An Oracle White Paper

September 2013

Oracle Enterprise Transformation Solutions Series

Big Data & Analytics Reference Architecture

ORACLE

Executive Overview

Data is often considered to be the crown jewels of an organization. It can be used in myriad ways to run the business, market to customers, forecast sales, measure performance, gain competitive advantage, and discover new business opportunities. And lately, a convergence of new technologies and market dynamics has opened a new frontier for information management and analysis.

This new wave of computing involves data with far greater volume, velocity, and variety than ever before. Big Data, as it is called, is being used in ingenious ways to predict customer buying habits, detect fraud and waste, analyze product sentiment, and react quickly to events and changes in business conditions. It is also a driving force behind new business opportunities.

Most companies already use analytics in the form of reports and dashboards to help run their business. This is largely based on well structured data from operational systems that conform to pre-determined relationships. Big Data, however, doesn't follow this structured model. The streams are all different and it is difficult to establish common relationships. But with its diversity and abundance come opportunities to learn and to develop new ideas – ideas that can help change the business.

To run the business, you organize data to make it do something specific; to change the business, you take data as-is and determine what it can do for you. These two approaches are more powerful together than either alone. In fact, many innovative solutions are a combination of both approaches.

For instance, a major European car manufacturer is collecting data via telematics from cars they produce. This data is used to influence offers they make to their customers. It is also used to better understand the conditions that the car has experienced, which in turn helps in root-cause failure analysis as well as in future automobile design.

The architectural challenge is to bring the two paradigms together. So, rather than approach Big Data as a new technology silo, an organization should strive to create a unified information architecture – one that enables it to leverage all types of data, as situations demand, to promptly satisfy business needs. This is the approach taken by a large worldwide bank. They are using a common information architecture design to drive both their real-time trading platforms and their batch reporting systems.

The objective of this paper is to define and describe a reference architecture that promotes a unified vision for information management and analytics. The reference architecture is defined by the capabilities an organization needs and a set of architecture principles that are

commonly accepted as best practices in the industry. It is described in terms of components that achieve the capabilities and satisfy the principles. Oracle products are mapped to the architecture in order to illustrate how the architecture can be implemented and deployed. Organizations can use this reference architecture as a starting point for defining their own unique and customized architecture.

Introduction

In order to approach Big Data and analytics holistically, it is important to consider what that means. The strategy used to develop this reference architecture includes three key points to set the context:

1. **Any data, any source.** Rather than differentiate Big Data from everything else (small data?), we want to view data in terms of its qualities. This includes its degree of structure, volume, method of acquisition, historical significance, quality, value, and relationship to other forms of data. These qualities will determine how it is managed, processed, used, and integrated.
2. **Full range of analytics.** There are many types of analysis that can be performed, by different types of users (or systems), using many different tools, and through a variety of channels. Some types of analysis require current information and others work mostly with historical information. Some are performed proactively and others are reactive. The architecture design must be universal and extensible to support a full range of analytics.
3. **Integrated analytic applications.** Intelligence must be integrated with the applications that knowledge workers use to perform their jobs. Likewise, applications must integrate with information and analysis components in a manner that produces consistent results. There must be consistency from one application to another, as well as consistency between applications, reports, and analysis tools.

This reference architecture is designed to address key aspects of these three points. Specifically, the architecture is organized into views that highlight three focus areas: universal information management, real-time analytics, and intelligent processes. They represent architecturally significant capabilities that are important to most organizations today.

Reference Architecture Conceptual View

The conceptual view for the reference architecture, shown in Figure 1, uses capabilities to provide a high-level description of the Big Data and Analytics solution.

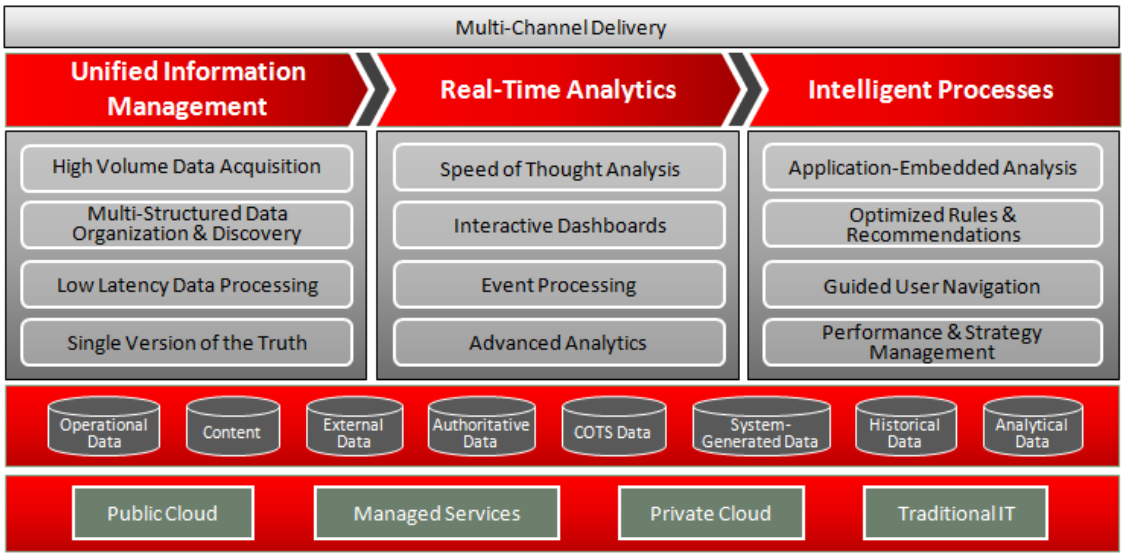

Figure 1. Big Data & Analytics Reference Architecture Conceptual View

The top layer of the diagram illustrates support for the different channels that a company uses to perform analysis or consume intelligence information. It represents delivery over multiple channels and modes of operation: stationary and mobile, (network) connected and disconnected.

Focus Areas

This paper concentrates on three important aspects of the Big Data and analytics architecture: Unified Information Management, Real-Time Analytics, and Intelligent Processes. Each of these focus areas is further detailed below.

It should be noted that although the reference architecture is organized into these three focus areas, the solution cannot be implemented as silos of functionality. Rather, the complete solution must incorporate all aspects of the reference architecture in a cohesive manner.

Unified Information Management

Unified Information Management addresses the need to manage information holistically as opposed to maintaining independently governed silos. At a high level this includes:

- High Volume Data Acquisition – The system must be able to acquire data despite high volumes, velocity, and variety. It may not be necessary to persist and maintain all data that is received. Some may be ignored or discarded while others are kept for various amounts of time.

- Multi-Structured Data Organization and Discovery – The ability to navigate and search across different forms of data can be enhanced by the capability to organize data of different structures into a common schema. Using this form of organization, the system can relate structured data such as model numbers and specifications, semi-structured data such as product documents, and unstructured data such as installation videos. In addition, new business opportunities can be discovered by looking at different forms of data in new ways.
- Low Latency Data Processing – Data processing can occur at many stages of the architecture. In order to support the processing requirements of Big Data, the system must be fast and efficient.
- Single Version of the Truth – When two people perform the same form of analysis they should get the same result. As obvious as this seems, it isn't necessarily a small feat, especially if the two people belong to different departments or divisions of a company. Single version of truth requires architecture consistency and governance.

Real-Time Analytics

Real-Time Analytics enables the business to leverage information and analysis as events are unfolding. At a high level this includes:

- Speed of Thought Analysis – Analysis is often a journey of discovery, where the results of one query determine the content of the next. The system must support this journey in an expeditious manner. System performance must keep pace with the users' thought process.
- Interactive Dashboards – Dashboards provide a heads-up display of information and analysis that is most pertinent to the user. Interactive dashboards allow the user to immediately react to information being displayed, providing the ability to drill down and perform root cause analysis of situations at hand.
- Advanced Analytics – Advanced forms of analytics, including data mining, machine learning, and statistical analysis enable businesses to better understand past activities and spot trends that can carry forward into the future. Applied in real-time, advanced analytics can enhance customer interactions and buying decisions, detect fraud and waste, and enable the business to make adjustments according to current conditions.
- Event Processing – Real-time processing of events enables immediate responses to existing problems and opportunities. It filters through large quantities of streaming data, triggering predefined responses to known data patterns.

Intelligent Processes

A key objective for any Big Data and Analytics program is to execute business processes more effectively and efficiently. This means channeling the intelligence one gains from analysis directly into the processes that the business is performing. At a high level this includes:

- Application-Embedded Analysis – Many workers today can be classified as knowledge workers; they routinely make decisions that affect business performance. Embedding analysis into the applications they use helps them to make more informed decisions.
- Optimized Rules and Recommendations – Automated processes can also benefit from analysis. This form of business process executes using pre-defined business logic. With optimized rules and recommendations, insight from analysis is used to influence the decision logic as the process is being executed.
- Guided User Navigation – Some processes require users to take self-directed action in order to investigate an issue and determine a course of action. Whenever possible the system should leverage the information available in order to guide the user along the most appropriate path of investigation.
- Performance and Strategy Management – Analytics can also provide insight to guide and support the performance and strategy management processes of a business. It can help to ensure that strategy is based on sound analysis. Likewise, it can track business performance versus objectives in order to provide insight on strategy achievement.

Information

The Big Data and Analytics architecture incorporates many different types of data, including:

- Operational Data – Data residing in operational systems such as CRM, ERP, warehouse management systems, etc., is typically very well structured. This data, when gathered, cleansed, and formatted for reporting and analysis purposes, constitutes the bulk of traditional structured data warehouses, data marts, and OLAP cubes.
- COTS Data – Custom off-the-shelf (COTS) software is frequently used to support standard business processes that do not differentiate the business from other similar businesses. COTS applications often include analytical packages that function as pre-engineered data marts. COTS analytical data, transformed from operational data, can also be incorporated into the data warehouse to support analysis across business processes.
- Content – Documents, videos, presentations, etc., are typically managed by a content management system. These forms of information can be linked to other forms of data to support navigation, search, analysis, and discovery across data types.

