Big Data in Financial Services and Banking

Architect's Guide and Reference Architecture Introduction

ORACLE®

Disclaimer

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Table of Contents

Executive Summary

The ability to access, analyze, and manage vast volumes of data while rapidly evolving the Information Architecture has long been critical to financial services companies as they improve business efficiency and their performance. Recently, bank profitability has been on the rise, especially in regions of the world where economic conditions are good. Financial services organizations will continue to focus on revenue growth and higher margins through operational efficiency, better risk management, and improved customer intimacy. Banks will also develop new revenue streams by entering new markets and service areas.

Using information technology to enhance the customers' experience in both retail and business banking can help grow interest-based and fee-based revenue. Many larger financial organizations are gravitating towards expansion of wealth management portfolios to ensure lower risk and consistent fee-based revenue. Differentiated services, cross-sell and up-sell initiatives, and expansion into emerging wealth-management markets around the world are on the rise. Analytics and information management play a central role in ensuring that these strategies are properly executed.

As financial services companies embark on a journey to gain a better understanding of customers and their household preferences in order to provide effective and differentiated services, the amount of data grows, data collection occurs more frequently, and data variety becomes more complex.   Today, these data sources can include:

» Traditional enterprise data from operational systems related to customer touch points such as:
  » ATMs
  » Call Centers
  » Web-based and mobile sources
  » Branches / Brokerage units
  » Mortgage units
  » Credit cards
  » Debt including student and auto loans
  » Volatility measures that impact the clients' portfolios

» Financial business forecasts from various sources such as:
  » News
  » Industry data
  » Trading data
  » Regulatory data
  » Analyst reports (internal and competing banks)
  » Alerts about events (news, blogs, Twitter and other messaging feeds)

» Other sources of data such as:
  » Advertising response data
  » Social media data

As the rate that this data is generated increases, business analysts who crave such data rapidly consume it. Information discovery tools enable them to rapidly combine various data sets leading to better insight. They often want more data to be ingested at higher rates and stored longer, and want to analyze the growing data volumes faster. "Big Data" solutions help financial services and banking institutions respond to these requirements.

This paper provides an overview for the adoption of Big Data and analytic capabilities as part of a "next-generation" architecture that can meet the needs of the dynamic financial services and banking industries.

This white paper also presents a reference architecture introduction. The approach and guidance offered is the byproduct of hundreds of customer projects and highlights the decisions that customers face in the course of architecture planning and implementation. The paper reflects the experience of Oracle's enterprise architects who work within many industries and who have developed a standardized methodology based on enterprise architecture best practices. Oracle's enterprise architecture approach and framework are articulated in the Oracle Architecture Development Process (OADP) and the Oracle Enterprise Architecture Framework (OEAF).

Key Business Challenges

Companies in the consumer banking and financial services industry typically have data warehouses and business intelligence tools for reporting on and analyzing customer behavior to better anticipate their needs, and for optimizing operations. By deploying Big Data Management Systems that include data reservoirs (featuring Hadoop and / or NoSQL databases), greater benefits in these areas can be achieved as the business gains more predictive capabilities and becomes more agile. The addition of Big Data systems enables organizations to gain much higher levels of insight into data faster and enables more effective decision making.

An Enterprise Modeling Platform

Financial services organizations are rethinking how they model their enterprises. This renewed focus is fueled in part by new regulatory requirements. In addition, financial institutions are increasingly incorporating analytical insights into their operational decision processes. Statistical modeling is taking on a wider role within the enterprise as institutions weave prediction, optimization, and forecasting models into their enterprise analytics fabric.

New challenges come as adoption increases. As output from the models becomes part of regulatory and other business intelligence processes, enterprise model management (much like enterprise data management) must become a priority. When modeling becomes more prevalent, the models are often deployed on centrally managed platforms in IT that are aligned to individual lines of business. However, there can be a chasm between the modeling and IT worlds.

Modeling platforms often contain copies of enterprise data. While a bank may have put in place sophisticated data governance policies around data in the enterprise warehouse, data used for models often falls outside the purview of these governance systems. The problem is exacerbated by the new sources of data that modelers want access to. The oft-repeated phrase that "the analytics problem is a data problem" underscores the need to closely link analytics and data management. Yet while banks have poured resources into enterprise level data management and governance programs, enterprise-level model management does not seem to have attracted the same level of attention. Just as regulatory requirements shaped the financial institutions' data management approach, we believe that regulators' demands for model management will be very similar.

Risk and Capital Management

Traditional enterprise architectures have served banks and financial services companies well for years. The architectures have enabled these institutions to manage credit, market liquidity, and operational risk. In addition, these systems have enabled the institutions to manage their capital and meet Basel regulatory requirements.

Credit and behavioral scoring to classify new or existing customers for credit worthiness required significant analysis of loan application data and data from credit bureaus by credit experts. Given the push of banks into micro-credit and the expansion into the emerging markets, the scarcity of available credit data is hugely problematic. This data scarcity can be overcome through predictive modeling using non-traditional input from peer groups, P2P payment data from mobile devices, utility consumption and payment data, prepaid mobile services purchase data, and other sources.

The valuation of complex and illiquid instruments and portfolios requires simulation of thousands of risk factors using stochastic models. Monte Carlo simulations are frequently used. As the number of risk factors increase, the compute power required for these simulations increases (exponentially to the number of risk factors). Compute grids or other expensive and proprietary solutions were previously used for such simulations, but the simulations were

often stopped after a period of time due their cost. Big Data technologies enable the simulations to run to conclusion at a low cost by introducing commodity hardware and largely open source software.

Banking analysts are building data models that can predict loans likely to become delinquent in order to initiate proactive action. Large banks in the United States are initiating transfer of servicing rights to sub servicing organizations for specific loans in default. Identifying loans that are to be moved to sub servicing organizations requires detailed analysis of various data elements such as history of the loan, the borrower, and the loan documents. Banks sometimes use predictive analytics to create heat maps that identify regions that are known for mortgage fraud, showing details at a zip code level and possibly at an individual level. This enables effective analysis when the appraisal process is being conducted for a new loan application and can help prevent potential property valuation, occupancy, and short-sale fraud.

Big Data technologies can help in the identification of household spending patterns by building a richer view of customers. Banks can analyze transaction logs for all their products and identify spending trends at the household level. This provides a better view of the customer's true ability to pay back loans and also helps identify future cross-selling and up-selling opportunities.

Wealth Management

As financial institutions seek to expand into new geographies and provide differentiated services to enhance a wealth management portfolio, they need a comprehensive view of customers, customer households & networks, and need to develop personalized solutions that can support growth.

Portfolio managers usually receive news alerts about companies in their holdings. In order to respond, the portfolio manager then needs to shift through massive amounts of data to determine if a shift in allocations is warranted. They would like to (1) detect changes in financial conversations about companies in the portfolio, (2) investigate the conversations to determine source of information, (3) compare social information with internal information, and (4) identify all funds containing the companies and also identify the allocations. These types of analytical capabilities can be gained quickly through Big Data and related solutions.

Improving Customer Intimacy

Banks and financial services companies seek to differentiate themselves by developing and delivering unique products and services for their customers. However in this very competitive industry, successful products are often copied and the customer's barrier to exit is very low. Ten years ago, a person was more likely to have a long term relationship with a bank enabling the bank to dictate the terms for current accounts, savings accounts, and mortgages. That person might have had another relationship with a discount brokerage, where the brokerage would have controlled the fee structure, the margin requirements and CD rates. The suppliers were at the center of these relationships. Today, this individual likely has multiple transient relationships with a number of banks, including an account at a bank that charges no fees, accounts in banks that offer the highest interest on savings, and mortgage loans from banks offering the lowest mortgage interest rate. The customer is now the center of attention, with the financial institutions being transient nodes.

