

Category Winner: Service Oriented Architecture
Company: RLPTechnologies
Submitters: Joe LaFeir, VP Product Development
Vertical: Automotive, Technology
Location: Farmington Hills, Michigan

1. Company description / size (basic company info).

R. L. Polk & Co. is the premier provider of automotive information and marketing solutions to the automotive world and its related industries—automotive and commercial vehicle manufacturers and dealers, automotive aftermarket companies, insurance companies, finance companies, media companies, advertising agencies, consulting organizations, government agencies and market research firms. A privately held global firm, Polk is based in Southfield, Michigan with over 1,300 employees located at operations in Australia, Canada, China, France, Germany, Japan, the Netherlands, Spain, the United Kingdom and the United States.

RLPTechnologies, Inc. is a wholly-owned research and development subsidiary dedicated to deploying world class data-driven technology products that support customers' needs to turn vast amounts of data into business value with accuracy, speed and security.

RLPTechnologies specializes in building industry-leading data solutions that serve as the foundation for focused, in-depth research, analysis and action across multiple industries, enabling the effectiveness of business intelligence tools and applications that “mine” intelligence from the data. Our company vision is nothing short of revolutionizing the way data is collected, standardized, enhanced and compiled into a Single Source of Truth. Our solution, the Enterprise Information Factory, does more than just build a consolidate view of data. It also:

- » Processes data faster
- » Improves data accuracy
- » Ensures compliance with regulations and reporting needs
- » Reduces the costs to process, support and maintain information assets

We've taken a unique approach to building the Enterprise Information Factory (EIF), applying the principles of lean and flexible manufacturing, along with IT industry standards including, Service Oriented Architecture (SOA).

2. Describe any business and/or technical challenges about the project.

In 2004 Polk's CEO, President and Executive Committee held a series of strategy meetings to discuss how Polk could first maintain and then improve its competitive advantage amid significant industry, regulatory and technology change.

Over the years, Polk has enjoyed a position as the market leader and is the “gold standard” for automotive vehicle and consumer data. This data is used by every automotive brand to make critical decisions about their businesses. Further, many automotive suppliers, dealers, and other automotive-related businesses (finance and insurance, media, research, government agencies) utilize Polk solutions. Polk's data and applications are used by its customers to help them make decisions about areas such as dealer and network planning, parts and inventory planning, customer segmentation and target marketing, and vehicle verification to name but a few. Having served the automotive market since 1922, Polk provides

data that is 'court-tested' to defend franchise decisions made by OEMs. Further, Polk's data is used for recall purposes to ensure that every vehicle owner is notified of recall campaigns.

Never wanting to "rest on its laurels," Polk has continuously improved its data management methods over the years. Given today's environment, in which privacy compliance is introducing even tighter restrictions on how data can be used, the time was right to move beyond continuous improvement to develop a innovative approach that would revolutionize Polk's core foundational data warehouse.

Polk's executive leadership had a healthy debate centered on two fundamental issues at the earliest stages of the re-FUEL project.

First, to be successful with a project of this significance, size and scope, the IT team members tasked with accomplishing success would need to be focused fully on this project, and not burdened with other daily demands. In other words, we didn't want to "change the tires on the car while it was moving."

Second, Polk realized it was not alone in facing the challenges and complexities inherent in large-scale data management, data warehousing, and application development/integration. According to Gartner, in 2004 organizations were faced with managing 30 times more data than in 1999. This trend is not likely to change.

With both issues in mind, Polk's senior management concluded that the appropriate course of action was to develop a new subsidiary, RLPTechnologies. The charter for this organization was to develop a working software solution for use by the parent that would also be viable for other organizations.

A plan was devised to totally re-engineer the core revenue generation engine that powers the company and to do it in such a way that it:

maintains and improves the current competitive advantage for the next 10 years, and creates a subsidiary company to "spin out" the technology innovations into the market place.

In December of 2004, the Polk Board of Directors approved the re-engineering program and the creation of RLPTechnologies, a wholly owned subsidiary of R.L. Polk & Co.

