A SIS-G White Paper
April 2009

Critical Success Factors to Achieve Operational Efficiency in Information Management for the Electric and Gas Distribution Value Chain

Strategic Information Systems Group

Improving Business Operations, Intelligence and Decision Profitability by Integrating and Aggregating Disparate Data

This white paper is intended to help Business and IT groups working within the Energy Utility Value Chain to improve Operational Efficiency by streamlining the delivery of high quality and low cost information. If you feel your business and Information Technology costs related to data management are too high relative to benefits, read on. Mission critical processes targeted in this analysis include Market Settlement, Regulatory Reporting, Advanced Metering Infrastructure, Deregulation and External Data Exchanges, but mastery of the following six critical success factors can be applied to many other subject areas dependent on the affordable and accurate delivery of information:

Enterprise Data Governance
Corporate Information Factory
Corporate Metadata Strategy
Data Integrity
Data Technology
Data Delivery

Data integration and data delivery processes, including Data Exchange and Data Warehousing, can be complex and expensive. They require specialized knowledge within Information Technology (IT) management to provide the appropriate strategic vision to accomplish objectives in a cost effective fashion. This paper is based on many years of experience in the electric and gas utility sector. It is intended to impart the knowledge gained to optimize data integration and data delivery processes within the firm, as well as expedite data exchanges with external regulators, suppliers, customers, partners and other third parties.

Enterprise Data Governance

Data Governance is the activity of managing data for the entire corporation with the intention of protecting, preserving, and improving its quality in order to facilitate business processes. By centralizing overall data strategy and data operations with a focus on increasing value and minimizing total cost of ownership of data assets the entire company will benefit. The goal of Data Governance is to coordinate and consolidate efforts that enhance data value throughout all data repositories in the company. A group of specialists from diverse organizational areas is necessary to identify data problems and devise avenues to solve these problems in an efficient way for the entire corporation. With top management's support, the Data Governance team is able to define and implement enterprise wide processes that correct the data problems in the right places to benefit many, replacing individual department data silos.

Data Governance begins with the actions of a Data Steering Committee. They sponsor the creation of a Data Roadmap to plan an iterative series of business process and system enhancements that successively build an increasingly strong data infrastructure, upon which future data delivery initiatives can build. The steering committee must critically evaluate the cost versus benefit relationship by understanding IT terminology and the true value of alternative project proposals, tool acquisition requests and high-level data architecture designs. The committee must be able to select strong project leaders who can accomplish enterprise wide data integration and business intelligence objectives within budget.

Data Governance also requires that the steering committee exhibit the wisdom to know when to faithfully follow the Roadmap, or when circumstances justify deviations from it in order to adjust to changing business or regulatory landscapes. The steering committee acquires this wisdom by partially staffing itself with a carefully selected blend of business and information technology members, trained in data integration and data warehouse disciplines. General purpose project management skills are not enough since the resolution of the complex technical, utility industry and regulatory issues must be addressed from a data asset investment perspective, trading off short term demands and risk with long term total cost of ownership as we will see later in this paper.

Corporate Information Factory

Present day electric and gas utilities are deluged by a plethora of raw data streaming from their Advanced Metering Infrastructure (AMI), Enterprise Resource Planning (ERP), Market Settlement systems, etc. Regulations require the flexibility to make considerable changes to operations, like Rate Restructuring and Customer Unbundling for Deregulation. A Corporate Information Factory (CIF) Architecture is an enterprise wide design whose purpose is to define the structures and paths of data, provided by business operations, which can be transformed into timely and appropriate information delivered to a variety of coordinated business activities, such as hour-by-hour decision making from analytics embedded in ERPs.

The CIF has proven to be a stable and enduring technical architecture for any size enterprise desiring to build strategic and tactical Decision Support Systems (DSS) or a data integration hub. By providing a well, defined purpose to each of its data layers and each layer building progressively on top of each other through focused transformations, the Corporate Information Factory provides flexibility through beneficial data redundancy. Its layered data structures are used to store progressively refined versions of the data so changes can be made in a single layer, saving overall effort, especially when new business processes create data integration issues in downstream systems. These layered data structures are a Staging Area for source system extracts, Operational Data Store, Data Warehouse and Data Marts. This architecture is commonly implemented in mature data warehouses and found in the designs of many packaged data warehouses (SAP BW, PeopleSoft EPM, Informatica Applications, etc.). It enables the integration and transformation of data (like those described in the Data Integrity section below) from diverse operational systems into a consolidated representation of the business, to be shared for many purposes, including internal integration and external data exchange.

