dashboards, ...)

Data Architecture
(Data Models, DB)

Data Integration Architecture
(Data capture/movement, Cleansing, Transformations ..)

Hardware and Software Architecture
(Servers, OS, Network, Databases, App. Servers, BI Tools, ETL Tools ...)

Governance
(Metadata, Performance Mgt, ..)

Overall Data Warehousing Design Considerations /1

- Plan for consolidation view of the data (aka “**Single source of truth**”)
- **Structured and unstructured** data
 - ▶ Should enable to search all data
 - ▶ Uncover all insights about customers, products, organization, etc
- Plan the solution to deliver **real time information**
 - ▶ At least have it as a **roadmap** when architecting the solution
 - ▶ In the future you **might need to be more dynamic** to support business decisions
- Start with the **right skills and organization** in place
 - ▶ DW requires **discipline to implement** (stakeholders ...)
- **Re-use of assets** (data, process, etc) as a standard and **consider IT standards** of the company

Overall Data Warehousing Design Considerations /2

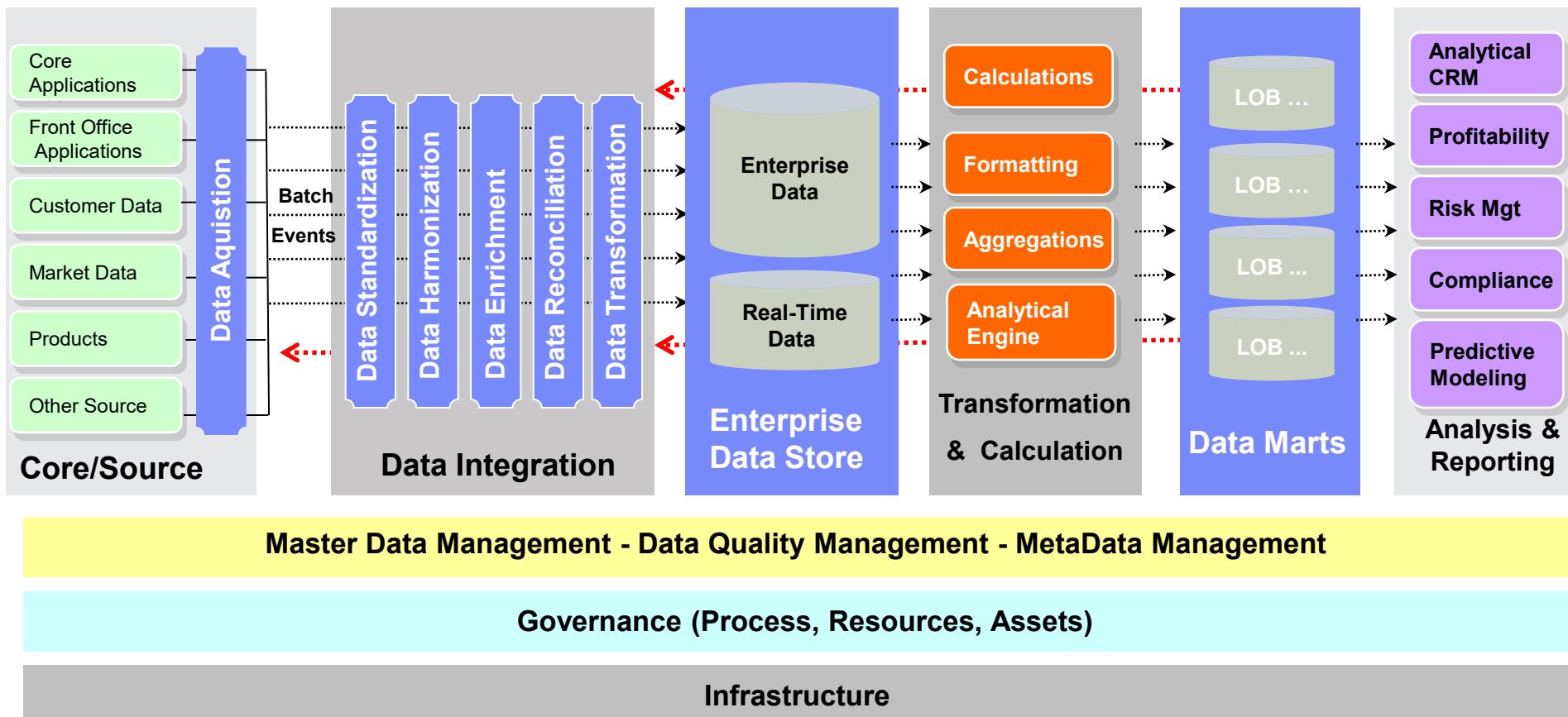
- **Modular growth with predictable cost**
 - ▶ Predictable costs based on business growth
 - ▶ Predictable performance for a predictable cost
- Choose a solution that **provides low adoption risk** (design)
- Seamless Scalability (“**start small, think big**”)
 - ▶ Linear scalability
 - ▶ Allow growth at the right time with minimum impacts and interruptions
- Reliability and integrity
 - ▶ Ensure **business continuity**
 - ▶ **Information integrity**

Overall Data Warehousing Design Considerations /3

- **Governance** (process, IT, systems management)
 - ▶ Resources consumption and utilization
 - ▶ Performance management
 - ▶ Workload Management
 - Guaranty Service Level Agreements (SLAs) and availability of resources due to business priorities
 - ▶ Traceability (data, process)
- **Security, privacy and data protection**
 - ▶ Ensure the security and privacy of data, information, systems and people with the right policies, methods, tools and overall governance

Data Integration Architecture

Data-Driven Reference Architecture



Data Integration and Data Latency

- Data Latency:

One of the most important drivers that affect design of the entire DW
According to Dick Hackathorn, data latency has three components

- **Data preparation latency:**

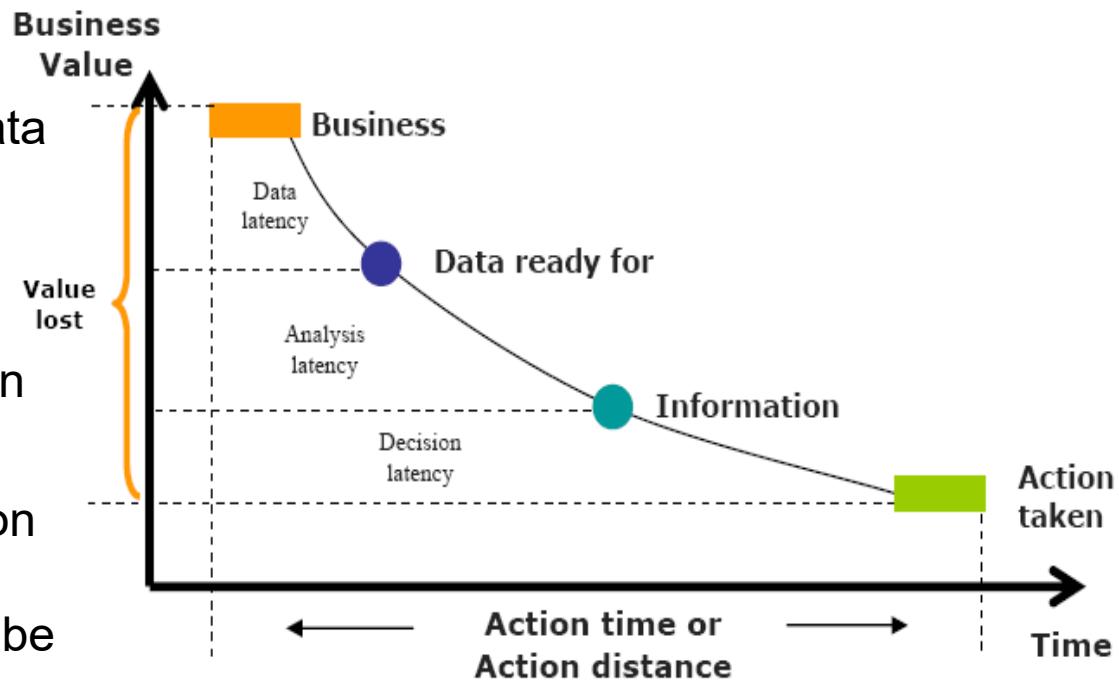
The time it takes to get the data ready for analysis

- **Analysis latency:**

The time it takes to get the results of an analytic operation

- **Decision latency:**

The time it takes for the person receiving the results to understand what action must be taken.