The re-engineering program was code-named re-FUEL (Re-engineering Functions with Urgency, Excellence and Leadership). The re-FUEL vision was described in the charter approved by Polk's Board of Directors as follows:

"The re-FUEL vision is nothing short of revolutionizing the way data is collected, standardized, enhanced and compiled into data warehouses.The solution will be designed with the awareness of today's security threats and data privacy issues.... The solution will be designed to incorporate a high level of quality automation and statistical trending to detect, and potentially predict, data quality issues....This effort should produce a world class data collection, enhancement, and compilation solution; a system that utilizes superior technologies and methods to produce superior results and profitability. It is not an exercise in continuous improvement, but a journey of discovery and innovation"

In essence the re-FUEL team was given the rare opportunity to take a clean sheet approach to designing the new systems, processes and organization. The business vision was established, and referred to as 50/50/100:

50 Percent More Efficient

» Lower Total Cost of Ownership

50 Percent Faster

» Improve data processing and timeliness and availability

100 Percent Quality

» Protect Polk's rich heritage as the industry standard, and provide improvements in identifying problems earlier in the process

The re-FUEL team evaluated and eventually embraced a standards based, service oriented architecture (SOA) as the foundation for the new system. As a new IT architectural paradigm, SOA provides significant benefits relative to protecting legacy investments, reducing costs, and providing accelerated time to development. The team also embraced the principles of lean manufacturing - continuous material flows, standardized process, and eliminating waste - which aligned closely with the 50/50/100 goals.

Over the next 14 months, a team led by the new subsidiary (RLPTechnologies) went through an aggressive project schedule to:

- » Build a world-class organization of data management and IT professionals
- » Perform business process re-engineering to define a future state process that leverages lean manufacturing principles at its core, applied to data management
- » Evaluate and select Commercial-Off-The-Shelf technologies to assist in the development of the end-state solution
- » Build the integrated solution, with significant intellectual capital developed by RLPTechnologies (RLPT), to create a single interface for business analysts and a data-driven dependency engine to enhance the accuracy and completeness of data
- » Leverage a service oriented architecture to protect legacy applications and investments made by Polk over its long history as a data provider. This approach also provides increased flexibility and agility for Polk as business conditions and compliance change.

The system has been built, and is being deployed in phases. The project is entering the final phase of parallel operation, which will occur from March through June, 2006. Following this phase, the Polk "data factory" will use the new solution exclusively to manage the wealth of data Polk collects.

A conversion of over 2.5 billion data records from the existing Polk data warehouse will be run through the system for consistency.

The program has delivered on both the business vision (competitive advantage & 50/50/100), and the technology vision of a true service oriented architecture (SOA) – enabling Polk to recognize significant benefits, while leveraging the new system to further strengthen its competitive advantage.

3. What was your solution?

The solution (The Enterprise Information Factory) was developed by RLPTechnologies as a comprehensive software application that manages how data is collected, standardized and enhanced, and compiled it into a Single Source Of Truth (SSOT) to feed use in analytical and operational applications. The Enterprise Information Factory innovates in two primary areas: business process and technology.

Business Process Innovation

We've learned from the principles of lean manufacturing, and applied those lessons to the area of data processing in the development of the Enterprise Information Factory (EIF). Key lean principles of continuous material flow, process automation, standardization, quality controls and continuous improvement are built into the core of the solution.

The solution handles incoming data in much the same way as a factory built on the principles of lean manufacturing handles raw materials. As soon as inbound data (the EIF's raw material) arrives, the factory immediately recognizes the availability of data and begins processing it. This automated collection and real-time processing of Polk's data reduces the overall time for the data to reach its key business intelligence and other transactional business systems. The Enterprise Information Factory has eliminated manual processes, allowing Polk to recognize overall improvements of up to 70 percent on processing inbound data.

