

# database

TRENDS AND APPLICATIONS

## Oracle's Ecosystem: *Accelerating* Ever Faster Into the Real-Time Enterprise 4

WWW.DBTA.COM

Moving From Maintenance to Innovation 2

The Dangerous Questions IT Never Wants to Hear 29

MongoDB Continues to Advance 32

# Data entry errors. Consolidation headaches. Quality data shouldn't be a pain.

Try a little Vitamin Q.



We supply Vitamin Q – “quality” – in data quality management solutions that profile, clean, enrich, and match your customer data – and keep it healthy over time. Add Vitamin Q to your data integration, business intelligence, Big Data, and CRM initiatives to ensure accurate data for greater insight and business value.

★  
**ACT  
NOW**  
★

Get a **FREE 30-day** trial!  
[www.MelissaData.com/vitaminQ](http://www.MelissaData.com/vitaminQ)



Solutions for  
240+ Countries



10,000+ Customers  
Worldwide



30+ Years  
Strong



Data Quality &  
Mailing Solutions



Cloud • On-Premise  
• Services



Germany  
[www.MelissaData.de](http://www.MelissaData.de)

United Kingdom  
[www.MelissaData.co.uk](http://www.MelissaData.co.uk)

India  
[www.MelissaData.in](http://www.MelissaData.in)

Australia  
[www.MelissaData.com.au](http://www.MelissaData.com.au)

**MELISSA DATA®**

Your Partner in Global Data Quality

[www.MelissaData.com](http://www.MelissaData.com) | 1-800-MELISSA



**PUBLISHED BY** Unisphere Media—a Division of Information Today, Inc.  
**EDITORIAL & SALES OFFICE** 121 Chanlon Road, New Providence, NJ 07974  
**CORPORATE HEADQUARTERS** 143 Old Marlton Pike, Medford, NJ 08055

Thomas Hogan Jr., Group Publisher  
609-654-6266; thoganjr@infotoday

Joyce Wells, Managing Editor  
908-795-3704; Joyce@dbta.com

Joseph McKendrick,  
Contributing Editor, Joseph@dbta.com

Stephanie Simone, Editorial Assistant  
908-795-3520; ssimone@dbta.com

Adam Shepherd,  
Advertising and Sales Coordinator  
908-795-3705; ashepherd@dbta.com

Celeste Peterson-Sloss, Lauree Padgett,  
Alison A. Trotta, Editorial Services

Tiffany Chamenko,  
Production Manager

Lori Rice,  
Senior Graphic Designer

Jackie Crawford,  
Ad Trafficking Coordinator

Sheila Willison, Marketing Manager,  
Events and Circulation  
859-278-2223; sheila@infotoday.com

DawnEl Harris, Director of Web Events  
dawnel@infotoday.com

## COLUMNISTS

Gerardo Dada, www.solarwinds.com  
Guy Harrison, guy.harrison@quest.com  
Kevin Kline, Kevin\_Kline@dbta.com

Craig S. Mullins, www.CraigSMullins.com  
Todd Schraml, TWSchraml@gmail.com

## ADVERTISING

Stephen Faig, Business Development Manager, 908-795-3702; Stephen@dbta.com

## INFORMATION TODAY, INC. EXECUTIVE MANAGEMENT

Thomas H. Hogan, President and CEO  
Roger R. Bilboul,  
Chairman of the Board

John C. Yersak,  
Vice President and CAO

Thomas Hogan Jr., Vice President,  
Marketing and Business Development  
Richard T. Kaser, Vice President, Content  
Bill Spence, Vice President,  
Information Technology

*DATABASE TRENDS AND APPLICATIONS* (ISSN: 1547-9897; USPS: 16230) is published bimonthly (Feb./Mar., Apr./May, Jun./Jul., Aug./Sep., Oct./Nov., and Dec./Jan.) by Unisphere Media, a division of Information Today, Inc., 143 Old Marlton Pike, Medford, NJ 08055 USA; Phone (609) 654-6266; Fax (609) 654-4309; Internet: infotoday.com. Registered in U.S. Patent & Trademark Office. Periodicals postage paid at Vincentown, NJ, and additional mailing offices.

© Copyright, 2016 Information Today, Inc. All rights reserved.

No part of this publication may be reproduced in whole or in part in any medium without the express permission of the publisher. PRINTED IN USA

POSTMASTER Send address changes to *Database Trends and Applications*, P.O. Box 3006, Northbrook, IL 60065-3006.

## RIGHTS AND PERMISSIONS

Permission to photocopy items is granted by Information Today, Inc. provided that a base fee of \$3.50 plus \$0.50 per page is paid directly to Copyright Clearance Center (CCC), or provided that your organization maintains an appropriate license with CCC.

Visit [www.copyright.com](http://www.copyright.com) to obtain permission to use these materials in academic coursepacks or for library reserves, interlibrary loans, document delivery services, or as classroom handouts; for permission to send copies via email or post copies on a corporate intranet or extranet; or for permission to republish materials in books, textbooks, and newsletters.

Contact CCC at 222 Rosewood Drive, Danvers, MA 01923; (978) 750-8400; Fax: (978) 646-8600; [www.copyright.com](http://www.copyright.com). If you live outside the USA, request permission from your local Reproduction Rights Organization. (For a list of international agencies, consult [www.ifro.org](http://www.ifro.org).)

For all other requests, including making copies for use as commercial reprints or for other sales, marketing, promotional and publicity uses, contact the publisher in advance of using the material. For a copy of our Rights and Permissions Request form, contact Lauree Padgett, lpadgett@infotoday.com.

ONLINE ACCESS Visit our website at [www.dbta.com](http://www.dbta.com)

Searchable archive of all articles with digital document delivery: [www.iti-infocentral.com](http://www.iti-infocentral.com)

Contents also available online under direct licensing arrangements with EBSCO, NewsBank, ProQuest, and Gale and through redistribution arrangements with information service providers including, Dow Jones Factiva, LexisNexis, OCLC, STN International, and Westlaw.

## SUBSCRIPTION INFORMATION

Subscriptions are available free to qualified recipients in the U.S. only. Nonqualified subscribers in the U.S. may purchase a subscription for \$69.95 per year. Delivery outside North America is \$135 via surface mail per year. All rates to be prepaid in U.S. funds. Subscribe online ([circulation@dbta.com](mailto:circulation@dbta.com)) or write Information Today, Inc., 143 Old Marlton Pike, Medford, NJ 08055-8755.

Back issues: \$17 per copy, U.S.; \$22 per copy, Canada and Mexico; \$27 per copy outside North America; prepaid only. Missed issues within the U.S. must be claimed within 45 days of publication date.

Change of Address: Mail requests, including a copy of the current address label from a recent issue and indicating the new address, to *DATABASE TRENDS AND APPLICATIONS*, P.O. Box 3006, Northbrook, IL 60065-3006.

Reprints: For quality reprints of 500 copies or more, call (908) 795-3703 or email [reprints@dbta.com](mailto:reprints@dbta.com).

**DISCLAIMERS** Acceptance of an advertisement does not imply an endorsement by the publisher. Views expressed by authors and other contributors are entirely their own and do not necessarily reflect the views of the publisher. While best efforts to ensure editorial accuracy of the content are exercised, publisher assumes no liability for any information contained in this publication. The publisher can accept no responsibility for the return of unsolicited manuscripts or the loss of photos. The views in this publication are those of the authors and do not necessarily reflect the views of Information Today, Inc. (ITI) or the editors.

EDITORIAL OFFICE 121 Chanlon Road, New Providence, NJ 07974

List Rental: American List Council. Contact Michael Auriemma, Account Manager, (914) 524-5238 or email [Michael.auriemma@alc.com](mailto:Michael.auriemma@alc.com)

## FEATURES

RESEARCH@DBTA

### 2 MOVING FROM MAINTENANCE TO INNOVATION

By Joe McKendrick

FEATURE STORY

### 4 ORACLE'S ECOSYSTEM: ACCELERATING EVER FASTER INTO THE REAL-TIME ENTERPRISE

By Joe McKendrick

## READERS' CHOICE AWARDS 2016 > SPECIAL SECTION

### 9 DATABASE TRENDS AND APPLICATIONS READERS' CHOICE AWARDS 2016

## DEPARTMENTS

TRENDS

### 27 GAINING SUCCESS WITH BIG DATA

By Joyce Wells & Stephanie Simone

APPLICATIONS

### 28 FIVE SQL SERVER DATABASE PERFORMANCE TIPS

By Vicky Harp

### 29 THE DANGEROUS QUESTIONS IT NEVER WANTS TO HEAR

By Keith Bromley

MULTIVALE SOLUTIONS

### 26 OPENQM ADDS PYTHON SUPPORT

### 26 ENTRINSIK PARTNERS WITH ATCORE

## COLUMNS

### 31 NEXT-GEN DATA MANAGEMENT > BY GERARDO DADA THE CORRELATION BETWEEN PERFORMANCE AND COSTS

### 32 BIG DATA NOTES > BY GUY HARRISON MONGODB CONTINUES TO ADVANCE

### 33 DBA CORNER > BY CRAIG S. MULLINS IMPROVING IT SECURITY WITH DATABASE AUDITING TECHNIQUES

### 34 IOUG OBSERVATIONS > BY MARIA ANDERSON PARTNERSHIPS AND COLLABORATION: A NEW IT DIRECTION

### 35 SQL SERVER DRILL DOWN > BY KEVIN KLINE SQL SERVER 2016 IS SIMPLY FASTER ... A LOT FASTER

### 36 DATABASE ELABORATIONS > BY TODD SCHRAML BEFORE WE GET TO NORMAL

MEDIA PARTNER OF THE FOLLOWING USER GROUPS



# Moving From Maintenance to Innovation

THERE IS A TUG-OF-WAR OF SORTS GOING ON IN ORGANIZATIONS TODAY. On one hand, there is pressure on IT staff to maintain systems' uptime and availability along with a growing focus on data security, while dealing with the multitude of maintenance-level tasks, such as applying upgrades, fixes, and patches. But on the other hand, there is a growing requirement for IT to support the business as it seeks to use data in new ways for strategic benefit. What's needed now are more efficient approaches to data management so that greater time, budget, and overall resources can be directed toward innovation and the holy grail of "competitiveness on analytics."

These are some of the insights gleaned from a new survey of data managers and pro-

fessionals, conducted by Unisphere Research, a division of Information Today, Inc. The survey finds that data managers and their organizations are hampered by routine database management activities, which demand a considerable amount of both time and budget. Conducted in partnership with EMC among members of the Independent Oracle Users Group (IOUG), the survey drew input from respondents at organizations of all sizes and across various industries.

The trade-off is straightforward: Constantly dealing with mundane tasks translates to less time and resources available for higher-level endeavors such as working on business problems or moving to digital platforms. In terms of IT budgets, the survey shows that budgets are tight, with

---

By Joe McKendrick

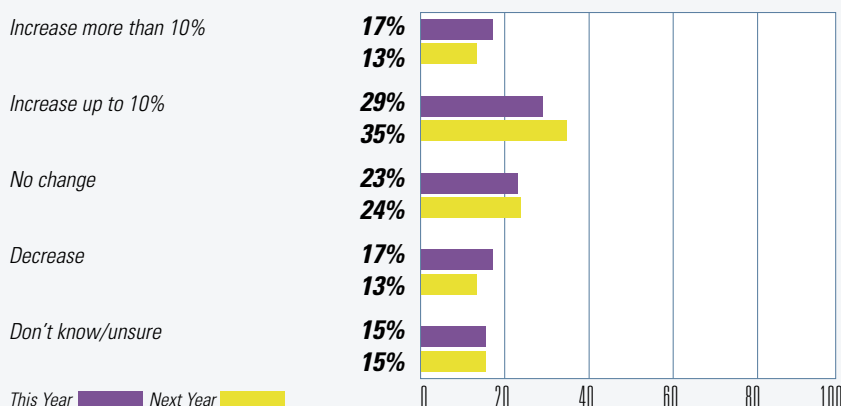
---

traditional database maintenance activities taking a large portion of resources. Budgets have risen only somewhat over the past year, with no significant uptick expected in the year ahead, meaning that IT staffs are caught in a squeeze (Figure 1).

The pressure is on. As decision makers seek to achieve competitive advantage by leveraging information from a panoply of connected systems, devices, and data sources, the job of managing the integration across myriad networks, data systems, and applications to deliver reliable information is becoming a greater challenge. Today, businesses want to be able to track and predict events and identify trends as they are happening or even before they occur and to be connected to an Internet of Things from which data can feed into analytics systems. Indeed, data managers themselves would prefer to be involved in projects that increase value to their businesses and enhance their own skills and careers, but they find that they are often mired in low-level tasks, with limited resources to focus on new initiatives (Figure 2).

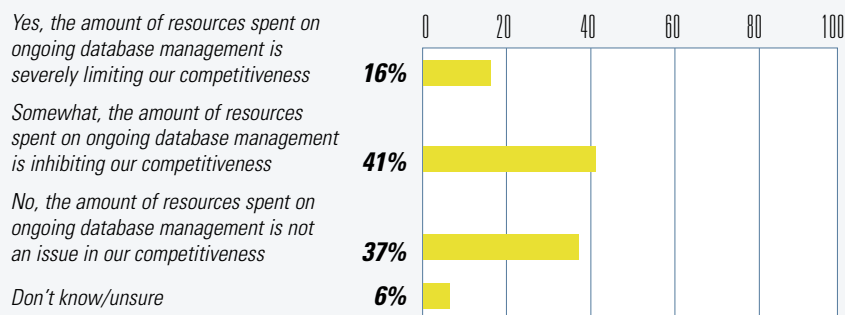
While data management departments must increase their output and productiv-

**Figure 1: How has your organization's overall IT budget changed?**

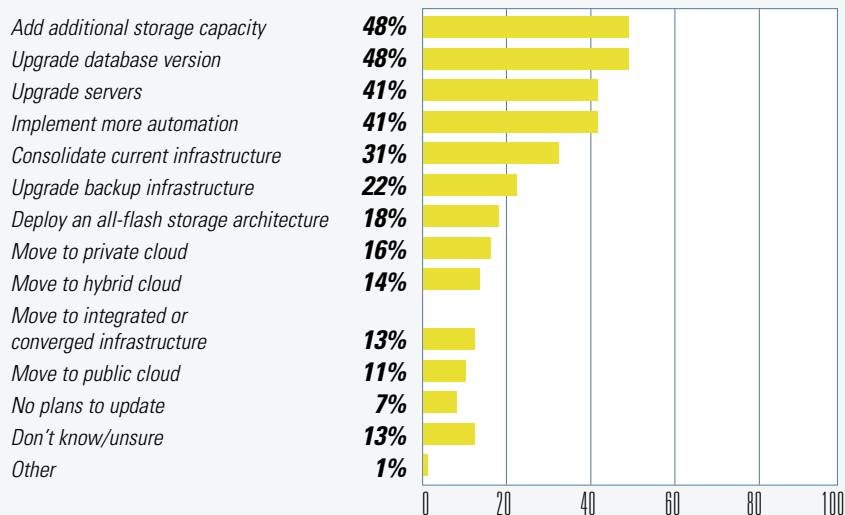




**Figure 2: Do you believe the amount of time, money, and resources spent on ongoing database management—versus new project development or new initiatives—is affecting the competitiveness of your organization?**



**Figure 3: How do you plan to update your database infrastructure over the next 12 months?**



ity, the survey also finds that efficiency on its own will not suffice. Instead, organizations need IT and data managers to be able to make innovation a greater part of their time and focus.

