

Data management in the utilities sector



An EXL whitepaper

Written by

Vikas Kumar

Assistant Vice President, Decision Analytics

lookdeeper@exlservice.com



Of all industries, utilities expect the biggest returns (73%) on its big data investment (against 39% for insurance and 38% for telecommunication). While 80% of utility player agree that big data provides new opportunities for business, only 20% of utilities have already implemented big data analytics.

It is estimated that more than 680 million smart meters will be installed globally by 2017, leading to 280 petabyte of data a year. This and other significant changes across value chain will make big data and real-time analytics a necessity for the majority of utility players. Interestingly, a study conducted by McKinsey¹ identified utilities as one of the few industries where the relative ease of capturing the valuable potential of data is high.

The commonly used term big data refers not just to the explosive growth in data that almost all organizations are experiencing, but also to the emergence of data technologies allowing data to be captured and leveraged. Big data is a comprehensive term used to describe the ability of any company, in any industry,

to find advantages in the ever increasing large amount of data that flows into those enterprises, as well as the semi-structured and unstructured data that previously was either ignored or too costly to deal with. As the data haystack grows larger, the needle, or useful information, becomes more difficult to find. So big data is really about finding the needles — gathering, sorting and analysing the flood of data to find valuable information on which sound business decisions can be made.

In the retail utilities context, data includes details corresponding to customer profile, consumption pattern, behavioural data and interaction data. Considering exponential



growth in the volume and complexities in each of these data segments, diversity of data sources, variety of uses and related latency requirements, developing an effective data management strategy presents a very significant challenge and opportunity.

Data management framework

A robust data management framework ensures that accountability is fixed, data ownership is assigned and a focused approach is taken to drive data improvement steps. This requires an organisation-wide approach to data strategy, ensuring the right data is captured at the beginning, the right set of tools and technologies are deployed and that data is kept consistent across the systems by

working on the required structure, process and enablers.

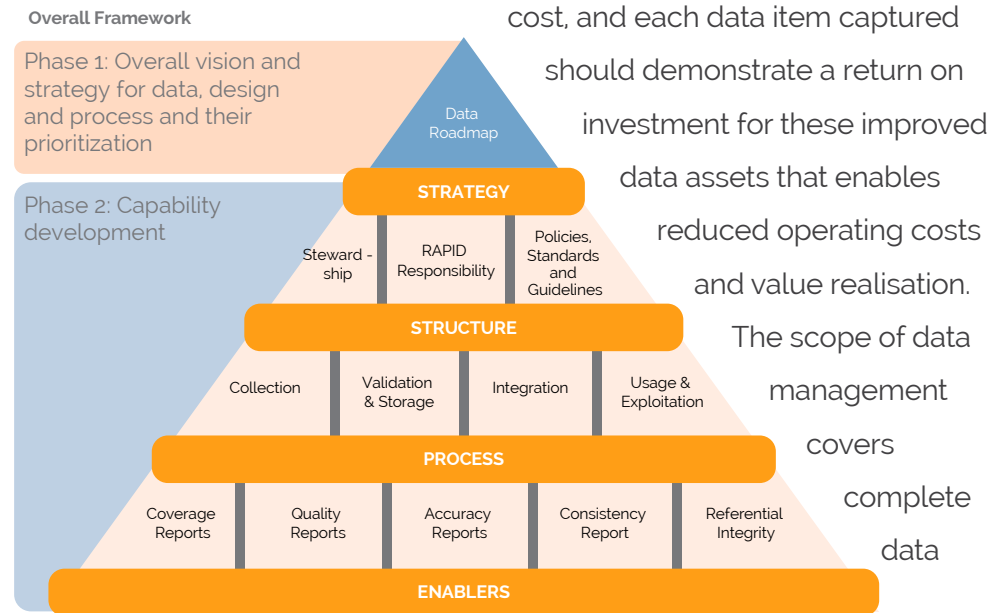
By listing down each block explicitly, the framework enables focus on collecting and managing the most valuable information for the enterprise as a whole.

