

SAP Sybase

PAGE 16

SAP SYBASE IQ —
A SCALABLE DATABASE FOR
BUSINESS ANALYTICS

Attunity

PAGE 20

MOVING DATA IN TODAY'S
REAL-TIME BUSINESS

Composite Software

PAGE 21

DATA VIRTUALIZATION
REDEFINES THE
STOCK EXCHANGE

Denodo Technologies

PAGE 22

TOP 4 DATA CHALLENGES
ADDRESSED BY DATA
VIRTUALIZATION

Tableau Software

PAGE 23

5 WAYS BEST-IN-CLASS
BI BUILDS ON THE
INDUSTRY'S HISTORY,
THEN THROWS IT OUT

Calpont

PAGE 24

WHY A COLUMNAR DATA
PLATFORM IS BEST FOR
BIG ANALYTICS

Objectivity

PAGE 25

GETTING BIG ANSWERS
OUT OF BIG DATA
*THE KEY TO BUSINESS
INNOVATION*

Progress DataDirect

PAGE 26

MAKE FAST AND INFORMED
DECISIONS BASED ON
UP TO THE MINUTE DATA

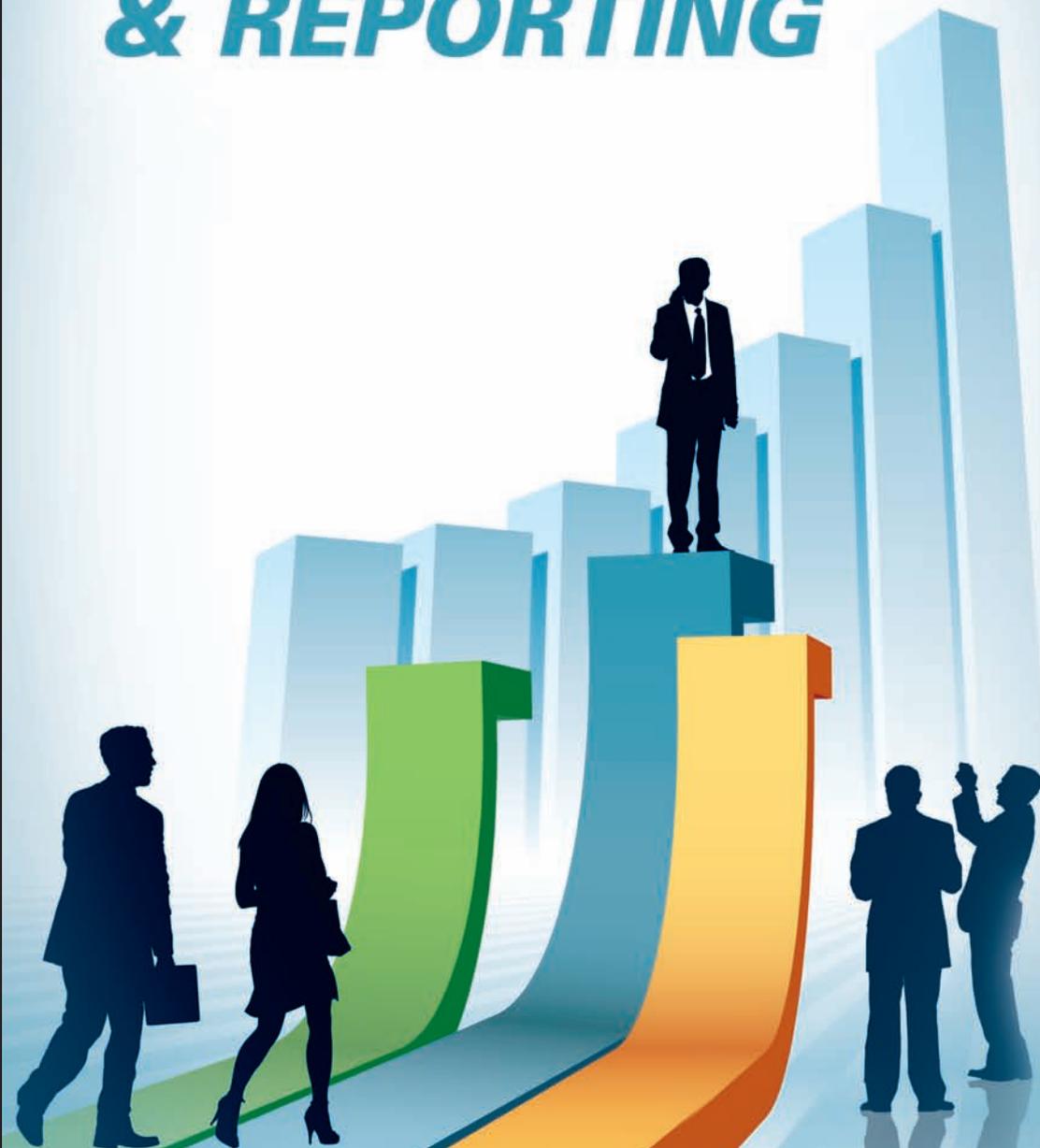
Hortonworks

PAGE 27

APACHE HADOOP:
A BIG DATA REFINERY
ENABLING BIG ANALYTICS

database[®]
TRENDS AND APPLICATIONS

ANALYTICS, BUSINESS INTELLIGENCE, & REPORTING



Best Practices Series

Big Data Opens New Frontiers for the Reporting and Analytics Industry



WITH THE RISE OF BIG DATA, the database and data management tools market is in a state of flux, the likes of which have not been seen in this sector before. Companies are now awash in big data, and end users are demanding greater capability and integration to mine and analyze new sources of information. As a result, organizations are supplementing their relational database environments with new platforms and approaches that address the variety and volume of information being handled.

Notably, big data is fueling the emergence of a new generation of database types, which no longer sit at the fringes of enterprises. While predictions of the demise of relational databases have abounded for more than a decade, a new survey of 264 data and IT managers finds that big data is adding a new wrinkle to this speculation. The survey, conducted among *DBTA* readers by Unisphere Research, the market research unit of Information Today, Inc. and sponsored by MarkLogic, confirms that these databases are not only being implemented for peripheral applications but also for core functions. The most prevalent platforms seen are the so-called NoSQL (“not only” SQL) and in-memory databases. Adoption of these non-relational databases is on the rise as big data expands. However, many respondents worry about the skill sets they will need to effectively deliver solutions on these platforms.

The value of being able to manage big data is not lost on major vendors, and

there has been a spate of acquisitions in recent years to fill any gaps that exist in product lines. For example, IBM recently entered into an agreement to buy Vivisimo, creator of big data analysis software. VMware bought Vetas, a big data analytics startup; Teradata purchased Aster Data Systems, an unstructured data analysis provider; HP purchased Vertica; IBM acquired Netezza; and EMC bought Greenplum. There have also been a number of strategic partnerships forged, including IBM’s partnership with Cloudera, a Hadoop provider. Cloudera is also partnering with both Oracle and Teradata. In addition, large, established data players have entered the big data domain with their own offerings, Vish Vishwanath, senior vice president of BI and analytics at Persistent Systems, relates to *DBTA*. “These have included the integration of big data with traditional BI and reporting solutions, appliance-based solutions that bundle big data technology with other software product offerings on a hardware platform, or as cloud-based big data solutions.”