Data Integration Architecture – Other Design Considerations

- **Scalability** and **Parallel** processing
 - ▶ Process large volume of data and transactions
 - ▶ Support execution of concurrent process
- **Flexibility**
 - ▶ Easy to maintain and adapt to new business demands
- **Optimization** of the processes
 - ▶ Avoid redundancy of process and maximize computer resources utilization

Data Architecture Design Considerations

- Includes **Enterprise Data Warehouse** and **Data Marts** (logical view)
- Needs to promote a **consolidated view** of the business - Avoid data redundancy
 - ▶ Consolidation of all business process - “**Single version of the truth**”
- “*Golden Copy of Data*”
 - ▶ Guarantee **single source** of information to be used in the decision process by different LOBs
- **Optimization Layers**
 - ▶ Allow optimization to support specific business need
- **Data Synchronization**
 - ▶ Allow batch and real-time updates of data structures
 - ▶ Allow concurrency of process (read and write due to continuous data ingestion)

Application Architecture Design Considerations

- Use state-of-the-art **visualization technologies** to bring KPIs to the end users
- Use state-of-the-art **delivery methods** to bring information to the right place
 - ▶ Web access on **any device**
 - ▶ E.g. Delivery via eMail
- **Incorporate** analytical results into business processes
- Support **different user types** with different information according to **requirements** previously defined

HW and SW Architecture for DW – Considerations /1

- Typical **multiple components** solution
 - ▶ Components of the solution should be configured/defined to avoid bottlenecks
 - ▶ Performance and functional aspects of the solution should be defined based on **integrated tests**
- **Scalability and Parallel** processing
 - ▶ Process large volume of data and transactions
 - ▶ Support execution of concurrent process
 - ▶ Linear scalability
- **Modular growth**
 - ▶ Vertical and horizontal growth to support new business demand

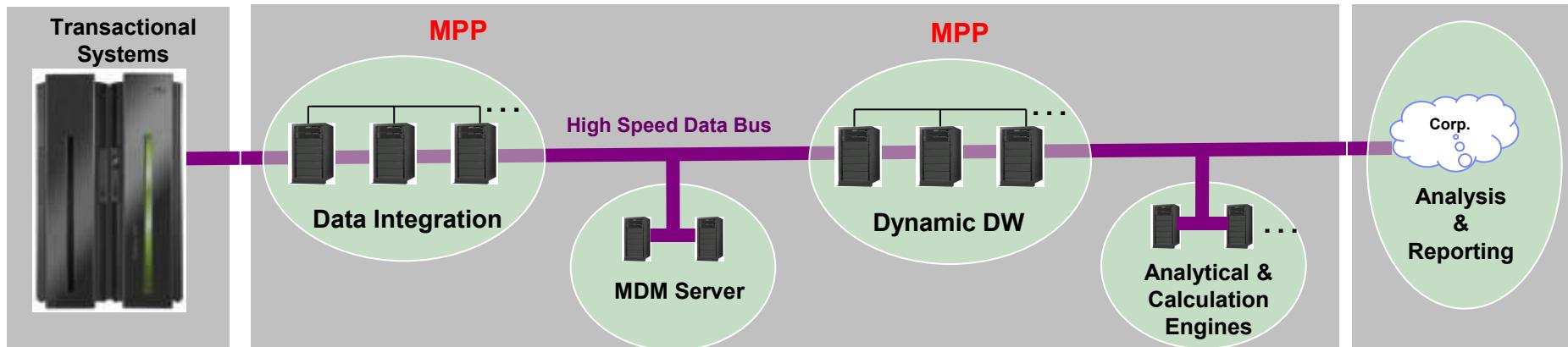
HW and SW Architecture for DW – Considerations /2

- Continuity of business operations
 - ▶ **Maintain business operations in the event** of an outage--with processes and infrastructures that are responsive, highly available and scalable
 - ▶ **Reliability** to ensure business continuity and resilience with **information integrity**
 - ▶ **High availability** to support mission critical applications
 - ▶ Process recovery
 - **Backup and Restore** due to processing failures/issues
 - ▶ Disaster and Recover
 - **Recover system** due to unexpected major issues into the infrastructure

SW and HW Architecture for Data Warehousing

- Data Integration Servers
- Data Repository Servers
 - ▶ DW + DM, MDM
- Analytical Servers
 - ▶ Cognos, MicroStrategy, SAS, etc
- Application Servers
 - ▶ Calculation engines (Fermat, ILOG, etc)
- Communication networks
 - ▶ Switches, etc
- Backup, Restore, Archiving

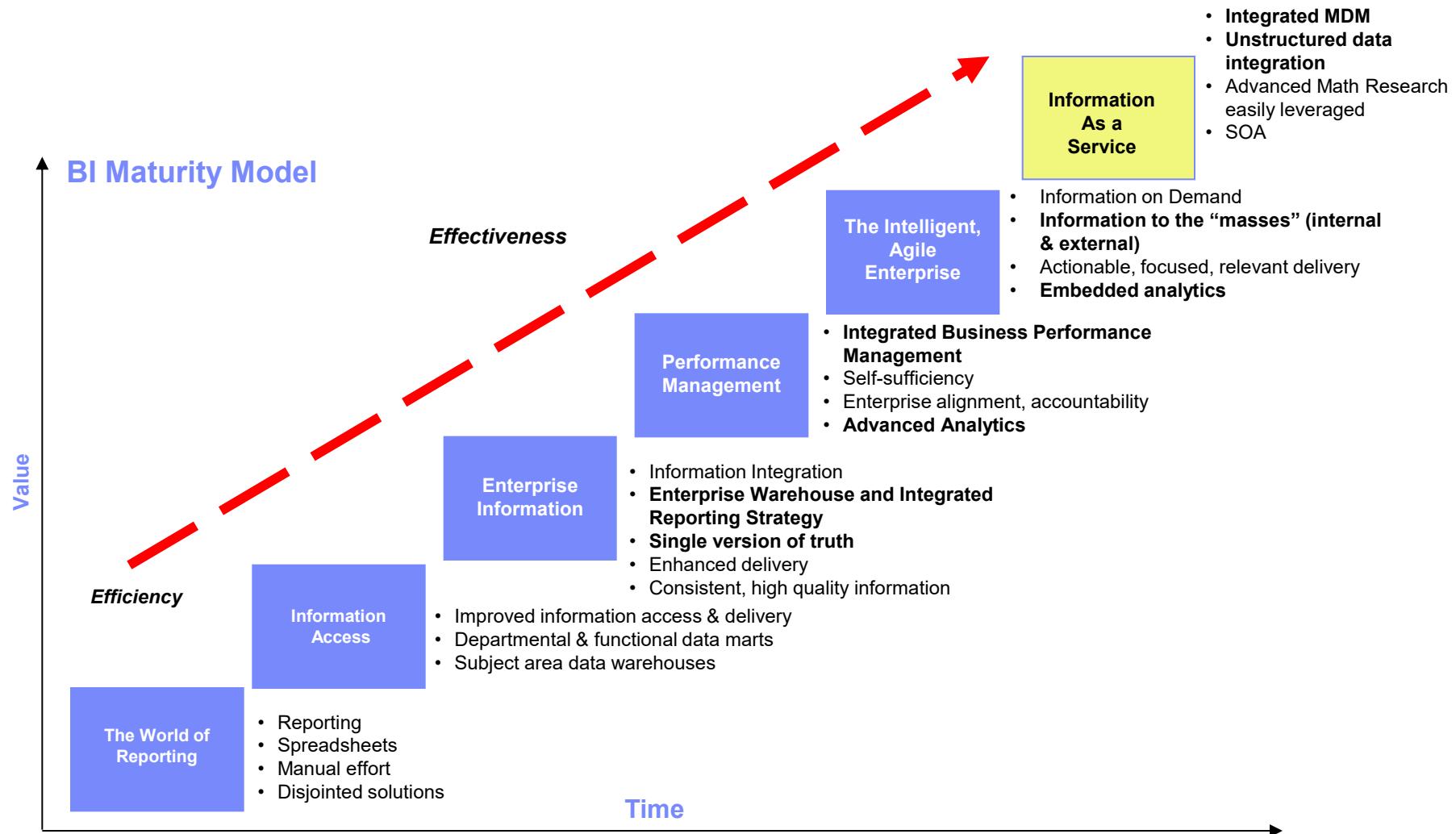
	OLTP	DW
Processing of large volumes of data	No	Yes
Predictable queries	Yes	No
Response time is a function of database size	No	Yes
Query Complexity	Simple	Simple to Complex
Aging of Data	Current	Current/Historical
Process type	Static	Dynamic



