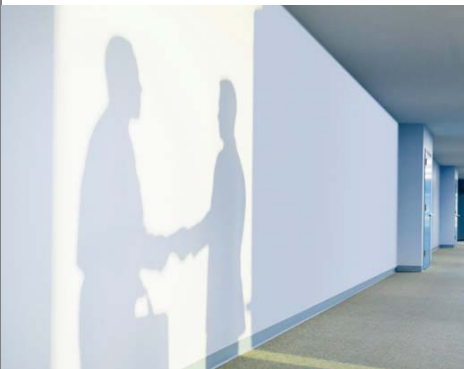


Powering Performance:

*LinkShare Achieves Massive Data Throughput in
Distributed Grid Architecture with Informatica PowerCenter*

CASE STUDY



This document contains Confidential, Proprietary, and Trade Secret Information (“Confidential Information”) of Informatica Corporation and may not be copied, distributed, duplicated, or otherwise reproduced in any manner without the prior written consent of Informatica.

While every attempt has been made to ensure that the information in this document is accurate and complete, some typographical errors or technical inaccuracies may exist. Informatica does not accept responsibility for any kind of loss resulting from the use of information contained in this document. The information contained in this document is subject to change without notice.

The incorporation of the product attributes discussed in these materials into any release or upgrade of any Informatica software product—as well as the timing of any such release or upgrade—is at the sole discretion of Informatica.

Protected by one or more of the following U.S. Patents: 6,032,158; 5,794,246; 6,014,670; 6,339,775; 6,044,374; 6,208,990; 6,208,990; 6,850,947; 6,895,471; or by the following pending U.S. Patents: 09/644,280; 10/966,046; 10/727,700.

This edition published December 2007

To fuel the long-term growth of the largest online pay-per action marketing network, LinkShare™ needed an infrastructure powerful and scalable enough to handle more advertisers, more publishers, more data, and more combinations and types of information. This online marketing services pioneer relies on Informatica® PowerCenter® for not only data warehousing, but to run the business. “Informatica PowerCenter is a mission-critical platform for our day-to-day operations,” said Dave Ramos, Director of Business Intelligence and Analytics at LinkShare.

Background

Eleven Million Partnerships in 11 Years

For LinkShare, a subsidiary of Rakuten Inc. (NASDAQ [4755]), success breeds success—and more data, more services, more advertisers, more publishers, and more business. Founded in 1996, the company pioneered a concept known as online affiliate marketing and provides a distribution network and ubiquitous platform for tracking and creating performance marketing and sales partnerships on the Web.

LinkShare’s comprehensive technology platform and services make it the premier affiliate marketing, search and lead generation network on the Internet. Headquartered in New York City with offices in San Francisco, Chicago, Tampa, London, and Tokyo, LinkShare has created over 11 million partnerships between more than 800 Fortune 500 and prominent companies, such as J.C. Penney, Wal-Mart and American Express, and more than 1 million publisher Web sites, such as UPromise, Ebates.com, and About.com.

BENEFITS

- Increased profitability and market share
- Reduced operating costs and improve efficiency
- Achieved cost-effective scalability in a distributed grid architecture
- Ensured future success with high performance and throughput to accommodate growing data volumes
- Enhanced reliability with high availability and automated failover
- Implemented new data-driven Web services to grow business

As Business and Data Grows, So Does the Demand for Performance

Information is LinkShare's stock in trade, and business intelligence is the core of the company's business model. The ability to leverage this information in new, high-value ways is the reason LinkShare's clients remain in the network. It's also the reason new clients join the LinkShare network, so it's key to the company's long-term competitive advantage.

