Improving Retail Performance with Big Data
Architect’s Guide and Reference Architecture Introduction

ORACLE ENTERPRISE ARCHITECTURE WHITE PAPER | FEBRUARY 2015
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 1
- Key Business Challenges 3
- Where to Find Business Cases that Justify Projects 4
- Establishing an Architectural Pattern 6
- IT Operational ETL Efficiency 9
- Oracle Products in the Information Architecture 10
- Additional Data Management System Considerations 13
- Extending the Architecture to the Internet of Things 15
- Keys to Success 18
- Final Considerations 20
Executive Summary

The ability to access, analyze, and manage vast volumes of data while rapidly evolving the Information Architecture is increasingly critical to retailers looking to improve business efficiency and performance. While operational efficiency, favorable customer experience, and loyalty and retention of customers remain keys to success, anticipating demand is important for more efficient inventory management, cash management and overall profitability. As retailers become larger and more diverse, the type of data that is managed becomes more complex. Analysis of this data can lead to better understanding of what products drive the highest profitability per square foot. For example, it might lead to a decision to stock more costume jewelry and fewer washing machines as the jewelry takes less space and turns over in sales more often, but the retailer must understand if there are enough jewelry sales to offset sales of the much higher priced washing machines. These are the types of analysis that retailers must make on a daily basis.

Retailers have long gathered customer data tied to loyalty cards, the majority of which show what customers previously purchased. The data illustrates past buying patterns, but might not be indicative of future demand. Utilizing additional data sources can help retailers gain a better understanding of future customer demand, as well as gain a better view of the customer and customer family/network buying patterns. These data sources can include:

» Social Media
» Web browsing patterns
» Traditional enterprise data from operational systems
» Data from data aggregators (Nielsen, IRI, etc.)
» Advertising response data
» Demographic data
» Weather forecasting and monitoring systems

The rate that this data is generated is rapidly increasing leading to higher rates of consumption by various business analysts who crave more information. This increase in data velocity and sources naturally drives an increase in aggregate data volumes. Business analysts want more data to be ingested at higher rates, stored longer and want to analyze it faster. “Big Data” solutions help to enable retailers to meet 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 in the dynamic retail market.
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 faced in the course of their architecture planning and implementations. Oracle’s advising architects work across many industries and government agencies and have developed 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

Retailers historically used data warehouses (including 3rd party data) and business intelligence tools to report on and analyze customer behavior and 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 such as:

» Up-selling / cross-selling to existing customers based on their family purchasing patterns
» Formulation of effective and targeted marketing campaigns and gain higher value from the money spent
» Effective and streamlined operations to meet customer demands, while providing a rich customer experience resulting in higher customer retention
» Better ability to predict demand and customer preferences to optimize retail operations

Improving Customer Intimacy

Big Data and advanced analytics solutions enable retailers to leverage data from their internal systems (Point of Sales or POS, Inventory, ships/receipts, loyalty and ERP systems) and external systems (weather, market share data, census/demographic data, etc.) by providing better visibility into individual customer purchase patterns with relevant contextual background information. That information can drive customer-centric offerings on a one-to-one basis. For example, it is possible to implement proactive 1:1 marketing programs that best fit the needs of EACH customer. In effect, retailers can create pricing programs (and virtual stores) that match the needs of each individual customer.

Analysis of individual customer spending patterns can set baselines for comparison to fraudulent or unusual usage patterns. Working in partnership with credit processing organizations, real time analytics can be used to alert customers and the retailer to abnormal activity that can either represent an untapped sales opportunity or a loss prevention opportunity. For example, if the retailer detects that expensive jewelry is being purchased with traceable tender associated to a known customer that doesn’t normally purchase such things, it could be for a special event such as a wedding, or it could be fraudulent. Either way, it is worth the retailer’s attention.

Improving Operational Efficiency

Predictive analytics can be used to minimize out of stock conditions and distribution reliability by anticipating demand and taking appropriate steps before this condition occurs. These same analytics can be used to identify trends and forecast demand.

For example, consumption of certain products (batteries, generators, canned foods, plywood, camping gear, etc.) is often strongly connected to weather patterns. The ability to predict an oncoming cold front could be used for proactive allocation of supplies and determination of the effect on storage distribution across regions. Understanding these patterns a couple of weeks in advance (as well as their possible impact on consumer behavior) could be used to influence retailer planning.