Its success requires customer-facing and back-end teams working together with a shared responsibility. This results in enhanced customer experience since data health checks are built into customer journeys wherever possible, enabling customers to self-verify pre-populated data to reduce journey

exceptions. It is critical to highlight that capturing data items have an associated cost, and each data item captured should demonstrate a return on investment for these improved data assets that enables reduced operating costs and value realisation.

The scope of data management covers complete data

lifecycle from cradle to the grave and across customers, processes, journeys, systems, organisation, behaviours, locations and other areas.





To realize the full potential of the framework, each component of the pyramid must be thought through and planned in detail before execution. This requires:

1. Strategy: The key objective is to lay out an organisation-wide vision and design journey. This is the most critical element of the overall framework and requires due consideration for balancing short-term and long-term timeframes, aligning with other business units to optimize the cost of iterations and defining the required structure, process and enablers to ensure quality execution.

2. Structure: This includes defined the governance structure, policies and guidelines to ensure harmony across various phases of execution to sustain

business continuity. It involves organizing formal company-wide accountabilities, processes and collaboration to achieve the goal of high-quality, integrated and trusted information through a well-defined RAPID (Recommend, Agree, Perform, Input and Decide) framework. Apart from a data leadership team, data strategy steering groups and data quality boards, due consideration should be given to having dedicated data stewards for each business or operational unit with the responsibility of establishing detailed data definitions within their data area, and develop plans using structure and strategy to deliver developed, robust and sustainable data that drives value and reduces operating costs within their business area.

3. Process: This refers to step-by-step guideline with a checklist to be followed for the execution of any data-related activities. For example, during data collection, all data should be validated against brand data standards, a consistent set of rules defined for information management, including data quality reporting, data capture, data validation and matching. Wherever feasible, this should be based on appropriate validation and adherence to the required security and compliance guidelines. Similarly, quality gate principles should be followed to validate data and ensure consistent validation independent of population methods such as migration, integration, data entry or acquisition.

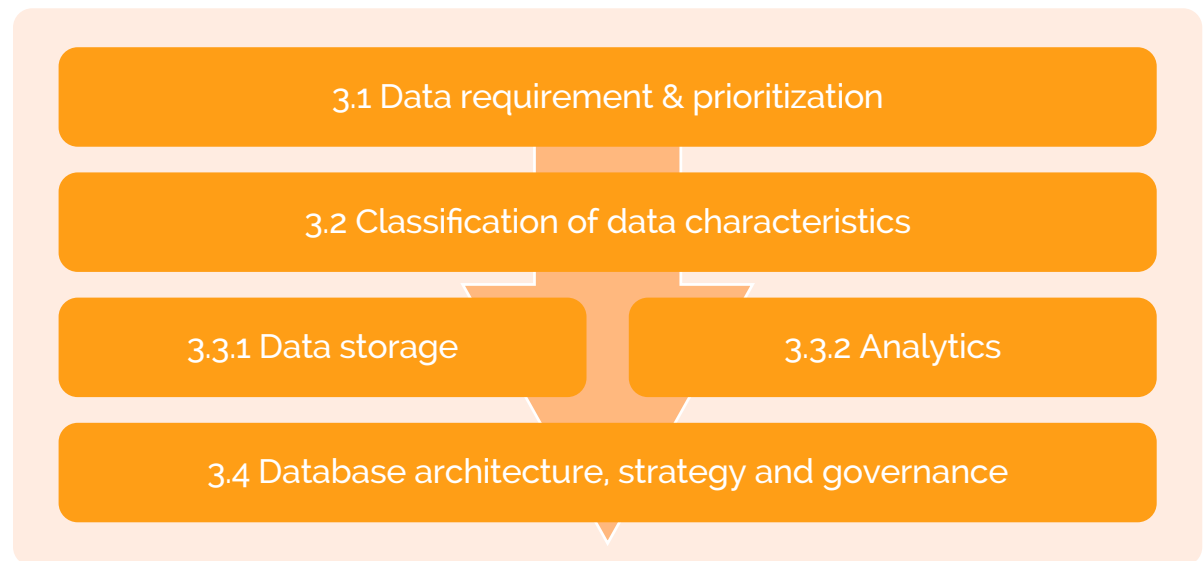


4. **Enablers:** This refers to supporting tools and techniques to ensure coverage, quality, consistency, accuracy and integrity. An example of this could be a standard report which clearly defines which application owns the primary set of each data type, and which applications are secondary. This will dictate how the data should be updated and in what sequence it should be synchronised in to ensure integrity. Similarly, an auto alert generated when spurious data comes in (e.g. phone number with nine digits) to the data owner will help maintain data quality.