The management challenge is to implement a long-term strategy, so individual data integration initiatives account for the additional cost and risk of integrating their new functionality into an already existing CIF system. The Data Governance Team must socialize this strategy across IT and Business groups so that stakeholders understand the benefits of this iterative, longer-term approach for cumulatively increasing the asset value of corporate data. A key to success is the assembly of a highly skilled team of data integration specialists, trained and equipped with the right technology to enable accurate impact analysis, efficient development and comprehensive automated regression testing.

A data model accurately identifying the analytic “Sweet Spots” (i.e. metered usage, billing charges) reduces the number and size of data structures needed, saving all around effort. Data Warehouse structures should be built at the lowest grain to maximize flexibility in creating and re-creating marts over time. If the Data Warehouse structures are atomic and have all the dimensional breadth of the source systems, marts can be re-constructed more easily with a dimensional mix that is more suitable for the dynamically changing queries needed to drive business decisions. For example, if detailed analysis of usage at a half-hour grain without respect for daylight savings time leads to questions about hourly trends, we may decide to re-build the mart from the warehouse detail at the hourly grain, adjusting for daylight savings time.

Intermediate details for highly transformed aggregates should always be saved for the inevitable need to drilldown. For example, if half hourly meter readings are transformed into billing usage by bill cycle, the transformed billing usage should be saved in the data warehouse, at the half hour grain, prior to aggregation by bill cycle for monthly billing. If Rate Case Reporting aggregates usage in incremental blocks, the detail by usage block should be saved to provide navigation between the different report levels. The data model should provide maximum flexibility, so that sweeping billing rate changes or other large-scale regulatory changes may be absorbed without significant impact to data loads, operational system integration or reports.

Corporate data architecture decisions have the greatest potential for cost optimization, or conversely, cost black holes. One of the most common and costly data delivery mistakes is to authorize an independent data mart (Silo) architecture for each major organizational domain (Finance, Rates, Field Services, Energy Purchasing, Retail Supplier Services, Marketing, etc.) in order to avoid sharing data that might impact pre-existing data warehouse applications “owned” by those departments. For example, a new project manager or data architect may decide that building new structures will save the effort of learning the existing system and risk of unwanted impacts to existing subscribers while retrofitting new integration. This approach generates costly redundancy in source system extracts, data storage, CPU processing, software license charges based on CPU processing, reporting solutions and IT maintenance. Other negative consequences for a silo strategy are:

- Data Mart Silos report different values for the same metrics, due to differences in their extract, transformation or load logic.
- Stand-alone data marts rarely have every data element needed to be able to satisfy a subsequent data delivery initiative.
- Enhancements to one data mart cannot provide benefit for the others.
- Independent marts are extremely costly to retire because they meet their funding user base's narrow needs and gives them a sense of sole proprietorship when it comes to sharing or consolidating their data. Significant IT maintenance costs are incurred indefinitely, while their value dwindles, until their ultimate, costly sunset.

It takes strong Data Governance embodied with enterprise vision, long-term commitment and skilled data integration specialists to champion a Corporate Information Factory architecture over data mart silos, but it does not take long (3-5 years) for the redundant mart costs to exceed those of a consolidated solution. Management's understanding of and confidence in the CIF architecture will enable the Data Governance Team to differentiate between good (CIF layers) and bad (silo marts) data redundancy.

Metadata Strategy - Corporate Wide Data Definitions

Metadata is defined as "Data about Data". It is comprised of the properties, attributes and characteristics that define data elements individually and collectively. Metadata describes enterprise terms like Customer, Usage or Tariff, as well as metrics like Billed Usage or Interval Consumption. When exchanging data between internal systems and external entities, clear metadata definitions help all parties have a common understanding of what each data element means. Consensus and shared knowledge of what the data means in the source systems and throughout the life cycle of the data in the downstream systems will lead to more accurate source data, more efficient development efforts and lower overall cost to deliver and maintain applications. Precise definition of metadata will enable the business to articulate what it needs and enable IT to understand it faster, facilitating more cost effective, targeted and higher quality data manipulation as it makes its way through the value chain.

