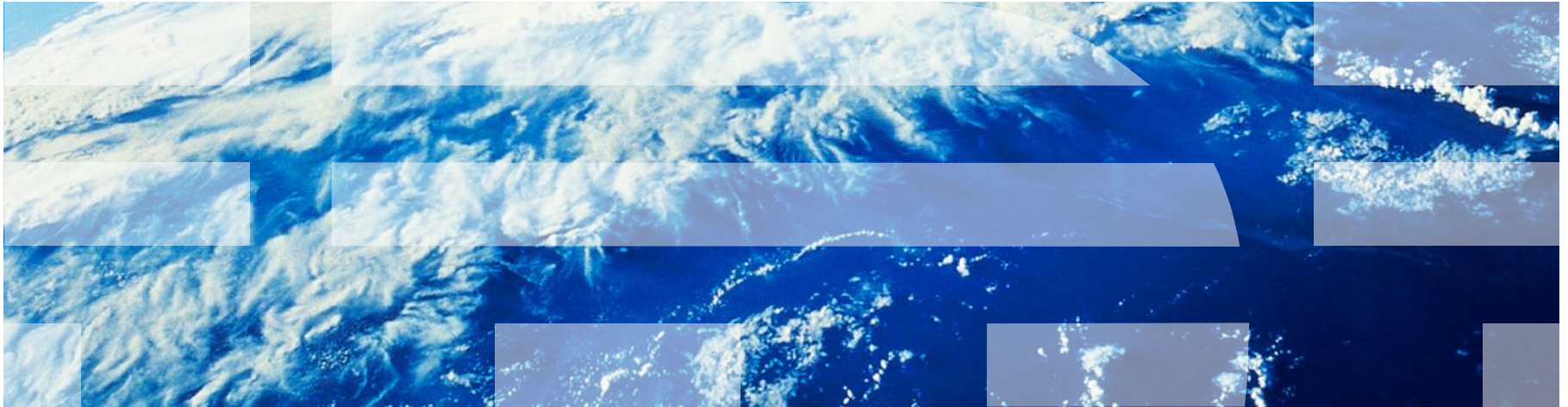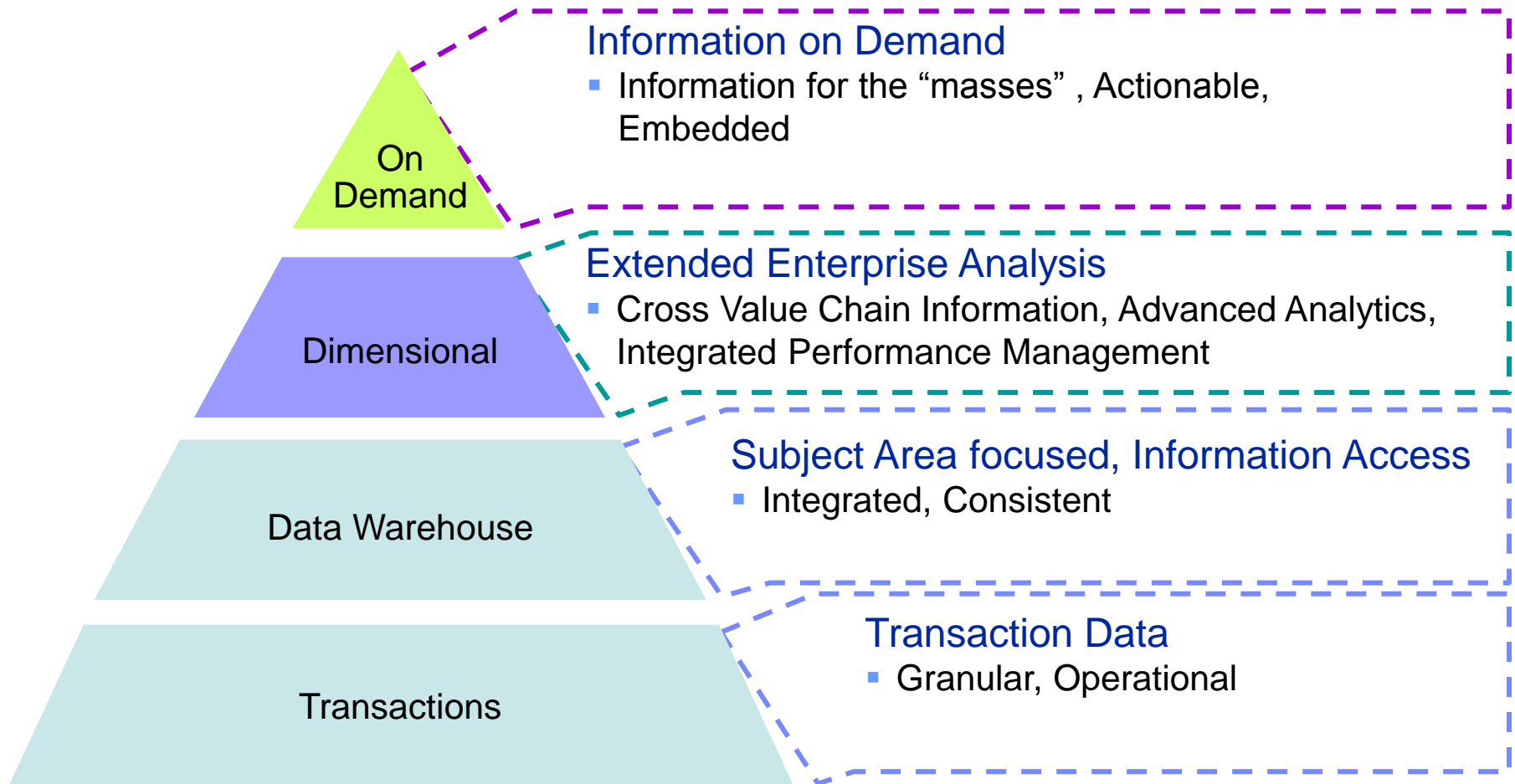Van Informatie op Orde naar Informatie van Waarde
BAO Reference Architecture
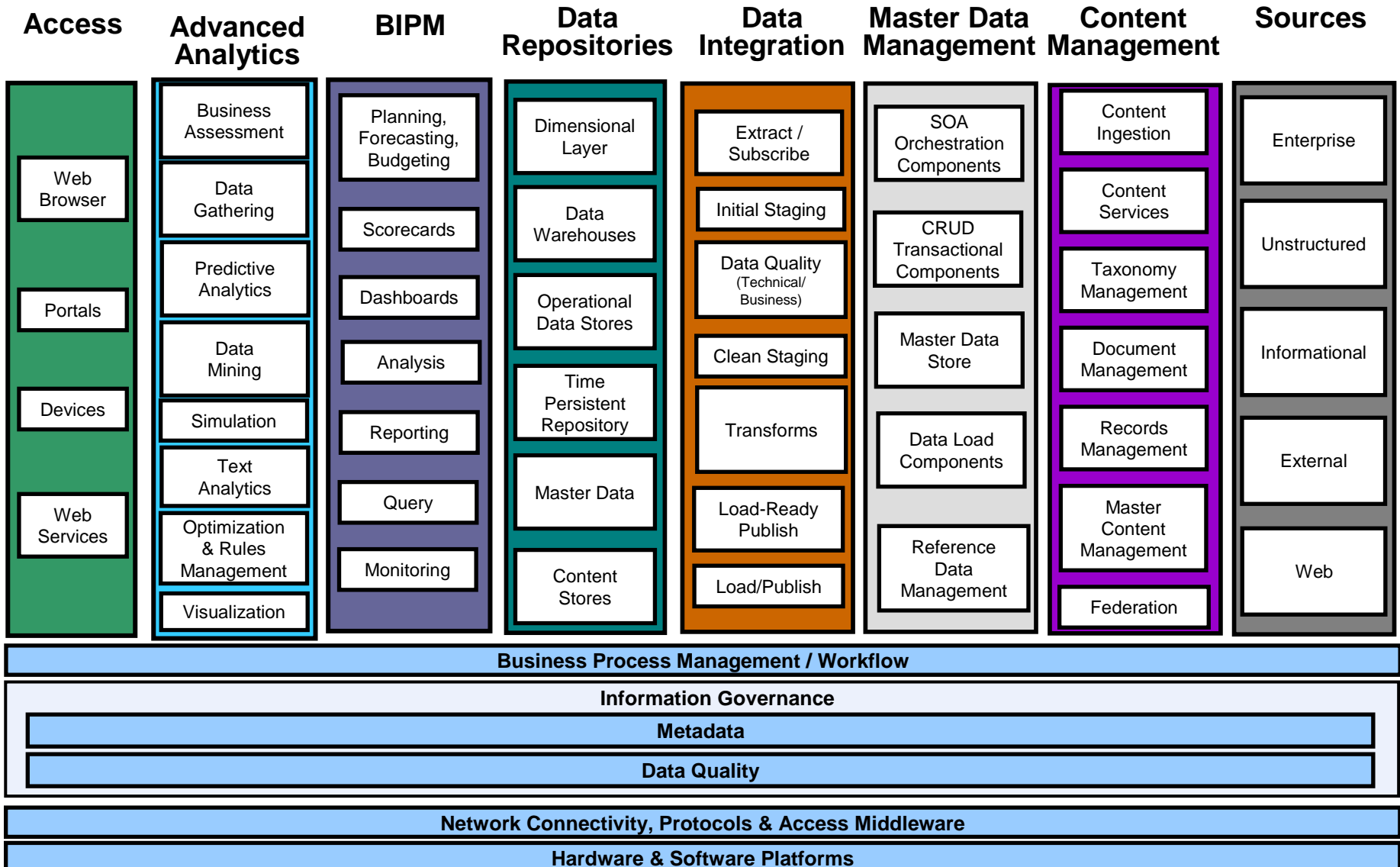
The IBM Business Analytics and Optimization Reference Architecture

