



California
DEPARTMENT OF TECHNOLOGY

California Enterprise Architecture Framework

Business Intelligence (BI) Reference Architecture (RA)

Version 1.0 Final
January 2, 2014



This Page is Intentionally Left Blank



TABLE OF CONTENTS

1	Introduction.....	1
1.1	Purpose	1
1.2	Limitations.....	2
1.3	Intended Users.....	2
1.4	Document Organization	2
1.5	Future Directions	2
2	Business Intelligence Overview.....	3
2.1	Definitions.....	3
2.2	Business Benefits of BI	6
2.3	Key Capabilities of BI Solution	6
2.4	Components of BI Solution	7
2.4.1	<i>Data Sources</i>	8
2.4.2	<i>Enterprise Information Integration (EII) Components</i>	8
2.4.3	<i>BI Data Repositories</i>	9
2.4.4	<i>BI and Performance Management Tools</i>	10
2.4.5	<i>Advanced Analytics Components</i>	11
2.4.6	<i>Operational Components</i>	12
3	Business Intelligence Reference Architecture Description.....	14
3.1	BI RA Conceptual View	14
3.2	BI RA Logical View	16
3.2.1	<i>Information Preparation Scenario</i>	16
3.3	BI RA Deployment View	18
4	Glossary	19
5	References	20
6	Document History	21



LIST OF FIGURES

<i>Figure 2-1 Business Intelligence Process</i>	4
<i>Figure 2-2 Information Preparation and Information Usage</i>	4
<i>Figure 2-3 Structured and Semi-Structured Data Sources for BI</i>	5
<i>Figure 2-4 BI Solution Components Overview</i>	7
<i>Figure 2-5 Data Sources Components in BI RA</i>	8
<i>Figure 2-6 EII Components in BI RA</i>	9
<i>Figure 2-7 BI Data Repositories</i>	9
<i>Figure 2-8 Business Intelligence and Performance Management Tools</i>	10
<i>Figure 2-9 Advanced Analytics Components in BI</i>	11
<i>Figure 2-10 Data Management Components in BI</i>	12
<i>Figure 2-11 Supporting Components in BI</i>	12
<i>Figure 3-1 BI RA Conceptual View</i>	15
<i>Figure 3-2 Service and Component Interactions in Information Preparation Scenario</i>	17



LIST OF TABLES

<i>Table 3-1 Component Interactions in Information Preparation Scenario</i>	18
<i>Table 6-1 Document History</i>	21



This Page is Intentionally Left Blank

1 Introduction

Business Intelligence (further abbreviated as “BI”) is a term coined by Gartner in 1989 so that one can put a label to all the organizational and technological facets of gathering and analyzing available data in order to improve business decision making and performance – taken as a whole.

When viewed in a historical perspective, BI brings together a number of concepts and approaches in data processing and management information systems, such as decision support systems and executive information systems (DSS and EIS), on-line analytical processing (OLAP) systems, and knowledge management systems (KMS). The traditional approach to providing BI functions, by piecing together those systems, has been facing a number of hurdles, including inherent complexity and integration challenges. This meant high risk and high cost.

Over the years, the fundamental need driving BI initiatives has remained the same – it is still the need to provide accurate and timely information for decision making in the enterprise. However, the complexity of making BI happen has been increasing. One factor is the volume of data that has been growing at an unprecedented rate. Another factor is the variety of data types and their formats as produced by systems implemented with disparate technologies and disparate ways of making them accessible. Yet another factor is ever changing technology and platforms. All of those factors, especially when taken together, drive the interest in better solutions for supporting Business Intelligence functions.

The current approach to BI solutions in the industry is shaped by the rapid adoption of Service Orientation and the growing interest in the Cloud. Adoption of SOA leads to considering BI as a shared service or a shared platform, rather than the traditional bundle of monolithic software products with problematic interoperability. Interest in the Cloud encourages considering deployment of these services in the Cloud rather than on-premises, including treating BI solution as a candidate SaaS (Software-as-a-Service) and BI platform as a candidate PaaS (Platform-as-a-Service). These developments provide the context for reconsidering the traditional blueprints for BI, so that the goal of delivering information to the right recipients at the right time becomes easier to achieve than before.

1.1 Purpose

The BI Reference Architecture (RA) document provides guidelines and options for making architectural decisions when implementing BI solutions.

The objectives for the document include the following:

- To introduce key terms and distinctions relevant for the topic
- To provide inputs for creating or evaluating architectures for BI from enterprise perspective
- To identify building blocks (architectural layers, services, components) for integrating elements of a BI solution
- To communicate the key architectural decisions relevant for creating or evaluating BI solutions
- To communicate opportunities for solution and/or platform sharing at agency, cross-agency and/or state levels



1.2 Limitations

This document focuses on BI and related concepts at the enterprise architectural level in the context of CEAF 2.0. It is not intended as a general guide for BI or an endorsement of specific products. Its overall perspective remains CEAF-centric.

1.3 Intended Users

The primary intended users of this document are Enterprise Architecture practitioners and other architects that contribute to enterprise architecture. This broad group includes architects from other domains/disciplines such as Security, Application, Information, Business, Technology, Infrastructure, and Solution Architects. It also includes Managers, at senior or operational levels, who are involved with BI or related areas, such as Master Data Management, Enterprise Application Integration, Cloud Computing, and similar areas.

