IBM Big Data

A collection of Big Data client success stories

IBM

Table of Contents

A financial software company

Projecting USD10 million revenue lift with the IBM Netezza data warehouse appliance

Overview

The need

A financial software company sought to analyze customer engagements to improve product quality and increase retention. It also wanted to increase marketing return on investment and targeting precision using behavioral variables.

The solution

The company deployed an IBM Netezza data warehouse appliance, which enables it to perform next-generation analytics in order to consistently and continuously improve its service.

The benefit

Higher customer satisfaction scores; projected USD10 million revenue lift within 12 months; ability to detect and correct cross-channel cannibalization.

A financial software company owes its success to high-quality products, innovative distribution channels and effective multichannel advertising programs. The company was an early adopter of the software as a service (SaaS) delivery model, while also reaching customers through promotions, retail channels and call centers. But credit is also due to the way it uses granular user behavior data to provide a more personalized and relevant customer experience.

This financial software company uses the Net Promoter Score, a popular methodology for gauging the loyalty of customer relationships, to confirm its customers are satisfied by asking them a simple question: Would you recommend us to a friend? There is no way for any company to fake its way to a good score: It must provide first-rate service.

This financial software company employs next-generation analytics to consistently and continuously improve its service through:

- In-channel optimization
- Cross-channel/platform engagement measurement
- In-product behavioral analysis
- Predictive analytics
- Testing for impact

"The company didn't want the people building analytic applications worrying about data distribution while the algorithms were running. Netezza takes that worry away."

—Brad Terrell, Vice President and General Manager, Netezza and Big Data Platforms, IBM

Easier said than done

There are several obstacles to achieving success in each of the areas noted above. First, digital media data pours in with unprecedented speed. The company needs the ability to load huge volumes of data into its data warehouse—hundreds of millions of rows and 100 million clicks daily—and to run complex, high-speed queries.

This has been accompanied by an explosion in the number of variables business users want to analyze thereby making segmentation, classification and regression exponentially complex. Then there are the related tasks of running complex analytics and building predictive models, both the regression and classification varieties.

The company created a five-point program for using data to create a better customer experience:

- Intelligent tracking of user behavior
- A scalable analytic platform
- Predictive modeling
- Hypothesis testing and validation
- Intelligent intercept interaction

One major component that would enable these action items: technology that would give the company a single view of the customer, and at the same time allow it to analyze petabytes of data.

After evaluating several leading data warehouse and analytics platforms, the software company selected and deployed the IBM Netezza® data warehouse appliance, an architecture that can rapidly query large volumes of data and handle computational complexity with low total cost of ownership (TCO) and operational simplicity. The company promptly put the IBM Netezza data warehouse appliance to work on three critical projects:

1. Website analysis
2. In-product discovery
3. Analysis of Google DoubleClick advertising data

Solution components

- IBM Netezza® 1000

"The company projected USD10 million in new revenue within 12 months. That's a non-trivial lift."

—Brad Terrell

Website analysis

The goal of this project was to increase the company's customer satisfaction score for its products with thorough website analysis. This software company's customers have the ability to use its products online and submit payment when they are finished with a given project. But there's a high price that the company pays if it fails to satisfy the customer: Frustrated users abandon the product, leaving the site without submitting payment. Thus, the company has to anticipate behavior and offer help before the customer asks for it. That means flagging screens, screen clusters, application flows and issues with higher-than-usual abandonment rates.

The company collects behavioral data from multiple sources. But those tools don't enable staff to track and analyze individual web behaviors. Nor do they facilitate behavioral segmentation, time-series analysis, predictive modeling or scoring. Furthermore, the company wanted to unite its user behavior and demographics data to gain deeper insights through analysis.

The IBM Netezza data warehouse appliance allows the company to combine these varied digital media data types and data sources, while also offering access to in-database analytic functions. It can now look at user paths through the product or the website, identify places where customers have trouble or drop-off, and use that information to optimize website navigation.

This granular understanding of user behavior on its site has allowed the company to make small changes that enhance the user experience. As a result, the company has achieved a higher customer satisfaction score and increased subscription revenues.

"With a market sizing tool powered by the IBM Netezza data warehouse appliance, the company can target even more precisely based on behavioral variables: who saw or clicked on an ad, the amount of time they have in business and whether they use competitors' online services."

—Brad Terrell

In-product discovery

One of the software company's online products serves as the main channel for two critical revenue streams. Managing these revenue streams requires advanced analytics which provide detailed usage information for that product.

The company collects in-product discovery (IPD) data using custom instrumentation built into its software. Prior to using the IBM Netezza data warehouse appliance, it could only get data from one percent of users, and it had trouble analyzing the data it had because the average query took five hours to run.

As a result, the product team faced challenges when making roadmap decisions. They couldn't base offers on "an understanding of the important characteristics and behaviors of that target audience," says Brad Terrell, vice president and general manager for Netezza and Big Data Platforms at IBM. "Response rates were low, and they were not achieving their goals." The company also needed to filter outliers and spot erroneous data coming from fraudulent clicks and click bots.

Since deploying the IBM Netezza data warehouse appliance, the company has increased its sample size from one percent of its users to 10 percent while reducing the average query time to a fraction of what it took before. Moreover, the product team now has a much deeper understanding of how customers are using the product. And with a new targeting analysis tool fueled by the IBM Netezza data warehouse appliance, the company conducts monthly digital media cross-sell campaigns generating additional revenues for that product.

"The company projected USD10 million in new revenue within 12 months," says Terrell. "That's a non-trivial lift." And with a market sizing tool powered by the IBM Netezza data warehouse appliance, the company can target even more precisely based on behavioral variables: who saw or clicked on an ad, the amount of time they have in business and whether they use competitors' online services.

"People expect instant gratification—with good reason. Real-time access to data is very critical in many applications. So that puts a premium on many new types of analytics."

—Brad Terrell

Advertising data analysis

The software company is a major display advertiser, and that's likely to continue. Its ad spend includes display impressions, television, email, affiliate marketing, mobile and search advertising. And this leads to staggering online traffic. The company received 18 billion impressions in the last year, and recently collected 16 billion records in 15 days. But when there are so many clicks, it's difficult to analyze them all, and the clicks that DoubleClick records can't be sampled. The company had to determine through regression analysis which exposures led to conversion and the impact of those conversions on the bottom line.

In its incumbent environment, the company had trouble detecting and correcting cross-channel cannibalization. But with transaction-level analysis facilitated by the IBM Netezza data warehouse appliance, this is no longer a challenge. In one campaign, the company justified additional advertising spend after proving that cannibalization had not occurred where the company thought it had.

In another case, the company discovered that people had been served digital media ads after they had converted. The company stopped these post-conversion impressions, leading to substantial savings while preventing customers from getting annoyed.

The company uses the IBM Netezza data warehouse appliance to track the full customer lifecycle from exposure to conversion (E2C). The IBM Netezza data warehouse appliance helps this company field more effective campaigns and make better data-driven decisions based, as indicated above, on in-channel optimization, cross-channel/platform engagement measurement, predictive analytics and testing for impact.

Big data, big math

Thanks in part to the IBM Netezza data warehouse appliance, the company is solving its core advertising challenge: Navigating through a complex ecosystem to achieve great outcomes.

The IBM Netezza data warehouse appliance enables the company to handle big data and big math. For example, customer experience analytics can be run without headaches thanks to massively parallel algorithms. "The company didn't want the people building analytic applications worrying about data distribution while the algorithms are running," says Terrell. "The IBM Netezza appliance takes that worry away."

The company had 15 terabytes of data in its IBM Netezza data warehouse appliance in 2011, and that number is constantly growing. For example, about 20 million impressions flowed through the IBM Netezza data warehouse appliance in one recent fiscal quarter.

Why the IBM Netezza data warehouse appliance? It combines storage, server and a high-performance database in a simple appliance that works well with other platforms and reporting tools. The IBM Netezza data warehouse appliance outperforms traditional systems on query execution, and it's scalable. It can be run with minimal management overhead. Freed from worrying about technology, the company can focus on attracting and serving customers.

The company's big data analytics platform complements its high performance IBM Netezza data warehouse appliances with a multinode Hadoop cluster. By moving data back and forth between the two systems, the company is able to take advantage of both technologies. Hadoop helps in establishing relationships between unstructured data elements before they are loaded into the IBM Netezza data warehouse appliances. There, the ability of the IBM Netezza data warehouse appliance to execute high-speed complex queries and integrate with a broad array of extract, transform and load (ETL) and reporting tools with minimal management gives the software company the power, flexibility and scalability it requires.

What's next? The company will surely have more complex analytical needs as time goes on. "People expect instant gratification—with good reason," says Terrell. "Real-time access to data is very critical in many applications. So that puts a premium on many new types of analytics."

The company will also find new ways to leverage what IBM's Terrell, calls "the digital exhaust trail of human activity" across its growing product suite.

It's no small task to conduct customer experience analytics and take action on the results when your customer base is as large as this company's. But IBM helps solve these challenges in ways previously not possible, and the results for the company can be easily summed up: better products, a more delightful customer experience and increased return on investment.

About IBM Netezza data warehouse appliances

IBM Netezza data warehouse appliances revolutionized data warehousing and advanced analytics by integrating database, server and storage into a single, easy-to-manage appliance that requires minimal set-up and ongoing administration while producing faster and more consistent analytic performance. The IBM Netezza data warehouse appliance family simplifies business analytics dramatically by consolidating all analytic activity in the appliance, right where the data resides, for industry-leading performance. Visit: **ibm.com**/software/data/netezza to see how our family of data warehouse appliances eliminates complexity at every step and helps you drive true business value for your organization. For the latest data warehouse and advanced analytics blogs, videos and more, please visit: thinking.netezza.com

About IBM Data Warehousing and Analytics Solutions

IBM provides the broadest and most comprehensive portfolio of data warehousing, information management and business analytic software, hardware and solutions to help customers maximize the value of their information assets and discover new insights to make better and faster decisions and optimize their business outcomes.

For more information

To learn more about the IBM Netezza data warehouse appliance, contact your IBM sales representative or IBM Business Partner, or visit the following website: **ibm.com**/software/data/netezza

To increase the business value of your IBM data warehouse appliance, participate in an on-line community. Join the IBM Netezza community at: www.enzeecommunity.com

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
July 2012

Please Recycle

IMC14728-USEN-00

Automercados Plaza's

Increases revenue by 30 percent with greater insight into operations

Smart is…

Leveraging information across the enterprise to understand consumer buying patterns and needs

By integrating information across all divisions and functions, executives at Automercados Plaza's were able to gain better insight into how its customers shopped—critical knowledge that helped them increase revenue by 30 percent and improve profits by US$7 million annually. Now, staff can see which products and promotions are most profitable to improve contract negotiations with suppliers and identify which items to place on sale each week. Company staff also has concrete information to help guide the placement of new stores, enabling the company to successfully grow its business.

What if you had a daily pulse on your business and saw immediately which sales promotions were working, which products delivered the highest profitability and which locations offered the most promise for new stores? Executives at Automercados Plaza's, a family-owned chain of grocery stores in Venezuela with more than 1,000 employees, knew that this type of insight could help them grow the business to become a US$1 billion company. However, with more than 6 TB of product and customer data spread across different systems and databases, they could not easily assess operations at each store.

"We had a big mess related to pricing, inventory, sales, distribution and merchandizing," says Jesus Romero, CIO, Automercados Plaza's. "We have nearly US$20 million in inventory and we tracked related information in different systems and compiled it manually. We needed an integrated view to understand exactly what we have."

Integrating information across the enterprise

Working with IBM and IBM Premier Business Partner IT Consultings, Automercados integrated data across its 15 stores and corporate systems to enable the sharing of trusted information and gain greater insight into operations. Doing so now enables corporate personnel to quickly review daily inventory levels, store sales, and cost of goods to see which products are selling, which are most profitable and which promotions are most successful. For example, staff can use information collected from the point-of-sale systems to see which products customers buy together and develop product promotions with suppliers accordingly.

"We have the right tool to talk with our suppliers and provide them with facts to discuss what kind of promotions we want to deliver to our customers," says Romero. "This helps us better negotiate prices with our suppliers."

Business benefits

- Increased revenue by 30 percent
- Achieved a US$7 million increase in annual profitability
- Prevented losses for 35 percent of its products
- Helped executives pinpoint the optimal locations for four new grocery stores, including new 'supercenters'

"In the next two years, we are planning to increase our income by nearly 55 percent to become a US$1 billion company. IBM and SAP provide us with the capability to drive our business in a more proactive and productive way to help us realize our goal."

—Jesus Romero, CIO, Automercados Plaza's

In Venezuela, the government mandates prices for many grocery staples, such as milk, eggs and bread. The seamless flow of information between the company's SAP and point-of-sale systems also enables Automercados to quickly update prices across all stores, as needed, and more easily confirm compliance with government requirements.

"Everyday, we have to update prices in our point-of-sale systems, and now it's much simpler to ensure we place the right price on products," says Romero.

Increasing revenue by 30 percent

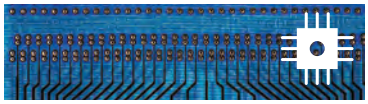
By integrating information across the enterprise, the grocery chain has realized a nearly 30 percent increase in revenue and a US$7 million increase in annual profitability. Romero attributes these increases to better inventory management and the ability to more quickly adjust to changing market conditions. For example, the company has prevented losses for about 35 percent of its products now that it can schedule price reductions to sell perishable products before they spoil.

Savings are also being realized with improved staff productivity. Previously, it could take nearly a month for finance staff to manually compile sales tax information. Now, the information is immediately available in the SAP system with a simple query—a more than 98 percent improvement.

Additionally, new insight has helped corporate personnel better understand sales by location to determine where to build new stores. In fact, in the past year, the company has successfully opened four new locations—including a new "supercenter"—based on consumer behavior and buying patterns.

Smarter solutions for retail

Assessing customer buying patterns to increase sales and profits



Instrumented

Sales information captured from each register in its 15 locations is automatically uploaded to the company's SAP business management system and data warehouse

Interconnected

Inventory, pricing, sales, customer and supplier information is integrated and shared among the company's business management, point-of-sale and data warehousing systems

Intelligent

Staff can review daily inventory levels, pricing changes, store sales, and customer buying patterns to identify successful promotions and determine where to build new stores

Solution components:

Software

- IBM® Informix® 11
- IBM InfoSphere® Information Server
  - IBM InfoSphere DataStage®
  - IBM InfoSphere QualityStage™
  - IBM InfoSphere Information Analyzer
  - IBM InfoSphere Information Server Pack for SAP R/3
- SAP ECC
- SAP for Retail

Hardware

- IBM System x® running Red Hat Enterprise Linux

IBM Business Partner

- IT Consultings

"IBM InfoSphere software helps us consolidate information so we can more easily comply with government requirements and better work with our suppliers and stores to apply the right prices at the right time."

—Jesus Romero

"In the next two years, we plan to increase our income by nearly 55 percent to become a US$1 billion company," says Romero. "IBM and SAP provide us with the capability to drive our business in a more proactive and productive way to help us realize our goal."

► The inside story: Getting there

To enable growth at the retail chain, IT Consultings first helped the company migrate from a legacy enterprise resource planning system to SAP ECC and SAP for Retail software running on IBM® System x® servers with Red Hat Enterprise Linux®. This change enabled staff to manage supplier relationships, inventory, sales and customer information on a single high-performance system.

Next, Automercados Plaza's leveraged IBM InfoSphere® Information Server software to enable the sharing of trusted information across the enterprise. IBM InfoSphere Information Analyzer helped staff profile and assess data to identify any anomalies, such as inconsistent customer information across different databases. InfoSphere QualityStage™ then helped standardize and cleanse customer data. InfoSphere DataStage® is used to integrate data daily across the company's SAP, point-of-sale and data warehousing systems. For example, using InfoSphere DataStage, data from each point-of-sale system in every store is loaded daily into the company's IBM Informix®-based sales consolidation system, helping company executives more quickly spot increasing demand for specific products.

With the InfoSphere Information Server Pack for SAP R/3, no manual coding was required for repeatable, reusable connectivity to SAP applications—key in helping reduce data integration time and costs.

"IBM InfoSphere software helps us consolidate information so we can more easily comply with government requirements and better work with our suppliers and stores to apply the right prices at the right time," says Romero. "We now have greater confidence in our information and can make better decisions."

For more information

To learn more about how IBM can help you transform your business, please contact your IBM sales representative or IBM Business Partner. Visit us at: **ibm.com**/software/data/integration

For more information about Automercados Plaza's, visit: www.elplazas.com

For more information about IT Consultings, visit: www.itconsultings.net

IBM Corporation
Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
July 2011

Barnes & Noble

Helping suppliers track sales and inventory in real time

Overview

The need

Create an online reporting tool that tells suppliers what books and other items have and have not sold.

The solution

Barnes & Noble improved communications with suppliers by deploying a web-based sales and inventory portal on a pre-integrated, pre-optimized data warehouse appliance from IBM. Publishers log on to get metrics on sales and inventory, then use the information to optimize inventory levels and avoid costly returns and stock-outs.

The benefit

Reduced inventory levels and inventory carrying costs; empowered small publishers to improve business results with real-time, sell-through status; decreased time to run queries from weeks to seconds.

Barnes & Noble, the No. 1 bookseller in the United States, has an unusual supply challenge: Its inventory is a mile wide and an inch deep. A store can carry up to 200,000 titles, including single copies that can sit on a shelf for a year. That makes it difficult to track results, especially for publishers.

How can they get accurate sales and inventory updates and avoid returns and stock-outs?

Barnes & Noble has built a solution, using the IBM Netezza® data warehouse appliance as its platform for a web-based sales and inventory portal that feeds metrics right to the publishers.

"Suppliers can log in on a daily basis and see sales and stock ratios," says Tom Williams, director of Web Services for Barnes & Noble. "It shows them what's selling and how, and the categories they're strong or weak in."

The goal of this service? To integrate the supply chain and help publishers reduce their costs. "We win if they win," Williams says.

Smarter Computing: The IT infrastructure of a Smarter Planet

Barnes & Noble transformed a four terabyte data warehouse with hundreds of billions of rows into a dynamic, agile platform that delivers real-time insight into product sales and stock ratios. Moving from a traditional data warehouse to a pre-integrated, pre-optimized platform enabled the company to perform analytics that were previously impossible. Barnes & Noble can now provide suppliers with a view of their inventory as never seen before, including big supply chain views, geographic views of the data, stack charts and heat maps showing entire groups of titles and subjects over time. The platform's simplicity has enabled Barnes & Noble to manage the environment with one part-time person, instead of 3.5 full-time equivalents (FTEs) that would be needed for other platforms.

Building better supplier relationships

A nationwide retailer with more than 700 stores, an online subsidiary (barnesandnoble.com) and an e-reader (the Nook), Barnes & Noble provides a comfortable environment whatever the channel. Readers can order a book online, load one on the Nook, browse through the racks or sip coffee in an in-store Starbucks. But one goal had not yet been achieved: to integrate with suppliers the same way it had with customers. "We wanted to make it easy and transparent for them to work with us," Williams says.

Specifically, publishers needed a better way to track results. "There isn't a really good way with the leading industry solutions to get a picture of how things are going," Williams explains. "Those solutions provide weekly data. You don't get to see inventory over time."

The consequences of this weak snapshot? Publishers absorb the losses on returns when they print too many books, and have stock-outs when they print too few. And the impact is worse on small publishers, those at the end of what Williams calls "the long tail" of the traditional supply-and-demand model.

Big publishers have resources; the smaller players don't. So Barnes & Noble sought to provide "one materialized view." But it had to be easy to use because the firm was in no position to train 10,000 independent publishers. And given the thin amounts of inventory and the wide range of products, there were unique business intelligence (BI) requirements.

"Peers from the apparel industry can have 10,000 SKUs," Williams says. "Here there are 20 million, and it's even greater with eBooks, where you don't have stock and can keep it as an active title indefinitely."

Solution components
- IBM Netezza® 1000
- IBM Netezza 100

"Suppliers can log in on a daily basis and see sales and stock ratios. It shows them what's selling and how, and the categories they're strong or weak in."

—Tom Williams, Director, Web Services, Barnes & Noble

In 2011, Barnes & Noble tested an early version of its portal, using Microsoft SQL Server as a platform. The system was tested in a limited pilot with 10 large publishers. "We learned quickly that the solution had reached the limits in terms of what could be accomplished with analysis services," says Williams. "There was simply too much data."

Seeking greater performance and simplicity, Barnes & Noble switched to an IBM Netezza data warehouse appliance in order to speed up the queries and deliver higher concurrency.

"The first version lacked the ability to say, 'There's still one copy on the shelf,'" Williams says. "It's now organized around transitions—did one come in, did one go out?—and constructing a snapshot of where things are in real time."

He adds, "We knew what we wanted. We could have looked at Vertica and [Oracle] Exadata, but Exadata would have put us where we were. We didn't want to manage an Oracle database putting cubes on top of it and running custom UDX [user defined extensions] for it."

The turnover went smoothly, and the suppliers in the pilot "didn't even know that happened," Williams says. "The IBM Netezza implementation started in the winter and went live in the spring," he adds. "With the Microsoft implementation, we were out there for over a year."

Given the scope of the challenge, all you need for a successful implementation is "the proper thought process upfront," says Williams.

Barnes & Noble now has about four terabytes of data on the IBM Netezza data warehouse appliance. "It's not a lot of data because it's stored in an unusual way," Williams says. "The effective row count is in the hundreds of billions, but they are relatively thin. Compression is gained by not storing things you don't need. The appliance idea was just perfect for a group like ours. And it's evolved, so we can do more than we intended to do."

"We can manage the IBM Netezza data warehouse appliance effectively with one person. And very soon, it will be only part of the person's job."

—Tom Williams

Stronger reporting leads to better supply decisions

Barnes & Noble is now rolling the portal out beyond the pilot stage. "We've got more than 50 percent of the publishing market seeing it, although we don't have more than 50 percent of the publishers," Williams notes. And the IBM Netezza data warehouse appliance—the "simplest and easiest-to-maintain solution" according to Williams—provides several benefits:

- **Insight**

"Our suppliers now get pictures that we have never seen before at Barnes & Noble: big supply chain views, geographic views of the data, stack charts and heat maps showing entire groups of titles and subjects over time," says Williams. "Having the IBM Netezza data warehouse appliance allows us to stop on a dime. We're more agile at development, and we can even offer features that we don't have for our own teams. In fact, while the system was originally designed for suppliers, it has been so successful, buyers are now using it as well."