Following a natural pattern, from Data to Information on Demand

On Demand

Information on Demand
- Information for the "masses", Actionable, Embedded

Dimensional

Extended Enterprise Analysis
- Cross Value Chain Information, Advanced Analytics, Integrated Performance Management

Data Warehouse

Subject Area focused, Information Access
- Integrated, Consistent

Transactions

Transaction Data
- Granular, Operational

The IBM Business Analytics and Optimization Reference Architecture Overview

Access	Advanced Analytics	BIPM	Data Repositories	Data Integration	Master Data Management	Content Management	Sources
Web Browser	Business Assessment	Planning, Forecasting, Budgeting	Dimensional Layer	Extract / Subscribe	SOA Orchestration Components	Content Ingestion	Enterprise
	Data Gathering	Scorecards	Data Warehouses	Initial Staging	CRUD Transactional Components	Content Services	Unstructured
Portals	Predictive Analytics	Dashboards	Operational Data Stores	Data Quality (Technical/ Business)		Taxonomy Management	
	Data Mining	Analysis	Time Persistent Repository	Clean Staging	Master Data Store	Document Management	Informational
Devices	Simulation	Reporting		Transforms	Data Load Components	Records Management	
	Text Analytics	Query	Master Data	Load-Ready Publish		Master Content Management	External
Web Services	Optimization & Rules Management	Monitoring	Content Stores	Load/Publish	Reference Data Management		Web
	Visualization					Federation	

Business Process Management / Workflow

Information Governance

Metadata

Data Quality

Network Connectivity, Protocols & Access Middleware

Hardware & Software Platforms

The IBM Business Analytics and Optimization Reference Architecture Content Management Layer