Many organizations continue to deal with demanding database and data center requirements by simply throwing more hardware at the problem, including stor-

age arrays and servers. However, there is an increasing focus on more innovative approaches that heighten efficiency (Figure 3).

Consolidation, automation, and virtualization are three approaches that many respondents feel are important initiatives for the future of their organizations. Database or data center consolidation was cited

as extremely (22%) or somewhat (48%) important; database and data center automation as extremely (30%) or somewhat (38%) important; and greater use of virtualization for mission-critical databases and applications was tapped as extremely (24%) or somewhat (36%) important. The funding that could be enabled by cost savings from automation, consolidation, and virtualization would be enough to fund more business-driven initiatives, according to more than two-fifths of respondents.

In addition, implementing a standardized approach for infrastructure across databases can play a major role in the reduction of repetitive and mundane tasks, improvement in application performance and reliability, and in decreasing costs.

The survey finds that the era of operating silos by database brand is winding down, and there is a trend toward consolidating databases to gain efficiencies and savings. More universal approaches to managing the overall asset base are gaining ground.

However, while convergence is underway at many organizations, survey responses also indicate that the availability of tools to cover multiple data environments lags behind the growing need for cross-platform database management. ■

# Oracle's Ecosystem: *Accelerating*

## Ever Faster Into the Real-Time Enterprise







**C**AN ORACLE AND ITS PARTNERS KEEP UP with the increasing demands of customers for real-time digital capabilities? Is the Oracle constellation of solutions—from data analytics to enterprise applications—ready for the burgeoning requirements of the Internet of Things (IoT) and data-driven businesses? For Oracle—along with its far-flung network of software vendors, integrators, and partners—times have never been so challenging. ▶▶

By Joe McKendrick



Enterprises are working overtime to address market requirements and fend off disruptors, while technology providers are working hard to keep up with these enterprises. Things are moving quickly for all levels of Oracle partners as they race to keep pace with their customers. How has Oracle been progressing on these fronts? “Speaking from public sector, we were ahead of Oracle years ago,” said Patricia Dues, president of the Oracle Applications User Group (OAUG). “We moved into mobility and data warehousing before Oracle had any products to support us. Now, Oracle is a viable technology provider for these areas and appears to be moving to the leading edge.”

As a result of these shifts, the entire database ecosystem “has had to rapidly embrace a new way of doing business,” said Rod Hamlin, vice president of global alliances and strategic partnerships with Redis Labs. “Not only are new applications demanding the price/performance gains of this new database technology, but old legacy applications must be migrated over to compete with the response times expected by today’s consumers.”

For systems integrators, this means “evolving our priorities to meet and to stay ahead of Oracle customers as they make their moves to become digital enterprises through adoption of technologies such as cloud, big data analytics, mobile, and the adoption of IoT,” said Sunder Singh, global head of Tata Consultancy Services’ Oracle Practice.

There is also an ever-expanding range of choices for enterprises, requiring constant research and

strategic choices to be made among vendors within the ecosystem. Ajay Khanna, vice president of product marketing at Reltio, said it all requires a “polyglot persistence strategy that utilizes the strengths of different storage types.” Along with a range of commercial products are “many new technology options, including Apache open source projects like Hadoop, Spark, and others like NoSQL, deep learning, and in-memory processing,” he said. “Data-driven applications built on modern data management platforms use a combination of technologies to attain big data scale for omnichannel interactions and transactions, flexible data models for agility, graphs to manage relationships, and in-memory and Spark for analytics. The right information in the right storage provides the right insights.”

As these choices proliferate, enterprise customers are looking to the Oracle ecosystem for solutions that address opportunities in big data, the cloud, and the real-time enterprise.

## BIG DATA

Nowhere is the shift to a hypercompetitive global marketplace more evident than in the challenge of managing data itself, especially as it transforms into big data. “New generations of Oracle databases are becoming adept at storing and analyzing unstructured as well as structured data,” said Singh. Big data solutions, of course, are an important piece of the Oracle ecosystem—especially when they integrate well with existing traditional data types. “The players that have been able to make their unstructured data talk to and truly inform their traditional data streams are the ones

who are easily outcompeting their peers,” agreed Bob Selfridge,

CEO and founder of TMMData.

“Making this happen—  
and,

ideally, automating the marriage of structured and unstructured data—is the holy grail of data analysis and often requires custom solutions for bridging the gaps between ETL and BI platforms.”

Perceptions about big data across the Oracle ecosystem have also evolved over time. “Big data has created an awareness in the technology community of just how valuable the data that we are collecting is, and how much we can do with it,” said Dues. “It has spread awareness to those in the business arena, and is generating innovative and productive business solutions.” Not only are business decision makers getting more involved with data management, but there’s been a convergence with open source solutions that would have been unthinkable even 5 years ago. Dues lauds Oracle’s integration with big data solutions such as Hadoop, Hive, HBase, Flume, and Kafka through the GoldenGate product. The solutions enable enterprises “to combine structured, semi-structured, and unstructured datasets in real time without compromising the performance of source systems.”

However, adept management is required to manage the deluge that comes with big data and its varied solutions—not only in terms of volume but also variety, velocity, and veracity. “Rather than focusing on the size of the data, business users should focus on the scale of the big problems that they need to solve,” Khanna said. “Data volumes, sources, and variety will keep increasing with time. IT and data-savvy users struggle with getting value out of big data projects because, despite the volume of data captured, data is unreliable, siloed, and unrelated. Once you build a reliable data foundation by bringing data together from all sources, matching, merging and cleaning it to ensure quality, relevant insights can be drawn from big data.”

Some vendors are reporting difficulties as a result of product shifts from Oracle. Reporting data “in a meaningful and useful format from the Oracle E-Business Suite can be challenging,” said Dean Jones, global vice president of Oracle EBS Solutions for Excel4apps. “The complexity of underlying Oracle data structures leaves average users dependent on IT or consulting resources to access and report on data they need, affecting report timeliness and accuracy.” Oracle’s discontinuation of premium support for Oracle Business Intelligence ►►



Discoverer also has made things difficult, Jones added. “Oracle discontinued premium support for Oracle Business Intelligence Discoverer, a commonly used reporting tool, in June 2015 and recommended that users migrate to the Oracle Business Intelligence Foundation Suite, which includes Oracle Business Intelligence Enterprise Edition, or OBIEE,” said Jones. However, OBIEE adds a great deal of complexity to BI reporting, he continued. “Implementation and the ongoing support costs and expertise needed for running OBIEE aren’t a possibility for all companies.” Discoverer’s demise and OBIEE’s complexity have actually created a window in the market for third-party tools, Jones said.

## CLOUD

Is cloud a natural next move for Oracle-related vendors as they advance into the big data realm? Many vendors think so. “Cloud adoption is a natural next step in Oracle users’ technology strategies,” said Walt Rossi, vice president of business development at Five9, which provides a cloud contact center solution integrated to Oracle Service Cloud. “Businesses previously had questions and concerns about the cloud, but reliability, security, and scalability apprehensions are minimized with Oracle’s drive to the cloud.”

Oracle Cloud’s varied services, which feature the cloud-aware Oracle 12c Database that enables movement of applications between clouds and on-premises systems, make up “a revolutionary concept that offers the advantages of managing many databases as one, yet retains the isolation and resource control of separate databases,” said Meenakshi Krishnamoorthy, general manager at Mindtree. Krishnamoorthy also lauded the features of Oracle Enterprise Manager Cloud Control 12c, which “provides a single console window to manage the entire cloud infrastructure and applications deployed in Oracle Cloud,” along with automation capabilities.

The rise of the cloud is creating a new set of challenges for ecosystem vendors—such as the handling of existing on-premise installations, and deciding when to make the cutover. “The simple fact remains that enterprises need to continue using their existing databases,” said Ken Rugg, CEO of Tesora. “Despite the strength of public cloud database offerings like Amazon Relational Database Service, incumbent data-

base providers remain in a good position to defend their respective installed bases and extend them with database as a service. At the same time, there are more database choices than ever before and the cloud is making those accessible to developers, which means that now Oracle, and every other database vendor, needs to consider how its technology fits alongside others.”

For example, Rugg continued, “it’s possible with database as a service using Open-

was tightly controlled for data ingested and even tighter for data extraction and reuse.” The move toward cloud, he said, has changed the game for Oracle and many of its counterparts, a major impact of which has been a dramatic shift in the customer base being able to use the tools and data.

ERP systems—well-known for their complexity and costs—are being usurped by cloud offerings. If anything, an ERP implementation a few years back “meant



*Enterprise customers are looking to the Oracle ecosystem for solutions that address opportunities in big data, the cloud, and the real-time enterprise.*

Stack, for both Oracle 12c and 11g to be deployed—along with a choice of 13 other different databases—while routine tasks like provisioning, and managing regular administrative tasks like clustering, replication, backup, and restore are handled in a simple, unified way. Oracle’s embrace of cloud is providing enterprises with lots of good choices for deploying their database.”

This is making the jobs of Oracle partners even more complicated, however. Oracle Cloud customers need to be able to manage what can quickly become a mixed bag of processes and technologies and to engineer integration between them so they all work together seamlessly, said Singh.

The cloud in and of itself is a huge opportunity that requires commitment and investment on the part of Oracle partners. “It’s natural for customers to be confused at first about their options when new technologies and solutions become available, and we are seeing that initial confusion among our user base,” said Dues. “But, Oracle’s increase in cloud offerings continues to provide choices to customers. Customers can now take advantage of a subset of cloud offerings—those that make sense for their business.”

In the process, the nature of Oracle’s offerings have shifted radically and opened up more opportunities for business users to build and work with solutions. Oracle “used to be deeply embedded in the realm of highly governed IT and engineering resources,” said Selfridge. “Access

a huge capital expenditure on infrastructure, followed by exorbitant support and maintenance costs due to dynamic business needs that demanded re-implementation of many modules as business changed forms through strategic transformations,” said Krishnamoorthy. “Today, the world of cloud ERP has opened up to various organizations, irrespective of their size and spending capabilities.” Traditionally, business leaders have been used to budgeting transformation and implementation programs separately, she said. However, today, because of the embedded transformation in Oracle Cloud applications, a single investment justifies two objectives of strategically re-engineering business processes and implementing Oracle Cloud systems in parallel through a consolidated cloud adoption program, she noted.

Some offerings, such as Oracle’s Financial Accounting Hub Reporting Cloud Service (FAHRCS), are a boon to Oracle EBS sites, Jones said. “FAHRCS lets users dabble in the Oracle Cloud environment by coexisting with Oracle EBS, helping to solve financial reporting challenges,” he explained. However, a downside “is that you have to schedule or run on-demand concurrent requests to push updated general ledger balances to cloud financials, meaning it’s not a real-time solution. Also, as security is not shared with EBS, users must be re-created and separately maintained in FAHRCS, along with a new security profile.” This creates openings ►►

for third-party tools that can provide such insights on a real-time basis, Jones said.

## REAL TIME

The increasing emphasis on real-time business is also a burgeoning area for Oracle ecosystem partners and represents another sea change in this market. Oracle's primary platform for addressing this shift to real time is its in-memory technology, which enables processing of large datasets within the random access memory of servers and PCs. In-memory database technology "is accelerating the speed to process data and quickly deliver answers, in ways that touching disk absolutely could not accomplish," said Peter Yared, CTIO and co-founder of Sapho. "In a lot of use cases, you want to synthesize that unstructured data into structured data that Oracle is very specialized at processing," he noted. "Dumping a ton of unstructured data into a traditional database versus a Hadoop cluster is challenging" and is an area "for Oracle to grow" its capabilities.

The potential impact of in-memory computing is particularly being felt among solutions providers serving companies with high volumes of data and complex operations, said Dues. Global competition, the rise of the internet, and customer impatience are all driving the need for businesses to respond quickly in order to remain viable, she said. "In-memory technologies are one vehicle for helping the business side of the house respond quickly and innovatively to pressing business needs," Dues said.

With the

greater need to parse large datasets, in-memory enables extreme compression and huge runtime performance gains for retrieving data much faster than before, she added.

The emergence of data on-demand,



*The increasing emphasis on **real-time** business is a burgeoning area for Oracle ecosystem partners, and represents another **sea change** in this market.*

which has been a long-sought goal for many technology providers, is now possible through in-memory technology, as well. These advances allow business users to do end-to-end planning, testing, optimization, and implementation, and repeat it in hours or days versus weeks or months, said Selfridge. "The ability to rapidly prototype scenarios has led to huge leaps in the field and has created an entirely new area of data modeling in real or near-real-time that simply couldn't have existed prior to advancements in in-memory technologies."

However, Dues cautioned, "businesses must examine the actual benefits of in-memory, as results may vary based on the use case and database architecture." There are also questions as to how deep an impact in-memory technology is having on the speed of business. "It's a mixed bag," said Krishnamoorthy. "What we have to understand here is that just by switching on this feature, we should not expect it to be a panacea for all the performance issues bogging down the applications of the business."

In addition, Krishnamoorthy added, "a lot of additional

work is required in terms of query tuning and the applications need to be further fine-tuned to take advantage of in-memory." This means additional costs, she explained. "Businesses need to understand the implications and

make a well-informed financial decision if they want to use the Oracle 12c in-memory feature." They would have to ensure that the benefits arising out of an enhanced acceleration of the application performance far outweigh the financial cost associated in implementing this feature, she noted.