Like many other industries, energy utilities operate with a plethora of data elements; many that are ambiguously named, similarly named, difficult to describe, or referred to with multiple synonyms. Aliases in data elements come from the past history of the firm, industry standards and competitive software vendors striving for differentiation. Long time users of operational systems, like billing, do not see the need for Metadata because they are so used to working with the data elements the way they exist. They do not value the perspective of downstream system users or external entities that must use their data. The result is costly disconnects and mass confusion about what is specifically meant by certain terms. For example, the term "usage" is frequently referred to as if it had a single meaning, when in fact there are many flavors of usage as it is collected by the Advanced Metering Infrastructure (AMI), refined by the billing system and selectively filtered, factored and aggregated as it is delivered to its various destinations supporting retail suppliers, accounting, energy suppliers, settlement, rate cases, regulatory reporting or internal operation analytics.

The reality is that the business rules and specific requirements of each usage subscriber may be slightly different. Requirements gathering must define a single system of record for each data element and its owner, definition, derivation logic, calculation and uses. It must ferrate out subtle differences like which loss factors or reading multipliers should be applied, or which meter types should be filtered out, netted or passed through. The regulators of different states may define these terms or metrics differently as well. It is difficult to find the very few skilled people who can articulate the business requirement, interpret it and translate it to appropriate and meaningful IT terminology, transformation specifications and test scenarios without defined common metadata definitions. Leveraging scarce human resources by capturing and storing their knowledge in metadata provides great benefit for the company, if done properly. A set of common metadata definitions will expedite the design, development and testing of all applications requiring data for their processes and reduce disconnects that are often discovered too late.

Metadata capture infrastructure must be treated as corporate assets. In reality, it is usually sacrificed from project scope early in the project. Ideally, incremental metadata definition required by each project should be estimated and included in each project. If this best practice is not attainable, metadata definition should be funded as a separate project. The reality is that utilities have been paying for Metadata to be collected repeatedly in requirement documents, design spreadsheets and email all over the company, but the fruits of these efforts have not been centrally captured, evolved over time and made widely available for leverage by others who need it. This is, therefore, not an issue about doing more work. It is an issue about the organization, methodology and tools we use to do this work. At a minimum, Metadata needs to reside in a simple central repository where it complies with standards, is reviewed and approved, can be made available to data analysts and shared with (exported to) data modeling, Extract, Transformation and Load (ETL) and reporting tools (where it is available to users). It should be well written, understandable and accurate, but should not be verbose or become a means to its own end. Once paybacks are realized from this minimal collection of metadata, it could be promoted through implementation of a packaged Metadata Management system, where the incremental value added will exceed the incremental cost of Metadata package implementation.

Data Integrity

Data integrity is the conformance of data across various systems to business rules and metadata definitions as defined by the company. It assures that information published from the underlying data is correct and ensures consistency when published to various subscribers in different ways.

Many energy utilities have antiquated source data systems like metering, billing and outage management, that may lack on-line entry or batch data edits for compliance with business rules to ensure data conforms to Metadata standards. Integrated systems may be hindered by data integration anomalies, like missing code values and descriptions, invalid characters, or timing disconnects, which cause the core systems to be attributed differently when storing customers, usage, demand, billing charges, etc. In lieu of replacing these old systems with new systems (i.e. Enterprise Resource Planning systems, etc.),special attention must be paid to the processes required to identify and correct the data integrity issues so business users can correctly publish information.