This type of real-time automated processing is possible because Polk's business analysts have the ability to setup, or "tool" the Enterprise Information Factory with custom business rules for the processing of any particular source or type of data. Once configured, the EIF runs as a fully automated system, requiring minimal manual activity. Like any highly automated system, the solution needed a robust monitoring and process control system. An operations management dashboard was built to provide visibility into the health of the factory. The operations management

portal displays real-time metrics of the EIF performance. It also provides access to exception queues, allowing analysts to resolve issues that occur during automated processing. When exceptions are encountered, the specific data in question is “pulled off the line” and alerts are sent out. This allows for the continuous flow of all other data moving through the EIF, and makes Polk business analysts immediately aware of the issue, so they can begin prompt resolution. Once an exception is resolved, the data is placed back into the Enterprise Information Factory’s workflow to complete processing.

The EIF may raise exceptions for a number of reasons, but a sophisticated data quality module is the primary source. The data quality module evaluates data content at various check points in the factory, from the time data arrives through to the delivery of finished data products. The data quality tool allows Polk analysts to establish rules on such factors as data consistency, completeness and value distribution. Polk analysts can use business rules to adjust thresholds up and down for acceptable variances in the data. As a result, Polk quickly identifies data quality issues and responds to them early in the data processing lifecycle.

Through the innovative use of lean manufacturing principles in the field of data processing, the Enterprise Information Factory has allowed Polk to recognize significant business process efficiencies in a once very manual process.

Technology Innovation

The re-FUEL program was structured to allow the team to architect the Enterprise Information Factory from a clean sheet of paper, without concern for the technology constraints of existing platforms. This allowed the team to develop a very innovative technology solution. The three key areas of innovation include: the creative use of an enterprise service bus (ESB) as the solution backbone, a custom Service Orchestration engine that provides dynamic integration to web services, and the implementation of a GRID computing platform.

Enterprise Service Bus (ESB)

The ESB serves as the JMS messaging backbone of the Enterprise Information Factory. The ESB is essentially the underlying foundation that holds together all modules of the factory, such as:

- » Data Capture,
- » Standardization,
- » Data Enhancement,
- » Quality profiling and
- » Assembly

And common foundation services, such as:

- » Logging,
- » Exception handling,
- » Scheduling and
- » Security.

A standard message structure facilitates communication between components in the EIF. This approach provided a tremendous amount of flexibility when developing and integrating components of the solution to create a large composite application.

Service Orchestration Engine

The Enterprise Information Factory is founded on a service oriented architecture. At the center of that SOA is a custom developed service orchestration engine. This engine manages all business services executed against the data moving through the factory. The Service Orchestration engine was specifically designed to handle high volume and a high degree of flexibility for the handling of changes to data and business services.

The solution provides an application that allows Polk business analysts to create and modify service orchestration profiles. These profiles are based on registered services and the type of data that is being processed.

The EIF service orchestration function is based on a data driven dependencies engine (D3E). At run time the service orchestration engine retrieves a profile that defines what services have been assigned to the data source. Through the use of Web Service Description Language (WSDL) in the service registry and the inbound data schema, the engine automatically derives an optimized execution path. Unique parsing, segmenting and aggregation routines developed by RLPTechnologies allow the engine to perform parallel processing and manage calls to and from all services. All communication with registered business services use common web service standards and protocols such as SOAP, JMS, HTTPS, XML and WSDL.

The capability provided by service orchestration allows Polk to quickly integrate business services provided through either the use of commercial off-the-shelf software (COTS), legacy applications, or external providers. Polk integrated business services and used COTS such as DataFlux for name and address cleansing, and iLog for sophisticated VIN rules processing. Additionally, Polk's legacy business logic in COBOL was wrapped with a web service interface and connected to the factory.

4. Describe your vendor selection process and reasons for choosing JEMS.

The selection of JEMS was decided very early in the project. During the early stages of the reFUEL project, the team completed a conceptual and logical architecture. Based on this target architecture several key foundation software and hardware components were identified, which included application server and object/relational persistence amongst others.

The aggressive time frame did not allow the team to do a broad sweep of available products, so the team quickly developed a short list based on market research firms such as Gartner. Based on this research the team evaluated the first round of candidate technology in a lab, to assess stability, maintainability, performance and interoperability. The results of this testing, coupled with the desire for RLPT to eventually develop a commercial product – JBoss was the clear winner.