For some layers of our BAO Reference Architecture, we have further defined based on engagement experience a detailed and proven Content Management Architecture with common conceptual, logical, and physical components. These components are designed to optimize the inherent strengths of the Content Management technologies.
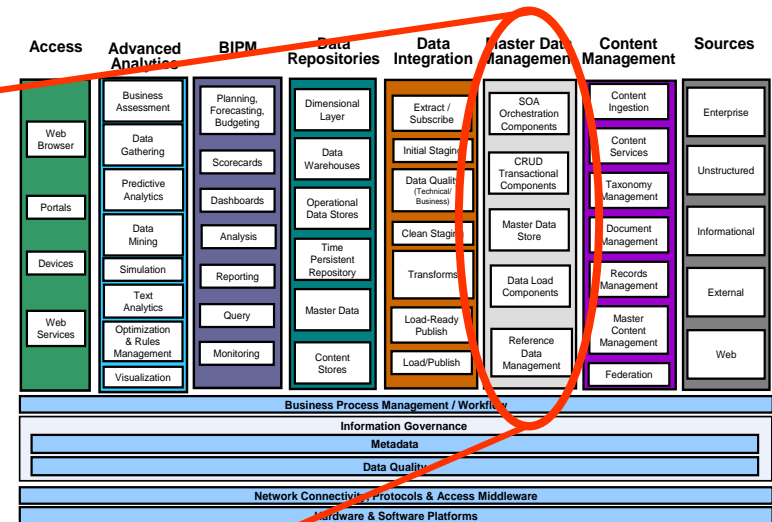
Content Management Layer

- Content Ingestion
- Content Services
- Taxonomy Management
- Document Management
- Records Management
- Master Content Management
- Federation

Access	Advanced Analytics	BIPM	Data Repositories	Data Integration	Master Data Management	Content Management	Sources
Web Browser	Business Assessment	Planning, Forecasting, Budgeting	Dimensional Layer	Extract / Subscribe	SOA Orchestration Components	Content Ingestion	Enterprise
	Data Gathering	Scorecards	Data Warehouses	Initial Staging	CRUD Transactional Components	Content Services	Unstructured
Portals	Predictive Analytics	Dashboards	Operational Data Stores	Data Quality (Technical/Business)	Master Data Store	Taxonomy Management	
	Data Mining	Analysis	Time Persistent Repository	Clean Staging		Document Management	Informational
Devices	Simulation	Reporting		Transforms	Data Load Components	Records Management	
	Text Analytics	Query	Master Data	Load-Ready Publish		Master Content Management	External
Web Services	Optimization & Rules Management	Monitoring	Content Stores	Load/Publish	Reference Data Management		Web
	Visualization					Federation	

Business Process Management / Workflow

Information Governance

Metadata

Data Quality

Network Connectivity, Protocols & Access Middleware

Hardware & Software Platforms
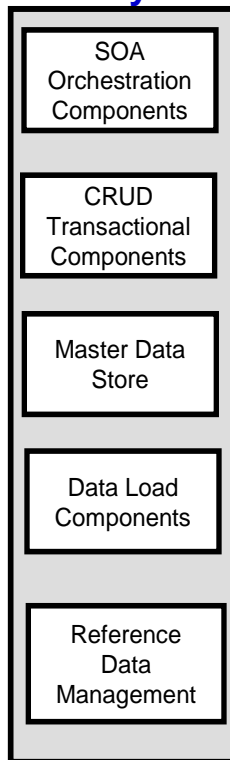
Content Management Layer Definition

Contains the services, technologies and processes used to capture, manage, store, preserve, and deliver unstructured content. It provides global access and management of digital assets used to collaborate and share information between a company and its customers, suppliers, employees and business partners.

The IBM Business Analytics and Optimization Reference Architecture Master Data Management Layer

Master Data Management is the process of consolidating enterprise master data from multiple systems for the purpose of validating and cleaning the data to support and improve various business processes. Master data includes business critical information about customers, suppliers, products, and accounts

Master Data Management Layer

SOA Orchestration Components	
CRUD Transactional Components	
Master Data Store	
Data Load Components	
Reference Data Management	

Access	Advanced Analytics	BIPM	Data Repositories	Data Integration	Master Data Management	Content Management	Sources
Web Browser	Business Assessment	Planning, Forecasting, Budgeting	Dimensional Layer	Extract / Subscribe	SOA Orchestration Components	Content Ingestion	Enterprise
	Data Gathering	Scorecards	Data Warehouses	Initial Staging	CRUD Transactional Components	Content Services	Unstructured
Portals	Predictive Analytics	Dashboards	Operational Data Stores	Data Quality (Technical/ Business)		Taxonomy Management	
	Data Mining	Analysis	Time Persistent Repository	Clean Staging	Master Data Store	Document Management	Informational
Devices	Simulation	Reporting		Transforms	Data Load Components	Records Management	
	Text Analytics	Query	Master Data	Load-Ready Publish		Master Content Management	External
Web Services	Optimization & Rules Management	Monitoring	Content Stores	Load/Publish	Reference Data Management	Federation	Web
	Visualization						

Business Process Management / Workflow

Information Governance

Metadata

Data Quality

Network Connectivity, Protocols & Access Middleware

Hardware & Software Platforms

Master Data Management Layer Definition

Master data management or MDM is a set of disciplines, technologies, and solutions to create and maintain consistent, complete, contextual, and accurate business data for all stakeholders across and beyond the enterprise.

The IBM Business Analytics and Optimization Reference Architecture Data Integration Layer

For some layers of our BAO Reference Architecture, we have further defined based on engagement experience a detailed and proven Data Integration Architecture with common conceptual, logical, and physical components. These components are designed to optimize the inherent strengths of the Data Integration technologies.
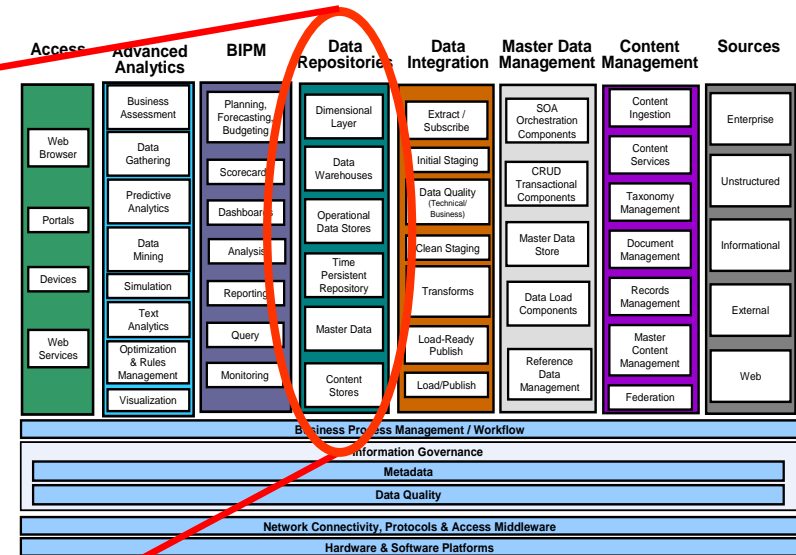
Data Integration Layer

| Extract / Subscribe |
| Initial Staging |
| Data Quality Technical/Business |
| Clean Staging |
| Transformation |
| Load-Ready Publish |
| Load/Publish |