Data management strategy

For companies wishing to take advantage of a big data solution and effectively manage data, a data management strategy well aligned with business situation is required. It starts by looking at data requirements and their

prioritization, followed by classification and characterization of data based on their segment, class, latency and storage type. This leads to the development of a solution matrix and performance measurement strategy which eventually enables the development of analytics architecture.





Data requirement and prioritization

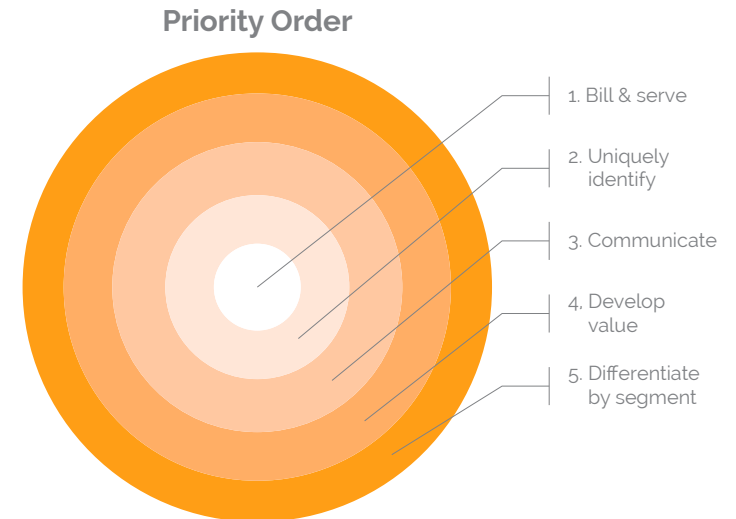
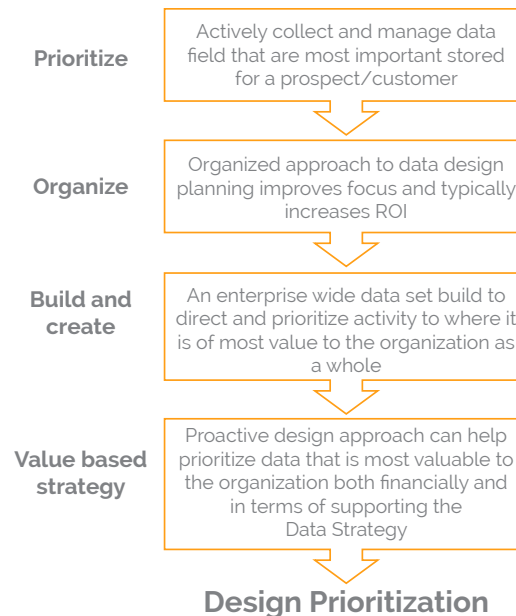
Data requirement refers to conducting detailed due diligence to identify the key fields required to conduct analysis and generate actionable insights. Overall, data requirement can broadly be classified into five buckets:

- > **Profile data:** Includes data required to create segments such as customer segment, business/property segments, appliance holding type and meter type.
- > **Subscription data:** Information related to customer subscription including contract, tariff and product holdings.

- > **Behavioral data:** Data clusters that help understand customer behavior like payment data, invoice data, debt data and digital transactions.
- > **Contact/Interaction data:** Data with regard to customer interactions including

call center data, CSAT/NPS, complaints and visit data.

- > **External data:** The different external data sources that would augment your internal database to derive meaningful insights, for example RPM Direct, Experian, Axiom etc.





It is not realistic or even feasible to actively collect and manage every data field that could be stored on a prospect or a customer. Doing so would dilute focus and reduce the return on investment, or at least the measurable return. Data prioritization refers to rank ordering requirement to focus initial efforts on the data that is most valuable to an organisation, both financially and in supporting the data strategy, and then move on to the next most critical details. Depending on available resource bandwidth, an organisation might want to cover some or all of these categories. However, beginning with prioritized data can help ensure that the most pertinent data is covered first.

