



---

## **Managing Business Complexity with Data Warehouse Consolidation at Pfizer Inc.**

---



A BTI Case Study



**Executive Summary**

The objective of this case study is to share the business motivation, approach, architecture and best practices from Financial Data Warehouse Platform Consolidation (FDWC) project at Pfizer Inc.

Over the past decade, the company has grown via many acquisitions. Given the complexity of the business, the benefits of economy of scale from the acquisitions are sometimes difficult to achieve. That complexity forces business units to be partially integrated, especially for financial matters. The company was facing massive increases in data volume while mandating significant reductions in expenses. In response, Pfizer needed to be more innovative and productive with its enterprise financial reporting system. Doing nothing was not an option.

The FDWC project consolidated several large financial data warehouses onto a single common platform. The technologies required were a MPP data warehouse platform from Teradata and Data Integration solution platform from Informatica. By building a scalable platform of shared services, the total cost of ownership was reduced for each new application developed. The cost justification was based on saving money by avoiding the predicted high costs of their current infrastructure in the following areas:

- Database technology that scales gracefully to handle higher data volumes without large cost increases
- Data integration solution framework that would allow sharing of common data across applications
- Consolidation of three application support teams into a single integrated team

This case study covers:

- Business requirements that drove the financial data warehouse consolidation
- Technical requirements for data warehousing and data integration that satisfied business requirements
- Architectural approach and best practices for utilizing the Informatica-Teradata solution
- Benefits that are realized and expected
- Lessons that were learned

Contents	
<b>Executive Summary</b> .....	<b>1</b>
<b>Company Background</b> .....	<b>2</b>
<b>Business Problem</b> .....	<b>4</b>
<b>Business Solution</b> .....	<b>6</b>
<b>Project</b> .....	<b>7</b>
<b>Benefits Realized</b> .....	<b>11</b>
<b>Lessons Learned</b> .....	<b>12</b>
<b>Synthesis</b> .....	<b>13</b>
<b>References</b> .....	<b>14</b>
<b>About the Methodology</b> .....	<b>15</b>
<b>About Bolder Technology Inc.</b> .....	<b>15</b>
<b>About the Sponsor</b> .....	<b>15</b>

## Company Background

Pfizer Inc. is a global biopharmaceutical company with a strong research emphasis. Pfizer's mission is to be valued by its customers, business partners, and the communities in which its business resides. They are dedicated to humanity's quest for longer and healthier lives through innovation.<sup>1</sup>

Pfizer does business in more than 150 countries. As of fiscal year 2009, the company employs 116,500 employees with annual sales of \$50 billion. Its headquarters is in New York City with facilities throughout the world.

The company operates in two broad segments: Biopharmaceutical and Diversified. Biopharmaceutical includes the Primary Care, Specialty Care, Established Products, Emerging Markets and Oncology customer-focused units. Diversified includes healthcare products for consumers and animals.<sup>2</sup>

In 2007, the new CEO and Chairman, Jeff Kindler, urged the company to become the "Health Information Company" to educate people about drugs and not just manufacture the drugs.<sup>3</sup> Then in 2009, Jeff Kindler summarized the nature of the biopharmaceutical business as:<sup>4</sup>

- Society has unmet medical needs throughout the world and must make large and risky investments required to find treatments for those needs.
- Drug discovery and development is enormously expensive and risky. As stewards of your capital, investments must be made wisely.
- Pfizer should not rely on a few highly profitable drugs for a major part of their revenue. A more balanced approach is necessary.

In response to this business environment, Pfizer stated that they will refocus and optimize their drug portfolio, capitalize on established (generic) products, grow in emerging markets like China, India, Brazil and Russia, and expand their culture of innovation, and do all this amid



**The IT fabric of Pfizer is unique and characterized by the high complexity of supporting a global corporation.**

aggressive  
cost  
reduction.

Pfizer has a distributed management culture where autonomy is the norm. However, it is autonomy with a focus on collective

performance. For instance, one team should create the wheel, rather than many teams creating many different wheels. The principle is... Don't reinvent the wheel, but think of a hundred ways to use that wheel. The end result is to enable people to be more productive and contribute to the collective performance of the company.

## Global Data Management

Information Technology (IT) within Pfizer is named the Business Technology (BT) group, which is composed of two parts: Shared Services (centralized special IT services) and Business Partners (business-facing IT groups aligned to each business unit). Within Shared Services is Global Data Management (GDM), which focused on the Data Management area and is the primary group within Pfizer responsible for the FDWC project.

The IT fabric of Pfizer is unique and characterized by the high complexity of supporting a global corporation.<sup>5</sup> To support the business, IT must understand how every business unit works and how the different groups work together. Further, business users within Pfizer around the world have high expectations of IT.

A sense of urgency to find innovative ways of coping with enormous complexity was driven by Pfizer's demand to operate a global biopharmaceutical business with more effective use of information assets amid the anticipated explosion in data volume. At the same time, IT costs were increasing drastically, especially in areas like data warehouse development and operation. It was apparent that strong cost reductions needed to be enforced. To accomplish cost reductions with shared services, business objectives were formulated within GDM and then executed through the FDWC and other projects.

Josh Raysman is the lead for Solution Engineering and Project Management for GDM and is "ultimately responsible from the functional deliverables for anything designed and built in our shop". He summed up the success factors for GDM:

*We are maximizing our investment in data management solutions by focusing on a long-term total cost of ownership. For long-term success, solutions should have scalability, extensibility, and supportability. In particular, to have supportability means low maintenance costs. A high percentage of*

## Managing Business Complexity at Pfizer

our IT budget is spent on application maintenance. We want to get that under control. If we build it right the first time, then we can reduce expenses of care and feeding over time.

In particular, GDM was “responsible for the design, build and operations of data warehousing, data integration, master data management, including reference data, and business intelligence solutions by providing the common processes, frameworks, and toolsets for effective data stewardship, governance and quality management.”<sup>6</sup>

As shown in Figure 1, GDM coordinates IT functions (design, build, and operate) from the perspective of four lines of business:

- Research and Development, which is the traditional driving force for Pfizer in its competitive marketplace
- Finance and Procurement, which is now Financial and Business Operations (a broader area)
- Commercial, which are the Sales and Marketing functions
- Manufacturing, which produces and distributes medical drugs worldwide for humans and animals
- Medical, Diversified, Corporate Functions (such as Legal and Human Resources)

Kevin Fleet is the Executive Director of GDM. He is a visionary who saw the need for consolidating IT functions into a set of Shared Services for use across the entire enterprise. He initially formed and chaired a Business Intelligence and Data Warehouse (BI/DW) council with representation from all lines of business and all regions. Over time, the Council has evolved into BI/DW Shared Services within Global Data Management (GDM) organization.