The growth in the power of the customer is recognized by new entrants in the market – institutions built around a complete digital footprint. These institutions have raised the expectations of the individual consumer. The consumer

now expects to have real-time responses to their queries, and expects to have this discussion in public. The consumer now expects to have full transparency about the products and services being offered.

For banks and financial services companies to keep customers for the long term, they must get closer to them. They must anticipate customer needs and be able to proactively position their products. When they fail, customers will choose to get financial services elsewhere. Over time, the customers might entirely leave the institution.

The Internet of Things (IoT) introduces new consumer options connecting customers with other service providers such as retailers, airlines, and hotels. Financial services companies are developing partnerships with many of these to extend their reach and integrate their products into all areas of their customers' lives. Creating a seamless, consistent experience across multiple channels can lead to a superior customer experience and drive enhanced revenue opportunities.

Consider the case of Jane Doe. Jane is a customer of the XYZ bank. Jane is single and direct deposits her pay into a checking account. If the bank was monitoring her social media activity, they might know that she just got engaged. Positioning a loan for a wedding might be something that Jane would be interested in. Monitoring Jane's social media activity enables the bank to anticipate Jane's needs. The relationship could later grow to offering a mortgage for Jane's first house, 529 plans for her children, college for her children, and continue all the way to Jane's retirement. When social media data for an individual is not available, data about the peer group that they belong to might be used instead (e.g. "people like you also liked xx").

Improving Fraud Detection

Fraud detection and prevention is facilitated by analyzing transactional data and interdicting an incoming real-time stream of transactions against a well known set of patterns. Big Data technologies enable correlation of data from multiple sources or incidents to determine fraud. The individual incidents by themselves may not signal a fraudulent event. An example of suspicious activity might occur when a trader consistently sends an email or calls a telephone number within a few minutes of making a large trade. Adding new data points such as geo-location data can enhance fraud prevention – an ATM card being used in New York to withdraw cash while the mobile device of the customer is active in London is an indication of a likely fraudulent event. Modern self-adaptive machine learning algorithms can learn and track behaviors of customers and devices enabling identification of fraud early.

Where to Find Business Cases that Justify Projects

Many existing business capabilities can be enhanced when more and varied data becomes part of the Information Architecture.   IT organizations at banking and financial services companies typically work with their lines of business to build solutions that deliver the following when defining Big Data projects:

1) Enterprise Modeling and Analytics Platforms: Banks are building data reservoirs as places to store data extracts from all operational and non-traditional data sources. Business users and analysts explore the data in the data reservoir and develop analytic business models in a self-service environment. Big Data technologies have been applied successfully in a number of financial services use cases, but the enterprise level use of Big Data for firm-wide analytic problems remains a challenge. Building an enterprise analytics platform gives users controlled access to all the data to explore it, build models, and deploy the models.

2) Mobility and Location based Services: As credit card interchange fees come under fire, banks are trying to offer value added services to merchants. Banks have an incredible amount of information available about the buying behavior of their customers. When combined with the location of the customer, it is possible to drive the customer to visit a merchant location. For example, a time bound offer for a local restaurant that has a relationship with the bank can be made to a customer via their mobile device as they walk into a movie theatre.

3) Increased Customer Wallet Share: Understanding the life cycle of a customer enables more services to be sold to the customer over time.   The value of the customer continues to grow as more and more services are sold to them.   Social media can be a good source of data to get a head start on life events including, graduation, first job, engagement, weddings, college costs and retirement. This insight can enable more products to be sold by getting the right product in front of the consumer at the right time.

4) Customer Intimacy: Every engagement with the customer can be a selling opportunity.   Better understanding of the customer, their traits, how they like to communicate, services they consume, and their value to the business enables the right product to be positioned to the customer at the right time for the right price.

5) IT operational efficiency: Not unique to banking and financial services companies and rarely driven from the lines of business (but a possible reason for embarking on extended architectures that include Hadoop) is the need to move data staging and transformation to a schema-less platform for more efficient processing and leveraging of IT resources. IT operational efficiency is often difficult to prove but is sometimes an initial justification that IT organizations gravitate toward when deploying these types of solutions.

On the next page, we show a table that summarizes several typical business challenges in financial services and banking companies and illustrates the opportunity for new or enhanced business capability when adding new analytic capabilities.

TABLE 1 – FINANCIAL SERVICES AND BANKING SAMPLE FUNCTIONAL AREAS, BUSINESS CHALLENGES & OPPORTUNITIES

FUNCTIONAL AREA	BUSINESS CHALLENGE	OPPORTUNITY
Revenue increases and improved profit margins	Increase revenues globally Enhance profit margins	Launch new services and expand into global markets with targeted services. Enhanced cross-sell/up-sell opportunities Reduce risk and costs to boost profitability
Risk Management	Reduce exposure and optimize asset utilization Significant compute capacity required for valuations	Improve credit scoring utilizing non-traditional sources of data Utilize open technologies such as Hadoop/Spark for massively parallel compute
Fraud Detection	Detect fraudulent transaction with greater accuracy to reduce costs	Correlate seemingly unrelated incidents to identify fraud Utilize machine learning to keep up with ever changing fraudulent activity
Broker and Trade Compliance	Monitor illicit trading activity	Identify trading misconduct by correlating unstructured content such as IM chats, emails, and telephone calls with trading activity
Anti-Money Laundering	Detect fraudulent transaction with greater speed and accuracy to reduce exposure	Utilize Big Data technologies to rapidly ingest large amount of data for anti-money laundering (AML) purposes Use machine learning to reduce the number of false positives
Marketing and Customer 360	Grow the business by selling more services and products to customers.	Better understand customer behavior and needs Identify customers with long term profitability potential Proactively make context relevant offers
Reputational Risk	Protect the brand	Monitor the web to understand customer sentiment towards the bank's products, employees, board members etc.

Establishing an Architectural Pattern

The following figure illustrates key components in a typical Information Architecture. Data is acquired and organized as appropriate and then analyzed to make meaningful business decisions. A variety of underlying platforms provide critical roles. Management, security and governance are critical throughout and are always top of mind in financial services and banking companies. These components are further described in the "Information Architecture and Big Data" whitepaper posted at http://www.oracle.com/goto/ea.
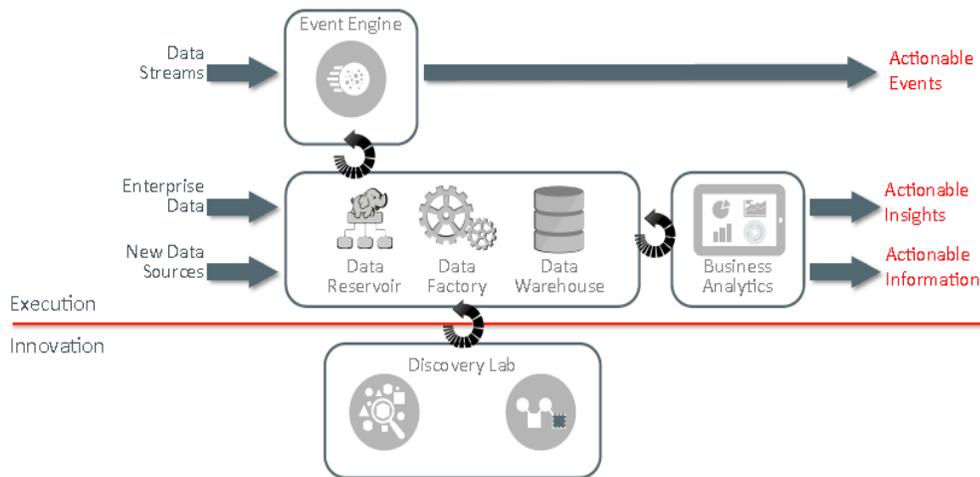
Figure 1: Key Information Architecture Components

How do we determine which of these components should be part of the architecture to meet the needs of a specific organization or company? If we create an information architecture diagram, and trace the data flow from the sources to the application (end-user), we can build a logical configuration of the components to support the functions.