Another technology that promises to speed things up dramatically is the use of all-flash storage arrays, which is becoming a key feature enticing to players within the Oracle ecosystem. The increase in input/output operations per second (IOPS)—upward of 80 times more IOPS based on some studies—"is so mind boggling that every enterprise has to consider shifting their Oracle workloads to flash," said Peter Eicher, director of product marketing at Catalogic Software. "The cost per raw terabyte is much higher, but all-flash reaps great savings on non-production use cases. Every Oracle Database instance can spawn as many as eight to 10 full copies used for non-production purposes. This enormous overhead can be eliminated with all-flash arrays because near zero-footprint snapshots can be used instead. You can't drive many simultaneous workloads off hard disk snapshots; you don't have the IOPs. Flash changes that."

However, as with many technology solutions, managing these environments is the key challenge, Eicher cautioned, adding that the management overhead may "blow the doors off your budget." Automation is crucial, said Eicher, who urged Oracle shops and providers to look to Oracle-aware copy data management solutions "that can automate the process of Oracle copy management from creation to access to clean up." ■





# database

TRENDS AND APPLICATIONS

**READERS' CHOICE AWARDS 2016**

THE RAPID EXPANSION OF THE IT MARKET shows no signs of abating. The growth of data in all its forms—from traditional data sources and newer sources such as social media and connected devices—is driving swift innovation. To address the need to secure, integrate, and draw meaningful insights from all this data, a steady flow of products and services, as well as new features to long-established offerings, continues to emerge.

How does an organization know which ones are worth exploring? In the IT industry, there are certainly many resources for learning about products and services. But sometimes, hearing about a product or service from an associate, a colleague—or a fellow *DBTA* reader—is also useful.

To support this flow of information, each year *Database Trends and Applications* magazine allows readers to vote for the *DBTA* Readers' Choice Awards. Unlike any other awards programs conducted by *DBTA*, this one is unique because the nominees are submitted and the winners are chosen by the experts—you, the readers.

Representative of the IT industry itself, the categories for the *DBTA* Readers' Choice awards are wide-ranging. Categories span newer technologies such as Hadoop distributions, NoSQL database products, and cloud products and services, in addition to well-established segments of the market such as relational database management, MultiValue, business intelligence, data integration, and security.

This year, there were 28 categories in which products could be nominated and ultimately voted upon. Here, we present the top three vote-getters in each category. Thanks to everyone who submitted nominations and voted!

—Joyce Wells





## Best Database Overall

**Microsoft SQL Server**

[www.microsoft.com](http://www.microsoft.com)

**FINALISTS**

**Oracle MySQL**

[www.oracle.com](http://www.oracle.com)

**Oracle Database 12c**

[www.oracle.com](http://www.oracle.com)

## Best Relational Database

**Microsoft SQL Server**

[www.microsoft.com](http://www.microsoft.com)

**FINALISTS**

**Oracle MySQL**

[www.oracle.com](http://www.oracle.com)

**Oracle Database 12c**

[www.oracle.com](http://www.oracle.com)

## Best Cloud Database

**Amazon RDS**

<http://aws.amazon.com>

**FINALISTS**

**Microsoft SQL Azure**

[www.microsoft.com](http://www.microsoft.com)

**Oracle Database 12c**

[www.oracle.com](http://www.oracle.com)

## Best NoSQL Database

**MongoDB**

[www.mongodb.com](http://www.mongodb.com)

**FINALISTS**

**Cassandra (DataStax)**

[www.datastax.com](http://www.datastax.com)

**HBase**

<http://hbase.apache.org>

## Best MultiValue Database

**Rocket D3**

[www.rocketsoftware.com](http://www.rocketsoftware.com)

**FINALISTS**

**Rocket UniVerse**

[www.rocketsoftware.com](http://www.rocketsoftware.com)

**Revelation Software OpenInsight**

[www.revelation.com](http://www.revelation.com)

## Best Analytical Platform

**Teradata**

[www.teradata.com](http://www.teradata.com)

**FINALISTS**

**Oracle Database 12c**

[www.oracle.com](http://www.oracle.com)

**Microsoft Parallel Data Warehouse**

[www.microsoft.com](http://www.microsoft.com)

## Best In-Memory Database

**Oracle Database 12c**

[www.oracle.com](http://www.oracle.com)

**FINALISTS**

**SAP HANA**

[www.sap.com](http://www.sap.com)

**MemSQL**

[www.memsql.com](http://www.memsql.com)

## Best Hadoop Solution

**Hortonworks**

<http://hortonworks.com>

**FINALISTS**

**Cloudera**

[www.cloudera.com](http://www.cloudera.com)

**MapR**

[www.mapr.com](http://www.mapr.com)

## Best Database Administration Solution

**Navicat Premium**

[www.navicat.com](http://www.navicat.com)

**FINALISTS**

**Dell Toad DBA Suite**

<http://software.dell.com>

**SQL Sentry PowerSuite**

[www.sqlsentry.com](http://www.sqlsentry.com)

## Best Database Development Solution

**Dell Software Toad Development Suite**

<http://software.dell.com>

**FINALISTS**

**Oracle Enterprise Manager 12c**

[www.oracle.com](http://www.oracle.com)

**Idera DB PowerStudio**

[www.idera.com](http://www.idera.com)





### Best Database Performance Solution

**SolarWinds Database Performance Analyzer**

[www.solarwinds.com](http://www.solarwinds.com)

#### FINALISTS

**Oracle Enterprise Manager 12c**

[www.oracle.com](http://www.oracle.com)

**SQL Sentry PowerSuite**

[www.sqlsentry.com](http://www.sqlsentry.com)

### Best Database Backup Solution

**Veritas NetBackup**

[www.veritas.com](http://www.veritas.com)

#### FINALISTS

**EMC Data Domain**

[www.emc.com](http://www.emc.com)

**Oracle Database Backup Logging Recovery Appliance**

[www.oracle.com](http://www.oracle.com)

### Best Database Security Solution

**Oracle Audit Vault and Database Firewall**

[www.oracle.com](http://www.oracle.com)

#### FINALISTS

**SolarWinds Log and Event Manager**

[www.solarwinds.com](http://www.solarwinds.com)

**McAfee Database Activity Monitoring**

[www.mcafee.com](http://www.mcafee.com)

### Best Data Modeling Solution

**erwin**

<http://erwin.com>

#### FINALISTS

**Navicat Data Modeler**

[www.navicat.com](http://www.navicat.com)

**SAP PowerDesigner**

[www.sap.com](http://www.sap.com)

### Best Data Quality Solution

**Trillium Software System**

[www.trilliumsoftware.com](http://www.trilliumsoftware.com)

#### FINALISTS

**Dell Toad Data Point**

<http://software.dell.com>

**BackOffice Associates Data Stewardship Platform**

[www.boaweb.com](http://www.boaweb.com)

### Best Data Governance Solution

**BackOffice Associates Data Stewardship Platform**

[www.boaweb.com](http://www.boaweb.com)

#### FINALISTS

**IBM InfoSphere Optim Data Management**

[www.ibm.com](http://www.ibm.com)

**Informatica Solution for Data Governance**

[www.informatica.com](http://www.informatica.com)

### Best Data Integration Solution

**Informatica PowerCenter**

[www.informatica.com](http://www.informatica.com)

#### FINALISTS

**Liaison Technologies ALLOY**

[www.liaison.com](http://www.liaison.com)

**IBM InfoSphere Information Server**

[www.ibm.com](http://www.ibm.com)

### Best Data Replication Solution

**Oracle GoldenGate**

[www.oracle.com](http://www.oracle.com)

#### FINALISTS

**EMC Replication Manager**

[www.emc.com](http://www.emc.com)

**IBM InfoSphere Replication Server**

[www.ibm.com](http://www.ibm.com)

### Best Change Data Capture Solution

**Attunity CDC**

[www.attunity.com](http://www.attunity.com)

#### FINALISTS

**Dell SharePlex**

<http://software.dell.com>

**BackOffice Associates HiT Software DBMoto**

[www.boaweb.com](http://www.boaweb.com)

### Best Data Virtualization Solution

**Red Hat jBoss Data Virtualization**

[www.redhat.com](http://www.redhat.com)

#### FINALISTS

**Cisco Data Virtualization Suite**

[www.cisco.com](http://www.cisco.com)

**Denodo Virtualization Platform**

[www.denodo.com](http://www.denodo.com)



**Best Cloud Integration Solution**  
**Progress DataDirect Cloud**[www.progress.com](http://www.progress.com)**FINALISTS****Dell Boomi AtomSphere**<http://software.dell.com>**Oracle Cloud Data Integration**[www.oracle.com](http://www.oracle.com)**Best Cloud Solution**  
**Rackspace**[www.rackspace.com](http://www.rackspace.com)**FINALISTS****Salesforce.com**[www.salesforce.com](http://www.salesforce.com)**VMware vCloud Suite**[www.vmware.com](http://www.vmware.com)**Best Streaming Data Solution**  
**MapR Streams**[www.mapr.com](http://www.mapr.com)**FINALISTS****TIBCO StreamBase**[www.tibco.com](http://www.tibco.com)**SQLstream s-Server**<http://sqlstream.com>**Best Data Storage Solution**  
**Amazon Simple Storage**<http://aws.amazon.com>**FINALISTS****NetApp**[www.netapp.com](http://www.netapp.com)**Oracle ZFS Storage**[www.oracle.com](http://www.oracle.com)**Best Business Intelligence Solution**  
**Tableau**[www.tableau.com](http://www.tableau.com)**FINALISTS****IBM Cognos**[www.ibm.com](http://www.ibm.com)**TIBCO Spotfire**[www.tibco.com](http://www.tibco.com)**Best Big Data Solution**  
**Hortonworks**<http://hortonworks.com>**FINALISTS****Teradata Aster**[www.teradata.com](http://www.teradata.com)**Oracle Database 12c**[www.oracle.com](http://www.oracle.com)**Best Data Mining Solution**  
**Teradata Warehouse Miner**[www.teradata.com](http://www.teradata.com)**FINALISTS****Oracle Advance Analytics**[www.oracle.com](http://www.oracle.com)**IBM SPSS Modeler**[www.ibm.com](http://www.ibm.com)**Best Query and Reporting Solution**  
**Tableau**[www.tableau.com](http://www.tableau.com)**FINALISTS****Microsoft SQL Server Reporting Services**[www.microsoft.com](http://www.microsoft.com)**TIBCO JasperSoft**[www.tibco.com](http://www.tibco.com)





## MapR Technologies



Anoop Dawar,  
Vice President,  
Products

THE MAPR TEAM would like to thank the DBTA readers for the "Best Streaming Data Solution" award and their continued support. There is an exciting paradigm shift happening in terms of how we process event data in real time to better react to business opportunities or risks. At MapR, we understand that to stay ahead of the competition, businesses need the ability to react to business-critical events as they happen. Real-time and reliable delivery of event data is one of the most critical challenges customers are facing, due to the ever-increasing volume and diversity of data as well as geographically dispersed data sources.

MapR Streams fills a critical gap when businesses need to depend on a truly reliable, performant, and scalable event delivery mechanism. A key component of the MapR Converged Data Platform, MapR Streams is a publish-subscribe event streaming system that supports reliable performance at IoT scale. Unlike traditional messaging queues, MapR Streams is built to support IoT scale messaging where millions of messages are expected to be delivered every second. With its real-time global replication mechanism, MapR Streams allows seamless failover in case a data center goes down, ensuring business continuity. Moreover, MapR Streams can be leveraged to converge streaming and batch applications into hybrid applications. This allows customers to analyze real-time streaming data in conjunction with historical data in Hadoop.

**MapR**

[www.mapr.com](http://www.mapr.com)

## SolarWinds



Gerardo Dada,  
VP, Product  
Marketing

**93 PERCENT!** That's how many business end users said that application performance and availability affects their ability to do their job in a recent Solarwinds survey. And, at the heart of nearly every application is a database.

In fact, another survey by Gleanster Research, found that 88 percent of respondents cited the database as the most common challenge or issue with application performance. That's why we built SolarWinds Database Performance Analyzer (DPA).

SolarWinds DPA is a complete database performance monitoring, analysis and optimization tool for Microsoft SQL Server, Oracle Standard and Enterprise Edition, MySQL, IBM DB2 and SAP ASE operations. And DPA works whether applications are hosted on-premises, in a virtualized environment or in the cloud.

DPA helps you pinpoint the root cause of performance issues quickly. It features multi-dimensional performance analysis, storage performance analysis, correlation and dynamic metric baselining with alerting and reporting.

But what really makes SolarWinds stand out above the crowd is that we don't stop at the database. Only SolarWinds through its integrated suite of products, provides easy-to-deploy, easy-to-use, complete, unified visibility into the role that each element of the stack has on application performance.

In short, we help IT drive peak application performance and reduce downtime. Isn't that what every DBA wants?

Combined with products such as SolarWinds Server & Application Monitor, SolarWinds Storage Resource Monitor and SolarWinds Virtualization Manager, only SolarWinds provides easy-to-deploy, easy-to-use, complete, unified visibility into the role that each element of the modern infrastructure stack has on application performance, helping IT drive peak application performance and reduce downtime.

**SolarWinds**

[www.solarwinds.com](http://www.solarwinds.com)



## Navicat



**NAVICAT IS ALL ABOUT INNOVATION.** We understand what time means to our customers—the need to leverage time in order to stay competitive in today's business world. We introduce Navicat iOS and Navicat Cloud, allowing you have the power to manage your databases any-

where you go, and nothing can ever stand in your way again.

Our products include:

- Navicat Premium
- Navicat for MySQL
- Navicat for MariaDB
- Navicat for PostgreSQL
- Navicat for SQL Server
- Navicat for Oracle
- Navicat for SQLite
- Navicat Data Modeler

All Navicat products accentuate an intuitive user interface and a wide array of powerful features that simplify database management, allowing you to produce high-quality applications and databases quickly and efficiently with ease.

With Navicat Cloud Collaboration, customers now have the ability to align your development team for better collaboration and across platforms, or organize your work through a UI-enhanced multi-tasking workflow.

With a more than 10-year history, Navicat has been downloaded over 3 million times and has a customer base of over 150,000 users. Our customers include numerous Fortune 500 companies, such as Apple Inc., Google Inc., Oracle, Ebay, JP Morgan, and many more. We will continue to offer world-class customer support and exciting new features. We are always working on new features and innovations to give your business the competitive edge.

### About Navicat

Navicat develops the leading database management and development software. One of its top-rated products, Navicat Premium, allows you to access up to 6 databases all-in-one, including MySQL, MariaDB, SQL Server, SQLite, Oracle, and PostgreSQL, eliminating workflow disruption to leverage users' time and increasing productivity and efficiency.