One happy user is a middle-tier publisher without the resources to develop its own tool. "They now receive pictures of their business that they wouldn't have had," Williams says. "The further publishers are in the long tail, the more excited they are."

- **Speed**

Queries are now answered in real time. Case in point: "A leading national publisher wanted to see sell-through, or the relationship between sales and stock in our distribution centers, over two years, but it took weeks to export data from the Microsoft warehouse into Access," Williams says. "We now do that for this leading publisher in seconds."

"Having the IBM Netezza data warehouse appliance allows us to stop on a dime. We're more agile at development."

—Tom Williams

- **ROI**

The IBM Netezza data warehouse appliance has helped the company reduce costs. Williams estimates that it would have taken 3.5 full-time equivalents (FTEs) to run the previous system, including part-time storage area network (SAN) administrators and database administrators (DBAs). "We can manage the IBM Netezza data warehouse appliance effectively with one person," says Williams. "And very soon, it will be only part of the person's job. I can't imagine another project of this scale that has required only half a headcount to keep running."

That's not the only potential savings. "Over time, we'll also spend less on hardware because we can scale out and still do the same things with one IBM Netezza system," Williams says.

Barnes & Noble could charge for a tool like this, but the company prefers to use it "as an incentive to do business with us," Williams adds.

What's next? "Forecasting, predicting out-of-stocks, clustering stores," Williams says. "It's nice to get the historic pictures. Now let's do a little more prediction."

For example, big publishers might identify stores with similar buying patterns—those strong in sports, say—and use that insight to plan smaller, more efficient print runs. "Whether you're a self-publisher or the biggest publishing house in New York City, we want you to be able to use this application," Williams says.

Whatever the future, Williams expects to be "throwing algorithms" at the IBM Netezza data warehouse appliance. "Am I happy with IBM Netezza? I am," he concludes. "Would I do something similar again? Absolutely."

About IBM Netezza data warehouse appliances

IBM Netezza data warehouse appliances revolutionized data warehousing and advanced analytics by integrating database, server and storage into a single, easy-to-manage appliance that requires minimal set-up and ongoing administration while producing faster and more consistent analytic performance. The IBM Netezza data warehouse appliance family simplifies business analytics dramatically by consolidating all analytic activity in the appliance, right where the data resides, for industry-leading performance. Visit: **ibm.com**/software/data/netezza to see how our family of data warehouse appliances eliminates complexity at every step and helps you drive true business value for your organization. For the latest data warehouse and advanced analytics blogs, videos and more, please visit: thinking.netezza.com

IBM Data Warehousing and Analytics Solutions

IBM provides the broadest and most comprehensive portfolio of data warehousing, information management and business analytic software, hardware and solutions to help customers maximize the value of their information assets and discover new insights to make better and faster decisions and optimize their business outcomes.

For more information

To learn more about the IBM Netezza data warehouse appliance, please contact your IBM sales representative or IBM Business Partner, or visit the following website: **ibm.com**/software/data/netezza

To increase the business value of your IBM data warehouse appliance, participate in an online community. Join the IBM Netezza community at: www.enzeecommunity.com

For more information about Barnes & Noble, visit: www.bn.com

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
May 2012

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions. It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs. THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

Actual available storage capacity may be reported for both uncompressed and compressed data and will vary and may be less than stated.

Please Recycle

IMC14749-USEN-01

Brigham and Women's Hospital

This research institution sets a new gold standard for processing computationally intensive pharmacoepidemiology studies when it implements a solution that delivers analytics 20 to 30 times faster, handling massive data volumes, including longitudinal data, with ease.

Boston, Massachusetts
www.brighamandwomens.org

"This solution will accelerate our ability to devise, test and publish new computationally intensive algorithms applied to large longitudinal healthcare databases that we hope become the 'gold standard' for researchers globally."

— Dr. Sebastian Schneeweiss, Vice Chief, Division of Pharmacoepidemiology

Brigham and Women's Hospital (BWH) is a 793-bed teaching affiliate of Harvard Medical School located in the Longwood Medical Area in Boston, MA. In addition to its biomedical research laboratories, BWH offers inpatient and outpatient services and clinics, neighborhood primary care health centers and state-of-the art diagnostic and treatment technologies.

The Opportunity

Traditional methods for analyzing large databases have become inadequate to deliver the potential that the research team at BWH envisioned for its growing trove of information. The institution was seeking an information management solution that would change the game, ultimately developing into a research tool that could learn over time, bringing the very latest drug effectiveness and interaction data right to patients' bedsides.

What Makes It Smarter

Drug risk awareness saves lives, which is why BWH recognizes the need to advance the complex analytics and computing power it deploys to help deliver "high-dimensional" pharmacoepidemiology research results. BWH implemented an information management solution that handles massive data volumes with ease and delivers analytics with unprecedented speed. The research team found it could immediately make full use of its large data sets and conduct multiple drug studies simultaneously. The solution is enabling researchers to design, test and apply brand-new algorithms to help identify drug risk warning signals far more quickly. The institution intends to use the solution to automate a process for continuous drug safety monitoring and evolve the solution into a system that learns from prior results to continuously improve its predictive accuracy.

What if you could uncover dangerous drug combinations faster than ever before and with more factors taken into account?

Solution Components

- IBM® Netezza® 1000 Data Warehouse Appliance
- IBM Netezza Analytics

Real Business Results

- Enables one of the department's novel algorithms, its high-dimensional propensity scoring, to run 20 to 30 times faster than in its previous relational database environment
- Allows the department to conduct basic analytic processing at two to three times previous speeds, with no change in code
- Enables research studies on larger databases and exploration of previously inconceivable new research avenues

For more information

Please contact your IBM sales representative or IBM Business Partner.
Visit us at: **ibm.com**/healthcare

To learn more about Brigham and Women's Hospital visit:
www.brighamandwomens.org

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States
June 2012

IMC14769-USEN-00

Fiserv

Saving USD8 million in five years and helping banks improve business outcomes using IBM technology

Overview

The need
Small and midsize banks and credit unions seek to attract, retain and grow profitable customer relationships while competing with the analytic capabilities of new mega banks.

The solution
Working with IBM, Fiserv is turning billions of transactions into actionable insights that help these banks better target offers and maximize their marketing dollars. The use of cloud technologies to consolidate and virtualize servers helps reduce costs and increase availability.

The benefit
Estimated increase of 100 percent or higher in the response rate to targeted marketing initiatives; estimated IT savings of USD8 million in five years.

With the trend toward consolidation in the banking industry, small and midsize banks and credit unions are seeking new ways to attract, retain and grow profitable customer relationships while competing with the resources of new mega banks. Many large financial institutions have embarked on aggressive programs to use predictive analytics technology to enhance their revenues. Smaller banks and credit unions frequently lack the scale and resources to develop these programs, hindering their ability to compete.

This is the challenge that Fiserv is helping small and midsize banks tackle.

As a leading financial technology services provider, Fiserv processes roughly 20 billion data transactions and moves more than a trillion dollars annually. The company supports more than 16,000 financial institutions. But Fiserv's goal isn't just to process data. The company has invested heavily in helping its customers gain critical and dependable analytical insights from the data—fast.

"It is all about helping clients make sense out of the petabytes of data that flow through Fiserv's systems annually," says David Rose, senior vice president of Decision Optimization Solutions at Fiserv.

Designed for Data
IBM information management and predictive analytic solutions helped Fiserv transform billions of raw transactions into actionable insight for its clients, giving small and midsize banks advantages previously available only to large financial institutions.

Tuned to the Task
Moving to a virtualized infrastructure on an integrated platform from IBM is expected to help Fiserv save an estimated USD8 million during the next five years, reduce the number of midrange servers under management by approximately 90 percent, and provide higher availability.

Managed with Cloud Technologies
Creates a foundation to build a self-service environment that will improve the agility of service delivery.

Driving Innovation
Enables Fiserv to provide its customers critical, dependable analytical insights from petabytes of data—and provide it fast—to differentiate itself in the market and gain a competitive advantage.

Targeting offers through predictive analytics

According to Rose, by applying predictive analytics, a bank can see a number of meaningful financial benefits. These include, he says, an increase of 100 percent or higher in the response rate to targeted marketing initiatives and a significant increase in debit card utilization.

"Today, many smaller financial institutions are offering the same product to everyone," he explains. "Our predictive analytics service will help financial institutions focus on their most profitable customers and create goal-oriented offers that are targeted to each customer's specific needs. For example, by mining the data, we can see if a customer's spending pattern has shifted to include children's stores, signaling a lifestyle change and an opportunity for the bank to provide a targeted offer for an education savings account. This type of focus enables a bank to get the maximum payback for their marketing investment."

Creating a platform for trusted information

To enable its clients to deliver what Rose calls "high-impact" customer experiences, Fiserv must integrate billions of transactions from core banking account processing, person-to-person payments, mobile payments, electronic funds transfers, and e-bill payments, and apply predictive models and analytics to determine what a consumer will likely need next. This information then must be fed into business intelligence dashboards that present the findings to bank staff.

Solution components

Servers
- IBM Power® 770

Software
- IBM AIX®
- IBM Banking Data Warehouse Model
- IBM Cognos® Business Intelligence
- IBM i
- IBM InfoSphere® DataStage®
- IBM InfoSphere Discovery
- IBM InfoSphere® Warehouse
  - IBM DB2® 9.7
- IBM PowerHA®
- IBM PowerVM®
- IBM SPSS® Modeler
- IBM SPSS Statistics
- IBM SPSS Collaboration and Deployment Services
- IBM System Director
- IBM Tivoli® Storage Manager 5.5
- IBM Tivoli System Automation for Multi-Platforms
- IBM WebSphere® Application Server

Services
- Data Center Services
  - IBM Server Product Services for Power Systems™
- IBM Software Services for Information Management

"Our predictive analytics service will help financial institutions focus on their most profitable customers and create goal-oriented offers that are targeted to each customer's specific needs."

– David Rose, Senior Vice President, Decision Optimization Solutions, Fiserv

It's a huge undertaking, one that Fiserv executives believed only one vendor could support.

"Fiserv has been a client and partner of IBM for many years," says Rose. "Our bill payment data warehouse, which contains billions of transactions, is built on IBM DB2® and is one of the larger implementations of a DB2 data warehouse in the industry. Most recently, we use IBM's full solution set—everything from data extraction and transformation tools to predictive modeling and analytic solutions—to provide the trusted information we need to turn billions and billions of raw transactions into highly usable, actionable insights."

In building Fiserv's predictive analytics service, Rose turned to Leroy Hill, manager of the company's Midrange Engineering organization. The Midrange Engineering team provides the computing platform that supports many of the bank's critical business applications, including e-banking, bill payment and investment services, along with the company's enterprise data warehouse.

The timing was perfect, according to Hill. His team was working on a significant IT infrastructure initiative that would not only reduce IT costs, but also provide a more agile, available and scalable infrastructure for initiatives such as the company's predictive analytics service.

"We have estimated a five-year, cumulative run rate reduction of about USD8 million with the server consolidation and virtualization project."

– Leroy Hill, Manager, Midrange Engineering, Fiserv

"We've seen data doubling almost every two years," says Hill. "And the challenge was: How do we lower the cost of managing the data while making sure users have access to it?"

For Hill, the answer was in creating a virtualized environment that improves server utilization and enables the sharing of server resources across business units. The company's existing midrange environment used a traditional client/server computing architecture that provided business units with individual servers for specific applications. This model required a transformation to gain the agility needed to support the company's new business opportunities.

"As we grew, we had to add more physical servers, which meant added maintenance, support and overhead costs. And getting the new systems in place was time-consuming," says Hill. "There was also the area of failover capability where more advanced solutions were needed."

The company's IT infrastructure initiative changes this paradigm, using cloud-like technologies to consolidate and virtualize servers, optimizing server utilization while providing the right platform for the right workload. The use of virtual machines that can be moved from one physical server to another—without human intervention—enables the organization to rapidly provide additional processing power when needed and reduces downtime due to scheduled or unscheduled maintenance.

"We're moving more toward 24x7 capability, in which if we have a hardware or operating system problem, or need to do maintenance, the virtualized environment enables us to failover to another machine without affecting the application," says Hill.

"We use IBM's full solution set — everything from data extraction and transformation tools to predictive modeling and analytic solutions — to provide the trusted information we need to turn billions and billions of raw transactions into highly usable, actionable insights."

– David Rose

The first phase of the server consolidation and virtualization project is expected to help reduce the number of physical servers under management from just over 150 servers down to approximately 15 servers and decrease the consumed data center space by approximately 2,500 square feet. Technology refreshes that typically were multi-year efforts could be accomplished in a fraction of the time now because of the fewer number of servers.

Five-year savings estimated at USD8 million

The expected benefits of moving to a virtualized environment are enormous.

"Costs for power and cooling, hardware maintenance, software licensing and support will all be reduced," says Hill. "We have estimated a five-year-cumulative run rate reduction of about USD8 million with the server consolidation and virtualization project."

Creating hardware profiles and standard configurations for different application types and resource requirements (for example, small, medium and large applications) enables Fiserv business units using the shared standard platform stack to reduce the IT spend.

The journey to Smarter Computing

Hill's team is working closely with IBM on its IT transformation journey.

"IBM is helping us to assess our strategy and technical requirements and see overall how cloud technologies can play into our environment," says Hill. "What that has done is helped us validate our approach and gain a better picture of how we can use these technologies to better serve each of the Fiserv business units we support."

"We're moving more toward 24x7 capability, in which if we have a hardware or operating system problem, or need to do maintenance, the virtualized environment enables us to failover to another machine without affecting the application."

– Leroy Hill

The new virtualized environment is based on IBM Power Systems™ servers running IBM AIX® and IBM i operating systems. According to Hill, the Power Systems provide the performance and reliability for the company's demanding database serving and transaction processing requirements.

"I've worked with IBM technology for quite a long time and the platform is among the highest performing we've found," says Hill. "Our enterprise data warehouse today houses billions of rows of data and on a nightly basis we're adding new data at a pace of over 120,000 I/Os [input/output transactions] per second. IBM hardware and software work really well together and deliver the performance and savings we need to handle this load."

The IBM PowerVM® solution enables continuous, dynamic resource adjustments across all partitions and operating environments to optimize performance while minimizing energy usage. The IBM PowerHA® solution is used to automatically detect hardware errors and failover operations to a standby virtual server. IBM System Director software provides technology to complement the team's system management tools and unify the management of physical and virtual server resources.

"This actually helps us move closer to a cloud platform because now we can build those standard VMs [virtual machines]," says Hill. "Our goal is to move more toward a self-service environment and this builds the foundation for getting there."

In creating the company's new midrange IT environment, Hill's team has added a "lab environment" that enables his clients, like Rose's team, to build and test new services. "The project we worked on with David Rose was built as a proof of concept in a new virtualized lab environment," says Hill. "It gave the organization the ability to test theories and implement modeling tools to conduct predictive analytics."

For Hill, the work completed so far represents only part of the story. "This year, we're supporting the first phase of the project," says Hill. "There are many areas of interest that we want to explore such as automated provisioning of virtual machines and forecasting demand. Our goal is to deliver IT services that support the quality and timeliness of Fiserv's business priorities and this approach will clearly help facilitate that."

For more information

To learn more about IBM Smarter Computing solutions, please contact your IBM sales representative or IBM Business Partner, or visit the following websites: **ibm.com**/software/data or **ibm.com**/smartercomputing

You can get even more out of Information Management software by participating in independently run Information Management User Groups around the world. Learn about opportunities near you at: **ibm.com**/software/data/usergroup

For more information about Fiserv, visit: www.fiserv.com

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
April 2012

Please Recycle

IMC14734-USEN-00

Harvard Medical School

Faculty researchers bring drug safety and effectiveness studies to enhanced levels with IBM® Netezza® data warehouse appliance technology

Overview

The need
The lab at Brigham and Women's Hospital was looking to find a platform for computational pharmacoepidemiologic analytics that would address rapidly emerging trends.

The solution
IBM Netezza data warehouse appliance

The benefit
By utilizing IBM Netezza data warehouse appliance technology in their pharmacoepidemiology research studies, the Harvard Medical School division will be able to:

- Increase the speed of computationally-intense analysis of claims data
- Accelerate testing of new, more sophisticated algorithms
- Facilitate automation of continuous drug safety and effectiveness monitoring

Introduction

The Harvard Medical School® Division of Pharmacoepidemiology and Pharmacoeconomics at Brigham and Women's Hospital in Boston, Massachusetts, is a globally recognized leader in drug safety and effectiveness research. Created in 1998, the division is led by Dr. Jerry Avorn, Professor of Medicine at Harvard. The division is part of Brigham and Women's Hospital's Department of Medicine, and performs advanced analytics on patient health claims data. Through computationally intensive 'big data' analytics, this research brings to light insights into how drugs compare to each other in terms of safety and effectiveness.

The charter of the Division of Pharmacoepidemiology and Pharmacoeconomics is to:

- Study the use of medications in very large populations in order to determine which drugs are most likely to effectively treat patients' medical conditions, and which are more likely to cause adverse events.
- Develop advanced, data-intensive methods in epidemiology, biostatistics and informatics to detect and evaluate important early signals of potential drug safety problems, as well as to compare the therapeutic results of alternative treatment strategies.
- Define patterns of physicians' prescription and patients' use and compliance of medications.
- Analyze the influence of health system factors (policies, coverage and reimbursement differences) on the quality of medication use.
- Design, implement, and test innovative programs to improve the appropriateness of medications prescribed by physicians and used by patients.

"We as a society cannot afford to be told, 'I can't run your algorithms… because our technology can't support it.' People need and want to know the information that our studies generate. We can't afford not to carry out these studies. It's too important to get these answers for drug developers, regulators, physicians, policy makers, payers and most importantly, for our patients."

— Dr. Schneeweiss, Director for Drug Evaluation and Outcomes Research Brigham and Women's Hospital

The challenge

The nature of Harvard's pharmacoepidemiology research has been highly complex and computationally intensive since its inception. Today, however, the sheer volume of information available – and the increasing detail in which the information is recorded – has dramatically increased the analytic and computational challenges.

The expansion of data – both in its depth and breadth – has also presented opportunities for more complex and robust drug safety and effectiveness studies. Dr. Avorn and his colleague, Dr. Sebastian Schneeweiss, Director for Drug Evaluation and Outcomes Research, in their 2009 editorial "Managing Drug-Risk Information-What to Do with All Those New Numbers" (*New England Journal of Medicine*, July 27, 2009), reveal some of the diverse challenges facing epidemiological research in the current era. The challenges include applying new, rigorous analytic techniques to enhance the levels of confidence in observed drug safety results, as well as challenges in organizational politics, patient privacy and administration. This diversity is precisely why the research group includes experts in fields such as specialty medicine, health policy, sociology, law and more.

With access to the markedly larger amounts of patient data, the more traditional 'reactive' approaches to adverse drug events can no longer suffice. Examining data after the fact to confirm a suspected drug safety problem will always be a crucial element of safety research. Powerful computation simultaneously opens up avenues for identifying issues after a drug has entered the market, but before it is in widespread use. The Harvard team envisioned the potential of proactive research techniques using very large databases. This way, they could identify "signals" of risk associated with the particular drugs themselves or with combinations of drugs.

"IBM Netezza will accelerate our ability to devise, test and publish new computationally intensive algorithms applied to large longitudinal healthcare databases that we hope become the 'gold standard' for researchers globally."

— Dr. Schneeweiss

The lab at Brigham and Women's Hospital was looking to find a platform for computational pharmacoepidemiologic analytics that would address these trends. Specifically, Dr. Jeremy Rassen, Assistant Professor of Medicine at Harvard, identified certain needs as critical:

- Rapid execution of computationally intense analysis of structured medical claims data.
- Capabilities for parallelized in-database analytics.
- Accelerated development and testing of new, more complex algorithms.
- Facilitated automation of continuous drug safety event monitoring.
- Capabilities for complex analytic SQL queries.
- Simplified database administration.
- Ability to work with semi-structured electronic healthcare record databases .
- Security features for the continued protection of patient privacy.

The solution

IBM partnered with the lab and installed an IBM Netezza data warehouse appliance. IBM Netezza data warehouse appliances are purpose-built to make advanced analytics on data simpler, faster and more accessible.

The IBM Netezza data warehouse appliance is:

- Designed specifically for running complex analytics on very large data volumes, at orders of magnitude faster than competing solutions.
- Simple to maintain – The appliance architecturally integrates database, server and storage into a single, easy to manage system that requires little on-going maintenance.
- Simple to deploy – As a purpose-built product, the IBM Netezza data warehouse appliance comes pre-tuned with the optimal architecture for fast, advanced analytics, making it ready to go right out of the box.
- Simple to try – The IBM Netezza TestDrive makes it easy to experience the high-performance data warehouse appliances on-site, with real data.
- Unmatched in value – The IBM Netezza data warehouse appliance offers faster time-to-value and lower total cost of ownership than competing products in the industry.

IBM Netezza data warehouse appliances deliver the proven performance, value and simplicity organizations need to dive deep into their growing data. Customers are able to easily and cost effectively scale their business intelligence and analytical infrastructure, to leverage deeper insights from growing data volumes, throughout their organization.

The results

According to Dr. Rassen: "The IBM Netezza data warehouse appliance put us ahead of the curve. With almost no system administration or index-building, we quickly saw orders of magnitude performance improvement. It also allowed us to conduct basic analytic processing at two to three times previous speeds, with no change in code. Furthermore, it enabled one of our novel algorithms, the high-dimensional propensity scoring, to run 20-30 times faster than in our previous relational database environment. The appliance also gave us the ability to explore previously inconceivable new research avenues, with the availability of highly parallel data access and analysis."

As a result, Dr. Rassen's team has been able to:

- Load data into the database more quickly.
- Run all previously developed SQL and SAS code with minimal changes.
- Know that their efforts to migrate certain SAS analytics and processes to the open-source R language would be easy.
- Dramatically increase query response.