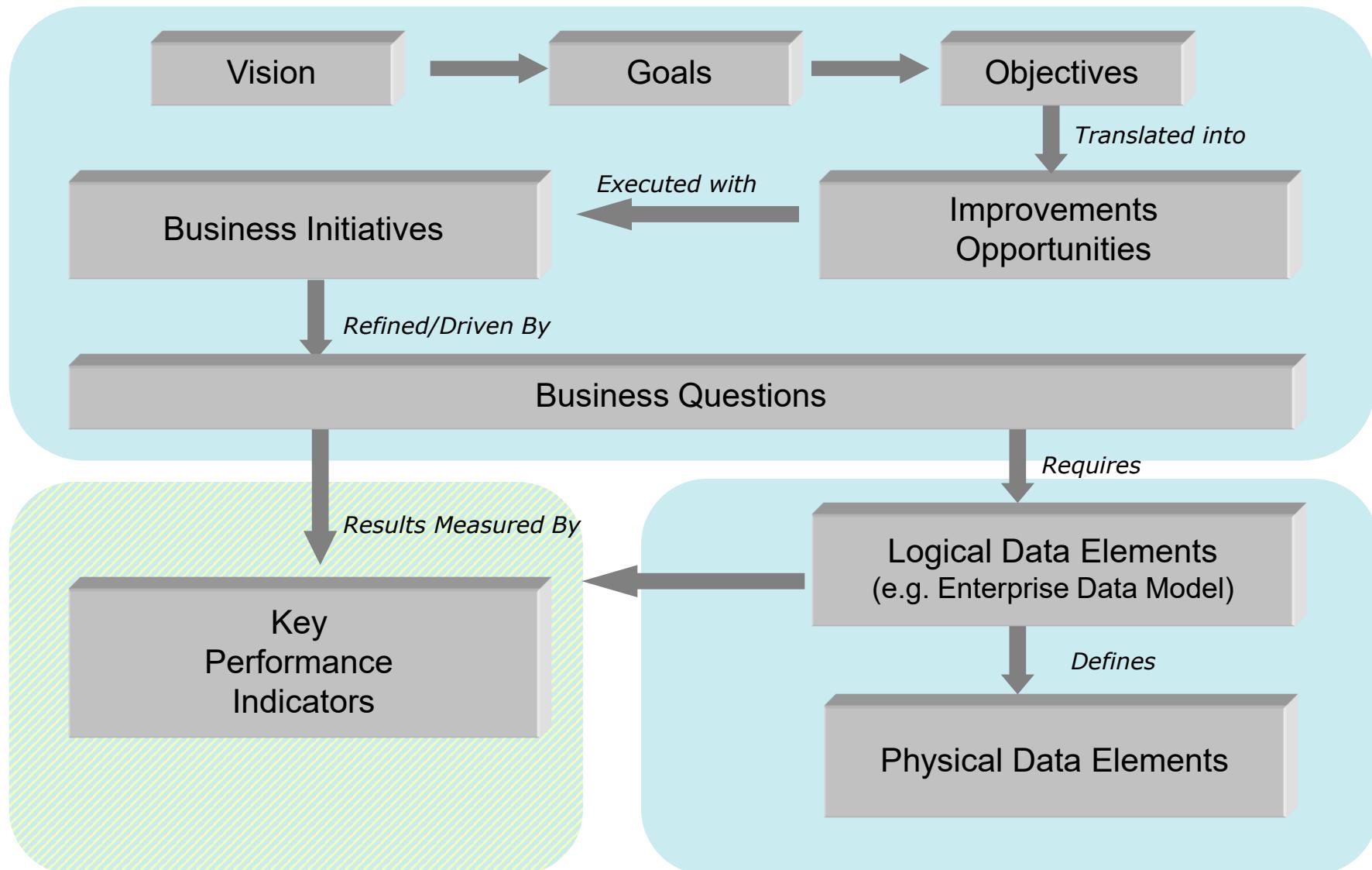
Agenda

- **Data Warehousing Architecture and best practices**
 - ▶ Key Design Considerations
 - ▶ Data Architecture, Data Integration Architecture, Application Architecture
 - ▶ Hardware and Software Infrastructure for a data warehouse
- **Business value to build a DW and BI solution**
 - ▶ BI Maturity model
 - ▶ Aligning DW and BI to strategic business goals
 - ▶ Key industry business drivers for Data Warehousing
- **Data Governance**
 - ▶ Metadata Management, Performance Management, Capacity Planning
- **Data Warehouse Lifecycle Management**
 - ▶ BI Program Management