The first step in defining a future state architecture is documenting the current state, its capabilities and any functional gaps. Typically a current state data warehouse environment might look something like Figure 2.
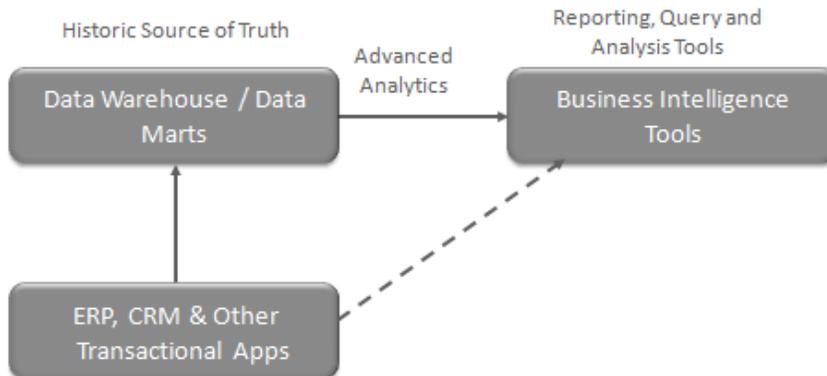
Figure 2: Typical Current State Data Warehouse

The first gap that typically has to be closed is a need to provide a more agile reporting and analysis environment where new data and ad-hoc reports are needed on an ongoing basis. Information and data discovery engines can provide this type of capability. When information discovery is incorporated into the architecture it would look something like the illustration in Figure 3.
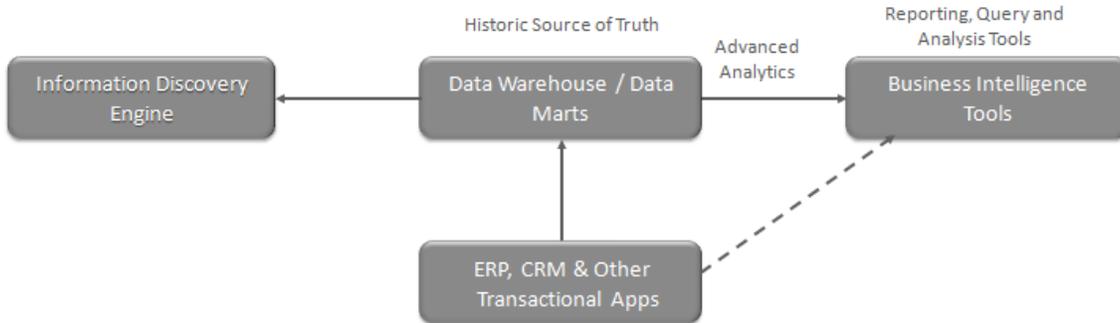
Figure 3: Typical Introduction of Information Discovery

Now that we're better able to analyze the data we have, the next step would be to explore bringing in new data and new data types. These data sets might be internal, 3rd party, structured, unstructured or of unknown structure. When storing data of unknown structure, the most efficient way to store data sets is often in a Hadoop-based data reservoir. Initially, such projects are often considered experimental in organizations and therefore they might be independent efforts separated from the traditional environments, as illustrated in Figure 4.
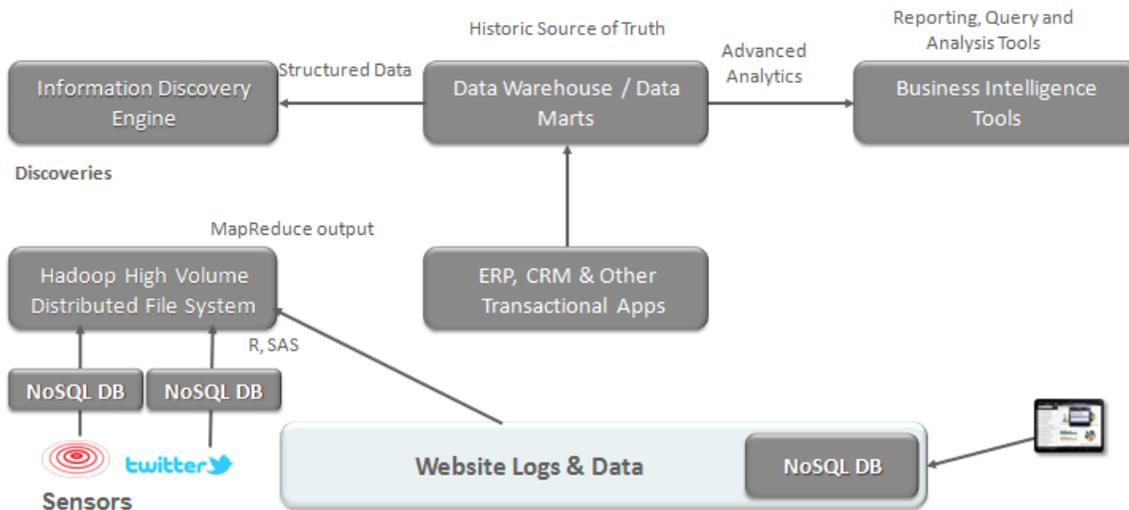
Figure 4: Typical Early Hadoop Environment separate from the Data Warehouse

The profile of the data such as how it is acquired, how it should be formatted, the frequency of updates and quality of the data will help us put the right technology in place best suited for the particular situation. We need to

understand whether real-time or batch processing is appropriate. We should understand the periodicity of processing required based on data availability. Below is a partial list of the characteristics that should be considered:

» Processing Method – prediction, analytics, query, ad-hoc reports
» Format and Frequency – external data feeds, real-time, continuous or periodic on-demand
» Data Type – web/social media, machine generated, human generated, biometric, legacy or internal, transactional
» Consumer Application – Web Browser, Intermediate processes, Enterprise Application

When business value is found in analyzing data in a Hadoop-based data reservoir, lines of business generally begin to see a need to link data there to historical data stored in their data warehouse. For example, a relationship manager might want to compare historical transactions for a commercial customer stored in the data warehouse to public data available about the industry the customer belongs to in the data reservoir. Various linkages are often established as pictured in Figure 5.
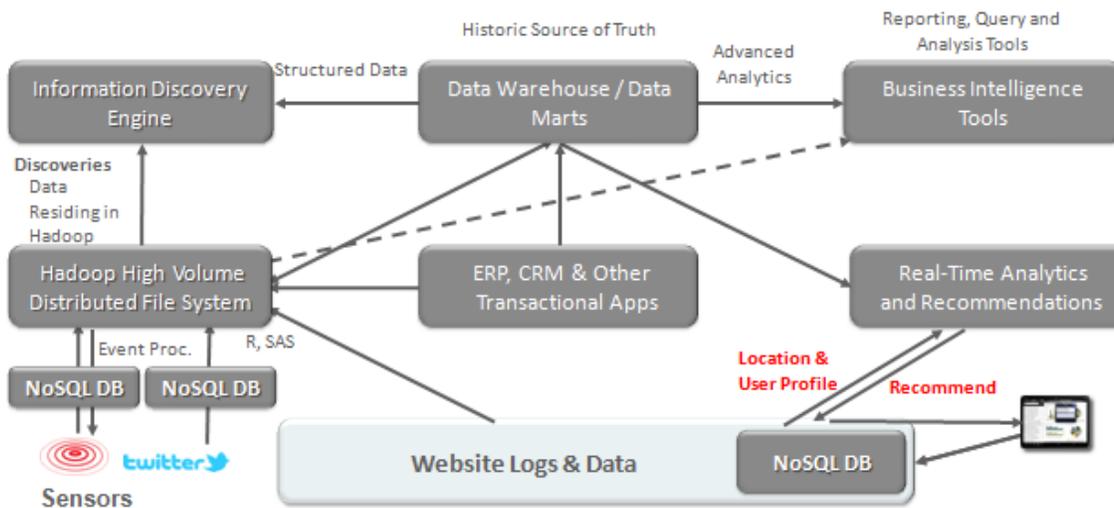
Figure 5: Integration of Hadoop Infrastructure and Data Warehouse

We also added something new to Figure 5, a real-time analytics and recommendation engine. In many situations, the latency inherent in the data movement pictured above means that the recommendation from analysis would come too late to take action in near real-time. A way around this is to perform periodic advanced analytics in the data reservoir and / or data warehouse and provide updates to a real-time recommendation engine that becomes more fine-tuned through self-learning over time.