1.4 Document Organization

The BI Reference Architecture documentation is organized as follows:

- Section “BI Overview” provides background for the BI RA by introducing descriptions and definitions of BI, discusses the main usage scenario types found in BI implementations, and identifies architectural components for respective usage scenarios
- The section “BI Reference Architecture Description” elaborates BI RA using the following architectural views:
 - The Conceptual View (in the subsection “BI RA Conceptual View”) introduces the necessary capabilities for a BI architecture and how they are supported by Architectural Building Blocks (ABBs)
 - The Logical View (in the subsection “BI RA Logical View”) describes key interactions among Layers and/or ABBs to realize functionality specific to BI systems
 - The Deployment View (in the subsection “BI RA Deployment View”) is intended to show system topologies and deployment facets of BI in the state
- The section “Glossary” provides description of the terms and abbreviations used in the document
- The section “References” lists publications used for preparation of the document. Note that literals in square brackets that appear in the document (e.g., [A], [5], [d]) are references to publications listed in this section

1.5 Future Directions

Future evolution of the document includes the following steps:

- Addition of existing best-practice-based realizations of the BI RA
- Identification and elaboration of solution sharing opportunities
- Formulation of implementation guidelines for BI RA

2 Business Intelligence Overview

This section provides a description of Business Intelligence (BI), including clarification of key terms and concepts. It identifies BI's intended business benefits and summarizes its main usage scenarios. A set of key capabilities of BI solution are identified and key components of the solution are described at a high level.

2.1 Definitions

Business Intelligence (BI) is “an umbrella term” that includes the applications, infrastructure and tools, and best practices that enable *access to* and *analysis of* information to improve and optimize decisions and performance [2].

The uses of BI in the public sector and the private sector are likely to be different in some areas, given that the private sector is concerned with competition in the market and its competitive analysis. These areas are not necessarily of primary importance in the public sector which is mainly concerned with service performance. Nevertheless, there are still vital areas for which BI can be used in the public sector, including obtaining a better understanding of the following areas: [6]

- Weak and strong areas in business processes and their execution
- High cost and/or low efficiency areas in the organization
- The capabilities available (or absent) in the organization
- Ramifications of (the changes to) the regulatory environment in which the organization operates
- The actions of other relevant organizations and the implications of these actions for the organization in question

In turn, **BI Systems** are described as systems combining data gathering, data storage, and knowledge management with analytical tools to present complex and competitive information to planners and decision makers [6].

One of the shortest characterizations of BI is to say that BI's objective is to transform *data* into actionable *information*. The difference between the two is important for understanding the core of BI, but it is rarely explained. The main difference between data and information is their meaning and – consequently – their applicability. In the context of BI, “data” has technical meaning, with facets like value, type, basic validity constraints and possibly how it is stored. For example, in case of a value such as “2/29/2016”, it is quite obvious the value is a date and also that it is a valid date.

In contrast, the meaning of “information” in BI must go beyond the technical meaning if the information is to be used for any form of decision making. Pieces of data contribute to new information by becoming linked together and against background knowledge (such as business rules and domain/relationship models). When the example date is linked to a contract (e.g., as the date for delivery of a product) and with specific business rules representing background knowledge (e.g., that delivering after that date involves specific legal and financial ramifications), then all these pieces together become actionable information.

The above perspective is present in what is referred to as the **Business Intelligence Process**, in which processing characteristic of BI is represented as two major phases. The first phase is transforming *data* (obtained from various sources of varying quality and of varying quality) into

reliable *information*. The next phase is mining and analyzing the information to produce *knowledge*, as depicted in the following figure:

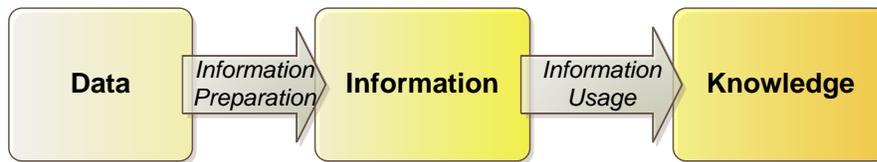


Figure 2-1 Business Intelligence Process

The above perspective distinguishes two major areas of activity in BI: *Information Preparation* and *Information Usage*. On occasion, the scope of BI is treated narrowly as limited to *Information Usage* and the *information preparation* is considered *Data Warehousing (DW)*. A broader and perhaps more common treatment of the scope of BI includes *Information Preparation*. CEAF 2.0 follows the broader approach and thus includes DW. The following figure shows the scope of BI as understood in CEAF:

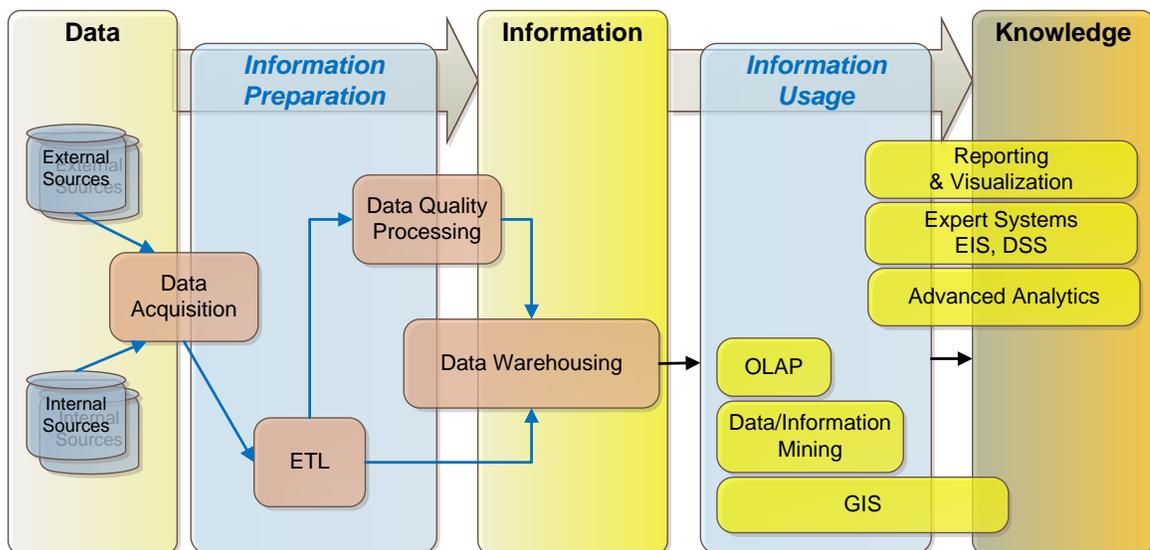


Figure 2-2 Information Preparation and Information Usage

As shown in the above figure, **Preparation of Information** involves processing and transforming data from various sources into analyzable information. This preparation includes the following steps:

- **Data acquisition** (from various sources, internal or external), based on determination of which data is to be loaded, processed, and analyzed
- **Extract, Transform, and Load** (or “ETL”) processing that involves moving the source data into the Data Warehouse
- **Data quality processing**, involving checking quality of the data, cleansing, filtering, and standardizing the data

- **Data warehousing** that connects electronic data from different operational systems so that the data can be queried and analyzed over time for business decision making. This area can be also treated as a subset of Information Management

BI typically operates on heterogeneous data of possibly many types, with varying levels of quality and reliability. In the context of BI, two main data types come to the fore: structured data and semi-structured data:

- The *structured* data have fully specified and fixed internal structure and are typically provided by systems such as OLTP (On-Line Transaction Processing), ERP (Enterprise Resource Planning), MDM (Master Data Management), and similar
- The *semi-structured*¹ data (such as multimedia files, web pages, memos and letters, or e-mails *together* with their meta-data) may have no internal structure or no fixed internal structure. The meaning of the data may be provided by the associated metadata. Semi-structured data are typically provided by Enterprise Content Management systems. The increasing popularity of the Internet as an interaction channel and an important source of information increases importance of semi-structured data for BI

The following figure (adapted from [15]) shows structured and semi-structured data as inputs to BI:

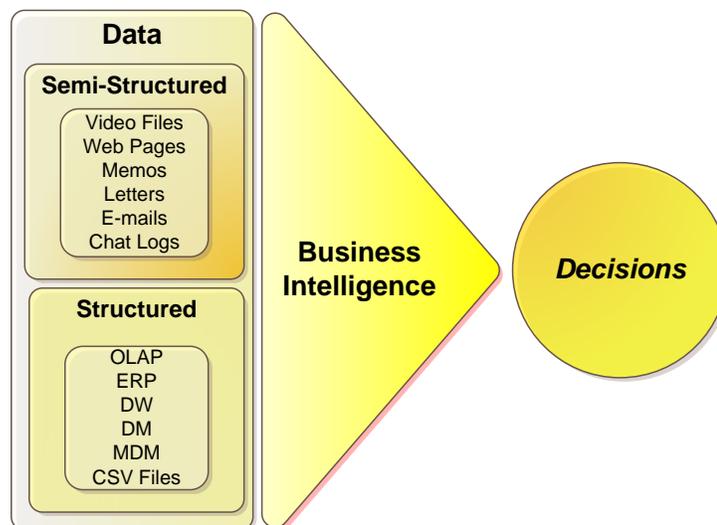


Figure 2-3 Structured and Semi-Structured Data Sources for BI

Information Usage involves accessing and analyzing information with different objectives in mind and using appropriate tools and techniques. The main areas of Information Usage are as follows:

¹ On occasion, *unstructured* data is also mentioned as possible input for BI, but this is questionable: in order for any piece of data to be usable in BI, it must have some information available about its structure and meaning, even if this information is supplied solely by the associated metadata, as it is typically the case even with multimedia resources. In practice then, what is usually meant by *unstructured* data is actually *semi-structured* data.

- Online Analytical Processing (OLAP), which allows users to analyze multidimensional data interactively from multiple perspectives and involving multiple dimensions of data
- Processing of events related to data processing and their lifecycle, including alerting when specific types of events of interest take place
- Data mining, in which multidimensional data sets kept in a data warehouse are analyzed to identify trends, outliers, and patterns
- Creation of canned and ad hoc reports, which may require sophisticated querying and reporting capabilities
- Geographic Information System (GIS), in the broad sense, as the capability of analyzing relationship between data or information and spatial phenomena
- Expert Systems, Decision Support Systems (DSS), Executive Information Systems (EIS), and Knowledge Management Systems (KMS)
- Visualizations are a part of the “presentation layer” of BI. They use different techniques to present the information and the findings and to aid in grasping their meaning and relationships. BI solutions typically offer dashboards and various graphical reports, and similar ways of facilitating interpretation of available information
- Business Analytics – a subset of BI based on statistics, prediction, and optimization based on the information available to BI repositories

2.2 Business Benefits of BI

The main goal for BI systems is to provide decision makers at various levels with reliable, consistent and actionable insights, which are used to achieve the following:

- Identifying and reducing operational costs and waste
- Identifying service performance and improvement/growth opportunities
- Making better, timely business decisions (strategic, tactical, or operational) and, consequently, course corrections, when required
- Reducing time-to-market for new services and products

Recent trends in BI include the broadening of access to the information made available using BI. This involves making the information available and digestible not only at the strategic level to select decision makers, but also for day-to-day activities at operational level to broader audiences. The broadening also involves wider access in terms of devices that can be used for convenient access to the available information, including e.g. mobile “on-the-go” access to the information of interest to specific business roles in the organization.

Market studies have shown that in the last decade the perceived importance of BI for enterprise has been steadily growing and is now comparable to that of security and identity management.

2.3 Key Capabilities of BI Solution

The key capabilities of a successful BI solution include the following:

- Pervasive reach to enterprise-wide users through multiple access channels, including support for mobile access and office integration
- Data acquisition and transformation capabilities, including extracting, cleansing, transforming, quality checking of data from disparate data sources, and supporting multiple data integration patterns

- Support for integration and transformations to visualize information in a way effective for the target audiences, including various forms of presenting information such as Dashboards, Balanced Scorecards, Decision Support and Planning, and Business Performance Management
- Reporting and analytical capabilities, including Drill Down/Roll Up capabilities, Slice/Dice capabilities, predefined and Ad Hoc Reporting and Querying, and OLAP
- Advanced Analytical Capabilities, including Multi-Dimensional Analysis, and Predictive Analytics and Simulation
- Integration capabilities with the key Architectural Building Blocks:
 - Enterprise Portal
 - Enterprise Content Management (ECM) solution
 - Identity and Access Management (IdAM) solution, especially in the area of role-based and/or attribute-based access control to information
 - Enterprise Application Integration (EAI) solution, including support for making BI Services available through the Enterprise Service Bus (ESB)
- Support for common enterprise-wide metadata
- Interoperability of platform components