The Business Intelligence maturity model

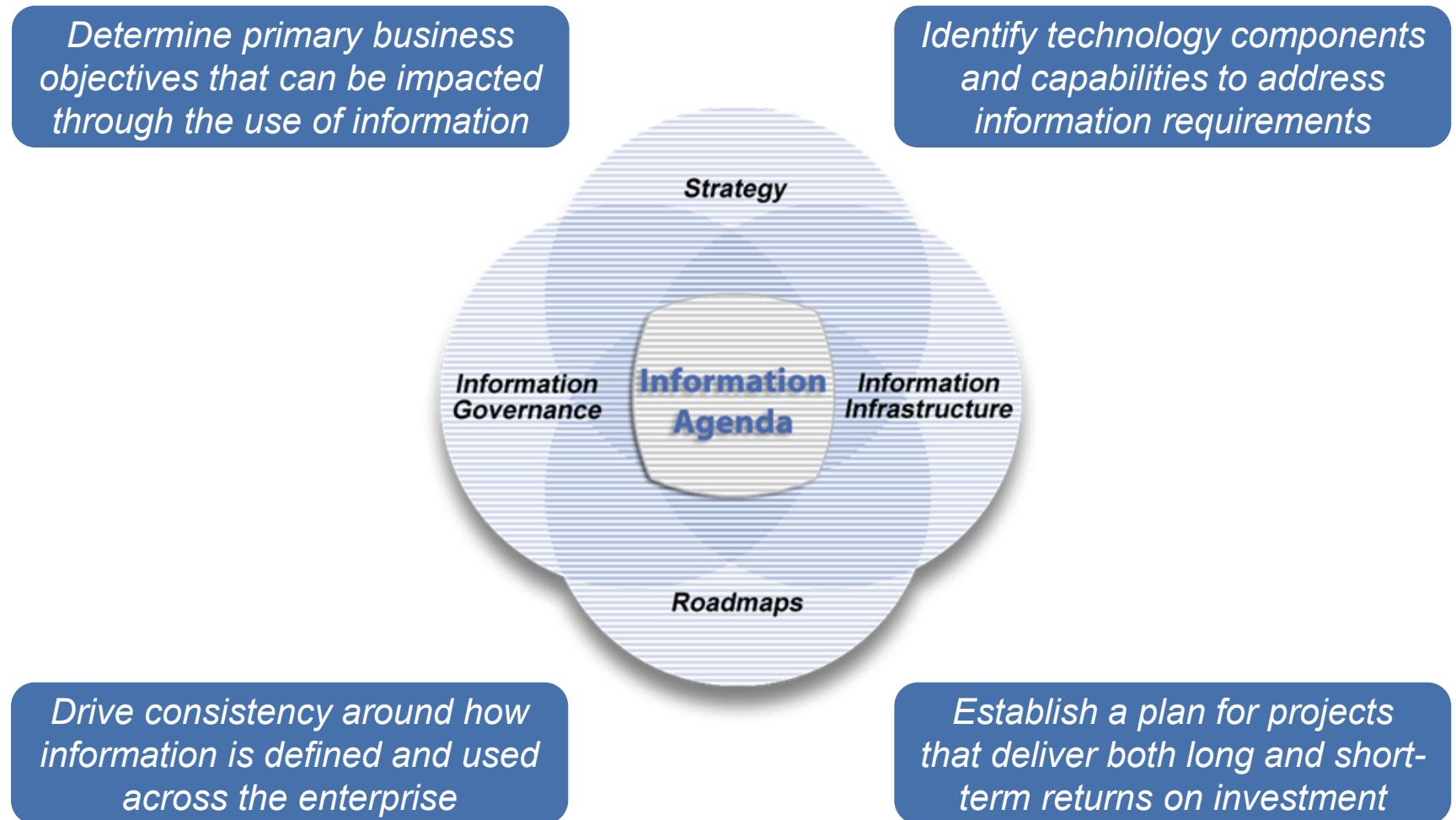


Aligning DW and BI to strategic business goals



Organizations Need an Information Agenda

An approach for unlocking the business value of information



Agenda

- **Data Warehousing Architecture and best practices**
 - ▶ Key Design Considerations
 - ▶ Data Architecture, Data Integration Architecture, Application Architecture
 - ▶ Hardware and Software Infrastructure for a data warehouse
- **Business value to build a DW and BI solution**
 - ▶ BI Maturity model
 - ▶ Aligning DW and BI to strategic business goals
 - ▶ Key industry business drivers for Data Warehousing
- **Data Governance**
 - ▶ Metadata Management, Performance Management, Capacity Planning
- **Data Warehouse Lifecycle Management**
 - ▶ BI Program Management

Data Warehouse Governance

- **Metadata** (“data about data”)

- ▶ **Technical metadata**

- Support documentation of processes and assets of the solution
 - Allow impact analysis of the environment - Critical for maintenance and improvements

- ▶ **Business metadata**

- Define business terms, calculations and formulas used in the decision process

- **Metadata** occurs on **every layer** of a data warehouse architecture

- ▶ Metadata should be captured in order to support

- **Impact Analysis:**
 - What else has to be changed if one objects changes?

- **Data Lineage:**
 - What is the information chain building up a single object?

Data Warehouse Governance Overview /2

▪ Resource Management

▶ Performance management

- Align I/T computing resources with business requirements

▶ Workload management

- Assign resources to high priority LOBs, prevent low priority work from taking resources
 - ETL, Queries, etc

▶ Capacity planning

- Provides historical information about utilization and growth of the environment

▶ Security, privacy and data protection

- Access control, Audit

Agenda

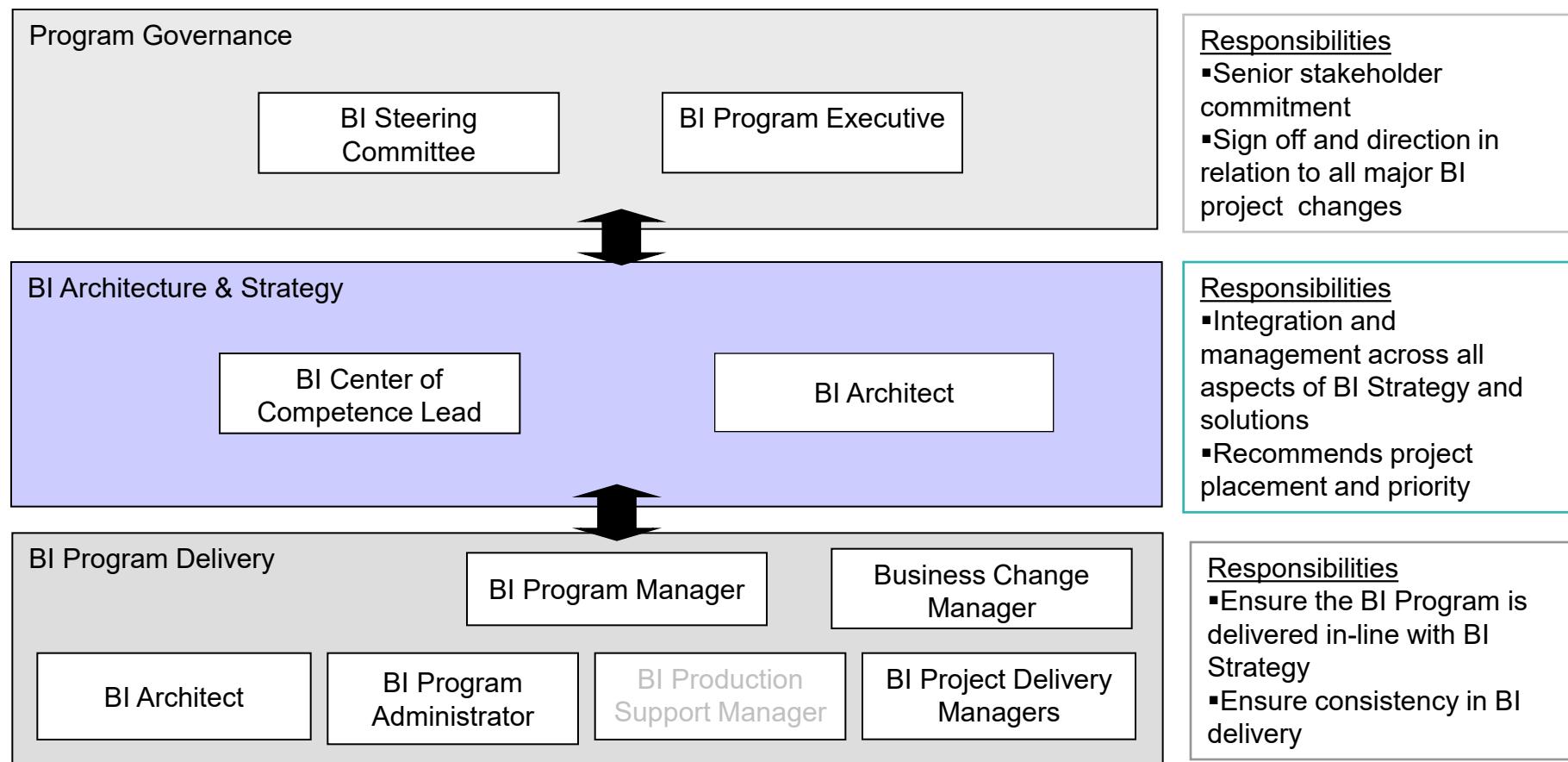
- **Data Warehousing Architecture and best practices**
 - ▶ Key Design Considerations
 - ▶ Data Architecture, Data Integration Architecture, Application Architecture
 - ▶ Hardware and Software Infrastructure for a data warehouse
- **Business value to build a DW and BI solution**
 - ▶ BI Maturity model
 - ▶ Aligning DW and BI to strategic business goals
 - ▶ Key industry business drivers for Data Warehousing
- **Data Governance**
 - ▶ Metadata Management, Performance Management, Capacity Planning
- **Data Warehouse Lifecycle Management**
 - ▶ BI Program Management