Ultimately, businesses understand the intrinsic value in the large data stores they are accumulating. “Big data should be thought of as a new natural resource,” Deepak Advani, vice president of IBM Business Analytics Products and SPSS, tells *DBTA*. “Data and analytics are abundant resources and companies are seeing how they can drive those into new solutions—in particular, retaining and satisfying customers and increasing operational efficiency.”

In addition, big data affords companies the opportunity to ask questions they never were able to ask before, adds Christian Hasker, director of product development for Quest Software. “How can they bring big data techniques to their current data and gain that competitive advantage? How can financial organizations find better ways to detect fraud? How can insurance companies gain a deeper insight into their customers to see who may be the most un-economical to insure? How does a software company find their most at-risk customers—those who are about to deploy a competitive product? These and countless other questions can be answered by adopting big data techniques.”

But amid all the hype and excitement around big data, it’s important to keep things in perspective as well. Big data itself may also carry some erroneous assumptions that data managers need to understand and convey to their businesses. “The manipulation of the data into truly holistic data points on top of which to base cogent decisions, is far harder than it looks,” Antonio Piraino, CTO of ScienceLogic, tells *DBTA*. “There is a ways to go before the variety of correlations and presentations of big data are refined and proven. It’s not always easy to confidently interpret what the data is telling you; not all values captured from big data are immediately relevant to the symptom of a problem or imminent opportunity.”

The rise of new and bigger data environments also brings the risk of lack

of cohesion within enterprises, many of which are still dominated by traditional data environments. “The big database vendors all follow the same general architecture,” noted Brian Clark, vice president of product management for Objectivity, Inc., in a webcast the company presented in partnership with DBTA. “Capture of the data coming from many sources at high speed, do some initial processing, then put it into some other storage for further processing and analysis, and then use that data in data analytics or visualization applications. The problem with this approach is yes, they can get the volume of data in, and yes, they can get the variety. But the data still ends up in different stores or different silos. And we believe there are hidden relationships in the data in those silos.”

In many organizations, “traditional data sources are not well integrated,” agrees Bill Abbott, principal for PwC, noting that “80% to 90% of the productivity cycle is spent gathering and organizing information.” Abbott recommends that organizations adopt a defined analytics strategy, focusing on the repeatability of valuable analytical processes.

Often, organizations attempting to manage big data have diverse environments in which it is difficult to actually take advantage of the data. “One of the problems of distributed big data is knowing what you’ve got and where it is,” says Clark, who recommends a federated approach to leverage data across the enterprise. “If you look at the ways big data is being handled, being managed, then typically it’s a combination of different storage, different databases, that store structured data.”

NEW TOOLS

NoSQL or nonrelational databases or technologies are being adopted by more than two-fifths, or 41%, of respondents to the DBTA-MarkLogic survey. Interestingly, the level of companies’ stores of unstructured data seem to have little bearing on decisions to go with these types of databases. About one-fifth of survey respondents are choosing in-memory databases as their solutions for

growing data demands. More than two-fifths, or 41% of respondents, indicate that at least some of their workloads are now running on these databases.

Mark Troester, CIO and IT market strategist at SAS, urges that these new breeds of databases be factored into a flexible data infrastructure. “It’s not one-size-fits-all, it’s not in-database versus in-memory, and it’s not simply leveraging an in-memory database,” he says. “It’s about using in-memory the right way—in some cases, such as data visualization and exploration, the architecture needs to support an in-memory model that provides a high level of concurrency because it might be necessary to open

‘Big data should be thought of as a new natural resource.’

big data up to many business users for exploration and analysis. Complementing this would be environments that require in-memory support to build analytical models based on a combination of complex and intensive analytics combined with huge data volumes.”

Another new option—cloud computing—is now in use or planned at seven out of 10 of the organizations in the DBTA-MarkLogic big data survey. A majority also report that at least some of their data-centric workloads are run in the cloud. Many are attracted to the cloud by the cost advantages but are still put off by data security concerns. Cloud-based environments won’t work for everyone, cautions Steve Woledge, senior director of product marketing for Teradata-Aster. “Cloud-based and web-based tools are easy to deploy, but usually not very performant. Companies will need to make the decision that balances ease of deployment versus speed of analytics. Near real-time analytics like live stock market ticker data feeds will not benefit from cloud-based deployments.”

Open source tools are also increasingly being used for addressing big data. Seven out of 10 organizations in the DBTA-MarkLogic survey use open source technologies within their data environments, and expect to keep

expanding their use of these solutions. “In addition to established commercial offerings from big IT vendors, open source tools now have the usability features that make them viable for the enterprise,” Richard Vermillion, CEO of Fulcrum, tells DBTA. “The ecosystem that’s developed around these open source applications—cloud-based offerings, third-party integration, and developer support—has further established them as a competitive factor. Today’s buyer has a wealth of purchasing options, from roll-your-own applications built on open source tools to cloud-based offerings to solutions sourced from traditional IT vendors.”

The ultimate open source tool for big data at this time is Hadoop, the file system that organizes massive data stores into a digestible format. About 15% of survey respondents either have or are planning to work with Hadoop. “Hadoop is emerging as the data store of choice to hold unstructured data such as weblogs and social data,” Hasker observes.

However, as the DBTA study found, there is still anxiety about support around open source software. Some industry experts also caution that open source may require more of an investment than initially bargained for. “Many of today’s open source models for big data require an excessive amount of programming to support big data analytics,” John Santaferro, vice president of solutions and product marketing from ParAccel, tells DBTA. Open source solutions for big data tend to be good for archiving, filtering and searching, but they are not well suited for complex, dynamic analytics required by companies.”

Ultimately, as data grows bigger and more complex, industry players will need to emphasize simplicity in their systems and solutions. “To make big data technology widespread, we need to see big data appliances: preconfigured systems that can be easily scaled and sized based upon volume,” Oliver Halter, principal with PwC, tells DBTA. “We should also see tighter integration between relational database and NoSQL big data vendors, with a focus on optimization of performance.”



SAP Sybase IQ— A Scalable Database for Business Analytics

THE FOLLOWING is an excerpt from a Winter Corporation white paper: *SAP Sybase IQ 15.4—An Elastic Platform for Business Analytics* by Richard Winter.

EXECUTIVE SUMMARY

Executives around the world are intensely focusing on business analytics. They see an analytical approach to business decisions—an approach based on more abundant data and mathematical analysis of that data—as the cornerstone of new strategies for profitable operation, profitable growth, new product development and customer engagement.