Access | Advanced Analytics | BIPM | Data Repositories | Data Integration | Master Data Management | Content Management | Sources
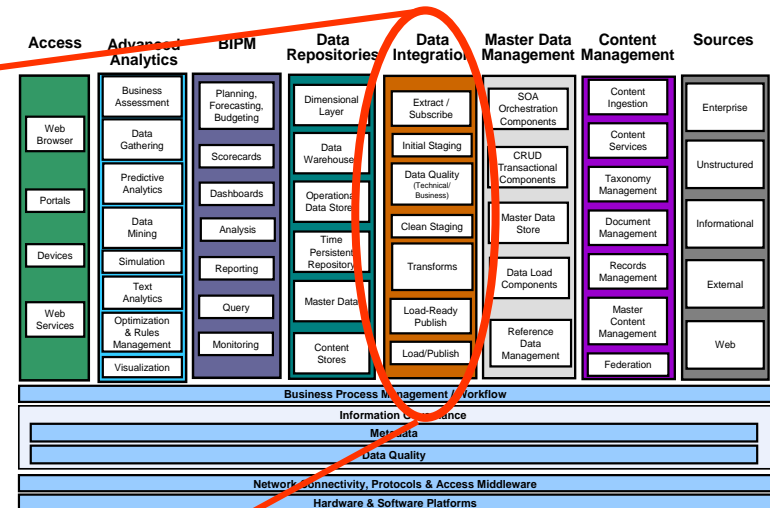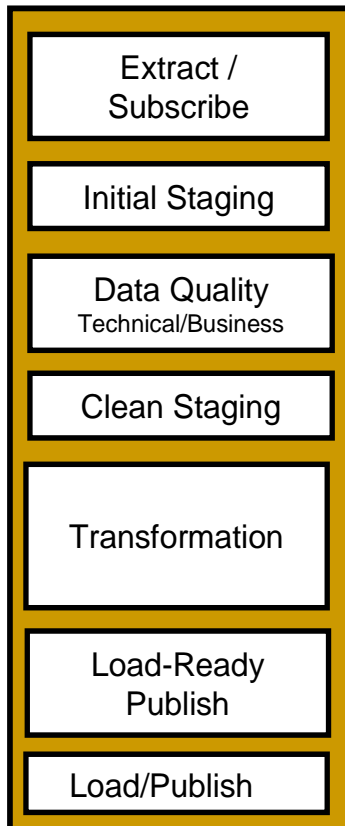
Data Integration Layer Definition

The Data Integration Architectural Layer focuses on the processes and environments that deal with the capture, qualification, processing, and movement of data in order to prepare it for storage in the Data Repository Layer, which is subsequently shared with the Analytical/Access applications and systems. This layer may process data in scheduled batch intervals or in near real-time/"just-in-time" intervals, depending on the nature of the data and the business purpose for its use.

The IBM Business Analytics and Optimization Reference Architecture Data Repository Layer

For some layers of our BAO Reference Architecture, we have further defined each layer based on our engagement experience.

Data Repository Layer

- Dimensional Layer
- Data Warehouses
- Operational Data Stores
- Time Persistent Repository
- Master Data
- Content Stores

Data Repository Layer Definition

The Data Repository Layer contains the databases and data stores and related components that provide most of the storage for the data which supports a BAO environment. The Data Repositories Layer's repositories are not a replacement or replica of operational databases which reside on the Data Source Layer, but are a complementary set of data repositories that reshape data into formats necessary for making decisions and managing a business. These database structures are represented by conceptual, logical, and physical data models and data model types (e.g. 3NF, star/snowflake schemas, unstructured, etc.)

The IBM Business Analytics and Optimization Reference Architecture Business Intelligence and Performance Management (BIPM) Layer

Empowers decision making and improved business performance through the timely access, analysis and reporting of actionable, accurate, and personalized information.

BIPM Layer

- Planning, Forecasting, Budgeting
- Scorecards
- Dashboards
- Analysis
- Reporting
- Query
- Monitoring
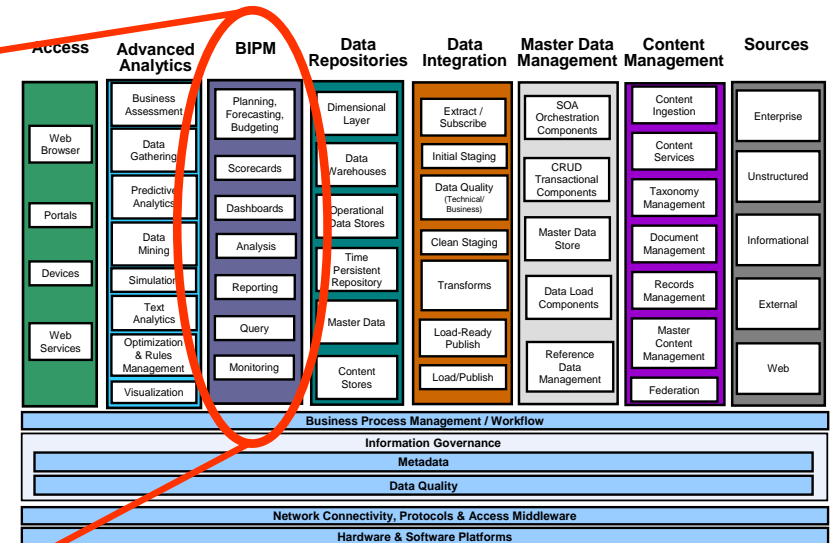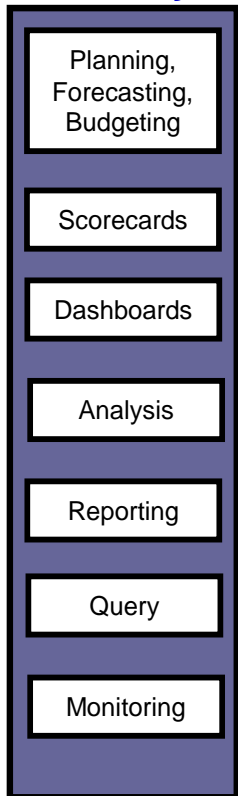
BIPM Layer Definition