IT is looking across organizations for common business functions so that shared IT services can reduce costs through economies of scale.

Each of the four LOB units has an IT person who is the primary liaison to that business function. To support these four units, the GDM Shared Services are divided into six functional areas:

- Solution Delivery- client engagement, requirements definition, project scoping and estimation
- Technology Architecture and Platform Engineering – technical reference architecture and standards, core platform engineering and design, new technology evaluation

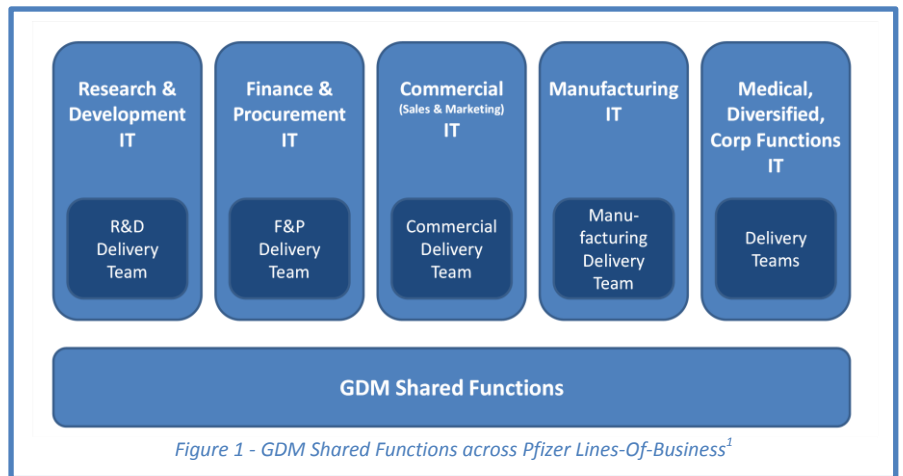


Figure 1 - GDM Shared Functions across Pfizer Lines-Of-Business<sup>1</sup>

- Solution Engineering – physical design of solutions, design patterns and standards
- Information Products – master data, metadata administration, data inventory, data architecture
- Service Operations and Support – application support, maintenance, deployment
- PMO and Business Operations – project management office, resource management, internal metrics, vendor management, internal audit coordination

For the four LOB units, there is a Delivery Team who is responsible for delivering all solutions to that type of business function. The members of the Delivery Team often do not actually perform the design and implementation work but rather lead and review the work being performed by outside contractors. There are usually five members whose roles are:

- Solution Delivery Manager
- Solution Engineer
- Information Product Lead
- Support Lead
- Solution Architect

For a proposed project, the Delivery Team would typically map the resource requirements for the project against the available resource pools, which are agreements with strategic vendor partners to provide a specific mix of skilled resources. If the resources are available, the proposed project will surface as demand against various parts of the resource pool.

A Project Team would be established consisting mostly of consultants specializing in the required skills and employed by the contractors. Under the supervision of the Delivery Team, the Project Team would design and implement the project. Using the resource pool, Pfizer avoids using new contractors on every project. Instead, Pfizer can draw on a “bench of knowledge people who have worked on previous projects”. This results in a reduced ramp-up time

## Managing Business Complexity at Pfizer

The FDWC project was slightly different in that a Request for Proposal was issued to tap into a set of new resources, which were later moved into the resource pool of Pfizer. In this context, the FDWC project was launched. The next section describes the business problem that drove the project objectives.

### Business Problem

Over the past decade, Pfizer has grown via many acquisitions, sometimes up to one hundred companies per year. The benefits of economy of scale from the acquisitions are difficult to achieve since the acquired companies continue to operate autonomously to minimize operational disruptions. In particular, the effort of integrating so many unique IT infrastructures and disjointed data models was enormous. Every acquired company had a different view of customer, product, revenue, and so on, all of which had to be rationalized into a common model.

In response, Pfizer needed to be more innovative and productive with its enterprise financial reporting system. In particular, Pfizer needed to develop a scalable infrastructure for their financial data that will consolidation all financial data across the corporation. Further, this infrastructure needed to be flexible to accommodate additions of new financial data while lowering the Total Cost of Ownership. These were the business imperatives that drove the conception and execution of the FDWC project.

In addition, customers of Pfizer are large-scale distributors, which put Pfizer at a distance from the individual pharmacists and physicians. Yet it is important for Pfizer to get closer to the consumers and understand what the consumers are thinking and needing. Moreover, several states in the USA banned the commercial use of syndicated prescription data, thus distancing Pfizer more from the consumer of its products.

Around 2008-2009, Pfizer created a direct channel to individual retail pharmacies under a program called "Direct to Pharmacy", in contrast with a fewer number of large sales order to a small set of wholesale distributors. This initiative results in an exponentially large number of sales orders from a large customer base, requiring the capability to support a huge

increase in data volume. To support this program, the Management Information Center (MIC) data warehouse was created in Europe using the Oracle e-Business Suite and the Oracle Warehouse Builder, with Oracle Professional Services.

### ***Data Tsunami is Coming!***

These business initiatives drove several imperatives for Pfizer IT to be more innovative, flexible, and productive. Pfizer needed a major IT transformation in the way that data was captured, managed, and analyzed...and to do so with less resources.



The most visible and understandable part of this IT transformation is the data tsunami – an explosion in the volume of data needed to fuel innovation at Pfizer. As Krishan Kumar observed, "We were seeing the data tsunami coming! And, the tsunami was coming from several fronts."

In 2007 at a conference, the VP of Worldwide Technology and Engineering was talking to his peers at other companies. They concurred that all were seeing a big data tsunami coming. He returned and challenged the IT organization to scale tenfold, both in data management capacity and in organizational capabilities. He probed, "Could we handle the 10x workload? Could we handle 10x demand for new applications and systems?"

Supporting internal research at Pfizer, data volume is driven by the combination of all research data, such as compounds, experiment results, and the like. For instance, new laboratory technologies like High Throughput Screening<sup>7</sup> (HTS) are causing data volumes to explode.