Classification of data characteristics

A majority of utility players need to differentiate between several categories of data found across customer life cycle and the time-related aspects related to data types. It is important to distinguish different types of data categories, including but not limited to customer experience, complaint management, subscription, renewal, journey management, customer switch, consumption pattern, payment history and contract tariff. To manage data effectively, it is essential to understand the differences of each data class, their potential applications and their respective latency considerations.

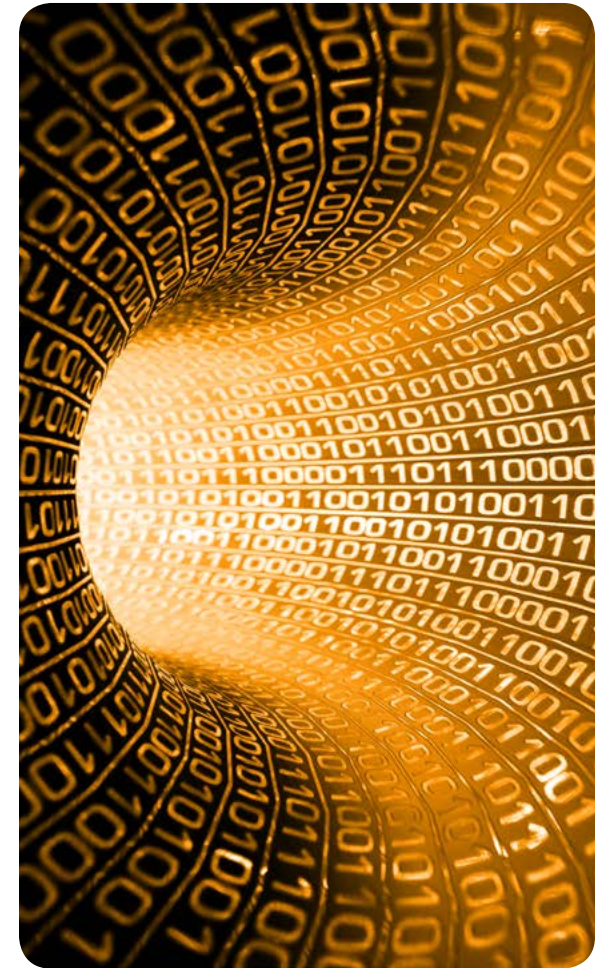
> **Data segment:** A data segment corresponds to various stages of the customer lifecycle where data is captured, including shop, join, use and help, change and leave. Such

categorization of data helps define stakeholders for data management and provides easy access to everyone for a more detailed deep-dive analysis. Ideally, the more refined the categorisation, the more insightful the analysis. For example, within the shop and join segments, having the flexibility to look into various aspects in a more details level corresponding to various levels of drop-outs, application submission, payment and installation might provide many more actionable insights instead of consolidated shop and join data.



> **Data class:** Data arising from smart meter devices needs a completely different form of storage and treatment compared to data for customer registration, regular use and pay cycles. It is critical for utility players to understand the differential treatment required and define the business value of each data class, perhaps to the point of subdividing the classes as appropriate for specific drivers and constraints, so that proper data management solutions may be derived.

> **Data latency:** Latency is defined as both the time interval between when data is requested by the system when it is provided by a source, and/or the time that elapses between an event and the response to it. Latency considerations must be included in the design of a data management platform. Otherwise, significant, and potentially fatal architectural issues will arise. Going forward, latency period can be customized to historic feed, identifying higher frequencies for cases where significant variations are observed from normal consumption pattern. This can help get granular details where required for further deep-dive analysis or proactive interventions.





> **Data lifespan:** Depending on how the data is to be used, there are various classes of storage that may have to be applied. For example, while a utility would like to store customer specific details forever, details corresponding to the exact consumption of a household within 30 minutes intervals might not be of interest beyond a few years. An ideal solution would be hierarchical data storage architecture, with different types of storage applied to different data sources, coupled with latency requirements. The table below captures some lifespan classes relevant to utilities.