Since 2001, LinkShare has relied on the Informatica PowerCenter platform for a scalable solution to deliver insight and intelligence back to advertisers and publishers regarding their marketing campaigns down to the transaction level. As business expanded, the data volumes generated by LinkShare's network grew to enormous scope and scale. Captured from more than 400 disparate sources, the company processes over 300 million aggregate transactions a day, of which 100 million are raw transactions. These transactions include:

- Publisher registrations
- Advertiser registrations
- Product inventory uploads of more than 150 million unique products from more than 200 advertisers
- Web creative inventory of more than 100 million banner images, text links, etc.
- Advertiser offers, including product groupings
- Millions of publisher/advertiser partnership applications
- Consumer interaction, including image displays and click tracking
- Advertiser order transactions
- Advertiser invoice line items and publisher commissions payments
- Feedback from outside vendors and "closed loop" reporting

Data resulting from these transactions is transformed by PowerCenter, staged in a near real-time operational data store (ODS) and extracted again to an enterprise data warehouse (EDW). Since LinkShare is a 24/7 global operation and advertiser customers can submit data feeds at anytime, processing large daily bulk loads has never been a feasible option. Instead, LinkShare uses PowerCenter grid architecture to load balance over 700 sessions that run in 5-minute batch windows on a 24/7 basis with up to 400 concurrent sessions.

With strict service-level agreements (SLAs) in place with customers, it was vital that LinkShare achieve reliable, high-performance processing of large data volumes to ensure customer satisfaction and to avoid potential revenue loss.

Through 2006, LinkShare had relied on PowerCenter 7.1 to meet its multiple high performance data integration needs. When Informatica announced the release of PowerCenter version 8.1 in 2006, LinkShare Technology executives took notice. As the most significant PowerCenter upgrade in Informatica's history, PowerCenter 8.1 offered breakthrough features and functionality that LinkShare officials envisioned would help the company build even greater performance, scalability, reliability, and growth headroom into its solution:

- Enhanced grid computing support, including multinode load balancing
- High availability with automated failover
- Dynamic and source partitioning to break down and speed processing in parallel environments
- Changed data capture for real time processing
- Linux integration in a 64-bit architecture
- Support for leading Web services protocols

As a longtime Informatica customer, LinkShare had confidence that PowerCenter 8.1 would live up to its billing. “You always want bigger, better, faster, and Informatica’s definitely been able to provide that for us,” said Ramos. “Without Informatica, we wouldn’t be successful—it’s as simple as that. It would be like trying to dig out from an avalanche with a shovel.”

Solution

Grid-Based Data Integration Delivers High Performance, Low Cost

In 2007, LinkShare migrated its entire production environment, which includes more than 500 mappings, 700 sessions, and 100 workflows, from PowerCenter version 7.1 to 8.1 with the PowerCenter Enterprise Grid Option. The upgrade was executed in tandem with a transition to newer versions of 64-bit IBM hardware with Intel dual-core 3.0 GHz processors and the introduction of an EMC storage area network (SAN), in part to supply shared storage for grid-based data integration.

Built on commodity x86 hardware and Red Hat’s Linux distribution, the grid architecture is engineered to deliver greater price/performance than might have been achieved in a traditional symmetric multiprocessing environment. The solution is comprised of:

- IBM DB2 Universal Database v9.1 ESE Partition Edition (EEE) on Suse Linux 9 underpinning the EDW and ODS
- IBM eServer Cluster1350, including:
 - eServer 3650: 64-bit x 18-node cluster hosting IBM DB2 v9.1 UDB
 - eServer 3650: 64-bit x 6-node cluster hosting Informatica PowerCenter v8.1
 - eServer 3650: 64-bit x 4-node cluster hosting Informatica PowerCenter v8.1
 - Node configuration:
 - Processors: 2 x 3GHz Dual-Core Xeon Woodcrest 1333 Mhz
 - Memory: 8 x 2GB PC2-5300 667Mhz ECC DDR2 FBDIMM
 - Disk controllers: 2 x Mega Raid 8840 SAS(2 ports x 4G bps HBA – DB2 Nodes Only)
 - Disk drives: 2 x 146GB 15Krpm Ultra320 SAS HDD
 - Ethernet: 1G bps
 - IBM storage: EXP3000 (36) 12 X 146GB 15Krpm Ultra SCSI HS HDD (2xH+P – DB2 Nodes Only)
- IBM System P for Siebel Business Analytics engine and WebSphere
 - System P P590:
 - Processors: 16 x 1.65GHz Power 5
 - Memory: 64 Gb ECC
 - Storage: 24 x 73GB 15Krpm Ultra320 SCSI HS HDD

LINKSHARE'S SOLUTION EVALUATION CRITERIA

Why PowerCenter data integration platform?