The opportunity to benefit from business analytics is especially large right now in part because businesses have access to “big data”—enormous, previously unavailable volumes of data on the actions, interests and sentiment of customers; on the movement of products, components and raw materials through the supply chain and the distribution chain; and, on many other aspects of the operation of businesses and their market environment. Perhaps surprisingly, the challenge of “big data” is not only the data volume. It is also that much of the new data is less structured and less regular than the tabular corporate data that has been the focus of data warehousing in the past. The new big data comes from new or greatly expanded sources: social media,

rapidly proliferating smart mobile devices, from vehicles and a dizzying array of new sensors and intelligent products.

Even beyond the challenges of big data, there are other obstacles to success with business analytics: data analysis can be a cumbersome, slow, frustrating and expensive process. First you have to find the data you want. Then you have to get it loaded into a repository where it is accessible. Then you need to cleanse it, organize it and integrate it with other data of interest. Then you have to conduct the analysis ... every step bedeviled by many practical difficulties, not least of which is often the difficulty of getting help from people with the right skills.

New open source technology has emerged and is being deployed for “big data”; new vocabulary includes terms such as “Hadoop clusters” and “MapReduce.” This technology brings new benefits for certain types of information and analysis. However, it also creates one more data silo in a world in which there are already too many silos. The complete analytical process thus gets enhanced in some areas but also becomes more fragmented: to get analytic results and business solutions, stakeholders must contend with a yet more complex environment with net new skill requirements.

The new, highly analytical business strategies place a particular emphasis on

prediction. Knowing what happened yesterday isn't enough—you need to predict which of the business actions in front of you is likely to produce the best result. And, as well as judgment, you need facts, data and analysis to back that decision. And, you must take into account the new data sources—the customer sentiment expressed on social media; the customer behavior evident from new data sources and devices; the subtle patterns that can be seen in purchase behavior, web browsing and many other sources; and, the supply chain realities now visible as parts, components, goods and materials move around the world and are affected by weather, catastrophes and human events.

Often, to the decision maker, the unfortunate reality is that predictions of which profitable customers are at risk may indeed be extremely valuable, but getting such predictions before it's too late is easier said than done.

For many enterprises, then, the key to the analytic opportunity is finding a way to make the entire analytic process work smoothly, conveniently, responsively and cost effectively—whether the analysis focuses on the tabular data most frequently used for the past 25 years; on newer data sources, such as sentiment expressed in social media; or, both.

In response to this challenge, SAP has introduced a new version of its flagship analytic DBMS product—SAP Sybase IQ 15.4—as a platform and an integrated



environment to support and facilitate the customer's entire analytic process.

In addition to a greatly enhanced DBMS engine for data warehousing, Sybase IQ 15.4 features significant new capabilities for business analytics and big data. Highlights are:

- A new analytic services layer that supports the use of MapReduce and many other analytic functions on data within Sybase IQ itself;
- Parallel interaction between Sybase IQ and Hadoop;
- Support of R, the open source language for statistical analysis;
- Support of new third party SQL-callable functions for data mining and predictive analytics;
- An expanded eco-system for the support of third-party applications for information lifecycle management, business intelligence and data integration, predictive analytics and system/data administration.

At the core of Sybase IQ 15.4 is the most mature column store DBMS for data warehousing on the market, with sophisticated capabilities for data compression, query processing and query optimization—an engine with a long record of exceptional query performance and efficiency.

While column storage and column-oriented data compression have been “hot trends” for the last few years, Sybase IQ was built from day one with these capabilities: its users have been benefitting from them for more than a decade. And, they contribute significantly to the efficiency of Sybase IQ for analytics.

In addition to the remarkably efficient storage and query processing technology at its core, Sybase IQ 15.4 features P1exQTM technology, a distinctive, elastic design that supports

highly parallel query processing and data loading along with independent scaling for data growth and workload growth.

ELASTIC FRAMEWORK

This paper examines the architecture and capabilities of SAP Sybase IQ 15.4 with a particular focus on demanding new requirements for business analytics and big data.

Business Analytics. People who have been involved with data warehousing for the last decade or more—especially those with a technical background in the field—are often puzzled by the new wave of executive interest in “business analytics.” A common question is, “Aren't we doing that already?” Surely, the reason all that data has been modeled, cleansed, integrated and stored in data warehouses for the last ten or twenty years is so that it can be analyzed!

Certainly there has been analysis going with data warehouse data. But, from the perspective of the business manager or business end user, data warehousing and business intelligence in practice has too often meant little more than routine-ized¹ reporting; extraction to other applications and systems; and, the occasional ad hoc query. Sure, business intelligence tools have steadily improved; data may be delivered on nicer looking, more functional electronic reports and dashboards; data access may be more interactive; and, data may even be

available on mobile devices. All of these advances add some value.

But most end users will still tell you the same thing: most of what they have been doing with the data warehouse has been “looking in the rear view mirror.” Often, business users learn what has happened from the data warehouse. They learn which products have been selling; which customers have been buying; which suppliers have consistently delivered on time ... these insights are treasured when good information was not previously available as a basis for decision making.

The problem is that the practice of business management has moved on from that point. Looking in the rear view mirror is no longer enough.

Increasingly, operating and strategic decisions must be based on forward looking analysis with a mathematically sound foundation. The analytical approaches to business exemplified in *Competing on Analytics* and a series of subsequent books—and in the best selling popular book and recent hit movie, *Moneyball*², have influenced business culture. These accounts and many others have shown how business performance can undergo radical improvement when the decision making process looks forward with analytics. At the heart of this revolutionary analysis is better prediction: whether of the performance of a baseball player, of a product, of a service—or the behavior of a customer.

¹ *Competing on Analytics, The New Science of Winning*, Thomas Davenport and Jeanne Harris, Harvard Business School Press, 2007 (www.tomdavenportbooks.com)

² *Moneyball, The Art of Winning an Unfair Game*, Michael Lewis, W. W. Norton & Company, 2003



And, while you may feel that your data warehouse already has the capabilities to support these analytics, there is more to the story.

Big Data. As predictive analytics have been gaining ever more significance in business circles, another trend—big data—has made a profound impact on business and data strategies.

“Big data” is a broad phenomenon encompassing the rise of social media; the seemingly sudden proliferation of machine generated data; the worldwide spread of mobile intelligent devices, including smart phones and tablets; the widespread use of GPS data, which attaches a location to many events in daily life; and, rapid decreases in cost associated with capturing, delivering and storing a wide range of previously costly varieties of data, including voice, image, video, etc.

Taking all of these phenomena together, we are witnessing an enormous explosion of data which is many times larger and faster growing than what we have seen in data warehouses over the last decade. While the transactional information about customers, products, stores and the like is still uniquely valuable—and plays a central role in understanding any business—there is now new and unprecedented information available that can provide business, engineering, scientific and medical insights never before available.

To provide one example, a useful technique in customer retention is to observe when a profitable customer’s activity with a credit card begins to decline and then react quickly to retain the customer before the account is cancelled. When this technique works it is much more efficient than acquiring a new customer that is equally or more profitable.