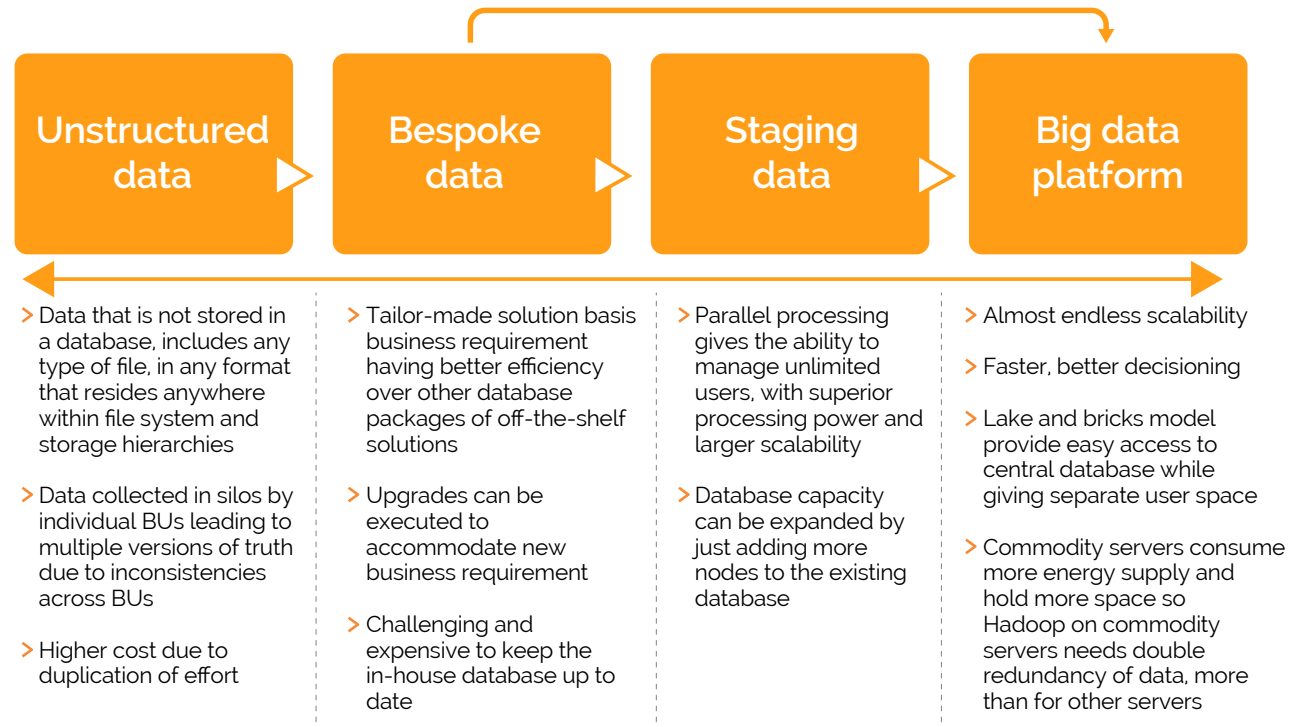
Data Lifespan	Key Characteristics
Transit	Data that exists for only the time necessary to travel from source to sink and be used, persisting only momentarily in the network and data sink before being discarded. Examples can include an event message used by protection relays and sensor data used in closed loop controls. In these cases, the persistence time may be microseconds.
Burst/Flow	Data that is produced in bursts or is processed in bursts may exist temporarily in FIFO queues or circular buffers until it is consumed or overwritten. Examples can include non-logged asynchronous event messages. Often, storage for this data is incorporated directly into applications, such as, complex event processing engine event buffers.
Operational	Data that may be used from moment to moment, but is continually updated with refreshed values so that old values are overwritten since only fresh values are needed. This can include meter state data such that may be updated every few seconds.
Transactional	Data that exists for an extended but not indefinite time typically used in transaction processing and business intelligence applications. Storage may be in databases incorporated into applications or in data warehouses, datamarts or business data repositories.
Archival	Data that must be saved for very long (even indefinite) time periods including meter usage data, customer details or log files. Note that some data may be retained in multiple copies depending on the requirement.