Purpose of a Governance Structure (BI Program Mgmt)

- A Governance Structure is defined and implemented to **manage and support the implementation of BI Solutions** that meet the business need for business intelligence. The governance structure also **addresses the ongoing management** of the BI solutions **after they are delivered**.
- The governance structure is developed **to ensure that all the key elements required to implement a successful business intelligence program are addressed**. These include, but are not limited to:
 - Business representation
 - Senior management commitment
 - Ongoing management and implementation of the BI Strategy
 - Data Integration and Management
 - Architecture blueprints and standards

BI Governance Structure

The diagram below illustrates a typical governance structure to successfully implement Business Intelligence solutions. The structure can be viewed in three distinctive but collaborating layers: Program Governance, BI Architecture & Strategy, and Program Delivery.



BI Program Governance – Group Charter

Program Governance Group

BI Steering Committee

BI Program Executive

Purpose

To ensure that all solutions requested by the business are channelled into the optimal delivery portfolio using appropriate standards to enrich Westfield's information assets. Ensure that the strategic alignment of the Business Intelligence Strategy is sustained and that appropriate levels of resources are provided to support the BI strategy and implementation.

Responsibility Statement

- Review recommendations for carrying out the BI Program, approving appropriate actions
- Ensure the BI Program remains aligned with the business strategy
- Approve the resources and executive commitment required to deliver valuable BI solutions
- Review and approve all major systems and technology investments and changes supporting BI

BI Steering Committee

- Responsibilities:
 - ▶ Review **program strategies** and provide **executive guidance**
 - ▶ Provide direction to program vision and leadership
 - ▶ Review **Program/Project status** and priorities
 - ▶ Recommend **project priorities** with the IT Portfolio Executive
 - ▶ Review and **approve** all major **systems and technology investments** and business changes supporting BI
 - ▶ **Approve the resources** needed to enable the strategy
 - ▶ **Confirm** that the **BI Strategy** is aligned with **business strategy**
 - ▶ **Hold business executives accountable** for achieving benefits
 - ▶ **Facilitate sustaining support** across business units

BI Program Executive

- Responsibilities:
 - ▶ Define the **BI strategy and vision**, and maintain alignment with the business strategy
 - ▶ **Communicate and collaborate** with the program's primary stakeholders
 - ▶ Team with the Program Manager to recommend projects and resource needs to the BI Program Steering Committee
 - ▶ **Secure the funding** required to run the program
 - ▶ **Win business line agreement** to implement **change management** initiatives to enable improved results using Business Intelligence capabilities
 - ▶ **Provide the executive leadership** required to plan, staff, and sustain a successful BI Program

BI Architecture & Strategy – Group Charter

BI Architecture & Strategy Group

BI Center of
Competence Lead

BI Architect

Purpose

To ensure that all solutions requested by the business are channelled into the optimal delivery portfolio using appropriate standards. Ensure that the strategic direction of Business Intelligence on the project is maintained and that appropriate levels of technology resources are provided to support the BI strategy and implementation.

Responsibility Statement

- Address all requested BI Projects
- Recommend how business requests will be met through either existing capabilities or new BI solutions;
- Provide these recommendations to the BI Executive, BI Program Manager and the IT Portfolio Process
- Ensure that BI project activities are in line with the BI Strategy
- Ensure standards for data, systems architecture, metadata and security are aligned with the BI Program
- Providing support and direction to the BI Program Delivery Team

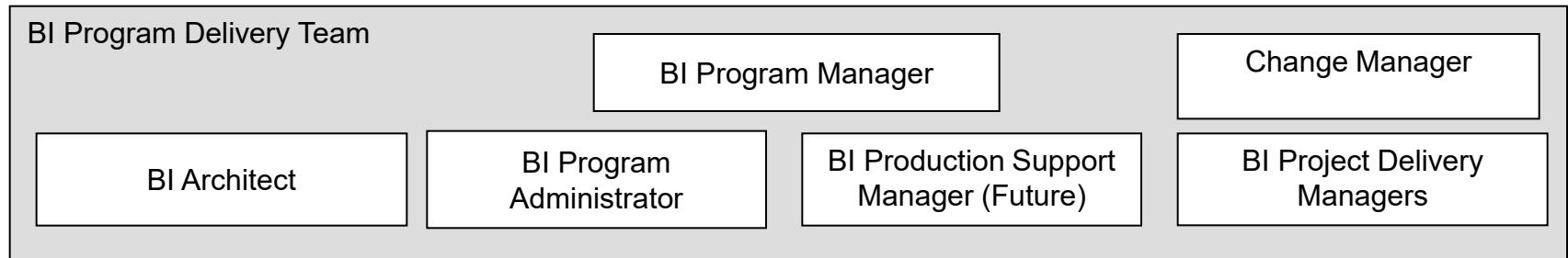
BI Center of Competence Lead

- Responsibilities:
 - ▶ Periodically assess the BI CoC competency (skills) inventory relative to current and planned BI projects to ensure proper alignment
 - ▶ **Ensure all CoC members understand and can apply the BI Method** for delivering BI projects utilizing appropriate work products and deliverables
 - ▶ **Create and maintain target BI Architecture**
 - ▶ Perform bi-annual skills assessments for each BI CoC member and provide feedback
 - ▶ Approve internal and external training and education in support of approved PDF goals

BI Architect

- Responsibilities:
 - ▶ **Responsible for all architectural alignment tasks** during the lifecycle of a project and that they align with **Information Standards and Enterprise Models** in the Phase-end QA Checkpoints
 - ▶ Supports the project manager, in determining the skill sets and level of effort required to design and construct a particular program, project, and/or applications to meet the business requirements
 - ▶ **Provides project-level data architecture planning and design** expertise on development projects
 - ▶ **Responsible for selecting the correct architecture** to fulfil those requirements (e.g. MOLAP, ROLAP, standard reporting)
 - ▶ **Keeps up to date on BI related advancements in technology arena** to suggest updates and enhancements to the BI architecture as appropriate.

BI Program Delivery – Group Charter



Purpose

To ensure that all solutions that are developed and managed by the BI Program, are delivered on time, within scope and budget in accordance with the BI Strategy and Roadmap.