For use cases such as fraud detection, AML, marketing, pricing, and credit scoring, having canonical data models that document the required data elements significantly reduces the amount of time required to deliver these solutions on Big Data technologies. Some of these data elements may be in a relational database, others in a Hadoop file system, while still others in a NoSQL database. However the ability to be able to identify precisely which elements are required for AML processing as mandated by the regulator can be a huge time-to-market advantage.

Giving business analysts and analytics data modeler's access to all the data can enable rapid monetization of the data reservoir. As long as the access is controlled, the data can be used for a variety of regulatory purposes

(BASEL, liquidity, AML), non-regulatory purposes (customer insight and fraud detection) and ad hoc exploratory purposes.

Emerging applications are being built upon Big Data technologies for the purpose of providing better customer insight, improved marketing, fraud detection, and other uses. The applications alleviate the burden of building the solutions from scratch. This approach offers a significant time-to-market advantage – the vendor supplies the application and specifications of the data requirements, the bank points the vendor to the bank's data sources or their data reservoir.

IT Operational ETL Efficiency

In Figure 5, you might have noticed a line pointing from the transactional sources to the Hadoop cluster. This is to illustrate a popular ETL alternative, leveraging Hadoop as a data transformation engine.

Let's now consider the type of data typically stored in today's data warehouse. Such warehouses are typically based on traditional relational databases using a "schema on write" data model. The data sources can vary, but the structure of the data is determined before the data in imported into the data warehouse. In the example below there are two data sources. These two data sources go through an ETL process to prepare the data to be loaded into the warehouse.
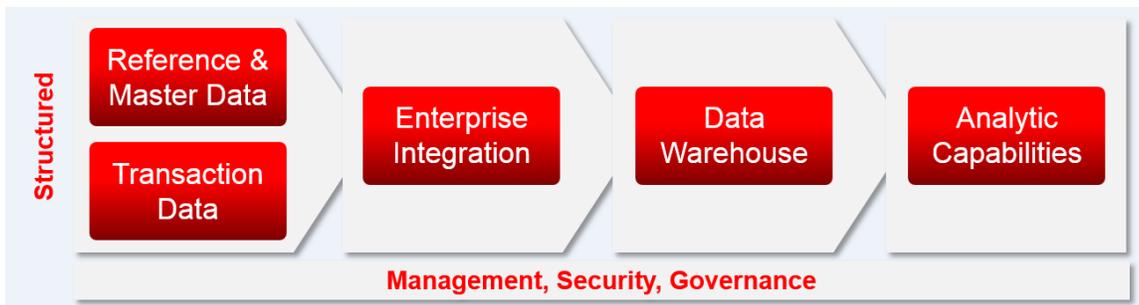
Figure 6: Structured Data and the Data Warehouse

Extending the architecture can enable a more agile workflow by incorporating data sets for which there is not rigid structure. This data model is best defined as "schema on read". That is, we store the data without the traditional ETL processing, as we don't know exactly how we want to access the data. In the example below we are using multiple data sources with varying structures.
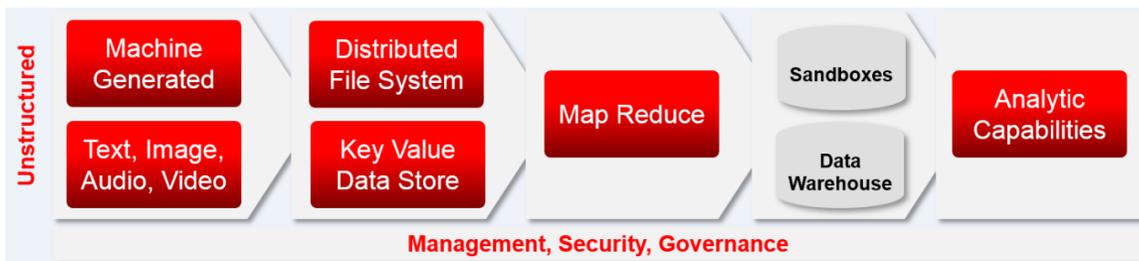
Figure 7: Unstructured / Streaming Data, Distributed File Systems and Key Value Data Stores

These two environments should not be separate and unique. Building an integrated Information Architecture that can handle data sets of known structure as well as unknown structure enables us to augment the capabilities of existing warehouses as well as leverage data center best practices that are already in place.

Oracle Products in the Information Architecture

In Figure 8, we illustrate how key Oracle products could fit in the generic architecture diagram previously shown.
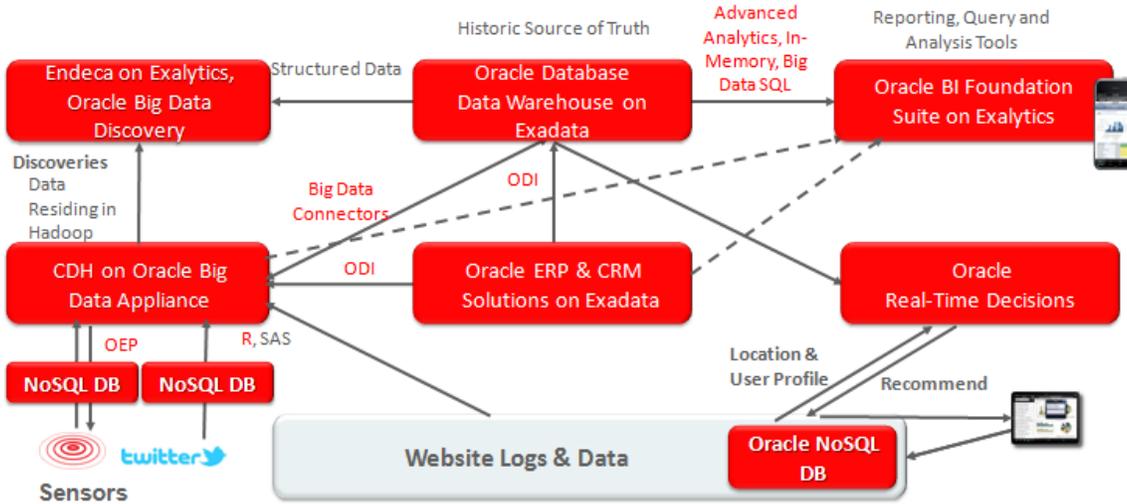

Figure 8: How Key Oracle Products Fit in the Generic Architecture

While Oracle can provide a more complete integrated solution, many organizations mix and match products from a variety of vendors. Therefore, such architecture diagrams often show such a mixture of products from Oracle and other vendors.

