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

ORACLE ENTERPRISE ARCHITECTURE WHITE PAPER | NOVEMBER 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 8
IT Operational ETL Efficiency 11
Oracle Products in the Information Architecture 12
Additional Data Management System Considerations 14
Extending the Architecture to the Internet of Things 16
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 successful operation of leading agribusinesses. Many agribusinesses are looking for ways to improve production techniques and yields and improve forecasting in order to better optimize supply chains – all leading to overall operational efficiency improvements. These operational improvements, combined with favorable public perception of the methodologies utilized, remain keys to driving success. Gaining new insight through information is critical to maintaining or growing market share for products. As agribusinesses become larger and more diverse, the growing volumes of data that must be managed are also becoming more complex. External data from social media outlets and supplier network channels combined with sensor and machine data coming from farm equipment and in the farm fields augment traditional sources of data.

Big Data solutions can help improve forecasting and operational efficiency and lead to improved and timely decision making. Big Data based architecture enables organizations to analyze a variety of data sources for improved insights. This, in turn, broadens the analytics and predictive options leading to better outcomes. Today these data sources can include:

» Traditional enterprise data from operational systems
» Farm field sensor data (e.g. temperature, humidity, rainfall, sunlight)
» Farm equipment sensor data (from tractors, plows, and harvesters)
» Harvested goods and livestock delivery vehicles (from farms to processing facilities) sensor data
» Commodities trade data
» Financial forecast data
» Weather data
» Animal and plant genomics research data
» Social media data

Agribusinesses have long lists of needed metrics. But what is needed now more than ever is the ability to obtain actionable information from these growing piles of data. Following is a list of top areas where Big Data technologies can impact an agribusiness:

» Improved forecasting of yields and production.
» Better optimized seeds and livestock and new methodologies that improve yields and production.
» Faster delivery of goods produced to distribution centers and consumers.
» Real-time decisions and alerts based on data from fields and equipment.
» Integrated production and business performance data for improved decision making.
» Rationalized performance data across multiple geographies.
» Supplier / distributor performance analysis and improved interactions and negotiations with suppliers and distributors.

Big Data could form the foundation for a variety of new capabilities, including identifying correlations between farm field and weather and commodity data for optimal irrigation, fertilization, and harvesting of crops and optimal feeding and shipping of livestock to market. More timely scheduling maintenance of equipment and minimization of energy usage can enable greater operational efficiency.

This increase in data velocity, variety, and value naturally drives an increase in data volumes companies must process. Data must be ingested at higher rates, stored longer, and analyzed faster to be of timely use. New system technology is required to enable the company to handle these requirements and remain competitive and responsive to customer demands.

This paper provides introductory guidance for the adoption of big data and analytic capabilities as part of a “next-generation” architecture that can meet the needs in the next phase of evolution in agribusiness. The 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 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

Agribusiness companies historically used data warehouses and business intelligence tools to report on and analyze production and yields, supply chains, 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 and the business can become more agile.

Modern agribusiness features tighter relationships between agriculture suppliers and the farming community. The supplier relationship is becoming one of a trusted advisor and also a research and development partner. Sensors are becoming prevalent on modern farms, gathering data in the fields and from farm equipment and providing useful information to the farmer and also to the suppliers. Sensors are also becoming common on shipping containers and delivery vehicles for optimization of delivery of goods.

Higher yield and reduced support costs are central to driving profitability and better customer experience for any major agribusiness. Big data technologies can enable improved analysis of yield and quality data, supplier’s quality data, and other critical measures for a rich and thorough root-cause analysis resulting in actions for enhanced quality and reduced overall cost. Data related to throughput, capacity utilization, and overall equipment effectiveness, can be combined for further analysis for improved quality.

Predictive analytics can be used to anticipate demand for seeds, fertilizers and animal feed and enable the agribusiness supplier to take appropriate steps to match production to demand. New pricing programs can be established to help manage demand consistent with available supply.