Further, Pfizer has acquired Warner-Lambert and Pharmacia) and then acquired another larger company Wyeth. The resulting data set for research is now the largest set of chemical compound data in the industry. In addition, new work being done to leverage the human genome will exponentially increase the volume of data. And, external clinical trials were also increasing rapidly, implying a need to manage more data.

## Managing Business Complexity at Pfizer

Driven by these initiatives, data within Pfizer drove requirements for: increased freshness (often more than daily updates), three times increase in volume, six times increase in workload processing, doing all of this faster, and doing all of this with less resources. However, their hardware/software

***Doing nothing was not an option, since doing nothing would result in much higher cost.***

infrastructure was already being run at maximum capacity and struggling to support current demands.

As was often mentioned during our interviews, as new data marts were spawned. Pfizer had to take bold action to transform their IT infrastructure or general business would suffer.

### ***Need for Massively Parallel Processing***

The Pfizer infrastructure was entirely composed of symmetric multiprocessing (SMP)<sup>8</sup> platforms running Oracle under UNIX with EMC storage units. This technology has served Pfizer well for most of a decade, but became unable to scale and satisfy the tenfold increase in data volume and workloads. Further, Pfizer has outsourced its infrastructure support, implying a reduced level of expertise to tune and optimize multi-tiered architecture composed of Oracle, Sun, EMC, etc. The conclusion of Pfizer's IT architects was the need for a MPP database platform.

Krishan Kumar described the limits of their current Oracle solutions:

*The complexity of using Oracle-based solutions was increasing as the data volumes were growing. Pfizer needed a new database platform that would scale and we knew that this could not be done with Oracle. Various vendors were releasing data warehouse appliance products that appeared to offer good cost performance advantages. Inspired by the emergence of data warehouse appliances from Netezza and others, Pfizer began surveying the market for MPP database technology for building data warehouses.*

Pfizer's IT experts reviewed their options and concluded that shared-nothing MPP platforms allowed the architectural integration of hardware, operating system, database management, and storage onto one system that could be optimized for their specific workloads. The MPP technology was proven to support ad hoc queries and complex analytics against multiple terabytes of data, a task that was very uncertain using their current SMP platforms.

Shared-nothing MPP platforms were also designed for high availability with component redundancy so that there was no single point-of-failure. If one component failed, another would take its place. The disk drives were able to be "hot swappable" meaning that if a drive failed, other drives would immediately take its function. And, a new disk drive could be physically swapped with the failed one, without causing the entire system to be shut down.

Finally, the most important characteristic of a MPP platform was its ability to scale in its capacity to store data and process queries. There was plenty of evidence from prior customer experiences that if a company has a 2-node MPP platform, they could buy additional two nodes and double their capacity without having to add technical support staff.

### ***Consolidation of Financial Information***

As discussion about the MPP platform was unfolding, consolidating all of the company's financial information in compliance with various government regulations became a critical business imperative.

Their legacy financial information was managed on several separate application systems with associated data warehouses,<sup>9</sup> such as:

- Financial Data Warehouse (FDW)
- Global Information Factory (GIF)
- Management Information Center (MIC), which was actually three data warehouse instances, due to scalability issues

Consolidating these financial systems was a very difficult task for IT. The task consisted of understanding the data contents of these data warehouses at a detailed level and then 'rationalizing' a new design that integrated common data (usually reference or master data about products, P&L units, regions and the like). It was also a good time to eliminate unused data that is no longer relevant to the business. In addition, consolidation required a redesign and redevelopment of the ecosystem around the financial data, such as data flows from source systems into warehouse and applications that maintain and report on data.

Each data warehouse environment migration effort varied in complexity and effort. For instance, the FDW integration layers used IBM Data Stage ETL, but consisted of a shell for Oracle PL/SQL scripts, which require substantial effort to rewrite for the Teradata data

## Managing Business Complexity at Pfizer

warehouse. Pfizer decided not to rewrite these scripts as part of the FDWC project but gradually transition the system over the coming 18 month period. On the other hand, the GIF data warehouse environments required moderate redesign. And, the MIC data warehouse required a full redesign due to solution architecture issues as well as the need to consolidate 3 warehouse instances to one to support the current business functionality. All environments required migration, re-write and/or upgrade of related components including ETL, PL/SQL and sometimes a consolidation of the web interfaces used to populate master data for the three data warehouse environments.

***Scalable framework of shared services, thereby reducing the total cost of ownership for each new***

In other words, the

FDWC project is one step in a long-term approach toward a single enterprise financial data warehouse. This first step focuses on consolidating several data warehouses onto a single platform. The subsequent steps will opportunistically redesign data structures as business value and technical resources permit, while driving toward a common data model and a single financial data warehouse in the long term.

### ***Finance One Standardized Reporting***

The original objectives of the FDWC project did not include any changes that would impact business users. All the above objectives were necessary infrastructure improvements that did not change the externals of business applications. As Krishan Kumar summarized, "The business units said only one thing to IT, which was to save three million dollars and do not bother us." In other words, the business units did not want procedural changes or different functionality.

Late in the game, there was, however, a significant expansion of the project scope with the process reengineering of financial reporting. Another initiative called Finance One (F1) was focused on standardizing all financial reporting and processes across the globe to comply with the increasingly complex regulatory environments. Standardized reporting for Finance One was added to the FDWC project objectives, which shifted the emphasis on infrastructure enhancement to

business reengineering. Support for Finance One implied that:

- Old reporting processing for several legacy systems must be analyzed
- New reporting formats mapped to this layer
- New reports implemented upon the new consolidated MPP data warehouse
- Bridge established to map the new reports back to the old report

Initially the F1 requirement specified thirty standard reports, which grew over six months to over forty reports.

## Business Solution

---

To respond to this data tsunami and other business problems, Pfizer's direction was to build a scalable framework of shared services, thereby reducing the total cost of ownership for each new application developed. The business case was based on saving money by avoiding the predicted high costs of continuing their current solution framework. The key assumptions were that (a) there was database technology that would scale gracefully without huge cost increases from higher data volumes, and (b) there was a solution framework that would allow the sharing of key services across applications so as to reduce the cost per application.