**Navicat**

[www.navicat.com](http://www.navicat.com)

## erwin Data Modeling



Danny Sandwell,  
Director of Product  
Marketing

**IT IS BOTH EXCITING AND** validating to be selected as the number 1 data modeling solution by DBTA's discerning readers for the third consecutive year in a row! At erwin, we are proud of the value we deliver to the market and to the 50,000 data professionals who rely on erwin as the foundation of their mission-critical data initiatives.

It's a very exciting time here at erwin, after our spin-out from CA Technologies in April. As a new, independent company, our customers are benefitting from a renewed focus on the erwin product line, deeper investment in innovation and broadening of our technology and commercial partner ecosystem.

While erwin has a rich history, what's most exciting is the role that we are coming to play in global Big Data initiatives. As our customers expand their data management requirements, erwin will provide a broader platform to address the complete data management process.

erwin's mission is to be the central reference hub for enterprise data, regardless of source or structure, making it governable, transparent and trusted across the business. erwin takes an ANY<sup>2</sup> approach to data management. No matter what type of data a business uses—structured or unstructured—and no matter where that data is stored—in-house relational databases or a public or private cloud—erwin ensures that the insights and results from Big Data projects are reliable and achievable.

**erwin**

[www.erwin.com](http://www.erwin.com)





## BackOffice Associates



Rex Ahlstrom,  
Chief Strategy  
Officer

**IT IS A GREAT HONOR** to accept DBTA's Readers' Choice Award for Best Data Governance Solution for the second year in a row. Our mission at BackOffice Associates is to help Global 2000 organizations utilize data stewardship and holistic information governance solutions to set, execute and enforce key data policies across their entire data and systems landscape. Our Data Stewardship

Platform (DSP) automates and orchestrates the process by ensuring that business processes around all corporate data are connected and driving maximized business value to the organization.

The speed and volume of enterprise data only continues to escalate, with both technical and business users interacting with multiple types of data on a daily and hourly basis—whether customer, material, HR, supply chain or sales data. As today's enterprises' needs have exceeded traditional data governance capabilities to simply ensure clean data, BackOffice Associates is leading the charge to empower enterprises with comprehensive information governance functionality designed to achieve their priority business goals and measurably improve overall operations.

The BackOffice Associates DSP is a proven solution to orchestrate enterprises' complex information governance initiatives across data migration, archival, quality, analytics, master data management and business process governance. Bringing technical and business users together, the DSP helps organizations design, set, execute and enforce data policies across all data and systems in the enterprise. The result is Business-Ready Data—achieved through reduced complexity and costs, increased speed and productivity, and mitigated risks of project delays, errors and non-compliance for better business outcomes.

**BackOffice Associates**  
[www.boaweb.com](http://www.boaweb.com)

## Amazon Web Services



Raju Gulabani,  
VP of Database  
Services

**OUR GOAL AT** Amazon Web Services (AWS) is to enable our customers to do things that were previously not possible, and make things that our customers can already do simpler and better at a much lower cost. Adhering to this goal has made Amazon Relational Database Service (Amazon RDS) what it is today. With six database engines to choose from and over 100,000 active customers, Amazon RDS is the new normal for running

relational databases in the cloud. Amazon RDS has consistently lowered the cost of managing relational databases and provided scalability and manageability that were difficult to achieve before and we're excited to continue innovating on behalf of our customers.

We've consistently heard from our customers that they wished they could get the performance offered by commercial databases at the price of open source engines. This propelled us to build Amazon Aurora, a MySQL-compatible database that innovates on the engine and storage layers to deliver five times the performance of MySQL at one-tenth the price of commercial database solutions, it's currently the fastest growing service in the history of AWS.

Another example of our approach is the new, fully managed AWS Database Migration Service which enables customers to migrate production Oracle, SQL Server, MySQL, and PostgreSQL databases to AWS with virtually no downtime. And, the AWS Schema Conversion Tool makes it easy for them to switch database engines.

These innovations help our customers build their applications without constraints so they can focus on running their businesses without having to constantly worry about database reliability or performance.

**Amazon Relational Database Service**  
<http://aws.amazon.com/rds>



## Progress DataDirect



Sumit Sarkar,  
Chief Data  
Evangelist

### AWARD-WINNING DATA CONNECTIVITY

The potential of your business intelligence or data processing application is limited without a comprehensive data connectivity solution. Staying competitive and relevant requires a breadth of data connectivity options.

At Progress, we are laser-focused on high-performance data connectivity for your application, cloud or platform. In fact, Progress® DataDirect® is the undisputed worldwide leader in data connectivity, thanks to our innovative approach and amazing partners—96 of the Fortune 100 and 350+ ISVs are distributing our technology. Our trusted products shield our partners from the high costs of poor data connectivity and set the stage for success.

The DataDirect suite offers unique solutions for enterprises needing to manage and integrate data across relational, big data and cloud data sources. Our portfolio delivers proven, high-performance connectivity, even in the most demanding data migration, integration and management scenarios.

Our latest innovation in hybrid connectivity will completely change the way cloud applications access data. Dubbed “Project Mustang” and available now for previews, this lightweight connector is designed to be embedded in your public or private cloud application, providing secure access to data residing on-premises behind a firewall or in the cloud.

This new DataDirect solution is compatible with industry-standard APIs—ODBC, JDBC, REST and OData—so you can easily manage access to data across cloud, RDBMS, Big Data and NoSQL. The result? Secure access to data behind the firewall from your cloud application—with almost no implementation effort for your customers.

Now, with Progress DataDirect, it's possible to get sharp, comprehensive business intelligence and unlimited data connectivity for your applications.

#### Progress

[www.progress.com/data-connectivity](http://www.progress.com/data-connectivity)  
[www.progress.com/projectmustang](http://www.progress.com/projectmustang)

## Attunity



Shimon Alon,  
CEO

### NEED DATA FAST? GET STREAMING DATA INGEST AND PROCESSING WITH KAFKA™ & ATTUNITY

To manage growing data volumes and pressing SLAs, many companies are leveraging Apache™ Kafka and award-winning Attunity Replicate with next-generation change data capture (CDC) for streaming data ingest and processing.

Attunity Replicate, high-performance data replication and loading software, delivers data to Kafka to enable real-time analytics, and can be used to:

- Feed live database changes to Kafka message brokers that in turn stream to Hadoop, HBase, Cassandra, Couchbase and MongoDB
- Automatically load data to Kafka in bulk or real-time via change data capture (CDC)—with no manual coding
- Ingest high data volumes at low latency from the industry's widest range of sources and targets using Apache Kafka APIs
- Enable Kafka to broadcast live data streams at high scale from multiple sources to multiple targets

Recently crowned for ‘Best CDC’ in the 2016 DBTA Readers’ Choice Awards, Attunity Replicate is designed for high-performance, user-friendly data replication and ingest, and solves your toughest business data challenges. Using the solution, you'll enjoy:

- **Remarkable Ease of Use & Set-up**—Drag & drop user interface with no scripting required
- **Higher performance**—In-memory streaming technology optimizes data movement and eliminates bottlenecks
- **Affordability**—Quicker time to value and lower TCO
- **Integration with Kafka**—Broaden your ecosystem by feeding many data sources into multiple Big Data targets through Kafka message brokers

Take the first step. Watch this on-demand webinar to learn more today: [Streaming data ingest and processing with Kafka](#).

#### Attunity

[www.attunity.com](http://www.attunity.com)



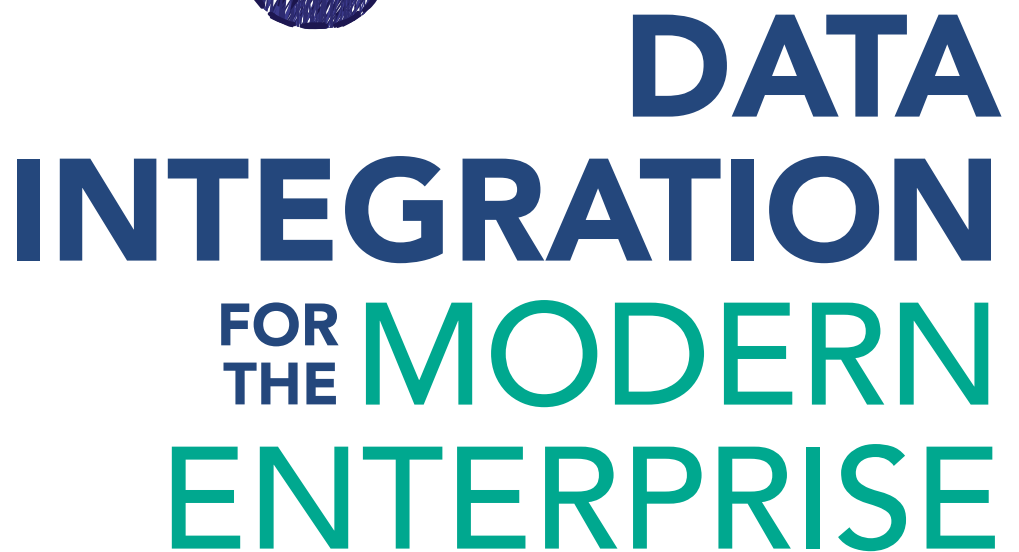
# ROLE OF DATA INTEGRATION IN UNLOCKING THE VALUE OF BIG DATA SOLUTIONS

## REAL-TIME INTEGRATION— THE FUTURE IS NOW

# EMBRACE DIGITAL DISRUPTION WITH DENODO PLATFORM IN THE CLOUD

## 10 NEW REQUIREMENTS FOR MODERN DATA INTEGRATION

# DATA INTEGRATION: A STEPPING STONE TO MODERN DATA APPS



## Best Practices Series

The blood, sweat, and tears that go into enterprise data integrations may not go away anytime soon, but lately, it has become easier to make data from any and all sources more available to people and applications that need it. The cloud—and more specifically, database as a service (DBaaS)—has shifted the challenge of data integration.

At the same time, the rise of cloud and DBaaS as an information management environment has shifted the balance of responsibility for enterprise data integration away from the confines of data centers to the enterprise as a whole. Suddenly, the entire business has a stake

A recent survey of 300 DBAs and IT professionals, conducted by Unisphere Research, a division of Information Today, Inc., finds growing interest in DBaaS as a viable approach to serving their enterprises' needs for greater agility and faster time to market with cloud computing. Many of the early hurdles in delivering enterprise capabilities for security and availability in the cloud become more evident with the reliance on hybrid cloud approaches and the need to move enterprise applications to the cloud and back on-premises based on the business requirements of the organization, their legacy investments, and regulatory requirements ("Database as a Service Enters the Enterprise Mainstream: 2016 IOUG Survey on Database Cloud," April 2016).

expected to triple over the next 24 months. There will be a significant amount of enterprise data shifting to the cloud over the next 24 months as well, as enterprises rethink data management in the cloud. Seventy-three percent of managers and professionals expect to be using DBaaS within their enterprises by that time, versus 27% at the present time.

SHOW THE BUSINESS  
POTENTIAL NEW WAYS DATA  
CAN BE LEVERAGED.

Moving to DBaaS is more than simply making data more accessible, it also opens new paths to innovation. Data is a tool, a means, to better engage customers and better understand markets. Plus, as new ideas and requirements



arise, DBaaS—especially if delivered by a cloud provider—serves as a testbed to help accelerate innovation and experimentation, since cloud providers will likely have all required features and services in place. Shifting more activities to cloud providers or shared service environments frees up enterprises and their IT staffs to provide higher-level support to the business.

### DEVELOP A DATA GOVERNANCE STRATEGY.

Data governance has long been a challenge for enterprises, and cloud or DBaaS doesn't make things any easier. Essential concerns such as data security, quality, and relevance will need to be dealt with at the enterprise level. In addition, there is a need to wrap governance around disparate data sources, which often were external to enterprises and therefore not under their purview. There are a range of business requirements that must be addressed, from real-time data streaming to analytics to customer relationship management. And, organizations must be able to move large and varied datasets at a high velocity through their systems—which raises issues for the handlers of this data, such as who owns it, who has access to it, and how and where it should be stored.

### GO HYBRID.

From a planning/spending perspective, the future belongs to more hybrid approaches. Many organizations continue to maintain an abundance of legacy or on-premises assets, and this is likely to be the case for some time to come. As long-standing legacy assets, these systems have proved their worth and resiliency, and continue to function well for their organizations. Mainframe systems, in particular, continue to be refreshed by IBM and are capable of supporting the largest cloud and DBaaS workloads and the latest protocols. As a result of this sizable legacy base, the largest percentage of organizations in the Unisphere/IOUG survey, 44%, see the establishment of hybrid cloud as their most important

priority as they enter the cloud space.

### TACKLE THE DATA SECURITY ISSUE HEAD-ON, AND AS EARLY AS POSSIBLE.

Data security isn't just about securing data from hackers, but also entails access control, as well as avoidance of any potential for third parties to mishandle the data. Half of the managers and professionals in the Unisphere-IOUG survey indicate that security and privacy concerns are the greatest inhibitors to their cloud initiatives. Data ownership and retention follow closely behind as the second-ranked concern. Often, trusting outside cloud providers with sensitive or mission-critical corporate data is seen as risky, not only in terms of potential breaches, but also in terms of the potential need for a relationship between a cloud provider and consumer to be modified or terminated. The fate of data held by a cloud provider may not be clear-cut.

### STANDARDIZE.

The beauty of cloud and DBaaS is that multiple standards, devices, and interfaces are supported. However, it's still important that all parts of the enterprise be on the same page. On a high level, standards are emerging to help simplify cloud-based integration. For example, the Open Data Protocol (OData) promises to replace the web services standards REST and SOAP to enable greater interoperability between enterprises and across the cloud. In many ways, enterprises are becoming API-driven, meaning applications, functions, and services can be interconnected, on-the-fly, to share data as required by the business demands of the day.

### ADOPT A SUPPORTIVE IT INFRASTRUCTURE.

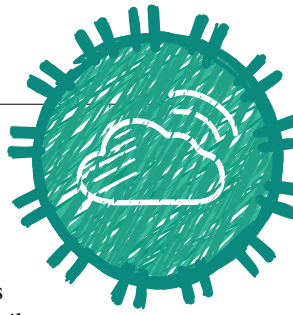
For the private cloud, a robust enterprise infrastructure is essential. The internal systems—particularly storage—need to be highly adaptable and elastic for unpredictable workloads. Public cloud services also offer compelling storage options. Services such as Amazon S3, OpenStack Swift, Microsoft

Azure Storage, and Google Cloud offer virtually unlimited storage resources that can serve as total storage sites or as bursting services for spikes in enterprise workloads. On-premises solutions—from open source frameworks to commodity hardware—also provide scalable storage solutions.

