Improving Utilities 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 successful operation of Utility companies. While operational efficiency and favorable customer experience through bi-directional communication remain keys to success, “Smart Grids” enable leaps forward for both in a new service world. As Utility companies become larger and more diverse, the type of data that must be managed becomes ever more complex. Data from Smart Meters and Smart Grids, becoming common-place, is augmenting data from enhanced customer management systems. Meanwhile, the Utility companies are looking for new ways to improve operational efficiency as well as reduce cost as they are faced with adding these new and wider sources of information. It is critical to leverage the vast data available to better understand customer needs and requirements (demand side) and to align that to energy generation and distribution (supply side) to maximize operational efficiency. This will also enable organizations to gain competitive advantage uniquely.

Utility companies have long gathered customer usage metrics, first through manual meter readings and later through meter transmissions. More recently, data sources also include:

» Smart Meter Data and Smart Grids
» Outage Management Systems
» Asset tracking for consumers, electric grids, and power generation
» Supervisory Control and Data Acquisition (SCADA History)
» Alternative Energy Sources
» Social Media
» Weather Monitoring Systems
» Wholesale Market 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 enable the Utility to handle these requirements, remain competitive and responsive to customer demands, and make progress in preparing for the “Smart Grid\(^1\) data” influx.

\(^1\)A Smart Grid is a modernized electrical grid that uses analog or digital information and communications technology to gather and act on information, such as information about the behaviors of suppliers and consumers, in an automated fashion to improve the efficiency, reliability, economics, and sustainability of the production and distribution of electricity.
This paper provides an overview for the adoption of Big Data and analytic capabilities as part of a “next-generation” architecture that can meet the needs of Utilities seeking to improve customer service and efficiencies while increasing profitability.

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

Utility companies historically used data warehouses 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 while providing customized and rich user experience and the business can become more agile, effective, and potentially generate higher revenues with lower costs.

Improving Customer Intimacy

Big Data and advanced analytics solutions enable Utilities to leverage data from the Smart Grid to improve customer service by providing better visibility into individual customer usage patterns with more detailed information. That information can drive customer-centric offerings on a one-to-one basis. For example, it is possible to implement proactive demand/response programs that best fit the needs of target customers. New pricing programs can be established to help manage demand consistent with available supply and usage patterns.

Analysis of individual customer usage of Utility supplied products can determine normal usage patterns and these can be compared to fraudulent or unusual usage patterns. Real time analytics can be used to alert customers and the Utility to abnormal usage spikes due to natural or unnatural circumstances. For example, a broken water faucet could be detected faster, saving time and money. The faster data can be analyzed and an actionable event created, the greater the potential for revenue recovery or loss prevention.

Improving Operational Efficiency

Predictive analytics can be used to minimize outages and improve power distribution reliability by anticipating demand and taking appropriate steps before an outage might be forced to occur. These same analytics can be used to identify trends and forecast demand.

For example, consumption of Utility products (gas, electricity, and water) is often strongly connected to weather patterns. The ability to predict an oncoming cold front could be used for pro-active 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 production.

Prediction is especially useful to better match supply to demand for electricity as it cannot really be readily stored for later use. Such analytics enable the Utility to better anticipate power demand and adjust supply, purchase or sell power as needed. For example, Smart Grids are generating data at enormous rates with some studies showing about 1.5 TB per household created on an annual basis. This data can be used to generate time-of-day usage profiles, and be useful in refining pricing models and commodity trading strategies. Better analytics also enable Utility to implement and/or improve efficiency programs as well as improve communications with regulators.
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. Utility IT organizations typically work with their lines of business to build solutions that deliver the following when defining Big Data projects:

1) **Revenue protection or loss prevention**: Used to determine unusual usage patterns and probe as to why a meter signature suddenly appears differently or for meter data consistency with the billing data. Also, leverage usage patterns to formulate demand based pricing based on peak usages on grid for improved revenue and better demand-supply management.

2) **Meter data acquisition and demand response**: While meter acquisition rates are typically set to 4-hour intervals, Hadoop-based architectures can handle much higher data rates. A meter could be sampled every 5 or 15 minutes instead of 4 hours to better analyze demand and manage the grid efficiently. Also, helps to streamline supply (power generation) to meet the demands effectively.

3) **Outage and load analysis**: Maintenance and operational pitfalls can be significantly reduced with predictive capabilities leading to reduced outages, better proactive maintenance of assets, and reduced costs. Operational efficiency can be improved by analyzing data related to power quality and from load balancing, load matching and daily peak demand. Predictive customer modeling and analyzing customer sentiment from sources like social media can provide insight into customer behavior providing another input for load forecasting. Sensors on equipment and vehicles can more quickly pinpoint problems and enable better logistical support.

4) **Improved financial forecasting**: More detailed histories can help determine, measure, and track the elasticity of demand and Electricity Tariff Analysis what-if scenarios can played-out in more detail, while asset optimization and asset planning can be improved by analyzing operational efficiency.

5) **Risk or Threat Management**: Utility companies face an unusually high number and variety of risks, including many that have serious health, safety and environmental implications. While these risks cannot be eradicated, more and varied data can be used to help manage and mitigate some of them. Finding unauthorized or illegal access or usage can save companies from suffering the consequences of these possibly malicious activities. Predictive analytics that can generate prescriptive actions can pay large dividends if major problems become avoidable.