2.4 Components of BI Solution

This section presents an overview of the main architectural components in a BI solution. The following figure shows main groups of components of BI solution:

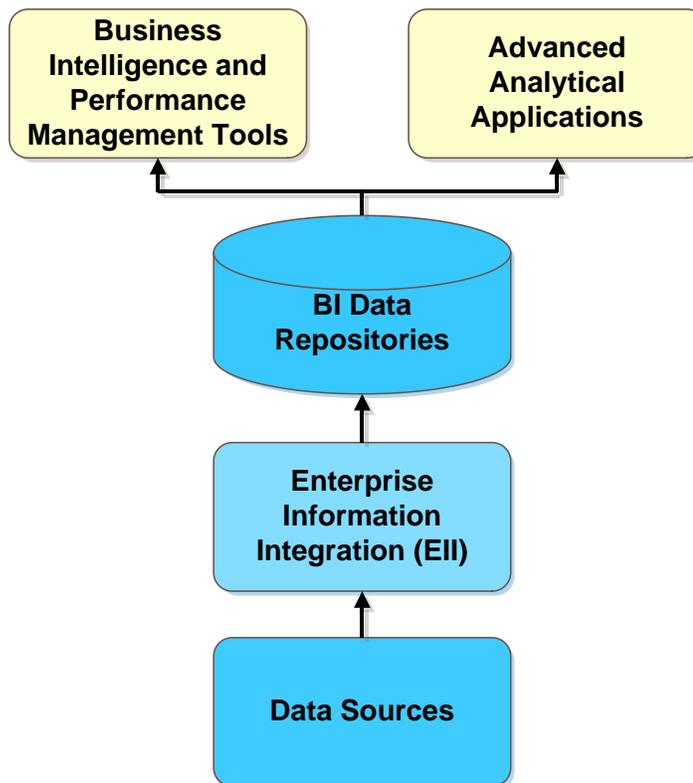


Figure 2-4 BI Solution Components Overview

The above figure shows the following main groups of components:

- Data Sources of various types
- Enterprise Information Integration (EII) platform
- BI Data Repositories
- BI and Performance Management Tools
- Advanced Analytical Applications

The above components are further broken down into their parts and described in the subsections that follow.

2.4.1 Data Sources

Data Sources in BI represent potential sources of business data and they include internal and external systems, and *also* data exchanged through messaging (EAI/EDI). The following figure shows typical types of Data Sources involved in an enterprise-level BI solution:

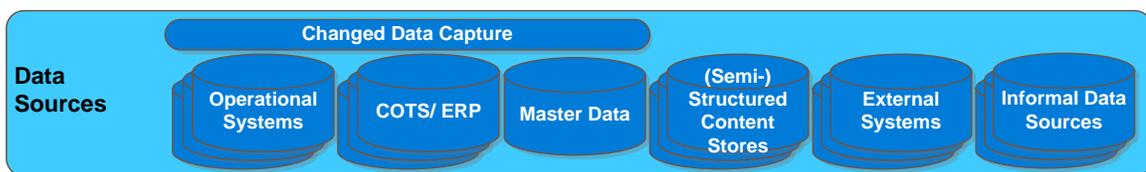


Figure 2-5 Data Sources Components in BI RA

The types of BI Data Sources are described below:

- *Operational Systems* including *COTS/ERP* Systems and *Master Data* Management Systems typically store *structured data* in relational databases. These data sources usually have provisions for capturing *changed data* either through built-in REDO log processes or through add-on components. In some cases, smaller (or older) operational systems may store data of interest to a BI solution in file systems and other *Informal Data Sources* (e.g., text files, spreadsheets etc.).
- *Semi-structured data* are usually persisted primarily as files of a type (or types) appropriate to the content, such as email files, spreadsheets, documents in various formats, etc. The metadata corresponding to these files may be stored in relational databases, content stores, or file systems
- *External Systems* represent the data sources outside of the organization. The data from External Systems is usually available to the organization through on-demand or pre-defined messaging (EAI/EDI), or through bulk data exchange at pre-defined intervals.

The data from these Data Sources may be provided or extracted through bulk data exports, subscription-based messaging, replication, file transfer, or changed data capture through database log mining or source application generated logs.

2.4.2 Enterprise Information Integration (EII) Components

Enterprise Information Integration (EII) platform provides common set of data integration capabilities across enterprise and is a critical component of overall Enterprise Information Management. The following figure shows key components in EII:

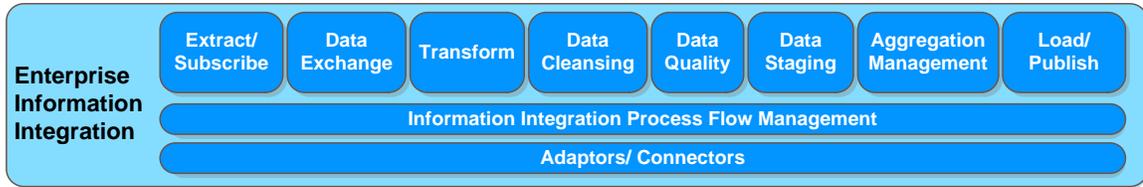


Figure 2-6 EII Components in BI RA

EII typically needs to support multiple styles of data integration, providing the following capabilities:

- Connectivity to various types of BI Data Sources, which typically involves Adaptors and/or Connectors to access these data sources
- Traditional ETL-based integration (bulk extract, transform, and load)
- Capture and apply of changed data, typically involving a form of subscribe and publish (“pub/sub”) mechanisms
- Data cleansing and quality management, based on defined business rules
- Summarization capabilities
- Information integration process flow management that allows for automated execution of the data integration steps (i.e., job) in a defined sequence
- Message-based data integration, which typically involve ESB together with Adaptors and/or Connectors to access various data sources
- Data staging when necessary, for example, to temporarily store data captured through changed data capture, to logically represent an informal data source such as a file or a spreadsheet, and to temporarily store intermediate data before aggregation.