- Authoritative Data – Authoritative data refers to very high quality data that is used to provide context to operational data. It includes master data - standardized key business entities such as Customer and Product, and reference data - classification data elements such as status codes and currency codes. Authoritative data is also used within the data warehouse.
- System-Generated Data – Data such as system logs, RFID tags, and sensor output are forms of Big Data that must be captured, organized, and analyzed. This data often originates from within the organization and has historically been overlooked in terms of business analytics value.
- External Data – Other common sources of Big Data tend to originate from outside of the organization. These include social media feeds, blogs, and independent product and service ratings.
- Historical Data – The data warehouse environment must maintain data for historical purposes. Historical Data refers to data that is organized to accommodate large volumes and structured to easily accommodate business changes without schema revisions.
- Analytical Data – The data warehouse environment also needs to support analytics. Analytical data refers to data that is structured to provide easy access using analytical tools and to perform well for analytical queries. For structured data analysis, analytical data often takes the form of dimensional data models and OLAP cubes. Although some types of analytics can be performed on historical data models, it is sometimes necessary to establish a subset of historical data that is filtered and optimized for analysis.

Historical Data and Analytical Data are two broad categories that describe data in a data warehouse. They are not specific to traditional, structured data or Big Data, rather they indicate a separation of concerns between historical record and efficient analytics. Often, organizations that ignore this separation of concerns tend to have difficulties accomplishing both objectives with a single-purpose solution.

Deployment

There are more options today for where to deploy a solution than ever before. At a high level the four options for deployment of architecture components are:

- Public Cloud – In the public cloud model, a company rents resources from a third party. The most advanced usage of public cloud is where the business functionality is provided by the cloud provider (i.e., software-as-a-service). Public cloud might also be used as the platform upon which the business functionality is built (i.e., platform-

as-a-service), or the public cloud may simply provide the infrastructure for the system (i.e., infrastructure-as-a-service).

- Private Cloud - Private cloud is the same as public cloud, but the cloud is owned by a company instead of being provided by a third party. Private clouds are ideal for hosting and integrating very large data volumes while keeping data secure behind corporate firewalls.
- Managed Services – In this model a company owns the components of the system, but outsources some or all aspects of runtime operations.
- Traditional IT – In this model a company owns and operates the system.

These various options for deployment are not mutually exclusive. The solution might be deployed in two or more different ways. Not only might the functional areas be deployed differently (e.g. data warehouse as Managed Services with certain operational systems deployed to a Public Cloud), but even a single functional area might span deployment options. For example, external Big Data capture may start with infrastructure hosted in a public cloud but then transition to an internal system (traditional IT) when business value and additional requirements make an infrastructure purchase more feasible.

Architecture Principles

In essence, architecture principles translate business needs into IT mandates that the solution must meet. Because architecture principles span the entire solution, they are at a much higher-level than functional requirements. Establishing architecture principles drives the overall technical solution. Some key architecture principles for the Big Data and Analytics solution are provided below.

Accommodate All Forms of Data	
Statement	The architecture must accommodate all forms of data that are of value to the business.
Rational	Business analytics can be performed using data in many different forms, from various sources, and with varying degrees of structure. The architecture must be flexible enough to support different forms of data in a manner that best supports analysis while being efficient and cost-effective.
Implications	• Capture, process, organize, and analyze all forms of data in order to meet existing business requirements and support discovery of new business opportunities. • Impose the proper amount of structure to each form of data. Some forms may be highly structured, which requires a relatively high degree of up-front modelling, cleansing, and formatting. Other forms may have

	minimal structure, which requires greater effort on the consumer for interpretation and processing. • Maintain relationships between different forms of data, and enable navigation between data of different structures.

Consistent Information and Object Model	
Statement	The system must present a single version of truth whereby the results of analysis are consistent between users and departments across the organization. In addition, the system must enable analysis to be shared in a manner that promotes a single version of the question.
Rational	Business analytics has much greater value when the results of analysis are consistent and can be duplicated. Likewise, analytics can be applied to a greater audience when analysis objects (graphs, charts, etc.) can be designed by subject matter experts and re-used by all knowledge workers.
Implications	• Virtualization capabilities will be required in order to aggregate data from multiple sources. • Maintain conformity of dimensions and facts across dimensional data stores. • Provide a means to catalog, define, and share analysis objects.

Integrated Analysis	
Statement	Information and analysis must be available to all users, processes, and applications across the organization that can benefit from it.
Rational	The reach of decision-making analysis must expand to include all knowledge workers in the organization and the applications they use.
Implications	• Integrated analysis into UIs, devices, and processes such that users gain insight where and when they need it. • Integrated analysis with business processes in a way to automatically leverage the available information to optimize operational processes. • Enable end users who are not familiar with data structures and BA tools to view information pertinent to their needs.

Insight to Action	
Statement	The system must provide the ability to initiate actions based on insight that is gained through analysis.
Rational	Analysis is most useful when it is actionable. It is important for an organization to link the results of analysis to actions that are taken. Otherwise, some of the value that analysis provides will be lost when

	users fail to take appropriate action. This may be for a variety of reasons including a breakdown in communications, negligence, or lack of knowledge.
Implications	• Support proactive forms of analysis such as monitoring data streams, detecting events, periodically querying various forms of information, and performing analysis to detect particular conditions of concern. • Alert users when events are detected and enable users to subscribe to different types of events. • Guide users to the appropriate applications, processes, and interfaces from which they can take action when action is necessary.

These are example architecture principles that a company might embrace. Ultimately it is up to each company to define the appropriate architecture principles to ensure that the Big Data and Analytics solution meets the needs of the business while furthering the strategic direction of IT.

Reference Architecture Logical View

The Oracle Reference Architecture (ORA) defines a multi-tier architecture template that can be used to describe many types of technology solutions. Figure 2 illustrates how major components of the Big Data and Analytics architecture fit within the universal ORA high-level structure.

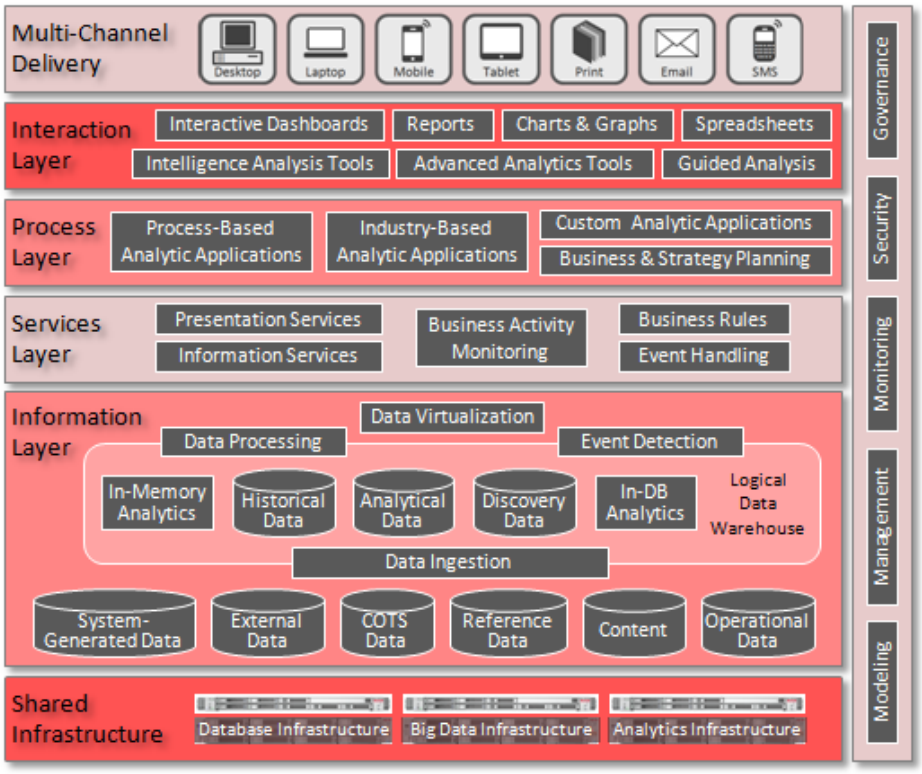

Figure 2. Reference Architecture High-Level Logical View

At the base of the reference architecture is the Shared Infrastructure Layer. This layer includes the hardware and platforms on which the Big Data and Analytics components run. As shared infrastructure, it can be used to support multiple concurrent implementations, in support of, or analogous to, Cloud Computing. This layer includes infrastructure to support traditional databases, specialized Big Data management systems, and infrastructure that has been optimized for analytics.

The Information Layer includes all information management components, i.e. data stores, as well as components to capture, move, integrate, process, and virtualize data. At the bottom are data stores that have been commissioned for specific purposes, such as individual operational data stores, content management systems, etc. These data stores represent sources of data that are ingested (upward) into the Logical Data Warehouse (LDW). The LDW represents a collection of data that has been provisioned for historical and analytical purposes. Above the LDW are components that provide processing and event detection for all forms of data. At the top of the layer are components that virtualize all forms of data for universal consumption.

The Services Layer includes components that provide or perform commonly used services. Presentation Services and Information Services are types of Services in a Service Oriented Architecture (SOA). They can be defined, cataloged, used, and shared across solutions. Business Activity Monitoring, Business Rules, and Event Handling provide common services for the processing layer(s) above.

The Process Layer represents components that perform higher level processing activities. For the purpose of Big Data and Analytics, this layer calls out several types of applications that support analytical, intelligence gathering, and performance management processes.