Poor data quality adds cost to data integration and data delivery efforts at each stage of transformation. It is a major cause of data integration rework, when uncovering a data problem after the requirements phase. Unfortunately, and at great expense, data integration routines may have to assume the data cleansing role. The costs of poor data quality are incurred in many direct and indirect ways - by the users of the operational system, the IT data integration efforts, testing, and the information subscribers who receive information that does not conform to data quality standards. If these subscribers make bad decisions based on the poor quality data, the cost goes up further. Data Analysts should profile data in the requirements phase of any data integration project, so data quality issues can be addressed in project scoping and estimating. Ideally, operational system data should be cleansed with on-line programs that validate the operational data being entered. For example, invalid accounting code values can be blocked at their source via a data edit module called upon where accounting code values are entered. The module compares the entered value with the valid values in the general ledger, and issues an on-line error message if the values are incorrect. If this on-line validation cannot be implemented on the source data, the data integration architecture must cleanse and conform the data to business rules, as early in the process as possible, to limit the number of downstream applications exposed to it. For example, cleansing should occur in the Operational Data Store, so the Data Warehouse, Data Marts or multiple subscribing applications do not incur the ill effects of invalid data or redundant cleansing cost.

In order to avoid discrepancies in end results, the data flow architecture should ensure data integrity by consistently applying business rules and limiting flows of data through a single path. For example, usage may be interfaced to the same target application (retail suppliers, settlement) more than once, possibly with different aggregations or frequencies. In this case, the data flows of each interface should pass through the same data structures as it flows to the target system, avoiding multiple data paths through different intermediate data stores whose different transformation logic could create inconsistencies.

Pre-implementation testing of new applications is usually not comprehensive enough to be relied on alone to ensure consistent data integrity in production. Automated audit processes derived from the most meaningful test cases should be run periodically in production in order to identify data quality breaches. These audits will catch problems caused by system failures, inappropriate system recovery/restarts, manual data cleansing errors or other infrequent system anomalies. The audit application prevents the spread of invalid data when it appropriately shuts down processes and generates alerts to notify support staff.

Data Technology

As in most human progress, our skills are enhanced and our accomplishments are greater if we are equipped with the appropriate tools. The data integration space is loaded with many useful tools to automate our work in requirements gathering, data profiling and manipulation, real-time integration, metadata, testing activities, document storage, reporting and other time consuming tasks. Below are descriptions of how examples of these tools can save time and improve data integration.

Extract Transformation and Load (ETL) software tools have reduced the programming phase of data warehouse and data integration projects to less than 15%. They include connectors to almost every likely type of data store and pre-built transformation objects for the most common data integration transformations. In high volume environments they can be run on multiple dedicated grid servers that share processing capacity.

With ETL costs reduced, testing of data integration and data warehousing applications has become the largest cost area, especially when the development methodology calls for iterative development imposing frequent major impacts on the existing production application. The larger an enterprise data warehouse becomes, the more serious the potential impact from new functionality. The solution is to automate the testing in each release, adding and adjusting the collective test cases, so that a large inventory of working test scripts is progressively accumulated over time. Once written, executed and passing, the test scripts should be automatically run as a batch each time code is migrated to testing environments. Each release adds new test conditions to the overall automated testing system, traceable to approved requirements. The set of scripts accumulated over the life of the application can be executed as regression tests without much effort before every code release, ensuring that the new release works per the specifications in each testing environment, and does not negatively affect the previous release functionality, preserving the high quality of the growing application in the production environment.

Mainframe Cobol or other 3GL coding language data extract programs are costly to specify, code, test and implement. They usually are fragile, inflexible and hard to maintain for changes that impact them. Nightly extracts consume expensive mainframe processing resources and shrinking batch window time. If the database is old, it may lack timestamps (create, update) which enable change capture, requiring full, high volume extracts each night that consume significantly more processing resources. Extracts using automated change capture tools that read database logs, provide extracts with reliable automated change capture without reading the database tables directly, saving significant cost in coding, testing, maintenance and processing cycles.

Complex data cleansing transformation should be performed with the assistance of specialized data cleansing software. For example, this is necessary for standardizing billing system names and addresses or for matching them to those in externally purchased demographic lists.

An investment in a large-scale performance environment and performance monitors may pay for itself in resources saved to tune application performance in production. Every significant release for a large Data Warehouse must be performance tested and tuned, to reduce production server resource use and decrease data latency.