This team will manage:

- Program delivery processes
- Project interdependencies
- Key decisions – business, technical and strategic
- Stakeholder and communication management
- Issues and Risk Management

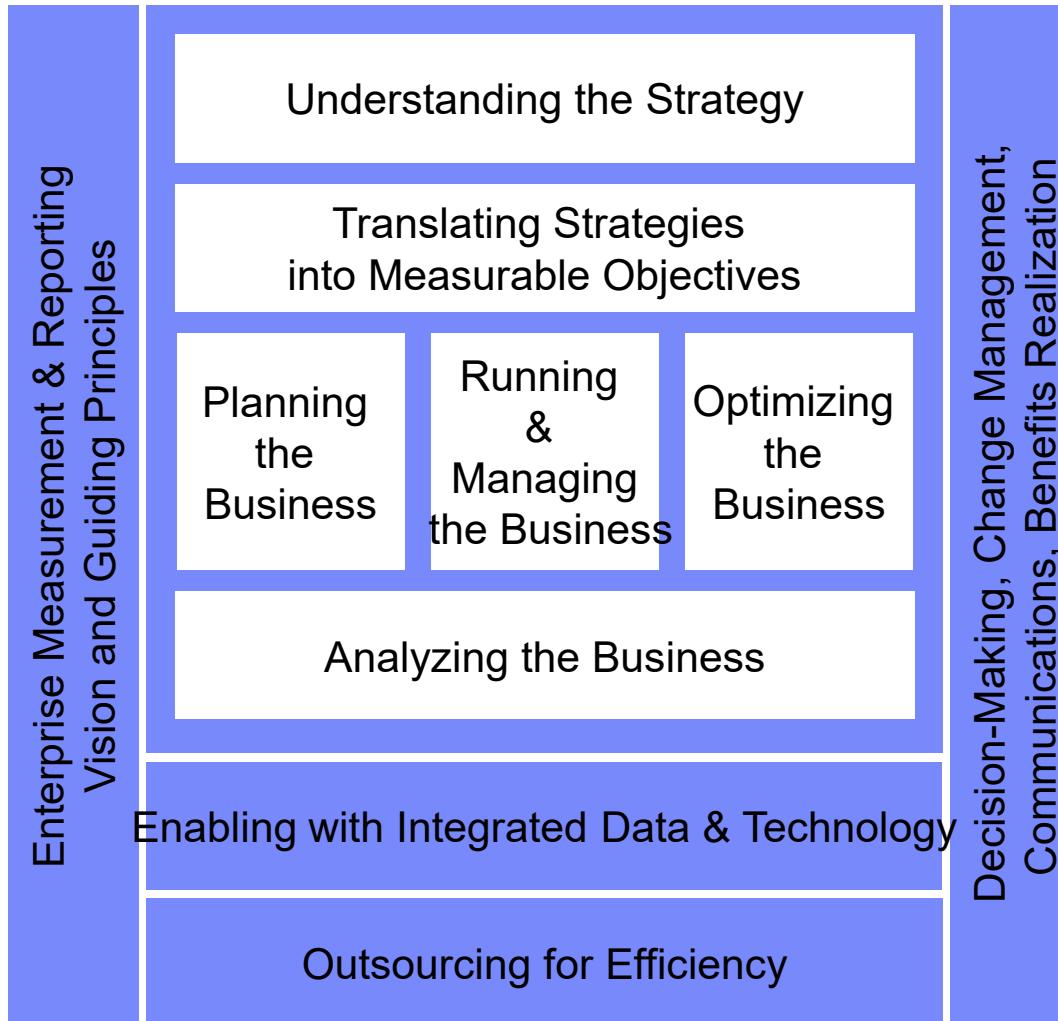
| Data Warehouse, 2025

Successfully building a Data Warehouse



The IBM Business Intelligence Framework

A strategic approach to information is imperative

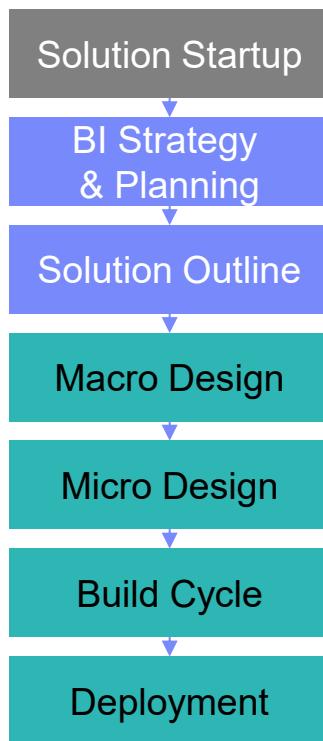
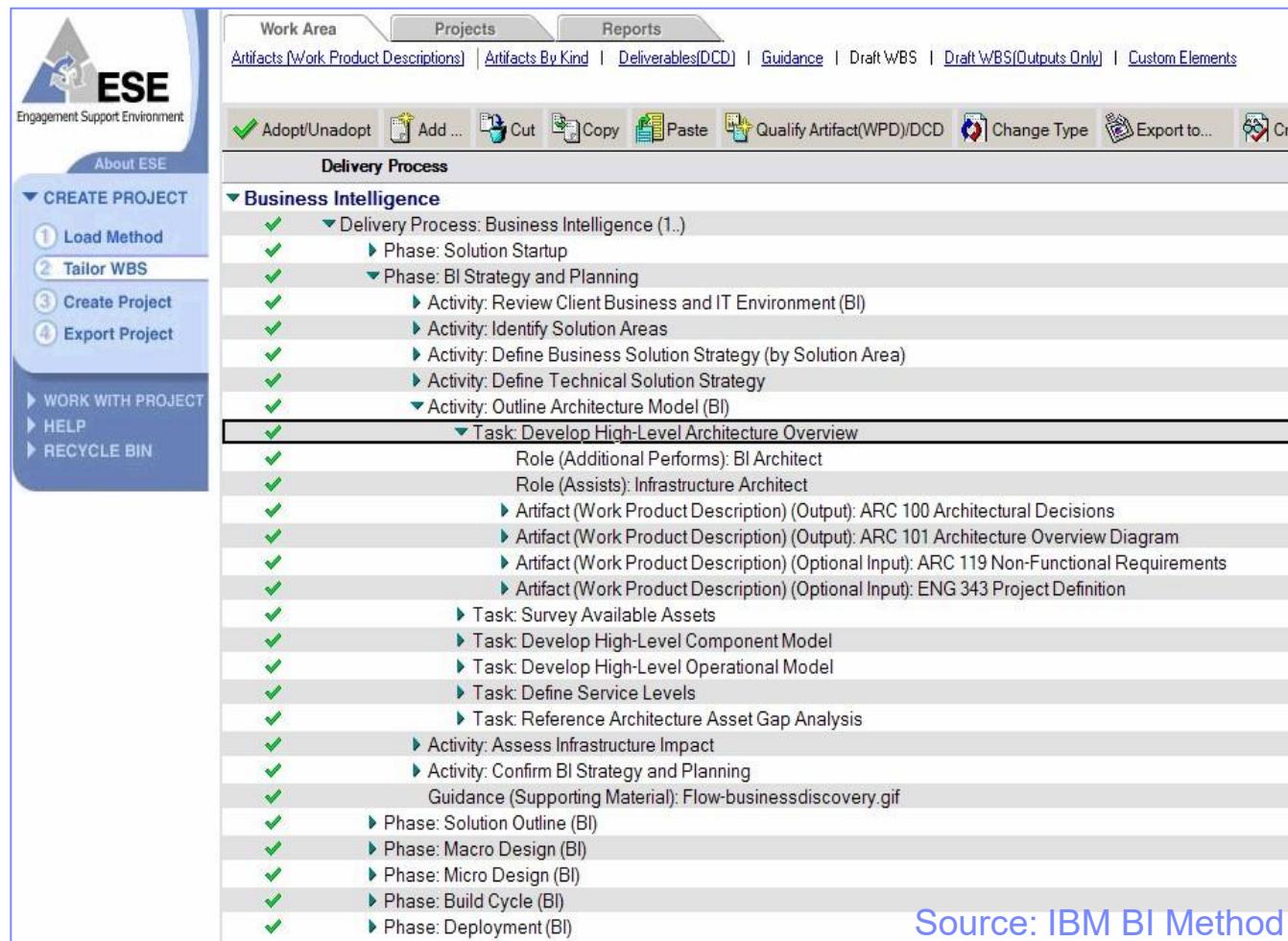