5. What role did JEMS products play in the solution?

The JBoss Application Server and Hibernate Object/Relational Persistence products were critical components of the technical foundation for the solution.

The JBoss application server is used to run all Java components of the Enterprise Information Factory (EIF) developed using Hibernate. The core EIF Java applications used to configure the operations of the data factory include: Data Capture, Reference Data Management, Data Quality, Service Orchestration Gatekeeper and Assembly functions. These are critical business applications used by Polk Analysts to perform their day to day jobs.

The performance and scalability of the factory was paramount for this project, and was achieved using the JBoss application server. The solution needed to scale to support over 100 transactions per second while processing through several business services. The EIF Service Orchestration Engine (developed in Java), is the foundation of the SOA architecture and manages all data movement, and calls to and from all registered services. In addition to running custom developed components, JBoss application server was also used to run several of the COTS products used in the solution. This included an implementation of iLog jRules business rules execution engine running in JBoss. The business rules implementation for Polk Vehicle decoding contains well over 600,000 rules in multiple rule sets deployed across multiple application servers, one of the largest implementation of rules for a solution using iLog.

6. What was the overall impact of the project on the business? (eg: ROI, competitive advantage, time to market?)

The Polk Executive Committee approved the re-FUEL project as the #1 priority for Polk's FY05 and FY06 business plan.

At its basis, the re-FUEL project focused on re-engineering and boosting the performance of Polk's core revenue-generating engine, the power driving the company's business success. Polk has realized significant business results from the re-FUEL project, including both revenue protection and generation, combined with equally significant cost-savings.

Revenue Generation

The project positively impacted Polk's revenue picture, both in terms of protecting current revenue streams as well as supporting additional revenue growth.

Revenue Protection

Over 50 percent of Polk's market-leading automotive data and analytical solutions are supplied by the new solution. Enhancements in the speed, accuracy, and quality of the data, combined with improved regulatory compliance capabilities, have enabled Polk to maintain a position of strength compared to its competitors. Two elements of the 50/50/100 plan—50 Percent Faster and 100 Percent Quality—are worth noting as drivers of significant business benefits for Polk.

50 Percent Faster – Tests to date show improvements of up to 70 percent in data-file processing speed (on average). For example, an average state registration file that previously would have required manual processing by as many as three full-time employees (FTE's) and four hours of processing time, now is processed in an automated fashion in roughly 23 minutes. Further, RLPT's approach to grid computing has allowed the solution to scale to process ~100 transactions per second, nearly four times Polk's average of 25 transactions per second - providing headroom to accommodate, processing spikes, future transactions or business growth.

100 Percent Quality – The standardization and enhancement functionality of the Enterprise Information Factory measurably improves the accuracy and completeness of the data, preventing quality problems that might impact customer satisfaction. Automated data quality checkpoints allow for earlier recognition of problems and enable the team to resolve issues before the data is delivered to Polk's business intelligence and operational applications. This functionality drives a focus on preventing issues--or at worst, recognizing them early--following the rule of thumb that "It costs \$1 to prevent a problem, \$10 to identify a problem, and \$100 or more to fix it."

Faster delivery of higher-quality information should translate into improved customer satisfaction, resulting in continued long-term business commitments. Given Polk's subscription-based models, this will enable continued positive financial returns for the 135 year-old company. Fending off any threats to the core business will allow Polk to **maximize new revenue generating opportunities and drive double-digit growth.**

Revenue Generation

A significant benefit delivered by the solution for Polk is their ability to shift focus from data management to product strategy and application development. Armed with the flexible environment provided by the Enterprise Information Factory, Polk's Product Strategy group can look for new data sources to enhance its offerings, while also developing new analytical and operational applications to leverage more timely and complete data. Polk expects these new capabilities to prime the company for future growth, and embolden managers with the knowledge that they will experience reduced time to market in future development efforts. The EIF solution will be rolled out in phases, and the team is currently working on how best to deploy it globally to further strengthen the consistency and completeness of Polk's data and product applications worldwide.