The research team noted that one of the very significant benefits of the technology was its simplicity of use. Dr. Rassen remarked that the system was situated in the data center, installed, and up and running in less than 48 hours. In addition, the team required no outside administration time, and system maintenance was handled by the researchers and analysts themselves. Other systems typically require weeks of set-up and tuning, not to mention frequent or constant administration.

The impact of the analytics

The stakes are high. In the U.S. alone, half of the population is taking one or more prescription drugs. Nearly four billion prescriptions are being filled annually. U.S. prescription drug sales are well over $300 billion per year, and according to IMS Health, global sales are expected to reach $1.1 trillion by 2014. Additionally, the Kaiser Foundation is reporting that prescription drugs account for well over 10 percent of

The high-dimensional data analysis approach for healthcare databases is based on the assumption that there is more information in such databases than is generally utilized. Such information may not immediately look relevant but our algorithms manage to identify previously-undetected risk factors in those data and improve adjustment to make treatment groups more comparable.

High Dimensional Pharmacoepidemiology
www.*hdpharmacoepi.org*

the total U.S. health expenditure (over $2.5 trillion). These figures highlight the positive impact that Harvard's advanced research can have on more rapid identification of unsafe and less effective drugs. The absence of more rapid risk awareness can be devastating to patients and continuously crippling to healthcare costs. Without steadily advancing complex analytics and tremendous computing power, the exceptional potential of Harvard's vision in pharmacoepidemiology research could face delays that global healthcare cannot afford.

The Division of Pharmacodpidemiology and Pharmacoeconomics at Brigham and Women's Hospital has developed research relationships with several *Fortune 500* companies to perform intervention studies on their medication benefit programs offered to employees. The lab is also a critical drug safety and effectiveness research training destination for many doctoral and post-doctoral students from all over the world. Division faculty have produced well over 300 notable papers on drug safety, drug costs and health policy. The quality and amount of accessible patient data for the group's research is critical.

The lab calls a major portion of what they do "High Dimensional" Pharmacoepidemiology. It seeks to take advantage of all the high-dimensional data space available in longitudinal insurance claims data, registries and electronic healthcare record databases. This will improve the validity of the research in the frequent cases where running a randomized trial is not possible or feasible.

Recognizing that the safety and effectiveness of new and existing prescription drugs plays an ever increasing role in the pursuit of improved healthcare worldwide, the research team designs, develops and executes studies to provide more reliable, actionable data that can:

- Bring greater clarity to issues of benefits versus risks.
- Evaluate the impact of prescription drug expenditures.
- Understand how medications are prescribed by doctors and used by their patients.
- Create action plans to help patients improve adherence to medications.
- Develop methods that optimize the use of prescribed drugs.
- Identify drug safety issues for specific subsets of the population.
- Assist public sector governing bodies in their decision making processes.
- Provide physicians with new insights when prescribing medications.
- Create a system of active drug safety surveillance, allowing proactive risk intervention.
- Help pharmaceutical companies bring safe, effective new drugs to market more smoothly.

"Things that we could not have done before…things we would not have considered, are now easily within our reach."

— Dr. Schneeweiss

The Harvard Medical School division has already achieved major accomplishments, one of which is developing new methodologies for computer-intensive analysis of very large datasets to identify drug risks. Another is identifying important "confounders" that can distort the results of non-randomized studies – these confounders can skew results and make incorrect findings appear robust. A third methodology seeks to assess the cost effectiveness of particular medications and drug-use strategies. As meaningful as these developments are, Dr. Schneeweiss envisions continuous advances in the analysis of ever-growing patient databases with even richer data. These advances will help automate and steadily improve society's ability to monitor drug safety and effectiveness.

What is a patient cohort?

In epidemiology, a cohort is a group of like individuals who are followed over time and analyzed for the incidence of diseases, adverse events or other health outcomes of interest.

> "As global healthcare evolves toward an ongoing learning healthcare system with a need for continuous patient-centered comparative effectiveness and safety research integrated in routine care, it is imperative that research methods evolve in parallel…IBM Netezza will accelerate our ability to devise, test and publish new computationally intensive algorithms applied to large longitudinal healthcare databases that we hope become the 'gold standard' for researchers globally…Accurate effectiveness information generated with little delay will help healthcare decision makers to improve the health of our patients." – Dr. Schneeweiss

The dramatic expansion of available patient data, and the ever-increased complexity of questions that need to be asked in more enriched studies, mean traditional methods have simply become too slow. With the research group's new algorithms and goals of drug safety automation, the computational demands on processing technology have increased exponentially but are now being met with IBM Netezza's data warehouse appliance.

Dr. Schneeweiss is fervent about the prospects of enhanced automation and more complex, large database studies that will be pursued with the IBM Netezza data warehouse appliance platform. His vision of a system that is continuously "learning" by incorporating data from ongoing drug safety and effectiveness research results is a "game changer" for the future of how prescription drugs are managed and developed.

"Our lab is focused on the best ways to automate the continuous evaluation of the safety of medications and we require fast systems to assemble and analyze data."

— Dr. Rassen, Assistant Professor of Medicine, Harvard Medical School

Dr. Schneeweiss fully expects that the ongoing system will be able to generate reliable, early warning signals of adverse drug effects that have previously taken months to years to discover.

The benefits

By utilizing IBM Netezza data warehouse appliance technology in their pharmacoepidemiology research studies, the Harvard Medical School division will be able to:

- Increase the speed of computationally-intense analysis of claims data.
- Accelerate testing of new, more sophisticated algorithms.
- Facilitate automation of continuous drug safety and effectiveness monitoring.
- Allow research studies on larger databases.
- Operate with little to no outside database administration.
- Make it possible to work on multiple databases at the same time.
- Add the ability to store and analyze electronic health record databases .
- Achieve higher levels of research result-confidence.
- Continue the Division's protection of patient privacy.
- Facilitate the development of an automated 'learning' healthcare system.
- Identify drug risk 'signals' far more quickly than previous systems could.
- Provide drug developers and regulators with superior large-scale, post-market data.
- Facilitate in-depth, active surveillance for post-market drug safety.
- Provide extensive data to assist with public sector health decisions.
- Offer physicians more detailed data for their prescription decision process.
- Create tools to improve adherence to beneficial medications.
- Present patients with more reliable information about their prescribed drugs.

The team

The Harvard team's drug safety and effectiveness research mission takes a uniquely inclusive approach by providing a deep interdisciplinary environment of experts for its studies. The division brings together clinicians from a wide-range of backgrounds including cardiology, geriatrics, primary care, rheumatology and nephrology; with experts in the quantitative sciences of biostatistics, epidemiology and complex analytics. Professionals in other pertinent disciplines such as health policy, law, decision analysis and the social sciences are also included. In doing so, they set their research focus sharply on the ultimate care of the patient.

> "Our high-dimensional pharmacoepidemiology takes an empirical, data-driven approach to solving the difficult problems of pharmacoepidemiology, drug safety and comparative effectiveness research. By making use of all the information contained in insurance claims, registries, and even electronic medical records, we seek to improve the validity of non-randomized research." – Dr. Schneeweiss

A small sample of publications from the Harvard faculty-researchers in the Division of Pharmacoepidemiology:

Risk of death and hospital admission for major medical events after initiation of psychotropic medications in older adults admitted to nursing homes.
Huybrechts K.F., Rothman K.J., Silliman R.A., Brookhart M.A., Schneeweiss S.
CMAJ. 2011 Apr 19; 183(7): E411-9

Anticonvulsant medications and the risk of suicide, attempted suicide, or violent death.
Patorno E., Bohn R.L., Wahl P.M., Avorn J., Patrick A.R., Liu J., Schneeweiss S.
JAMA 2010; 303: 1401-9

High-dimensional propensity score adjustment in studies of treatment effects using health care claims data.
Schneeweiss S., Rassen J.R., Glynn R.J., Avorn J., Mogun H., Brookhart M.A.
Epidemiology 2009; 20: 512–22

Privacy-maintaining propensity score-based pooling of multiple databases applied to a study of biologic agents.
Rassen J.A., Solomon D.H., Curtis J., Harrington L., Schneeweiss S.
Med Care 2010; 48: S83-9

Variation in the risk of suicide attempts and completed suicides by antidepressant agent in adults: A propensity score-adjusted analysis of 9 years of data.
Schneeweiss S., Patrick A.R., Solomon D.H., Metha J., Dormuth C., Miller M., Lee J., Wang P.S.
Arch Gen Psychiatry 2010; 67: 497-506

Post marketing studies of drug safety.
Schneeweiss S., Avorn J.
BMJ. 2011 Feb 8; 342: d342. doi: 10.1136/bmj.d342

The epidemiology of prescriptions abandoned at the pharmacy.
Shrank W.H., Choudhry N.K., Fischer M.A., Avorn J., Powell M., Schneeweiss S., Liberman J.N., Dollear T., Brennan T.A., Brookhart M.A.
Ann Intern Med. 2010 Nov 16; 153(10): 633-40

Cardiovascular outcomes and mortality in patients using clopidogrel with proton pump inhibitors after percutaneous coronary intervention.
Rassen J.A., Choudhry N., Avorn J., Schneeweiss S.
Circulation 2009; 120: 2322-9

About IBM Netezza data warehouse appliances

The IBM Netezza data warehouse appliance revolutionized data warehousing and advanced analytics by integrating database, server and storage into a single, easy to manage appliance that requires minimal set-up and ongoing administration while producing faster and more consistent analytic performance. The IBM Netezza data warehouse appliance family simplifies business analytics dramatically by consolidating all analytic activity in the appliance, right where the data resides, for blisteringly fast performance. Visit netezza.com to see how our family of data warehouse appliances eliminates complexity at every step and lets you drive true business value for your organization. For the latest data warehouse and advanced analytics blogs, videos and more, please visit: thinking.netezza.com.

IBM Data Warehousing and Analytics Solutions

IBM provides the broadest and most comprehensive portfolio of data warehousing, information management and business analytic software, hardware and solutions to help customers maximize the value of their information assets and discover new insights to make better and faster decisions and optimize their business outcomes.

IBM Corporation
Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
October 2011

Please Recycle

IMC14685-USEN-00

IBM

Applies emerging technologies to deliver instantaneous people searches

Overview

The need

With over 600,000 names in BluePages, IBM's employee directory, and over 500,000 queries daily, the average search session takes two minutes. IBM needed a faster, more efficient application.

The solution

Using Apache open source technologies, the IBM CIO Lab Analytics team developed a new people-search application that allows flexible queries and returns as many results as possible, as fast as possible. Additional capabilities include quick browsing and photo images.

The benefit

The new *Faces* application offers instantaneous response time, saving on average over a minute for each search session—and thousands of hours daily for IBM employees.

With an enterprise population of over 600,000 people worldwide, how do IBM® employees find and connect with their colleagues? For over a decade, IBM BluePages has been the primary source. This high-demand, intranet application provides information on all IBM employees and contractors, including areas of expertise and responsibilities. And with IBM's focus on innovation and emerging technologies, positive changes are always on the horizon.

"BluePages is one of the most used applications at IBM," says Sara Weber, manager of IBM's CIO Lab Analytics team. "At one time, BluePages was state-of-the-art; however, over the years it was not updated to keep up with new advances in Internet technology. With over 500,000 BluePages searches done every day, and with BluePages accessing huge volumes of data, an average search session can take up to two minutes. When multiple results are returned they do not show individual photo images, and incorrect spelling may yield no results. My team was tasked with addressing the question: 'How can we build a better and faster people search?'"

The goals for this project, aptly named *Faces*, were to support flexible queries and return as many results as possible, as fast as possible. Results that more closely matched the query would appear first. Additional capabilities would permit quick browsing and photo images.

"At IBM, when we find an open source technology that has potential, we experiment with it to understand how to use it to bring the most business value to IBM. For example, IBM InfoSphere BigInsights is a new class of analytics platform based on Hadoop and innovation from IBM. It can store raw data 'as-is' and help clients gain rapid insight through large scale analysis."

—Sara Weber, Manager, IBM's CIO Lab Analytics team

Applying emerging technologies to deliver innovation

Weber's CIO Lab Analytics team identifies problems that IBM employees are experiencing and finds ways to apply emerging technologies to develop solutions. "We had to process tremendous amounts of data, and then store it in a way that it could be accessed quickly," says Weber. "For this project, we selected Apache Hadoop and Apache Voldemort; both are open source technologies. My development team has extensive expertise in using Hadoop technology. The *Faces* application was developed by two members of our team over a five month period."

Apache Hadoop allows developers to create distributed applications that run on clusters of computers. Organizations can leverage this infrastructure to handle large data sets, by dividing the data into "chunks" and coordinating the data processing in the distributed, clustered environment. Once the data has been distributed to the cluster, it can be processed in parallel. Apache Voldemort is a distributed key-value storage system that offers fast, reliable and persistent storage and retrieval. Specific keys return specific values. If no additional query power is needed, a key value store is faster than a database.

"At IBM, when we find an open source technology that has potential, we experiment with it to understand how to use it to bring the most business value to IBM," says Weber. "For example, IBM InfoSphere® BigInsights is a new class of analytics platform based on Hadoop and innovation from IBM. It can store raw data 'as-is' and help clients gain rapid insight through large scale analysis."

For *Faces*, Hadoop preprocesses data from the IBM Enterprise Directory and Social Networks and sends this information to the Voldemort Person Store (2.2 GB). Voldemort, in turn, sends data to Hadoop processing for the Person ID fetcher, Reports Loader, Query Expander, and Location Expander. These results are saved to Voldemort's Query Store (5.5 GB). Hadoop also receives images from BluePages that are saved in Voldemort's image store to remain available for Hadoop's montage generator.

Solution components

Servers

- IBM® BladeCenter® servers

Software

- Apache Hadoop
- Apache Voldemort Key Value Storage System

"We could not have developed Faces without the distributed processing capabilities Hadoop provides. The Faces application has really highlighted the power of Hadoop and has helped us address a major pain point for all IBMers."

—Sara Weber

"We placed all 600,000 names into memory for immediate access," says Weber. "Preprocessing with Hadoop directly improves performance. Each time you type a letter in a name, results are immediate. We have precomputed the search process to retrieve every employee name that matches what is entered. Every time you type another letter, scoring retrieves people who are more relevant to the search criteria. The information is available and, from a performance perspective, everything is ready to go. Memory and storage are inexpensive and nightly processing takes only a few hours."

Weber adds, "We run Hadoop on ten, five-year-old IBM BladeCenter® servers. These Blades are low powered, but Hadoop distributes the workload and takes advantage of the hardware to the fullest. If more computation is needed, we can add machines and improve performance without modifying the code."

Measuring business value

According to Weber, the new *Faces* application enables employees to receive instantaneous search results. "Conservatively speaking, we are saving on average over a minute for each search session," says Weber. "Searches are faster and easier. The information is timely and accurate. With over 500,000 searches daily, IBMers are saving thousands of hours each day."

For IBM employees, the improvement is noticeable. "To gain user acceptance or change user behavior, we know any new solution we create has to be significantly faster and better," says Weber. "As far as I know, *Faces* is the fastest growing innovation ever introduced at IBM. In the first two weeks, *Faces* went from zero to 85,000 users with continued viral growth throughout the entire IBM organization. What used to take minutes now takes milliseconds. We provide a feedback button on all our applications so users can report errors or issues. With *Faces*, IBMers were using the feedback button to say, 'Thank you for making my job so much easier.'"

Weber concludes, "We could not have developed *Faces* without the distributed processing capabilities Hadoop provides. The *Faces* application has really highlighted the power of Hadoop and has helped us address a major pain point for all IBMers."

For more information

To learn more about IBM Information Management solutions, please contact your IBM sales representative or IBM Business Partner, or visit the following website: **ibm.com**/software/data

To learn more about IBM InfoSphere BigInsights, visit: **ibm.com**/software/data/infosphere/biginsights

Additionally, financing solutions from IBM Global Financing can enable effective cash management, protection from technology obsolescence, improved total cost of ownership and return on investment. Also, our Global Asset Recovery Services help address environmental concerns with new, more energy-efficient solutions. For more information on IBM Global Financing, visit: **ibm.com**/financing

IBM Corporation
Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
October 2011

Please Recycle

IMC14698-USEN-00

KTH – Royal Institute of Technology

Analyzes real-time data streams to identify traffic patterns

Stockholm, Sweden
www.kth.se/?l=en_UK

"Analyzing large volumes of streaming data in real time is leading to smarter, more efficient and environmentally friendly traffic in urban areas."

— Haris N. Koutsopoulos,
Head of Transportation and Logistics,
Royal Institute of Technology,
Stockholm, Sweden

The Royal Institute of Technology (abbreviated KTH) is a university in Stockholm, Sweden. KTH was founded in 1827 as Sweden's first polytechnic and is with Aalto University School of Science and Technology in Espoo, depending on definition, Scandinavia's largest institution of higher education in technology and one of the leading technical universities in Europe.

The Opportunity

Researchers at KTH, Sweden's leading technical university, gather real-time traffic data from a variety of sources such as GPS from large numbers of vehicles, radar sensors on motorways, congestion charging, weather, etc. The integration and analysis of the data in order to better manage traffic is a difficult task.

What Makes It Smarter

Collected data is now flowing into IBM InfoSphere Streams software—a unique software tool that analyzes large volumes of streaming, real-time data, both structured and unstructured. The data is then used to help intelligently identify current conditions, and estimate how long it would take to travel from point to point in the city, offer advice on various travel alternatives, such as routes, and eventually help improve traffic in a metropolitan area.

Real Business Results

- Uses diverse data, including GPS locations, weather conditions, speeds and flows from sensors on motorways, incidents and roadworks
- Enters data into the InfoSphere Streams software, which can handle all types of data, both structured and unstructured
- Handles, in real time, the large traffic and traffic-related data streams to enable researchers to quickly analyze current traffic conditions and develop historical databases for monitoring and more efficient management of the system

Solution Components

- IBM® InfoSphere™ Streams
- IBM BladeCenter® HS22
- IBM BladeCenter H Chassis
- IBM System Storage® DS3400
- Red Hat Linux®

For more information

Please contact your IBM sales representative or IBM Business Partner.
Visit us at: **ibm.com**/education

To learn more about KTH – Royal Institute of Technology visit:
www.kth.se/?l=en_UK

IBM Corporation
1 New Orchard Road
Armonk, NY 10504
U.S.A.

Produced in the United States
March 2011

BLC03060-USEN-00

macys.com: Focusing on each customer as the brand goes national

Overview

The need

macys.com sought to create a more personalized shopping experience.

The solution

macys.com works with IBM to establish a foundation for a more dynamic, data-driven and integrated website.

What makes it smarter

Macy's is laying the groundwork for customized content arbitration as well as more granular pricing and market strategies that increase margin.

The result

"We see the strength of IBM's data management solutions as a critical part of our success."

—Darren Stoll, Group Vice President of Marketing Operations and Analytics, macys.com

In the world of retail, Macy's is one of very few brands that can credibly be called an icon. Its flagship store in Herald Square in New York City, together with its long-running Macy's Thanksgiving Day Parade, are perhaps its most prominent commercial and cultural legacies. In recent years, the company behind the Macy's brand, Macy's, Inc., has undertaken an extensive—and by all accounts successful—effort to transform Macy's from a regional to a national brand. Though enormous in scope, the physical, logistical and organizational dimensions of the change—in which hundreds of regionally branded stores across the United States took on the Macy's name—represent only the first stage of the company's plan to redefine itself.

Although Macy's national branding strategy might have given it a bigger footprint, CEO Terry Lundgren is betting that the future lies in offering a more localized, personalized—and ultimately smarter—retail customer experience. Lundgren is the charismatic force behind My Macy's, a sweeping initiative designed to embed a customer-centric orientation into every aspect of the company's operations. More than just a catchy name, My Macy's reflects the way Macy's wants its customers to view their relationship with the company. It is a relationship marked by a level of personalization that manifests—and grows stronger—each time customers interact with the brand, whether online at macys.com, at a local store or traveling in another city. The aim is consistency and the key is data.

Building a data foundation

macys.com implicitly understood that to successfully implement the kind of customer-centricity envisioned, a comprehensive customer data strategy was a prerequisite—and developing one would not be easy. Its first order of business was to consolidate the various data repositories that had accumulated from predecessor companies. Realizing the foundational importance of this work to its long-term customer data strategy, macys.com turned to IBM—and its 20-plus years of experience working with Macy's—to provide the guidance and tools it needed.

Now some two years down the road and with the first stage of its transformation virtually complete, macys.com is working toward putting in place the architectural elements it needs for its most fundamental capability: a "360-degree view" of its customers to support a strategy of multichannel integration. This, notes Darren Stoll, Group Vice President of Marketing Operations and Analytics at macys.com, will mark an important milestone in Macy's journey. "Aggregating our customer data and achieving the 360-degree view of the customer will put macys.com at a starting point for customer-centric innovation," explains Stoll. "Once we achieve that view, we'll be in a position to operationalize it in a way that personalizes and enriches the customer experience."

Smarter solutions for retail: Macy's gets personal for a compelling experience

Instrumented

Capturing customer information at a variety of brand touch points enables Macy's to interact with its customers on a more personal level.

Interconnected

Macy's SOA-based integration framework is enabling macys.com to move toward a comprehensive, 360-degree view of the customer.

Intelligent

The ability to use real-time site information at the customer level enables the company to develop more targeted—and ultimately more profitable—pricing and promotional programs.

Business benefits

- Expected increase in sales and revenue per customer due to closer engagement, stronger retention and targeted promotions
- Strengthening of brand via a more consistent and integrated cross-channel customer experience
- Increase in macys.com volume facilitated by distribution center optimization
- Improved decision making via the seamless integration of consistent, high-quality data from across the enterprise.

Putting customer data to work

In the context of Macy's' strategy, putting customer data to use requires the means and expertise to aggregate it, analyze and extract insights from it, and then use it to shape the customer experience at macys.com. IBM data management technology plays a crucial role in every stage. For example, macys.com is employing the IBM InfoSphere™ Information Server platform featuring IBM InfoSphere DataStage® to integrate its multiple sources of customer data across the enterprise, while IBM InfoSphere QualityStage™ ensures that the data is always accurate, consistent and complete, and thus can be trusted to drive the company's core processes. IBM InfoSphere Warehouse Enterprise Edition provides macys.com with a powerful data warehouse framework for real-time analytics.