The Interaction Layer is comprised of components used to support interaction with end users. Common artifacts for this layer include dashboards, reports, charts, graphs, and spreadsheets. In addition, this layer includes the tools used by analysts to perform analysis and discovery activities.

The results of analysis can be delivered via many different channels. The architecture calls out common IP network based channels such as desktops and laptops, common mobile network channels such as mobile phones and tablets, and other channels such as email, SMS, and hardcopy.

The architecture is supported by a number of components that affect all layers of the architecture. These include information and analysis modeling, monitoring, management, security, and governance.

Subsequent sections in this white paper further detail the logic view of the reference architecture. Each of the three primary focus areas from the conceptual view (Figure 1) is shown in greater detail to illustrate and describe the components that are required to fully support the capabilities.

Information Management Components of the Logical Architecture

Figure 3 focuses on the unified information management aspects of the Big Data and Analytics reference architecture. It is divided into three layers: Information Provisioning, Information Delivery, and Information Consumption. This layering promotes a separation of concerns between the way information is acquired and managed and the way it is accessed. It allows information to be managed holistically while enabling access to appropriate users via the best means possible.

Note that this layering represents an information-centric view of the architecture, whereas ORA, described in the previous section, is solution-centric. As a result, there isn't a one-to-one relationship between the layers despite the similarity in component layout.

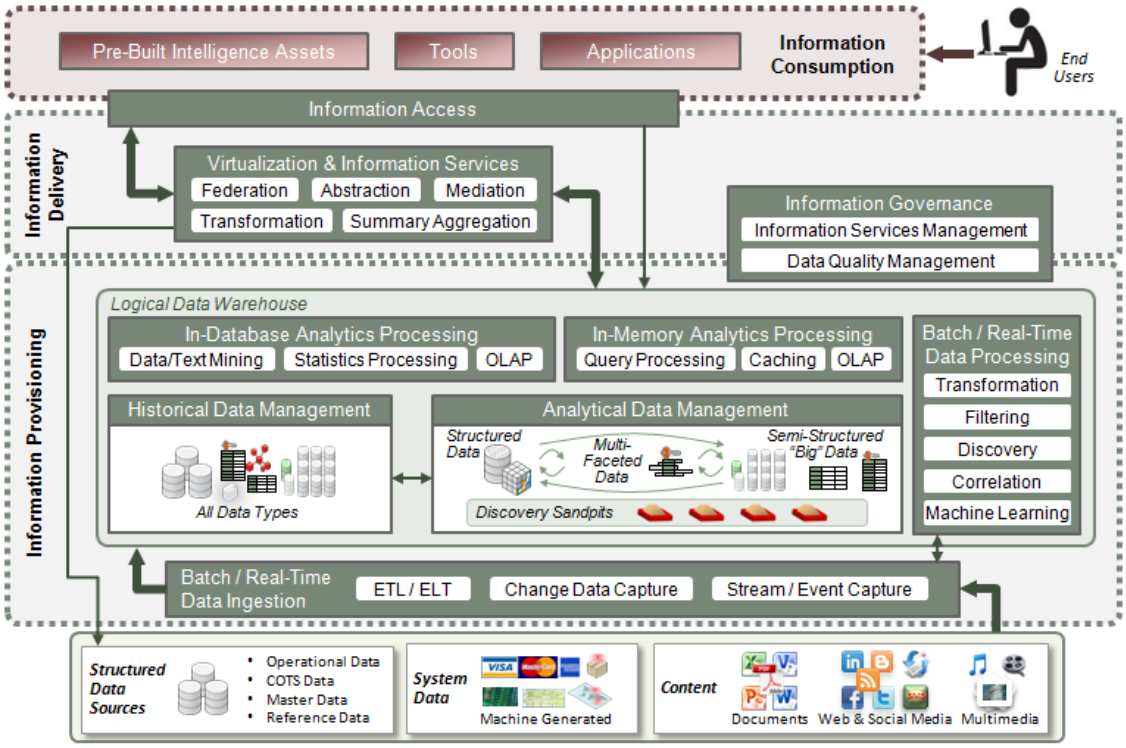

Figure 3. Unified Information Management Components of the Logical Architecture

Central to the architecture are the components that comprise the Logical Data Warehouse (LDW). This includes all types of data stores that maintain information for historical and analytical purposes. The historical data stores consolidate large quantities of information in a manner that best maintains historical integrity. Information may be stored using many different technologies and schema designs, such as relational, spatial, and graph databases, distributed file systems, and NoSQL databases. Analytical data stores are designed to support analysis by maximizing ease of access and query performance. Technologies and schema designs will often consist of dimensional data models, OLAP cubes, multi-faceted data models, distributed file systems, and NoSQL databases.

Dimensional data models and OLAP cubes are generally used for structured data analysis. They support most of the operational and performance reporting requirements of the business. Distributed file systems and NoSQL databases are most frequently used to store semi-structured or unstructured data such as system-generated data, blog data, content, etc. These types of data stores impose little or no requirements on data formats, thus supporting the collection of many different types of data without intrusive data formatting and restructuring processes. In addition, they tend to be low-cost, high volume data stores, further enabling the storage of Big Data.

Multi-faceted data models are designed to handle both structured and semi-structured data. They support the organization of data using a common set of keys, while allowing each data record to contain unique data elements. This form of record linking enables search and navigation across data types without the need to restructure data into a fixed schema.

The LDW is populated with many different types of data from a variety of sources. This includes structured data sources such as operational data, COTS application data, reference data, and master data, as well as system-generated data and some forms of content. Data can be moved (ingested) into the LDW using a combination of batch or real-time methods. Traditional extract, transform, and load (ETL) processes, or the extract, load, transform (ELT) variant, are frequently used for batch data transfer. Change data capture (CDC), supported by some relational database management systems, enables the propagation of changes from source systems to the LDW in near real time. Stream and event capture systems can be used for real-time data collection, filtering, and capture.

Once data is loaded into the LDW it can be moved from one area to another or from one type of storage to another. Typically, data will be cleansed and moved into the historical data management area for long term storage. There it is maintained in a format that is ideally suited for resilience, (to avoid schema changes over time as the business changes and evolves), and space utilization, (to minimize storage costs). Data is frequently copied, filtered, and transformed to populate data stores in the analytical data management area. There it is maintained in a format that is best suited for analysis. Derived data – such as quartiles that are generated as a result of analysis - may be copied back into the historical data management area to be saved as part of the historical record.

Data may be loaded into one type of storage, processed, and moved to another type of storage. For example, sensor data may be captured in a distributed file system, processed, filtered, and loaded into a relational database or NoSQL data store where they can be analyzed using structured data analysis tools. Likewise, structured data can be extracted from relational databases and loaded into distributed file systems for batch processing along with other types of data.

The LDW includes both data management and data processing capabilities. This combination of features allows processing to occur as close to the data as possible. Since some functions involve large quantities of data, it is important to perform them without transferring data to and from the data stores.

Processing can be divided into two categories: higher level analytics processing and lower level data processing. The higher level analytics processing is either performed within the database or within an in-memory database. In-database processing includes common

analytical functions such as data mining, text mining, statistical analysis, and OLAP. In-memory capabilities include high speed query processing, results caching, and OLAP. Lower level data processing can be performed in support of analytical processing, data ingestion, or other functions such as data cleansing and discovery processing.

While the LDW primarily handles information provisioning and processing, other components of the architecture are tasked with information delivery. These components are labeled Information Services and Virtualization. As mentioned, the separation of provisioning from delivery is intentional – it underscores the intent to separate the way in which information is physically collected and organized from the ways in which it is consumed. This separation of concerns promotes an environment where changes in one part of the architecture are isolated from other parts of the architecture, e.g. the ripple effect of changes is minimized.

Virtualization and Information Services represent two powerful models that support query federation, abstraction, and mediation. Virtualization is achieved via a logical-to-physical data mapping layer that allows multiple physical data stores, using various access protocols, to be combined into a single logical schema. Consumers can query this component via standard protocols and have the query federated to wherever the needed data physically reside. Usually, queries will access one or more data stores within the LDW, however, it is also possible to access structured data stores outside of the LDW provided that issues such as access rights and data quality are addressed in an acceptable manner. Virtualization can also perform common dimensional analysis functions such as calculating aggregations. Queries and results can be cached for performance gains.

The Information Services components provide a service-oriented approach to information consumption. They promote the definition, classification, governance, and reuse of information services using standard interface technologies. These components provide service level abstraction such as message level transformation, mapping, routing, and interface and protocol mediation. In addition, they provide a means to classify and catalog services in order to promote reuse of common services and help manage resources in a shared environment. Information Services are a fundamental building block to a Service-Oriented Architecture.

Information is accessed via common channels such as JDBC/ODBC, Web Services (WS*), REST, custom APIs, and adapters. The preferred (virtualized) access path will flow through the Virtualization and Information Services components. This ensures that information consumers and providers are not directly connected. Information consumers include analysis artifacts, such as reports and dashboards, analysis tools, and various types of applications.

For some forms of access it is reasonable to bypass the virtualization components and access the LDW directly. Common situations include advanced analysis and discovery activities that operate on data and models that are being maintained by end users. These data stores and models are often referred to as sandboxes since they are intended for single-purpose usage and the contents are managed at the discretion of individual end users. They are logically included in the analytical data management area of the LDW, however they may or may not share physical resources with other LDW components.

Information governance is included in the architecture to support capabilities such as information services management and data quality management. Information services management is used to manage the life cycle of, and access to, information services. It includes components to advertise available services, download service artifacts, and maintain service versioning.

Data quality management guides the flow of data into and through the LDW to protect the integrity and consistency of each area. It manages data cleansing, where appropriate, and enables the system to advertize data quality metrics so that consumers are aware of the quality of the data they are using.

