Improving Logistics & Transportation 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 ....................................... 2
- Where to Find Business Cases that Justify Projects ........ 3
- Establishing an Architectural Pattern ......................... 5
- IT Operational ETL Efficiency .................................. 8
- Oracle Products in the Information Architecture ............ 9
- Additional Data Management System Considerations ....... 12
- Extending the Architecture to the Internet of Things ....... 14
- Keys to Success .................................................. 17
- Final Considerations ............................................ 19
Executive Summary

The ability to access, analyze, and manage vast volumes of data while rapidly evolving the Information Architecture has long been critical to logistics management and transportation companies as they improve business efficiency and performance. While operational efficiency and favorable customer experience and differentiation remain keys to success, anticipating demand and optimizing the routing of goods and services also helps to maximize overall profitability. As logistics management and transportation networks become larger, more complex and driven by demands for more exacting service levels, the type of data that is managed also becomes more complex.

Today, these data sources can include:

» Traditional enterprise data from operational systems
» Traffic & weather data from sensors, monitors and forecast systems
» Vehicle diagnostics, driving patterns, and location information
» Financial business forecasts
» Advertising response data
» Web site browsing pattern data
» Social media data

The rate that this data is generated is rapidly increasing leading to higher rates of consumption by the business analysts who crave such 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 logistics and transportation companies 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 logistics and transportation 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

Companies focused on logistic management and transportation historically used data warehouses and business intelligence tools to report on and analyze customer behavior, optimize operations, and build advanced routing solutions. 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.

Improving Customer Partnerships & Levels of Service

Logistics management and transportation companies seek to differentiate themselves by becoming a reliable partner in the delivery of goods and services. When successful, they are often seen as delivering a critical portion of the supply chain. They must deliver in a predictable and safe manner and offer transparency into the delivery process. As business needs change over time, they must also intelligently forecast the likely impact so as to have the right transport vehicles, crew and facilities available. Hence, predictive analytics has an important role in assuring proper plans are made.

As sensors become more prevalent in transportation vehicles, shipping, and 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.

Improving Operational Efficiency

Predictive analytics is typically used with data warehouse solutions today to optimize routing of vehicles, crew and goods. Such routing is fundamental to providing the lowest possible cost of transportation to customers while maintaining margins. It is also critical for understanding potential maintenance challenges and safety issues.

Sensor driven data will increasingly provide more information about the vehicles and goods. In some cases, monitoring this data and taking automated actions (such as stopping a locomotive too close to another train) will be mandated by government regulations. However, there is also an opportunity to better understand when maintenance needs to occur and, by monitoring the changing state of key components in near real-time, affording even greater efficiencies and cost savings. So predictive analytics solutions deployed across Big Data Management Systems (including Hadoop) will likely become common practice to increase overall safety, reliability, understand potential warranty issues, and reduce cost.
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 logistics management and transportation companies typically work with their lines of business to build solutions that deliver the following when defining Big Data projects:

1) **Accurate Business Demand Planning**: Network resource planners need to accurately forecast future business demand to assure they have the right fleet resources and crews in the right place to meet promised delivery schedules. They also need to assure they have the right inventory in the right distribution centers and warehouses.

2) **Route Optimization**: The key to on-time delivery and optimal load factors is optimal crew and equipment routing. Airlines, trucking companies, and railroads need to understand the implications of changing weather conditions, missed connections, crew work time limits, maintenance schedules, and a host of other factors. Sensors are becoming a significant new source of critical data. Telematics can be applied to insurance to introduce new data monetization opportunities. The claims-to-premiums ratio can be reduced by analyzing risk profiles based on real-time driving statistics for people. Also, additional location based services can be provided based on routing and route optimization.

3) **Increased Customer Wallet Share**: Customer service professionals must deliver goods in time to meet service level agreements and be seen as an integrated partner in supply chains and delivery. Where customer needs go outside the network, partnerships become critical to offering a more complete solution.