But what if you could know earlier—before the usage declined—that the customer was at risk? Perhaps the retention rate would become yet higher and the retention cost lower, particularly if you could discover the reason that the customer relationship was threatened. If you knew the reason, then your actions to deal with it could be yet more efficiently directed at the root cause.

But how could you know earlier? One possibility is social media. If you are engaged with your customers on social media, they may tell you what they are thinking: that they like the service or the incentives or the prices offered by a competitor; that they don’t like your call center or your fees. Or, if they have opted into your social media program, they may let you see what they are saying to others about your product or service.

The enormous flood of data pouring out of social media is one of many examples of big data. Data is also pouring out of a growing tide of products that we use every day, and to the extent that we opt in, manufacturers can gain precious knowledge about how, when, and where we use products—and what problems we have with them. This is clearly the case today with smartphones and tablets. Vehicles are becoming more intelligent and more connected and will increasingly provide similar capabilities (more expensive commercial vehicles, such as helicopters, already provide telemetry data that is used to optimize safety and maintenance). The trend will spread to many other products that we use every day, in every case generating yet more “big data” for analysis.

New Tools and Technologies. The concurrent rise of predictive analytics and big data has generated interest in new tools and technologies for several reasons.

First, much of the big data does not fit closely with the relational database model. Much of the significance of the data is not revealed by fitting it into a tabular structure. Social media data has textual, image, audio, video and other components that must be analyzed primarily by specialized or procedural functions—SQL solves a relatively small part of the problem here. Embedded in the data is a social graph which is most readily analyzed outside of SQL.

In general, a significant element of the new, more predictive analysis—especially of the newly varied and highly voluminous “big data”—is best attacked with tools other than SQL. In connection with this, interest has grown in MapReduce, a parallel data analysis framework, and Hadoop, an open source engine for running MapReduce jobs.

Some data analysis jobs can be readily performed in a Hadoop cluster. Others may require the services of a data warehouse, such as SAP Sybase IQ. Yet others may best be handled with a combination of the two.

Regardless of where the data is stored, interest has also grown rapidly in other analysis tools, such as the open source statistical analysis language, R. In general, the new business analytics will use SQL and the data warehouse, but will also create a strong demand for other tools.

Data Strategies. As enterprises grapple with this rapidly changing world of big data, they need a data infrastructure that will enable them to implement analytic business strategies. Especially with regulatory and governance requirements enforcing longer periods of data retention, enterprises need a convenient, flexible, cost effective process for solving analytic data problems from beginning to end.



Sybase seeks to address that customer need—for a comprehensive approach to business analytics—through its new capabilities in SAP Sybase IQ 15.4.

CONCLUSIONS

Over the course of its last five rapid releases in 3 years—from 15.0 through the present 15.4—SAP Sybase IQ has been transformed to a platform for large scale data analytics and big data. It has significantly advanced in:

- **Scalability**, with the development of its elastic “PlexQ” grid that adds highly parallel execution of large queries and loads; previously, such operations could run in parallel over a single node of the grid; now they can run in parallel over multiple nodes; this is a major architectural advance, highly significant for larger data and workload requirements;
- **In-database analytics**, with a major generalization and extension of the user defined function (UDF) facility in Sybase IQ; with these new capabilities, UDFs can be written in Java as well as C++; they can read and write bulk data in the form of tables and files; they can be run in a protected mode, increasing system reliability and data availability; and, they can be executed in parallel over multiple nodes of the grid;
- **In-database MapReduce, enabling** end users and partners to run MapReduce routines and libraries against data in place and in a highly parallel fashion in Sybase IQ, and opening Sybase IQ up to a large range analytic tools and applications from many vendors and sources;
- **Interface to Hadoop**, enabling the many customers who are investing—or will invest—in an open source data repository in a Hadoop cluster—to

leverage that investment in combination with data and analysis in Sybase IQ;

- **Other analytic application services leveraging in-database MapReduce and new, more powerful UDFs**; these include an expanded, more efficient and more highly parallel version of the Fuzzy Logix data mining and analytics library; a simulator for testing analytic applications; and, other features.
- **Partner Ecosystem**—Other analytical, management and business intelligence tools and functions available from partners, certified by Sybase IQ and providing analytical solutions and capabilities to customers; these include support for the SAP BusinessObjects tool set, the R statistical language; a PMML plug-in for data mining from Zementis; social network analysis from KXEN; query and administration tools from Quest TOAD; and, of other capabilities.

These advances are evidence of a significant reorientation of the product direction and a significant enhancement of the product line to focus on the major drivers of change in business today. Organizations everywhere are grappling with the implications of a much larger volume and variety of data and a much increased focus on business strategies driven by fuller analysis of that data. Mobility (tablets, smartphones, other devices), social media and machine generated data are all changing our data environments.

Sybase IQ now claims more than 4,500 installations of Sybase IQ across the globe, following a rapid growth in revenue and a large expansion of the development organization.

In addition to the recent advances in releases 15.0 through 15.4 described here, Sybase IQ retains its established

advantages in column storage, indexing and compression. These features—present since the earliest versions of Sybase IQ—work in combination to confer benefits that are unique to Sybase IQ. While other products offer column storage and compression, no other product has the sophistication of Sybase IQ in integrating these features with advanced indexing and query optimization. The result is that Sybase IQ is particularly efficient in reducing the amount of data that must be read to satisfy queries. These fundamental strengths are now combined with increased parallelism and other features to deliver product benefits in a wider range of applications, now including those that use advanced analytic methods, including MapReduce and that involve interaction with big data in Hadoop clusters.

For the complete Winter Corp white paper please visit:
<http://www.sybase.com/sybaseiq-elastic>

For more information on SAP Sybase IQ go to: <http://sybase.com/SybaseIQ>

WinterCorp is an independent consulting firm expert in the architecture and scalability of big data and analytic database solutions. Through their in-depth knowledge and experience, they deliver unmatched insight into the issues that impede scalability and into the technologies and practice that enable business success. ■

WINTER CORPORATION

For more information on Winter Corporation and Richard Winter please visit www.wintercorp.com



Moving Data In Today's Real-Time Business

**New Study Finds 1 in 4 Face Confidence Crisis
Due To Data Timeliness**

GET YOUR 2012 DATA INTEGRATION REPORT

Resources.Attunity.com/GetReport

ACCORDING TO A RECENT SURVEY conducted by Attunity and Unisphere Research, one-fourth of companies face a confidence crisis among BI and analytics users in terms of data timeliness.

And in companies where larger volumes of data must be moved, that confidence level drops even further. One-third of organizations that move more than 1TB of information per day report low confidence among data consumers.

To reach business insights faster, decision makers need access to data as rapidly as possible. Some IT teams fear that enabling real-time data for BI and analytics is an unattainable dream.

That is simply not reality. Look to Attunity for the solution.

TRADITIONAL ETL SOLUTIONS AREN'T THE ANSWER FOR REAL-TIME BI AND ANALYTICS...