Although a data management foundation is essential, it is just as important to know how to weave customer data and insights into all of the processes that surround—and ultimately determine—the customer experience. For example, it is seen in the way a customer's preferences (both implicit and explicit) are combined with recent purchase data to create dynamically customized recommendations (such as a complementary clothing accessory or color) or personalized promotions. It is also seen in the seamless consistency of the customer experience across channels, whether it's in a Macy's store or on macys.com. On a more nuanced level, it is seen in the use of social media—from blogs to gift registries—to establish a deeper level of engagement between the customer and the Macy's brand. To implement these changes successfully, macys.com is counting on the breadth of IBM's best practices and cross-channel retail experience.

Solution components

Software

- IBM WebSphere® Application Server ND V6.1.x
- IBM DataPower® XI50
- IBM ITCAM for WebSphere
- IBM DB2® for z/OS®
- IBM InfoSphere™ DataStage® for System z®
- IBM InfoSphere QualityStage™ for z/OS
- IBM InfoSphere Warehouse Enterprise Edition
- IBM Rational® Software Architect
- IBM Rational ClearQuest®
- IBM Rational AppScan®
- IBM Rational Policy Tester™

Servers

- IBM System z
- IBM System p®
- IBM System x®

Services

- IBM Global Business Services®
- IBM Global Technology Services

► The inside story: Getting there

Changing the mindset

As the Macy's transformation program unfolds, it will ultimately affect each of the company's 161,000 employees, from the CEO to the sales associate on the floor. To soften resistance and encourage buy-in, Macy's turned strong executive sponsorship to its advantage by broadly publicizing the company's commitment to becoming more customer-centric.

Changing the metrics

Macy's realized that one of the key implications of its customer-centric strategy was the need to redefine the way it measures success, from more traditional by-the-numbers metrics such as event sales and item sell-through rates, to longer-term measurements of customer engagement, retention and ultimately wallet share. "For customer-centric retail strategies to work over time, there needs to be a fundamental change in how we think about things, what we consider important and how we evaluate our performance along with measures," Stoll explains. "It's a change that will affect every part of our business."

Aligning IT with the business

Given the central importance of customer data management to its new strategy, macys.com is acutely aware of the need to align its business strategy with its long-term technology road map. While establishing a solid data management foundation is the immediate imperative, says Stoll, "We need to be sure we at the company can translate our evolving business needs into the right decisions on technology and architecture—three years, five years…10 years down the road. If we don't address that now, it will create roadblocks for us in the long run by reducing our flexibility. That's where a company like IBM can be very helpful in pointing us forward as an organization."

The Macy's customer-centric strategy is about growing its core business by strengthening customer loyalty and ultimately increasing wallet share. The company's unflagging commitment to the transformation program—even in the face of major retail downturn—is a measure of its confidence in its strategic direction. A bright spot is the IBM-hosted macys.com site, whose revenues have grown substantially since 2005 on the strength of faster navigation and broader selection. In fiscal 2009, online sales at Macy's, Inc. grew by 19.6 percent. macys.com sees the ability to create a more dynamic customer experience as a key enabler in its goal to continue to grow revenue significantly in the years to come.

Smarter retail, bigger profits

As Macy's continues to refine and expand its multichannel capabilities, it expects the macys.com channel to play an increasingly prominent role in driving growth. To make this expansion possible, Macy's worked with IBM to optimize its network of fulfillment centers and supporting processes to enable broader and more efficient fulfillment of online sales while lowering costs.

To Darren Stoll, the clear benefits already resulting from recent strategic initiatives are a sign that the company is on the right track and well positioned for strong growth. "macys.com is committed to delivering the kind of dynamic and personalized experience that engages customers and strengthens the Macy's brand," says Stoll. "We see the strength of IBM's data management solutions—and its ability to put insights into action—as a critical part of our success."

For more information

To learn more about how IBM can help you transform your business, contact your IBM sales representative or IBM Business Partner.

Visit us at: **ibm.com**/smarterplanet/retail

IBM Corporation
1 New Orchard Road
Armonk, NY 10504
U.S.A.

Produced in the United States of America
August 2010

ODC03172-USEN-00

Marine Institute Ireland

Putting real-time data to work and providing a platform for technology development

Overview

The need

The Marine Institute sought to establish SmartBay as a research, test and demonstration platform for new environmental technologies—paving the way to commercialization and the development of new markets for Irish-based companies.

The solution

The Institute, working with IBM, developed a pilot information system to feed environmental data into a data warehouse, where it is processed, analyzed and displayed in new ways.

What makes it smarter

The project yields greater insight into the bay environment, as well as providing practical value—from understanding how water quality impacts fisheries to predicting hazard locations and more.

When sensors become pervasive, entirely new and unexpected uses for the flood of information they produce often arise, yielding benefits far beyond those originally envisioned. Seeing the world in a new way—via technology—generates an inventive spark, prompting people to devise new uses for information that they may never have considered before.

That's exactly what is happening in Ireland's Galway Bay, as part of the SmartBay project initiated by the Marine Institute Ireland. In support of its advanced technology platform, which seeks to make Ireland a major player in the development of smart ocean technologies, the project's initial purpose was to develop a platform for testing environmental monitoring technologies, and the idea was simple: Deploy a series of radio-equipped "smart buoys" in the bay containing sensors that could collect data such as sea state (wave height and action) and other weather conditions, water data such as salinity, and similar environmental information.

A basis for economic transformation

When the Marine Institute learned of the IBM Big Green Innovations initiative to find ways to use technology to promote and enable environmental science, the idea of a collaboration on the SmartBay project was born. The IBM Advanced Water Management Centre Dublin built upon the domain expertise of the Marine Institute, complimenting it with its deep computing intelligence.

While the synergy with the IBM Smarter Planet™ strategy's drive towards Smart Green technology was clear, the real impetus behind the decision to expand SmartBay is largely economic. Beginning in the 1990s, the Irish economy became a global growth powerhouse. Wise policy decisions and forward-thinking investment had transformed Ireland into a manufacturing phenomenon.

More recently, with the global economy encountering difficulty, Ireland's prosperity began to wane. The government saw the need to change course, moving the country towards a knowledge-based economy. Investment in projects that showcase Ireland as a technological leader would not only create new commercial opportunities,

Business benefits

- Enables the creation of a vast array of diverse applications that goes far beyond the original purpose of the project, from technical research to tourism promotion
- Real-time access via the web delivers valuable insight quickly to remote users
- Open architecture enables new applications to be brought on line easily, combining data from both SmartBay sensors and other sources, such as geographical information systems
- Add-on effect of the project promotes education and stimulates economic development in the Irish economy

attract talent and additional capital investment, but also prompt a new generation of Irish citizens to pursue careers in knowledge-based industries.

Taking SmartBay to a new level

The Marine Institute, working in conjunction with government agencies, research institutions and the private sector, is working together to leverage the significant R+D capacity that exists in Ireland to help drive economic development. There is clear potential to expand SmartBay into an international platform demonstrating new approaches to environmental challenges and delivering new technological solutions for a range of global markets.

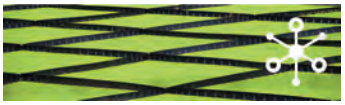
IBM is working with the Marine Institute to speed the process of innovation, starting with an assessment of existing capabilities. The team saw that if the data could be centralized, processed and accessed in the right way, it could become far more useful—the information already available could be turned into intelligence and put to work to create real practical value that impacts the lives of citizens directly.

IBM designed and deployed an enterprise-scale data warehouse using IBM InfoSphere™ Warehouse, that is connected to the SmartBay sensors, as well as external sources such as mapping databases and sensors beyond the bay. An open-standards application layer processes and analyzes the data in a variety of ways, making it available via a Web interface enabled by IBM WebSphere® Portal and WebSphere Application Server. Additional WebSphere products, including WebSphere MQ and WebSphere Sensor Events, provide a key middleware layer that integrates the sensors with the data warehouse. To ensure reliability and scalability, the system is housed on IBM System x® 3950 servers.

Smarter water: Creating new value from environmental data

Instrumented

Sensors deployed on buoys in Galway Bay transmit key data on ocean conditions and water quality.



Interconnected

Sensor data is fed into a central data warehouse for aggregation and processing, and can be accessed by diverse groups using customized web applications to generate targeted value.

Intelligent

Combining real-time data with a flexible technology platform creates near-limitless new uses for information—from environmental research to predictive monitoring, technology validation and much more.

Solution components

Software

- IBM DB2® Alphablox® v9.5
- IBM DB2 Enterprise Server Edition v9.5
- IBM InfoSphere™ Streams
- IBM WebSphere® Application Server v6.1
- IBM WebSphere MQ v5
- IBM WebSphere Sensor Events
- IBM WebSphere Portal Server v6.1

Servers

- IBM System x® 3950

Services

- IBM Global Business Services®

"The immediate benefits of SmartBay, whether it's helping and supporting industrial development or promoting marine safety, are tangible, direct and worthwhile."

—John Gaughan, project coordinator, SmartBay

The system design makes it easy to combine data from the sensors with other online databases—such as geographical information—as needed to create new functionality. Rapid development, enabled by IBM DB2® Alphablox® is an important feature, giving project managers the ability to deploy new applications quickly and easily.

The project yields greater insight into the bay environment and can provide real-time information feeds to a range of stakeholders, while at the same time enabling commercial technology developers to test new environmental product and service offerings. The project is now moving into a new phase, with higher bandwidth and powered cabled sensors being deployed that will enable more information to be gathered. IBM is also working with Irish-based companies on an advanced initiative to add stream (i.e., real-time) computing capabilities to the project, with the goal of increasing its capacity utilizing the real-time analytical processing capacity of InfoSphere Streams.

Applications limited only by imagination

As the IBM and Marine Institute team began to map out the possibilities for delivering information and services via the SmartBay portal, more and more potential new uses began to spring up. Stakeholders—the harbormaster, fishermen, researchers, tourism officials and others—were all part of the brainstorming process. The SmartBay vision was quickly expanding far beyond its initial goals.

The variety of applications either deployed or under consideration for SmartBay is strong testament to the power of creative thinking enabled by the right technological tools. The critical element is the ability to analyze, process and present the data in a useful form, tailored to the needs of specific users. For example:

- Technology developers can conduct a variety of sophisticated studies remotely and in near real time, instead of retroactively. Climate researchers, using sensors on land paired with sensors in the bay, can learn about the exchange of CO_2 across the land-sea interface, and marine biologists can use acoustic sensors deployed throughout the bay to assess marine mammal populations.
- Alternative energy developers can access real-time wave data and use it to determine the effectiveness of prototype wave-energy generators, and developers of new sensor technologies can deploy prototypes on the buoys to find out how well the hardware holds up in a harsh marine environment, with continuous monitoring.
- The project can also promote commercial interests. Fishermen can use environmental data to tell them when to put to sea. Fishery managers can monitor and track water quality issues, gaining a comprehensive view of actual conditions throughout the bay.

- Applications developed as part of the SmartBay project can also help increase public safety. Mariners who spot floating objects that pose a hazard to navigation can report the location, and the system will combine this information with geographic data, real-time weather, current, and tide data to predict the path and position of the hazard hours in advance. Collaboration with the Galway harbormaster has also enabled the creation of an expert system based on human expertise that can issue flood warnings more promptly and accurately than he can himself, based on real-time weather, sea state and tidal information.

Gaughan says the project provides a positive benefit in many areas. "The immediate benefits of SmartBay, whether it's helping and supporting industrial development or promoting marine safety, are tangible, direct and worthwhile."

For more information

To learn more about how IBM can help you transform your business, please contact your IBM sales representative or IBM Business Partner.

Visit us at:

- **ibm.com**/government
- **ibm.com**/smarterplanet/water

IBM Corporation
1 New Orchard Road
Armonk, NY 10504
U.S.A.

Produced in the United States of America
November 2010

ODC03150-USEN-00

MediaMath

Category leader innovates with data and analytics

Overview

The need
Best of breed data analytics to enable the largest, most sophisticated ad buyers to get and use all information needed from every aspect of every ad campaign.

Select and deploy new solution within three months with minimal internal resources.

The solution
A purpose-built, high-performance data warehouse appliance that makes advanced analytics on very large data volumes simpler, faster and more accessible.

The benefit
Transparent view of every impression and factor affecting performance of more than 13 billion ad impressions per day; one client achieved campaign goals while reducing CPA from $170 to $80; requires half the manpower to deliver 10x the output with 3x more advertisers, more media channels and clients.

The supply of online display advertising inventory skyrocketed in the mid-2000s with the rapid growth of social media networks, blogs and content-sharing web sites – everything from Facebook, YouTube and Flickr to millions of individual blogs. An urgency to monetize billions of new ad impressions spawned a slew of new ad networks and ad technology innovation that culminated in the creation of the online advertising exchange.

By enabling many networks, publishers and advertisers to connect with one another on a unified, auction-based media trading platform, the ad exchange promised to automate much of the complexity out of the online display advertising process. The hope was that automated impression-level bidding might be just what sellers of display advertising needed to ignite demand from more buyers for their surplus inventory. More exchanges, new networks and tech providers rushed in, initially to help publishers – the "supply side" – generate more revenue from abundant and inexpensive inventory.

Enter MediaMath, a New York City startup with an idea to help ad buyers capitalize on the rapidly evolving marketplace by providing services and tools for ad agencies – the most active "demand side" players – to identify, bid on and buy just those impressions most likely to yield the results they sought for their clients. Implementing the idea in 2007, MediaMath created a hot market segment now known as demand side platforms, or DSP.

Exchanges had also made it easier to use anonymous cookie data and third-party sources to track and reach millions of visitors across the internet. The ensuing avalanche of data started a mind shift among ad buyers away from the simple procurement of commodity impressions to the algorithmic buying of audiences likely to be most receptive to an advertiser's messages. MediaMath needed serious analytical power to allow sophisticated buyers to harness and channel data to drive optimum performance for any advertiser, campaign or marketing objective, using any combination of data inputs. Ultimately the new platform had to ingest and analyze massive volumes of data from multiple sources to make ad optimization and delivery decisions in milliseconds.

"Terabyte-scale data analysis is the new weapon for competitive advantage on Madison Avenue, and the IBM Netezza data warehouse appliance provides MediaMath with the infrastructure to optimize and serve billions of daily impressions – establishing a technical foundation for their long term success in this industry."

—Said Brad Terrell, IBM Vice President and General Manager, Digital Media

Early in the start-up's life, MediaMath used and outgrew MySQL. They tried Oracle Standard Edition with about five terabytes of data and found it wanting. According to Chief Technology Officer Roland Cozzolino, it was difficult to ingest and store data from 50 million daily transactions, let alone handle their growth to 350 million transactions per day. It took "tons of partitions to summarize and break data into vertical buckets by advertiser for analysis," said Cozzolino, and the Oracle platform limited any critical ad hoc analysis capabilities that were required to understand data value and to "gain a horizontal view of the business."

That horizontal view is essential in online ad decision making, and speed is critical to ad agencies responsible for investing their clients' budgets to achieve specific performance objectives. If something in a campaign goes wrong – or if something wildly exceeds expectations – the sooner the buyer knows, the sooner he or she can optimize and adjust how and where to direct the flow of ad dollars. If gaining and acting upon campaign performance insights could be automated substantially, all parties to the transaction would benefit.

Getting to speed-of-thought analysis

Seeking a solution to their data challenges, Tom Craig, VP of Information Strategy at MediaMath, tested databases while Cozzolino tested in-memory coding. Code was fast, but difficult to maintain. They needed something they could deploy quickly, and that would scale to meet the demands of their fast-growing business.

From an initial consideration set that included Aster Data, Hadoop, Infobright, Oracle, Teradata and Vertica, the finalists were Netezza and Greenplum. Cozzolino reported, "We selected Netezza because it offered the best ROI, the fastest time to market of any solution, ease of use and a low total cost of ownership." TCO was key, because there would not be a lot of internal resources available to build applications and support the solution selected. Plus, competition was heating up and MediaMath insisted on bringing the new platform live that quarter with few resources allocated to it.

Craig noted, "We knew where we wanted to take this market, but were unable to execute on that vision with our current tools. Through our proof of concept, Netezza technology clearly demonstrated it provides speed of thought analysis that helps us extend our 'market leader' status." In addition to fast deployment and low resource requirements, MediaMath needed their new solution to enable:

- Flexible reporting, including dashboards and campaign diagnostics
- Capability to use all of their data in the company's proprietary optimization algorithm for fast, thoroughc decision-making

"We knew where we wanted to take this market, but were unable to execute on that vision with our current tools. Through our proof of concept, Netezza technology clearly demonstrated it provides speed of thought analysis that helps us extend our 'market leader' status."

—Tom Craig, VP of Information Strategy, MediaMath

- Special applications such as cross-channel attribution analysis enabling advertisers to gather and de-dupe all user data across display, email, and search, from a multitude of sources and technical platforms
- Internal/financial reporting
- Dynamic interval reach and frequency
- Purchase funnel analysis
- Deep site analysis and classification
- Fraud detection at an IP level
- Near real time (e.g. 15 min) reporting and attribution

The IBM Netezza data warehouse appliance met MediaMath's requirements while adding analytic capabilities that were impossible previously. The IBM Netezza data warehouse appliance could recast and strengthen MediaMath's deliverables because "we could give it more data, and faster," Cozzolino said.

"Before Netezza, we had to be far more deliberate in our data tracking and reporting because our analytic computing power was limited, ," according to Craig, who had worked with Netezza previously at AOL. "With Netezza, we pour through hundreds of millions of rows, with as many dimensions as are available, to look at and consider all the data. This intensive data mining enables better ad decision making. Forget about aggregating and bucketing data. It makes us and our customers smarter. You can pull out all the facts you want."

Ad exchanges provide access to billions of buying opportunities each day. MediaMath has proven to be the best at matching those buying opportunities to their client goals. With the IBM Netezza data warehouse appliance powering the algorithmic trading engine, MediaMath is able to listen to and act on more of those opportunities and drive up campaign performance. "Netezza enables MediaMath to deliver on the promise of impression level bidding in real time," Craig said.

Cozzolino added, "We asked our clients for every possible fact they wanted to see from campaigns running on our platform. The top of the list is simply, 'transparency', visibility into the very granular data exhaust from the buying process. This data could only be delivered with the capacity and capabilities of Netezza." This includes, for example, identifying patterns in the data over much longer time periods to understand true statistical flow. Without the IBM Netezza data warehouse appliance, MediaMath could not view 30 days of data at once. "Now we can look at, for example, what happened 12 months ago and see how it relates to today."

"Before Netezza, we had to be far today: more deliberate in our data tracking and reporting because our analytic computing power was limited. With Netezza, we pore through hundreds of millions of rows, with as many dimensions as are available, to look at and consider all the data. This intensive data mining enables better ad decision making. Forget about aggregating and bucketing data. It makes us and our customers smarter. You can pull out all the facts you want."

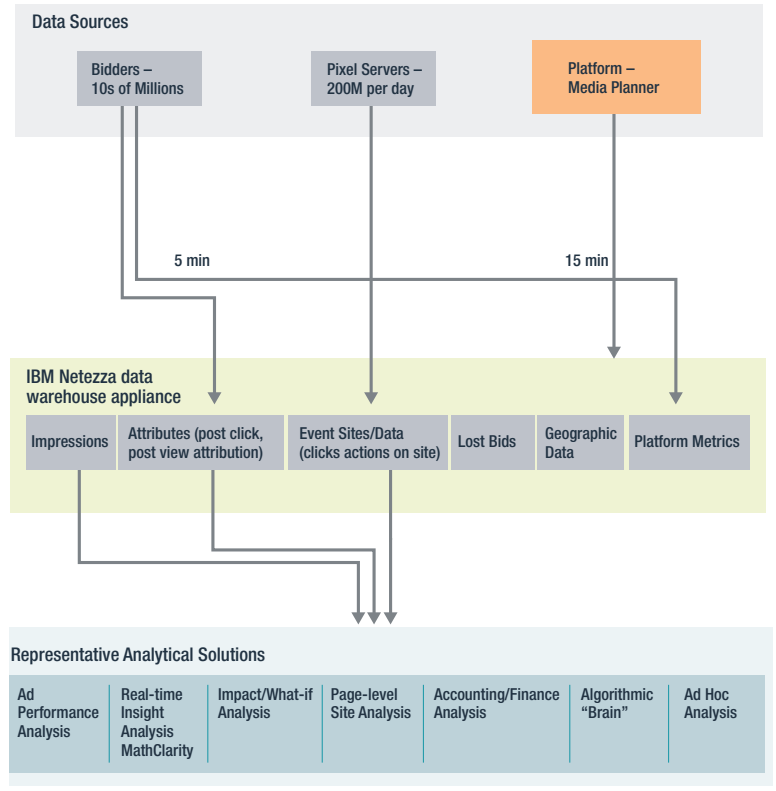
—Tom Craig

On any given day MediaMath may see more than 13 billion impressions (and growing). "We can tell you about every single impression we see every day. Seeing and knowing where every impression originated helps us create the most effective machine learning algorithm and best overall performance of any DSP," according to Cozzolino.

With the IBM Netezza data warehouse appliance, "everything is done in real time," Craig said. "We can value impressions and determine pricing on raw data. We can tell clients if they're seeing a difference in average price or standard deviation or median for different time periods. We couldn't capture and report that data before."

Said Cozzolino, "Netezza today is the hub of everything we do. If I removed it and there was nothing in its place, the only knowledge we would have of what is happening would come from staring at log files. We're the first and largest DSP and we innovate continuously. Netezza reduces cycle times and the learning curve for faster development."

Here is a view of MediaMath's infrastructure:



"We asked our clients for every possible fact they wanted to see from campaigns running on our platform. The top of the list is simply, 'transparency', visibility into the very granular data exhaust from the buying process. This data could only be delivered with the capacity and capabilities of Netezza."