- Scalability on grid: native clustering
- Performance: parallel operation, real time, partitioning
- High availability with automated failover
- Easy administration and management
- Rapid development environment
- Product vision, support, open architecture (APIs) and SDKs

Why IBM database?

- Scalability: native clustering, data clustering
- Performance: multidimensional clustering (index), comprehensive aggregate functionality
- Security: product vision, support, open APIs and SDKs

Why Siebel Business Analytics engine?

- Scalability: native clustering
- Performance: comprehensive caching capabilities, multipass SQL and parallel operation, aggregate awareness
- Support for very large dimensions (10 million rows)
- Zero-footprint Web client
- Security
- Product vision, support, open APIs and SDKs

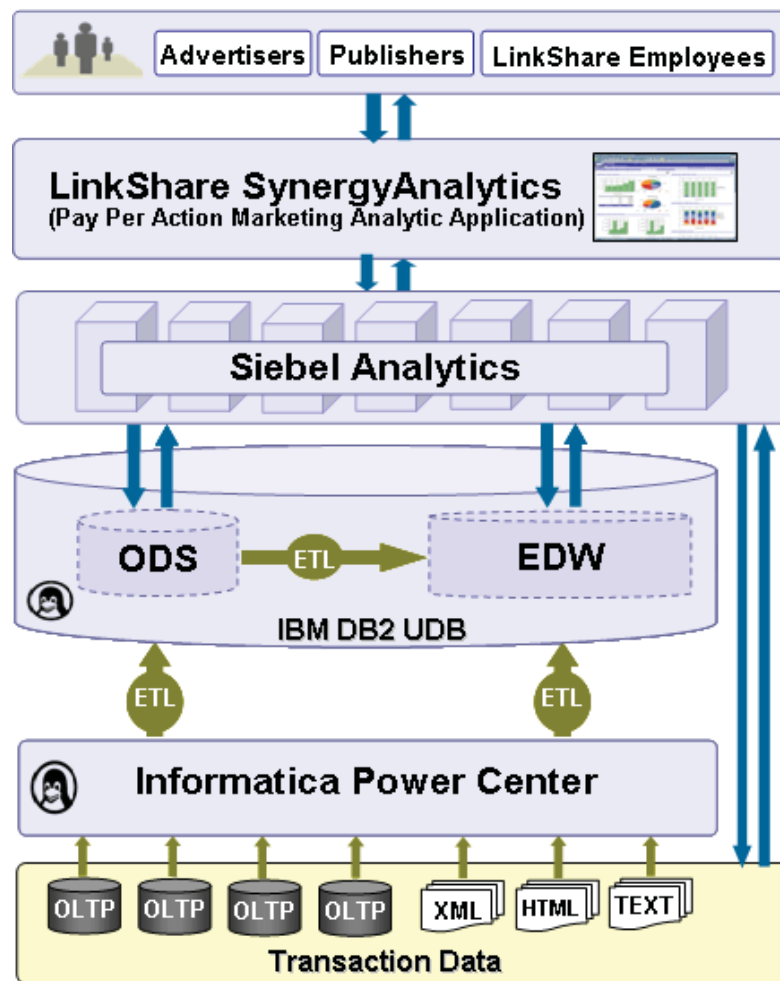


Figure 1. LinkShare's business intelligence solution architecture. Informatica, IBM, and Siebel Provide the Core Engines for Managing, Monitoring, and Measuring the Activity and Success of the LinkShare Network.