Despite the overwhelming desire for timely information, half of the companies in Attunity's survey report that most of their enterprise data is more than a day old. This suggests that many organizations are still loading data in batches overnight.

ETL tools aren't suitable for delivering large amounts of data in a timely way. Batch processing is time consuming and

operationally complex. When IT teams use ETL products for real-time data movement, it results in labor intensive processes.

CHANGE DATA CAPTURE (CDC) AND REPLICATION TOOLS ARE ONE ANSWER FOR LOW-LATENCY DATA MOVEMENT

IT teams are recognizing that CDC and data replication tools are good solutions for enabling real-time data for BI and analytics.

The Unisphere Research survey found that half of organizations moving 100GB to 1TB of data daily are using data replication technologies.

Three key points to remember:

- *CDC and replication tools are a good complement to batch processes.*
- *Log-based CDC tools are best for low-latency data movement.*
- *CDC and data replication are natural choices for supporting cloud-based BI and analytics.*

ATTUNITY

For more information, visit www.Attunity.com or call (800) 288-8648 (toll-free) / +1 (781) 730-4070.

'We've doubled our revenues thanks to Attunity and improved our reporting to support business operations.'

—IT Director,
Atlantic Detroit Diesel-Allison

'With Attunity, it now takes about 10 minutes to produce a report—an improvement of about 86 percent over what users faced in the past.'

—IT Director,
AUTOonline



Data Virtualization Redefines the Stock Exchange

REMAINING INNOVATIVE IN A COMPETITIVE ENVIRONMENT

NYSE Euronext operates the world's leading and most liquid equities and derivatives exchanges. The company is comprised of six equity exchanges and eight derivatives exchanges located in the U.S. and Europe, including the New York Stock Exchange (NYSE), NYSE Arca, NYSE Liffe, Euronext and American Stock Exchange (AMEX).

Over 8,500 issues are listed on NYSE Euronext exchanges and cover an extensive and diverse set of products, such as stocks, bonds, exchange-traded funds (ETFs), exchange-traded notes (ETNs), options, open funds, warrants, commodity futures and other derivative products. The exchanges handle over four billion transactions per day with an average daily value of \$153 billion.

DEALING WITH COMPLEXITY

The Global Data Services team at NYSE Euronext evaluates, designs, develops, and implements services, technologies, and architectures for the entire firm. Emile Werr, Vice President of Global Data Services and Global Head of Data Architecture and Identity Management, stresses the demand for change: "The need to remain innovative in a competitive environment while

transforming the business exacerbates the inherent overall level of complexity. In addition, all of the information generated by these new business services has to funnel into the NYSE Euronext data delivery environment."

A major challenge for NYSE Euronext is the sheer complexity of its business and operating environment. There are several contributing factors:

1. Many mergers and acquisitions in its evolution.
2. NYSE Euronext trades 14 different types of products, from the equities market to complicated derivatives, plus commodities and futures all having different data structures.
3. NYSE Euronext deals with massive data volumes, producing an aggregate of 2TB per day across all of its exchanges and markets.
4. The need to meet rigid service-level agreements (SLAs) with business units for delivering and retaining data.

COMPOSITE SOFTWARE DATA VIRTUALIZATION PLATFORM SELECTED

The NYSE Euronext solution is an enterprise-wide data virtualization layer, built using Composite Software's Data Virtualization Platform, which functions as a virtual data warehouse to provide

access to post-trade data for analysis and reporting. The solution has been named TORQCA (pronounced "torkah"), an acronym for the major data transactions that comprise the business: trades, orders, reports, quotes, cancels and administration (admin messages).

The NYSE Data Virtualization Solution Architecture is shown in Figure 1.

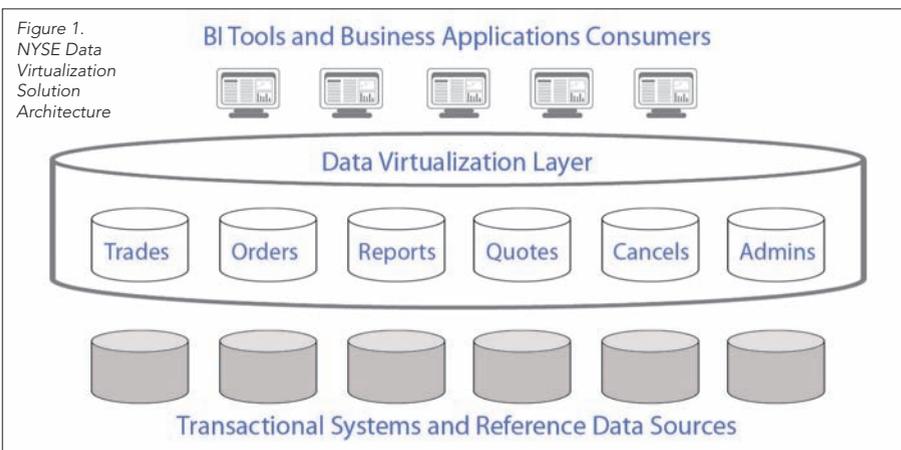
DELIVERING COMPLETE, HIGH QUALITY, ACTIONABLE INFORMATION

"External customers need fast access to this information. With data virtualization, we have the opportunity to engineer the application properly, use the appropriate technology stack, reduce cost, and improve performance. These enhancements will result in increased value and good visibility for the firm," says Werr.

In the NYSE Euronext environment, the data virtualization layer has many functions and roles. It provides all of the following:

- Standardized data access for connectivity to all data sources.
- Virtual data warehouse for federating data through logical views.
- Centralized repository for common metadata, application/ business logic and connectivity, and data services.

Changes to any of these centralized components are made in the data virtualization layer, eliminating the need to redeploy code to applications. The data virtualization layer also makes the environment extensible and provides significant opportunities for reuse. These are all important drivers for faster systems development going forward. The NYSE Euronext is anticipating savings of over \$4.5 million annually. ■



COMPOSITE SOFTWARE, INC.

is the data virtualization gold standard.
www.compositesw.com



Top 4 Data Challenges Addressed by Data Virtualization

DATA IS NOW BIG! Business users have access to an unprecedented volume of data that can potentially enable them to create new products and reach new customers and markets at unbelievable speeds. But ironically, the same data poses critical challenges that stop them from getting there. Data Virtualization helps companies to create a unified view of data and information by addressing 4 key challenges.

1. **“I cannot get to it”**—No access to the wealth of information in disparate internal silos and new and unstructured sources like big data.
2. **“It’s not current”**—Data delivered is not current, but a day or a week old.
3. **“I cannot understand it”**—Semantic views of data are not available and it is in a format that is hard to use.
4. **“It’s taking too long”**—Divergent priorities of IT and business lead to data requests taking weeks or months to clear and hamper many engagements.

In spite of substantial investments, traditional data integration and management tools do not address these challenges as they are too rigid and cannot deal with increasing data complexity, unnecessary replication and long batch windows. There is a burning need for a flexible integration solution that can access all required data from any source, transform it and deliver it to its destination on time and in a consumable format.