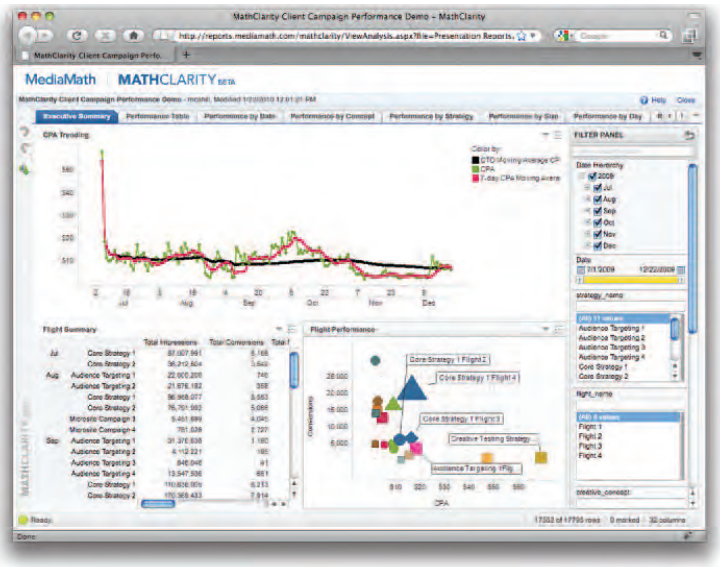
—Roland Cozzolino, Chief Technology Officer, MediaMath

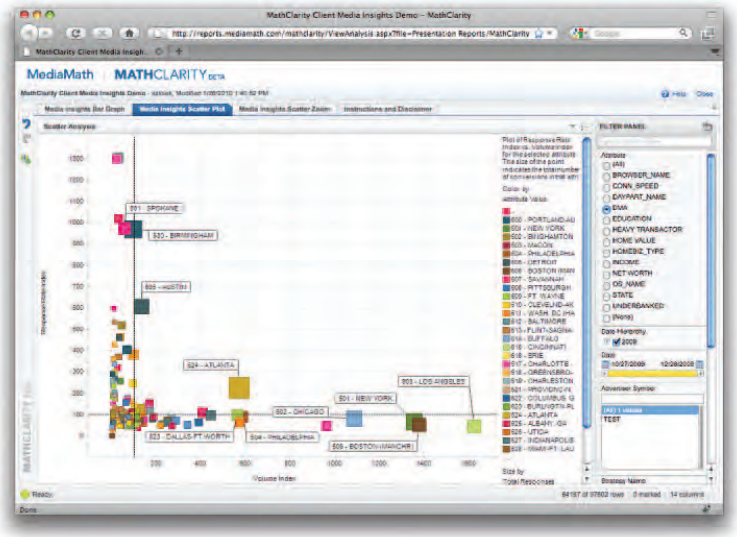
"Everything is done in real time. We can value impressions and determine pricing on raw data. We can tell clients if they're seeing a difference in average price or standard deviation or median for different time periods. We couldn't capture and report that data before."

—Tom Craig

Data and analytics as competitive weapons

The DSP market has become crowded with companies wanting to help advertisers invest an estimated $8 billion in online display-related ad spending (IAB/PriceWaterhouseCoopers Internet Ad Revenue Report, 2009 Full Year Results). MediaMath executives believe that their proven results and unmatched scale have kept them ahead of the pack. Now they are also "the insight DSP," according to Craig, which enables them to continue to set the pace for the category as the one platform that most effectively lets large, sophisticated ad buyers derive maximum value from data. "Our MathClarity product [illustrations below] brings in all the data you want and presents it exactly as you want to consume it. This enables clients to make intelligent databased decisions that improve performance and profits."





"We selected Netezza because it offered the best ROI, the fastest time to market of any solution, ease of use and a low total cost of ownership."

—Roland Cozzolino

A few of MediaMath's actual client use cases and the benefits they've achieved with MediaMath's solution are summarized in the table below.

MediaMath's ambitious roadmap includes more industry firsts that tap the power of the IBM Netezza data warehouse appliance, such as introducing predictive analytics for replaying past events while changing one variable at a time to see how it would have changed the outcome. Said Brad Terrell, IBM Vice President and General Manager, Digital Media, "Terabyte-scale data analysis is the new weapon for competitive advantage on Madison Avenue, and the IBM Netezza data warehouse appliance provides MediaMath with the infrastructure to optimize and serve billions of daily impressions – establishing a technical foundation for their long term success in this industry."

Challenge	Solution	Benefits
Optimization / deaveraging with Client data A longstanding financial services client seeks to further differentiate product offering based on consumer lifetime value (LTV)	• Key LTV scores and indicators passed to MediaMath in real time using MediaMath's flexible customer data integration service for use in optimization • Market basket linkage established to enable integration of client's offline data • Robust predictive and look-alike models developed to extend prospecting audience	• Rapid watermarking period allows client to deaverage a generic conversion and drive the best offer to each consumer • Best customer profiling enabled across exchanges • On a campaign that was already top of plan, the rapid customer data integration resulted in 50 percent + lift in both the upper and lower market campaign performance
Funnel advance rate / cost analysis In a large branding initiative, a CPG client must gain visibility into the advance rate and effective cost of consumers through every stage of the funnel	• User level analysis and advance rate metrics provided to clients at every stage in the lifecycle • Standard process established to provide to all clients with even the most basic conversion process	• True ROI analysis of each stage in the funnel – awareness, familiarity, consideration, and action • Unique user tracking and exposure analysis informs more effective spend • Full transparency into site and key behavioral segments lead advancement through the stages
Trade area exposure and ROI analysis Large CPU client needs to understand the impact of a broad reach campaign to market level while supporting new product rollout in test markets	• Dark market campaign strategy constructed to isolate and compare like markets pre and post campaign • Hundreds of millions of campaign impressions poured over, providing daily access to reach, frequency and gross rating point (GRP) metrics at the zip and trade-area levels • Industry standard market research data integrated to provide SKU level metrics pre and post campaign	• Advertiser receives meaningful reach and audience metrics • Metrics delivered in a meaningful and actionable way using the client's language (e.g. 'trade area') • Client is able to control and monitor reach and frequency on a daily basis to ensure goals are met; integration of market data provides a perfect translation layer across the CPG company's marketing channels

About IBM Netezza data warehouse appliances

IBM Netezza data warehouse appliances revolutionized data warehousing and advanced analytics by integrating database, server and storage into a single, easy-to-manage appliance that requires minimal set-up and ongoing administration while producing faster and more consistent analytic performance. The IBM Netezza data warehouse appliance family simplifies business analytics dramatically by consolidating all analytic activity in the appliance, right where the data resides, for industry-leading performance. Visit **ibm.com/** software/data/netezza to see how our family of data warehouse appliances eliminates complexity at every step and helps you drive true business value for your organization. For the latest data warehouse and advanced analytics blogs, videos and more, please visit: thinking.netezza.com.

About IBM Data Warehousing and Analytics Solutions

IBM provides the broadest and most comprehensive portfolio of data warehousing, information management and business analytic software, hardware and solutions to help customers maximize the value of their information assets and discover new insights to make better and faster decisions and optimize their business outcomes.

For more information

To learn more about the IBM Data Warehousing and Analytics Solutions, please contact your IBM sales representative or IBM Business Partner or visit: **ibm.com**/software/data/netezza

IBM Corporation
Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
December 2011

Please Recycle

IMC14699-USEN-00

Merkle

Helping clients process hours worth of data in minutes

Overview

The need
Influence customer behavior and increase sales with precisely targeted marketing campaigns.

The solution
Using a high-performance data warehouse appliance from IBM for advanced analytics, Merkle is transforming its clients' raw data into unprecedented insight that influences the marketing process and helps staff predict customer preferences with incredible accuracy. It's a strategy that has helped one of the biggest names in business dramatically increase marketing effectiveness and gain a revenue lift of 25-90 percent.

As chief information and technology officer for Merkle—the nation's largest, privately held customer relationship marketing (CRM) agency—Russ Pearlman is responsible for applying the right technologies to help some of the biggest names in the consumer packaged goods, retail, financial services, insurance, nonprofit, and travel and entertainment industries understand and engage their customers and constituents. However, for Pearlman, the focus isn't just on technology. It's on data. Specifically, how to effectively transform petabytes of raw data into useful information that can better influence marketing processes and predict customer preferences with much greater accuracy.

"One of Merkle's biggest value propositions to our clients is applying rich analytics to their marketing problems," says Pearlman. "It's about understanding who their customers (and donors) are, why they are buying, and what their total lifetime value is to our clients — the customers' 'currency'— so that we can encourage better customer acquisition and servicing. The more data that we can provide and the more accurate and timely that data is, the better we can do our job."

Analyzing petabytes of data daily

Like many companies, Merkle's approach to managing data has evolved as the amount of data it must analyze for clients has grown. Today's companies have access to a wealth of data from call centers, point-of-sale systems, web sites, social media sites, email campaigns, newsletters and many other online and offline sources. The challenge lies in the ability to aggregate and analyze this data at lightning speeds so marketers can better target their campaigns and increase the return on their marketing investments.

Business benefits:

- 25-90 percent revenue lift for one client through use of new analytic models
- Regularly received a 70 percent reduction in processing time for complex marketing campaigns—decreasing time from hours to minutes
- Up to 25 percent decrease in the cost of managing clients' environments
- Dashboards and business intelligence reports, such as market basket analysis, can be executed significantly faster
- 50 percent decrease in end-to-end run time for marketing campaign execution—from sample to test to final version

"One of Merkle's biggest value propositions to our clients is applying rich analytics to their marketing problems. The more data that we can provide and the more accurate and timely that data is, the better we can do our job."

– Russ Pearlman, Chief Information and Technology Officer, Merkle

"There are petabytes of data in our data center," says Pearlman. "We have hundreds of millions of new transactions coming in daily and it simply became very difficult to consolidate and analyze all the information for some of our bigger clients using traditional data warehouse environments."

As Merkle looked to transform its analytics infrastructure, staff saw that moving from a traditional data warehouse to a purpose-built data warehouse appliance was critical in handling complex analytics on very large data volumes. Today, the IBM® Netezza® data warehouse appliance is Merkle's "go-to" platform to perform advanced data analytics and execute highly focused marketing campaigns for clients in very short periods of time.

"The IBM Netezza data warehouse appliance has become our standard for MPP [massively parallel processing] technologies," says Pearlman. "It gives us the best value-to-performance ratio of any of the technologies we've used and enables us to help clients target their marketing spend for the best economic outcome."

New analytic models drive revenue lift of up to 90 percent

What do faster analytics on larger data volumes mean to the companies Merkle serves? Consider the firm's work with one leading retailer of clothing, accessories and housewares.

"Operationally, the retailer's database gets refreshed on a nightly basis and helps us provide business users with their latest customer data," says Shambho Krishnasamy, vice president, chief solution architect, Merkle. "Not only can we manage our data load operations within a tight window, but we are also able to create around 1,600 aggregates at the customer and household level for all their brands. This would not have been possible without the IBM Netezza data warehouse appliance. From a marketing operations perspective, the retailer segments its customers at a much finer grain than most companies do. This results

Solution components

- IBM® Netezza® Performance Server
- IBM Netezza 1000
- IBM Unica® Enterprise

"Our clients need solutions that provide them with their latest customer data refreshed on a nightly basis with adequate levels of summarization and aggregation. They expect their ad hoc campaign management, business intelligence and modeling operations to run seamlessly and effectively. A properly sized Netezza environment helps us accomplish that."

– Shambho Krishnasamy, Vice President, Chief Solution Architect, Merkle

in campaign operations that create a lot more concurrent processes—anywhere from 50 to 200 queries running at the same time. The IBM Netezza data warehouse appliance helps us deal with the volume and capture click stream transactions at the lowest level and aggregate it into a single customer view. Combined with IBM Unica® campaign management software, we can now help the brands quickly understand customer preferences and communicate the relevant messages in both online and offline campaigns."

The retailer's business intelligence (BI) platform offers their brands both a comprehensive set of packaged dashboard reports and the ability to run and create reports in an ad-hoc manner. The IBM Netezza data warehouse appliance facilitates execution of concurrent reports more quickly. For example, the traditional market basket and product affinity reports run seven to ten times faster, enabling the retailer to have more timely reports to see consumer behavior.

Additionally, the IBM Netezza data warehouse appliance provides for consistent, reliable and faster execution times for the most complex campaigns and analytic models. This translates into the ability to fine-tune campaigns and customer segment assignments multiple times so that the right audience is selected for the right media for each campaign. In fact, two of the retailer's brands saw revenue lifts ranging from 25 percent to 90 percent through the application of new models.

Smarter Computing by design

A platform for advanced analytics

In delivering its analytic services, Merkle has found that the IBM Netezza data warehouse appliance can process data and run reports faster than the company's previous environment. This, in turn, provides Merkle clients ample time to get their campaigns exactly right. Taking the time to focus on getting the best possible results per campaign enables companies to be more effective in targeting and executing their marketing services.

Designed for Data

By consolidating traditional data warehouse environments (hardware, memory and SAN) on the IBM Netezza data warehouse appliance, Merkle has realized a 25 percent decrease in its IT footprint.

Tuned to the Task

The IBM Netezza data warehouse appliance enabled Merkle to process data 10 times faster and reduce the time to complete nightly data loads by 300 percent, while decreasing the operational maintenance cost by 66 percent. Queries are now completed in seconds and analytic models returned in minutes, as opposed to hours.

Managed for Rapid Service Delivery

With the solution's simplicity, speed and scalability, Merkle can now build precisely targeted campaign lists for its clients in just a few hours and deliver 24x7 availability of information to marketers to support their global operations.

Driving Innovation

Unmatched visibility into customers and their purchasing habits helps Merkle to predict customer preferences with incredible accuracy, increasing campaign returns and maximizing sales.

"From a querying perspective, we can return responses in seconds, and from a modeling perspective, in minutes, which is phenomenal," says Krishnasamy.

Additionally, for one of its clients, this incredible processing power enables the Merkle staff to complete its nightly load of data into the warehouse in less than three hours—a reduction of more than 70 percent. By creating two identical databases on the appliance, and updating each in sequence, the IT staff can deliver 24x7 database availability to global organizations that must provide access to information across many time zones.

"We have clients that need us to create nimble solutions," says Krishnasamy. "From a data platform point of view this translates to us ensuring higher throughput both in terms of operations and marketing execution. Our clients need solutions that provide them with their latest customer data refreshed on a nightly basis with adequate levels of summarization and aggregation. They expect their ad hoc campaign management, business intelligence and modeling operations to run seamlessly and effectively. A properly sized Netezza environment helps us accomplish that."

What's more, this quantum leap in speed enables Merkle to easily support the expanding needs of its clients. "We are now promising clients that we can refresh the data in the warehouse on a nightly basis, regardless of the volume or the type of data," says Krishnasamy. "It gives companies a huge competitive advantage in terms of being able to market to the latest snapshot of their customers. It also enables us to take a batch-oriented marketing solution and evolve that into a real-time operation to support websites and call centers."

Breaking through barriers

The increased performance will also fuel incredibly precise segmentation, modeling and forecasting of marketing opportunities. For example, with the IBM Netezza data warehouse appliance using IBM Netezza Analytics, Merkle can, for the first time, perform in-database analytics, allowing analysts to run algorithms against a client's entire data set in the same timeframe that they previously could analyze only a small subset of the data. As a result, the company can deliver more accurate results faster and create more data models in the same period of time. The business implications of this are potentially massive.

"We can apply it to problems that we just couldn't address with our other technologies," says Pearlman. "For example, we can compare an entire subscriber list from a media source with a client's full customer list to more accurately evaluate the return on investment for an advertising campaign. Consumer purchase transactions now show up in the client's data warehouse in hours as opposed to days—with those customers fully integrated, segmented, analyzed and placed into the appropriate campaigns. This enables companies to send a "thank you" email that also provides the opportunity to up-sell accessories or complementary services."

Reducing the cost of service delivery

With the IBM Netezza data warehouse appliance, Merkle has also experienced a 66 percent decrease in the cost of managing its clients' environments. "If you look at a traditional data warehouse infrastructure, we would need multiple servers and have to manage all the components (the operating systems, storage devices, etc.)," says Pearlman. "It's a more complex environment. Now all we need to think about is network connectivity and power to the appliance. Our overall management costs for an equivalent non-IBM Netezza solution would be two to three times higher."

Lessons learned

According to Pearlman, one of the lessons the team has learned is that training and best practices are important in the move to a new technology infrastructure. To become most effective, Merkle has dedicated resources to managing its IBM Netezza environment and has spent time to properly train and educate them.

"Some people may take a relational database concept and think, 'I now have a relational database appliance,'" says Pearlman. "And that's not what it is. It is a different environment and requires a different mindset. Managing it like a relational database is like buying a Ferrari but always driving in a school zone."

Transforming data into an enterprise-wide asset

By transforming its approach to analytics, Merkle has set a new standard for CRM services. However, as Pearlman looks to the future, he sees even greater opportunity.

"We want to help our clients transform customer information into a truly enterprise-wide asset," says Pearlman. "And by that I mean, it's no longer just a marketing asset—it's a sales asset; a supply-chain asset; a finance asset. So as we go into the future of what Merkle is becoming and what our customers are demanding of us, the information has to be available right away—comprehensive and better integrated across the organization. The architecture we've chosen gives us the capability to meet those needs."

About IBM Netezza data warehouse appliances

The IBM Netezza data warehouse appliance revolutionized data warehousing and advanced analytics by integrating database, server and storage into a single, easy-to-manage appliance that requires minimal set-up and ongoing administration while producing faster and more consistent analytic performance. The IBM Netezza data warehouse appliance family simplifies business analytics dramatically by consolidating all analytic activity in the appliance, right where the data resides, for blisteringly fast performance. Visit netezza.com to see how our family of data warehouse appliances eliminates complexity at every step and lets you drive true business value for your organization. For the latest data warehouse and advanced analytics blogs, videos and more, please visit: thinking.netezza.com

IBM Data Warehousing and Analytics Solutions

IBM provides the broadest and most comprehensive portfolio of data warehousing, information management and business analytic software, hardware and solutions to help customers maximize the value of their information assets and discover new insights to make better and faster decisions and optimize their business outcomes.

About Merkle

Merkle, a customer relationship marketing (CRM) firm, is the nation's largest privately-held agency. For more than 20 years, Fortune 1000 companies and leading nonprofit organizations have partnered with Merkle to maximize the value of their customer portfolios. By combining a complete range of marketing, technical, analytical and creative disciplines, Merkle works with clients to design, execute and evaluate Integrated Customer Marketing™ programs. With more than 1,500 employees, Merkle is headquartered near Baltimore in Columbia, Maryland with additional offices in Boston; Chicago; Denver; Little Rock; Minneapolis; New York; Philadelphia; Pittsburgh; Seattle; Hagerstown, MD and Shanghai. For more information, contact Merkle at 1-877-9-Merkle or visit: www.merkleinc.com

For more information

To learn more about IBM data warehousing and analytics solutions, please contact your IBM sales representative or IBM Business Partner, or visit the following website: **ibm.com**/software/data/netezza

To increase the business value of your IBM data warehouse appliance, participate in an on-line community. Join the IBM Netezza community at: www.enzeecommunity.com

IBM Corporation
Software Group
Route 100
Somers, New York 10589
U.S.A.

Produced in the United States of America
November 2011

Please Recycle

IMC14712-USEN-00

Premier

Helping healthcare providers deliver the best possible care to their patients

Smart is...

Sharing and analyzing healthcare information to help physicians identify the best treatments for their patients

Serving more than 2,600 U.S. hospitals and 84,000-plus other healthcare sites, the Premier healthcare alliance is undertaking a groundbreaking initiative that will help hospitals, doctors and other health providers to identify which treatments benefit patients the most so that no matter where patients live, or are treated, they can be assured the best care. The potential benefits for saving lives, helping people enjoy healthier lives, and reducing healthcare costs are staggering. In one Premier project, 157 participating hospitals saved an estimated 24,800 lives while reducing healthcare spending by USD 2.85 billion.

For Keith Figlioli, senior vice president of healthcare informatics at the Premier healthcare alliance, finding a way to deal with fragmented healthcare information isn't just a job—it's personal.

"When I think about my children, I think about what it will mean for them to live in a society that has solved the complexities of the healthcare system, so that no matter where they live, no matter what they do, no matter what condition they have, they can have the best possible care," says Figlioli.

Premier, which serves more than 2,600 U.S. hospitals and 84,000-plus other healthcare sites, exists to help its members improve the cost and quality of the care they provide the communities they serve. Premier also assists its members to prepare for and stay ahead of health reform, including accountable care and other new models of care delivery and reimbursement.

As Premier executives looked to execute this vision, they recognized that the company's existing technical infrastructure could not support the new model. Over the years, Premier had developed a series of "siloed" applications, making it difficult for members to connect different data sources and metrics and see the "big picture" of how to drive healthcare transformation.

Business benefits

- Offers potentially enormous benefits for saving lives, helping people enjoy healthier lives, and reducing healthcare costs. In one Premier project, 157 participating hospitals saved an estimated 24,800 lives while reducing healthcare spending by USD 2.85 billion.
- Helps providers better identify which treatments will enable their patients to live longer, healthier lives
- Supports Premier members' work to address healthcare reform and other legislative requirements

These platforms and associated software systems also lacked the scalability required to support the massive transaction volumes that were needed. At the same time, as Premier integrates data in new ways, it needs to ensure that the historic high level of data privacy and security is maintained. Moving forward with new technology, Premier had to confirm that it can isolate each healthcare organization's information to continue to meet patient privacy requirements and prevent unauthorized access to sensitive information.

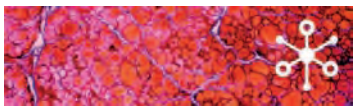
Bridging the information gap

Premier's "re-platforming" effort represents groundbreaking work to enable the sharing and analysis of data from its thousands of member organizations. The new data architecture and infrastructure uses IBM software and hardware to deliver trusted information in the right context at the right time to users based on their roles.

Using the new platform, Premier members will be able to use the portal to access the integrated system for various clinical, business and compliance-related applications. From a clinical aspect, they will have access to best practices from leading hospitals and healthcare experts across the nation and can match patient care protocols with clinical outcomes to improve patient care.