Key capabilities of EII include the following:

- Strong support for meta-data management, in order to support identification, classification, and effective querying of information
- Extensibility, in order to easily accommodate modifications to existing data integration styles or addition of the new required styles
- Scalability, in order to sustain increasing data volumes and loads
- Enterprise level, shareable and multi-tenant capable platform

2.4.3 BI Data Repositories

Data Repositories in BI group all components that provide persistence of data and information relevant to BI. The following figure shows these:

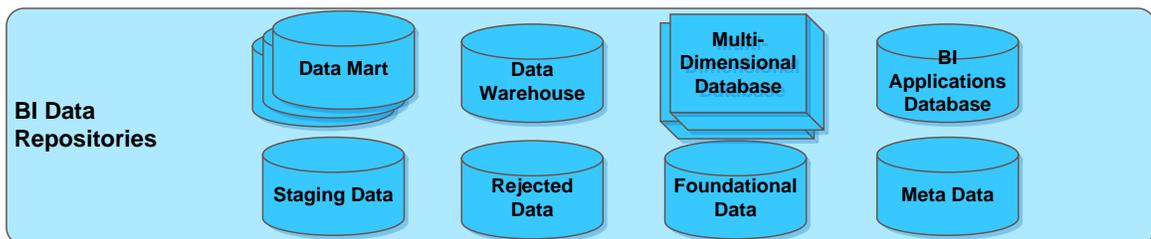


Figure 2-7 BI Data Repositories

As depicted in the above figure, BI Data Repositories include the following:

- **Enterprise Data Warehouse (EDW)**, which integrates operational data from various parts of the business organization and functions as the “single source of truth”. Typically, EDWs are built using (multi-)dimensional data models (star, snow-flake)
- **Data Marts**, which are subsets of the enterprise data warehouse, and which contain information specific to a subject area
- **Multi-Dimensional Databases**, which are subsets of the enterprise data warehouse created by pre-filling aggregations along multiple dimensions (data cubes) at multiple levels of granularity. These databases support rapid multi-dimensional online analytical processing (MOLAP)
- **BI Applications Databases**, which are responsible for a given subject area, typically COTS-specific BI application database, with pre-defined data structures
- **Foundational Data**, which represents an application-neutral enterprise-level relational database. This allows for storing detail-level data warehouse data that will persist over time. Data in this database is typically modeled using normalized (3NF) relational data models. Implementing a Foundational Data repository in a BI solution can simplify the population of the data warehouse
- **Meta-data Stores**, provide a robust way to store metadata objects such as dimensions, hierarchies, measures, performance metrics and report layout objects
- **Staging and Rejected Data**, which are used to temporarily store incremental data captured during Changed Data Capture (CDC) and to temporarily store intermediate data to assist data quality checks and aggregations during ETL processing. This includes *external tables* for querying informal data sources such as flat-files and data rejected during ETL processing.

2.4.4 BI and Performance Management Tools

Business Intelligence and Performance Management tools provide pervasive enterprise-wide reach to BI. The following figure shows main logical components in this group:

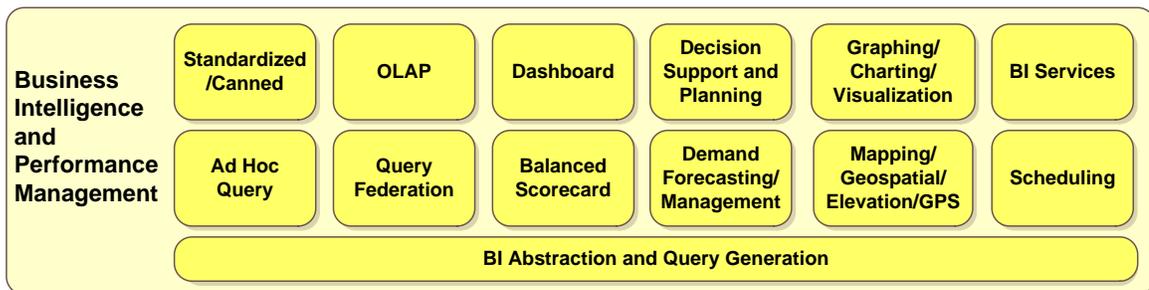


Figure 2-8 Business Intelligence and Performance Management Tools

Brief descriptions of the components shown in Figure 2-8 are provided below:

- **BI Abstraction and Query Generation** provides BI abstraction through an implementation neutral data model using business terminology, and isolates reporting and query functions from physical data structures. This implementation neutral data model forms a robust *semantic layer* to allow users to navigate available data sources and construct ad hoc queries using business terminology. This component uses metadata to map the business terminology to the physical data sources to generate executable queries.
- **Ad Hoc Query** enables users to ask their own questions of the data, without relying on IT team(s) to create reports

- *Standardized/ Canned* reports provide pre-defined reports that can be executed by end users by supplying parameter data. These reports may also be pre-executed with results cached to support high performance and interactive viewing of the reports. The report definitions are usually stored as metadata
- *Online Analytical Processing (OLAP)* enables a style of analysis known as "slicing and dicing" with fast query and calculation performance. It allows users to easily navigate multidimensional drill paths
- *Query Federation* provides the ability to combine data from the enterprise data warehouse to be combined with other data sources to deliver near real-time business intelligence
- *Dashboards* are data visualization tools that display the current status of metrics and key performance indicators (KPIs) for an enterprise. Dashboards provide an at-a-glance view by consolidating and arranging KPIs (usually) on a single screen. They can be configured to display metrics targeted for a single point of view or for a specific role. The essential features of a BI dashboard component include a customizable interface and the ability to pull real-time data from multiple sources
- *Balanced Scorecards* take the metrics displayed in a dashboard a step further by applying them to a strategy map that aligns key performance indicators (KPIs) with a strategic objective. Scorecard metrics should be linked to related reports and information in order to do further analysis
- *Decision Support and Planning* capabilities are the functionalities pertaining to the effective application of BI in support of decision making. Decision making is a fundamental activity to the success of any organization and improving it is one of the main business drivers of a BI investment. The performance measurements provided by the Dashboards and the Balanced Scorecards, and the OLAP capabilities are just one critical piece of the decision process. Support for user designed scenarios/models to simulate possible performance outcomes and collaboration capabilities are the other capabilities provided by the BI and Performance Management tools for decision support and planning
- Supporting *Demand Forecasting and Management*
- Supporting *Graphing/Charting and Visualization* facilities
- Supporting *GIS service* consumption
- Supporting *batch scheduling* of pre-defined reports
- Providing *BI Services* through integration with Enterprise Application Integration Bus/Hub