A choice was made strategically to centralize common BT functions, including data management and then to share a common support model for the infrastructure to support those functions. Therefore, Pfizer would have the ability to scale both operations and delivery capacity, which was critical in light of the anticipated volume increases in data.

### ***MPP Data Warehouse Platform***

Therefore, Pfizer's approach was to acquire a suitable MPP platform and consolidate the financial data warehouses by migrating them onto the MPP platform.

Acquiring the MPP platform was the simpler of the two tasks. After considering several vendor alternatives, Pfizer decided to acquire Teradata as their MPP platform and extend their existing Informatica data integration environment to support the MPP platform.

Josh Raysman has high expectations for the use of the Teradata MPP platform and puts a

## Managing Business Complexity at Pfizer

different twist on the motivations for selecting Teradata:

*We selected Teradata less on its high performance but as a bridge to more simplicity and agility. With the Teradata platform, we expect to support less code by evolving toward simpler systems. For instance, we expect to retire thousands of code lines as we move from ETL to ELT.*

*And, we need to be more agile ...more nimble. Business definitions are constantly changing, which in our current infrastructure can happen in many places (table definitions, ETL scripts, query tools, etc). We aim to do it once in one place! ...and have that definition use in all other places. This is important to us. We hope that this one place will be in the Teradata semantic layer. Define it once and make available to all data consumers, human or application.*

The second task of migrating several financial systems to the Teradata platform was more difficult. The objective was to reduce IT infrastructure costs and to secure capability for significant growth. The program objective was stated as:

*The overarching goal is to increase Pfizer's ability to acquire and leverage new and existing data sources, increased volumes and while at the same time drive down data warehouse processing complexity and cost. ...on a single strategic platform that offers increased performance while at the same time reduces support cost and overhead. ...along with*

*supporting near real-time analytical capabilities.*<sup>10</sup>

The goals to be attained were:

- Implement solution architecture that meets emerging business needs, aligns with the information strategy, and retire existing end-of-life infrastructure.
- Implement a solution that will achieve a 10x-50x performance improvement and reduce overall TCO over the current environments.
- Introduce massive parallel processing capabilities that are tightly coupled with the underlying database, storage and operating system components achieving significantly enhanced performance and streamlined data processing.

Based on the responses to the RFP, Pfizer decided to perform the work in-house and limit the scope to GIF and MIC, leaving the FDW platform as is. All three DW environments were to be redesigned at the presentation layer with reporting that was integrated across the enterprise.

## Project

The FDWC project unfolded over several years in various stages and continues to evolve, so it is difficult to place sharp boundaries on project duration and resources. The major milestones for the project are shown in Figure 2 below:

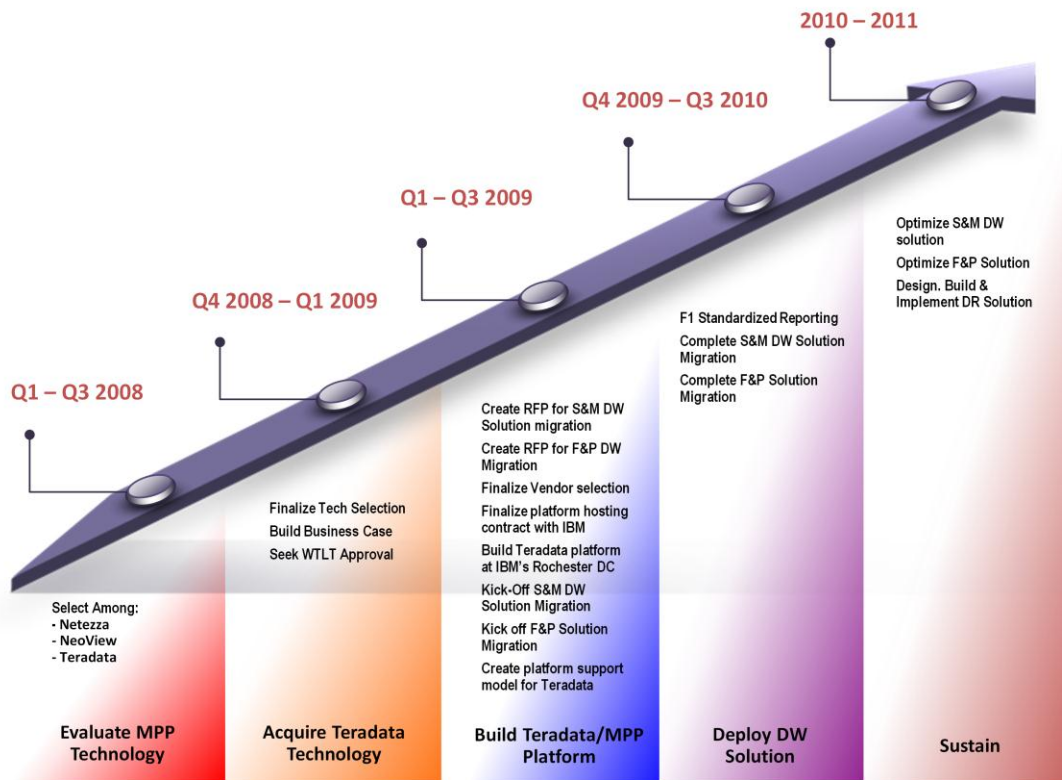


Figure 2 - Roadmap for FDWC Project<sup>1</sup>

# Managing Business Complexity at Pfizer

## Project Status

The first two phases consisted of evaluating, selecting and acquiring the MPP platform from Teradata, which was completed in late 2008. The third phase of designing and implementing the Pfizer financial system occupied most of 2009. The tasks for the third phase consisted of Analysis, Design, Development, Test, and Parallel Run.