Applications on the new platform will run the gamut from retrospective analysis of patient populations focused on identifying how to reduce readmissions and hospital-acquired conditions to near real-time identification of patients receiving sub-therapeutic doses of an antibiotic.

Business users within the alliance will be able to compare the effectiveness of care locally and with national benchmarks, which will help them improve resource utilization, minimizing waste both in

Smarter Healthcare:	Near real-time insight and collaboration improve patient health
 **Instrumented**	Clinical, operational and supply chain data from member health systems is collected and integrated in near real time
 **Interconnected**	A single data model joins the delivery of healthcare services from patient to provider to payer
 **Intelligent**	Comprehensive analytics—spanning clinical, operational and supply chain data—provide a deeper, more correlated view of cost and quality of care, and thus a broader informational platform for evidence-based clinical practices

Solution components

Software

- IBM Cognos® Business Intelligence
- IBM Connections
- IBM DB2® Advanced Enterprise Server Edition
- IBM DB2 pureScale®
- IBM Enterprise Service Bus for Healthcare
- IBM FileNet® Content Manager
- IBM Health Integration Framework (HIF)
- IBM InfoSphere® Information Server
  - IBM InfoSphere DataStage®
  - IBM InfoSphere QualityStage®
  - IBM InfoSphere Information Analyzer
- IBM InfoSphere Master Data Management
  - IBM Initiate® Exchange
  - IBM Initiate Master Data Service
  - IBM InfoSphere Master Data Management Server
- IBM Netezza® 1000
- IBM Power® 770
- IBM Power 750
- IBM Power 740
- IBM Rational® software
- IBM Tivoli® Identity Manager
- IBM Tivoli Access Manager for e-business
- IBM Web Content Manager
- IBM WebSphere® DataPower®
- IBM WebSphere Business Events
- IBM WebSphere ILOG® JRules
- IBM WebSphere Portal

Services

- IBM Software Services: Information Agenda® Consulting

IBM Business Partner

- Perficient, Inc.

healthcare delivery and in administrative costs. Additionally, this integrated data will help healthcare organizations contract with payors in support of integrated, accountable care.

Premier's commitment to improving healthcare extends beyond its member organizations. As part of its work, it teamed with IBM to create an integrated set of data models and templates that would help other organizations establish a comprehensive data warehouse of clinical, operational and outcomes information. This data model, called the IBM Healthcare Provider Data Warehouse (HCPDW), can help healthcare organizations provide their staff with accurate and timely information to support the delivery of evidence-based, patient-centric and accountable care.

Journey to Smarter Computing

Fundamental to helping Premier turn its vision into reality is an Information Agenda® strategy that transforms information into a strategic asset that can be leveraged across applications, processes and decisions.

"In its simplest form, Premier's platform brings together information from all areas of the healthcare system, aggregates it, normalizes it, and benchmarks it, so it impacts performance while the patient is still in the hospital or the physician's office," says Figlioli.

With assistance from IBM and IBM Premier Business Partner, Perficient, Inc., the Premier healthcare alliance developed a completely new IT strategy, using a software as a service (SaaS) model based on the IBM Health Integration Framework.

"We wanted a flexible, nimble partner because this is not a cookie-cutter kind of project," says Figlioli. "Premier and IBM brought to the table an approach that was best of breed and included a cultural and partnering dimension that was fundamentally different from other vendors."

"When I think about my children, I think about what it will mean for them to live in a society that has solved the complexities of the healthcare system so that no matter where they live, no matter what they do, no matter what condition they have, they can have the best possible care."

—Keith Figlioli, Senior Vice President, Healthcare Informatics, Premier

The organization's IT division is building its new infrastructure from the ground up. This includes replacing its existing x86 servers from a variety of hardware vendors with IBM POWER7® processor-based systems to gain greater performance at a lower cost. In fact, an early pilot showed up to a 50 percent increase in processing power with a reduction in costs.

Additionally, the company is moving its core data warehouse to IBM DB2® pureScale, which is highly scalable to support the growing amount of data that Premier is collecting from its members. As part of Premier's platform, DB2 pureScale will help doctors gain the information they need to avoid patient infections that are common in hospitals, and will help pharmacists ensure safe and effective medication use.

Data from facility admission, discharge and transfer (ADT) systems along with departmental systems, such as pharmacy, microbiology and lab information systems, will be sent to Premier's core data warehouse as HL7 messages, with near real-time processing of this data occurring at a rate of 3,000 transactions per second. With the high performance DB2 data software provides, Premier members can quickly learn of emerging healthcare issues, such as an increased incidence of MRSA (a highly drug-resistant version of staphylococcus aureus bacteria) in a particular area.

Data from IBM DB2 database software will be loaded into the IBM Netezza® data warehouse appliance to enable members to conduct advanced analytics faster and easier than was previously possible. IBM Cognos® Business Intelligence will be used to help members identify and analyze opportunities and trends across their organizations.

IBM InfoSphere® software is used to acquire, transform and create a single, trusted view of each constituent or entity. The data is then integrated and validated, and clinical or business rules management is applied through WebSphere® ILOG® software. These rules can help automatically notify clinicians of critical issues, such as the appropriate dosing of anti-coagulation medication.

"In its simplest form, Premier's platform brings together information from all areas of the healthcare system, aggregates it, normalizes it, and benchmarks it, so it impacts performance while the patient is still in the hospital or the physician's office."

—Keith Figlioli

IBM Tivoli® software provides security and service management. Application development is built upon Rational® software and a common user experience and collaboration are provided through IBM Connections software.

The inside story: Getting there

To help bring its members closer to an accountable, coordinated care model, Premier needed to position itself not for tomorrow, but for three, five and even 10 years into the future. And this vision required a new way of thinking.

Over the next five years, Premier plans to provide its members with many new applications in support of healthcare reform and other legislative requirements. As capabilities are added, the SaaS model will enable the platform to support its commitment to keep all 2,600 hospital members on the same page, and even expand its user community. With its new approach, Premier IT staff can develop, test and launch new applications from a central location to provide users with updates concurrently. This is a lower-cost way to give Premier members an analytics solution with a shorter time to value.

For more information

To learn more about how IBM can help you transform your business, please contact your IBM sales representative or IBM Business Partner.

For more information about IBM's Information Agenda, visit: **ibm.com**/software/data/information-agenda

For more information about Perficient, visit: www.perficient.com

For more information about the Premier healthcare alliance, visit: www.premierinc.com

IBM Corporation
Software Group
Route 100
Somers, NY 10589 U.S.A.

Produced in the United States of America
June 2012

Please Recycle

IMC14754-USEN-00

Large Gene Interaction Analytics at University at Buffalo, SUNY

Giving researchers the ability to speed computations and increase data sets

Overview

The need

Researchers required the ability to quickly build models using a range of variable types and run them on a high-performing environment on huge data sets.

The solution

The solution helped the SUNY Buffalo researchers consolidate all reporting and analysis in one location to improve the efficiency, sophistication and impact of their research.

The benefit

The SUNY researchers were able to reduce the time required to conduct analysis from 27.2 hours without the IBM Netezza data warehouse appliance to 11.7 minutes with it.

The State University of New York (SUNY) at Buffalo is home to one of the leading multiple sclerosis (MS) research centers in the world. MS is a devastating, chronic neurological disease that affects nearly one million people worldwide. The disease causes physical and cognitive disabilities in individuals and is characterized by inflammation and neuro-degeneration of the brain and spinal cord. From the beginning, the genetics of MS were known to be complex and it was apparent that no single gene was likely causative for the disease.

Since 2007, the SUNY team has been looking at data obtained from scanned genomes of MS patients to identify genes whose variations could contribute to the risk of developing MS. New technologies now enable hundreds of thousands of genetic variations, called single nucleotide polymorphisms (SNPs), to be obtained from single samples. According to research lead, Dr. Murali Ramanathan, a critical fact in the study of MS is that “gene products work by interacting with both other gene products and environmental factors.”

"There are lots of good reasons to use Revolution Analytics in an analytics appliance super computer like Netezza. It's faster and easier to program. It will speed up our computation. And, we have more data sets available to us because of how flexible and quick Revolution Analytics makes it to add and delete variables in our model."

—Dr. Murali Ramanathan, University at Buffalo, State University of New York

"Because of this, researchers have postulated that multiple SNPs—combined with environmental variables—would better explain the risk of developing MS. Finding these combinations is analogous to the proverbial search for a needle in a haystack," says Dr. Ramanathan. Identifying a candidate's environmental factors that could be used to prevent the disease from progressing in patients was of great importance. Examples of this include sun exposure and vitamin D levels, Epstein-Barr virus infection and smoking.

The researchers needed to create algorithms that would efficiently identify interactions and attempt 'parsimony'—an extreme efficiency. The team was looking for the fewest number of SNPs, environmental and phenotypic variable combinations that would help explain the presence of MS.

The researchers at the State University of New York developed an approach they called AMBIENCE that is distinctive in its use of new information theoretic methods along with its versatility and scalability. The theoretic underpinnings of AMBIENCE enable the detection of both linear and non-linear dependencies in the data. The AMBIENCE algorithm is capable of conducting an efficient search of the large combinatorial space because of the unique nature of the information theoretic metrics, which allows for greedy search identification of the most promising combinations.

Solution components

- IBM Netezza® 1000-24
- IBM Netezza Analytics Appliance

IBM Business Partner

- Revolution R Enterprise for IBM Netezza

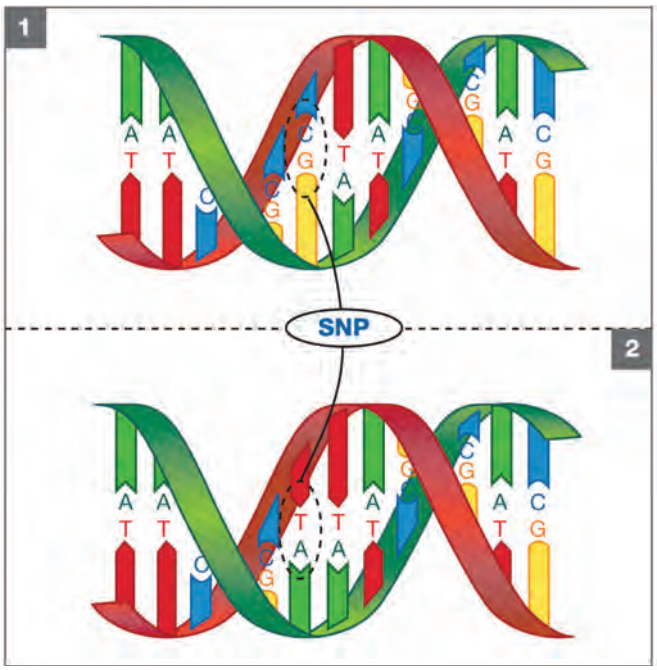


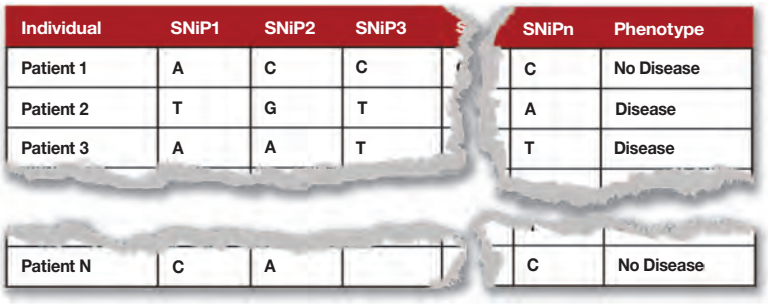
Figure 1: A single nucleotide polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide—A, C, G or T—in the shared genome differs in an individual. If many individuals with a disease (or 'phenotype') have the same polymorphisms that healthy individuals do not, this may be a clue to finding out genetic clues to what causes a disease (Wikipedia).

The need

The data sets used in this type of multi-variable research are very large and the analysis is computationally very demanding because the researchers are looking for significant interactions between thousands of genetic and environmental factors. There are two issues to overcome: crunching through the immense data set and building analytic models that allow the team to look at more than simply first order interactions. The researchers want to see not only which variable is significant, but also which pairs of variables or which three variables are significant. This requires the ability to quickly build models using a range of variable types and run them on a high-performing environment on huge data sets. It also requires the ability to include an almost limitless variety of dependent variable types.

The computational challenge in gene-environmental interaction analyses is due to a phenomenon called 'combinatorial explosion.' Considering that there are thousands of SNPs, the number of combinations of SNPs that have to be assessed for uncovering potential interaction becomes incredibly large.

Before they could run the analysis using all of the variables presented in the data set, the researchers needed an analytic framework that would allow them to add and remove variables from the model quickly and easily, without having to write hundreds of lines of code. The sheer number of SNPs combined with environmental variables and phenotype values mean that the amount of computations necessary for data mining could number in the quintillions (18 zeros).



Individual	SNiP1	SNiP2	SNiP3	SNiPn	Phenotype
Patient 1	A	C	C	C	No Disease
Patient 2	T	G	T	A	Disease
Patient 3	A	A	T	T	Disease
Patient N	C	A		C	No Disease

Figure 2: A heuristic example of the SUNY data table

The solution

SUNY Buffalo is using Revolution R Enterprise for IBM Netezza® in conjunction with the IBM Netezza Analytics Appliance to dramatically simplify and speed up very complex analysis on large data sets.

The organization's researchers knew they needed the level of processing only available through a high-performance computer (HPC). They also needed capabilities found in relational databases—not often included

in HPC platforms. HPCs typically run their processes only in parallel, breaking up calculations to run simultaneously across multiple processors. In addition, they also often used field-programmable gate array (FPGA) architectures for additional speed.

The IBM Netezza data warehouse appliance offered the features and performance the researchers needed and much more. From an analytics perspective, the SUNY Buffalo researchers were able to write a version of their software tools in Revolution R Enterprise for IBM Netezza, so that all their reporting and analysis were consolidated in one location. This prevented the need to move very large amounts of data in and out of the IBM Netezza data warehouse appliance, which would cause further delays. They were also able to use a wider variety of data sets.

Due to the nature of SUNY Buffalo's research work, there is immense value in now being able to use a wide variety of data sets to study the interactions among a greater range of variables. The solution improved the efficiency, sophistication and impact of the research. SUNY's team adopted R—the core of Revolution R Enterprise for IBM Netezza—because it offered the flexibility to include diverse types of variables such as categorical, discrete Poisson-dependent, or continuous normally-distributed variables by simply adding a few lines of code. In the past, the SUNY team would have to rewrite entire algorithms, requiring a great deal of staff time. Now, scientists can change the algorithm themselves by adding a new entropy function, and focus less on writing algorithms and more on the science.

The benefit

Once SUNY Buffalo deployed the IBM Netezza data warehouse appliance as its research analytics infrastructure along with Revolution R Enterprise for IBM Netezza, and the genetic data were assembled, the environmental and phenotype data were combined, and the algorithms were customized, the researchers were empowered to look for potential factors contributing to the risk of developing MS.

The SUNY researchers were able to:

- Use the new algorithms and add multiple variables that before were nearly impossible to achieve.
- Reduce the time required to conduct analysis from 27.2 hours without the IBM Netezza data warehouse appliance to 11.7 minutes with it.
- Carry out their research with little to no database administration. (Unlike other HPC platforms or databases available, the IBM Netezza data warehouse appliance was designed to require a minimum amount of maintenance.)
- Publish multiple articles in scientific journals, with more in process.
- Proceed with studies based on 'vector phenotypes'—a more complex variable that will further push the IBM Netezza data warehouse appliance platform.


AMBIENCE: A Novel Approach and Efficient Algorithm for Identifying Informative Genetic and Environmental Associations With Complex Phenotypes

Figure 3: One of the articles published by SUNY researchers

For more information

To learn more about the IBM Netezza data warehouse appliances, please contact your IBM representative or IBM Business Partner, or visit the following website:
ibm.com/software/data/netezza

For more information about Revolution Analytics, visit www.revolutionanalytics.com

For more information on SUNY Buffalo, visit: www.buffalo.edu/, or for more information on SUNY's research, consult the following sample publications:

AMBIENCE: a novel approach and efficient algorithm for identifying informative genetic and environmental associations with complex phenotypes. Chanda P, Sucheston L, Zhang A, Brazeau D, Freudenheim JL, Ambrosone C, Ramanathan M. Genetics. 2008 Oct; 180(2): 1191-210. Epub 2008 Sep 9. PMID: 18780753 [PubMed - indexed for MEDLINE]

Information-theoretic gene-gene and gene-environment interaction analysis of quantitative traits. Chanda P, Sucheston L, Liu S, Zhang A, Ramanathan M. BMC Genomics. 2009 Nov 4; 10:509. PMID: 19889230.

Comparison of information-theoretic to statistical methods for gene-gene interactions in the presence of genetic heterogeneity. Sucheston L, Chanda P, Zhang A, Tritchler D, Ramanathan M. BMC Genomics. 2010 Sep 3; 11:487. PMID: 20815886.

The interaction index, a novel information-theoretic metric for prioritizing interacting genetic variations and environmental factors. Chanda P, Sucheston L, Zhang A, Ramanathan M. Eur J Hum Genet. 2009 Oct;17(10): 1274-86. Epub 2009 Mar 18. PMID: 19293841.

Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
April 2012

Please Recycle

IMC14675-USEN-01

TEOCO

Helping communication service providers (CSPs) save hundreds of millions of dollars

Overview

The need

TEOCO wanted to analyze 500 TB of data from call detail records and inter-carrier invoices daily to help communication service providers (CSPs) identify cost savings and improve services.

The solution

TEOCO's assurance and analytics solutions, powered by the IBM Netezza data warehouse appliance, enable CSPs to access and analyze massive amounts of data to uncover the source of cost and network issues.

The benefit

Helped one Tier 1 mobile operator save over USD400 million in four years; reduced management costs by decreasing server sprawl; enabled clients to proactively respond to network issues before customer service was affected.

When industry experts talk about Big Data, they typically discuss the three "Vs" of data: volume, velocity and variety. But ask Jonjie Sena, Senior Director of Product Management at TEOCO about Big Data, and his list includes a fourth element: value.

"Value is where we come from," says Sena. "It's about making sure that we apply those bits and bytes so companies gain meaningful feedback that solves business issues. Because of this, most of our customers see value in the first 100 days following implementation of our solution."

TEOCO service assurance, customer analytic and margin assurance systems help international communication service providers (CSPs) better manage their costs, expenses and revenue structures. With TEOCO's solutions, CSPs can quickly identify unnecessary costs, optimize call routing through real-time traffic management, and proactively respond to network issues before they affect customer service. More than 100 of the top CSPs worldwide have selected TEOCO, an IBM Business Partner, to be their analytics and assurance solutions provider. TEOCO's solution is frequently used to identify, and help dispute, millions of dollars in inaccurate billings between telecommunications operators.

"We can view all the network data and can validate it to make sure that our clients' costs are correct and the corresponding revenue is accurate," says Sena. "And as we analyze the costs, we can also identify and recommend alternate means, alternate partners, and alternate business models that can increase revenue."

Designed for Data

By offering its clients an all-in-one assurance and analytics solution, powered by the IBM Netezza data warehouse appliance, TEOCO enables communication service providers to access and analyze the massive amounts of data required to see emerging cost issues that are typically not visible at the summary levels.

Tuned to the Task

Consolidating processes on an optimized and integrated appliance helped TEOCO gain greater performance and efficiency for each workload—a move that enables the company to capture, load, enrich, process and analyze 500 TB of data daily for its clients.

Managed with cloud technologies

The platform's simplicity enables TEOCO to manage approximately 24 racks with only two full-time administrators, allowing IT staff to focus on delivering new business applications instead of database optimization efforts.

Driving Innovation

With the speed, simplicity and scalability of its IT environment, TEOCO can provide its clients a 3 - 5 times return on investment (and greater)—with savings often over USD100 million.

For example, one large operator uses TEOCO's solution for transaction-level costing and accruals, daily and monthly gross margin forecasting, and analysis by product and partner for all transactions on its network. This allows the operator to react when actuals are trending outside of projections and correct it immediately before it has a negative business (or customer) impact.

Another operator uses TEOCO for M2M (Machine-to-Machine) and Device Analytics for its emerging devices such as e-readers, GPS, telematics, and other non-handset devices. By looking at the usage behaviors, TEOCO generates information used for both network planning as well as financial forecasting, and forms the basis of the settlements between many content partners.

A third example: TEOCO's solutions are used to reconcile and forecast usage, revenues and costs of mobile virtual network operator (MVNO) subscribers. MVNOs represent a significant amount of traffic and revenue for this operator, but the available mix of services may be quite different from the operator's direct subscribers. TEOCO's analytics help optimize current and future offers for MVNO customers by helping ensure the offered services remain profitable.

The financial impact is enormous. "As part of the Proof of Concept with this client, we identified a potential USD100 million savings as a direct result of our solution," says Sena. "And this savings was realized in just three months. We helped another client save more than USD400 million in four years and we expect, in the next year, the savings to exceed a half-billion dollars."

Solution components
- IBM Netezza® 1000

"We helped another client save more than USD400 million in four years and we expect, in the next year, the savings to exceed a half-billion dollars."

—Jonjie Sena, Senior Director, Product Management, TEOCO

Analyzing Big Data

To accurately identify the real cost savings within any network, operators need a 360-degree view of their network-to-bill lifecycle with several months of usage history at their fingertips. Gaining this view requires processing and analyzing a staggering amount of data from usage detail records and complex settlement invoices from ecosystem partners.

"Typically, we're analyzing approximately 500 terabytes of data from call detail records daily," says Sena. "You make a phone call, you get a record; you send a text message, you get a record; you click on a web page; you get a record. As those calls are set up, you get additional records. And each of these actions generates some type of network event as well. We're looking at all this data to help them not only make sense of it, but take action based on it."

According to Sena, processing this data in a traditional data warehousing environment would not be possible. TEOCO offers its clients an all-in-one appliance, powered by the IBM Netezza® data warehouse appliance, that makes it possible for CSPs to access and analyze the massive amounts of data required to see cost issues that are typically not visible at the summary levels offered by TEOCO's competitors.

"We delivered our applications on the very first Netezza appliance [now IBM Netezza data warehouse appliance]," says Sena. "It provides us with the speed, parallel processing and scalability we need to deliver these business benefits to our clients. We have a billion records coming in each day, but because we have to present it in many ways, we may be processing 10 billion records daily."