Along with the relevant technology components, Oracle is also a supplier of a canonical physical data model, the Financial Services Data Foundation or FSDF. It is a platform using in managing the data lifecycle and enables modelers to create models and operationalize them rapidly. The Analytical Applications Infrastructure or AAI is the basis for a series of applications built using FSDF as well as AAI for risk, capital adequacy, regulatory reporting, finance, treasury, profitability, marketing, and financial crime and compliance. Figure 9 illustrates FSDF and AAI.
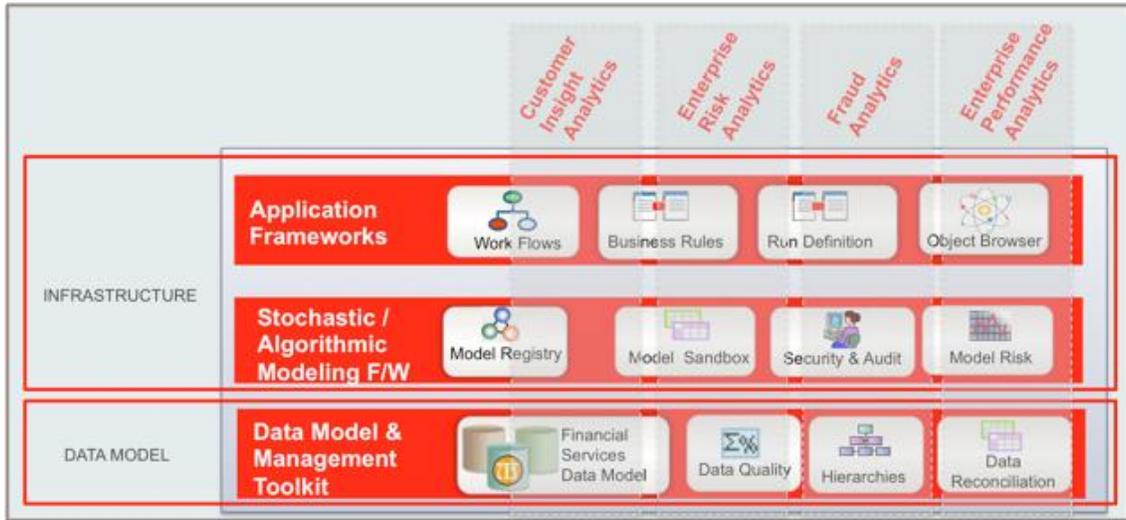
Figure 9: Oracle Financial Services Data Foundation (FSDF), Analytical Applications Infrastructure (AAI)

FSDF can be hosting using an Oracle Database, a Hadoop file system, or a combination of both. AAI is capable of ingesting any data model – including FSDF – and it can be implemented on an Oracle Database or Hadoop. It allows data movement and transformation, defining of data quality checks and their execution, deterministic rule execution, building and operationalization of models written in "R", Mahout and/or Custom C++/Java (and soon Apache Spark) and creation of reports for business intelligence Figure 10 illustrates this platform and shows the analytical workflow orchestration for the Oracle Financial Services Analytics Applications (OFSAA). Figure 11 illustrates the analytics applications functional areas.
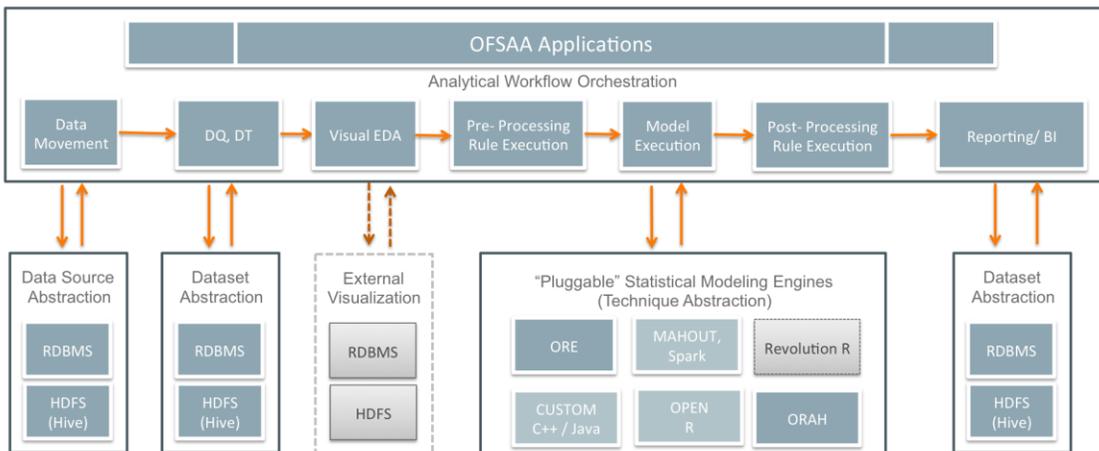
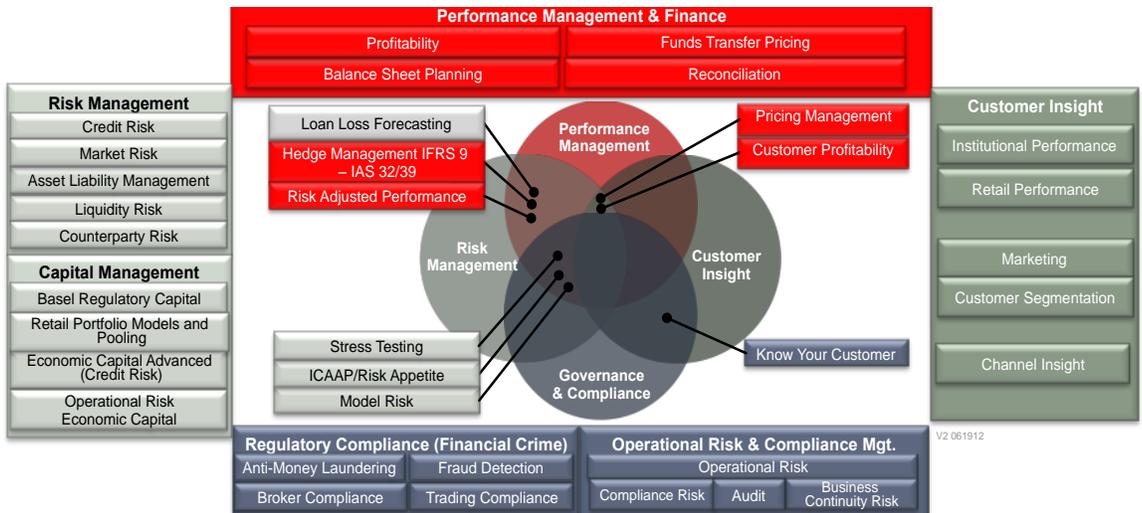Figure 10: Using AAI as an Enterprise Modeling and Analytics Platform

Figure 11: Sample Analytical Applications functional coverage

The various software capabilities required in a typical architecture might include these components:

» Relational Database Management System (RDBMS): Oracle Database 12c Enterprise Edition is designed for performance and availability, security and compliance, data warehousing and analytics, and manageability. Key data warehousing options often include In-Memory, OLAP, the Advanced Analytics Option, and Partitioning.

» A business intelligence platform that delivering a full range of capabilities - including interactive dashboards, ad hoc queries, notifications and alerts, enterprise and financial reporting, scorecard and strategy management, business process invocation, search and collaboration, mobile, integrated systems management and more.

» Financial Services Applications and Analytics

» A real-time recommendation engine.

» Hadoop Distributed File System (HDFS): A scalable, distributed, Java based file system that is the data storage layer of Hadoop. Ideal for storing large volumes of unstructured data.

» Flume: A framework for populating Hadoop with data via agents on web servers, application servers, and mobile devices.

» ETL software to provide a data integration platform that covers all data integration requirements: from high-volume, high-performance batch loads, to event-driven, trickle-feed integration processes, to SOA-enabled data services.

» Metadata Management: Data governance and metadata management tool providing lineage and impact analysis, and model versioning for business and technical metadata from databases, Hadoop, business intelligence tools, and ETL tools.

» An information discovery tool and visualization engine.

» An integrated enterprise platform management single tool used to manage both the structured and unstructured data environments and BI tools.

» An OLAP (Online Analytical Processing) Server that provides an environment for deploying pre-packaged applications or developing custom analytic and enterprise performance management applications.

Obviously, many variations are possible. Financial services companies generally rely on a mixture of ERP and other applications from vendors and highly customized data sources. Routing algorithms are often seen as the secret sauce used to run the business at the most optimal levels enabling competitive pricing and delivery.