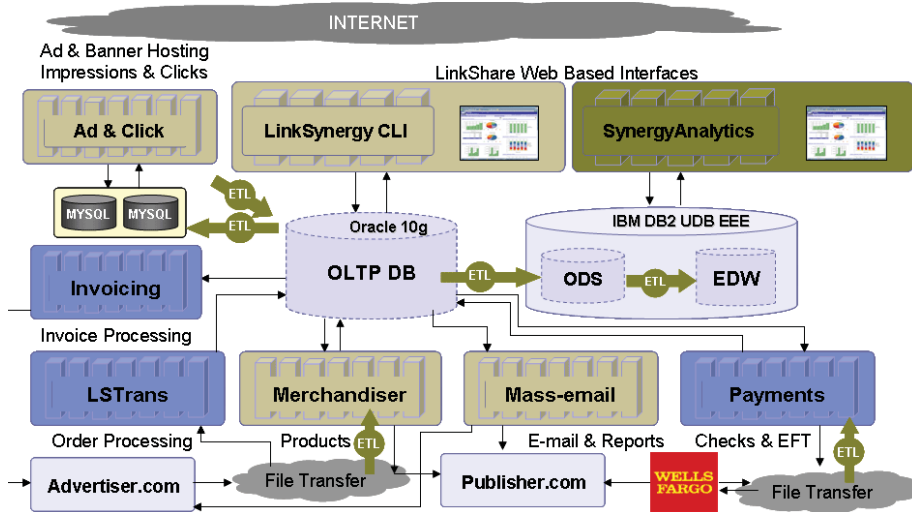


Figure 2. How PowerCenter provides the extract, transform, and load (ETL) processes that drive interaction between the various LinkShare systems. PowerCenter Drives the Data Integration Processes that Drive Interaction Between LinkShare Systems

Re-Engineering the Grid for Greater Throughput and High Availability

Taking advantage of enhanced grid computing support in PowerCenter 8.1, LinkShare re-engineered what had been an eight-node grid in a single cluster in pursuit of greater performance and cost-effective scalability. Its IT organization implemented a two-cluster, 10-node grid, with one cluster comprising six IBM eServer 3650s with an EMC Clarion CX3-20 SAN shared disk array to support PowerCenter grid features such as shared cache and session on grid. The second cluster comprises four IBM eServer 3650s with an EMC Clarion CX3-20. The two clusters use Global File System and are located in two separate data centers connected via Metro-Area Ethernet.

The two clusters have been designed to have separate primary load responsibilities: one for OLAP loads, and the other for OLTP data integration loads. But the clusters are also designed to enable LinkShare the ability to seamlessly migrate all loads to a single cluster in the event of data center outage or maintenance needs. LinkShare is now exploiting automated load balancing in PowerCenter 8.1 to optimize performance based on available memory, processor usage, disk speed, partitioning configuration, and other resource utilization characteristics.

In the previous iteration, LinkShare administrators had to manually configure load balancing, monitor resource utilization, and adapt on the fly. “We had been doing load balancing ‘by guess and by golly’ to figure out which process should go off to another server, and this makes it much, much more automated and efficient and improves performance substantially,” Ramos said.

LinkShare is also making the most of improved and automated high availability and failover features in PowerCenter 8.1 by applying them both inter- and intra-cluster. In other words, if a node fails within the OLTP cluster, the data services running on it will fail over to available nodes within that cluster, or if the OLTP cluster were to fail, the data services running on it would fail over to the ODS/EDW cluster.

“That gives us very high redundancy with automated failover—that’s definitely a big step,” Ramos said. “We had failover before in the sense that we’d have to manually go into the system and reassign sessions to another node. Now along with the SAN shared storage, if a node goes down, all those jobs fail over to an available node without any intervention.” This high availability functionality is critical in helping LinkShare meet its SLA obligations and in avoiding revenue loss.

Boosting Grid Performance with PowerCenter Partitioning

Dynamic and source partitioning and session-on-grid are other PowerCenter features that are helping LinkShare achieve high performance and massive throughput. Though LinkShare had, in the past, used database partitioning on the target side, PowerCenter-driven partitioning on data sources give performance an additional boost. Session-on-grid and dynamic partitioning enable session processing to be distributed across multiple nodes and executed in parallel for even greater throughput—a key consideration as the number of concurrent sessions executed can exceed 400.

“We are able to really, really increase our throughput substantially through the combination of load balancing, source partitioning, dynamic partitioning, session on grid—all in a 64-bit architecture,” Ramos said. LinkShare is currently monitoring the load behavior on the new clusters to identify potential candidates for tuning but has already experienced a tenfold increase in session performance by migrating to the new cluster.

“Performance and scalability are mission-critical to LinkShare,” Ramos said. “We’re growing in markets internationally in Japan and in Europe. Along with data growth, we’re building out additional systems that need to support transactional processing overseas. So the processing power we get from Informatica is very, very important.”