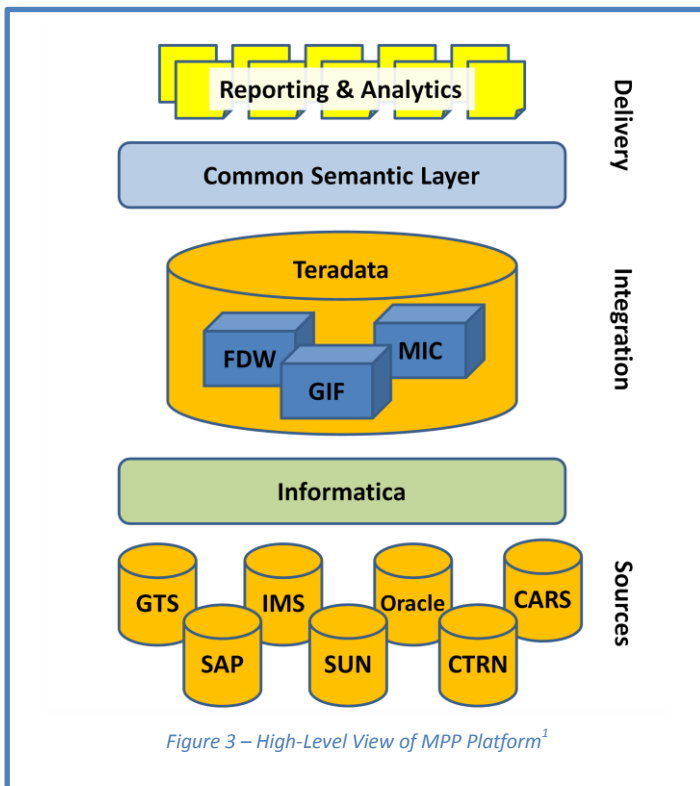
The project is currently in the fourth phase of deploying the solution for use by the business units. Josh Raysman summarized the status as:

*We are about a year into [the fourth phase] and feel very good about the prospects. There have been several user acceptance sessions. We are currently doing system testing and performance optimization. The Go-Live event will be in October. We will initially run the old and new systems in parallel.*

An emphasis is to build credibility for the information generated by the new system. Methods for comparing financial figures between the new and old systems were implemented to support this credibility.

## Architecture Overview

The high-level architecture for the FDWC project is shown in Figure 3. Starting at bottom are the various transactional Systems-of-Record that are the data sources for the data



warehouse. Informatica serves as the data integration solution required to extract data from the source systems, load the data into the data warehouse, and then transform the data by verifying its accuracy and integrating with related data.

Next layer is the MPP data warehouse supported on the Teradata platform. Within the data warehouse are several logical collections of financial data, such as FDW, GIF and MIC. Over the data warehouse is a Common Semantic Layer that will present the columns in tables to business users in business terms like customer or region. Finally, the top layer is the presentation layer that supports a variety of reporting and analytics tools.

This architecture is designed to accomplish the following:<sup>11</sup>

- Simplifies data integration with a multitude of current and future new data sources
- Enables unification of data definition by eliminating multiple views and interpretations
- Forms foundation for consistent reporting and analytics, as a single version of truth
- Provides overall lower platform Total Cost Ownership
- Fosters economies of scale by migrating multiple data warehouse onto a common platform
- Provides a common point of integration for future endeavors
- Provides robust scalability to support future capabilities beyond current technical constraints
- Enables the opportunity for eliminating data redundancy in solutions having common data sources

## Migration of Financial Data Warehouses

The center of the FDWC architecture is a collection of financial data warehouses or, as some would say, separate (but large) data marts that focus on a small set of subject areas to support a particular organization unit. The FDWC project is the migration of three of those financial data warehouses from their current Oracle platforms to a Teradata platform. These data warehouses were:

- Three Management Information Centers (MIC), each for Americas, Asia & Europe
- Global Information Factory (GIF)
- Financial Data Warehouse (FDW), which was later deferred as explained below

It is not a simple migration since some changes are necessary to leverage the capabilities of the Teradata platform, such as physical data models, use of indexes, view definitions, ETL scripts. The effort varied by data warehouse.

## Managing Business Complexity at Pfizer

There are two different issues with this migration effort. First, there is the effort required to co-locating these data warehouses on the Teradata platform leveraging the capabilities of Teradata. Second, there is the effort of integrating data among the data warehouses achieving data consistency and eliminating data redundancy. The emphasis was definitely on the first effort as an initial step toward an enterprise data warehouse.

**An important aspect of this architecture is Pfizer's strategic partnership with Informatica and Teradata.**

Josh Raysman stated, "Long-term I want to

do more data integration on the Teradata platform. Currently we have four logically different data warehouse co-located on the Teradata platform. Over time, we will gradually integrate the data, consolidating the redundant data and merging the rest into well designed tables." So IT chose not to do a massive data integration all at once, instead rehosting their data warehouses for scalability and later peeling off integration projects to avoid disruption to the user communities.

As the FDWC project progressed, it was decided to retain FDW on the current Oracle platform instead of migrating to the Teradata platform. FDW is nearing the end of its life because of the planned SAP roll-out. Further, since the ETL scripts were written 95% in Oracle and 5% in IBM DataStage, a complete redesign of the old ETL would have been required. Further, there was no business benefit to be attained, except for standardization. Finally, there was no anticipated performance gain and no new business requirements. Support costs would be the same. The standardized reports (F1) could be implemented on both platforms equally as well.

### **Strategic Partnership with Informatica and Teradata**

An important aspect of this architecture is Pfizer's strategic partnership with Informatica and Teradata. Both vendors are supplying an array of products to enable this architecture. In addition, they are interacting with Pfizer on a continuous basis, offering advice at multiple levels.

Over the last decade, Informatica has evolved to be the standard for data integration within Pfizer GDM. The benefits of this single solution from Informatica include:

- Cost-effective to manage and partner with one vendor who is a leader in data integration
- Platform can be extensible to multiple projects without having to configure new environments
- There is no need to cross-train developers for multiple platforms
- Support issues addressed faster with a single point of contact
- High availability of trained developers in the industry

As an important part of the partnership, Pfizer has had a presence on Informatica's Customer Advisory Board for the past six years. Besides building relationships with the Informatica executive team, Pfizer has benefitted from priorities for new product functionalities as well as being consulted on relevant strategic decision, such as product acquisitions.

Krishan Kumar discussed with us the various facets of this partnership:

*Both Informatica and Teradata are industry leaders in their space. In data integration, we have not seen a product that is as reliable and comprehensive as Informatica. Pfizer has had a partnership with Informatica since late 1999. And, we (Pfizer and Informatica) have matured together.*

*The use of Informatica was growing in various different parts of the company. Managed as a common platform for data integration, over thirty applications emerged. Informatica was an evolution for us. Not a rethinking but more evolving our experience of further optimizing this shared service approach. With the increased demands upon Pfizer within the highly regulated pharmaceutical industry, Pfizer has expanded our use of Informatica using additional products, such as the Information Lifecycle Management (ILM) product.*

The Informatica and Teradata joint solution uniquely enables Pfizer to adopt common architectural standards, which have reduced the effort on five projects so far. As part of Pfizer's corporate standards, these vendors have created a pool of resources for Pfizer from which they can tap for future projects.