Additional Data Management System Considerations

In defining the Information Architecture, it is important to align the data processing problem with the most appropriate technology.

When considering the choices you have in database management systems to include in an Information Architecture, you might consider if the form of the incoming data or ACID properties or fast data availability is most important. Other considerations should include manageability, interoperability, scalability, and availability. Of course, you should also consider the skills present in your organization.

Some of the various data management technologies in a typical architecture include:

Relational Databases

Typically already in use at most companies, RDBMS' are ideal for managing structured data in predefined schema. Historically they excel when production queries are predictable. Support of dimensional models makes them ideal for many business intelligence and analytics workloads. They frequently house cleansed data of known quality processed through ETL workloads. Relational databases also excel at transactional (OLTP) workloads where read / write latency, fast response time, and support of ACID properties are important to the business.

These databases can usually scale vertically via large SMP servers. These databases can also scale horizontally with clustering software.

Example RDBMS Product: Oracle Relational Database

MOLAP Databases

Typically used for highly structured data, MOLAP databases are ideal when you know what queries will be asked (e.g. facts and dimensions are predefined and non-changing) and performance is critical. These databases excel at certain business intelligence and analytics workloads.

Example MOLAP Product: Oracle Essbase, Oracle Database OLAP Option

NoSQL Databases

NoSQL databases are without schema and are designed for very fast writes. Often, they are used to support high ingestion workloads. Horizontal scale is most often provided via sharding. Java and Java scripting (JSON) are commonly used for access in many of the commercial varieties.

NoSQL databases are sometimes described as coming in different varieties:

Key Value Pairs: These databases hold keys and a value or set of values. They are often used for very lightweight transactions (where ACID properties may not be required), and where the number of values tied to a key change over time.

Column-based: These databases are collections of one or more key value pairs, sometimes described as two dimensional arrays, and are used to represent records. Queries return entire records.

Document-based: Similar to column-based NoSQL databases, these databases also support deep nesting and enable complex structures to be built such that documents can be stored within documents.

Graph-based: Instead of structures like the previous types, these databases use tree-like structures with nodes and edges connecting via relations.

Example NoSQL Database Product: Oracle NoSQL Database

Distributed File System

Not a database per se as the name would indicate, highly distributed file systems have the advantage of extreme scalability as nodes are added and frequently serve as a data landing zones or data reservoirs for all sorts of data. Read performance is typically limited by the individual node of the "system" when accessing data confined to that node, however scalability to a huge number of nodes is possible driving massive parallelism. Write performance scales well as data objects can be striped across nodes.

The most popular distributed file system used today is Hadoop. Given its role as a data reservoir, it is increasingly a location for performing predictive analytics. SQL access is available via a variety of interfaces though various levels of standards support are offered.

Example Distributed File System Product: Cloudera Hadoop Distribution (featuring the Cloudera Hadoop Distributed File System and other features)

Big Table Inspired Databases

There is an emerging class column-oriented data stores inspired by Google's BigTable paper. These feature tunable parameters around consistency, availability and partitioning that can be adjusted to prefer either consistency or availability (given these are rather operationally intensive.

 A typical use case might be where consistency and write performance are needed with huge horizontal scaling. HBase (deployed on a Hadoop Distributed File System) in particular has been deployed to 1,000 node configurations in production.

Example Big Table inspired Product: Cloudera Hadoop Distribution (Cloudera HBase)

Extending the Architecture to the Internet of Things

Thus far, we've focused on the analytics and reporting and related data management pieces of the Information Architecture. Where sensors are providing key input, the architecture for data capture, security, and linkage to the rest of the Information Architecture can require additional consideration. Banks and financial services companies put themselves in a favored position when they can connect their customers to merchants as they are credit card suppliers to their customers. The customers have smartphones with banking and other financial applications installed. They can deliver mobility services by leveraging the smartphone as a "sensor". Using this technology, they can connect the customer to nearby merchants and collect additional revenue when brokering the "connection".

The following illustrates what is often described as an Internet of Things footprint for connected customers in financial services companies:
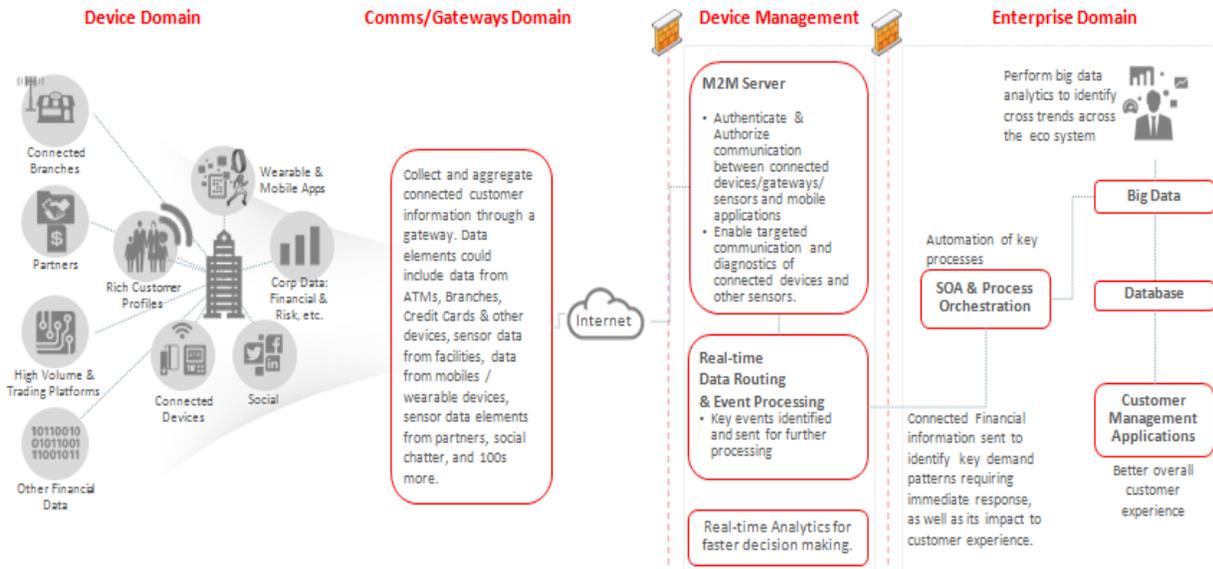
Figure 12: Connected Banking and Financial Services

Items to the far right of Figure 12 have largely been previously discussed in this paper. Many of the other items pictured are what Oracle typically describes as Fusion Middleware components. For example, much of the sensor programming today takes place using Java. Security is extremely important since most would not want unidentified third parties intercepting the data provided by the sensors. Applications closer to the sensors themselves are often written using Event Processing engines to take immediate action based on pre-defined rules. There are also various message routing, provisioning, and management aspects of such a solution.