2.4.5 Advanced Analytics Components

Advanced Analytics groups components required to support complex data analysis scenarios and uses cases. The following figure shows these components:



Figure 2-9 Advanced Analytics Components in BI

Advanced Analytics contain the following components:

- Data Discovery and Mining
- Modeling facilities
- Simulation capabilities

- Analytic Applications, which are dedicated applications for specific data analysis scenarios
- Predictive Analytics
- Mathematical Analysis of information, including statistics
- Multi-dimensional Analysis

2.4.6 Operational Components

Several operational components support a BI solution. These are further classified into the following subgroups:

- Data Management components
- Security components
- Development and Administration tools

Data Management components are responsible for standard operations on data in the enterprise. The following figure shows main Data Management components relevant to BI:



Figure 2-10 Data Management Components in BI

The above figure shows the following components:

- Data Backup
- Archive and Restore
- Data Recovery
- Data Classification
- Meta-Data Management

A BI implementation requires a number of other supporting components. The following figure shows those:



Figure 2-11 Supporting Components in BI

As data and information are organization's key assets, they have to be protected from unauthorized access. Protection of BI solutions should be done through integration with enterprise IdAM facilities without requiring separate mechanisms to create identities and enforce access.

Information produced by BI can be potentially routed through a number of channels, including direct routing to users through interfacing with portals, web sites to make the information available to desktop or mobile clients.

BI solutions typically require dedicated administration tools and associated development tools. These include:

- Data modeling tools
- Information integration flow creation and implementation tools



- Metadata creation and management tools
- Performance and availability monitoring tools

Some of the above tools support other types of solutions in addition to BI. Therefore, enterprise thinking in the selection of these tools is important to control technical diversity and simplify overall administration.



3 Business Intelligence Reference Architecture Description

This section describes BI Reference Architecture using three views:

- Conceptual View, which provides a summary of logical-level building blocks for BI as presented in the Section 2 above
- Logical View, which provides an overview of relationships and interactions between components in a BI solution for specific usage scenarios
- Deployment View, which illustrates the distribution of processing and components across nodes in the system.

Each of the above views is presented in the subsections that follow.

3.1 BI RA Conceptual View

The following Conceptual View figure brings together all major components of a BI solution that have been already described in the section “Components of BI Solution”. The diagram shows key Business Intelligence components organized in subsequent horizontal layers. The layers placed vertically represent components that can be used in a number of horizontal layers (such as e.g. Data Management components or security- and access management-related components).