ADDRESSING THE DATA CHALLENGE: DATA VIRTUALIZATION

Data Virtualization is a technology which combines disparate data sources into a single “virtual” data layer that provides unified access and integrated on-demand data services to consuming applications in real-time (or right-time).

Data Virtualization is helping companies shift focus to using any and all data rather than storing, replicating, controlling etc. This is seeing fast

adoption by large firms to solve key data related challenges with new capabilities:

Access any data: Irrespective of whether the data is internal or external to the company, on the cloud, in a big data source or on a social media website, the Data Virtualization tool can access it, structure it and conform it to existing patterns making it easier to use. For example, a marketing executive should be able to extract reviews of a product from popular discussion forums and combine this information with standard data like leads, responses for analysis and reporting.

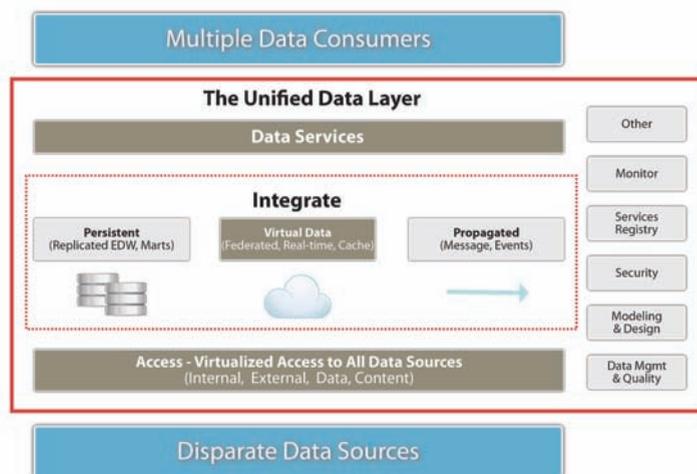
Deliver data in real-time, scheduled or cached mode as required: Business users have dynamic data needs and Data Virtualization has the flexibility to deliver in multiple modes depending on how current the data needs to be. For example, pricing analysts need real-time sales and turnover information on Thanksgiving (Black Friday Sale) while a one day delay may be acceptable on other days.

Provide meaningful information: Data Virtualization can enable business users to get a semantic view of data and also access it in multiple formats using BI tools or any enterprise application. The presence of a shared data services layer is critical in enabling use of data across analytical and operational uses with coherency.

Increase agility in data management:

The Data Virtualization capability of abstracting data from sources and publishing to multiple data consumers in real-time allows business and IT to collaborate and function iteratively, thereby reducing turnaround time for data requests considerably.

The unified data layer enabled by Data Virtualization also provides strong data governance and data security features. Data governance capabilities like monitoring and data lineage aid in quick impact analysis and error tracking, especially when data is being pulled from a variety of sources. Highly granular data security features (row and column level) ensure that there is no unauthorized access and leakage of confidential information. Thus, Data Virtualization helps a company to meet its business goals and ensures compliance with regulatory frameworks while also easing data management processes. ■



NEXT STEPS

Learn how Data Virtualization can add value to your company. Please contact info@denodo.com or call +1 877.556.2531 to discuss your next project, request a demo or evaluation copy.



5 Ways Best-in-Class BI Builds on the Industry's History, Then Throws It Out

AFTER A 30-YEAR HISTORY of the BI industry, best practices are finally emerging. And for the most part, those best practices involve throwing out the first 25 years of the industry.

Perhaps “throwing out” is too strong: the early days of BI gave us concepts and tools to build on. But a BI implementation was typically complicated, heavy, expensive and slow. Companies spent millions of dollars and thousands of hours but never attained the promise of intelligence through data, never mind intelligence when and where you need it.

Some companies are becoming more competitive by using analytics strategically. These companies follow best practices that build on the legacy of BI, then throw it out. Here are five.

GET RID OF THE QUEUE WITH SELF-SERVICE TOOLS

It seems simple: let people answer their own questions. But legacy BI systems forced people to answer basic questions about their data by submitting requests to developers who would then prioritize them, write code, and eventually deliver an answer. Any changes went through the same process. Hence the famous BI “queue” which subjected business users to long wait times and frustration.

Best practice? Get out of the queue. The people who can best answer questions are the very people asking them. The promise of self-service business intelligence is to let people create ad-hoc analytics to communicate a result, answer a question or just satisfy their own curiosity. Because the pace of business today is such that anyone standing in a queue gets left behind.

MAKE DASHBOARDS PEOPLE WANT TO USE

Legacy Business Intelligence meant creating hard-to-use dashboards and analytics tools. People had to go to

training to understand how to use even the tools for end-users, never mind developers. This created an inventory of dashboards and reports that people didn't use. The consumerization of software has taught us that people will use tools that are attractive and usable. People readily adopt new apps on their phones and new online tools if they find them useful. Why should BI be any different?

Best-in-class BI includes a focus on interactive, beautiful, and easily accessible analytics. It delivers dashboards and reports in a web browser and via mobile devices, so users don't have to jump through hoops simply to see their data. It allows business to design analytics that feel delightful rather than oppressive. As a result, people use those analytics—and the business gets smarter.

ACKNOWLEDGE THE REALITY OF DATA

Legacy business intelligence had a concept of a single data architecture that was fast, scalable and held only clean data. The problem is that no business exists with this ideal data architecture. Real businesses have multiple data sources of varying types and capabilities.

The best practice for the next generation of BI is to provide a way to work with all of your data, from spreadsheets to the most sophisticated databases and data warehouses, and even new data sources like Hadoop. And even to blend different data sources, since the answer to a question is rarely well-behaved enough to sit obediently in a single database. Blending sales data with operational and finance data yields answers that are much richer and more useful than analyzing single data silos.

WRANGLE BIG DATA

Here's the promise of big data: to find intelligence in the outliers that aggregated data loses. To understand new customer

behaviors with new sources of data like sensors and web analytics. To work with all your data, rather than reduce it into a schema that might prevent you from asking new questions in the future. But if legacy BI forced the normal user into a queue for a report, working with big data meant big projects, new data warehouses and a lot of time and money.

Best practice BI provides tools for working with big data just like any data. Fast in-memory extracts, filtering, and connections to optimized data sources mean that users keep the ability to ask and answer questions—even when the questions are big.

INTEGRATE WITH YOUR EXISTING SYSTEMS

A traditional BI implementation was such a grand undertaking that companies would sometimes rip out existing systems or accept painful tradeoffs that limited the potential of their BI or other infrastructure.

Best practice BI is friendly. It works with your existing security and data architecture, and lets you choose whether you want to live connect to data or work in-memory. Because it's hard to achieve best-in-class analytics when you're hung up implementing it.