Example:

The IBM Business Intelligence framework – a component of the BI method

—
provides a common framework to describe our client's BI needs, solutions and approaches. It also includes the vision, roadmap and governance for designing a fully integrated business intelligence environment, and approaches for applying enabling technologies.

The IBM Business Intelligence Method - A proven method to deliver Business Intelligence Solutions

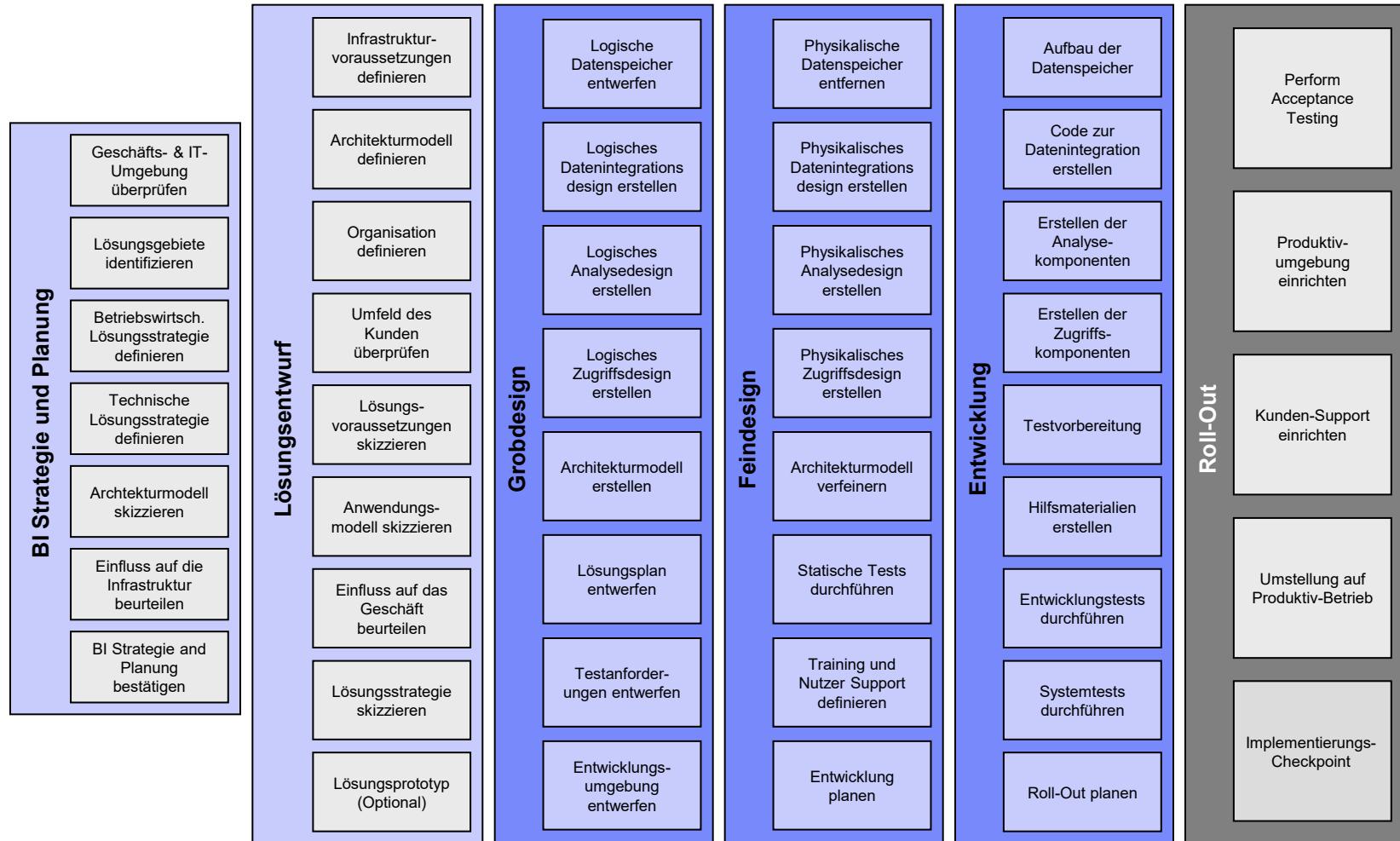
The screenshot shows the ESE interface with the following navigation tabs: Work Area, Projects, and Reports. Under Work Area, there are links for Artifacts (Work Product Descriptions), Artifacts By Kind, Deliverables(DCD), Guidance, Draft WBS, Draft WBS(Outputs Only), and Custom Elements. Below these are toolbars for Adopt/Unadopt, Add ..., Cut, Copy, Paste, Qualify Artifact(WPD)/DCD, Change Type, Export to..., and Create.

The main content area displays the "Delivery Process" for "Business Intelligence". The process starts with "Delivery Process: Business Intelligence (1..)" which includes "Phase: Solution Startup", "Phase: BI Strategy and Planning", and "Activity: Review Client Business and IT Environment (BI)". It then branches into "Activity: Identify Solution Areas", "Activity: Define Business Solution Strategy (by Solution Area)", "Activity: Define Technical Solution Strategy", and "Activity: Outline Architecture Model (BI)". A specific task, "Task Develop High-Level Architecture Overview", is highlighted with a black border. This task involves roles like BI Architect and Infrastructure Architect, and artifacts such as ARC 100 Architectural Decisions, ARC 101 Architecture Overview Diagram, ARC 119 Non-Functional Requirements, and ENG 343 Project Definition. Other tasks listed include Survey Available Assets, Develop High-Level Component Model, Develop High-Level Operational Model, Define Service Levels, Reference Architecture Asset Gap Analysis, Assess Infrastructure Impact, Confirm BI Strategy and Planning, and Flow-businessdiscovery.gif.