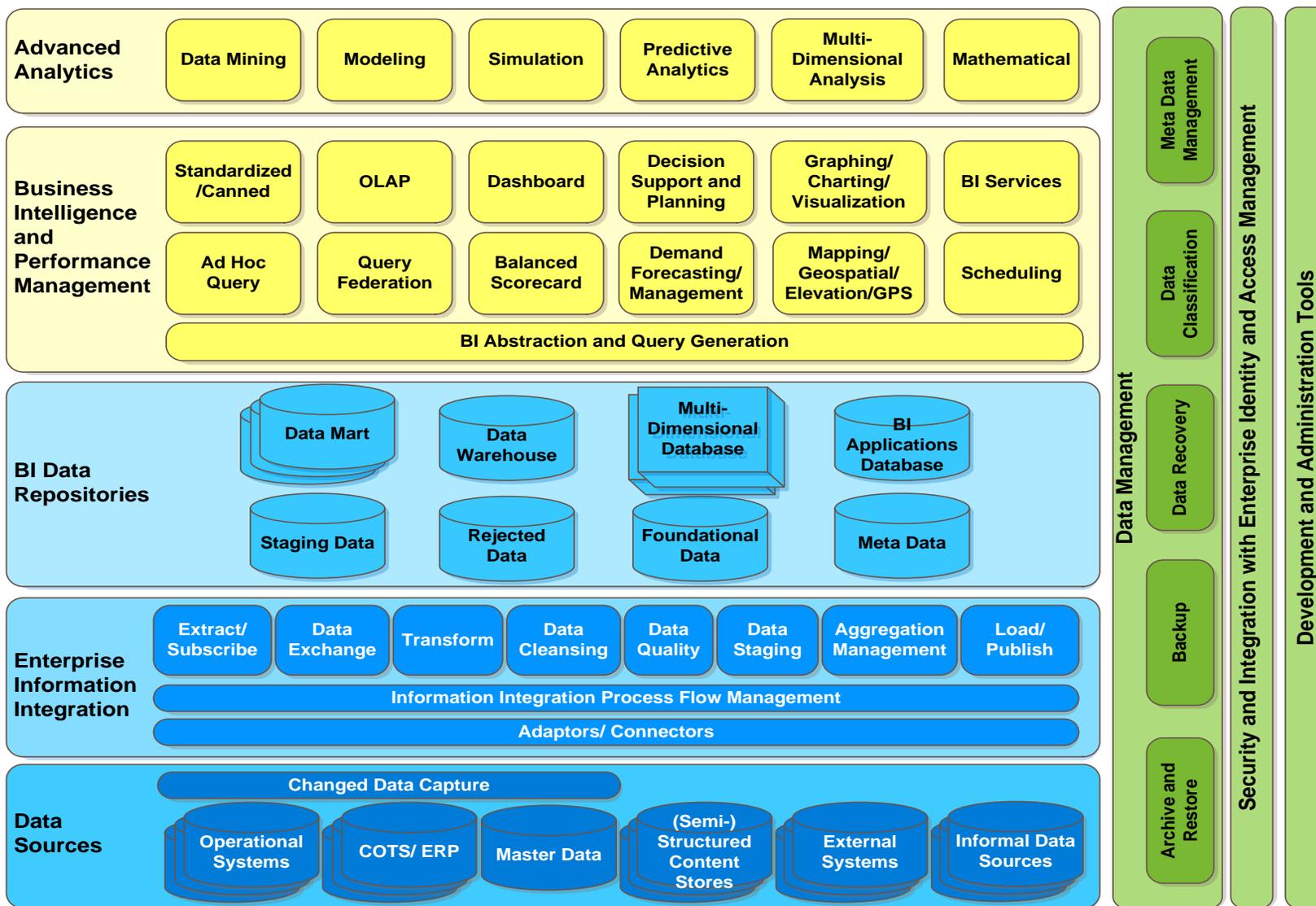


Figure 3-1 BI RA Conceptual View



Please note that the conceptual view of BI should not be considered as separate from other Reference Architectures presented in CEAF 2.0. There are aspects of BI Reference Architecture that involve a connection with other RAs, including the following:

- SOA RA provides the overall underlying pattern of service orientation and composition of services
- Identity and Access Management (IdAM) RA provides elaboration of security aspects
- Enterprise Content Management (ECM) RA discusses management of unstructured and semi-structured data and making it available using various access channels
- Enterprise Application Integration (EAI) RA focuses on interoperability and the role of ESB
- Master Data Management (MDM) RA provides insight into master data management and data quality.

3.2 BI RA Logical View

This section shows relationships and main interactions between main components of a BI solution. The interactions are shown, at a high level, using a generic Information Preparation scenario.

It should be noted that the interactions shown in this section are logical abstractions representing typical interactions. In practice, however, due to the physical packaging of BI logical components and services in the *specific system software products* selected, the component interactions (and also the names of components and services) may change. Therefore, this section is intended to enable the reader understand how BI components logically realize a given scenario or use case, and use that knowledge to evaluate specific technology choices to provide desired capabilities.

3.2.1 Information Preparation Scenario

As already discussed (in the section “Definitions”), *Preparation of Information* involves processing and transforming data from various sources into analyzable information, typically stored in an Enterprise Data Warehouse (EDW). The process includes the following steps:

- Data acquisition
- Extract, Transform, and Load
- Data quality processing
- Data warehousing

The following diagram shows key interactions between BI components for realizing the *Information Preparation* scenario.

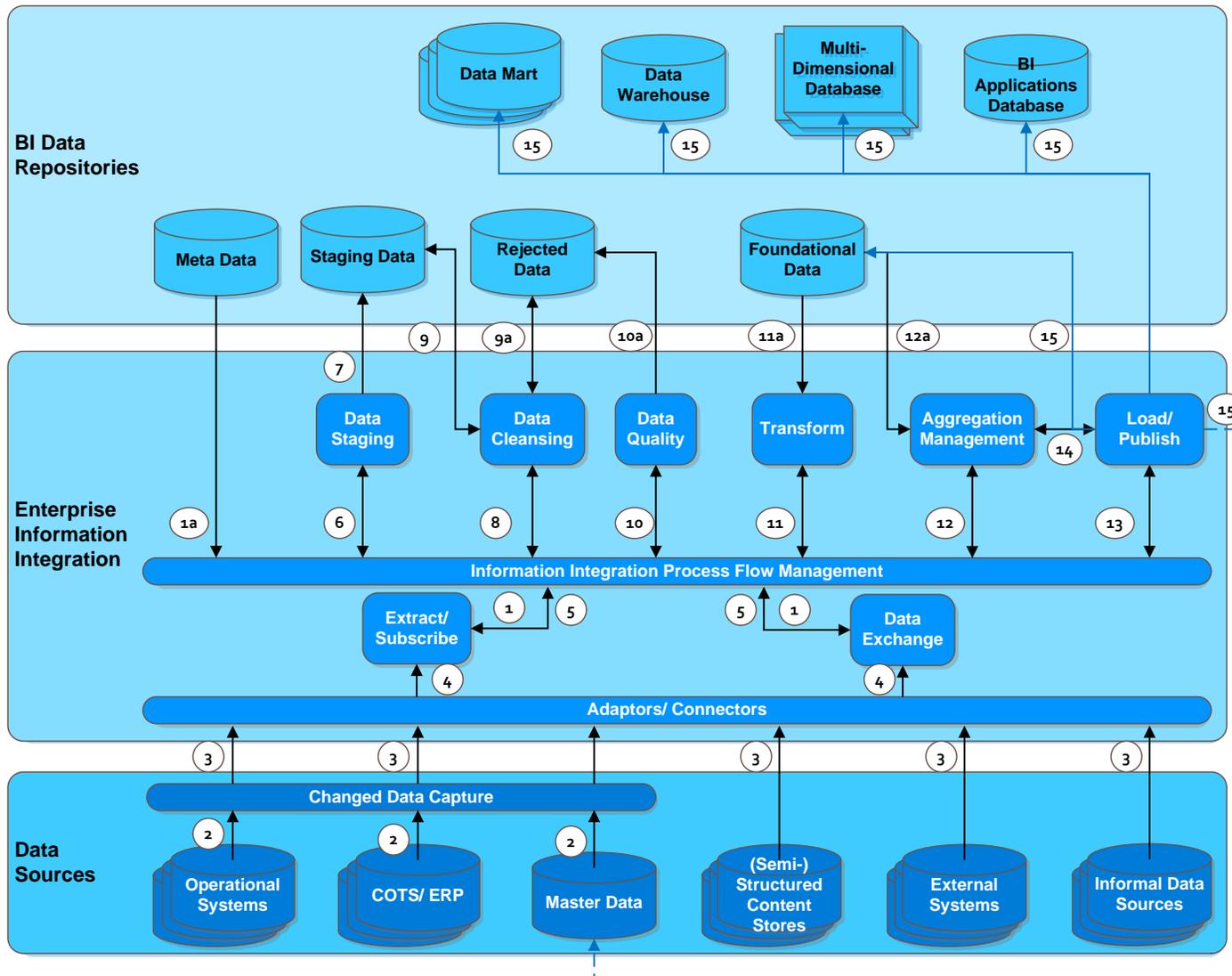


Figure 3-2 Service and Component Interactions in Information Preparation Scenario