Real-Time Analytics Components of the Logical Architecture

Figure 4 focuses on the real-time analytics components of the Big Data and Analytics reference architecture. It builds upon the components of unified information management, (in the lower right corner), presented in the previous section.

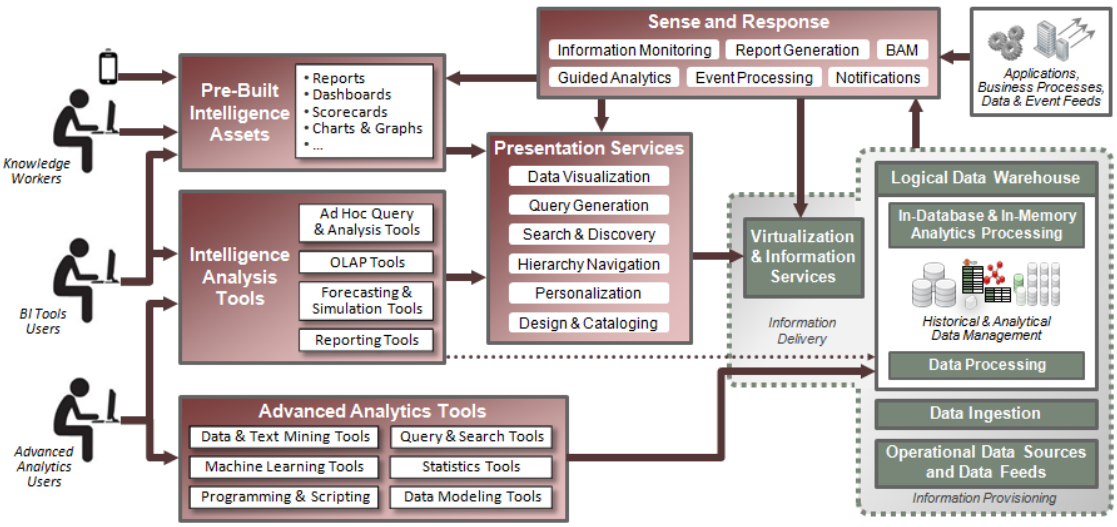

Figure 4. Real-Time Analytics Components of the Logical Architecture

The actors in this scenario are divided into three logical groups: knowledge workers, business intelligence / analysis (BI) tools users, and advanced analytics users. Although the grouping is somewhat arbitrary, it serves to illustrate that different types of users have different needs. Likewise, the tools and interfaces one uses must be tailored to the roles and skills of the user community.

Today, many people in an organization can be classified as knowledge workers. Although relatively few have access to BI tools and possess the background to perform analysis, almost everyone can benefit from the results of analysis. The challenge is to make pertinent information available to knowledge workers, when and where they need it, without them having to design, test, run, and validate their own forms of analysis. This is accomplished by having BI experts perform the 'design', 'test', and 'validate' functions to create a catalog of assets that the knowledge workers can easily use. Knowledge workers use, and reuse, these pre-built assets as they pertain to their job functions.

The users of standard BI tools includes most managers, executives, and planners. They are comfortable with routine operational and performance reporting, forecasting, and planning. BI tools, such as OLAP, SQL-based query building, report generation, and simulation are designed to be user-friendly in order to appeal to this group. The tools tend to include relatively robust and intuitive user interfaces. Often, the tools are either incorporated into common desktop applications such as Microsoft Excel or operate within a standard Web browser.

Both of these groups can take full advantage of Presentation Services. These are services designed to bridge the gap between the intuitive graphical interface and the intricate SQL or MDX data query languages that are used by the information management components. They help to turn data into a personalized visual experience and in turn convert drag-and-drop interface gestures into queries. They allow hierarchies of information to be searched and navigated. They also support the creation of analysis assets that can be designed and cataloged by expert users, and later discovered and used by knowledge workers.

Presentation Services are layered on top of the Virtualization and Information Services. This combination of components helps to establish both a "single version of the truth" and a "single version of the question". The virtualized and governed logical data model defines a common set of information semantics, thus establishing a single version of truth. Presentation Services, via the asset catalog, semantically define a common set of analysis queries, i.e. a single version of the question. Users from across the organization that use these assets will be asking questions the same way and will be getting answers in the same way, from the same source(s) of data. For technical reasons there may be cases where BI

tools must access the LDW directly, however, this should be avoided and these cases should be treated as exceptions.

Advanced analytics users represent the high end of the analysis spectrum. They are most familiar with analysis tools and techniques as well as data modeling, processing, programming, and scripting. Their tools are designed to provide the most capabilities and the freedom to explore, and as such are often considered “expert-friendly”. These tools include routines for data mining, text mining, machine learning, and statistical analysis. They support low level data queries, data modeling, programming, and scripting.

Since advanced forms of analysis tend to work directly on specific data stores, Presentation Services and virtualization capabilities are much less important. Rather, these tools tend to have direct access to data stores in the LDW. In fact, these tools can directly interact with in-database analytics and data processing capabilities within the LDW. This allows advanced users to design analysis routines which are executed directly on data within the warehouse without transmitting data back and forth from the warehouse to the users’ computers.

Real-time analytics involves a set of components that can monitor information and interactions, sense conditions, and respond to events as they occur. These components include information monitoring, business activity monitoring (BAM), and event processing. They provide updates to reports and dashboards as conditions change, alert users about events as they occur, and guide users toward the best course of action in response to conditions and events. Guided analytics can even redirect users to view a specific problem and offer interactive reports that allow the user to drill down into problem areas.

BAM and event processing are complementary technologies. BAM functions as a monitor, receiving data from sources such as business processes and applications and querying other sources on a scheduled basis. BAM looks for pre-defined conditions based on analysis of the data and takes action when a condition is found. Actions include alerting users, initiating processes, and generating business level events. Event processing is used to sift through events of all types and notify subscribers (users, business processes, or other event handlers) when a specific type of event occurs. The event can be based on the contents of a single message or a specific correlation of messages (based on order, frequency, and/or content).

Sense and Response components change the game of analytics from historical analysis to real-time “insight-to-action”. They connect users with the most current information about conditions that are important at the moment.

Intelligent Process Components of the Logical Architecture

Figure 5 focuses on the components of the Big Data and Analytics reference architecture that pertain to intelligent processes.

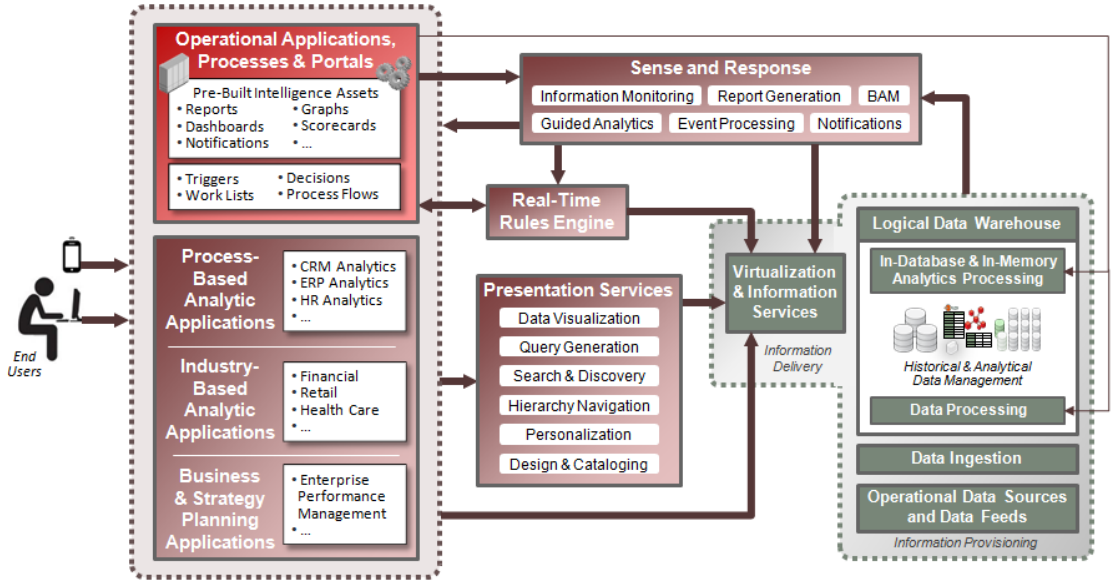

Figure 5. Intelligent Process Components in the Logical Architecture

There are many types of applications that can be used to perform analysis. In the broadest sense, they have been divided into two groups. One group represents applications that support business operations. This includes automated business processes, business services, portals, and various applications that are used to run the business. These applications can leverage pre-built intelligence assets that were described in the previous section. For example, graphs and notifications can appear within an application screen or portal to provide context for making decisions. Reports and dashboards can also be available within the flow of business operations.

In addition, operational applications can programmatically access certain in-database analytics and data processing capabilities. These include statistical analysis, data mining, and machine learning algorithms that can be useful for marketing purposes, intelligence search routines, risk profiling, etc.

The other group represents applications that are primarily designed for analytics. This includes process-based analytic applications and industry-based analytic applications. These applications are often designed to complement specific operational applications, e.g. CRM analytics to analyze data from a CRM application. Both process-based and industry-based

applications tend to be created for data models and analysis that are either standard or common for a specific process and/or industry.

Other types of analytic applications are often designed around certain business functions, such as business and strategy planning. Enterprise performance management applications are an example of this type of application. They support business functions that rely heavily on reporting or analysis.

Several components have been carried forward from the previous architecture scenarios, including the unified information management components, sense and response components, and Presentation Services. They provide the same set of capabilities to applications as they do to analysis tools. In addition, a new component has been added – a real-time rules engine. It evaluates decision logic and provides decisions and recommendations based on real-time information and analysis. The rules engine makes it possible to alter a decision based on current conditions, even if the process itself is completely automated.