Essential for making this possible is the platform's ability to support in-database analytics, relieving the need to move data back and forth between a database and separate analytics applications.

"The IBM Netezza data warehouse appliance has helped us gain greater performance and efficiency for each workload and simplified our lives."

—Jonjie Sena

"In traditional systems, you get the data, you transform it, you load it, you pull it out, you analyze it, and then you stick it back in," says Sena. "With the IBM Netezza data warehouse platform, we can perform in-database analytics and in-database processing, which means that all the work happens within the box itself. This gives us the ability to process billions of records multiple times per customer, per device, per network, per region, and so on. One of the questions I am usually asked is 'how much faster is this over alternative approaches?' From our perspective, this may not be possible with other approaches because of the volume of data and the number of in-depth domain processes that we have to apply to each record."

Keeping infrastructure costs low

According to Sena, using a traditional data warehousing approach to process tens of billions of records daily also would have resulted in considerable "server sprawl"—a huge concern for any company.

"Traditionally, you had a lot of servers, a lot of discs eating up storage area, which would drive costs up," says Sena. "The IBM Netezza data warehouse appliance has solved that challenge. It has helped us gain greater performance and efficiency for each workload and simplified our lives. We have approximately 24 racks [of Netezza equipment] in our environment and we only need two full-time people to manage the environment. For the kind of data volumes we're processing, that's just not possible using traditional database methodologies."

About IBM Netezza data warehouse appliances

IBM Netezza data warehouse appliances revolutionized data warehousing and advanced analytics by integrating database, server and storage into a single, easy-to-manage appliance that requires minimal set-up and ongoing administration while producing faster and more consistent

"We have approximately 24 racks [of Netezza equipment] in our environment and we only need two full-time people to manage the environment. For the kind of data volumes we're processing, that's just not possible using traditional database methodologies."

—Jonjie Sena

analytic performance. The IBM Netezza data warehouse appliance family simplifies business analytics dramatically by consolidating all analytic activity in the appliance, right where the data resides, for industry-leading performance. Visit: **ibm.com**/software/data/netezza to see how our family of data warehouse appliances eliminates complexity at every step and helps you drive true business value for your organization. For the latest data warehouse and advanced analytics blogs, videos and more, please visit: thinking.netezza.com

About IBM Data Warehousing and Analytics Solutions

IBM provides the broadest and most comprehensive portfolio of data warehousing, information management and business analytic software, hardware and solutions to help customers maximize the value of their information assets and discover new insights to make better and faster decisions and optimize their business outcomes.

For more information

To learn more about IBM Data Warehousing and Analytics Solutions, please contact your IBM sales representative or IBM Business Partner or visit: **ibm.com**/software/data/netezza

To increase the business value of your IBM data warehouse appliance, participate in an online community. Join the IBM Netezza community at: www.enzeecommunity.com

For more information about TEOCO, visit: www.teoco.com

IBM Corporation
Software Group
Route 100
Somers, NY 10589 U.S.A.

Produced in the United States of America
May 2012

The content in this document (including currency OR pricing references which exclude applicable taxes) is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions. It is the user’s responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs. THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided. TEOCO’s solution is not an IBM product or offering. TEOCO’s solution is sold or licensed, as the case may be, to users under TEOCO’s terms and conditions, which are provided with the product or offering. Availability, and any and all warranties, services and support for TEOCO’s solution are the direct responsibility of, and are provided directly to users by TEOCO.

Actual available storage capacity may be reported for both uncompressed and compressed data and will vary and may be less than stated.

Please Recycle

IMC14753-USEN-00

TerraEchos

Streaming data technology supports covert intelligence and surveillance sensor systems

Overview

The need
A U.S. Department of Energy National Lab needed a solution to detect, classify, locate and track potential threats to secure its perimeters and border areas.

The solution
IBM Business Partner, TerraEchos, implements an advanced security and covert surveillance system based on the TerraEchos Adelos S4 System with IBM InfoSphere Streams and an IBM System x3650 server.

The benefit
Reduces time to process 275 Mbit of data from hours to just one-fourteenth of a second; captures and analyzes huge volumes of data in real time, providing unprecedented insight to detect, classify, locate, track and deter potential threats.

A leading provider of covert intelligence and surveillance sensor systems, TerraEchos, Inc., provides organizations with advanced security solutions for critical infrastructures and extended borders. One TerraEchos client is a science-based, applied engineering national laboratory dedicated to supporting the mission of the U.S. Department of Energy (DOE) in nuclear and energy research, science and national defense.

Securing the scientific intelligence, technology and resources related to these initiatives is vital. To this end, this national lab recognized the need for a technology solution that would detect, classify, locate and track potential threats—above and below ground—to secure its perimeters and border areas. This solution would provide lab personnel and security staff with more situational awareness and enable a faster and more intelligent response to any threat.

Detecting and analyzing a wide range of sounds—even from miles away

The requirements of the ideal solution were considerable. The solution would have to continuously consume and analyze massive amounts of digital acoustic data from biological, mechanical and environmental objects-in-motion. In addition, because lab personnel lacked time to record the data and listen to it later, the solution had to gather and analyze information simultaneously.

"We are capturing and reducing a tremendous amount of digital acoustic data, which is what InfoSphere Streams does well, and we are running intensive computational statistical analysis, all in one-fourteenth of a second."

—Dr. Alex Philp, Founder and CTO, TerraEchos, Inc.

The analysis could extract meaningful intelligence, as well as verify and validate the data, such as distinguishing between the sounds of a trespasser versus a grazing animal. To put the sophistication of the needed technology into perspective, the data consumption and analytical requirements would be akin to listening to 1,000 MP3 songs simultaneously and successfully discerning the word "zero" from every song—within a fraction of a second.

The solution would also serve as the lab's central nervous system and would have to meet strict technical requirements, including:

- Interoperability, enabling lab personnel to collect and analyze an array of data from video, acoustic, and other types of sensors to create a holistic view of a situation
- Scalability to support new requirements as the lab's fiber-optic arrays, surveillance areas, and security perimeters change
- Extensibility, serving as a framework to fit into the lab's existing IT architecture and integrating with signal processors and mobile and mapping applications

Advanced fiber-optics combine with real-time streaming data

To meet these requirements, the lab turned to IBM Business Partner, TerraEchos, to implement an advanced, covert security and surveillance system, based on the TerraEchos Adelos® S4 System, an IBM System x® 3650 server and IBM InfoSphere® Streams software, part of the IBM big data platform. The TerraEchos Adelos S4 solution offers advanced fiber-optic acoustic sensor technology licensed from the United States Navy. InfoSphere Streams is the engine that processes digital acoustic data-in-motion continuously from fiber-optic sensor arrays.

Solution components

Software
- IBM InfoSphere® Streams

Servers
- IBM System x® 3650

"Given the data processing and analytical challenges addressed using our Adelos Sensor Array, InfoSphere Streams is the right solution for us and our customers. We look forward to growing our strategic relationship with IBM across various sectors and markets to help revolutionize the concept of 'Sensor as a Service.'"

—Dr. Alex Philp

Serving as the underlying analytics platform, the processing capacity of InfoSphere Streams enables the Adelos S4 solution to analyze and classify streaming acoustic data in real time. InfoSphere Streams collects data from multiple sensor types and enables associated streams of structured and unstructured data to be integrated into an intelligence system for threat detection, classification, correlation, prediction and communication by means of a service oriented architecture (SOA). Based on this technology, TerraEchos provides one of the most robust surveillance classification systems in the industry, and is the first fiber-optic sensor company to incorporate InfoSphere Streams as the computational platform for sensor data analytics.

"InfoSphere Streams captures and feeds approximately 275 Mbit of acoustic data from 1,024 individual sensor channels into the Adelos S4 solution, where a computationally intensive Bayesian Data Reduction Algorithm performs calculations to classify the data in real time," says Dr. Alex Philp, Founder and CTO of TerraEchos. "We are capturing and reducing a tremendous amount of digital acoustic data, which is what InfoSphere Streams does well, and we are running intensive computational statistical analysis, all in one-fourteenth of a second."

"Before implementing our solution, many of our government customers required hours to process 275 Mbit of data, not in one-twelfth or one-fourteenth of a second," Dr. Philp adds. "Our solution is deployed on a single IBM System x3650 server, running eight cores and capable of processing up to 42 terabytes every day, which would be costly to store. The cost-benefit ratio for using InfoSphere Streams to process and analyze streaming data in real time is phenomenal."

Extending the security perimeter creates a strategic advantage

Because the solution captures and transmits real-time, streaming acoustical data from around the lab premises, security staff has unprecedented insight into any event. The system enables lab and security personnel to "hear" what is going on—even when the disturbance is miles away. In this way, it is possible to confidently identify and classify a potential security threat—and take appropriate action.

"We use the fiber-optic cable as a sensing array," says Dr. Philp. "The listening devices are actually virtual segments of the cable, so think of a cable a mile long. We break it down digitally into individual microphones or individual listening areas of one meter, and these distances can change. The beauty of this extended perimeter security system is that it is completely passive. Using miles of fiber-optic cables and thousands of listening devices buried underground, the lab can extend its perimeter security and gain a strategic advantage."

Correlating sensor data delivers a zero false-positive rate

The solution is part of a more comprehensive security system. With the ability to integrate and collect data from video and airborne surveillance systems, lab personnel gain a holistic view of potential threats and issues—or nonissues.

"In addition to detection, classification, localization and tracking, correlating the analysis from acoustic sensors and video cameras provides for verification and validation and a zero false-positive rate," says Dr. Philp. "With these results, security staff can make confident decisions about responding to a threat—such as how many officers to deploy and which tactics to use—and can also thwart any plans intruders have to breach the property."

Finally, in addition to meeting the lab's requirements for extensibility, interoperability and scalability, the solution saves the lab costs associated with data storage because data does not have to be stored before being analyzed. "Capturing approximately 42 terabytes of data each day adds up fast and would be challenging and costly to store," says Dr. Philp. "InfoSphere Streams offers advantages, especially when you have to capture data continuously in real time and analyze it as it passes by. Organizations can realize huge savings in storage."

"Given the data processing and analytical challenges addressed using our Adelos Sensor Array, InfoSphere Streams is the right solution for us and our customers," notes Dr. Philp. "We look forward to growing our strategic relationship with IBM across various sectors and markets to help revolutionize the concept of 'Sensor as a Service.'"

For more information

To learn more about IBM InfoSphere Streams, please contact your IBM sales representative or IBM Business Partner, or visit the following website: **ibm.com**/software/data/infosphere/streams

To learn more about IBM big data, visit: **ibm.com**/software/data/bigdata

To get involved in the conversation:
www.smartercomputingblog.com/category/big-data

For information on TerraEchos, Inc. visit: http://www.terraechos.com

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
May 2012

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions. It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs. THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided. TerraEchos Adelos S4 System is not an IBM product or offering. Adelos S4 System is sold or licensed, as the case may be, to users under TerraEchos' terms and conditions, which are provided with the product or offering. Availability, and any and all warranties, services and support for Adelos S4 System is the direct responsibility of, and is provided directly to users by, TerraEchos.

Actual available storage capacity may be reported for both uncompressed and compressed data and will vary and may be less than stated.

Please Recycle

IMC14726-USEN-01

University of Ontario Institute of Technology

Leveraging key data to provide proactive patient care

Overview

The need

To better detect subtle warning signs of complications, clinicians need to gain greater insight into the moment-by-moment condition of patients.

The solution

A first-of-its-kind, stream-computing platform was developed to capture and analyze real-time data from medical monitors, alerting hospital staff to potential health problems before patients manifest clinical signs of infection or other issues.

What makes it smarter

Early warning gives caregivers the ability to proactively deal with potential complications—such as detecting infections in premature infants up to 24 hours before they exhibit symptoms.

The rapid advance of medical monitoring technology has done wonders to improve patient outcomes. Today, patients are routinely connected to equipment that continuously monitors vital signs such as blood pressure, heart rate and temperature. The equipment issues an alert when any vital sign goes out of the normal range, prompting hospital staff to take action immediately, but many life-threatening conditions do not reach critical level right away. Often, signs that something is wrong begin to appear long before the situation becomes serious, and even a skilled and experienced nurse or physician might not be able to spot and interpret these trends in time to avoid serious complications.

Unfortunately, the warning indicators are sometimes so hard to detect that it is nearly impossible to identify and understand their implications until it is too late. One example of such a hard-to-detect problem is nosocomial infection, which is contracted at the hospital and is life threatening to fragile patients such as premature infants.

According to physicians at the University of Virginia,[1] an examination of retrospective data reveals that, starting 12 to 24 hours before any overt sign of trouble, almost undetectable changes begin to appear in the vital signs of infants who have contracted this infection. The indication is a pulse that is within acceptable limits, but not varying as it should—heart rates normally rise and fall throughout the day. In a baby where infection has set in, this doesn't happen as much and the heart rate becomes too regular over time. So, while the information needed to detect the infection is present, the indication is very subtle; rather than being a single warning sign, it is a trend over time that can be difficult to spot, especially in the fast-paced environment of an intensive care unit.

Business benefits

- Holds the potential to give clinicians an unprecedented ability to interpret vast amounts of heterogeneous data in real time, enabling them to spot subtle trends
- Combines physician and nurse knowledge and experience with technology capabilities to yield more robust results than can be provided by monitoring devices alone
- Provides a flexible platform that can adapt to a wide variety of medical monitoring needs

The monitors continuously generate information that can give early warning signs of an infection, but the data is too large for the human mind to process in a timely manner. Consequently, the information that could prevent an infection from escalating to life-threatening status is often lost.

"The challenge we face is that there's too much data," says Dr. Andrew James, staff neonatologist at The Hospital for Sick Children (SickKids) in Toronto. "In the hectic environment of the neonatal intensive care unit, the ability to absorb and reflect upon everything presented is beyond human capacity, so the significance of trends is often lost."

Making better use of the data resource

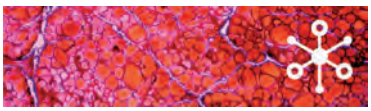
The significance of the data overload challenge was not lost on Dr. Carolyn McGregor, Canada Research Chair in Health Informatics at the University of Ontario Institute of Technology (UOIT). "As someone who has been doing a lot of work with data analysis and data warehousing, I was immediately struck by the plethora of devices providing information at high speeds—information that went unused," she says. "Information that's being provided at up to 1,000 readings per second is summarized into one reading every 30 to 60 minutes, and it typically goes no further. It's stored for up to 72 hours and is then discarded. I could see that there were enormous opportunities to capture, store and utilize this data in real time to improve the quality of care for neonatal babies."

With a shared interest in providing better patient care, Dr. McGregor and Dr. James partnered to find a way to make better use of the information produced by monitoring devices. Dr. McGregor visited researchers at the IBM T.J. Watson Research Center's Industry Solutions Lab (ISL), who were extending a new stream-computing

Smarter healthcare: Using streaming data to help clinicians spot infections

Instrumented

Patient's vital-sign data is captured by bedside monitoring devices up to 1,000 times per second.



Interconnected

Monitoring-device data and integrated clinician knowledge are brought together in real time for an automated analysis using a sophisticated, streamlined computing platform.

Intelligent

Detecting medically significant events even before patients exhibit symptoms will enable proactive treatment before the condition worsens, eventually increasing the success rate and potentially saving lives.

Solution components

Software

- IBM InfoSphere™ Streams
- IBM DB2®

Research

- IBM T.J. Watson Research Center

"I could see that there were enormous opportunities to capture, store and utilize this data in real time to improve the quality of care for neonatal babies."

—Dr. Carolyn McGregor, Canada Research Chair in Health Informatics, University of Ontario Institute of Technology

platform to support healthcare analytics. A three-way collaboration was established, with each group bringing a unique perspective—the hospital focus on patient care, the university's ideas for using the data stream, and IBM providing the advanced analysis software and information technology expertise needed to turn this vision into reality.

The result was Project Artemis, part of IBM's First-of-a-Kind program, which pairs IBM's scientists with clients to explore how emerging technologies can solve real-world business problems. Project Artemis is a highly flexible platform that aims to help physicians make better, faster decisions regarding patient care for a wide range of conditions. The earliest iteration of the project is focused on early detection of nosocomial infection by watching for reduced heart rate variability along with other indications. For safety reasons, in this development phase the information is being collected in parallel with established clinical practice and is not being made available to clinicians. The early indications of its efficacy are very promising.

Project Artemis is based on IBM InfoSphere™ Streams, a new information processing architecture that enables near-real-time decision support through the continuous analysis of streaming data using sophisticated, targeted algorithms. The IBM DB2® relational database provides the data management required to support future retrospective analyses of the collected data.

A different kind of research initiative

Because SickKids is a research institution, moving the project forward was not difficult. "The hospital sees itself as involved in the generation of new knowledge. There's an expectation that we'll do research. We have a research institute and a rigorous research ethics board, so the infrastructure was already there," Dr. James notes.

Project Artemis was a consequence of the unique and collaborative relationship between SickKids, UOIT and IBM. "To gain its support, we needed to do our homework very carefully and show that all the bases were covered. The hospital was cautious, but from the beginning we had its full support to proceed."

Even with the support of the hospital, there were challenges to be overcome. Because Project Artemis is more about information technology than about traditional clinical research, new issues had to be considered. For example, the hospital CIO became involved because the

system had to be integrated into the existing network without any impact. Regulatory and ethical concerns are part of any research at SickKids, and there were unique considerations here in terms of the protection and security of the data. The research team's goal was to exceed provincial and federal requirements for the privacy and security of personal health information—the data had to be safeguarded and restricted more carefully than usual because it was being transmitted to both the University of Ontario Institute of Technology and to the IBM T.J. Watson Research Center.

After the overarching concerns were dealt with, the initial tests could begin. Two infant beds were instrumented and connected to the system for data collection. To ensure safety and effectiveness, the project is being deployed slowly and carefully, notes Dr. James. "We have to be careful not to introduce new technologies just because they're available, but because they really do add value," says Dr. James. "It is a stepwise process that is still ongoing. It started with our best attempt at creating an algorithm. Now we're looking at its performance, and using that information to fine tune it. When we can quantify what various activities do to the data stream, we'll be able to filter them out and get a better reading." The ultimate goal is to create a robust, valid system fit to serve as the basis for a randomized clinical trial.

Merging human knowledge and technology

The initial test of the Project Artemis system captured the data stream from bedside monitors and processed it using algorithms designed to spot the telltale signs of nosocomial infection. The algorithm concept is the essential difference between the Artemis system and the existing alarms built into bedside monitors. Although the first test is focused on nosocomial infection, the system has the flexibility to handle any rule on any combination of behaviors across any number of data streams. "What we've built is a set of rules that reflects our best understanding of the condition. We can change and update them as we learn more, or to account for variations in individual patients. Artemis represents a whole new level of capability," Dr. James notes.

The truly significant aspect of the Project Artemis approach is how it brings human knowledge and expertise together with device-generated data to produce a better result. The system's outputs are based on algorithms developed as a collaboration between the clinicians themselves and programmers. This inclusion of the human element is critical,

because good patient care cannot be reduced to mere data points. Validation of these results by an experienced physician is vital since the interpretation of these results has to do with medical knowledge, judgment, skill and experience. As part of the project, the rules being used by Project Artemis are undergoing separate clinical research to support evidence-based practice.

Artemis also holds the potential to become much more sophisticated. For example, eventually it might integrate a variety of data inputs in addition to the streaming data from monitoring devices—from lab results to observational notes about the patient's condition to the physician's own methods for interpreting information. In this way, the knowledge, understanding and even intuition of physicians and nurses will become the basis of the system that enables them to do much more than they could on their own.

"In the early days, there was a lot of concern that computers would eventually 'replace' all health care providers," Dr. James says. "But now we understand that human beings cannot do everything, and it's quite helpful to develop tools that enhance and extend the physicians' and nurses' capabilities. I look to a future where I'm going to receive an alert that provides me with a comprehensive, real-time view of the patient, allowing me to make better decisions on the spot."

Broadening the impact of Artemis

The flexibility of the platform means that in the future, any condition that can be detected through subtle changes in the underlying data streams can be the target of the system's early-warning capabilities. Also, since it depends only on the availability of a data stream, it holds the potential for use outside the ICU and even outside the hospital. For example, the use of remote sensors and wireless connectivity would allow the system to monitor patients wherever they are, while still providing life-saving alerts in near-real time.

"I think the framework would also be applicable for any person who requires close monitoring—children with leukemia, for example," says Dr. James. "These kids are at home, going to school, participating in sports—they're mobile. It leads into the whole idea of sensors attached to or even implanted in the body and wireless connectivity. Theoretically, we could ultimately monitor these conditions from anywhere on the planet."

For more information

To learn more about how IBM can help you transform your business, contact your IBM sales representative or IBM Business Partner.

Visit us at: **ibm.com**/smarterplanet/healthcare

IBM Corporation
1 New Orchard Road
Armonk, NY 10504
U.S.A.

Produced in the United States of America
December 2010

This case study illustrates how one IBM customer uses IBM products. There is no guarantee of comparable results.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.

[1] P. Griffin and R. Moorman, "Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis," Pediatrics, vol. 107, no. 1, 2001.

Please Recycle

ODC03157-USEN-00

UMBC
AN HONORS UNIVERSITY IN MARYLAND

University of Maryland, Baltimore County (UMBC)

Developing accurate forecasts for wildfire smoke dispersion

Smart is...

Accurately forecasting wildfire smoke patterns to improve strategies that promote public safety.

Researchers at the University of Maryland, Baltimore County are conducting a real-time assessment of wildfire smoke patterns to promote informed decisions for public evacuations and health alerts. Air quality analysis has traditionally been limited to using frontline observations, weather forecast data updated every six hours, and low-resolution satellite imagery. With IBM InfoSphere Streams, part of IBM's big data platform, it is now possible to collect and analyze massive amounts of data instantly from drone aircraft, high-resolution satellite imagery, and air quality sensors to develop more accurate smoke dispersion forecasts.