The following table provides description of responsibilities for the services shown in the above figure:

Table 3-1 Component Interactions in Information Preparation Scenario

Label	Description
(1), (1a)	A pre-defined Information Integration Process Flow (or job) is initiated by the EII platform. The process configuration is obtained from the metadata.
(2), (3)	Data is acquired from various data sources using data source specific Connectors/ Adaptors through Changed Data Capture, Replication, Query or other data acquisition mechanisms.
(4)	Represents Data Extraction and Data Exchanges.
(5)	Extracted data is available for further processing in the Information Integration Process Flow.
(6), (7)	Extracted data is typically (temporarily) staged for further processing
(8), (9), (9a)	Data Cleansing rules are applied to validate and cleanse the extracted data. Some extracted data may be rejected during this step.
(10), (10a)	Data Quality rules are applied to the extracted and staged data. Some extracted data may be rejected during this step.
(11), (11a)	Extracted Data is transformed to the target format. If a foundational data source is implemented, then the transformation is typically performed to convert the source formats to the foundational data model.
(12)	During this step, defined aggregations are performed.
(13), (14)	The load step is initiated to load the quality checked and transformed source data into the BI repositories.
(15)	Various BI repositories as shown in the figure are populated during this step. Additionally Master Data repository may be updated with any new or changed master data.

3.3 BI RA Deployment View

This subsection is to be completed in a future release. It is intended to show best-practice-based system topologies and implementation patterns of BI, based on existing realizations of the BI RA in the state.



4 Glossary

Business Intelligence is “an umbrella term that includes the applications, infrastructure and tools, and best practices that enable *access to* and *analysis of* information to improve and optimize decisions and performance”.

Decision Support System (DSS) is an information system designed to support activities of decision makers by providing interactive capabilities to analyze and visualize available information.

Executive Information Systems (EIS) are specialized form of DSS tailored for the needs of senior executives to identify opportunities and problems.

Competitive Intelligence (CI) is legal and ethical form of gathering, analyzing, and distributing information about the competitive market – other organizations, available products, and customers.

Geographic Information Systems (GIS) are information systems that support querying, analysis, visualization of data involving spatial phenomena in order to reveal relationships, patterns, and trends.

Knowledge Management System (KMS) are information systems used in an organization to identify, create, represent, distribute, and enable adoption of insights and experiences relevant to functioning of the organization.

On-Line Analytic Processing (OLAP) is an approach in information systems to support execution of multi-dimensional queries, usually interactively.

Reference Architecture models the abstract architectural elements in the domain independent of the technologies, protocols, and products that are used to implement the domain.

Service Component is an actual application, program or subsystem providing implementation of a Service treated as a contract and performing specific responsibilities.



5 References

State and Federal Documents

- A. State of California, California State Information Technology Strategic Plan, 2013 Update
- B. State of California, California Enterprise Architecture Framework, Version 2.0

Books and Papers

1. Tang Liyang; Ni Zhiwei; Wu Zhangjun; Wang Li, "A Conceptual Framework for Business Intelligence as a Service (SaaS BI)," *Intelligent Computation Technology and Automation (ICICTA)*, 2011 International Conference on , vol.2, no., pp.1025,1028, 28-29 March 2011, doi: 10.1109/ICICTA.2011.541
2. Liya Wu; Barash, G.; Bartolini, C., "A Service-oriented Architecture for Business Intelligence," *Service-Oriented Computing and Applications, 2007. SOCA '07. IEEE International Conference on* , vol., no., pp.279,285, 19-20 June 2007, doi: 10.1109/SOCA.2007.6
3. "Business Intelligence as a Shared Service. An Oracle White Paper", March 2010, <http://www.oracle.com/us/corporate/insight/bi-shared-service-wp-171717.pdf>
4. Liyang, Tang; Zhiwei, Ni; Zhangjun, Wu; Li, Wang, "A Conceptual Framework for Business Intelligence as a Service (SaaS BI)", 2011 International Conference on Intelligent Computation Technology and Automation (ICICTA), doi: 10.1109/ICICTA.2011.541
5. Watson, H.J.; Wixom, Barbara H., "The Current State of Business Intelligence," *Computer*, vol.40, no.9, pp.96,99, Sept. 2007, doi: 10.1109/MC.2007.331
6. S. Negash, "Business Intelligence", *Communications of the Association for Information Systems (Volume13, 2004)* 177-19
7. Magic Quadrant for Business Intelligence Platforms, Gartner, 6 February 2012
8. Magic Quadrant for Data Integration Tools, Gartner, 17 July 2013

Referenced Web Sites and Resources

- a. Gartner IT Glossary, <http://www.gartner.com/it-glossary>
- b. Power, D.J. *A Brief History of Decision Support Systems*. DSSResources.COM, World Wide Web, <http://DSSResources.COM/history/dsshhistory.html>, version 4.0, March 10, 2007



6 Document History

Table 6-1 Document History

Release	Description	Date
Version 1.0 Draft	Initial creation	05/31/2013
Version 1.0 Second Draft	Revised based on internal review comments	07/22/2013
Version 1.0 Final Draft	Addressed EAC review comments	10/21/2013
Version 1.0 Final	Final version	01/02/2014