### DEPLOY DBAAS AND CLOUD WHERE IT FINANCIALLY MAKES SENSE.

Not all business cases may be suitable for DBaaS or cloud. To determine the cost/benefit of cloud sites, enterprises need to weigh the traditional on-premises costs of maintaining hardware, systems, networks, and storage, versus that of a shared, multitenant enterprise private cloud, versus subscribing to a public cloud provider. There are also management and labor costs that pertain to project management, development, oversight, monitoring, and quality management—costs that are likely to be incurred regardless of whether the data is managed on-premises or by an outside cloud provider. Even with the public cloud, there are time and expense requirements related to migration or integration between various systems, or between cloud and on-premises systems.

Other expenses that need to be measured include the cost of software licenses versus subscriptions, or the costs of upgrades and maintenance versus pay-by-the-sip monthly fees. In some cases, it may make sense just to leave things as they are; in other cases, the savings may be significant. But, ultimately, the true value comes from the enhanced opportunities for innovation and growth. ■



—Joe McKendrick



# Role of Data Integration in Unlocking the Value of Big Data Solutions

THREE YEARS BACK, you would have had to follow secret directions to the basement to find out where your organization's big data lab was situated. If lucky, you might even have spotted a white coated data scientist. Fast forward to today, and you would be hard pressed to find a meeting room where senior executives are not huddled around a white screen with big data written in bold red.

This is not surprising. The big data ecosystem is forever changing and it plays such an important role that it figures in many strategic projects with some of the best brains in the enterprise heading them. Compounding the problem is that the big data technology landscape is no longer clearly partitioned into data management and data visualization. It has expanded to include data acquisition, data wrangling, data movement, data processing, deep data analytics, data visualization and discovery, and finally data governance solutions.

Big data integration and big data governance touch all these subdivisions of data integration.

## EMERGENCE OF DATA WRANGLING IN THE BIG DATA ERA

The criteria to invest in big data solutions have seen a marked shift from technology-based decisions to those based on business use cases and benefits. Where earlier, purchasing decisions were made based significantly on addressing technical limitations of current solutions, today organizations have started viewing big data technologies as a way to solve business problems or even transform the way they do business. In other words, big data systems have become another data source or target that serves key business initiatives.

Data preparation or data wrangling has emerged as a leading use case which was almost non-existent before this shift. Data preparation and data set creation has

traditionally been a bottleneck for quickly unlocking value from business data. Data preparation solutions use a big data-based processing engine like Spark for a better user experience. Data wrangling services have brought together traditional ETL and data discovery features by providing easy data set creations, data enrichments, publishing and operationalizing (i.e., automating the entire data preparation process) capabilities. Data curation is no longer the sole domain of IT experts. This decoupling of business dependency on IT resources and expertise has enabled faster data analysis, enabling customers to unleash their business savvy onto unsuspecting data sets. Data wrangling technologies have also merged the line between data quality through data enrichment; ETL, through easy data acquisition; and data discovery, through intelligent and rich recommendations.

Leaders in data preparation solutions use advanced machine learning algorithms and natural language processing to glean and bubble up insights from large and varied data sets that cannot be manually discerned because of the scale and complexity of the data sets.

This does not mean that the classic data analyst roles, those of the business analytics and ETL experts, are becoming obsolete. In fact, their roles have evolved to enabling enterprise data management where they are called upon for more strategic initiatives like moving the infrastructure to the cloud and ensuring data for Software as a Service (SaaS) applications.

What really differentiates the multitude of big data projects within organizations is not just the functional sophistication of the projects, but the speed of delivery of insights. While democratizing data and enabling business users to interact directly with data are irreplaceable to help make data-based decision making ubiquitous within the organization, it is

the actual technological and architectural advancements in big data technologies that enables faster decisioning.

## HOW FAST IS FAST ENOUGH FOR ANALYTICS—STREAMING ANALYTICS

The short answer is pretty fast. Real-time data is critical to making any business decision more pertinent. Streaming data right into cloud data warehouses, which are in turn plugged with data visualization solutions downstream, ensures that the data being used for business analytics is the latest data. Successful organizations invest in not just data visualization tools, which are typically the tip of the data management iceberg. They understand and implement a responsive and governed network of data delivery pipeline to capture, filter, aggregate and correlate data before it reaches the analytics platform.

Streaming data first captures data right at the source of the data creation. However, most of the events that generate the data happen on business-critical applications and systems. Real-time data integration technologies should balance the need to capture data in real-time without slowing the performance of the systems.

Combining big data with streaming data capabilities presents infinite possibilities. Data from transactional systems that is captured in the organization's database combined with user generated data in real-time finds many applications in the real world. Businesses with data-driven marketing initiatives can improve customer experience by generating customized promotional offers based on various historic factors such as purchase history combined with real-time data such as location, and data from their social media footprint. Today, real-time data integration technologies are used mainly to deliver data to the analytics data warehouse or data marts. There is an opportunity to embed



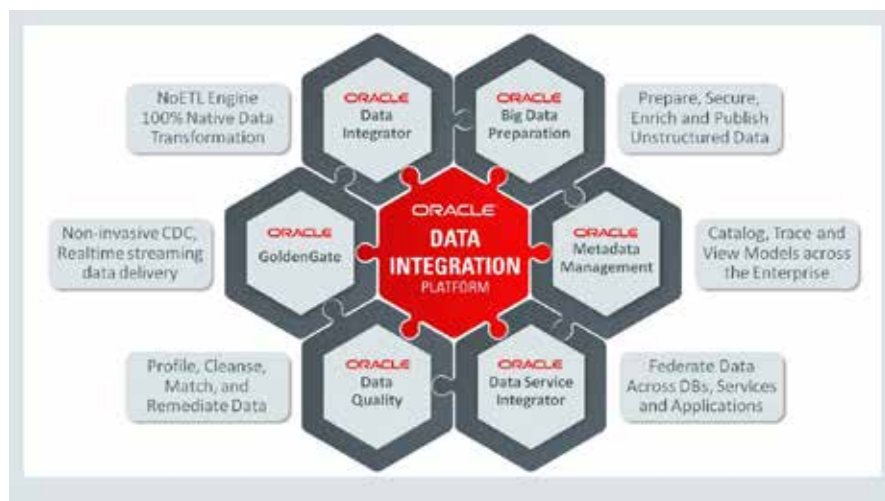


analytics in the data streams that deliver the data to accelerate insights and action.

Streaming analytics that include event stream processing shortens the time from data creation to decision drastically. Stream analytics empowers a business audience in any industry that is looking to create solutions that embrace real-time, instant insight across data delivery infrastructures. A good

deep data storage provides the best of both worlds. Lambda architectures, as such combined streaming and deep data storage architectures are called, provide a single platform that enables enterprises to perform both real-time streaming analytics, and refine the analytics with insights from mining for richer and more complex data recommendations from the deep data storage reservoir.

delivery solution between the cloud and on-premise systems and applications. Oracle's Big Data Preparation Cloud Service is a next-generation data wrangling service that helps business users unlock data quickly from complex business data. Oracle Big Data Preparation Cloud Service is built on Apache Spark, combines natural language processing and machine learning and bridges the line of business—IT divide when extracting insights from the data and operationalizing them into enterprise data integration flows. Meanwhile, the Oracle Stream Analytics platform provides a compelling combination of an easy-to-use visual façade to rapidly create and dynamically change real-time event stream processing applications, together with a comprehensive run-time platform to manage and execute these solutions. This tool is business user-friendly and solves the business dilemmas by completely hiding and abstracting the underlying technology platform. Oracle Data Integrator brings together big data platform portability, the ability to switch between multiple big data platforms seamlessly, and powerful data transformation capabilities to enterprise big data development teams. Oracle Data Integrator provides a unified and common interface to build data transformations irrespective of the underlying big data technology. This ensures the data integration developers can utilize the latest big data platform and language without having to compromise on productivity and with minimal disruption. Oracle's entire integration platform is governed and audited by Oracle Metadata Management. Oracle big data integration offerings are flexible, robust and complementary to maximize big data investments and unlock value from these investments now and in the future. ■



Oracle offers a full set of products for cloud, big data and on-premise data integration requirements

stream analytics solution is designed to handle large data volumes with subsecond latency, while also providing a business-friendly, easy-to-use interface. They are designed with drag-and-drop interfaces that help model data streams replicate business models and behaviors. Streaming analytics finds great use in scenarios that rely on very low latency business decisioning. Some examples include fraud detection in the financial industry to automate stock trading based on market movement, monitoring the vital signs of a patient and setting preventative triggers in health care, and detecting security issues and fraud in transportation industries by finding anomalous patterns as they happen to initiate immediate investigation.

Stream analytics combined with

Because big data technologies underpin the data storage, the cost-to-benefit ratio is extremely appealing for organizations looking to make a difference with their big data investments.

### THE ORACLE ADVANTAGE

Oracle provides a wide range of products that help with all the moving parts of building a differentiated and forward-looking big data integration, management and analytics platform. As part of the data integration portfolio, Oracle GoldenGate and Oracle GoldenGate Cloud Service ensure real-time data capture and streaming from heterogeneous business-critical transactional systems with minimal impact to the performance of the source systems. Oracle GoldenGate provides the most secure and reliable big data

ORACLE

[oracle.com/goto/dataintegration](http://oracle.com/goto/dataintegration)



# Real-Time Integration— The Future is Now

WITH ALL THE cloud-based, mobile and niche applications, plus the Internet of Things (IoT), it's clear that the days of one-size-fits-all, monolithic applications are over. It's now much easier to find a third-party solution to fill a specific need or to "go deep" into an area that cannot be fulfilled by your primary application.

That's led to an increase in application integration. However, the growing need for instantaneous up-to-date data is motivating companies to transition from traditional batch-oriented techniques to real-time data integration. Although achieving real-time integration can be challenging, and accomplished using a variety of technologies, the goal is the same: to transfer accurate, timely data from point A to point B in real-time so users can make better-informed business-critical decisions.

## THE NEW REAL-TIME WORLD ORDER

One technology has emerged as the dominant Web service design model for real-time integration: REST (Representational State Transfer). Why? RESTful Web services are easier to use and the resource-oriented model is more flexible than previous SOAP, RPC and WSDL-based interfaces. REST has already been adopted by providers such as Google, Netflix and Twitter and many other enterprise organizations. RESTful architectures and implementations provide these characteristics/benefits:

- **Easy Web integration** – uses standard HTTP methods
- **Increased Scalability** – stateless interaction and caching semantics
- **Reliability** – separation between client, server and data
- **Security** – via the transport layer (SSL) and message-level mechanisms
- **Standard Language** – XML and JavaScript Object Notation (JSON)

## SIMPLIFIED REST DEVELOPMENT

Kourier Integrator and Kourier's REST Gateway are Kore's easy-to-use and versatile REST integration solutions

for MultiValue (MV) systems. Kourier Integrator streamlines and simplifies the process of building and testing bi-directional integrations using RESTful Web Services; the REST Gateway provides secure, real-time access to MV applications via REST APIs from outside the firewall.

Developers are more productive working within Kourier's REST framework because they can focus on the application interface instead of low-level protocol details such as data validation, resource security and transaction logging.

REST APIs are primarily created via specification pages and typically require minimal programming. Powerful Event Handlers make it easy to leverage existing application business logic or add special instructions within REST resources at specific timing intervals.

Other developer-friendly features of Kourier's REST framework:

- Automatic data validation
- Standard HTTP status code support
- Dynamic query parameters
- Create REST APIs without coding:
  - JSON and XML
  - Query parameter validation
  - Query wildcards
  - Pagination of large result sets
  - Field limiting
  - Result filtering and sorting
- Automatic transaction logging, history and metrics
- API versioning

## SECURE, RATED AND MEASURED ACCESS

The Kourier REST Gateway is a critical piece of the integration architecture because it's responsible for providing secure access to applications from outside the firewall while it monitors, manages and measures REST API usage. Connection pooling is supported for enhanced performance. Policies can limit (rate) the maximum number of requests per minute/hour/day for each user. Gateway administrators can feel confident about exposing their system to the outside world

without worrying about a user consuming all of their resources. Kourier's REST Gateway makes it easy to:

- Create policies to manage resource access
  - Define the maximum number of requests per hour, minute and day
  - Define the availability for each resource (CRUD)
- Define users and server/database security
  - Associate policies and databases
  - Routes requests to server / database
- Visualize performance with the interactive dashboard
  - Graph transaction history
  - View REST resource statistics
  - Drill down into request headers, parameters and timers

## THE REST OF THE STORY

The quest for tighter integration between enterprise applications, third-party solutions and the IoT will drive companies to use more real-time integration via REST to extend and modernize their business operations. Developers will be able to retain the core functionality and value of their MV enterprise applications while extending it as needed via integration to other solutions.

Kore is helping its partners and clients meet this challenge by implementing Kourier Integrator, our award-winning enterprise integration and data management solution, and the Kourier REST Gateway. These products facilitate the building, managing and deployment of secure, scalable, real-time integrations to best-in-class applications via RESTful Web Services.

To learn more about our integration solutions or to schedule a demonstration, please visit our website or call 866-763-KORE (5673). ■

KORE TECHNOLOGIES  
www.koretech.com



# Embrace Digital Disruption with Denodo Platform in the Cloud

IN ORDER TO REMAIN competitive in today's digital economy, data-driven businesses are seeking agile, rapid data integration capabilities that support real-time decision making. Data Virtualization provides the necessary speed and agility by accessing a wide variety of data from multiple internal and external data sources, and transparently combining (without moving) data to provide business users with a unified view of all relevant data for analysis.

Data Virtualization is becoming mainstream for digital enterprises. This is underscored in the May 2016 *Cool Vendors in Pervasive Integration, 2016* report in which Gartner states that the "Internet of Things (IoT), digital business and logical data warehouse use cases require data virtualization approaches for integration to achieve fast time to value for supporting analytics and operations."

**Data Virtualization in the Cloud** provides additional advantages, delivering unparalleled business agility, modernization and cost effectiveness.