The BIPM Layer empowers decision making and improved business performance through the timely access, analysis and reporting of actionable, accurate, and personalized information. A variety of applications may be supported, from static reporting to balanced scorecards to monitoring tools that are embedded within an operational process. This layer is typically composed of various technological components destined to meet specific needs, and are usually built from "best-of-breed" software and tools such as data access models, query tools, reporting tools for OLAP.

The IBM Business Analytics and Optimization Reference Architecture
Advanced Analytics Layer

Enhances organizational performance by applying advanced mathematical modeling, deep computing, simulation, data analytics and optimization techniques to improve operational efficiency.
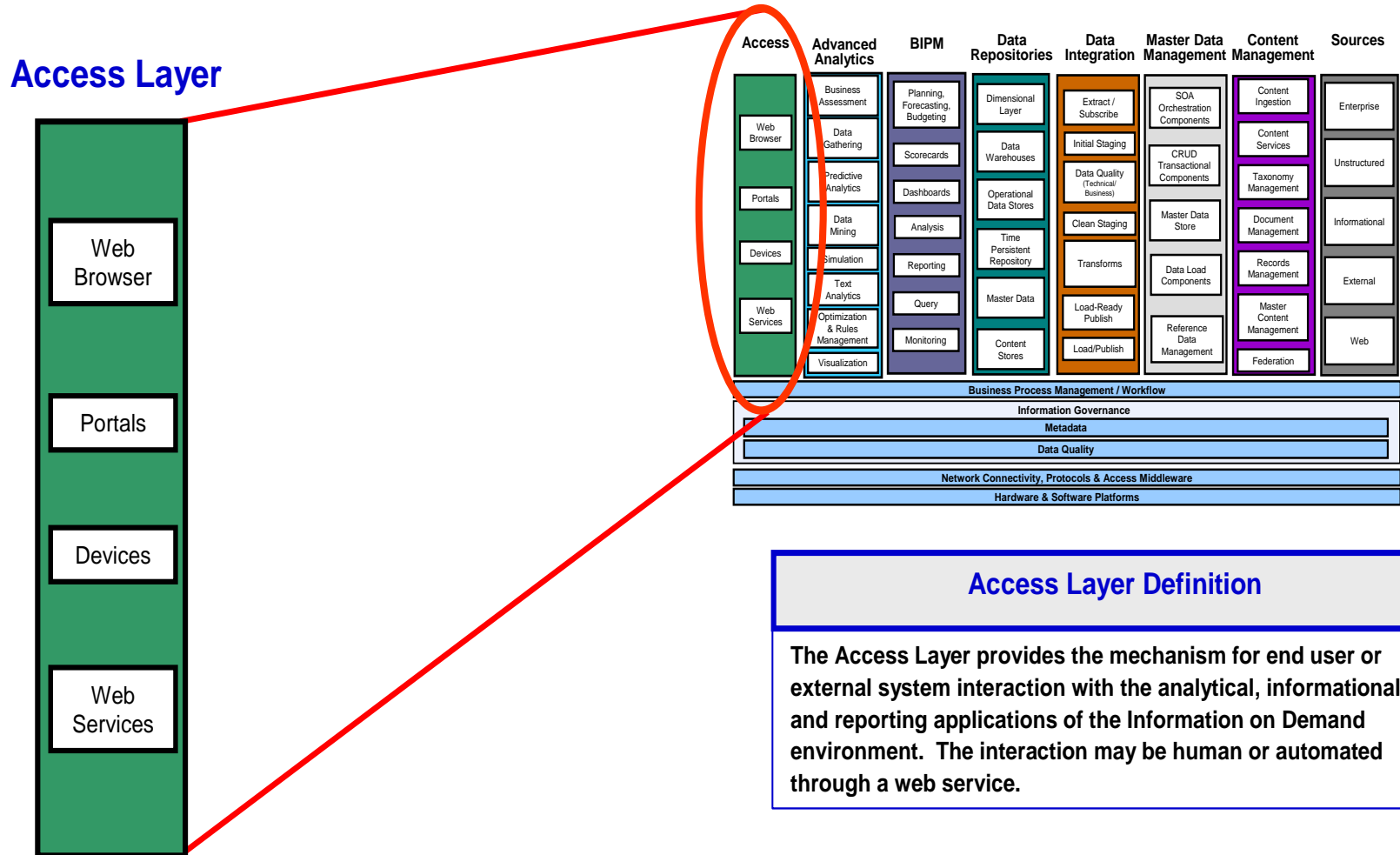
Advanced Analytics Layer

- Business Assessment
- Data Gathering
- Predictive Analytics
- Data Mining
- Simulation
- Text Analytics
- Optimization & Rules Management
- Visualization

Access	Advanced Analytics	BIPM	Data Repositories	Data Integration	Master Data Management	Content Management	Sources
Web Browser	Business Assessment	Planning, Forecasting, Budgeting	Dimensional Layer	Extract / Subscribe	SOA Orchestration Components	Content Ingestion	Enterprise
	Data Gathering	Scorecards	Data Warehouses	Initial Staging	CRUD Transactional Components	Content Services	Unstructured
Portals	Predictive Analytics	Dashboards	Operational Data Stores	Data Quality (Technical/Business)	Master Data Store	Taxonomy Management	
Devices	Data Mining	Analysis	Time Persistent Repository	Clean Staging		Document Management	Informational
	Simulation	Reporting		Transforms	Data Load Components	Records Management	External
Web Services	Text Analytics	Query	Master Data	Load-Ready Publish	Reference Data Management	Master Content Management	Web
	Optimization & Rules Management	Monitoring	Content Stores	Load/Publish		Federation	
	Visualization						