Data storage

As depicted in the figure below, utility players have already started moving from high cost unstructured data captured in silos to well-organized big data platform with significantly higher scalability, speed and other capabilities.

To manage big data, it is essential to apply technology solutions appropriate to the data class and intended business processes. In most cases, it is not a question of one technical approach versus another, but of what is the best combination to meet the specific business need. Many types of storage and database technologies are useful in a smart utility context.





[Data management in the utilities sector]

As an example of why it is important to understand relationships among data classes, persistence models and data store types, consider the present interest in the Hadoop "big data" storage model. As such, the Hadoop model can be very good for enterprise-level business data repositories. However, for operational data, it has several drawbacks, such as a centralized data store model that cannot satisfy the needs of low latency multi-objective/multi-controller (MO/MC) systems where analytics must often be consumed close to the point of data generation. Similarly, the Hadoop Distributed File System (HDFS) coherency model does not work for dynamic operational state information and burst event message data flows that are huge components of the big data challenges of smart physical systems.

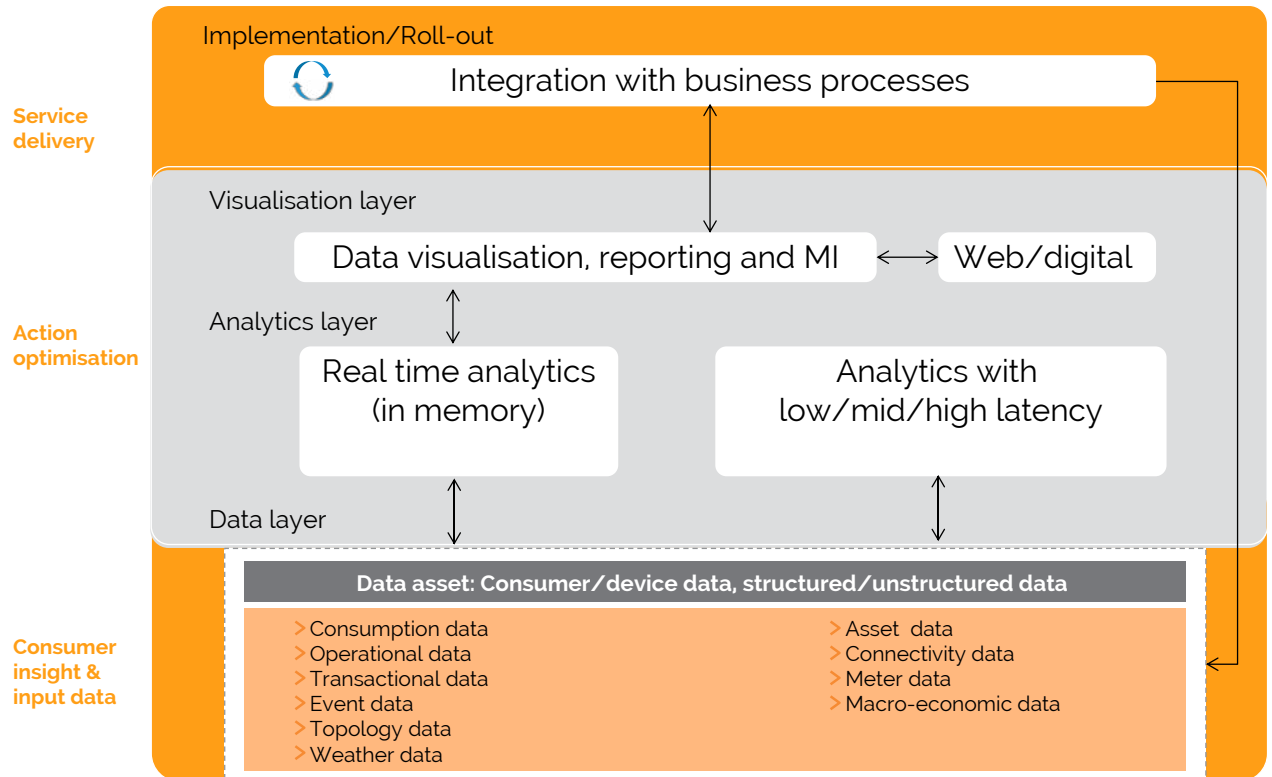
Societal change calling for more sustainable and economical energy is driving utility industries through a fundamental shift where distributed and renewable energy generation, energy storage by customers and electric vehicles are becoming popular. To manage the associated data volume and complexities, traditional methods need a paradigm shift towards an approach where customers have higher visibility, control and participation. Cloud computing provides a highly automated, dynamic and cost-effective solution to this challenge because it offers massive scalability and collaboration capabilities. Furthermore, it can be used to deploy new services with greater speed without significant capital investment.