The formation of RLPTechnologies was founded on the knowledge that the Enterprise Information Factory could also solve challenges facing other large organizations--and generate new revenue streams in the process. Market trends support this approach. A November 2005 Gartner study on data integration, for example, shows that in North America, 21% of respondent companies expect constant investment in data integration, with 60% expecting to increase such investments in 2006. Further, the same study asked respondents about the "degree to which their SOA initiatives included the service orientation of data assets and creation of data services." Gartner concludes that "with only 37 percent indicating a strong focus on this topic, most organizations appear to be at risk for failure in their SOA efforts because they are probably not addressing fundamental issues, such as consistent transformation, delivery, and quality improvement of the data."

With its comprehensive approach to data integration and management and with a service oriented architecture at its core, RLPTechnologies solution (the Enterprise Information Factory) is well positioned to capture continued investments by businesses in these areas.

Supported by strong partnerships with industry-leading consulting firms and software technology companies, RLPTechnologies will provide significant growth for the parent company. The three-year business projections are expected to deliver **between 5 and 10 percent top-line revenue growth for Polk.**

Cost-Savings

The first element of the business vision that was established in the 50/50/100 plan was 50 percent more efficient. The re-FUEL project allowed this goal to be met, with significant cost benefits to be realized by Polk in two core areas:

Leaner, Better Aligned Team

Prior to the re-FUEL project, Polk's previously-named Data Operations team had moved from Cincinnati to Polk Headquarters in Detroit in 2003, resulting in a centralization of IT functions. The re-FUEL project transformed the structure and size of this group, creating a more cost efficient and focused unit. Renamed the Data Factory, the group is now **43 percent smaller**, and team members have significantly different roles and responsibilities. The group is structured more efficiently, with roles that align directly to the functions of the solution (Data Capture, Standardization & Enhancement, Reference Data Management, Single Source of Truth/Operational Data Store, Assembly, and Operations Management). Further, the reduction in manual processes has enabled the group to focus on strategic management and analysis of data, including areas such as issue resolution and handling.

Lower IT Operating Costs

The implementation of RLPT's grid computing model will result in significant savings for Polk. By moving away from a mainframe-based system, the grid will operate with **hardware costs that are 65 percent less**. This change amounts to savings of millions of dollars per year for Polk. Finally, improvements made to the open systems environment are leading to **additional savings of 30 percent per year compared to prior operating budgets**.

7. With the savings gained from implementing JEMS, how did you reallocate your cost savings within your company?

The savings gained from the project including those from the implementation of JEMS are invested back into the business to drive product development efforts to strengthen Polk's competitive advantage as the market leader in business intelligence for the automotive industry.

8. Technical description of implementation/size of deployment? (Hardware, applications, o/s, databases, etc.)?

GRID Computing Platform

The technology stack for the Enterprise Information Factory operates in a grid computing environment running Linux Redhat on Intel Xeon processors. The target production grid is comprised of 49 servers and 118 processors. The database holding the Single Source of Truth contained in a 4.5TB database with over 2.5 Billion transactions. The operation and management of the grid is accomplished through the combination of JBoss clustering, the EIF Service Orchestration and Oracle 10g GRID.

Embedded in the EIF solution is leading commercial off-the-shelf software (COTS); Oracle 10g database grid, portal and Oblix, Tibco BusinessWorks, Dataflux dfPower Studio and iLog jRules to accelerate the time to market for the solution. The EIF application that wraps all of these technologies together is a series of J2EE applications running in clustered JBoss application servers.

The grid based computing platform has allowed both significant cost savings and flexible scalability options to provide capacity on demand.

9. Did you leverage JBoss support services, training or consulting? If so, how was your experience?

Yes, we engaged JBoss for support, training and consulting. The development team ramped up very fast and a number of outside contractors were used. To ensure proper use of Hibernate we purchased training programs for our team. Additionally, we used JBoss consulting to provide assistance with tuning activities.

10. Advice to other companies considering JEMS.

Establish a sound SOA architecture up front and stay true to it as much as practical.