Big Data and Analytics components add intelligence to business processes in a number of ways, such as:

- Embedded analysis assets that provide up-to-the-minute intelligence information to decision makers where and when it is needed.
- Real-time decision logic to provide intelligence to automated processes.
- Sense and response capabilities that perform analysis on information from historical data stores, operational systems, and real-time data feeds, and make the results known to knowledge workers.
- Event processing capabilities that can either trigger or alter business processes.
- In-database analytics that can leverage machine learning algorithms and provide capabilities such as product recommendations, targeted advertisements, and fraud detection.

Oracle Product Mapping View

This section describes how the logical architecture can be implemented using Oracle products. The relationship between architecture components and products is not intended to reflect a one-to-one mapping since products have multiple features. Products are mapped to the logical component(s) with the greatest affinity. Likewise, components will map to multiple products that support different deployments or unique sets of capabilities.

The list of products presented in this section is not intended to be a comprehensive list of all products available from Oracle that could be applied to Big Data and business analytics. Rather, they represent a best match to the scope of the architecture that is addressed by the conceptual and logical views. Likewise, not all products listed are required for any particular solution. The actual set of products required depends on the individual company's functional requirements as well as their existing IT assets.

For more information on Oracle products, or further information on product features, please consult the Oracle Big Data website or an Oracle product specialist.

Information Management Product Mapping

Figure 6 maps the Oracle products onto the unified information management components of the logical architecture to illustrate how the capabilities required for information management can be realized using Oracle products.

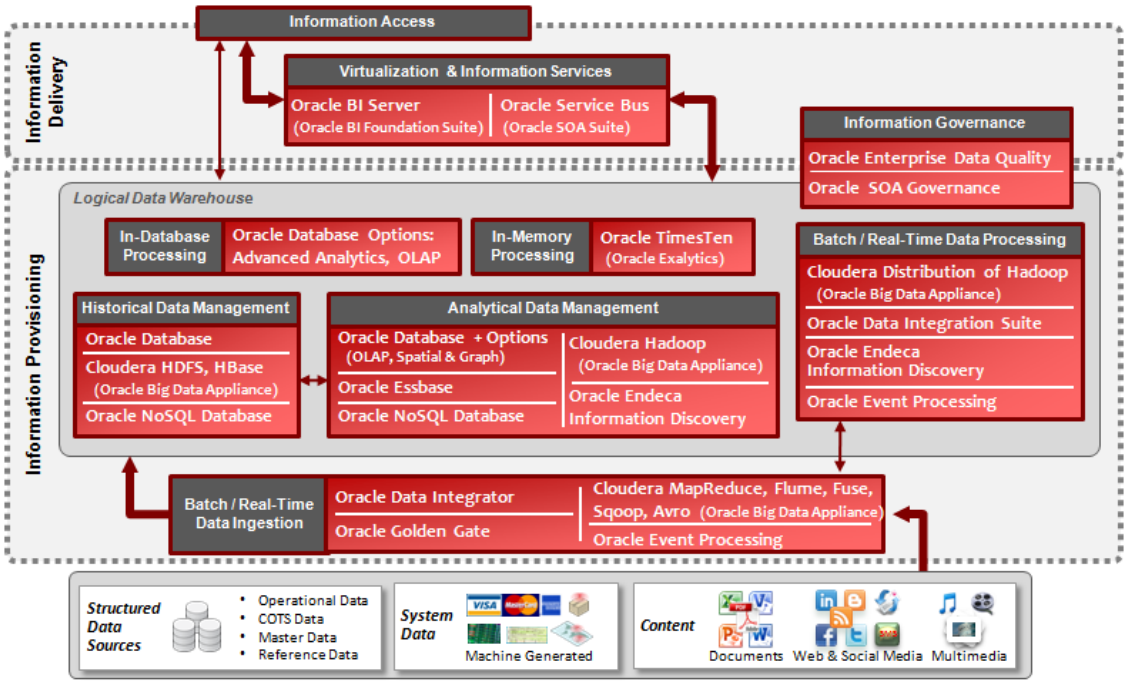

Figure 6. Unified Information Management Product Mapping

Oracle Data Integrator offers unique Extract Load and Transform (ELT) technology that improves performance and reduces data integration costs—even across heterogeneous systems. Hot-pluggable Knowledge Modules provide out-of-the-box modularity, flexibility, and extensibility. Unlike conventional Extract Transform and Load (ETL) tools, ODI delivers the productivity of a declarative design approach and the benefits of an active integration platform for seamless batch and real-time integration.

Oracle Golden Gate is a comprehensive software package for enabling the replication of data in heterogeneous data environments. The product set enables highly available solutions, real-time data integration, transactional change data capture, data replication, transformations, and verification between operational and analytical enterprise systems.

Cloudera Distribution including Apache Hadoop (CDH), included in **Oracle Big Data Appliance**, includes the following products for ingesting, processing, and managing Big Data:

- **Apache Hadoop** consists of the Hadoop Distributed File System (HDFS) and MapReduce.
  - **HDFS** is the primary storage system used by Hadoop applications. HDFS creates multiple replicas of data blocks and distributes them on compute nodes throughout a cluster to enable reliability, and extremely rapid computations.
  - **MapReduce** is a framework parallel processing of large data sets across a large number of nodes. Computational processing can occur on data stored either in a file system (unstructured) or in a database (structured). MapReduce can take advantage of locality of data, processing data on or near the storage assets to decrease transmission of data.
- **Apache Flume** is a distributed, reliable service for efficiently collecting, aggregating, and moving large amounts of log data from many different sources to a centralized data store. It has a simple and flexible architecture based on streaming data flows. It is robust and fault tolerant with tunable reliability mechanisms and many failover and recovery mechanisms.
- Project **FUSE** (Filesystem in Userspace) allows HDFS to be mounted on supported UNIX/LINUX systems as a standard file system. This allows access to HDFS using ordinary file system access commands and utilities, thus making data import and export easier.
- **Apache HBase** is a distributed NoSQL column-oriented store built on top of HDFS providing the capability to perform consistent real-time random read/write access to very large data sets.
- **Apache Sqoop** is a tool designed for efficiently transferring bulk data between Apache Hadoop and structured data stores such as relational databases. Sqoop can be utilized to import data from external structured data stores into HDFS or related systems like Hive and HBase. In addition, Sqoop can be used to extract data from Hadoop and export it to external structured data stores such as relational databases and enterprise Data Warehouses.
- **Apache Avro** is a data serialization system that provides rich data structures in a compact, fast, and binary data format that is simple to integrate with dynamic languages.

Oracle Big Data Connectors (not shown) is a software suite that integrates Apache Hadoop with Oracle software, including Oracle Database, Oracle Endeca Information Discovery, and Oracle Data Integrator. The suite includes:

- **Oracle SQL Connector for HDFS** is a high-speed connector for accessing data in HDFS directly from Oracle Database. With this connector, users have the flexibility to either query or import data from HDFS at any time, as needed by the application.
- **Oracle Loader for Hadoop** is a MapReduce utility used to optimize data loading from Hadoop into Oracle Database. Oracle Loader for Hadoop sorts, partitions, and converts data into Oracle Database formats in Hadoop, then loads the converted data into the database.
- **ODI Application Adapter for Hadoop** provides native Hadoop integration within ODI. ODI generates optimized HiveQL which in turn generates native MapReduce programs that are executed on the Hadoop cluster.
- **Oracle R Connector for Hadoop (ORCH)** enables R scripts to run on data in Hive tables and files in HDFS – seamlessly leveraging the MapReduce framework. Hadoop-based R programs can be deployed on a Hadoop cluster without needing to know Hadoop internals, command line interfaces, or IT infrastructure. ORCH can optionally be used with the Oracle Advanced Analytics Option for Oracle Database.

Oracle Event Processing is a complete, open architecture that enables the sourcing, processing, and publishing of complex events. This allows filtering, correlation, and processing of events in real-time so that downstream applications are driven by true, real-time intelligence. Oracle Event Processing provides the ability to join incoming streaming events with persisted data, thereby delivering contextually aware filtering, correlation, aggregation, and pattern matching.

Oracle Endeca Information Discovery is an enterprise data discovery platform for rapid, intuitive exploration and analysis of information from any combination of structured and unstructured sources. It enables organizations to extend their existing business analytics investments to unstructured data – such as social media, websites, content systems, files, email, database text, and Big Data. The core search-analytical database organizes complex and varied data from disparate source systems into a faceted data model that is extremely flexible and reduces the need for up-front data modeling.

Oracle NoSQL Database provides multi-terabyte distributed key/value pair storage with predictable latency. Data is stored in a very flexible key-value format, where the key consists of the combination of a major and minor key (represented as a string) and an associated value (represented as a JSON data format or opaque set of bytes). It offers full Create, Read, Update and Delete (CRUD) operations, with adjustable durability and consistency guarantees.

Oracle Database delivers industry leading performance, scalability, security, and reliability on a choice of clustered or single-servers. It provides comprehensive features to easily manage the most demanding transaction processing, business analysis, and content management applications. Oracle Database comes with a wide range of options to extend the world's #1 database to help grow your business and meet your users' performance, security, and availability service level expectations.