Business Process Management / Workflow
Information Governance
Metadata
Data Quality
Network Connectivity, Protocols & Access Middleware
Hardware & Software Platforms

Advanced Analytics Layer Definition

The Analytics Layer uses data and models to provide insight to guide decisions. It empowers clients to make more effective decisions and build more productive systems based on more complete data, consideration of all available options, and careful predictions of outcomes and estimates of risk. This layer is typically composed of various technological components destined to meet specific needs, and are usually built from "best-of-breed" software and tools such as SPSS and ILOG.

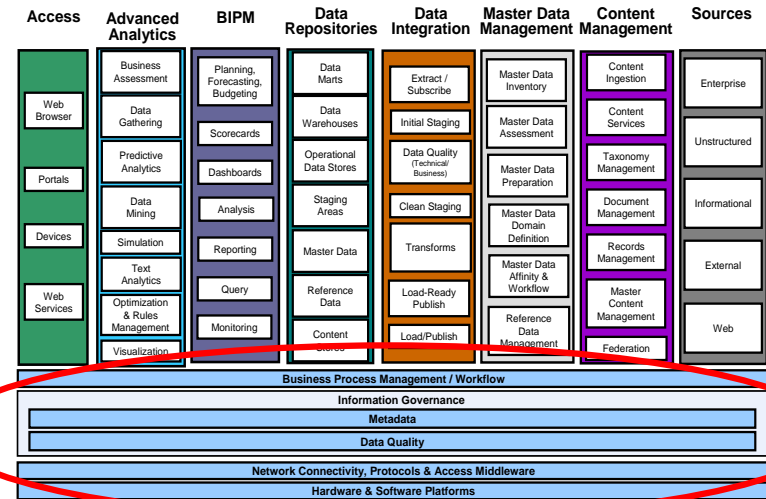# The IBM Business Analytics and Optimization Reference Architecture Access Layer

For some layers of our BAO Reference Architecture, we have further defined each layer based on our engagement experience.
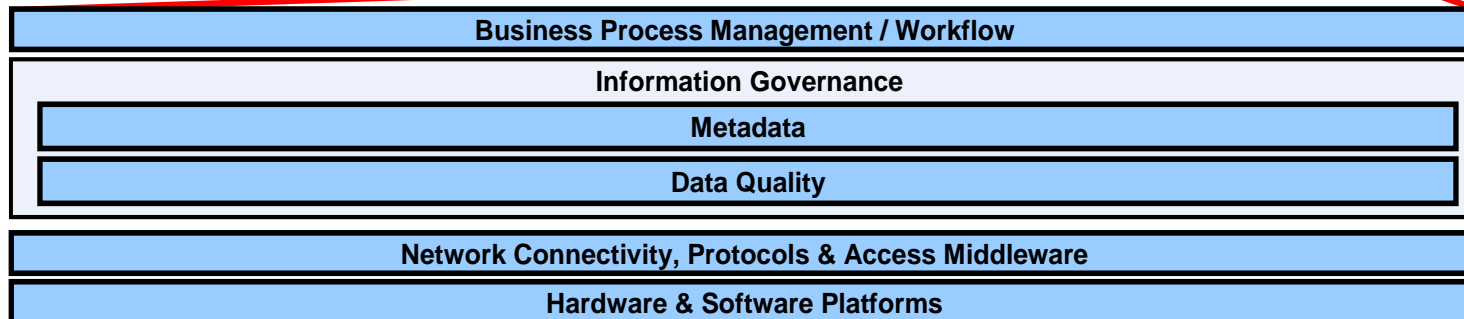
Access Layer

Web Browser
Portals
Devices
Web Services

Access	Advanced Analytics	BIPM	Data Repositories	Data Integration	Master Data Management	Content Management	Sources
Web Browser	Business Assessment	Planning, Forecasting, Budgeting	Dimensional Layer	Extract / Subscribe	SOA Orchestration Components	Content Ingestion	Enterprise
	Data Gathering	Scorecards	Data Warehouses	Initial Staging	CRUD Transactional Components	Content Services	Unstructured
Portals	Predictive Analytics	Dashboards	Operational Data Stores	Data Quality (Technical/ Business)	Master Data Store	Taxonomy Management	
	Data Mining	Analysis	Time Persistent Repository	Clean Staging		Document Management	Informational
Devices	Simulation	Reporting	Master Data	Transforms	Data Load Components	Records Management	External
	Text Analytics	Query		Load-Ready Publish		Master Content Management	
Web Services	Optimization & Rules Management	Monitoring	Content Stores	Load/Publish	Reference Data Management	Federation	Web
	Visualization						

Business Process Management / Workflow

Information Governance

Metadata

Data Quality

Network Connectivity, Protocols & Access Middleware

Hardware & Software Platforms

Access Layer Definition

The Access Layer provides the mechanism for end user or external system interaction with the analytical, informational, and reporting applications of the Information on Demand environment. The interaction may be human or automated through a web service.

The IBM Business Analytics and Optimization Reference Architecture Cross Layer

For some layers of our BAO Reference Architecture, we have further defined each layer based on our engagement experience.

Cross Layer

Business Process Management / Workflow
Information Governance
Metadata
Data Quality
Network Connectivity, Protocols & Access Middleware
Hardware & Software Platforms

Cross Layer Architecture Definition

The Cross Layer provides the cohesive foundation whereupon the BAO pillars can function in an integrated and seamless fashion.

11

Glossary

The IBM Business Analytics and Optimization Reference Architecture Glossary – Content Management

- **Content Ingestion:** Collects, classifies, analyzes, assigns metadata to and stores content into Content Stores.