The University of Maryland, Baltimore County (UMBC) maintains a high standing in national rankings and a notable reputation for research in many areas. Since poor air quality can greatly impact health and life, timely observations and reliable forecasts are important for hazard control and protecting public health. In 2008, two professors and a Ph.D. graduate student in the Department of Computer Science and Electrical Engineering at UMBC initiated a research project to analyze wildfire smoke patterns. The ability to provide fire and public safety officials with a real-time assessment of smoke patterns during a fire would support more informed decisions on public evacuations and health warnings.

Dr. Kostas Kalpakis, Associate Professor in the Department of Computer Science and Electrical Engineering at UMBC, and principal researcher on this project, explains: "We saw the availability of new data and the demand for better forecasts, but we had limited means to integrate the new data to improve dynamic model forecasts."

The UMBC research team began a search for technology that would help process large volumes of real-time data streams from various atmospheric sensors and weather reporting agencies and provide input for forecasting air quality and improving the accuracy of predictions.

IBM

Business benefits

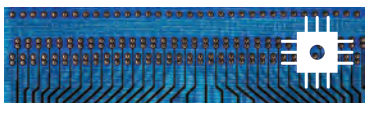
- Fuses air quality data from multiple sources to provide public officials with real-time data about fire and smoke status
- Enables an improvement of about 16 percent in forecast accuracy in the first three days of an emission event based on prototype results
- Accelerates expected times for wildfire analysis from hours to minutes to promote timely decisions concerning firefighting and public safety
- Delivers perpetual analytics to reconcile differences between real-time data and associated predictions to continuously refine forecasts

"Our project required collecting and analyzing massive amounts of air quality data from a variety of sources in real time or near real time," says Dr. Yaacov Yesha, Professor at UMBC's Department of Computer Science and Electrical Engineering. "Our immediate goal was to demonstrate a prototype system that could fuse near real-time observations from multiple sensors and improve the accuracy of near-term forecasts for pollutant dispersion. The longer-term goal is to develop an operational system for air quality forecasting. We knew that the idea for our Data Assimilation research project would align nicely with IBM's Smarter Planet initiatives and saw that we could benefit from IBM InfoSphere Streams to overcome problems in analyzing and correlating the data."

Developing a prototype that improves forecasting accuracy

The Data Assimilation research project comprises techniques used to fuse real-time data obtained through observations with results from mathematical models, while considering the uncertainties of both, as well as any constraints on consistency resulting from the underlying physical system. A complete data assimilation process requires observational data obtained from a variety of air quality sources, a processing model and an assimilation algorithm.

Smarter Research: Bridging the gap between predictions and real-time observations

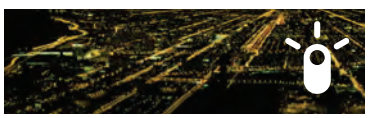


Instrumented

Obtains real-time air quality data available from government environmental agencies and weather services for analysis.

Interconnected

Analyzes real-time air quality data using a scientific algorithm and processing model designed to use atmospheric conditions to generate air quality forecasts.



Intelligent

Reconciles the differences between the actual measurements and the associated predictions, and recalibrates model algorithms to continually improve forecast accuracy.

Solution components

Software

- IBM® InfoSphere® Streams

Servers

- IBM BladeCenter® HS22

"The preliminary results on our Data Assimilation prototype indicate an improvement of about 16 percent in forecast accuracy in the first three days of an emission event. It's preliminary, but represents quite a bit of an improvement over traditional forecasts."

—Dr. Kostas Kalpakis, Associate Professor, Department of Computer Science and Electrical Engineering, UMBC

First, real-time air quality data is collected from multiple observation sources to more accurately predict the concentration of particulate matter in the air. Next, UMBC researchers feed this data into the HYSPLIT (Hybrid Single Particle Lagrangian Integrated Trajectory) model developed by the NOAA (National Oceanic and Atmospheric Administration) Air Resources Lab. This model provides a complete system for computing particle dispersion patterns to indicate how particles travel through the atmosphere. It also allows for performing complex dispersion and deposition simulations.

Lastly, a data assimilation algorithm (known as LETKF or Local Ensemble Transform Kalman Filter) provides an ensemble analysis of the data to deliver more accurate initial conditions used to create forecasts. The algorithm runs on IBM® InfoSphere® Streams, which provides parallel and high-performance stream processing and continually receives and analyzes data from the variety of sources. InfoSphere Streams is part of IBM's big data platform, and UMBC is using InfoSphere Streams version 1.2, with plans to upgrade to version 2.0.

The new system enables researchers to use streaming data in real time from sources, such as geostationary and orbiting satellites and drone aircraft, to help improve forecasting accuracy.

Dr. Kalpakis explains: "To generate a forecast, first we have historical weather and air quality data from prior experiments conducted by NOAA while developing the HYSPLIT model for pollutant dispersion. It is critical to note that HYSPLIT is the operational forecasting model that NOAA uses to advise government agencies. The same types of data used to validate the HYSPLIT model are being used to assess the quality of our fusion architecture and the algorithms that we have to fuse measurements with the forecast. And we have seen quite an improvement in terms of the forecasting accuracy."

"We knew that the idea for our Data Assimilation research project would align nicely with IBM's Smarter Planet initiatives and saw that we could benefit from IBM InfoSphere Streams to overcome problems in analyzing and correlating the data."

—Dr. Yaacov Yesha, Professor, Department of Computer Science and Electrical Engineering, UMBC

Stationary satellites provide measurements approximately every half-hour, based on goals set by NOAA. The orbiting satellites, operated by NASA (National Aeronautics and Space Administration) provide measurements twice a day. "Going forward, we want to obtain data from UAVs (Unmanned Aerial Vehicles) that can be flown over wildfire areas to obtain higher spatial and temporal resolution data than is available today with the satellite instruments," says Dr. Kalpakis.

After the Data Assimilation prototype generates a forecast, the objective is to validate it. "The most important contribution of our project is bridging the gap between air quality forecasts and real-time measurements to improve wildfire smoke prediction and monitoring," says Dr. Kalpakis. "Reconciling the differences between the actual air quality measurements and the associated predictions enables us to recalibrate model algorithms to improve future forecasts. By making continuous improvements, we can bring predictions and observed measurements closer together."

Building a platform for computationally intensive models

The process of revising forecasts based on actual measurements can be computationally intensive. For example, when the weather service runs its numerical models, it is actually using big super computers to do those calculations because the process is computationally demanding. As a result, selecting an appropriate platform for this research project was critical.

The IBM Systems and Technology Group, University Alliances team worked with the IBM University Relations team to provide UMBC researchers with an IBM BladeCenter® H and 13 IBM BladeCenter HS22 servers comprising 104 cores (8×13) and a few hundred gigabytes of random access memory (RAM). These servers support a broad range of workloads. The Data Assimilation prototype runs on a University cluster, called BlueGrid, which hosts eight IBM BladeCenter HS22 servers with Intel Xeon processors running Red Hat Linux (RHL) 5.5, a 64-bit operating system.

"InfoSphere Streams is quite suitable for scientific calculations since it integrates complex data types for vector and matrix operations, as well as extended libraries, which we need to use extensively in the Data Assimilation system."

—Dr. Kostas Kalpakis

InfoSphere Streams offers advantages in scaling across a wide range of hardware platforms and can be reconfigured automatically in response to changes in data and system resource availability. With InfoSphere Streams, UMBC can respond to events and trends immediately, while it is still possible to improve outcomes.

Dr. Kalpakis comments, "Using the LETKF algorithm and InfoSphere Streams, multiple localized and independent processes can happen simultaneously to generate results much faster."

Promising results from the Data Assimilation prototype

"To my knowledge, this is the first time anyone has done air quality forecasting in a systematic way using an operational model like HYSPLIT from NOAA and a diverse set of measurements to improve the accuracy of sensor forecasts," says Dr. Kalpakis. "We managed to get this project off the ground in a little over a year and also provided the capability to improve forecast accuracy by fusing these observations into the system. The preliminary results on our Data Assimilation prototype indicate an improvement of about 16 percent in forecast accuracy in the first three days of an emission event. It's preliminary, but it represents quite an improvement over traditional forecasts. That is the thing that makes our research interesting and exciting for us to continue."

Looking clearly toward the future

If made available in a production environment, UMBC's Data Assimilation system would help improve the way wildfires are tracked and extinguished and improve decision making for issuing health and safety alerts.

"We are fortunate that we started this project with IBM, and so far, our activity has been well received by some of the major groups, for example NOAA," says Dr. Yesha. "Our long-term goal is to move the prototype design into an operational system for air quality forecasting. One factor of paramount importance is the accessibility and interoperability of the prototype system. A fully operational system will need to interface with many other systems."

"InfoSphere Streams is quite suitable for scientific calculations since it integrates complex data types for vector and matrix operations, as well as extended libraries, which we need to use extensively in the Data Assimilation system," Dr. Kalpakis adds. "InfoSphere Streams offers scalability and strong support for data stream manipulation that will make developing an extensible system and adding functionality much easier. Moreover its mixed-mode code design also helps us add generic code to handle different situations and hence makes the code easier to manage. In addition, the support InfoSphere Streams provides for the operator graph structure allows us to conceptualize and visualize the system design, reduce the development effort and facilitate the design of a clear extensible system."

IBM continues to provide strong technical support in promoting InfoSphere Streams," concludes Dr. Kalpakis. "We are excited to be part of demonstrating the Data Assimilation prototype. IBM's participation in this effort was crucial in terms of resources and technical guidance. Having a close collaboration with the InfoSphere Streams team was extremely critical for us to reach this point. We are very grateful for their guidance, support and encouragement."

► The inside story: Getting there

According to Dr. Kalpakis, the Data Assimilation project started with an idea to align environmental research with IBM's Smarter Planet initiatives.

"The inspiration for this project was driven by a few key factors," Dr. Kalpakis explains. "First, we have the increasing availability of near real-time measurements and various sensors, including space borne sensors and those aboard unmanned aircraft vehicles (UAVs). Second, we have increasing demand for continuous, near real-time forecasts about the future state of the world and our environment. So we had the availability of new data and demand for better forecasts, but limited means to integrate observations with our dynamic mathematical system models to gain the actionable insight. These factors aligned closely with IBM's Smarter Planet initiative. So we started talking to the IBM InfoSphere Streams group and we got the project started."

The biggest incentive was to demonstrate the value of the data assimilation techniques. The researchers proposed starting with a simple prototype that exhibits most of the characteristics of the target application for analyzing pollutant dispersion.

"By working out the details using the prototype, we were able to gain support from the various University stakeholders for the project," says Dr. Kalpakis.

For more information

To learn more about how IBM can help you transform your business, please contact your IBM sales representative or IBM Business Partner.

To learn more about IBM big data, visit: **ibm.com**/software/data/bigdata

To learn more about IBM InfoSphere Streams, visit: **ibm.com**/software/data/infosphere/streams

To get involved in the conversation: www.smartercomputingblog.com/category/big-data

For more information about the University of Maryland, Baltimore County, visit: www.umbc.edu

IBM Corporation
Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
December 2011

IMC14727-USEN-00

Vestas

Turning climate into capital with big data

Smart is...

Pinpointing the optimal location for wind turbines to maximize power generation and reduce energy costs.

Precise placement of a wind turbine can affect its performance and its useful life. For Vestas, the world's largest wind energy company, gaining new business depends on responding quickly and delivering business value. To succeed, Vestas uses one of the largest supercomputers worldwide along with a new big data modeling solution to slice weeks from data processing times and support 10 times the amount of data for more accurate turbine placement decisions. Improved precision provides Vestas customers with greater business case certainty, quicker results and increased predictability and reliability in wind power generation.

For centuries, sailors have seen how fickle the wind can be. It ebbs and flows like the tide and can allow ships to travel great distances or remain becalmed at sea.

But despite the wind's capricious nature, new advances in science and technology enable energy producers to transform the wind into a reliable and steadfast energy source—one that many believe will help alleviate the problems of the world's soaring energy consumption.

"Wind energy is one of today's most important renewable energy sources," says Lars Christian Christensen, vice president, Vestas Wind Systems A/S. "Fossil fuels will eventually run out. Wind is renewable, predictable, clean and commercially viable. By 2020 as much as 10 percent of the world's electricity consumption will be satisfied by wind energy and we believe that wind power is an industry that will be on par with oil and gas."

Producing electricity from wind

Making wind a reliable source of energy depends greatly on the placement of the wind turbines used to produce electricity. The windiest location may not generate the best output and revenue for energy companies. Turbulence is a significant factor as it strains turbine components, making them more likely to fail. Avoiding pockets of turbulence can extend the service life of turbines and lower operating costs, which reduces the cost per kilowatt hour of energy produced.

"We can now show our customers how the wind behaves and provide a solid business case that is on par with any other investment that they may have."

– Lars Christian Christensen, Vice President, Vestas Wind Systems A/S

Business benefits

- Reduces response time for wind forecasting information by approximately 97 percent—from weeks to hours—to help cut development time
- Improves accuracy of turbine placement with capabilities for analyzing a greater breadth and depth of data
- Lowers the cost to customers per kilowatt hour produced and increases customers' return on investment
- Reduces IT footprint and costs, and decreases energy consumption by 40 percent—all while increasing computational power

Selecting wind turbine sites is a science that Vestas understands well. Since 1979, this Danish company has been engaged in the development, manufacture, sale and maintenance of wind power systems to generate electricity. The company has installed more than 43,000 land-based and offshore wind turbines in 66 countries on six continents. Today, Vestas installs an average of one wind turbine every three hours, 24 hours a day, and its turbines generate more than 90 million megawatt-hours of energy per year—enough electricity to supply millions of households.

"Customers want to know what their return on investment will be and they want business case certainty," says Christensen who heads the company's division responsible for determining the placement of wind turbines. "For us to achieve business case certainty, we need to know exactly how the wind is distributed across potential sites, and we need to compare this data with the turbine design specifications to make sure the turbine can operate at optimal efficiency at that location."

What happens if engineers pick a sub-optimal location? According to Christensen, the cost of a mistake can be tremendous. "First of all, if the turbines do not perform as intended, we risk losing customers. Secondly, placing the turbines in the wrong location affects our warranty costs. Turbines are designed to operate under specific conditions and can break if they are operating outside of these parameters."

For Vestas, the process of establishing a location starts with its wind library, which incorporates data from global weather systems with data collected from existing turbines. Combined, this information helps the company not only select the best site for turbine placement, but also helps forecast wind and power production for its customers.

Smarter Energy: **Increases wind power generation through optimal turbine placement**

Instrumented

Determines the optimal turbine placement using weather forecasts and data from operational wind power plants to create hourly and daily predictions regarding energy production.

Interconnected

Combines turbine data with data on temperature, barometric pressure, humidity, precipitation, wind direction and velocity from the ground level up to 300 feet.

Intelligent

Precisely models wind flow to help staff understand wind patterns and turbulence near each wind turbine and select the best location to reduce the cost per kilowatt hour of energy produced.

Solution components:

Software

- IBM® InfoSphere® BigInsights Enterprise Edition

Hardware

- IBM System x® iDataPlex® dx360 M3
- IBM System Storage® DS5300

"In our development strategy, we see growing our library in the range of 18 to 24 petabytes of data. And while it's fairly easy to build that library, we needed to make sure that we could gain knowledge from that data."

– Lars Christian Christensen

"We gather data from 35,000 meteorological stations scattered around the world and from our own turbines," says Christensen. "That gives us a picture of the global flow scenario. Those models are then cobbled to smaller models for regional level called mesoscale models. The mesoscale models are used to establish our huge wind library so we can pinpoint a specific location at a specific time of day and tell what the weather was like."

The company's previous wind library provided detailed information in a grid pattern with each grid measuring 27x27 kilometers (about 17x17 miles). Using computational fluid dynamics models, Vestas engineers can then bring the resolution down even further—to about 10x10 meters (32x32 feet)—to establish the exact wind flow pattern at a particular location.

However, in any modeling scenario, the more data and the smaller the grid area, the greater the accuracy of the models. As a result, Christensen's team wanted to expand its wind library more than 10 fold to include a larger range of weather data over a longer period of time. Additionally, the company needed a more powerful computing platform to run global forecasts much faster. Often company executives had to wait up to three weeks for feedback regarding potential sites—an unacceptable amount of time for Vestas and its customers in this competitive industry.

"In our development strategy, we see growing our library in the range of 18 to 24 petabytes of data," says Christensen. "And while it's fairly easy to build that library, we needed to make sure that we could gain knowledge from that data."

Turning climate into capital

Working with IBM, Vestas today is implementing a big data solution that is slicing weeks from data processing time and helping staff more quickly and accurately predict weather patterns at potential sites to increase turbine energy production. Data currently stored in its wind library comprises nearly 2.8 petabytes and includes more than 178 parameters, such as temperature, barometric pressure, humidity, precipitation, wind direction and wind velocity from the ground level up to 300 feet, along with the company's own recorded historical data. Future additions for use in predictions include global deforestation metrics, satellite images, historical metrics, geospatial data and data on phases of the moon and tides.

Journey to Smarter Computing

Designed for Data
Implementing a big data solution enables Vestas to create a wind library to hold 18 to 24 petabytes of weather and turbine data at various levels of granularity and reduce the geographic grid area used for modeling by 90 percent for increased accuracy.

Tuned to the Task
Working with IBM, Vestas can increase computational power while shrinking its IT footprint and reducing server energy consumption by 40 percent. Today, twice the number of servers can be run in each of its supercomputer's 12 racks.

Managed for Rapid Service Delivery
Processing huge volumes of climate data and the ability to gain insight from that data enables Vestas to forecast optimal turbine placement in 15 minutes instead of three weeks. This in turn shortens the time to develop a wind turbine site by nearly a month.

"We could pose the questions before, but our previous systems were not able to deliver the answers, or deliver the answers in the required timeframe," says Christensen. "Now, if you give me the coordinates for your back yard, we can dive into our modeled wind libraries and provide you with precise data on the weather over the past 11 years, thereby predicting future weather and delivering power production prognosis. We have the ability to scan larger areas and determine more quickly our current turbine coverage geographically and see if there are spots we need to cover with a type of turbine. We can also assess information on how each turbine is operating and our potential risk at a site."

IBM® InfoSphere® BigInsights software running on an IBM System x® iDataPlex® system serves as the core infrastructure to help Vestas manage and analyze weather and location data in ways that were not previously possible. For example, the company can reduce the base resolution of its wind data grids from a 27x27 kilometer area down to a 3x3 kilometer area (about 1.8x1.8 miles)—a nearly 90 percent reduction that gives executives more immediate insight into potential locations. Christensen estimates this capability can eliminate a month of development time for a site and enable customers to achieve a return on investment much earlier than anticipated.

"IBM InfoSphere BigInsights helps us gain access to knowledge in a very efficient and extremely fast way and enables us to use this knowledge to turn climate into capital," says Christensen. "Before, it could take us three weeks to get a response to some of our questions simply because we had to process a lot of data. We expect that we can get answers for the same questions now in 15 minutes."

For customers, the detailed models mean greater business case certainty, quicker results and increased predictability and reliability on their investment.

"Our customers need predictability and reliability, and that can only happen using systems like InfoSphere BigInsights," says Christensen. "We can give customers much better financial warrantees than we have been able to in the past and can provide a solid business case that is on par with any other investment that they may have."

"IBM InfoSphere BigInsights helps us gain access to knowledge in a very efficient and extremely fast way and enables us to use this knowledge to turn climate into capital."

– Lars Christian Christensen

Smarter Computing by design

Tackling big data challenges

Vestas and IBM worked together to implement IBM InfoSphere BigInsights software, designed to enable organizations to gain insight from information flows that are characterized by variety, velocity and volume. The solution combines open source Apache Hadoop software with unique technologies and capabilities from IBM to enable organizations to process very large data sets—breaking up the data into chunks and coordinating the processing across a distributed environment for rapid, efficient analysis and results.

"IBM gave us an opportunity to turn our plans into something that was very tangible right from the beginning," says Christensen. "IBM had experts within data mining, big data and Apache Hadoop, and it was clear to us from the beginning if we wanted to improve our business, not only today, but also prepare for the challenges we will face in three to five years, we had to go with IBM."

Maintaining energy efficiency in its data center

For a company committed to addressing the world's energy requirements, it's no surprise that as Vestas implemented its big data solution, it also sought a high-performance, energy efficient computing environment that would reduce its carbon footprint. Today, the platform that drives its forecasting and analysis comprises a hardware stack based on the IBM System x iDataPlex supercomputer. This supercomputing solution—one of the world's largest to date—enables the company to use 40 percent less energy while increasing computational power. Twice the number of servers can be run in each of the system's 12 racks—reducing the amount of floor space required in its data center.

"The supercomputer provides the foundation for a completely new way of doing business at Vestas and combined with IBM software delivers a smarter approach to computing that optimizes the way we work," says Christensen.

"Before, it could take us three weeks to get a response to some of our questions simply because we had to process a lot of data. We expect that we can get answers for the same questions now in 15 minutes."

– Lars Christian Christensen

▶ The inside story: Getting there

According to Christensen, the idea for this project began with the collaboration among his team, the company's global research department and its sales business units.

"We needed to know where the goldmines of wind are hidden, and we needed to have more information to aid our decisions," says Christensen. "We quickly formed a project group that took the idea forward and set out some key performance indicators that had to be met in order to proceed to the stage where we are today."

For Vestas, the opportunity that a big data solution could provide made the decision easy. "Once we had the business potential of having these capabilities, it was fairly easy to gain acceptance," says Christensen. "We were able to show the cost of a system alongside the near-term and long-term benefits, so it was really a no brainer."

For more information

To learn more about how IBM can help you transform your business, please contact your IBM sales representative or IBM Business Partner.

To learn more about big data solutions from IBM, visit:
ibm.com/software/data/bigdata

To learn more about IBM InfoSphere BigInsights, visit:
ibm.com/software/data/infosphere/biginsights

To increase your big data knowledge and skills, visit:
www.BigDataUniversity.com

To get involved in the conversation:
www.smartercomputingblog.com/category/big-data

For more information about Vestas Wind Systems A/S, visit:
www.vestas.com

IBM Corporation
1 New Orchard Road
Armonk, NY 10504
U.S.A.

Produced in the United States of America
October 2011

IMC14702-USEN-00

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
October 2012

Please Recycle

IMM14097-USEN-00