Figure 13 illustrates a typical capability map of this architecture for connected banking and financial services:
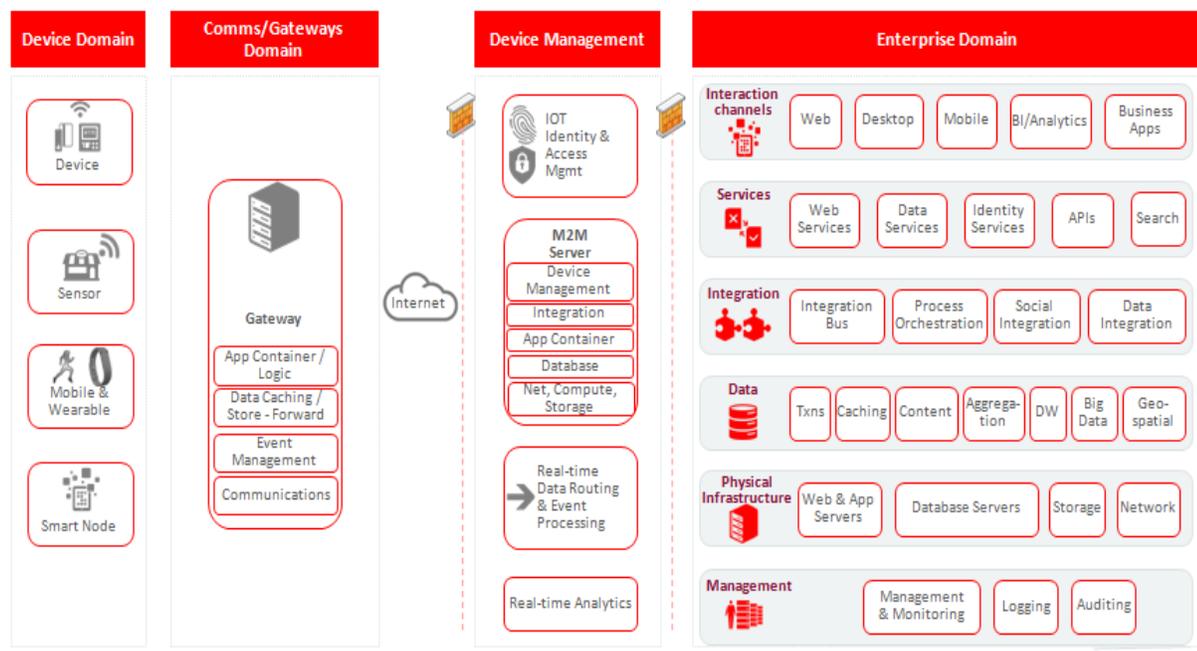
Figure 13: Connected consumers and Devices Capability Map for Financial Services

Mobile devices and sensors are increasingly providing critical usage patterns and customer preferences, and increasingly being integrated into banks and financial institutions. They are also used to enhance partner offerings by monitoring customer needs and preferences enabling improved customer experiences. The ability to respond to customer events in real-time and ability provide advice beyond traditional banking networks can help ensure higher revenues in the long term.

Today's smartphones, with various software applications, act as sensors. They tell us where they are, how they are being used. There is a wealth of data that can be gathered from cellphones, including location data. The data will continue to grow and enable financial services companies to better assess and manage the products and services that are being offered.   With this information, financial institutions can connect consumers to merchants and collect fees for the connection or collect a percentage of the transaction.

Figure 14 illustrates some of the Oracle products aligned to the previously shown capability map:
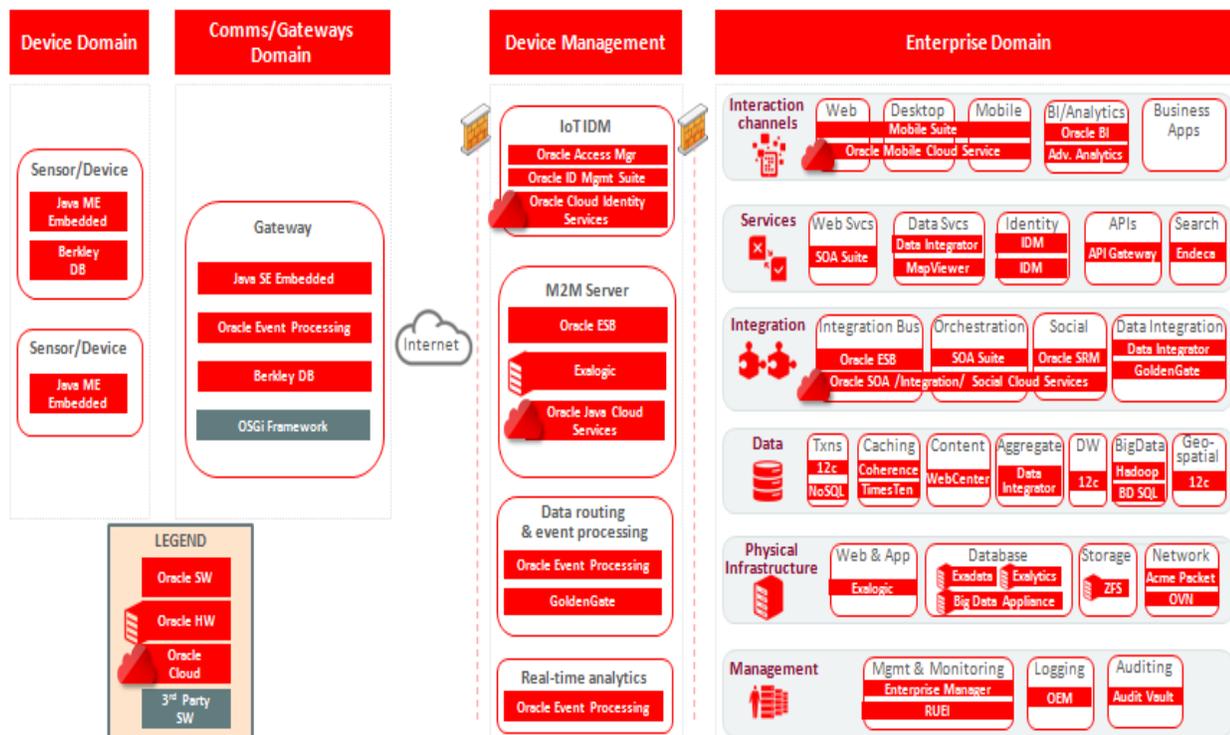
Figure 14: Oracle Products aligned to Capability Map

Keys to Success

One of the most significant keys to success in a large project undertaking is to gain alignment between the business needs and goals and with the IT architecture design and deployment plans. Key business sponsors must be engaged and active in all phases.

Methodologies based on phased approaches are almost always the most successful. To start, you'll need to understand the current state and its gaps so that you can better understand how to build towards the future state. You will need to modify the architecture as business needs change. Therefore, a common method to help assure success is to deploy quickly in well scoped increments in order to claim success along the way and adjust the plan as needed. A complete Information Architecture is never built overnight, but is developed over years from continued refinement.

Figure 15 illustrates such an approach, beginning with defining an initial vision, then understanding critical success factors and key measures tied to use cases, defining business information maps based on output required, linking the requirements to a Technical Information Architecture, defining a Roadmap (including phases, costs, and potential benefits), and then implementing. Of course, an implementation leads to a new vision and requirements and the process continues to repeat. Pictured in the Figure are some of the artifacts Oracle often helps deliver during Enterprise Architecture engagements and Information Architecture Workshops.
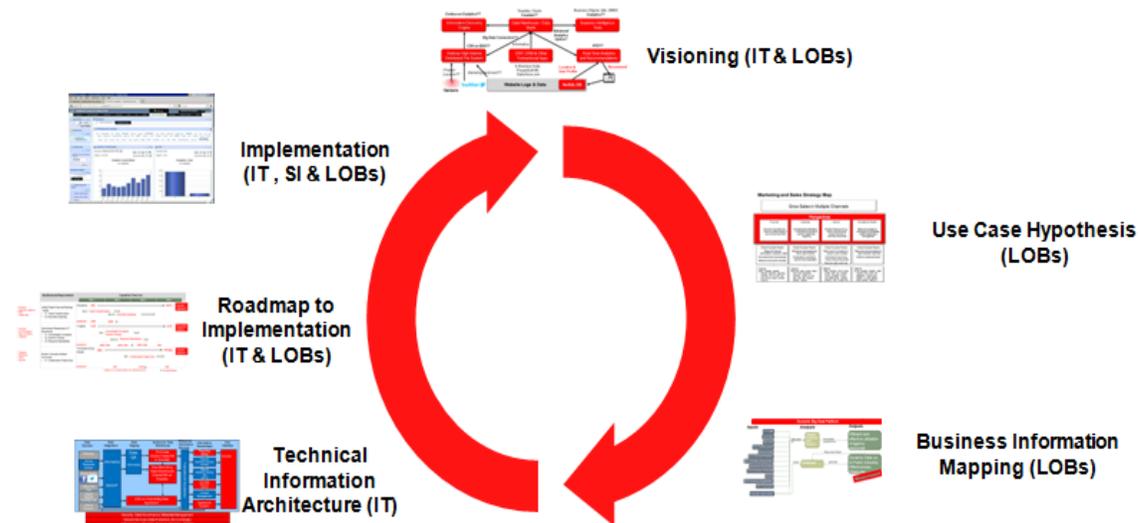

Figure 15: Typical Methodology for Information Architecture Projects

Usability needs will drive many of your decisions. Business analysts will likely have a variety of business requirements and possess a variety of analysis and technical skills. They could require solutions ranging from simple reporting to ad-hoc query capability to predictive analytics. You'll need to match the right tools and capabilities to the right users. One size does not usually fit all. While new features in the data management platforms can provide more flexibility as to where you host the data for such solutions, the data types, volumes and usage will usually determine the most optimal technology to deploy. A common best practice is to eliminate as much movement of data as possible to reduce latency.