For example, demand for some products is often strongly connected to commodity pricing. The ability to better predict pricing changes could be used for pro-active allocation of supplies and determination of the effect on storage distribution across regions. Having insight and understanding the correlations and effects of even minor weather patterns on the supply and demand could be used to influence major business decisions.
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 typically work with their lines of business to build solutions that deliver the following when defining Big Data projects:

1) Farm Production:
   For growers and livestock producers, a critical goal is to optimize farm operations for maximum yield and while reducing cost. This requires detailed analysis of field and equipment data from sensors combined with other critical data elements for an integrated analysis. Farmers raising livestock want to optimize the size of their herds and the value. Today, agribusinesses supplying farmers are creating tools that help optimize production and analyze the results.

2) Supply Chain Management:
   Common and key dependency points can be a major source of risk and broadly disrupt supply in the value chain if the agribusiness is not properly prepared. While such impact can’t necessarily be fully avoided, minimizing the initial impact and other ripple effects during recovery is important. The overall supply risk or threat can only be minimized with careful planning, processes, and tools that provide objective data to make the right decisions.

   As sensors become more prevalent in transportation throughout the supply chain, they can provide data enabling greater transparency than has ever been possible. Such data will dwarf today’s data warehouses and require Big Data Management systems for processing and reporting.

3) Support:
   Support provided by agribusiness suppliers is increasingly critical to building partnerships with the farming community. Providing data and analyses builds the partnership and must be delivered in an easily consumable manner. Many of the data sources often consist of streaming data requiring a Big Data Management system for analysis.

4) Product Development:
   Agribusinesses spend tremendous sums of money on research and development. For example, today many are using advanced plant and animal genome analysis and use Big Data Management systems for analysis.

5) Sales and Marketing:
   Closer relationships between agribusiness suppliers and farmers results in growing data volumes from web logs and other interfaces into data that is provided and enables more targeted sales and marketing. Such data is also used for new business development and product planning by the agribusiness supplier.

6) Transportation and Logistics:
   Delivery of seeds, feeds, and fertilizers to farms in a timely and efficient manner requires careful planning and logistics management. Pickup and delivery of harvested crops, farm animals, and other products must also be efficient to avoid spoilage and other problems. Sensors on vehicles and shipping containers provide information on location and temperature and can be tied to other information such as capacity of vehicles.

7) IT operational efficiency:
   Not unique to agribusiness 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. IT operational efficiency is often
difficult to prove but is sometimes an initial justification that IT organizations use when deploying these
types of solutions.
Table 2 - Functional Areas, Business Challenges & Opportunities

<table>
<thead>
<tr>
<th>FUNCTIONAL AREA</th>
<th>BUSINESS CHALLENGE</th>
<th>OPPORTUNITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farm Production</td>
<td>Quality and yield challenges</td>
<td>Improved quality and yield</td>
</tr>
<tr>
<td></td>
<td>Crop revenue fluctuation impacting growers</td>
<td>Link commodity pricing forecasting to crop planning</td>
</tr>
<tr>
<td></td>
<td>Lack of growing efficiency</td>
<td>Optimal plowing, seeding, fertilizing</td>
</tr>
<tr>
<td></td>
<td>Suboptimal livestock production and reproduction</td>
<td>Bigger, stronger, more productive (and reproductive) animals</td>
</tr>
<tr>
<td></td>
<td>Changing climate conditions</td>
<td>Proactive adjustment of fertilization, irrigation, and timing of key events such as harvesting</td>
</tr>
<tr>
<td>Supply Chain</td>
<td>Forecasting mismatch resulting in suboptimal availability of supplies</td>
<td>Improved demand forecasting resulting in optimized supply</td>
</tr>
<tr>
<td></td>
<td>Reduction of supplier costs and optimal supplier contract negotiations</td>
<td>Enhanced negotiation capabilities with supplier with enhanced supplier performance information</td>
</tr>
<tr>
<td>Support</td>
<td>High supplier support costs</td>
<td>Reduced support costs</td>
</tr>
<tr>
<td></td>
<td>Inconsistent quality of grower experience and supplier responsiveness</td>
<td>Become trusted partner to growers and deliver information in an easily consumable manner</td>
</tr>
<tr>
<td></td>
<td>Inconsistent safety procedures in the usage of products</td>
<td>Guide growers on how to safely utilize products</td>
</tr>
<tr>
<td>Product Development</td>
<td>Suboptimal seed for growing conditions including soil types &amp; weather patterns</td>
<td>Better forecasting of field conditions and better seeds through genome research &amp; testing</td>
</tr>
<tr>
<td></td>
<td>Suboptimal fertilizers for growing conditions and current pricing of available fertilizer components</td>
<td>Better forecasting of weather / field conditions and fertilizer commodity component pricing</td>
</tr>
<tr>
<td></td>
<td>Suboptimal livestock feed &amp; heredity producing in lower valuation</td>
<td>Better testing of feeds, better understanding of heredity impact (e.g. genome research)</td>
</tr>
</tbody>
</table>
| Sales & Marketing | Sales inefficiency and demand response  
| Sales forecasting inconsistency  
| Sales campaign and spend inconsistency | Targeted sales / marketing based on grower and geographic profiles  
| Effective marketing campaigns  
| Gain new customers, cross-sell / up-sell to existing, optimal spend |
| Business Development | Sales and revenue growth challenges | Identification of new product and market opportunities  
| Focused efforts to higher business impact |
| Transportation & Logistics | Suboptimal delivery of seed & fertilizers to growers  
| Suboptimal pick-up and delivery of farm products | Delivery of right seeds, fertilizers to right locations at right time  
| Optimized pick-up and efficient delivery of harvest and farm products to processors |
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 companies dealing with business sensitive data and industry regulations. Clearly these challenges apply to many agribusinesses where compliance along with rapid innovation is a key driver of IT. These components are further described in the “Information Architecture and Big Data” whitepaper posted at [http://www.oracle.com/goto/ea](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 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.