6) **IT operational efficiency**: Not unique to Utility 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, a table summarizes several typical business challenges in Utility companies and illustrates the opportunity for new or enhanced business capability when adding new analytic capabilities.
## TABLE 1 – UTILITY COMPANY FUNCTIONAL AREAS, BUSINESS CHALLENGES & OPPORTUNITIES

<table>
<thead>
<tr>
<th>FUNCTIONAL AREA</th>
<th>BUSINESS CHALLENGE</th>
<th>OPPORTUNITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Billing Exceptions</td>
<td>Better prioritize and manage exceptions</td>
<td>Billing exceptions</td>
</tr>
<tr>
<td></td>
<td>Tracking usage information</td>
<td>High bill prioritization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low bill prioritization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bill cycle monitoring</td>
</tr>
<tr>
<td>Call Center Support</td>
<td>Providing relevant usage information to support customer inquiries</td>
<td>Monitor consumption</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Provide usage comparison</td>
</tr>
<tr>
<td>Meter Operations</td>
<td>Managing problematic meters</td>
<td>Determine defective meter IDs</td>
</tr>
<tr>
<td></td>
<td>Tracking metering issues, and power grid and generation assets</td>
<td>Determine new meter health, and provide preventive maintenance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Provide AMI contract validation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Geospatial data maps for analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Track meter and other assets inventory</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preventive maintenance on all equipment</td>
</tr>
<tr>
<td>Energy Efficiency/Demand Response</td>
<td>Targeting programs to the appropriate customers</td>
<td>Targeted marketing based on customer information and load profiles</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Matching supply and demand for maximum revenue</td>
</tr>
<tr>
<td>Operational / Outage Response</td>
<td>Dispatch right equipment and crews as fast as possible to restore service delivery</td>
<td>Match nearest crew and proper equipment to problem</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improve service and overall customer experience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reduce cost of service</td>
</tr>
<tr>
<td>Safety</td>
<td>Theft resulting in unsafe situations</td>
<td>Determine gas leaks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Detect electrical arc flashes</td>
</tr>
</tbody>
</table>
Establishing an Architectural Pattern

The following Figure illustrates key components in a typical Information Architecture. Data is acquired, organized as appropriate to nature of the data, and then analyzed to make meaningful business decisions. A variety of underlying platforms provide critical roles. Management, security and governance policies are critical aspects across the entire architecture and are always top of mind in the Utilities’ decision makers. 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 a 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.

![Figure 3: Typical Introduction of Information Discovery](image)

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](image)
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.

![Structured Data and the Data Warehouse](image)

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.

![Unstructured Data, Distributed File Systems and Key Value Data Stores](image)

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.

One of the key components in Information Architecture is the data model. Utilities often will want to leverage industry standards and an industry standard Common Information Model (CIM). EPRI has been working to advance a CIM.

CIM is a standard from IEC TC57, and it comes from three working groups within TC57: WG13 for transmission - IEC 61970; WG14 for Distribution - IEC61968; and WG16 for Market Operations - IEC 62325. Working Groups are the groups that focus more on specific areas within the CIM model to help define where and what should be undertaken moving forward and also the progression of the model from a standards perspective. Oracle sees CIM as a value added contributor for data standardization across the enterprise. As such, we are actively involved in the organization in areas such as board participation, standards definitions, sponsoring and hosting CIM events, Working Group participation, and successfully participating in CIM IEC Interoperability Testing.

The Oracle Utilities Data Model (OUDM) is a pre-built, standards-based data warehouse solution designed and optimized for Oracle database and hardware. OUDM is CIM-based and has within its foundation layer many of the attributes necessary for retail utility operations. OUDM also has many features available in Oracle Communications Data Model (OCDM) around the business to customer relationship. OUDM can be used in any applications environment and is easily extensible. OUDM enables utilities to establish a foundation for business and operational analytics across the enterprise, allowing users to leverage a common analytics infrastructure and pre-defined cross-domain relationships, which drive unprecedented levels of intelligence and discovery. With it, utilities can jumpstart
the design and implementation of enterprise information management strategies quickly to achieve a positive return on investment (ROI).

Defining the Information Architecture for your company is all about the linkage to the specific use cases of interest. For example, a use case that includes the analysis of weather data and Smart Meters might drive the need for a footprint something like Figure 9:

Figure 9: Typical Footprint for Analyzing Smart Meter and Weather

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 Utility industry that Oracle can provide. The following figure shows how the Oracle Utilities Data Model might access a variety of data sources and fit a variety of business intelligence needs.

Figure 10: The Oracle Utilities Data Model (OUDM) and its role in the Information Architecture
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 or other electronic machine devices provide the data inputs, the architecture for data capture, security, and the linkage to the rest of the Information Architecture requires additional consideration. The following diagram illustrates what is often described as an “Internet of Things” for smart meters:

![Diagram](image)

Figure 11: Connected Devices / Smart Meters in Utilities

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.

Many Utility companies 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 rules and actions taken. The information gathered using sensors depends on the specific vendors engaged, and so that portion of the architecture may not be out-of-scope for some projects.

When we extend this concept to connected homes and additional monitoring that is possible, we see other sources of data being added that could lead to new business possibilities:
Figure 12: Production Side & Consumer Side Connected Devices and Smart Meters

Figure 13 illustrates a typical capability map for a utility company considering deploying either of the earlier examples:
Sensors are increasingly providing critical weather and transmission data needed by utility companies. They are also used to monitor equipment state and provide logistical information. This data will continue to grow and enable companies to better manage people, equipment, and services that are being offered and provide a rich experience to both customers and corporate partners.

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

Figure 14: Oracle Products aligned to Capability Map
Keys to Success

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

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

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

![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. Utility 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 in Utility 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:

![Typical Conceptual Future State Diagram](image)

Figure 16: 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 utilities 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.

Utility companies are at a pivotal moment in history where more data is available to them than any time in history, and yet much more can be gathered. 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.