Oracle Database options include:

- **OLAP -** The OLAP Option to Oracle Database is a full featured online analytical processing server embedded within the Oracle Enterprise Edition Database. It runs within the kernel of the database, which by default allows it to benefit from standard Oracle Database features such as scalability, reliability, security, backup and recovery, and manageability.
- **Spatial and Graph -** The Oracle Spatial and Graph option to Oracle Database includes full 3-D and Web Services support to manage all geospatial data including vector and raster data, topology, and network models. It's designed to meet the needs of advanced geographic information system (GIS) applications such as land management, utilities, and defense/homeland security. Oracle's open, native spatial format eliminates the cost of separate, proprietary systems, and is supported by all leading GIS vendors.
- **Oracle Advanced Analytics** provides capabilities including data-mining algorithms, SQL functions for basic statistical techniques, and integration with open-source R for statistical programming and access to a broad set of statistical techniques. Business analysts, data scientists, and statisticians can access these analytics via SQL or R languages and apply these algorithms directly against data stored within the Oracle Database. Oracle Advanced Analytics reduces complexity and speeds development and deployment of analytics by providing all core analytic capabilities and languages on a simple, powerful, in-database architecture.

Oracle Essbase Oracle Essbase is the market leading OLAP server for enterprise performance management (EPM) applications. Designed specifically for business users, Oracle Essbase supports forecasting, variance analysis, root cause identification, scenario planning, and what-if modeling for both custom and packaged applications.

Oracle Business Intelligence Foundation Suite is a complete, open, and architecturally unified business intelligence solution for the enterprise that delivers best in class capabilities for reporting, ad hoc query and analysis, OLAP, dashboards, and scorecards. The Oracle BI Server component of this suite is a highly scalable and efficient, query, reporting, and analysis server that provides services to support information virtualization. It exposes its services through standard ODBC and JDBC-compliant interfaces. Clients of the Oracle BI Server see a logical schema view independent of the physical database schemas. Oracle BI Server clients submit

"Logical" SQL, which ultimately gets translated by the server to native, source-specific data source query languages like SQL and MDX.

Oracle SOA Suite simplifies connectivity by providing a unified experience to integrate across cloud, on-premise, and business-to-business. Additional components included within the unified platform are the enterprise service bus as the foundation for shared services, process orchestration for business optimization, business rules for business agility, and business activity monitoring to deliver role-based business visibility.

Oracle Enterprise Data Quality provides organizations with an integrated suite of data quality tools that provide an end-to-end solution to measure, improve, and manage the quality of data from any domain, including customer and product data. Oracle Enterprise Data Quality also combines powerful data profiling, cleansing, matching, and monitoring capabilities while offering unparalleled ease of use.

Oracle SOA Governance (OSG) is a complete solution for delivering a broad range of governance capabilities including SOA Management, API Management, and SOA Security. OSG includes the Oracle Enterprise Repository, to help manage service assets throughout the SOA lifecycle, and the Oracle API Gateway, to provide security for all types of services interfaces.

Oracle Exalytics In-Memory Machine is the industry's first engineered in-memory analytics machine that delivers no-limit, extreme performance for BI and EPM applications. The hardware is a single server that is optimally configured for in-memory analytics for BI workloads and includes powerful compute capacity, abundant memory, and fast networking options.

The Oracle Exalytics In-Memory Machine features an optimized **Oracle BI Foundation Suite** and **Oracle TimesTen In-Memory Database for Exalytics (TimesTen)**. The TimesTen In-Memory Database for Exalytics is an optimized in-memory analytic database, with features exclusively available on Oracle Exalytics platform. TimesTen is a proven memory-optimized full-featured relational database with persistence. It stores all its data in memory optimized data structures and supports query algorithms specifically designed for in-memory processing. Using the familiar SQL programming interfaces, TimesTen provides real-time data management that delivers blazing-fast response times and very high throughput for a variety of workloads.

Real-Time Analytics Product Mapping

Figure 7 maps the Oracle products onto the real-time analytics components of the logical architecture to illustrate how the capabilities required for real-time analytics can be realized using Oracle products.

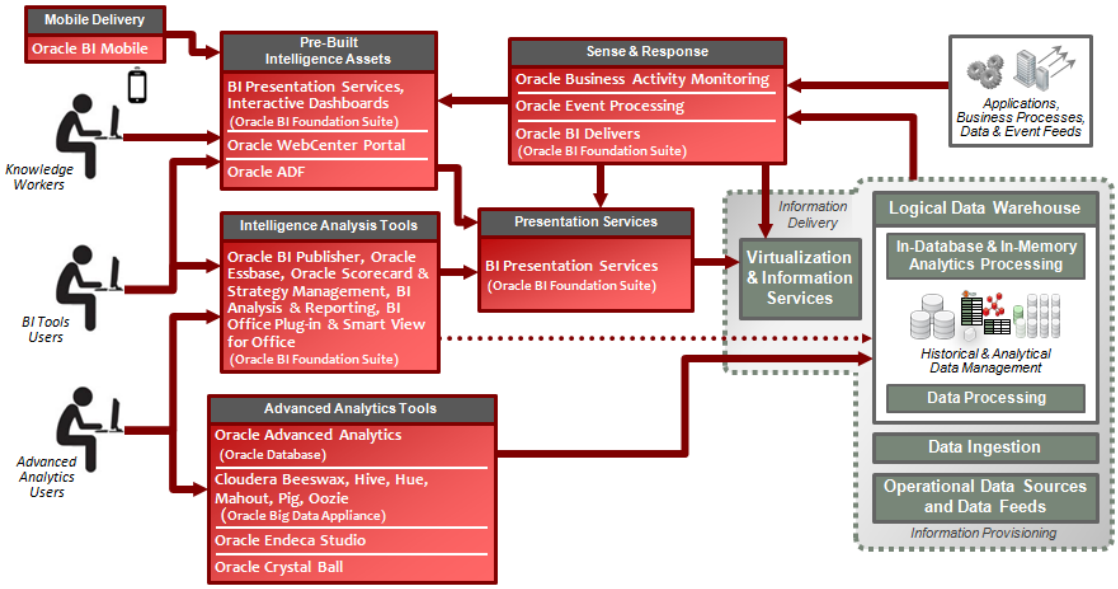

Figure 7. Real-Time Analytics Product Mapping

Real-time analytics builds upon the architecture for unified information management, which is represented as an icon in the lower right corner. The product mapping for these components is available in the previous section.

Some Oracle products offer features for multiple layers of the architecture and are therefore applicable to this scenario as well as the previous scenario. Rather than duplicate previous product descriptions, this section will expand on the descriptions of these products and describe specific features as they apply to this scenario.

Oracle Business Intelligence Foundation Suite (OBIFS) includes several products and features that support real-time analytics, such as:

- **BI Presentation Services** generates the user interface such as dashboards and analyses which are used to visualize data from the Oracle BI Server. With OBIFS, all BI content resides in a common catalog enabling search, archiving, migration, and the re-use of common objects across any number of personal and shared catalog items. BI Presentation Services provides a browser-based administration tool to administer all functions in the catalog.

- **Oracle BI Publisher** is an enterprise reporting solution for authoring, managing, and delivering highly formatted documents such as operational reports, electronic funds transfer documents, government PDF forms, shipping labels, checks, and sales and marketing letters.
- **Oracle Essbase** simplifies cube construction by delivering a single environment for performing tasks related to data modeling, cube designing, and analytic application construction. With a wizard-driven user interface, Essbase Studio supports modeling of the various data source types from which Essbase applications are typically built. Essbase enables line-of-business personnel to simply and rapidly develop and manage analytic applications that can forecast likely business performance levels and deliver "what-if" analyses for varying conditions.
- The **BI Analysis and Reporting** feature provides end users with broad ad-hoc query, analysis, and reporting capabilities. It is a pure Web-based environment that is designed for users who want to create new analyses from scratch or modify and change existing analyses that appear on dashboard pages. Users can create a range of interactive content types which can be saved, shared, modified, formatted, or embedded in the user's personalized dashboard or enterprise portal.
- **Oracle Scorecard and Strategy Management** extends OBIFS with capabilities intended to communicate strategic goals across the organization and monitoring their progress over time. It provides capabilities to establish specific goals, define how to measure their success, and communicate that information throughout the organization.
- The **Interactive Dashboards** feature provides users with an interactive experience where information is filtered and personalized to a user's identity or role. End users can create their own dashboards from pre-defined content in the catalog, e.g. reports, prompts, graphs, tables, and pivot tables. Users interact with dashboard content by selecting prompted values and filtering data, and drilling on graphs or tables to access detail and related content.
- **Oracle BI Delivers** provides the ability to proactively monitor business information; identify patterns to determine whether specific problems are occurring; filter the data based on data and time-based rules; alert users via multiple channels such as email, dashboards, and mobile devices including text messages and mobile phones; and take action in response to alerts that are received.
- **BI Office Plug-in** and **Smart View for Office** allow business users to add business intelligence information into Microsoft Office documents such as Word and Excel. Smart View provides the ability to integrate EPM & BI data directly from the data source into Microsoft Word, Microsoft PowerPoint, and Microsoft Outlook and the capability to synchronize information between documents.

Cloudera Distribution including Apache Hadoop (CDH), included in **Oracle Big Data Appliance**, includes the following products for analyzing Big Data:

- **Apache Hive** is a system that facilitates easy data summarization, ad-hoc queries, and the analysis of large datasets stored in Hadoop compatible file systems. Hive provides a mechanism to project structure onto this data and query the data using a SQL-like language called HiveQL.
- **Beeswax** provides a user interface to Hive, and allows Hive queries to be exported to spreadsheets and flat files.
- **Hue** is a browser-based environment that enables users to interact with a Hadoop cluster. Hue includes several easy to use applications that assist users in working with Hadoop MapReduce jobs, Hive queries, and user accounts. The Hue applications run in a Web browser.
- **Apache Mahout** offers a number of core machine learning algorithms such as clustering, categorization, and collaborative filtering that are implemented on top of Apache Hadoop using MapReduce.
- **Apache Oozie** is a scalable, reliable, and extensible workflow/coordination system to manage several types of Apache Hadoop jobs (e.g. Java MapReduce, Streaming MapReduce, Pig, Distcp, etc.).
- **Apache Pig** is a platform for analyzing large data sets that consists of a high-level language called Pig Latin coupled with an infrastructure that consists of a compiler that produces sequences of MapReduce programs. Pig programs are amenable to substantial parallelization, which in turn enables them to handle very large data sets.