CONCLUSION

The new world of business intelligence builds on and breaks from the old. These best practices are followed by organizations working at the pace of modern business and getting smarter every day. ■

TABLEAU SOFTWARE helps people see and understand data. Used by more than 7,000 organizations worldwide, Tableau's award-winning software delivers fast analytics and rapid-fire business intelligence. For a free trial, visit www.tableausoftware.com



Why a Columnar Data Platform is Best for Big Analytics

BIG DATA TRANSLATES TO BIG ANALYTICS

With the exponential data growth the world has experienced over the past several years, the variety and velocity of new data has affected transactional, operational, and analytic processes, tools, and data environments. Recent research has uncovered that fifty percent or more of Big Data deployments are born out of the need for better, deeper analytics. This phenomenon is creating new business opportunities, as companies begin to see the tremendous value in analyzing the data for competitive and strategic advantage. It is also putting more pressure on IT departments, as they are tasked with improving their computing infrastructure, making it faster, more scalable, and cheaper to reap the benefits that data analytics promise.

THE POWER AND SCALE OF COLUMNAR DATA PLATFORMS

Big Data and the associated big analytics that are fueling these new business opportunities have created some challenges in the day-to-day work of DBAs. Demands for more data from a wider pool of business users are on the rise, equating to more queries against larger data sets.

To support the emergence of “analytics for everyone,” purpose-built columnar analytic platforms have been developed to address differing data workloads, improving the value of OLTP, operational uses, and BI/analytics. Columnar data technology is one of many new approaches taking workloads off the star schema data warehouse, which is where many of the I/O bound queries are today. Heterogeneity in post-operational systems is going to be the norm for some time, and columnar data platforms are a major reason because they can outperform many of the queries executed in the traditional data warehouse.

This new level of scalable performance enables analytics against billions of rows of data while returning results in seconds.

INFINIDB®—PERFORMANCE AT SCALE LIKE NEVER BEFORE

InfiniDB is a columnar data platform optimized for high-performance analytics against massive data sets. Whether data discovery, predictive analytics, or dimensional analytics, InfiniDB scales performance linearly with growing data. It executes queries extremely fast, returning results that used to take hours, in minutes or seconds. MapReduce, NoSQL, and row RDMS-only offerings are not optimized for such workloads. InfiniDB’s unique columnar design eliminates all wasted I/O to disk, executes SQL queries as parallelized map and reduction operations to achieve linear scale, and optimizes queries across all CPU threads—and it does so out-of-box and on any hardware.

THE COLUMNAR DATA PLATFORM IN ACTION

The implications of such analytic power and scale are far-reaching and can be applied to any organization in any industry with Big Data challenges. Imagine real-time analysis of point-of-sale data in retail, behavior analysis in digital and social media for more targeted advertising, stock and ticker data analysis in financial services for risk management, or analysis of call records in telecommunications to avoid churn.

ASI Group (ASI) is a software company specializing in the aviation and postal industries that has turned to columnar data technology to address analytics. ASI receives detailed logistics data about cargo, baggage, and mail in real-time, stores the information, and makes analytics on this data available over the web to its customers. Determining when

to release a flight full of mail means taking into account weather conditions, arrival time, the volume and pick-up location of that mail, contracts and service levels, tariffs, possible penalties, and costs. Analysts scour the data to make the right decision for the airline. It is imperative that analysts have their questions answered in real-time so they can make accurate decisions quickly to meet on-time demands and profit goals.

To address their analytics challenges, ASI adopted InfiniDB. InfiniDB brought the fastest performance to data loading and query execution of the alternatives ASI researched. It removed the bounds to dimensional limiters or limits of scale, while also providing easy management for its data administrators. The InfiniDB instance enables query scans on billions of records in minutes or seconds, depending on the query. The scan queries address the thousand-plus columns in the tables, focusing on a several columns at a time to continually check the status of all mail coming in and out of a city, for example. This is an ideal situation for a columnar data platform like InfiniDB, as ASI’s row-based database could neither meet its query performance nor data load requirements. With InfiniDB, ASI is able to deliver meaningful analytics with the scale and performance its customers need to make immediate decisions that impact their business. ■

InfiniDB
Scalable. Fast. Simple.

CALPONT

InfiniDB—The Simple, Powerful Platform for Big Data Analytics
www.calpont.com/ @Calpont



Getting Big Answers Out of Big Data

The Key to Business Innovation

DATA IS A COMMODITY. Not only is it easy to gather massive amounts of data about almost anything and everything; it has also become relatively inexpensive to store this data. Today, the average consumer can purchase an external terabyte hard drive for about \$100. The challenge is no longer collecting data; the challenge is how to decipher all the data that has been collected. How do we leverage this information to support mission-critical objectives; make better business decisions and predictions, prevent crime, catch bad guys and more accurately turn this information into insight and advantages towards understanding future opportunities?

CHANGING HOW WE THINK ABOUT DATA

Traditionally, data has been stored in rows and columns, or focused on structured information. By storing data in a structured format, we potentially limit our ability to search out information deep within the database which may be relevant to our business needs. Today, with NoSQL technologies, more and more data is being stored in a semi- or un-structured format, allowing for deeper queries within the database to access relevant details specifically associated to the business decisions at hand. Gartner estimates upwards of 80% of data in the next 5 years will be stored this way.

THE RIGHT TOOL FOR THE JOB

With the use of NoSQL technology, databases have the ability to become more dynamic and flexible. Searches can become limitless and with the ease of semi- and un-structured data ingest, it is faster to sort through new information in real-time. NoSQL also enables users to discover unforeseen outcomes by querying with greater complexity. Simply put, you get answers to what you

don't know. By storing all data as first class objects or "nodes" with relationship or "edge" information, users are able to ask ambiguous questions within large datasets and get actual answers in return. Technologies such as Object Oriented and Graph Databases improve our chances of finding hidden advantages and opportunities within existing datasets.

Organizations now have the opportunity to bring together large amounts of any type of data to be analyzed immediately. Being able to see the forest, versus focusing on a section of trees, greatly enhances analysis and intelligence capabilities by enabling deeper access to greater details. Connecting the dots in real-time yields amazing advances in knowledge, especially for time-sensitive applications such as defense and security, banking, oil and gas, and medical.

THE BIG DATA CONNECTION STORE

Objectivity, Inc., the leader in NoSQL technologies, enables organizations to develop massively distributed and scalable mission-critical applications to handle Big Data and countless billions of distributed data objects on a global scale, to help get answers and make decisions, faster. The unique, complementary nature of Objectivity's products allows customers to capitalize on existing infrastructure while still taking advantage of the latest NoSQL technologies to realize the benefits of real-time Big Data and Relationship Analysis.

InfiniteGraph is the only commercially available, massively distributed and



scalable Graph database available today. **Objectivity/DB** is a distributed and

scalable data and object management system which supports advanced government and enterprise systems.