Through cloud computing, utilities can shift from on-premise, slow and sub-optimal data management methods to online, optimized virtual data centres. Cloud computing's pay-per-use model helps organisations avoid capital expenditure with flexibility to instantly scale up or down. For utilities, a hybrid model combining best of private cloud (owned by the organisation on a private network and highly secured) and public cloud (owned by cloud service provider and highest level of efficiency) would work optimally.



Analytics

The objective for analytics is precisely automating high-volume decisions on a unified platform in a consistent, scalable, fast and economical manner that allows a high degree of adaptability. Applying the right solution to a particular business process can be challenging, given the variety of data characteristics and related business processes across customer journeys. For example, a complaint management system might require a completely different set of tools to facilitate text analytics against those analysing customer journeys on a utility website. Mapping business needs against the capability of various analytical tools can help identify the optimal application. Further, consistent use of common applications across business units not





only reduces cost, but also enhances efficiency through synergies by saving time spent on non-core data transformation and migration activities and avoiding overlapping activities. It also provides the flexibility to create an in-house repository of common tools and techniques which can plug-and-play across business units. The figure below shows different layers corresponding to analytics and their inter-linkages.

In the future, the use of real-time, in-memory analytics tools like SAP HANA, Apache, Actian, Spark, Kognitio, Oracle Times Ten and others will quickly replace legacy static data analysis, to decreasing turn-around-time between insights and actions. This will address a regions' peak-hour consumption breaching a sustainable

threshold level proactively. At the same time, traditional modelling tools and techniques will also need to evolve with more sophisticated models, scenario building capacity and ability to manage a much higher volume and variety of data.

The use of real-time data feeds can help utilities proactively identify potential customer pain-points and address them. For examples, a sudden abnormal spike in consumption patterns recorded through smart meters can be used as an indicator of potential break-down at household level, and utilities can ask their service team to proactively reach out to such customers. In this case, the real-time use of such data eliminates the possibility of accumulating it into a large store to be processed offline, enabling on-going operator decision support.

In order to work more efficiently with real-time data, utilities should adopt a two-stage data management and analytics architecture. Complex Event Processing (CEP) is a technology that has found wide use in industries as varied as financial systems, homeland security, and sensor data processing. In each of these cases, the common element is that data from various devices must be processed on the fly,, whether coming in streams or asynchronous bursts. The CEP technology is capable of applying complex queries to multiple data stream processing simultaneously to detect specified conditions, thus triggering the appropriate actions in real time. For example, a sudden



jump in posts related to pricing on a social network platform for a specific region can be used as a signal of competitor price change.

CEP can support a range of relevant utility business functions. These include meter data management, fault detection, outage management and remote device/system monitoring. CEP is a flexible tool that, when included in an overall data management strategy and architecture, can significantly augment the flexibility needed to implement modern utility technical solutions.

Database architecture, strategy and governance

The utilities information universe is fast evolving from relatively high-latency batch processing to low-latency real time analytics. This places a new emphasis on the need for updated data management architecture that move from batch to even-driven real time operations.

Further, it is also important to consider the inter-relationship between business processes, business applications, data management and overall infrastructure to have a positive effect on customer experience. To facilitate the development of good analytics architecture, it will be helpful to develop classifications for the structure of data and analytics classes. By combining this taxonomy with business strategy and requirements for protection

and control along with other applications, utilities can develop a full database architecture and analytics strategy.