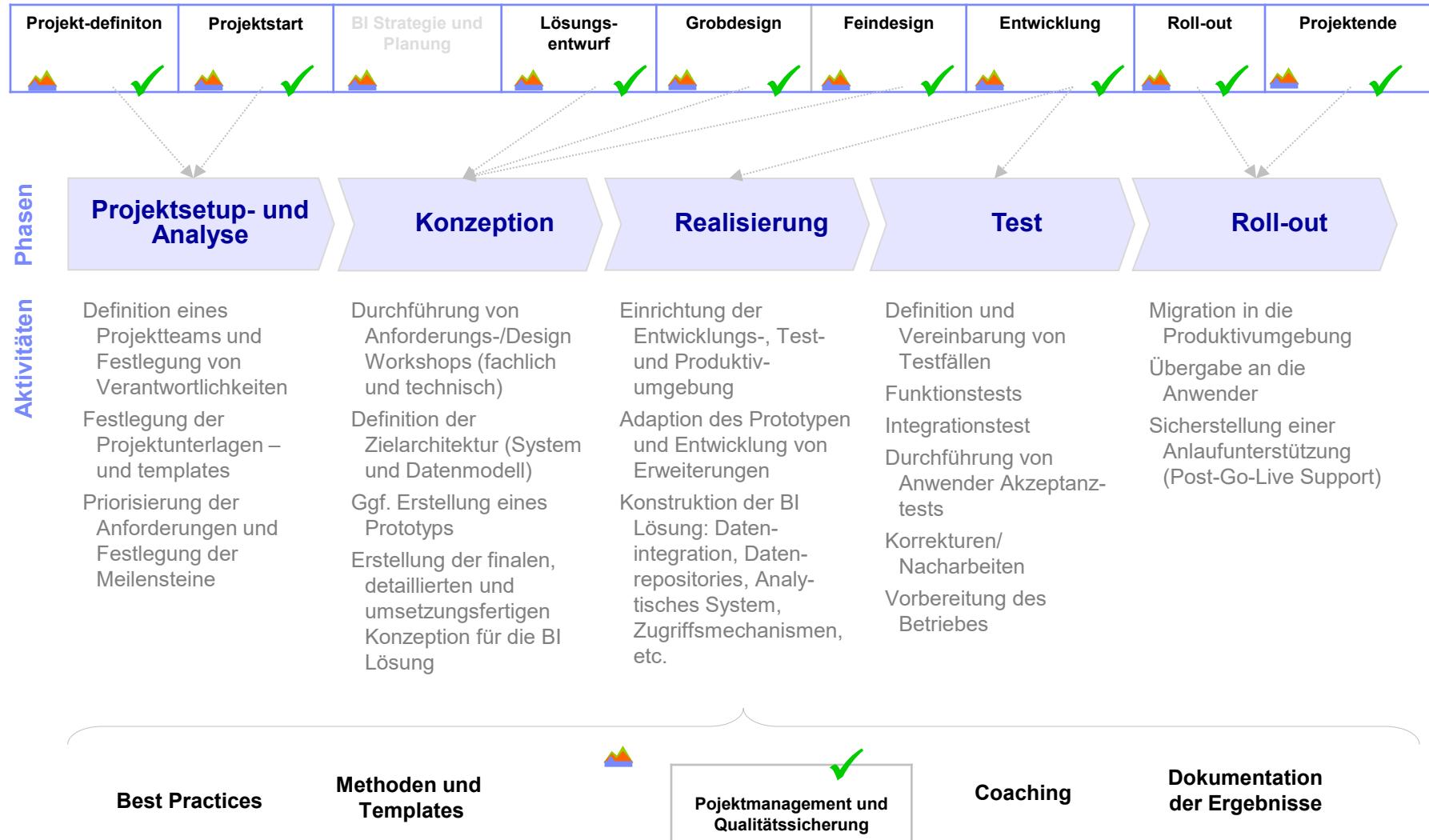
The process continues with "Phase: Solution Outline (BI)", "Phase: Macro Design (BI)", "Phase: Micro Design (BI)", "Phase: Build Cycle (BI)", and "Phase: Deployment (BI)".

Source: IBM BI Method

The IBM BI Method Project Approach



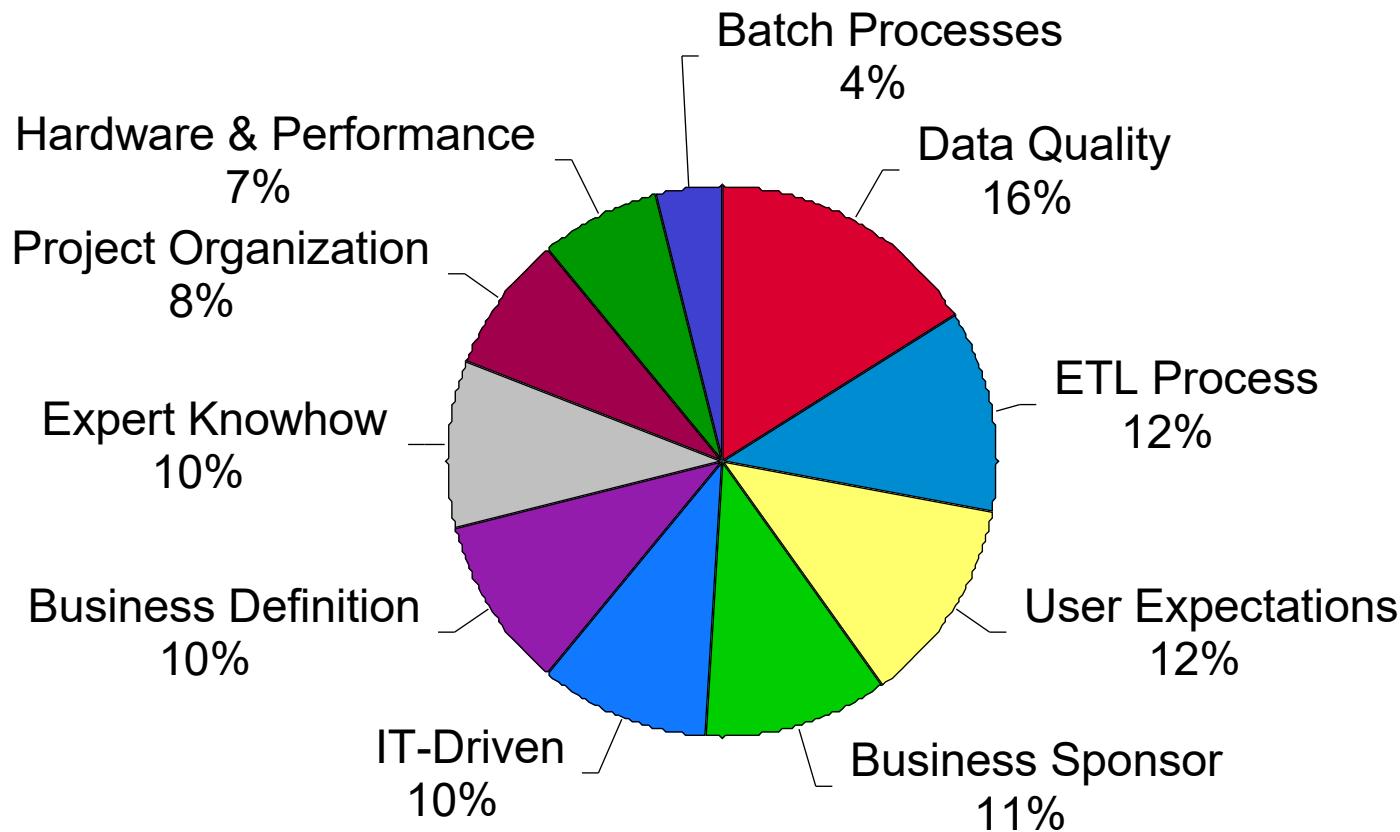
Sample Taylored Project Approach



Critical success factors for building a data warehouse

- Answer most important questions of participating business units
- Provide high-quality data
- Introduction in time
- Usage of modern technology
- Assure high return on investment
- Business orientation
- Easy to use
- Executive sponsor
- Patience – user acceptance evolves over time
- “Quick wins”

Why do DW projects fail?



What can we do?

- Align DWH initiative to business strategies
- Follow Design Considerations
- Use Methods, Reference Architectures and Data Models
- Follow an iterative approach („Think Big, Start Small“)
- Establish a Governance Structure (Project Organization, Programm Management)