Prediction is especially useful to better match supply to demand for perishables in order to avoid spoilage and waste. Such analytics can enable the retailers to better anticipate demand and adjust supply. Accurate hourly demand predictions for perishables can enable a retailer to decide to take a scheduled delivery of items about to sell, keep their inventory fresh, and utilize optimal shelf and storage space. Better analytics also enable retailer to implement and/or improve efficiency programs as well as improve communications with their suppliers.
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 retail companies typically work with their lines of business to build solutions that deliver the following when defining Big Data projects:

1) **Improved Margins through Better Localized Merchandise**: Store managers want differentiated merchandise assortments based on local customer profile information. Transactional data, data available via social media, local weather forecasts, and other data sources are used to assure the right products are available at the right time. The retailer monitors local promotion success, wants to predict the success of such promotions as quickly as possible, and adjust promotions accordingly when sales are not meeting expectations.

2) **Reduced Stock-outs**: The key to reducing stock-outs is accurate demand forecasting driving optimal inventory planning not only by analyzing the goods being sold and in warehouses, but also by understanding which suppliers are best able to react to changing demands. Reducing stock-outs is critical to increased sales and satisfied customers.

3) **Optimal In-store Execution**: Store managers need optimal store staffing to provide desired level of service, monitor shelf availability and product placing impacting sales, and predict potential inventory loss due to spoilage and theft. Intelligent markdown and clearance policies must be established based on the ability to predict the fair market price for products as they age and understand the pricing of similar products at competitors.

4) **Increased Customer Wallet Share**: Retailers seek to understand customer segments, perform market basket analysis, and target promotions across brick and mortar stores and web sites (leveraging both channels together where appropriate). Significant sales increases can be gained when channels are optimized around customer behavior.

5) **IT operational efficiency**: Not unique to Retail 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 retail companies and illustrates the opportunity for new or enhanced business capability when adding new analytic capabilities.
<table>
<thead>
<tr>
<th>FUNCTIONAL AREA</th>
<th>BUSINESS CHALLENGE</th>
<th>OPPORTUNITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Store Operations &amp; Loss</td>
<td>Understand the financial health of a store</td>
<td>Increase comparable same store sales</td>
</tr>
<tr>
<td>Prevention</td>
<td></td>
<td>Increase comparable sales over time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Understand competitors store sales</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measure Current vs. planned sales</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Understand goods selling price and margin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Increase shopper conversion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Match employee performance / commissions</td>
</tr>
<tr>
<td>Merchandise &amp; Category Management</td>
<td>Maximize demand for products</td>
<td>Perform market basket analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Compare store / dept. sales</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measure time on shelf for goods, turns</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measure top and bottom sellers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Determine items for promotions / discounts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Create competitive pricing / positioning</td>
</tr>
<tr>
<td>Point of Sales</td>
<td>Optimal conversion of shopping into sales</td>
<td>Measure items sold in market basket</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Determine customer wait time / shopping cart abandonment</td>
</tr>
<tr>
<td></td>
<td>Payment processing</td>
<td>Determine payment fraud detection (in partnership with financial partners)</td>
</tr>
<tr>
<td>Order Management &amp; Inventory</td>
<td>Right amount and mix of products in the right stores / channel</td>
<td>Determine stock in stores, inventory</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Optimize space utilization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improve warehouse efficiency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improve product quality / safety</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improve supplier compliance</td>
</tr>
<tr>
<td>Promotions</td>
<td>Optimal advertising and pricing to drive maximum sales</td>
<td>Improve advertising effectiveness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improve cross-sell / up-sell</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Determine loss leaders / impact on spend</td>
</tr>
<tr>
<td>Workforce Management</td>
<td>Right level of personnel</td>
<td>Associate sales / loss</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Determine employee efficiency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improve employee safety</td>
</tr>
<tr>
<td>Customer</td>
<td>Growing customer repeat spend</td>
<td>Customer satisfaction / loyalty</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Right category mix &amp; brand affinity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Demographic profiling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Customized customer experience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single view of customer across channels</td>
</tr>
</tbody>
</table>
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 retail companies. These components are further described in the “Information Architecture and Big Data” whitepaper posted at http://www.oracle.com/goto/ea.