### **Project Management**

Project management centered on Delivery Team framework, which was based upon the GDM (Global Data Management) operating model for each of the various lines-of-business, as described earlier. During the FDWC project, the Delivery Team for Finance & Procurement consisted of five members, shown in Figure 4.

## Managing Business Complexity at Pfizer

The Delivery Team oversees the project team who performs the actual work. Pfizer has a strategy of hiring outside contractors on a project basis and funding each project by the various business units while using the core team of full-time employees to manage contractors and projects. The expectation is that the use of outside specialists would support the exponential increase in projects since 2009, while managing costs.

Most members of the project team are specialists employed by outside contractors,

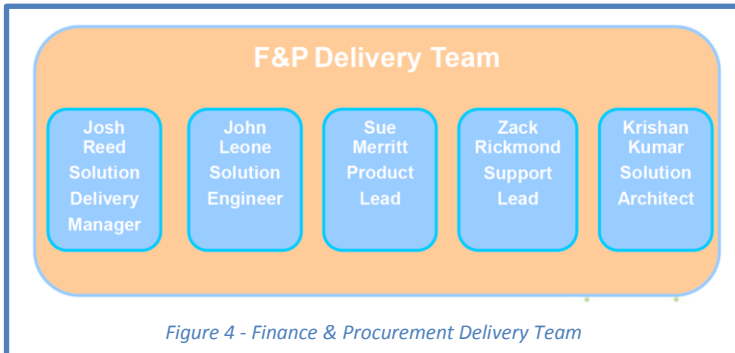


Figure 4 - Finance & Procurement Delivery Team

such as Deloitte. Pfizer has organized its IT development around the use of outside contractors, such as Deloitte<sup>12</sup>, PricewaterhouseCoopers<sup>13</sup> (PWC) and others. There are 29 full-time Pfizer employees who are managing the work of over 230 consultants. In addition, support for sustaining production applications is from PWC and Cognizant<sup>14</sup>. For Pfizer, this approach has been flexible and scalable.

### **Data Integration as Critical Part of Solution**

Central to the FDWC project was the massive data integration effort required to provide the data warehouse with the timely and accurate data. The FDWC project needed to catalog all the existing data and its structure. This data needed to be rationalized in business terms, so that, for instance, this column in this table means a specific type of revenue. Then, the new database needs to be designed incorporating the relevant parts of the old, eliminating redundancies and consolidating common data. In other words, the team must understand what each data item means to the business. To persuade dozens of powerful business units to agree on one data element's meaning and use requires extensive analysis of the data's lineage, contents, and usage across the company. It takes patience and acute

attention to detail, along with solid data modeling skills and knowledge of the business.

The phrase "lift & shift" is sometimes used to describe the work required to migrate a data warehouse from one database platform to another. Krishan admitted that the phrase implies that this effort is trivial, when in fact it is not. Conceptually this migration from one database platform to another does seem simple. Both platforms are using similar SQL statements. Both platforms structure tables and indexes in similar ways. Both platforms have similar view definitions. However, the devil is in the details!

You need a thorough analysis of the data, its meaning and its usage. You need to catalog and rationalize the data for its new context. Some of the data may be not useful and should not be migrated. Other data should be consolidated and shared. As Krishan Kumar summarized, "A simple Lift & Shift approach completely falls apart, and you end up with garbage." Clearly the "migrate it but don't change anything" approach is not possible, although some business managers continue to believe it maybe be possible.

The following Informatica capabilities have been adopted to leverage the joint Teradata-Informatica environment for data integration:

- Using Informatica's Pushdown Optimization tailored to Teradata. Increase the use of CPU power within Teradata for transformation processing
- Leverage Teradata Parallel Transporter to integrate Informatica and Teradata environments. Optimize the usage of Teradata utilities with the Informatica Power Exchange interface
- Support the mission-critical operation with Informatica Enterprise Grid and High Availability options. Apply systematic method/procedure to restore/recover from failures without any re-work
- Use Team-based development to ensure version control and secure deployment from QA to production. Ensure design consistency and code compliance across the in-house resources and contractors
- Optimization in error-handling techniques. Allowing constraints be caught at the loaders and database layer at Teradata

An innovative practice was the use of Informatica Pushdown Optimization (PDO), which pushes down into the database engines the processing of transformation, rather than performing those transforms on a separate server. Hence, PDO leverages the MPP capabilities of Teradata to improve performance for the Extract-Load-Transformation (ELT) processing. It starts with

## Managing Business Complexity at Pfizer

the transfer of data from the source databases (such as the financial data warehouses) into a staging area on the Teradata platform. Under direction of the PowerCenter server, the transfer performs only light transformations on the data to maximize the transfer rate.

PowerCenter utilizes the Teradata Parallel Loader (TPT) so that multiple high-speed streams of data loads can be performed concurrently. TPT generally provides a 20% improvement over the traditional stand-alone load utilities. Once loaded into the staging area, PowerCenter can then pass SQL-based transformations to be performed by the Teradata database engine in parallel. Transformed staging data is then applied to the actual base tables. Because PowerCenter's pushdown leverages the parallel database, runtime for many ELT jobs is 2X to 10X faster than a traditional ETL job.

### Benefits Realized

The FDWC project will be in production over the coming months. Even so, cost reductions are already emerging from building a shared platform that implements a common solution patterns and frameworks across multiple projects. Here are some specific examples of these cost reductions:

#### Technology Standardization

By standardizing on Teradata, Informatica, and SAP BusinessObjects (for reporting), Pfizer has adopted common corporate architectural standards, which have reduced the development effort on five projects so far. In particular, a tangible benefit has been the reduction in software licensing fees, which previously were paid to dozens of vendors. The effort of troubleshooting a problem has been reduced with a small set of vendors. And, the consolidation of platforms has resulted in smaller hardware footprint in the data centers and simplified maintenance. Finally, these vendors have created a pool of resources from which Pfizer can tap for future projects.

#### Data Scalability

With the linear scalability in processing data volumes by Teradata and Informatica, Pfizer has avoided future technology investments

that could be on the order of \$10M every few years. Results from performance testing confirmed the scalability of the MPP system.