**The Denodo Platform for Data Virtualization in the Cloud** offers a unique combination of capabilities:

- ✓ **Self-service data discovery and search** empowers users with knowledge about the data, including data lineage.
- ✓ **Real-time intelligence** provides users with the ability to tap into real-time data for analysis.
- ✓ **Multi-structured data** support for wide variety of data sources such as Hadoop, Spark, NoSQL, Relational and SaaS applications.
- ✓ **Hybrid implementation model** enables a single virtual view of data, combining on-premises and cloud data sources.
- ✓ **Enterprise-ready data governance** allows data-access management from a single point for all data sources.

## WHO IS REAPING THE BENEFITS?

Two case studies serve to illustrate the benefits of the Denodo Platform for Data Virtualization in the Cloud.

### Logitech

Logitech is a global provider of personal computer and tablet accessories. The company was seeking a cost-effective solution for moving its on-premises data to cloud, and integrating this data with cloud data sources. Logitech also faced hurdles associated with time-to-deliver, redundant data and siloed data, as well as security issues related to unauthorized access to underlying data sources which raised governance concerns.

Logitech replaced its on-premises data warehouse with Amazon Redshift, and then implemented an LDW architecture utilizing Denodo Platform for Data Virtualization to unify data across on-premises and cloud data sources and provide a single virtual data access layer. Denodo Platform was used as a business layer provisioning data to all other enterprise analytical tools such as Tableau, Pentaho BA, and other data interfaces and Web services.

Denodo's LDW architecture is the cornerstone for enabling cloud-based reporting and analytics, and played a critical role in the success of Logitech's Cloud BI and analytics strategy, helping Logitech achieve faster access to data via data virtualization without traditional ETL; a business access layer for unified view of on-premises and cloud data; governance enforcement through Denodo's single virtual data access layer; the flexibility to add new data assets; and data virtualization in the cloud supporting Logitech's cloud strategy to bring innovation and agility.

### Mobile Device Protection Firm ("The Firm")

The Firm provides device protection and support services for smartphones, tablets, consumer electronics, appliances, and satellite receivers. Driving the challenges faced by The Firm were an initiative to move analytics from on-premises to the cloud on the Amazon Web Services (AWS) platform, and an initiative by the CSO organization to implement enterprise-ready authentication to manage data access with appropriate security policy.

In order to enact these initiatives, The Firm needed a solution that would enable a business layer as well as a security access layer. The Firm was building a data lake in the cloud for which a virtual data access layer on top of the data lake was needed to provision the data for analytics. The security access layer was needed to comply with enterprise-wide and legal data access requirements.

The Firm selected the Denodo Platform to satisfy both the business layer and security access layer requirements. The cloud analytic solution uses AWS S3 as the data lake environment and the Denodo Platform as the virtual data access and governance layers on top of the data lake to provision data to the analytical tools such as Oracle Business Intelligence.

The Denodo Platform enabled The Firm to ensure that the "right people" have access to the "right data"; that governance requirements are met across cloud and on-premises environments; and was instrumental in the continued success of The Firm's cloud analytic strategy. ■

**DENODO** is the leader in Data Virtualization. Please contact us at [info@denodo.com](mailto:info@denodo.com), or download Denodo Express using the link: <http://www.denodo.com/en/denodo-platform/denodo-express> to get started





# 10 New Requirements for Modern Data Integration

DON'T LET YOUR legacy data and application integration technology be your legacy. Here are 10 new integration requirements that will accelerate your enterprise cloud, big data, and IoT adoption.

## 1) APPLICATION INTEGRATION IS DONE PRIMARILY THROUGH REST AND SOAP SERVICES

To be effective, modern data integration platforms must provide easy and robust ways to consume REST and SOAP. They need to provide an easy way to abstract the complexities of these APIs into business actions and objects to enable an application administrator to rapidly integrate these services with the rest of the enterprise.

## 2) LARGE-VOLUME DATA INTEGRATION IS AVAILABLE TO A HADOOP-BASED DATA LAKE OR TO CLOUD-BASED DATA WAREHOUSES

Enterprise IT organizations are moving away from bespoke data warehouses to data lakes that are repositories of all data based on a Hadoop cluster. Spark is used as the compute framework for data transformation of large amounts of data in this environment. Cloud data warehouse technologies such as Amazon Redshift and Microsoft Azure SQL Data Warehouse are alternatives to expensive specialized data warehouse appliances. Data integration tooling has to have a native understanding of newer storage and compute frameworks based on large-scale distributed frameworks such as HDFS and Spark.

## 3) INTEGRATION HAS TO SUPPORT THE CONTINUUM OF DATA VELOCITIES

Last-generation data integration engines were either optimized for batch processing of large volume data or for low latency handling of small messages. Modern integration platforms should be able to provide the necessary velocity regardless of size of data. This means that the engine has to be able to stream large data such as sensor data from the Internet of Things

just as easily as it can consume and deliver responses to discrete business events.

## 4) INTEGRATION IS EVENT-BASED RATHER THAN CLOUD DRIVEN

Responding to a business event as it happens is expected. For example, increasing the stock inventory on an item based on sentiments expressed in social media or entering a support case automatically when a failure is detected at a device. In either case, polling after the fact for these conditions means a frustrated or lost customer and an inefficient process in today's real-time enterprise.

## 5) INTEGRATION IS PRIMARILY DOCUMENT-CENTRIC

This is a corollary to the fact that integration is based on SOAP/REST APIs that send and receive hierarchical documents rather than row sets or compressed message payloads of the previous generation client server-based technologies. Transforming hierarchical documents into row sets or into compressed payloads at the edges to make the internal engines run efficiently is the biggest impediment to streamlined repurposing of the previous generation of data integration tooling.

## 6) INTEGRATION IS HYBRID

In today's hybrid, multi-cloud environment, modern data integration technology has to be able to respect data's gravity and handle both on-premises and cloud-based applications and data sources with the same efficiency and ease.

## 7) INTEGRATION HAS TO BE ACCESSIBLE THROUGH SOAP/REST APIS

Integration technology has to interoperate with other services in the enterprise such as monitoring, provisioning, and security. For example, enterprises might want to monitor the success or failure of integration flows through their own monitoring tools, and they might want to add new users automatically as they get added to the enterprise integration group.

## 8) INTEGRATION IS STILL ALL ABOUT CONNECTIVITY

By definition, integration is about connecting disparate systems each with its own API set, and an integration platform needs an effective framework to adapt these APIs to efficiently process the data. In addition, a large set of pre-built connectors speeds up the implementation and increases agility in responding to new integration scenarios.

## 9) INTEGRATION HAS TO BE ELASTIC

Reserving capacity to handle the worst case computation/storage needs is costly, and not having sufficient capacity when necessary is even more so. This means that the integration framework has to be able to scale up and scale down resources on demand.

## 10) INTEGRATION HAS TO BE SELF-SERVICE AND DELIVERED AS A SERVICE

In a world that is increasingly cloud-based and data-driven, integration technology has to be delivered as a service that's widely accessible. A new class of users has made self-service essential and only a cloud-based approach with simplified design, management, and monitoring interfaces can meet the broad spectrum of requirements.

These new requirements have given rise to the requirement for a converged integration platform as a service (iPaaS), which should be built from the ground up to address the new and legacy enterprise data and application integration needs. ■

**SNAPLOGIC**  
www.snaplogic.com



# Data Integration: A Stepping Stone to Modern Data Apps

MODERN DATA INTEGRATION in the enterprise is about deriving meaningful insights from a variety of data sources and data structures that can be quickly turned into actionable business information. The data can be structured, semi-structured and/or unstructured data; it can contain business, technical and operational metadata; and its origins can be from within or from outside the enterprise, including everything from ERP systems, legacy and modern databases, to social media posts (tweets, blogs, etc.), system and application logs, machine sensor data, etc. Data integration tools available on the market today are generally viewed as adequate for helping with ingesting, preparing, transforming and provisioning data for where they eventually are consumed in the business or where they need to be fed into business processes. They have come a long way from supporting manual coding to providing more intuitive, visual interfaces, and from increasingly handling real-time data delivery in addition to bulk (batch) data movement. They typically now also facilitate the scale required with integrating data sources from cloud-based applications and data storage. But in a world where big data and applications are converging rapidly, is the value traditional data integration tools are providing to the business going far enough, or are they merely a stepping stone—albeit an important one—to modern data-driven applications?

Generally speaking, a good, modern data integration solution must deal with a wide variety of data types and sources, offer customizable data preparation and cleansing, and support different modes of data delivery (batch, micro-batch, streaming). It should offer easy and timely access to data, help with data governance, and support both existing and emerging use cases, such as IoT. In the world of big data, one of the most commonly expressed challenges is how to get data from sources into an application or a data lake, where it can generate value. But the challenges don't end there: How do you track where the data goes, and who has access to it

once it is in the data lake? And how do you accelerate the time to value for the business by not just solving the data integration and governance problems, but also promoting reuse of components to prevent repeated integration efforts? How do you get on a path to rich, data-driven applications—such as recommendation engines and anomaly detection systems—which require application and cloud integration and often turn out taking a long time due to siloed efforts within the enterprise? Merging the data flow with the application flow in one cohesive integration approach can help reduce the amount of custom coding and manual processes, while speeding up the development and deployment of modern data applications.

This is where the Cask Data Application Platform (CDAP) comes in. CDAP provides a unified integration platform that allows developers, data scientists and IT/operations teams to use a consistent set of tools for



Image 1: Cask Data App Platform

data integration, application integration, operations management and governance. As an integrated framework, it has been designed from the ground up for building, deploying and operating self-service data applications and data lakes on Hadoop and Spark. It is 100% open source and highly extensible, and it supports all major Hadoop distributions offering complete portability within and between the distros.

Through its open source extensions, Cask Hydrator and Cask Tracker, CDAP offers capabilities and user interfaces specifically designed to help with data integration challenges, such as how to quickly “hydrate” a data lake, and how to easily “track” data movement within data lakes and data applications. Cask Hydrator is a self-service and extensible framework designed to develop, run, automate and operate data pipelines. Its intuitive drag-and-drop interface integrates with Hadoop and non-Hadoop storage, and it has the ability to switch between different processing technologies—MapReduce, Spark, and Spark Streaming. Cask Hydrator can prepare, blend, aggregate and apply science to create a complete picture of an enterprise’s business data in order to drive actionable insights.

Cask Tracker is another open source, self-service framework that automatically captures rich metadata and provides users with visibility into how data is flowing into, out of, and within a data lake. It enables IT to oversee changes, while delivering trusted, secure data and an audit-trail for compliance in a complex data lake environment. Cask Tracker provides access to structured information that describes, explains, locates, and makes it easier to retrieve, use and manage datasets, including rich lineage and provenance information.

CDAP, with its extensions Cask Hydrator and Cask Tracker, is the de facto, open source big data application and integration platform for building, deploying and operating data-centric applications and data lakes; it also enables IT organizations to implement well-governed data-as-a-service environments designed to quickly unlock the value of data. For more information about these products, please go to the Cask website, and to stay updated on product and company news, follow us on Twitter @caskdata. ■

# MV SOLUTIONS ▶▶

## OpenQM Adds Python Support

THE LATEST release of OpenQM introduces integration with Python. OpenQM is a high-performance, self-tuning database supporting MultiValue applications on Linux, Windows Server, and IBM AIX.

The latest release, OpenQM 3.4-6, boosts technical support from Zumasys to help organizations keep their applications up and running 24x7. "Our relationship with Zumasys as a worldwide sales and marketing partner since January 2016 is very important to us," said Martin Phillips, master developer of OpenQM. "It has given QM a greater presence in the MultiValue marketplace and brings a highly skilled team of support and development staff to the product."

Along with tech support, the update extends the QMClient API to allow applications written in Python to access most features of QM, including reading/writing data files and executing QM programs and commands. The new release supports both the structured programming and object-oriented programming paradigms,

loads modules dynamically at runtime, and uses variant data type variables.

There is also full support for locking, transactions, select lists, alternate key indices, instantiation, and execution of QM object-oriented programs and use of QM's connection pooling system.

Additionally, the release adds a new feature called command logging. This audit tool will track changes to code to be used for diagnostic purposes.

*OpenQM 3.4-6 provides  
new **development and  
migration tools***

"This release provides some new development and migration tools," Phillips said. "QM is not a clone of any other MultiValue database product and never will be, but it has good compatibility. Users looking to move from other MultiValue platforms should find that this process is becoming easier."

## Entrinsik Partners With Atcore

ENTRINSIK INC. is partnering with Atcore Systems to provide users of Atcore with reporting, real-time analysis, and dashboarding tools. The partnership will focus on enhancing Atcore's SugarCRM with capabilities users need from an operational BI solution.

Entrinsik Informer enables organizations to perform ad hoc reporting and analysis, blending data from multiple sources to create interactive reports and visualizations, while Atcore Systems designs and implements CRM solutions to help businesses attract prospects, convert prospects to customers, and retain the most profitable customers.

Atcore will introduce Informer software into the CRM solutions it deploys to help clients gain insights from real-time reporting and advanced analytics regardless of where the data is stored, said Tad Buck, director of Informer Solutions. ■

## Enterprise eCommerce & Integration Solutions



### Reach the B2B/B2C Market

**\*Powered by** KommerceServer

- **Create** a B2B or B2C Web Experience
- Connect on any Mobile Device
- **Integrate** Back Office Product Data
- Generate Customer-Specific Prices
- **Drive** Web Traffic with Ads & SEO
- Design Engaging Web Content
- **Support** Omni-Channel

**kore**  
Technologies



### Extend the Enterprise

**\*Powered by** Kourier Integrator

- **Integrate** Best-in-Class Applications
- Implement Real-Time Updates
- **Connect** Disparate Databases
- Support Change Data Capture
- **Share** Data Across the Enterprise
- Create "One Version of the Truth"
- **Enhance** Business Decisions

**Solutions that work.  
People who care.**

Let us assist you with your next project.  
Call **866.763.KORE** or visit **www.koretech.com** for more information.



# Gaining Success With Big Data

## 10 Takeaways From Data Summit 2016 Presentations

By now it is well-known that the value of big data comes from its variety, volume, and velocity. But, with the increase in data sources, data types, and data management platforms, new obstacles can also appear, creating challenges to leveraging data for meaningful insights.

During educational presentations on industry trends and technologies, keynotes, discussions, and hands-on workshops at the recent Data Summit 2016 in New York City, the combination of philosophies and technical approaches that can help organizations be successful with data was addressed.

Here are 10 key takeaways from Data Summit 2016 presenters about what to keep in mind when embarking on initiatives to put data to work.