OBJECTIVITY TECHNOLOGY BENEFITS:

- Enables real-time relationship analytics, network analysis, data fusion, and metadata index management.
- Provides a single logical view and common operating picture.
- Supports any data type and format, including key-values, RDF triples, documents, structured, semi / unstructured, text, media and other data files.
- Beyond petabyte scale and data volumes.
- Efficient real-time performance with reduced hardware and resource overhead.
- Offers a scalable solution capable of working in a stand-alone environment, through networked systems up to cloud enabled enterprise systems all using the same product and programming interface.
- InfiniteGraph uses memory and cache to help you get the best performance on your live data, but also persists relationship information to disk, for durability and large scale analytics.

We've evolved beyond data collection to data comprehension—the key to real innovation. Answering unknown questions and connecting the dots between disparate, but relevant information leads us to the next "big" opportunity.

Objectivity delivered on this promise through its work in the defense, healthcare, financial services, telecom, and technology industries. ■

OBJECTIVITY

To learn how you can discover your data, visit www.Objectivity.com

Progress | DataDirect.

Make Fast and Informed Decisions Based on Up to the Minute Data

IN THE COMING YEARS, enterprise organizations will realize significant business benefits from Big Data, but unfortunately few organizations today are capable of accessing the full scope of Big Data. Most have implemented Business Intelligence (BI) Solutions to exploit Big Data for critical business insights that translate into competitive advantages. But BI Solutions are only as good as their access to your critical data. Here are some things to consider when selecting the data connectivity solution to make the most of your BI investment:

FLEXIBILITY

DataDirect Solutions for BI make it easy to access all of your data, no matter how obscure or where it is located. This data gives you a complete picture of your business, helping you to make the smartest business decisions. Here's how it works:

- Easily access any kind of data across your organization, no matter its type, format, or location.
- Both internal and external data sources can be integrated into your BI queries.
- Leverage your BI application deployments from any platform.
- Avoid paying for limited, low-quality, high-cost data connectivity solutions from the BI vendor.

PERFORMANCE

Your BI Solution should be able to expand and contract along with your business needs. DataDirect products help minimize the CPU and memory footprint of your BI application while allowing it to support more simultaneous users and database connections with the same or less hardware. Here are some of the performance-tuning features:

- Connection pooling
- Thread pooling
- Statement pooling

- Network packet management
- Caching and data buffering optimization for reducing network roundtrips
- TCP/IP socket management for greater CPU efficiency
- Promotion of parameter array, batch insert operations to bulk load for fastest available performance for mass insert operations

SECURITY

Your information is among your most valuable assets, and its protection is vitally important. DataDirect products provide complete end-to-end security for your data, and ensure compliance with key industry regulations and security initiatives. User authentication and access authorization features help effectively manage risk.

DataDirect supports the industry's best database and application security protocols and features including:

- Network data encryption via SSL
- OS authentication via Kerberos
- Secure architectural framework

SUPPORT

Your BI Solution should run effortlessly and require minimal maintenance to address memory leaks, performance problems and other bugs. DataDirect products ensure that any data analyzed by the BI solution isn't corrupted or mangled. DataDirect assures product quality by using the industry's largest and most robust test suites. Our testing framework features:

- Millions of test cases ensure reliability on all supported database and platform versions
- All major client operating systems, including Windows, Sun Solaris, HP-UX, IBM AIX, and Linux
- Multiple hardware chipset architectures including Intel Itanium, and AMD/Intel x64

- All major database versions, including all supported versions of Oracle and DB2 on Windows, Linux, UNIX, AS/400, and z/OS
- Over 50 CPUs on UNIX, Windows, Linux, AS/400, including a mainframe
- Constant regression testing to guarantee every fix is included in every product version

ENHANCED FUNCTIONALITY

You need to be able to add important new functionality to your BI Solution without having to wait and schedule time for complicated upgrades. DataDirect products make it easy to add advanced features without the need for application patches, changes, or costly database options.

Here are just some of the enterprise quality features that are 100% driver implemented:

- Application failover
- Load balancing
- Bulk loading
- Advanced connection pooling
- Statement pooling

PROGRESS DATADIRECT

Big Data is here to stay—there is no denying this fact. And as enterprise organizations attempt to reap the benefits of Big Data by using BI Solutions, they must come to grips with the inherent limitations of the existing data connectivity tools on the market today.

Progress DataDirect Connect is the industry leader in high-performance, scalable, and reliable data access. Our solutions for BI will improve your decision making abilities, providing key information efficiently and quickly. ■

PROGRESS DATADIRECT

For more information, visit www.datadirect.com/platinumODBC



Apache Hadoop: A Big Data Refinery Enabling Big Analytics

THE TERM **BIG DATA** has become a popular way to describe data that has a volume, velocity or variety that overwhelms traditional enterprise data infrastructures. New technologies have emerged to handle Big Data that complement the existing technologies. The leading technology is an open source project called Apache Hadoop. Hadoop provides large-scale storage and processing capabilities for nearly any kind of data. It was born out of the large web properties because they needed a low cost way to process data that was being collected from the Internet.

One of the main benefits of Hadoop is the unmatched flexibility for storing and processing data. Hadoop is broadly applicable to unstructured data problems and is being used for a wide variety of use cases from fraud detection, risk analysis and pattern recognition to recommendation engines, oil exploration and proactive policing.

At the core of Hadoop are two frameworks, which, when combined provide significant new capabilities to the enterprise. The first is a scale-out file system called the Hadoop Distributed File System (HDFS). This presents a large number of servers with direct attached disks as a large share storage pool with built in redundancy and failover. The second component is a distributed processing framework called MapReduce. The combination of these two capabilities makes Hadoop extremely powerful for preparing data for analysis.

A DATA REFINERY

Analyzing Big Data presents new challenges and often involves combining data with vastly different structures, think video, audio and click stream and transactional data, in order to find new insights. Hadoop is an excellent engine to take this multi-structured data and prep it for further analysis in tools that

*One of the
main benefits
of Hadoop is the
unmatched
flexibility
for storing and
processing data.*

require more structure in the data. In this case you can think of Hadoop as a data refinery, combining and processing data to varying levels of structure so that it is readily accessible by analytics tools.

Capturing the value of Big Data requires turning raw data into new usable formats for further analysis, just like the capturing the value in a natural resource like crude oil depends on our

ability to turn it into usable end products. Hadoop takes data and transforms and refines it into intermediate forms or into fully refined and analyzed forms.

HADOOP AND THE LARGER ENTERPRISE DATA ARCHITECTURE

Despite a significant amount of hype to the contrary, Hadoop complements existing infrastructures like data warehousing and ETL. By refining unstructured data it will make it possible to process more data through these technologies.

Hadoop also complements existing analytics and BI platforms, in essence extending them to the new world of Big Data. Virtually every vendor has announced support for Apache Hadoop, typically via a connector that enables refined data to be analyzed via the same tools that business analysts, data scientists and data engineers use on a daily basis.

If you have not yet started to address the Big Data opportunity to analyze your business in new ways, I strongly suggest that you begin doing your homework on Apache Hadoop. Given the current nature of Hadoop adoption, your competitors probably already have. ■

HORTONWORKS

For more information send an email to info@hortonworks.com.