Figure 5 shows the duration required to generate typical daily reports. The first column

Report Name	Single User Legacy	Single User MPP	100 Users MPP	Perf. Imp.
Revenue by Product	209	13	34	6.1x
Revenue by Product	n/a	12	32	
Expenses by Account	132	30	87	1.5x
Commitment by Product	190	25	73	2.6x
Commitment by Project	211	28	71	3.0x
Detailed Transactions	290	45	28	10.4x
Standard P&L	n/a	21	47	

Figure 5 - Report Processing (seconds)

indicates the type of report, while the second column indicates the duration for the legacy system to generate the reports. The next two columns show the duration for a single user and for 100 concurrent users. Testing initiated that number of report instance at the same instance, so that they were all processing concurrently. The time was the duration of the last report to complete. The last column is the performance improvement of the new MPP system for 100 concurrent users over the legacy system for a single user.

Note that, in general, reports are now taking 30-90 seconds to complete, while the old legacy system required 3-5 minutes. In particular, the detailed transactions report went from 5 minutes to less than 30 seconds, implying that users will more likely to request detail financial information in the future.

Figure 6 shows the duration required to complete the nightly ETL processing. There are three regions worldwide that load data into the MIC data warehouse, plus an additional 30 minutes for indexing and aggregation (for only the MPP system). The legacy system required that each region be processed sequentially so that the three durations are added for the total. The MPP system can process the regions concurrently so that the longest duration (188 minutes for Europe) plus index refresh determines the total duration.

## Managing Business Complexity at Pfizer

Note that the MPP system now take 3-4 hours per night, while the legacy system previously took almost 14 hours. These dramatic performance improvements in load times have

Region	Legacy Sequential	MPP Concurrent
Europe	369	188
Americas	245	116
Asian Pacific	202	139
Index Refresh	-	30
Total Time	816	218
Performance Improvement		3.7x

*Figure 6 - Nightly ETL Processing (minutes)*

enabled Pfizer to provide financial management reporting capability in each region based on their local business hours, which was not possible with the legacy system.

### Responsive Service Levels Globally

With the higher performance of the MPP platform, Pfizer has the capacity to load and deliver timely data globally, so that fresh reports are available to business users before they arrive at their office, regardless of their local time-zone. Previously, service levels were set regionally by maintaining three separate instances of the data warehouse for the Management Information Center.

## Informatica Usage across Pfizer

Pfizer has been deploying the Informatica platform since 2001. Pfizer started its partnership with Informatica for the deployment of the data integration technology in the R&D data warehouse. For the last 10 years since its first deployment, Pfizer has expanded its implementations for the following areas:

- Research information factory – expanded version of the initial R&D data warehouse
- Pharmeco Risk and Safety data warehouse under Chief medical officer for drug safety
- Global manufacturing data warehouse
- Japan sales/distribution data warehouse
- clinical development information factory
- Legal group product quality assurance DW (PQA)
- Financial data warehouse
- Worldwide biopharmaceutical business unit - CDW

Pfizer chose Informatica to be the enterprise standard and to run its mission critical operations. Informatica is the basis for its GDM shared service practice because of the maturity and the reliability of data integration platform. Currently, Pfizer is actively integrating its critical information assets from the Wyeth acquisition on Informatica's enterprise data integration platform to ensure that they can continue to do more with less and maintain its leadership in the biopharmaceutical industry.

### Support cost

Ongoing support cost has been reduced by 25% to 30% especially in the staffing of database administration and application-level support/services (monitoring user problem and performing minor enhancements). The support costs were reduced from \$960K to \$860K per year (or 10%) for the Financial Data Warehouse and from \$2.3M to \$1.6M per year (or 30%) for the Management Information Center.

### Overall Project Saving

The FDWC project is generating an estimated savings of several millions of dollars per year, with an investment of under ten million dollars, composed of 80% in capital and 20% in one-time costs.<sup>15</sup>

Kevin Fleet summarized the benefits, "So far so good [in realizing reduced support costs], but the real value will come once these systems have been implemented, tuned, and ready to respond to the next big business challenge."

## Lessons Learned

Through the interviews, we asked about the lessons that were learned during the course of the project.

### Success Factors

Krishan Kumar shared his view of the key success factors, as...

*Definition of 'success' has changed over time. Finance One requirements are the focus now. Success is really to consolidate the silos of multiple financial solutions to a common platform that will result in year-over-year saving for support, resulting in reducing the physical footprint of boxes. Then, with current budget and time available with saving realistic for time period.*

### Do Technology and Organizational Changes Separately

In the discussion with Josh Raysman, he suggested that an IT project should not attempt major technology changes at the same time as a major organizational change:

*There is a confluence of initiatives embodied in the FDWC project. The original project was a change in the technical infrastructure, which is the easy part. We are now supporting a change to the business model for the corporation. Thus, the FDWC project is attempting three major changes concurrently: organizational change, business process change,*

## Managing Business Complexity at Pfizer

*and technology change. Do the platform migration first, and then do the organizational change later.*

### **Transitions to new Technology is Hard**

Do not underestimate the amount of change in thinking required in system transitions from an Oracle environment to a Teradata one. There is a different way of thinking about basic database parameters and operations. Kevin Fleet mentioned, "We had 'bumps'. And, it took some time to get the team up to speed. We under-estimated the time required." While Teradata provides assorted tools and experts to help with the migration, it is not unlike moving from one house to another – some things don't fit, some things must be discarded, and some things must change to exploit new capabilities.

## Synthesis

---

To summarize this case study and generalize the lessons learned, here are the challenges that the FDWC project faced and resolved.

***Data integration is tough! However, data integration has huge business value.***

### **Data Integration Is Tough**

Data integration is tough! However, data integration has huge business value. Pfizer chose a stepwise approach balancing payoffs in business value with constraints in technical resources. The approach is to consolidate each domain into a common environment, then rationalize each domain, and finally integrate the common elements into a single enterprise-wide view of the business. This is a long-term incremental strategy that has a high probability of success and has flexibility to respond to unexpected industry changes.

### **Business Changes + Technology Changes**

Any project that makes major business changes and technology changes at the same time is a risky project. But, incurring this risk is sometimes necessary for the business. Further, the FDWC project with its dual business/technology changes was a key catalysis for the paradigm shift in architecting data warehouse environments for Pfizer.