Ramos and other LinkShare officials were also impressed with the breadth and depth of grid functionality incorporated into PowerCenter 8.1. For instance, a feature called resource reservation enables administrators to give higher priority to vital data integration services, as well as to specify thresholds to help avoid overloading nodes and impacting performance.

As shown in Figure 3, PowerCenter runs in two separate clusters across 10 IBM eServer 3650 cluster nodes with two EMC Clarion CX3-20s. It's important to note that PowerCenter helps LinkShare keep costs down by distributing 20 CPUs (40 cores) of processing power across 10 physical servers.

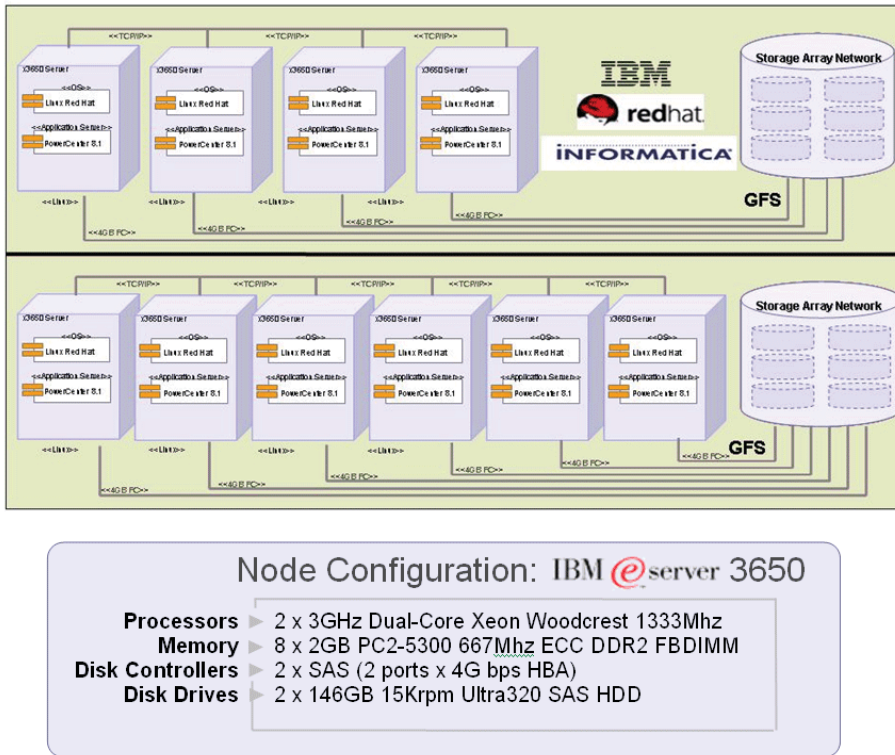


Figure 3. PowerCenter's Logical Architecture Helps LinkShare Support a Cost-Effective Grid Environment.

The platform provides built-in parallel performance to handle high volumes of complex tasks (for example, data transformations, mappings) in near real time, supporting the on-demand nature of LinkShare's environment. PowerCenter's architecture, which avoids hard connections between logical design and physical implementation, prevents mappings from having to be changed between different environments.

Increasing Real-Time Loading Efficiencies

LinkShare has 10 years' worth of data totaling 6 TB and warehouse capacity for 40 TB. With an average row size of 360 bytes, the company has about 9.77 billion rows and an average daily load of about 15 GB.

LinkShare has a 36-CPU (72 cores) IBM BCU DB2 EEE cluster, which is the main target for the data warehouse and is used for the operational data store database. The company uses a combination of loading techniques including daily bulk load, hourly batch, and near real time batch loading (10-, 5-, and 2 minute batch windows). The biggest loads move click and impressions from the source, aggregates them, and then moves the data to the OLTP application (Oracle) as well as to an ODS and an EDW. The ODS and EDW are shared DB2 database instances. Both are loaded in near real time using continuous sessions. The data must be

available in the EDW within 60 minutes, so PowerCenter runs in a continuous 10 minute batch loops (for example, there is a 60-minute reporting service level agreement, or SLA, back to the advertisers and publishers). The data integration between OLTP systems is loaded in near real time using continuous sessions. The data must be available in these systems within 10 minutes; therefore, PowerCenter runs in a continuous 2-minute batch loops. LinkShare uses incremental aggregation and some key range partitioning to accomplish the transformation requirements and to load to the target database in the most efficient manner.