Oracle Advanced Analytics provides capabilities including data-mining algorithms, SQL functions for basic statistical techniques, and integration with open-source R for statistical programming and access to a broad set of statistical techniques. Oracle Advanced Analytics provides multiple user interfaces designed for every business audience, including:

- **Oracle Data Miner** graphical user interface (GUI), an extension to Oracle SQL Developer, provides analysts with an easy to use work flow environment for predictive modeling. The GUI generates SQL scripts for analytical methodologies so analysts can rapidly move from analytical concept to enterprise-wide deployment—saving time and money.
- **SQL, PL/SQL and R APIs.** Algorithms are accessible by SQL and R APIs. Users who know the R statistical programming language can use R's console, RStudio or other R IDEs to work directly with their Oracle data. Data scientists familiar with R can write, test, and deploy R scripts and optionally integrate them with Oracle Data Miner.

Oracle Endeca Studio is a discovery application composition environment for the Oracle Endeca Information Server. Studio provides drag and drop authoring to create highly interactive, visually-rich, enterprise-class information discovery applications.

Oracle Crystal Ball (OCB) provides capabilities for simulation and predictive analytics. OCB works with Microsoft Excel spreadsheet models. Users define their models within Excel and use OCB to define assumptions, forecasts, and decision variables, and to run the simulations.

Oracle BI Mobile allows users to view, analyze, and act on Oracle BI content on supported mobile devices such as Apple iPhone and Apple iPad. It supports BI functionality including notifications and alerts, reporting, ad hoc query, OLAP analysis, dashboards, and scorecards. It also supports in-context embedded actions such as invoking business processes and Web Services.

Oracle Business Activity Monitoring (BAM) provides a single view of key business metrics from across the organization. It integrates with key business systems and information sources. Oracle BAM monitors complex changing conditions in real-time based on user-defined rules. It takes a variety of actions in response to those changes, including notifying users, generating reports, invoking processes, and calling other applications via Web Services.

Oracle Event Processing is used to capture, process, generate, and distribute business events based on streams of data. Events can be processed directly and/or fed into BAM, where BA-relevant meaning can be applied, and appropriate actions can be taken.

Oracle WebCenter Portal provides social collaboration and information sharing. It supports consumption and interaction with Oracle BI content. WebCenter Web 2.0 Collaboration features include: search, tagging, tag clouds, linking and document association, discussion forums, chat, presence, real-time collaboration, Workspaces, and Community lists.

Oracle Application Development Framework (ADF) is an end-to-end Java EE framework that simplifies application development by providing out of the box infrastructure services and a visual and declarative development experience. Oracle BI provides BI view components for easy integration of objects such as reports, dashboards, scorecards, and SQL view objects into ADF-based applications.

Intelligent Processes Product Mapping

Figure 8 maps the Oracle products onto the logical architecture for intelligent processes to illustrate how the capabilities can be realized using Oracle products. Several products used to support this scenario have been introduced in previous sections. This section describes the products as they apply to intelligent processes and introduces some new products that are specific to this scenario.

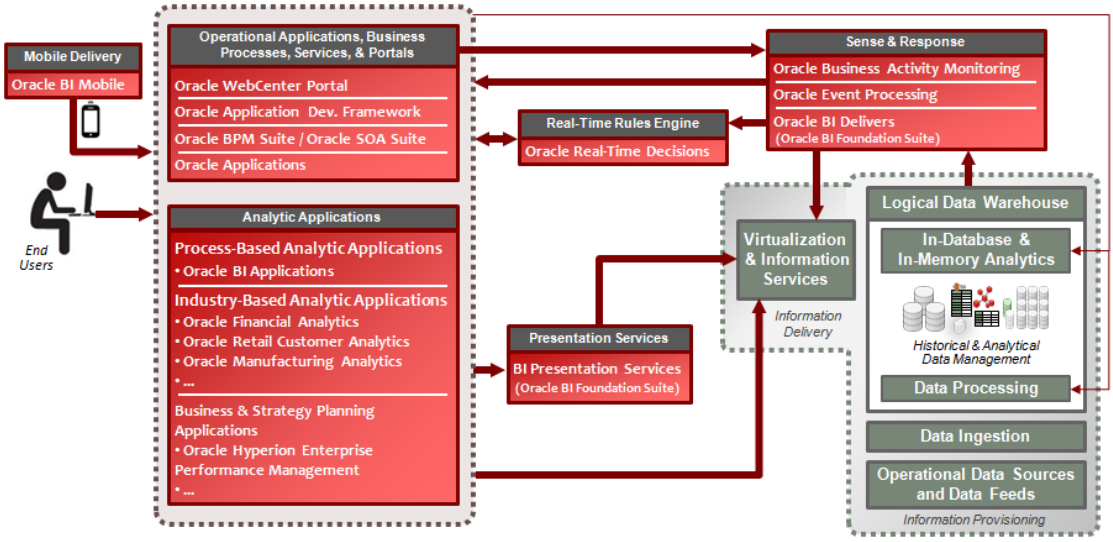

Figure 8. Intelligent Processes Product Mapping

Figure 11 includes products that support intelligence-gathering processes as well as products that add intelligence to other operational processes. The left side of the diagram illustrates this concept. Operational applications, processes, services, and portals are supported by products that easily integrate and/or interact with intelligence assets. This includes products such as:

- **Oracle WebCenter Portal**, which supports consumption and interaction with Oracle BI content via portlets.
- **Oracle Application Development Framework**, which provides BI view components for easy integration of objects such as reports, dashboards, scorecards, and SQL view objects into ADF-based applications.
- **Oracle BPM Suite and Oracle SOA Suite,** which provide the ability to model and execute business processes and support Service-Oriented Architecture environments. These capabilities provide two-way interaction with Big Data and Analytics.
- **Oracle Applications,** which can interact with the architecture through all of the above channels.

In addition to the analysis tools presented in the previous section, intelligence-gathering processes are supported by several types of analytic applications. These applications can be classified as:

- Process-based analytic applications, which align with a specific type of business process. **Oracle BI Applications** support many standard processes related to ERP, CRM, and HR. These applications include pre-integration with Oracle Applications,

pre-defined data models with conformed dimensions, role-based reports and dashboards, security integrated with operational applications, metadata-driven integration capabilities, and mobile access. They greatly reduce the time and cost of developing data warehouse and analysis capabilities as compared with custom-built solutions.

- Industry-based solutions are available for many types of industries including retail, financial services, health sciences, and manufacturing. Solutions such as **Oracle Retail Merchandising Analytics, Oracle Retail Customer Analytics, Oracle Financial Analytics,** and **Oracle Manufacturing Analytics** include data models, reports, dashboards, and tools that are designed and optimized for specific industries.
- **Oracle Hyperion Enterprise Performance Management Applications** address critical processes such as strategy management, planning, budgeting, forecasting, financial close and reporting, profitability, and cost management.

Intelligent processes are supported by the same infrastructure that is used to perform analysis. For instance, Presentation Services, provided by **Oracle BI Foundation Suite**, are used for visual rendering and query generation. Likewise, the architecture for information services, virtualization, and management is common across all use cases.

Furthermore, the Sense and Response capabilities are also integrated into intelligent processes. For example, **Oracle Business Activity Monitoring** and **Oracle Event Processing** can be used to initiate processes in response to conditions and events that occur within the system. They are particularly effective when used in conjunction with business process management (BPM) and SOA architectures, including **Oracle BI Applications**.

In addition, **Oracle Real-Time Decisions (ORTD)** can be used to influence rules for real-time decision management. **ORTD** is a complete decision management solution that delivers real-time decisions and recommendations and automatically renders decisions within a business process to create tailored messaging for every customer interaction**.**

Oracle Engineered Systems

The product mapping sections above describe the Oracle products that can be used to realize the Big Data and Analytics solution. Most of the products are software. In addition, Oracle has developed hardware systems with advanced features for Oracle software – engineered systems. These engineered systems incorporate hardware and software designed and built to provide the best possible performance and scalability.

Oracle Engineered Systems benefit from a long list of optimizations that have been made that are unique to the combined hardware / software solutions. Additionally, they provide top to

bottom management and monitoring (via Oracle Enterprise Manager) and, due to the predefined configuration, include industry leading product support. Thus, although the Oracle software products can be hosted on a wide variety of hardware systems, the Oracle Engineered Systems provide a uniquely capable hosting platform.

Figure 9 illustrates several engineered systems that provide a superb platform for the Big Data and Analytics solution.