4) **Risk Analysis**: Logistics companies must understand risk associated with lengthy and complicated crew work schedules and potential losses caused when perishable items are delayed in delivery or subject to temperature extremes, or when items are mishandled and damaged. Vehicle and driver risk analysis based on sensor and geo data will improve driver safety and reduce warranty costs by avoiding potential costly recalls.

5) **IT operational efficiency**: Not unique to logistics management and transportation 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 logistics management and transportation companies and illustrates the opportunity for new or enhanced business capability when adding new analytic capabilities.
TABLE 1 – LOGISTICS & TRANSPORTATION FUNCTIONAL AREAS, BUSINESS CHALLENGES & OPPORTUNITIES

<table>
<thead>
<tr>
<th>FUNCTIONAL AREA</th>
<th>BUSINESS CHALLENGE</th>
<th>OPPORTUNITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution Center &amp; Warehouse Management</td>
<td>Properly sized and located distribution centers and warehouses</td>
<td>Increased inventory turnover&lt;br&gt;Decreased inventory write-downs&lt;br&gt;Consistent available floor space&lt;br&gt;Consistent time to delivery&lt;br&gt;Reduced stock moves to centers</td>
</tr>
<tr>
<td>Network Resource Planning</td>
<td>Ability to scale network up or down based on demand</td>
<td>Accurate demand forecasting&lt;br&gt;Optimal fleet utilization&lt;br&gt;Optimal driver /crew availability&lt;br&gt;Improve safety</td>
</tr>
<tr>
<td>Route Optimization</td>
<td>Most optimal pick-up and delivery</td>
<td>Predictable time to delivery&lt;br&gt;Optimize fuel utilization&lt;br&gt;Account for traffic conditions&lt;br&gt;Reduce mileage /maximize load</td>
</tr>
<tr>
<td>Customer service &amp; loyalty</td>
<td>Service innovation and differentiation from competition leading to repeat business</td>
<td>Improved customer satisfaction&lt;br&gt;Extended network partnerships&lt;br&gt;Service level satisfaction&lt;br&gt;Reduced cost of service</td>
</tr>
<tr>
<td>Promotions and marketing</td>
<td>Optimal spending on promotions to drive profitable business results</td>
<td>Gain new customers&lt;br&gt;Cross-sell / up-sell to existing&lt;br&gt;Measure future sales impact&lt;br&gt;New location based services</td>
</tr>
<tr>
<td>Risk analysis</td>
<td>Ability to function as a reliable part of the customer’s supply chain / meet customer’s needs</td>
<td>Understand potential liability&lt;br&gt;Help customers grow business&lt;br&gt;Provide unique service offerings</td>
</tr>
<tr>
<td>Maintenance</td>
<td>Optimal maintenance scheduling for on-time delivery and cost containment</td>
<td>Minimize breakdowns&lt;br&gt;Minimize potential liability&lt;br&gt;Improve service levels&lt;br&gt;Reduced warranty costs</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 logistics and transportation 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](image)

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](image)
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 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.

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.

![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 operational sources for the airline industry used to understand passengers and incorporating the Oracle Airline Data Model might be represented similar to 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. Logistics management and transportation companies generally rely on a mixture of ERP 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. The following illustrates what is often described as an Internet of Things footprint for connected vehicles in logistics and transportation companies:

![Figure 10: Connected Vehicles in Logistics and Transportation Companies](image)

Items to the far right of Figure 10 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 11 illustrates a typical capability map of this architecture for logistics and transportation:
Sensors are increasingly providing critical weather and traffic data to all sorts of logistics management and transportation companies. They are also used to monitor equipment state as well as the state of cargo in transit. Sensors are also an important part of government mandated safety solutions (such as the emerging solutions for automation of railroads in the United States). This data will continue to grow and enable companies to better assess and manage people, equipment, and goods and services that are being offered.

Figure 12 illustrates some of the Oracle products aligned to the previously shown capability map:
Figure 12: 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 13 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 13: Typical Methodology for Information Architecture Projects](image)

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. Logistics and transportation 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 logistics and transportation 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 14: 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). Aside from security considerations, most logistics and transportation 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 logistics management and transportation 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 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.