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 the current state data warehouse environment might look something like Figure 2.
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.

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 tapes. 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.
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 business analyst might want to compare historical transactions for a shipment stored in the data warehouse to sensor data tracking that shipment in the data reservoir. Various linkages are often established as pictured in Figure 5.

![Figure 5: Integration of Hadoop Infrastructure and Data Warehouse](image)

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.
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 is 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.

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.

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.

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.

Defining an Information Architecture is all about linking it to a specific use case. For example, a use case that includes retail operational sources and the Oracle Retail Data Model for analyzing various aspects of retail operations might look like Figure 9:
The various software capabilities required in a typical architecture might include these Oracle components:

- Oracle 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.

- Oracle Business Intelligence Enterprise Edition (OBIEE): A business intelligence platform that delivers 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.


- 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.

- Oracle Data Loader for Hadoop: A connectivity toolset for moving data between the Oracle RDBMS and the Hadoop environment.

- ODI: Oracle Data Integrator is a comprehensive 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.

- Oracle Enterprise 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.

- Endeca: An information discovery tool and engine.

- Oracle Big Data Discovery: A Hadoop-based information discovery tool.
Oracle Big Data SQL: An optimal solution for querying an Oracle Database on Exadata and combining the results with data that also answers the query and resides on Oracle’s Big Data Appliance.

ORE: Oracle R Enterprise enables analysts and statisticians to run existing R applications and use the R client directly against data stored in Oracle Database (Oracle Advanced Analytics Option) and Hadoop environments.

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

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

The software products listed above can be deployed in an integrated environment leveraging these engineered systems:

- Big Data Appliance (BDA): Eliminates the time needed to install and configure the complex infrastructure associated with build-out of a Hadoop environment by integrating the optimal server, storage and networking infrastructure in a rack.
- Exadata: Streamlines implementation and management while improving performance and time to value for Oracle relational database workloads by integrating the optimal server, storage and networking infrastructure.
- Exalytics: Provides an in-memory server platform for Oracle Business Intelligence Foundation Suite, Endeca Information Discovery, and Oracle Essbase.

Obviously, many variations are possible. For example, a solution might be focused primarily on relational data and leverage a data model specific to the retail industry that Oracle can provide. The following figure shows how the Oracle’s retail solutions can provide a wide breadth of retail business intelligence for managing the business.
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. The following illustrates what is often described as an Internet of Things footprint for connected Retail:

![Figure 11: Connected Devices in Retail](image)

Items to the far right of Figure 11 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 12 illustrates a typical capability map of this architecture for connected Retail:
Many retailers are simply gathering data from devices where all of these software components are already embedded in a purchased solution. However, there can be opportunities to customize the actions taken and the information gathered using sensors depending on vendors engaged, and so that portion of the architecture may not be out-of-scope for some projects. Some of the examples of Internet of Things applications being explored by retailers today include monitoring of refrigeration units (for possible failure and proactive maintenance), power utilization and control in stores and warehouses, analysis of data from cameras mounted in stores for stock-out situations and product placement, and monitoring of perishable items during transportation among facilities for possible temperature extremes.

Sensors are increasingly providing critical data regarding the retail operations and customers. This data will continue to grow and enable retailers to better determine the status of and manage people, equipment, and services that are being offered.

Figure 13 illustrates some of the Oracle products aligned to the previously shown capability map:
Figure 13: 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 14 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.

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. Retail companies gather sensitive data that in the wrong hands could lead to lawsuits and loss of customer trust. 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 retailers. 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:

![Typical Conceptual Future State Diagram](image)


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.
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 retailers 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.

Retailers are at a key moment in history where more data is available than any time in history and much more can be gathered. Those companies that lead the industry will take advantage of this data to invent new and better business processes and efficiencies and they will do so by evolving their Information Architecture in an impactful manner. Some are even leveraging the advanced footprints and data to start their own 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

Copyright © 2015, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 0215

February 2015
Oracle Enterprise Architecture White Paper – Improving Retail Performance with Big Data
Author: Art Licht, Robert Stackowiak, Louis Nagode, Venu Mantha