A new information technology, Service Oriented Architecture (SOA) may be installed as an enterprise information layer above the data warehouse, meter data management system and ERP to provide an information hub to function initially as universal source for the Data Delivery features discussed below. It consolidates and synchronizes data for all data subscribers. In more advanced implementations it can also syndicate updates between operational systems.

These productivity tools may be expensive and regulated energy utilities have limited budgets, high return hurdles for tool justification, and limited IT support staffs. Also, these technologies require robust standards, methodologies, best practices and Centers of Excellence to realize their promise. If the IT staff fails to master new software technology, the benefits of the technology investment are not realized, or worse, the tool becomes shelf ware. Higher return hurdles will then be required for future justifications and potential cost savings will not be realized. This risk emphasizes the importance of honest cost-benefit analysis and evaluation of the staff's ability to successfully implement and maintain new technology, as components of the tool justification process.

The quicker and better a project team can justify and master the implementation of these tools, the more project cost reductions can be realized in a downward cost trend, increasing return on project investment and freeing funds for re-investment. The opportunities to drive down project cost are only limited by the project team's ability to justify and master their implementation. Therefore, IT organizations must emphasize training and encourage tool mastery to drive costs out of data integration and data delivery processes, which historically had been too expensive. Identification and support of technology champions is critical to maintaining the downward trend of cost reductions.

Data Delivery

Data delivery is the timely presentation of appropriate data to subscribing systems and users. It is the culmination of the information management process discussed in this paper and is not effective without successful implementation of the critical success factors above.

An electric and gas distribution utility should focus its investment in data delivery on usage metering and customer billing. Its regulated nature requires it to be very concerned with hundreds of fundamental operating metrics to manage revenue, account for energy, control costs and improve service quality. Some of these are: number of customers billed, average call wait time, energy consumption, net-metered usage, active meters with zero usage, on-peak demand, billed usage and billed charges. Performance metrics are presented via standardized report dashboards, ad-hoc queries, drill down analysis and automated user alerts. The dashboards should be broken down by organizational area, such as billing, field meter services, outage management and collections.

Data is also delivered externally to customers, suppliers and business partners. This is typically through EDI or web sites. Deregulation may require these interfaces to upgrade in order to handle the increased volumes of data being requested by customers, including residential customers, wanting to manage their demand and energy suppliers bidding for residential retail customer business. The use of web based query tools, Service Oriented Architecture and database performance tuning is needed to eliminate the need for inefficient tools like screen scrapers, BOTS and other repetitive data requests. The residential market requires special web page filtering and logic to respect privacy and implement Opt-Out preferences.

The complex business of an electric or gas distribution utility also requires data delivery to Analytic Applications that provide statistical analysis like fraud detection and outage management, extrapolation for rate cases, load forecasting, load back-casting and load scheduling, and predictive modeling for capacity planning. Some of this analytic capability needs to be embedded in operational systems for highly productive real time integrated access by users.

Conclusion

Electric and Gas Utilities must rapidly improve their ability to manage the coming data tsunami from Advanced Metering Infrastructure, Enterprise Resource Planning and other new technologies being implemented. They must enable energy conservation, utility cost reductions, demand management, demand response, deregulation, residential unbundling and capacity planning, simultaneous with increased frequency and difficulty of regulatory changes, including Sarbanes-Oxley compliance. With the new federal administration funding “Smart Grid” infrastructure and “Green” energy technologies to accelerate change, top management must take action to become more agile by implementing data management systems that create and sustain a downward trend in project costs and an upward trend in value derived from data assets and data integration projects. To accomplish this, utilities must adopt an enterprise approach to data governance, data architecture, data definition, data integrity, and data technologies in order to comply with regulatory requests. To manage the total cost of data ownership, they must transform data integration and delivery processes from a departmental and individual project perspective to an enterprise wide orientation.

IT must assure consistent success in technology adaptation to deliver the downward trend of project costs discussed above. While there are no guarantees in IT implementation today, engagement of proven, specialized consulting resources, focused on new technology installation, process implementation, training and adaptation by the in-house IT department is the least cost, least risky approach.

David Getty
Strategic Information Systems Group, LTD
701 Carlyle, Northbrook, IL 60062
888-372-1815
www.sisg.com
dgetty@sisg.com