Data security and governance are also a key consideration.   Financial services and banking companies gather sensitive data that in the wrong hands could lead to liability claims and worse. So securing access to the data, regardless of data management platforms, tools, and data transmission methods used, is critical. Data governance needs regarding the meaning of data as well as its accuracy and quality will often require close coordination with and among multiple lines of business.

Finally, as fast time to implementation important to the success of any business driven initiative, you will want to leverage reference architectures, data models and appliance-like configurations where possible. These can speed up the design and deployment and reduce the risk of incomplete solutions and severe integration challenges. Leveraging engineered systems and appliances where possible can simplify the architecture, reduce time to value and improve architecture reliability.

Final Considerations

This paper is intended to provide an introduction to applying Information Architecture techniques for banking and financial services companies. These techniques guide the extension of current architecture patterns to meet new and varied data sources that are becoming part of the information landscape. Oracle has very specific views regarding this type of information architecture and can provide even more of the individual components than were described in this paper.

The following diagram provides a conceptual future state that can encompass all types of data from various facets of the enterprise:
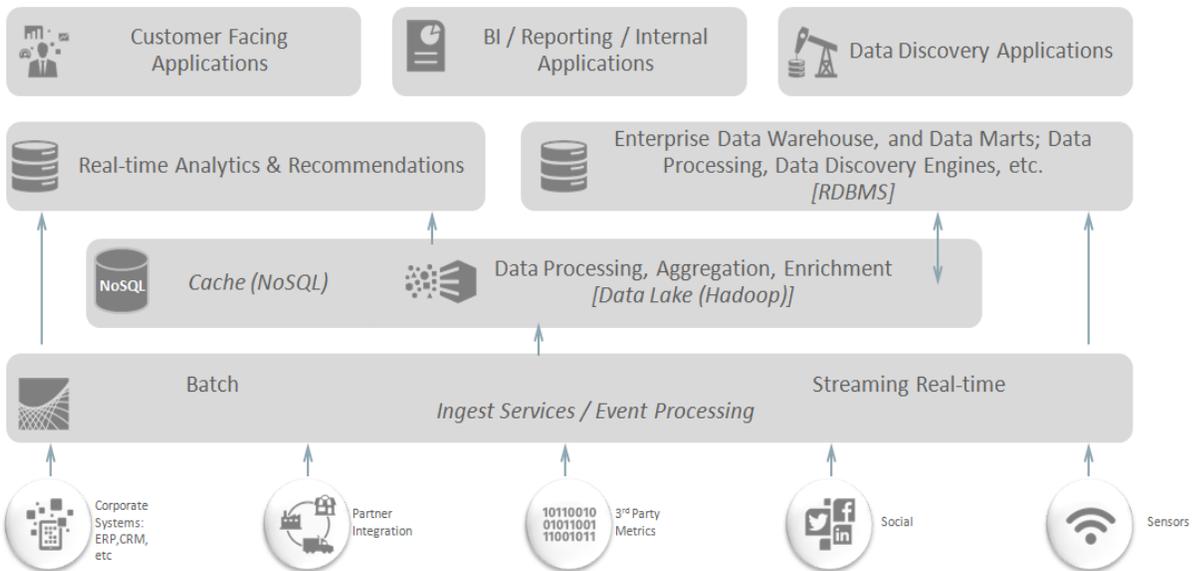
Figure 16: Typical Conceptual Future State Diagram

A more detailed look at "Business Analytics" reference architectures appears in documents posted to the Oracle Enterprise Architecture web site at http://www.oracle.com/goto/ITStrategies.

The following is a figure from one of the just referenced documents to give an idea as to the level of detail that might be considered around information delivery and provisioning.
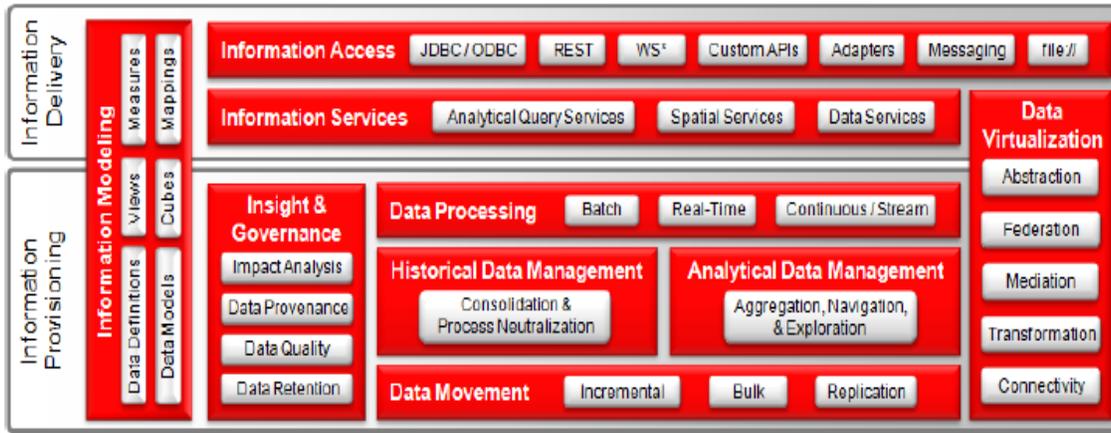
Figure 17: A more detailed Reference Architecture Diagram for Information Delivery and Provisioning

Often, the architecture discussion also leads to consideration on where to host and analyze the data (e.g. in the cloud versus on-premise). Aside from security considerations, most banking and financial services companies come to the conclusion that another motivating factor to storing the data on-premise is the volume of data being produced and a desire to minimize network data traffic. In other words, most organizations are coming to the conclusion that it makes sense to analyze the data where it lands. And once it lands, reporting and predictive analytics often take place in the data management system holding the data.

An additional consideration not addressed in this paper is the availability of skills needed by the business analysts and the IT organization. A future state architecture evaluation should include an understanding as to the degree of difficulty that a future state might create and the ability of the organization to overcome it.

The highly competitive nature of companies involved in banking and financial services will assure that those that take advantage of these new data sources to augment what they know about their business will continue to be leaders. They will continue to invent new and better business products, processes and efficiencies and they will do so by evolving their Information Architecture in an impactful manner. Some will likely leverage their advanced footprints to offer data subscriber networks, thereby going into competition with data aggregators and further monetizing their IT investments.

Oracle Corporation, World Headquarters

500 Oracle Parkway

Redwood Shores, CA 94065, USA

Worldwide Inquiries

Phone: +1.650.506.7000

Fax: +1.650.506.7200

ORACLE®

CONNECT WITH US

B blogs.oracle/enterprisearchitecture

f facebook.com/OracleEA

y twitter.com/oracleEAs

o oracle.com/EA

Hardware and Software, Engineered to Work Together

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