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.

![Figure 8: How Key Oracle Products Fit in the Generic Architecture](image)

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.

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.

- **Oracle Real-time Decisions:** A real-time recommendation engine.

- **Oracle Event Processing / Stream Explorer:** Real-time event-driven engine for applying rules based on streaming data.

- **Hadoop Distributed File System (HDFS):** A scalable, distributed file system that is the data storage layer of Hadoop. Ideal for storing large volumes of unstructured data. Key options available from Oracle include the Big Data Connectors and the Spatial and Graph Option.

- **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.

---

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](image)

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.

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.

- **Oracle Real-time Decisions:** A real-time recommendation engine.

- **Oracle Event Processing / Stream Explorer:** Real-time event-driven engine for applying rules based on streaming data.

- **Hadoop Distributed File System (HDFS):** A scalable, distributed file system that is the data storage layer of Hadoop. Ideal for storing large volumes of unstructured data. Key options available from Oracle include the Big Data Connectors and the Spatial and Graph Option.

- **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, Oracle Big Data Discovery, Endeca Information Discovery, and Oracle Essbase.
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. Next generation agribusinesses are built upon a connected ecosystem. Connecting sensors, via the next generation Internet of Things (IoT) based architectures, will further enable better and faster analyses and better outcomes. 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. Figure 9 illustrates what is often described as an Internet of Things footprint in agribusiness.

Figure 9: Internet of Things for Agribusiness

Items to the far right of Figure 9 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 10 illustrates a typical capability map of this architecture for connected agribusinesses:
Sensors are increasingly providing critical information gathered in farm fields, the transportation network, research, and for nearly every aspect of agribusiness.

Figure 11 illustrates some of the Oracle and other products aligned to the previously shown 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 12 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. 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 agribusiness. 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)

Figure 13: Typical Conceptual Future State Diagram


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). Many agribusinesses have come to the conclusion that it makes sense to analyze and store data where it initially lands to minimize network data traffic. And once it lands, reporting and predictive analytics often take place in the data management system holding the data. Cloud initiatives are becoming commonplace for test and development. Full scale production might take place either in the cloud or on-premises depending on future growth characteristics, anticipated costs, and business goals.

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 increasingly competitive nature of institutions recruiting a limited number of students will assure that those that take advantage of these new data sources to augment what they already know will continue to be leaders. They will continue to invent new and better business processes and efficiencies and they will do so by evolving their Information Architecture in an impactful manner.
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.

November 2015
Oracle Enterprise Architecture White Paper – Improving Agricultural Business Performance with Big Data
Author: Robert Stackowiak, Jeff Pohlmann, and Venu Mantha.