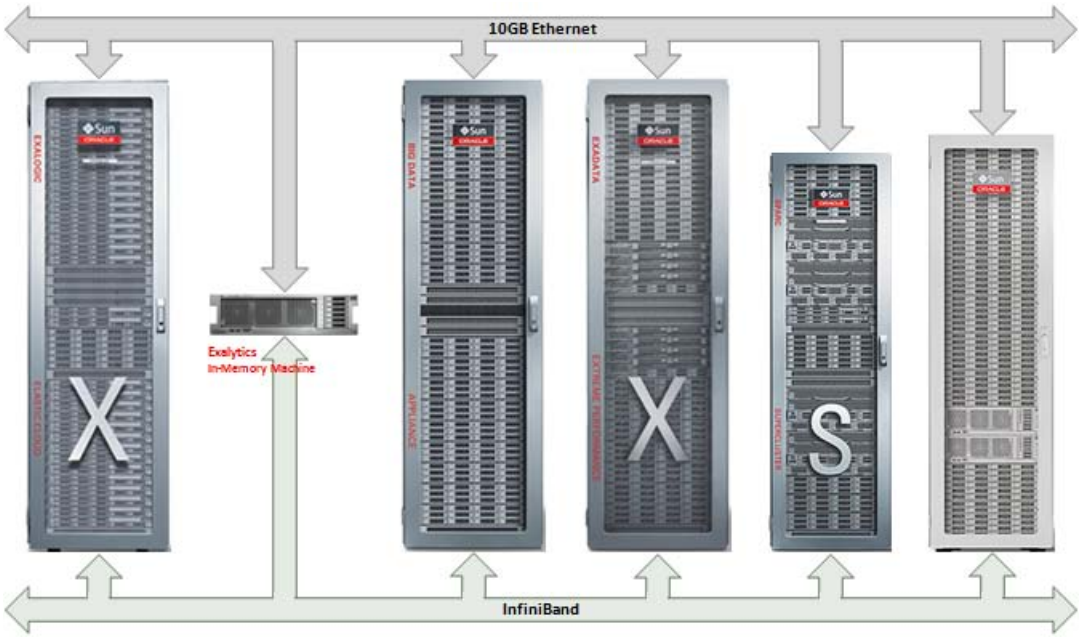

Figure 9. Oracle Engineered Systems for Big Data and Analytics

From left to right, the systems shown in Figure 13 are:

Oracle Exalogic Elastic Cloud is designed, optimized, and certified for running Oracle applications and technologies and is ideal for mission-critical middleware and applications from Oracle and third-party vendors.

Oracle Exalytics In-Memory Machine is the industry's first in-memory BI machine that delivers the fastest performance for business intelligence and planning applications. The Oracle Exalytics In-Memory Machine features an optimized Oracle BI Foundation Suite and Oracle TimesTen In-Memory Database for Exalytics. Oracle BI Foundation Suite takes advantage of large memory, processors, concurrency, storage, networking, operating system, kernel, and system configuration of the Oracle Exalytics hardware. This optimization results in better query responsiveness, higher user scalability and markedly lower TCO compared to standalone software.

The TimesTen In-Memory Database for Exalytics is an optimized in-memory analytic database, with features exclusively available on Oracle Exalytics platform. Also included is Oracle Essbase with in-memory optimizations for Exalytics. These two data management engines are leveraged in the following four techniques to provide high performance in-memory analytics for a wide variety of business intelligence usage scenarios:

- In-Memory Data Replication. Many BI implementations, including pre-packaged Oracle BI Applications, may be able to fit entirely in memory. In such cases, the Oracle BI Server for Oracle Exalytics can replicate the entire data warehouse into the TimesTen In-Memory database.
- In-Memory Adaptive Data Mart. Most BI deployments have workload patterns that focus on a specific collection of “hot” data from their enterprise data warehouse. In such cases, TimesTen for Exalytics can be used to create a data mart for “hot” data. Query response times have been reduced by 20X using this strategy.
- In-Memory Intelligent Result Cache. Oracle Exalytics Result Cache is a completely reusable in-memory cache that is populated with results of previous logical queries generated by the server. In addition to providing data for repeated queries, any result set in the result cache is treated as a logical table and is able to satisfy any other queries that require a sub-set of the cached data.
- In-Memory Cubes. Oracle Essbase with its in-memory optimizations for Oracle Exalytics supports both read and write operations to cubes stored in memory. Cubes can be created from data that is extracted from the semantic layer of Oracle BI Server.

Oracle BI Foundation components feature a hardware acceleration option that enables optimizations that specifically exploit the particular configuration of the Oracle Exalytics machine from the processor architecture to the concurrency and memory. These optimizations have shown to provide up to 3X improvement in throughput at high loads and thus can handle 3X more users compared to similar commodity hardware.

Oracle Big Data Appliance combines optimized hardware components with new software to deliver the most complete solution for acquiring, organizing, and loading unstructured data into an Oracle Database. Oracle Big Data Appliance uses Cloudera's Distribution Including Apache Hadoop (CDH) and Oracle NoSQL Database as data management capabilities. It includes a combination of Oracle Enterprise Manager and Cloudera Manager for both hardware and software cluster administration and monitoring. For deep analysis of Big Data, it also includes an open-source distribution of the statistical environment R.

Oracle Exadata Database Machine is a complete package of servers, storage, networking, and software that is massively scalable, secure, and engineered for redundancy that provides

extreme performance for both data warehousing and OLTP applications. Exadata Storage Server Software includes features such as:

- SmartScan, which improves query performance by maintaining storage indexes within the storage server cell memory. It eliminates disk I/O by avoiding read operations to data blocks that do not contain entries that satisfy the query.
- Smart Flash Cache, also contained within the storage cells, is used to transparently store data that will be reused in other queries.
- Hybrid Columnar Compression reduces storage space requirements by 10x while also improving the speed of database queries.

Oracle SuperCluster is Oracle's fastest engineered system for running both database and enterprise applications. Combining powerful virtualization and unique Oracle Exadata and Oracle Exalogic optimizations, Oracle SuperCluster is ideal for consolidation and enterprise private cloud deployments. Oracle SuperCluster is a complete system integrating Oracle's servers, storage, network and software. It delivers extreme performance, no single point of failure, and high efficiency while reducing risk.

Oracle's Sun ZFS Backup Appliance provides an integrated, high-performance backup and recovery solution for Oracle's engineered systems. The InfiniBand fabric provides high bandwidth between the Oracle Exadata database servers, storage cells, and the Sun ZFS Backup Appliance.

Figure 9 illustrates how the engineered systems can leverage the included InfiniBand network capabilities to provide a high-throughput, low-latency connection between the engineered systems. Where latency and throughput are less critical (e.g. connecting to the rest of the datacenter), the engineered systems also include 10GB Ethernet capability.

Only one of each type of engineered system is shown in the illustration but, depending on the computational load, multiples of the engineered systems might be used to host the Big Data and Analytics solution. For smaller workloads, Exalogic and Exadata also come in quarter and half rack configurations.

Implementation

For most organizations, the journey to a complete, modern, and truly differentiating Big Data and Analytics solution is best taken in steps. Given the scope of the effort, this is also a path best not traveled alone. To help you devise a plan, quickly achieve success in each phase, and build with a consistent architecture from start to finish, Oracle Services and the Oracle

Partner Network offer expertise and experience to complement your in-house skills and competencies.

Oracle's services span all phases of the lifecycle. From early questions about what to do first and what to deploy on-premise or in the cloud, Oracle can provide decision models and best-practice-based recommendations. Oracle Enterprise Architects can help you define a vision for Big Data and Analytics, based on your particular business opportunities, and create an enterprise architecture roadmap to guide the build-out of your environment. Oracle can also help you develop a center of excellence to promote best practices and advance the maturity of your analytics program.

Regardless of whether your solution involves upgrades, migrations, new installations, integration, or all of the above, Oracle Consulting is uniquely qualified to help – delivering Oracle expertise and a single point of accountability with strong connections to Oracle Development, Oracle Support, Oracle Technology Partners, Oracle Integration Partners, and other important players in your implementation. Oracle provides your staff with the knowledge and assistance needed to get the most of your Oracle solution through world-class product training and global enterprise support. With Oracle as your strategic partner, you can count on continuous innovation and ongoing enhancements fueled by Oracle's unparalleled investment in research and development. As a result, your solution can evolve even as we extend it to bring even more capabilities to enterprise computing.

Conclusion

Companies today are always looking to gain a competitive advantage. As compute power and storage capacities continue to rise, and costs continue to decline, Big Data and analytics are playing an increasingly important role in this quest. But rather than deploy ever more systems and create new information silos and integration challenges, a holistic approach should be taken. This approach should incorporate new forms of data into a universal and extensible architecture.

The Big Data and Analytics Reference Architecture described in this paper delivers:

- An approach to information management that unifies all forms of data including structured, unstructured, and semi-structured data
- High performance in-memory and in-database analytics
- Ability to handle batch and real-time data feeds
- "Single version of the truth" coupled with "single version of the question"
- Sense and response capabilities that drive insight to action and infuse intelligence into business processes

- A complete, full-featured, high speed analytics platform

The Oracle products that provide the capabilities were described and mapped onto the logical architecture to illustrate how Oracle products can be used to realize the Big Data and Analytics Reference Architecture. The breadth of Oracle's products related to Big Data and analytics is extensive, which enables Oracle to craft and implement a complete solution.

Further Reading

IT Strategies from Oracle

IT Strategies from Oracle (ITSO) is a series of documentation and supporting material designed to enable organizations to develop an architecture-centric approach to enterprise-class IT initiatives. ITSO presents successful technology strategies and solution designs by defining architecture concepts, principles, guidelines, standards, and best practices.

There are several documents in the ITSO library that are particularly relevant to the Big Data and Analytics Reference Architecture including:

- ***Oracle Reference Architecture Information Management***
- ***Oracle Reference Architecture Business Analytics Foundation***
- ***Oracle Reference Architecture Business Analytics Infrastructure***
- ***Oracle Reference Architecture Service Orientation***
- ***Oracle Reference Architecture Security***
- ***Oracle Reference Architecture Engineered Systems***

All of these topics are important to Big Data and analytics but were only briefly discussed in this paper. Please consult the ITSO web site for a complete listing of documents as well as other materials in the ITSO series.

Other References

Further information about the products described in this paper can be found on Oracle's web site at: http://www.oracle.com/us/products/index.html

Further information about the Big Data and Analytics Solution can be found on Oracle's web site at: http://www.oracle.com/us/technologies/big-data/index.html

Big Data & Analytics Reference Architecture
September 2013

Author:
Dave Chappelle

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200

oracle.com

Oracle is committed to developing practices and products that help protect the environment

Hardware and Software, Engineered to Work Together