- **Content Services:** Repository services that provide check in/out, versioning, permission management and enforcement, as well as higher-level application services.

- **Taxonomy Management:** Defines and rationalizes the document hierarchy and definitions across the enterprise.

- **Document Management:** Provides lifecycle management, structured and collaborative authoring, automatic publishing, and support for multiple languages.

- **Records Management:** Allows content and associated metadata to be treated as business records which can be held or disposed of according to business needs.

- **Master Content Management:** A set of processes and tools that consistently defines and manages the content entities of an organization. MCM provides processes for collecting, aggregating, matching, consolidating, quality-assuring, persisting and distributing the content throughout an organization to ensure consistency and control in the ongoing maintenance and application use of this information.

- **Federation:** Consolidated search and retrieval of documents and metadata across multiple disparate content stores.

The IBM Business Analytics and Optimization Reference Architecture Glossary – Data Integration

- **Extract / Subscribe:** Extract/Subscribe is the set of processes that capture data, transactional or bulk, structured or unstructured, from various sources and lands it to an Initial Staging Area. It follows the architectural principle of "read once, write many" to ensure that impact on source systems is minimized, and data lineage is managed.

- **Initial Staging:** Initial Staging is an optional "landing zone" where the copy of the data from sources is landed as a result of the extract/subscribe processing. One of the purposes for the Initial Staging Area is to persist source data in non-volatile storage to achieve the "pull it once from source" goal. Data from real time sources that is intended for real time targets only is not passed through Extract/Subscribe and may not land in Initial Staging Area.

- **Data Quality:** Data Quality are the processes that qualifies and cleanses the data, based upon Technical and Business Process rules. Regardless of the data quality rules the Data Quality Layer should provide the following functionality:
  - **Cleansed Data Files:** using the data quality criteria
  - **Reject Data Files:** data records that fail the cleansing logic
  - **Reject Reports:** a tabular report of the records that failed with reason codes for review and renovation

- **Clean Staging:** The Clean Staging Area in the next optional landing zone, which contains records that have passed all DQ checks. This data may be passed to processes that build load ready files. The data may also become input to transformation processes which, in turn produce new data sets. The Data Integration architecture should include an archiving facility for the files in the clean staging area

- **Transformation:** A set of well-defined processes that transform the data from a source format to a common target and include Calculation & Split components, Processing & Enrichment components and Filtering components

- **Load-Ready Publish:** Load-Ready Publish is a optional staging area that is utilized to store target-specific load-ready files. If a target can take a direct output from a data integration tool without storing the data first, storing it in Load-ready Staging Area may not be required.

- **Load/Publish:** Load/Publish is a set of standardized processes. Loads are structured by subject area by data store, for example, Subject Areas in the Data Warehouse such as Involved Party. There are five types of physical load architectures, they include:
  - **FTP to Target** – in this type of load, the process is only responsible for depositing the output to the target environment.
  - **Piped data** – the process executes a load routine on the target that takes the data directly piped from the Target Specific Filter.
  - **RDBMS Utilities** – e.g. DB2's Bulk Loader on the target, but the source is Load-Ready Staging Area.
  - **SQL** – SQL writes directly to the target database.
  - **Message Publishing** – for loading real time data feeds to message queues

The IBM Business Analytics and Optimization Reference Architecture Glossary – Data Repository

- **Dimensional Layer:** Dimensional models are developed to support a single business function or process. It is usually a subset of information found in the data warehouse, further transformed and re-shaped for a specific analytical application. The dimensional model can contain both current and historical data, and typically contains summarized and aggregated data. They are also:
  - Usually modeled as a star schema, or snowflake models.
  - Can be instantiated as views, materialized views, or tables.

- **Data Warehouse:** A Data Warehouse (DW) is the main store of analytic information. The DW provides business subject area orientation in order easily rationalize data from multiple subjects (such as different lines of business) and source systems. The data model is structured through its key designs to be able to accommodate updates either by traditional change data capture processes or snapshot updates. Because of the dynamic for multiple source system and business subject area integration, and change data capture, the optimal data model approach is 3rd Normal Form. These 3rd normal form Data Warehouse Models are not transactional in format, but designed for optimal loading and reading of analytic information. Data Warehouse may contain historical data, as well as rationalized transactional data, aggregated data, and derived/calculated data.

- **Operational Data Store:** A Operational Data Store (ODS) is the main repository and integration point of operational data from disparate systems. The ODS stores transaction level detailed data used to satisfy common, integrated enterprise-level operational data needs. The ODS contains current, non-redundant detailed data common across multiple systems or organizational units. ODS's are often the data store structures used for Master Data Management Hub environments instantiating core subject areas such as Customer and Product.

- **Time Persistent Repositories:** Time Persistent Repositories (TPR) are structures that are modeled after the application data stores and databases. It is a raw (no data quality flagging), siloed by application, source of transactional data. A TPR is used for operational reporting, data quality profiling, and sourcing for Advanced Analytic applications such as SAS and SPSS. They are "lightly conformed" enough in order to provide a common key structure in order to provide connectivity to data warehouses and dimensional layers.

.

- **Master Data:** Master Data contains the core business concepts and hierarchies of an organization such as: suppliers, products, customers, and organization. These data stores are often are designed as an ODS data structure and are referenced by both analytic and transactional systems.

- **Content Stores:** Content stores contain unstructured data such as text, video, documents and non-traditional relational formats such as XML. The design of these data stores using concepts such as content modeling are dependent on the planned business usage and underlying technology.

The IBM Business Analytics and Optimization Reference Architecture Glossary – BIPM

- **Planning, Forecasting, Budgeting:** Planning, Budgeting & Forecasting leverages analytics to align financial and operational plans, understand target values for key categories of revenue and expenditure, and evaluate expected business outcomes. It measures progress against leading industry best practices for the purpose of identifying opportunities to better link strategy to action, optimize budget allocations, and perform what-if analysis.