### **More than a Data Tsunami**

Data tsunami is too simple of analogy. It is the complexity of the business that consumes IT resources. That complexity can come in the form of increased data volume, but that is easy problem. Complexity from inconsistent business procedures is the killer, especially across a large global corporation. Kevin Fleet remarked, "It is not just volume. There are different types...or classes...of data that are so difficult to integrate."

### **Outsourcing Infrastructure Support**

Pfizer outsourced their infrastructure support, resulting in technical resources consisting of "lowest common denominator" skill sets. The benefit was that cost reductions from the operations level lead to a fundamental change in the architectural solution framework for applications.

### **Strategic Technology Partners**

Strategic technology partners (such as Informatica and Teradata) proved to be essential in providing 'good' IT for Pfizer. These partners understand the unique characteristics of Pfizer's environment (both organizational and technically) and institutionalize that knowledge in applying the proper resources to projects and in aligning their efforts to Pfizer's long-term objectives. The partners are more likely to drive toward long-term "win-win" situation rather than short-term "transaction-by-transaction" situation.

### **Organizational Autonomy and Centralized IT Services**

There is a natural tension between organizational autonomy and centralized IT services. To be a healthy tension, Kevin Fleet pointed out, "It is a tough transition into a shared resource model. And, the governance of shared assets is the key to its success."

### References

---

- <sup>1</sup> Paraphrased from the Financial Data Warehouse MPP Migration, Request for Proposal, February 3, 2009.
- <sup>2</sup> Wall Street Journal company profile about Pfizer Inc.
- <sup>3</sup> <http://media.pfizer.com/files/annualreport/2007/annual/review2007.pdf>
- <sup>4</sup> Annual Review 2009 at [http://www.pfizer.com/investors/financial\\_reports/annual\\_reports/2009/](http://www.pfizer.com/investors/financial_reports/annual_reports/2009/).
- <sup>5</sup> Taken from interview with Josh Raysman on July 21, 2010.
- <sup>6</sup> Financial Data Warehouse MPP Migration, Request for Proposal, February 3, 2009.
- <sup>7</sup> High-Throughput Screening allows testing of millions of compound reactions per day at a millionth of the cost of conventional testing methods. [http://en.wikipedia.org/wiki/High-throughput\\_screening](http://en.wikipedia.org/wiki/High-throughput_screening)
- <sup>8</sup> Symmetric multiprocessing with multiple processors sharing the same main memory. [http://en.wikipedia.org/wiki/Symmetric\\_multiprocessing](http://en.wikipedia.org/wiki/Symmetric_multiprocessing)
- <sup>9</sup> Financial Data Warehouse MPP Migration, Request for Proposal, February 3, 2009.
- <sup>10</sup> Financial Data Warehouse MPP Migration, Request for Proposal, February 3, 2009.
- <sup>11</sup> Teradata Technical Overview, Global Data Management, May 20, 2010.
- <sup>12</sup> <http://www.deloitte.com/>
- <sup>13</sup> <http://www.pwc.com/>
- <sup>14</sup> <http://www.cognizant.com/>
- <sup>15</sup> Teradata Technical Overview, Global Data Management, May 20, 2010.

### About the Methodology

---

The objective of this case study is education—to share insights with other IT professionals so that we can mature as an industry, amid escalating business challenges and rapidly evolving technology. An on-site visit evolving a full day of interviews was conducted to adequately document the details of this case. Prior to the on-site visit, we had several telephone discussions that narrowed the scope to a specific IT project with its business requirements, timeline, resources, and results. As synthesized from the on-site interviews, several drafts were circulated for review. When the participants were satisfied with its contents, the document was submitted for publication approval by the company. All discussions and collected materials were considered confidential until the company had approved the case study for publication.

**Richard Hackathorn** of Bolder Technology and **Julianna DeLuca** of Informatica conducted the interviews. **Dan Graham** of Teradata also reviewed the drafts. The persons interviewed at Pfizer were:

- **Kevin Fleet**, Executive Director, Global Data Management: He was the sponsor for this case study and provide valuable context to the case.
- **Krishan Kumar**, Associate Director, Business Intelligence & Data Warehousing: He was our primary contact and was invaluable in explaining the details behind the project.
- **Josh Raysman**, Director, Solution Architecture and Engineering, GDM, BT Group: He was the IT sponsor for the FDWC project, acting as mentor/coach to many of the team members.
- **John Leone**, Solution Engineer: He was the technical lead for the data warehouse and ETL flows and managed the deliverables for the project. In addition, he is responsible for the unit and system testing for the project.
- **Sil Crino**, Financial System Architect with Deloitte: As an outside contractor, he provided keen insights into the financial complexity of a global corporation and the needed IT requirements. He also shared his expectations about the impacts upon business users from the FDWC project.

### About Bolder Technology Inc.

---



Bolder Technology is a twenty year old consultancy focused on Business Intelligence and Data Warehousing. The founder and president is Dr. Richard Hackathorn, who has over thirty years of experience in the Information Technology industry as a well-known industry analyst, technology innovator, and international educator

Richard was a member of Codd & Date Associates and Database Associates, early pioneers in relational database management systems. In 1982, he founded MicroDecisionware Inc. (MDI), an early vendor of database connectivity products, growing the company to 180 employees and was acquired by Sybase, now part of SAP, in 1994. He has published numerous articles in DM Review and BeyeNETWORK. He is a member of the IBM Gold Consultants and the Boulder BI Brain Trust. He has written three books and has been a professor at the Wharton School and the University of Colorado. He received his degrees from the California Institute of Technology and the University of California, Irvine.

### About the Sponsor

---



Informatica Corporation is the world's number one independent provider of data integration software. Organizations around the world gain a competitive advantage in today's global information economy with timely, relevant and trustworthy data for their top business imperatives. More than 4061 enterprises worldwide rely on Informatica to access, integrate and trust their information assets held in the traditional enterprise, off premise and in the Cloud.

Informatica and the Informatica logo are registered trademarks of Informatica Corporation and/or its affiliates in the U.S. and worldwide. Teradata and the Teradata logo are registered trademarks of Teradata Corporation and/or its affiliates in the U.S. and worldwide.

© Bolder Technology, Inc. 2010. All rights reserved. No portion of this report may be reproduced or stored in any form without prior written permission.