**1** Co-located data is not the same as integrated data: It's a red flag when you go into a company and the No. 1 thing they say is, "Let's get our data in one place, then we can do something." The idea that co-located data automatically becomes integrated is false. Also remember that the best results from data management can't be achieved if there are two different sets of goals in mind by two sides of an organization. While business leaders want to jump into projects, sometimes throwing caution to the wind, IT personnel take into account how to do more with less while controlling costs.

— Anne Buff, business solutions manager and thought leader for SAS Best Practices

**2** Keep it simple: There are lots of new terms and technologies associated with big data, but it doesn't need to be complicated. "Polyglot persistence," for instance, has become a neologism in big data, but it simply means selecting the best technology for the problem. And, remem-

ber that NoSQL systems don't replace relational systems—they augment them.

— Craig S. Mullins, principal of Mullins Consulting

**3** Relieving the DevOps data constraint: Creating, distributing, and managing test data has become a bottleneck in the increasingly fast world of agile and DevOps. Twenty percent of a software development lifecycle (SDLC) is lost waiting for data, and 60% of development and test time is consumed by data tasks. One way to manage test data management is with virtual data.

— Kyle Hailey, technical evangelist at Delphix

**4** Understand how cloud can help: Cloud has changed the equation and introduced the ability to handle large datasets. Cloud allows us to take petabytes of data and analyze it, while 15 years ago, it would not be possible to do this kind of analysis unless you had highly skilled data experts. Cloud as a whole is breaking down barriers.

— Kalev Hannes Leetaru, a Forbes columnist and founder of the GDELT Project

**5** The future lies in cognitive computing: These are systems that are programmed to learn, adapt, and discover with us. Industries are beginning to look at cognitive computing as a necessary business competency to deal with existing challenges. The gathering of massive amounts of data requires that humans have more assistance; and achieving the greatest value in unstructured data requires that machines help with analysis.

— John O'Brien, principal analyst and CEO at Radiant Advisors

**6** Data as a valuable asset: Increasingly, data is being recognized and appreciated as an asset, and even a kind of capital, and needs to be treated as such. Think

of Airbnb's rental business, Uber's surge pricing, and Alibaba's online marketplace.

— Nick Chandra, vice president of Cloud Customer Success at Oracle

**7** The emergence of blockchain: If you don't know anything about blockchain, do yourself a favor and do some research now.

— Charles Pack, technical director, CSX Technology (with credit to Chris, Anant, Kuassi, Madhu; Coleman & the IOUG Team)

**8** Data virtualization tackles the problem of proliferation: Data virtualization can reduce unnecessary copies, the root of data proliferation.

— BJ Fesq, chief architect and chief data officer of CIT Group

**9** The challenge of IoT: As more connected devices come online and generate more data, the hurdle will not simply be getting the data from the devices; the big challenge will be integrating it and using it for meaningful advancements.

— IoT panel discussion with Radiant Advisors' John O'Brien; Joe Caserta, president and CEO of Caserta Concepts; and George Corugedo, CTO of RedPoint Global

**10** The value of in-chip analytics: In-chip analytics is simpler for business users and IT, enabling ad hoc data mashups faster and at greater scale gain more value from data.

— Jeremy Sokolic, vice president, product, at Sisense

Many presentations from Data Summit 2016 have been made available for download at [www.dbta.com/DataSummit/2016/Presentations.aspx](http://www.dbta.com/DataSummit/2016/Presentations.aspx).

**Save the Date for Data Summit 2017—May 16–17 at the New York Hilton Midtown in New York City.** ■

—Joyce Wells and Stephanie Simone

## APPLICATIONS ▶▶

# Five SQL Server Database Performance Tips

WHEN IT COMES TO assigning value to different components in today's IT environment, it would be a serious mistake to underestimate the value of data and databases. Whether you are looking to reduce downtime and speed up your database-centric application or to save money and time on optimizing your database servers, the following best practices can help you improve the performance of your SQL Server environment and, ultimately, your business service delivery process.

## 1 Invest in and maintain your infrastructure.

Organizations with high-performing SQL Server environments give great consideration to the infrastructure underlying and surrounding their databases. This means more than just investing in newer and faster hardware (though that undoubtedly makes a difference). It also means maintaining development and test environments so that potential performance problems can be identified before reaching production. It means architecting solutions for high availability and disaster recovery such that preventative maintenance and updates can be performed regularly without impacting user workloads. And it means reaching across functional boundaries to ensure that the network, storage, virtualization, and application layers which communicate with your databases are properly configured and kept up-to-date.

## 2 Use the features of the engine.

The number of features available to DBAs and developers continues to mount. Do not let your fear of upgrading—to a new version or to a new edition—prevent you from testing and then utilizing the features which will improve the performance of

your environment. Availability groups, columnstore indexes, memory optimized tables, snapshot isolation—the list of features that can drastically speed up certain patterns of work goes on and on.

## 3 Eliminate the low-hanging fruit.

With the wealth of information available online and through user groups, DBAs should be able to easily correct the most common performance-sapping misconfigurations and anti-practices that dog their environment. Overlapping maintenance windows, single-file tempdb topology, outdated server configuration options, excessive log file VLFs, over- and under-indexed tables, and many basic issues of blocking and deadlocking influenced by query hints and isolation levels may all be examined and corrected with minimal interaction with application code.

## 4 Understand your workload.

It is difficult to improve the performance of workloads that are not well-understood. Knowing the business cycles, user patterns, and maintenance windows of the primary applications that are consuming database resources will help any DBA to make the right choices before a problem ever arises.

A skilled administrator will further become familiar with the query plans and data access patterns of the most critical parts of an application and thus be able to anticipate possible consequences at the data layer when there is a change in procedure.

## 5 Monitor and maintain a baseline.

Many issues can be detected and corrected before they rise to a human-perceptible level of concern, and the users' reports are often nonspecific and so subjective that they can be impossible to troubleshoot. SQL Server environments should be monitored and their performance baselined so that deviations can be isolated and investigated. With a baseline report, not only can improvements be achieved, but they can be documented and demonstrated to other team members.

## The Bottom Line

There is a wide range of opinions as to which best practices are the most important to optimizing database server performance. Regardless of the factors that may affect your decision, keeping your database servers operating efficiently is vital to meeting business goals. ■

Vicky Harp is corporate strategist at Idera.



# APPLICATIONS

## The Dangerous Questions IT Never Wants to Hear

By Keith Bromley

BEING ABLE TO DELIVER a rapid response to problems is a core need for the modern enterprise. While many network “issues” can suffer a delay in response time, true problems such as security breaches and network outages cannot. Let’s look at two situations that require a rapid response, five questions that IT never wants to hear, and one example of a solution that IT can implement.

In the case of a suspected security breach, you need to know if you were actually breached or not, where the breach occurred, and what was compromised. According to the 2016 “Verizon Data Breach Investigations Report” (DBIR), almost 68% of breaches happen over the course of several days, so a rapid response to security threats can definitely help minimize the cost of a breach.

If you have just been breached, this is not a situation in which there is nothing you can do. The faster you isolate the attack vector, the faster you can limit the amount of damage to your business. The key is that you need to know right away when you are breached. Unfortunately, the Ponemon Institute’s 2015 Cost of Cyber Crime Study shows that it is actually taking businesses longer to resolve cyberattacks. It now takes an average of 46 days to resolve a cyberat-

tack, which is one day longer than it took last year. This represents an increase of 30% over the last 6-year period and results in a corresponding increase in the cost for a cyberattack.

The situation is also serious for network outages. According to the 2016 “Cost of Data Center Outages” study conducted

In both situations, you need information and you need it fast. The question is, where and how do you capture the critical data that you need and then how fast can you do it? While time is money, time also affects cus-

tomers’ satisfaction and reputations (especially IT department reputations).



*When a **crisis** happens, whether it's a cyberattack, component failure, or application failure, IT needs to quickly **respond** to the problem.*

by the Ponemon Institute, the average cost of a data center outage is \$740,357 and lasts for about 95 minutes. This results in a cost of approximately \$7,793 per minute of downtime. A rapid response is needed in this situation as well to control costs and keep your mean time to repair as short as possible.

When a crisis happens, whether it's a cyberattack, component failure, or application failure, IT needs to quickly respond to the problem. This includes using diagnostic tools to analyze the problem.

If the tools aren't already available in the network, or if they need to be ►►



# APPLICATIONS ►►

moved, the IT engineer is faced with the five following questions that need an immediate response:

- 1. *Will this require approval?*
- 2. *Where should I insert the tools?*
- 3. *Will the insertion of the tools cause other problems?*
- 4. *Will the tool data actually be useful?*
- 5. *Will the data I collect even be accurate?*

Change board approvals require time, and, even in emergency situations, this can take hours. If it is not an absolute emergency (or at least not perceived to be an absolute emergency), the delay can be weeks. Even if you have test equipment on hand, you will face a time delay to get approval to modify the network.

“Where should I insert the tools?” is the next common question. This often becomes an architectural exercise to determine where to get the correct information. In an emergency, you may not have the time to sit down and create a visibility architecture—it needs to already be in place. Otherwise, it is a guessing game of where to collect the data.

The next question is, “Will the insertion of the tools cause other network problems?” This is especially important when inserting inline tools that affect all network data from that point on. The last thing you want is to break something else.

Another concern is whether the tool data will be useful. Different threat vectors can hide their tracks and even delete data after the attack is over. So, are the tools you plan to insert in the network going to capture the critical data needed? If the attack is ongoing, then the answer may be yes. If not, the best case scenario might be to sift through log data instead of inserting new tools into the network.

The fifth fundamental question is whether the data collected is even going to be accurate. For out-of-band monitoring situations, if you’re not using TAPs (Test Access Points), then SPAN (Switched Port Analyzer) ports will only give you summarized data—



when you need it. As a bonus, once you have designed and implemented a

*A **visibility architecture** provides a coherent plan for how to gather monitoring and security data from your **network**.*

not the full data. This means that critical (bad) data right before an incident happens is probably omitted from the SPAN data, and you may not have the data you need to accurately diagnose the source of a problem/attack. SPANs also have the lowest priority for data-switching functionality. Therefore, if the switch is running at full processor capability (due to a

DDOS attack), packets on the mirroring port will get dropped and you will miss important data.

The most cost-effective solution to these issues is to include a visibility architecture in your network design. A visibility architecture is simply a coherent plan about how to gather monitoring and security data from your network. It typically consists of TAPs, network packet brokers, and monitoring/security tools that are arranged in key positions across your network to collect the type of data you want. This helps maximize monitoring effectiveness by ensuring proper access to the data you need,

visibility architecture, there are typically no further change board approvals or other delays in your acquisition of monitoring data. This gives you access to the data you need 24x7. ■

**Keith Bromley** is a product marketing manager for Ixia, Inc., which provides testing, visibility, and security solutions.





GERARDO DADA

In addition to being a long-time geek, **Gerardo Dada** is vice president of SolarWinds' database and applications business globally ([www.solarwinds.com](http://www.solarwinds.com)).

# The Correlation Between Performance and Costs

**■ WHEN YOU THINK ABOUT THE ROLE OF A DATABASE PROFESSIONAL,** you probably don't include "cost savings" in the list of responsibilities. In fact, you probably sum up the role with something along the lines of "keeping applications running smoothly."

While that is a critical responsibility—especially given how important applications are to modern business and, in turn, how important databases are for applications—by and large, databases are misunderstood and underappreciated. And often, so are database professionals.

Maybe if there were a clearer correlation between the work of database professionals and money, people would pay more attention. Well, it turns out, there is.

## The Financial Impact of DBAs and Databases

Generally speaking, there are three ways database professionals can have a direct financial impact:

**1. Proactive database management prevents downtime:** If the database stops functioning, everything stops working. A slow database decreases productivity for every user of an application. How many times have you heard a bank teller, hotel reception clerk, or any other customer service representative say, "Sorry, our computers are a little slow today?"

**2. Moving workloads to lower cost alternatives:** Database professionals can help transition applications from high-license-cost DBMS systems to lower-cost or even open source systems. Even though license costs are only a percentage of total costs, for certain applications, moving from an "enterprise" database to MySQL or PostgreSQL can result in significant savings.

**3. Improving database performance:** Yes, database professionals can also have an impact on the financial health of a company through improving database performance. There is a direct connection between the two, even if far too many are not yet aware of it.

The first two here should not come as a surprise, but the third is much less well-understood. Let's discuss it further.

## Improving Database Performance to Reduce Costs

For most SaaS businesses, as an example, a significant percentage of their expenses is the cost of infrastructure technology—most often in the cloud. We also know that the most critical component of an application is the database. Of course, it's also the least understood.

When an application needs more performance, it's easy to scale in the cloud by adding more hardware. If a database is struggling to meet end-user expectations, too often, the default reaction is to provision larger, faster hardware. You can add memory, move to a database instance with higher CPU capacity, pay for provisioned input/output operations per second (IOPS), or move to solid-state disk storage (SSD). In the cloud, each one of these results in a larger bill at the end of the month.

Unfortunately, all too often, these decisions are made without specific knowledge of whether or how they will help with database performance because few people really understand what happens inside a database and the specific bottlenecks that may be holding performance back. The result is continuously increasing infrastructure spending. Most companies accept these increasing costs as a fact of running a technology-based business.

In these situations, a database professional can be the performance guru and the person who changes the financial profile of the organization. With the right tools and information, he or she can see what happens inside the database to tune queries, remove bottlenecks, and identify exactly what is slowing down an application. The result is a more efficient application that consumes less infrastructure resources and therefore costs less to operate, and is a more intelligent investment overall.

A company with such an approach only adds CPU when they know CPU is the bottleneck, and they know the exact performance impact they will see with the additional investment. Likewise, they only use SSD drives when they know storage read/write is a significant contributor to performance. And so on and so forth.

Such a company has achieved *performance certainty*, meaning they know how their system performs, why it performs that way, what the drivers of performance are, and that the system has been optimized to run at top speed while consuming the minimum resources needed.

As you can see, in the cloud there is a clear connection between performance and resources, and between resources and costs. Well, you know what? The same connection exists on-premises, but the way companies operate often obscures the correlation. However, that doesn't mean these opportunities to drive cost savings and make resources available for projects that truly need them don't exist. It just takes a proactive database professional to find them and to educate management of their value. ■





GUY HARRISON

Guy Harrison is an executive director of R&D at Dell and author of the *Oracle Performance Survival Guide* (Prentice Hall, 2009). Contact him at [guy\\_harrison@dell.com](mailto:guy_harrison@dell.com).