Data governance is a set of processes which ensure that important data assets are formally managed throughout the enterprise. It is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models describing who can take what actions with what information, when, under what circumstances and using what methods. Data governance ensures that data can be trusted and people are made accountable. Data stewards are the key links in data governance structure practiced.





[Data management in the utilities sector]

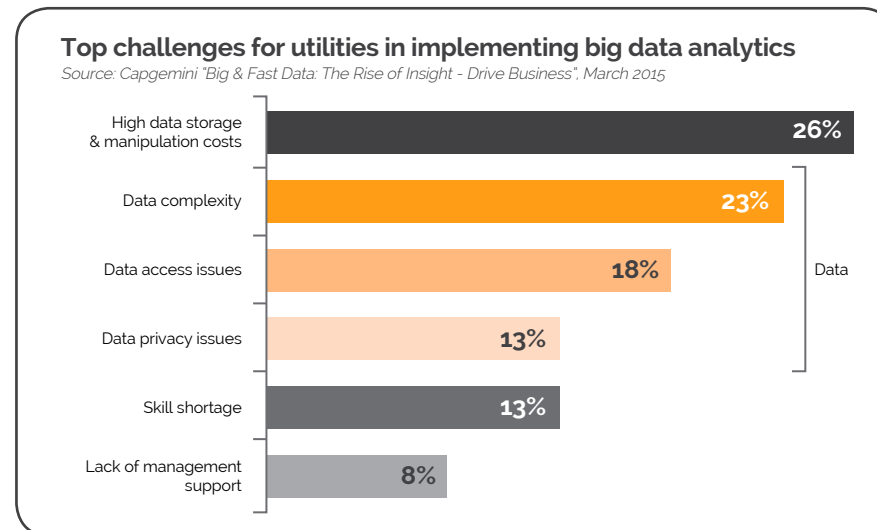
Data stewards are typically members of a data quality board who review progress and plans to improve data quality. Their key responsibilities include:

- › Establishing detailed data definitions within their data area. For example, the energy data steward will be responsible for energy operations data domain and its associated data.
- › Developing plans to deliver a developed, robust and sustainable data domain that drives value and reduces operating costs

Conclusion

The implementation of a sophisticated database management process can yield significant operational and financial benefit for utilities. However, 41% of analytics initiatives undertaken by utilities are still

limited to basic analytics with reporting functionality. The top challenges faced in implementation are summarized in the graph below.



However, there is a light at the end of the tunnel. The market research firm GTM Research predicts that the global utility company expenditure on data and analytics will grow to USD \$3.8B in 2020, with gas, electricity and water suppliers

in all regions increasing their investments. This investment is expected to lead a reinvention of utilities business.

Utility players can best leverage these reinventions by capturing the maximum benefit through a well-defined strategy that's suited for their data management and analytics requirements. This framework would take into account the prioritization of data, its classification,

required storage, analysis and architecture. Doing so would help utilities realize significant cost savings while developing a state of art data management process.



EXL (NASDAQ: EXLS) is a leading operations management and analytics company that helps businesses enhance growth and profitability in the face of relentless competition and continuous disruption. Using our proprietary, award-winning Business EXLerator Framework®, which integrates analytics, automation, benchmarking, BPO, consulting, industry best practices and technology platforms, we look deeper to help companies improve global operations, enhance data-driven insights, increase customer satisfaction, and manage risk and compliance. EXL serves the insurance, healthcare, banking and financial services, utilities, travel, transportation and logistics industries. Headquartered in New York, EXL has more than 25,000 professionals in locations throughout the United States, Europe, Asia, Latin America, Australia and South Africa.

© 2016 ExlService Holdings, Inc. All Rights Reserved.

For more information, see www.exlservice.com/legal-disclaimer

Email us: lookdeeper@exlservice.com

On the web: EXLservice.com



GLOBAL HEADQUARTERS

280 Park Avenue, 38th Floor, New York, NY 10017

T: +1.212.277.7100 • F: +1.212.277.7111

United States • United Kingdom • Czech Republic • Romania • Bulgaria • India • Philippines • Colombia • South Africa