- **Dashboards & Scorecards:** Dashboards & Scorecards provide a mechanism to translate corporate strategy into measures, targets and initiatives across an organization and to achieve the visibility required to manage corporate performance. Dashboards are used by business managers to take immediate actions and improve day to day business performance. They have a limited time horizon and are updated weekly and/or daily. Dashboards typically use leading indicators, provide drilldown capabilities, and leverage business activity monitoring and exception alerts. Scorecards are used by executives to perform cross functional monitoring of progress towards achieving business strategy. They usually leverage historical indicators and provide limited drilldown capabilities. Scorecards have a longer time horizon and are updated monthly/quarterly

- **Analysis & Reporting:** Business Analytics & Reporting provides the ability to connect disparate, disconnected, and non-integrated data from departmental and functionally siloed sources into a consistent, commonly defined and governed reporting format to enable timely and accurate analytical and reporting capabilities. It helps personalize information delivered to the user community and defines the "why" and "how" behind historically focused "what happened" analysis. Business Analytics & Reporting includes setting future direction, defining measures, targets, managing value drivers and analytic dimensions, enabling insight and vision around business events. It evaluates business requirements and current enterprise wide reporting processes in order to leverage the information architecture in the most effective and efficient manner across the organization

- **Query:** Relational ad doc query and reporting capabilities

- **Monitoring:** Operational reporting and real-time monitoring of key performance indicators (KPIs) and operational metrics

The IBM Business Analytics and Optimization Reference Architecture Glossary – Advanced Analytics

- **Business Assessment:** Involves understanding the business problem to be solved and determining the Advanced Analytics & Optimization technique best suited to address the client need. During this phase, business users, modelers, and domain experts work closely together to define the project objectives and the requirements from a business perspective. The project objective is then translated into a problem definition for predictive analytics & data mining, optimization & rules management, simulation, and/or visualization.

- **Data Gathering:** Identifies and collects existing source information and prepares it for usage by subsequent analytical processes. During the data gathering, understanding the client business processes and analytical needs within context of their information environment is essential.

- **Data Mining:** Focused on extracting patterns and previous unknown facts from large volumes of data. It helps businesses uncover key insights, patterns, and trends in data, then uses this insight to optimize business decisions. Data mining techniques can be divided into major categories including classification (arranging data into predefined groups), clustering (similar to classification but the groups are not predefined), and regression (statistical analysis between a dependent variable and one or more independent variables).

- **Optimization:** Helps companies improve operational efficiency while also providing centralized business rules in robust repositories that can be used across applications. Business Optimization leverages advanced mathematical techniques to find the best solution to a complex problem with many decision options and constraints. It is a powerful analytical tool for calculating the best possible utilization of resources to help achieve a desired business result, such as reducing cost or processing time, or increasing profit, serviceability and throughput.

- **Predictive Analytics:** Analyzes patterns found in historical and current transaction data as well as attitudinal survey data to predict potential future outcomes. The core of predictive analytics relies on capturing relationships between explanatory variables and developing models to predict future outcomes.

- **Rules Management:** Used to define, deploy, execute, monitor, and maintain the variety and complexity of business logic that are used by operational systems within an organization. Business rules are typically written using IF/THEN statements, decision tables, decision trees and scorecards. Business rules describe the operations, business logic, and constraints that apply to an organization in achieving its goals.

- **Simulation:** Replicates a system, process, behavior, or business problem using advanced analytical techniques. It is used to perform "What If" analysis based on a set of parameters and input variables. Simulation models a business process to estimate the impact of management decisions or changes. Simulation enables companies to reproduce the dynamic behavior of a business process, in order to analyze workloads and potential bottlenecks. Simulations are an example of how technology can aid smarter decision-making. It can predict, for example, what will happen to an area if a new major facility is built and lead to improved planning of roads and public transportation. By providing insight into the impact of decisions and design alternatives, simulation can help companies determine the optimal path forward.

- **Text Analytics:** Analysis of textual patterns to provide business insight. Infers meaning from unstructured data

- **Visualization:** Allows clients to gain insight through diagrams, maps, schedules, charts, and images. Visualization techniques manipulate, transform, and render data based on points, lines, areas, volumes, images or geometric primitives in any combination. Visualization is a powerful tool for understanding the behavior of complex systems. By building graphical displays users can easily understand and interpret large volumes of data. Visualization is applicable to a wide range of business problems including financial trend analysis, monitoring of traffic and communication system, analysis of social networks, and arrangement of large-scale text and image data.

The IBM Business Analytics and Optimization Reference Architecture Glossary – Access

- **Web Browsers:** A software front end application used for retrieving and presenting information on the internet

- **Portals:** A method of consolidating multiple information sources on one site with common access control

- **Devices:** Any device attached to another computer or computing device that extends its functionality. The device has the ability to transmit and receive information

- **Web Services:** A service or set of services that can be accessed via the internet or other computing processing methods and provide content, information or processing services

The IBM Business Analytics and Optimization Reference Architecture Glossary – Cross Layer

- **BPM:** Business Process Management is a management approach based on continually improving and optimizing business processes using business value and technology as drivers utilizing a systematic approach for continuous improvement

- **Information Governance:** The orchestration of people, process, and technology to enable an organization to leverage data as an enterprise asset. Information Governance allows organizations to be more consistent, coherent, and comprehensive in the way they define, discuss, analyze. and leverage data in their environment to achieve business objectives

- **Metadata:** Metadata is often described as "information about information", but it can also be explained as another layer of information created to help people use data. The process to create metadata involves simplifying, describing, distilling, and organizing information in order to provide capabilities to search and explore data. Consistency in the metadata is necessary to keep information organized. Consistent terminology helps communicate metadata, and it helps applications process the metadata.

- **Data Quality:** Data quality is defined by how effectively the data supports the transactions and decisions needed to meet an organization's strategic goals and objectives, as embodied in its ability to manage its assets and conduct its core operations.

- **Network, Connectivity, Protocols & Access Middleware:** The platform and technology that allows communication between components, formal description of message formats and the rules for exchanging those messages. Security plays an important role in this sub-layer in that the appropriate controls and accesses need to be in place and continously monitored.

- **HW & SW Platforms:** The physical platform that includes the hardware components and the software applications that run on the platform