Results

An Extensible Foundation for New Services and New Business

LinkShare's upgrade to PowerCenter 8.1 with the Enterprise Grid Option has helped it fortify its position as a global leader in performance-based Web marketing solutions. The unmatched data integration scalability, performance, and reliability that LinkShare has been able to realize by its innovative and resourceful use of Informatica technology is a distinct competitive advantage that translates into more business and greater revenue.

It has also supplied LinkShare with an extensive foundation upon which to build out new data-driven Web services. For instance, LinkShare plans to roll out data services by which it can supply to customers the raw transactional Web tracking data on consumer Internet activity and navigation patterns, as well as exchange information with partners. With the raw data, customers and partners will be able to do their own reporting and analysis.

Instrumental in this initiative is PowerCenter and its support for leading Web services protocols and changed-data capture technology that will enable LinkShare to supply only data updated since the last exchange to customers, thereby accelerating its delivery and minimizing operational impact on both LinkShare and customer systems. "Being able to provide a Web-based platform to deliver transactional data to our partners is a huge advantage," Ramos said. "The alternative of data files and the ftp process is slow and asynchronous. It wouldn't support our needs to be able to roll out a new product, meet customer demands, and grow our business."

Cost-Effective Ease of Use and Reusability Across Multiple Projects

Now featuring an improved, single-interface Web-based administration console, PowerCenter 8.1 provides powerful, streamlined management capabilities essential to LinkShare's ability to effectively administer its diverse data integration environment.

The PowerCenter platform allows LinkShare to extract, transform, and integrate data from practically any data source in a codeless point-and-click development environment. PowerCenter extracts data from more than 400 disparate sources, including OLTP systems and XML, HTML, and text files. It also allows LinkShare to reuse objects across multiple projects, drastically reducing the time and cost otherwise required to developing mappings from scratch, thereby increasing the company's ability to address new data demands. As a result, LinkShare is able bring new advertisers and publishers online more rapidly to accelerate time to revenue.

The Informatica-based solution's ability to deliver data in a functionally rich, zero-footprint Web client enables LinkShare to provide real-time insight and intelligence about marketing programs back to more than 1 million customers. More than just a competitive advantage, 24/7 global, secure access to this data is integral to the company's business model. With LinkShare's pay per action network, the company's clientele can optimize their performance and thus maximize revenues to publishers, to advertisers, and to LinkShare.

A Key Component to LinkShare's Long-Term Competitive Advantage

With PowerCenter, LinkShare has received significant and measurable results directly related to its ability to scale and expand the utilization of critical business information by more people. In terms of increased use of existing information assets, the payback has been significant. This solution is viewed as such a competitive advantage that specific revenue increases and cost savings are highly confidential. It's safe to say, however, that the information insight and analysis that the Informatica solution provides is the foundation for the company's ability to grow and serve its customers.

"The value of a LinkShare affiliate marketing, search or lead generation program to drive commerce online continues to attract leading companies," said Ramos. "With hundreds of thousands of affiliate partners in the LinkShare Network, our clients have access to one of the largest distribution channels online to reach their end consumers. PowerCenter has been indispensable in helping us grow and continue to able to offer both advertisers and publishers opportunities to innovate and develop unique ways to drive greater revenue."



Worldwide Headquarters, 100 Cardinal Way, Redwood City, CA 94063, USA
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871 www.informatica.com

Informatica Offices Around The Globe: Australia · Belgium · Canada · China · France · Germany · Japan · Korea · the Netherlands · Singapore · Switzerland · United Kingdom · USA

© 2008 Informatica Corporation. All rights reserved. Printed in the U.S.A. Informatica, the Informatica logo, and The Data Integration Company are trademarks or registered trademarks of Informatica Corporation in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.
First Published: 2007

6627 (09/19/2008)