# MongoDB Continues to Advance

WITH 20 MILLION DOWNLOADS TO DATE, MongoDB is arguably today's fastest-growing database technology. MongoDB's rapid growth has been driven primarily by its attractiveness to developers. By using JavaScript Object Notation (JSON) documents as the native database format, MongoDB reduces the impedance mismatch between program code and database, allowing more agile and rapid application development.

However, the MongoDB database is not as easy to use for non-programmers who find the JavaScript-based query syntax difficult, and running MongoDB as a distributed database at large scale creates performance and administrative challenges that are not for the faint of heart.

At the June MongoDB World event in New York City, the company announced a number of significant innovations that attempt to address these concerns.

Improvements to the MongoDB Compass UI allow users to interrogate the database without constructing the sometimes complex JavaScript code that otherwise would be required. The tool displays a graphical representation of an attribute's data distribution and allows users to create a query filter by clicking on a data point or selecting a range of data. Unfortunately, Compass is part of the enterprise version of MongoDB, so it's not available to the open source MongoDB user.

For those who are comfortable with the JavaScript language, enhancements to the aggregation framework—the MongoDB API that enables users to construct the equivalent of SQL GROUP BY queries—now allow for the development of recursively hierarchical queries that can perform graph-style traversals similar to those often performed in social or computer networking applications.

This aggregation framework now also supports the definition of read-only views. Views can restrict the attributes or documents available to nominated users, providing the equivalent of row-level security that is commonplace in relational databases.

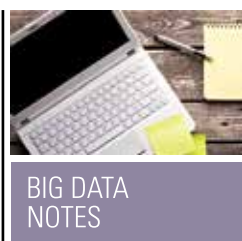
In addition, there are enhancements to MongoDB sharding—the standard means of scaling MongoDB beyond a single writable server. The improvements provide for the establishment of zones, which allows data to be routed to the nodes of the cluster located in local data centers, or allows data to be tiered so that hot/recent data can be placed on systems with an expensive solid-state disk, while colder, older data can be placed on systems with a cheaper disk.

MongoDB has provided cloud-based monitoring, backup, and provisioning services for several years. At MongoDB World, the immediate availability of a database as a service (DBaaS) offering was announced. This service—named Atlas—provides a cloud-based, on-demand MongoDB database cluster with configurable sharding, replication, and server sizing. Atlas runs initially only on Amazon AWS but will also be available eventually on Azure and Google Cloud. A small, three-node replica set cluster (the effective minimum high availability solution) costs around \$300/month.

As with most DBaaS offerings, the service offers rapid deployment and a reduction in administrative overhead. However, there are potential performance issues for applications that are not themselves resident in the same cloud, as well as some concerns about security, though the Atlas security features seem pretty robust. Finally, a cloud-based database may cost more in the long run—at least with respect to simple hardware costs—when compared to on-premise hardware.

MongoDB's appeal to developers has largely been based on simplicity of use and alignment with development platforms, particularly those based on JavaScript. However, as MongoDB becomes more prevalent in the data center, the continuing focus on operational manageability, access to data for analysts and non-programmers, and on security is very welcome.

As with many open core software companies, MongoDB relies



*As MongoDB becomes more prevalent in the data center, the continuing focus on operational manageability, access to data for analysts and non-programmers, and on security is **very welcome**.*

on sales of commercial tools to generate revenue. For the long-term success of open source databases such as MongoDB, these business models must succeed. MongoDB has been an amazing success story from an adoption standpoint, and I hope the company can establish an enduring revenue stream—perhaps based on initiatives such as Atlas and Compass—to support the continuing evolution of the technology. ■





CRAIG S. MULLINS

**Craig S. Mullins** is president of Mullins Consulting, Inc. He's an IBM Gold Consultant and the author of two best-selling books, *DB2 Developer's Guide* and *Database Administration: The Complete Guide to DBA Practices & Procedures*. Website: [www.mullinsconsulting.com](http://www.mullinsconsulting.com)

# Improving IT Security With Database Auditing Techniques

**REGULATORY COMPLIANCE IS A CRITICAL ASPECT** of the IT landscape these days, and the ability to audit database activities showing who did what to which data when is a specific requirement of many industry and governmental regulations.

Different types of database activities may be required to be tracked to ensure compliance. Typical categories of activity that need to be audited include the DDL (or Data Definition Language) for database structure changes, DML (or Data Manipulation Language) for data retrieval and modification, DCL (or Data Control Language) for authorization grants and revokes, security exceptions, and other types of access (such as database utilities that load and unload data).

Another significant aspect of database activity that needs to be monitored is privileged user auditing. This means tracking all of the activities of super users, such as the DBADMIN or SYSADMIN, because these users have high-level access to the system. In addition, many regulations specifically require tracking the actions of privileged users.



## So How Can This Be Done?

There are six primary methods that can be used to accomplish database auditing:

**Audit using DBMS traces.** Database systems typically enable DBAs to start traces to track specific activities. For example, DB2 provides an AUDIT trace that can be started to track multiple categories of events, particular AUTHIDs or programs, and other system details. The benefit here is that the capability is provided by the DBMS at no additional cost. The drawback is that it can generate a significant number of trace records, cause performance slowdowns, and is difficult to report on without a reporting tool that understands the format of the trace output.

**Audit using temporal capabilities.** Modern DBMS offerings have begun to support temporal data management. The system time temporal option can be used to offer a form of data modification auditing. System time support stores every change made to the data in a related history table. Support for managing system changes enables users to be able to query the database as of a point in time and returning the value of the data as of that time frame. The benefit is the simplicity of implementing the solution if the DBMS offers temporal capabilities. The problem is that it is useful only for tracking modifications and cannot tell you explicitly who made each change.

**Audit using database transaction log files.** Every database logs the changes that are made to the data in a transaction log file. Using the information on the log, it is possible to track who modified which data and when. Again, the benefit is that the capability is built into the DBMS. But there are several problems with this approach. You will need a tool that can interpret and report on complex log records, you may need to modify the retention period for your log files, the volume of data can be significant, not every modification may be logged depending on your database settings, and again, this method cannot track read access, only modification.

**Audit over the network.** Sometimes called network sniffing, this approach is used by several tools vendors to trap SQL calls on the network as they are sent over the network. But be careful, because not all SQL requests go across the wire. A DBA can log directly onto a server and make requests that will not be directed over a network. And, on the mainframe, many applications that use CICS, IMS, and batch may never go over a network connection.

**Hand-coded audit trails.** Sometimes the approach is to add "audit columns" to tables, such as LAST\_MODIFIED\_DATE and LAST\_MODIFIED\_USER, that must be modified programmatically whenever data is changed. But this is a problematic "solution" because it is easy to miss a program or a modification request, ad hoc modifications will not be tracked, and data read access is not tracked. Auditors do not like this approach because audit trails should be kept outside of the database (if you delete the row you lose the audit data) and the audit trail is easily corrupted.

**Audit access directly on the server.** The sixth and final approach is to audit directly against the DBMS server control blocks. Sometimes called a "tap," this approach is beneficial because it can capture all SQL requests directly at the server, without starting a potentially costly trace or relying on log files. The potential concern is that this approach interfaces directly with DBMS internals, and bugs can cause outages. Additionally, it requires purchasing ISV software.

## An Emerging Requirement

Database auditing is increasingly becoming a requirement for ensuring data protection and compliance with industry and governmental regulations. Be sure to study the auditing capabilities of your DBMS and to examine any third-party database auditing tools to understand which of the methods discussed here are used to bolster the auditability of your databases. ■



MARIA ANDERSON

**Maria Anderson** is president of the Independent Oracle Users Group and has more than 20 years of experience in various technical and leadership roles.

# Partnerships and Collaboration: A New IT Direction

■ IT IS INTERESTING to see where your mind can take you. I am writing this from Toronto, Canada, where I am attending the Microsoft Worldwide Partner Conference. On my flight to Toronto, I started thinking about partnerships. From a personal perspective, we enter into many different partnerships over our lives—not only partnerships between organizations designed to deliver better solutions to our customers but also within organizations to address the needs of our internal business clients. As organizations, are we engaging in partnership opportunities as often as we should?

As IT decision making moves out of the IT department and into the functional areas of organizations, partnerships and collaboration become even more critical. According to an article in *strategy+business* on why CEOs must become more technology savvy, “the majority of technology spending (68%) is now coming from budgets outside of IT, a significant increase from 47% in 2014.” What this means is that many critical technology decisions are being made without the consultation of IT professionals.

Gilles Freund outlined a slightly different perspective on this shift in a recent LinkedIn post, based on the premise that IT is more than a cost center—though many organizations still perpetuate outdated opinions in this regard. IT has seemingly become a dial tone—we only hear grumbling when something breaks. The rest of the time, IT is scrambling to meet the varied and unique needs of their clients. Because IT professionals truly want to help, we sometimes create situations that are challenging to support and maintain.

Freund indicates that the way to change this is to implement three things: alignment with the organization’s strategy, transparency with key performance indicators, and accountability. These are all very good strategies. However, I believe there is one

important element missing: developing partnerships with your clients. Sit with your clients and talk to them about what they do on a daily basis, learn to speak their language, discuss their pain points, and become a trusted advisor. The solution your client is looking for may not only require collaboration with internal providers of IT functions, it may also require developing a partnership with an external provider. The same rules of engagement apply—before you enter into a partnership, understand where there is common ground and how a better solution can be delivered together.

The news here is not dire—some organizations have started and are continuing a dialogue with IT, according to a recent article in *The Wall Street Journal* titled, “CIOs Get New Stature in Digital Economy.” In this article, Target’s CEO, Brian Cornell, said, “I view technology today very differently than I did 5 or 10 years ago. It’s no longer a backroom function.” Cornell also stated that his CIO is very involved in strategic planning. According to the *WSJ*, 53% of CIOs in Fortune 500 companies now report to the CEO—up from 46% 5 years ago! Ultimately, this article emphasizes the message on collaboration and partnership, which involves strategic discussions. According to Target’s CIO, the life of a CIO is no longer about deployment of hardware and software. What is important to him now, and likely other CIOs as well, “are user experience, speed to market, and putting technology in front of customers.”

Once again, this underpins my opinion that IT professionals must evolve their skill sets to include leadership capabilities that will enable them to reach out and start a dialogue with others, regardless of whether they are clients or at other organizations. Collaboration is the new way of doing business—there is an emphasis on global perspectives and outcomes, which are challenging to achieve without trusted partners. This all came together for me when I walked into the Microsoft conference exhibit hall and saw a number of partner quotes shared on large posters. One of them stated, “Every partnership made is a step closer to the solution.”

How will you start initiating partnership conversations to help your clients? ■



*Collaboration is the new way of doing business—there is an emphasis on global perspectives and outcomes, which are challenging to achieve without trusted partners.*

# SQL Server 2016 Is Simply Faster...A Lot Faster



KEVIN KLINE

**Kevin Kline**, a longtime Microsoft SQL Server MVP, is a founder and former president of PASS and the author of *SQL in a Nutshell*. Kline tweets at @kekline and blogs at <http://kevinkline.com>.

I HAD THE PLEASURE TO SPEND SOME TIME with my old friend Mark Souza, a general manager in the Data Platform team at Microsoft, while speaking at the SQL Saturday event in Dublin, Ireland. Now keep in mind that Mark and I have known each other since the 1990s when SQL Server was just being ported to a brand new operating system called Windows NT. Mark and I were having a laugh and more than a twinge of nostalgia about how much SQL Server has improved over the decades and now sits atop the heap on most analysts' "best database" reports. This isn't just two old-timers sharing a few war stories though. This is a living, breathing transformation that is still in process.

## The Cream of the Crop

One source of information you should absolutely make a part of your regular reading is the blog "SQL Server According to Bob" (<https://blogs.msdn.microsoft.com/bobsq1>). It's written by not one, but two guys named Bob—Bob Dorr and Bob Ward. These individuals are perhaps the most widely respected deep technology experts on the SQL Server team. They're the most senior members of the SQL Server support organization. Rest assured, if you have a support call about SQL Server referred to one of the Bobs, there is no higher authority.

## Best of the Blog

Back to their blog. They take a lot of time to provide lucid and detailed explanations of how the internals within SQL Server work and, most pertinent to this discussion, reveal the mysteries of why SQL Server 2016 is so much faster than previous versions. Just look for the tag "It Just Runs Faster" to see all of the relevant articles.

*There are many subtle and **deeply internal improvements** that make everything else in the database engine faster and better balanced.*

There are literally dozens and dozens of deep-code improvements in SQL Server 2016, so let me run down a few highlights of things that get a lot faster in the newest release. (These are arbitrary and personal preferences. You probably will have other favorites.)

- **DBCC**, SQL Server's internal consistency checking utility, scales up by seven times. And that improvement happens despite having much additional consistency and logical checks. That means preventative maintenance operations are much faster.

- **Tempdb**, a file system where SQL Server does most of its internal processing, has better default handling of the underlying I/O subsystem.

- **Transaction Logs**, the primary method through which SQL Server ensures durability in ACID-compliant transactions, gets an improved "stamping" algorithm to maximize modern hardware, improve multi-threaded processing, and optimize storage reclamation and cleanup.

- **Automatic Soft NUMA**, seldom seen on older hardware, is now the norm for better memory and CPU partitioning. This provides a series of cascading benefits to other internal structures, such as spinlocks, latches, mutexes, and semaphores. Gains of 10% to 30% are not uncommon on certain OLTP workloads.

- **Better Thread Scheduling** enables SQL Server to better schedule worker tasks and balance the workload for higher scalability. This, combined with the Soft NUMA improvements, means that many background SQL Server processes can run within a NUMA-node rather than outside of the NUMA-node.

These are subtle and deeply internal improvements, but they're ones that make everything else in the database engine faster and better balanced. An analogy I like to use is that many improvements are flashy and easy to spot, such as a cherry-red sports car. But to truly make most people's commute a lot faster and smoother, you have to redesign the roads and traffic patterns. These are the sorts of improvements we're seeing delivered at internet-speeds with SQL Server.

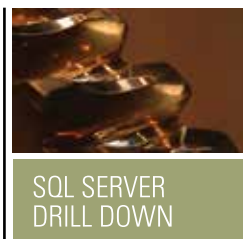
## Free SQL Server 2016!

SQL Server 2014 and 2016 Developer Editions are now truly free. Now, there's just no excuse not to get serious about SQL Server. Read all of the details at <https://blogs.technet.microsoft.com/dataplatforminsider/2016/03/31/microsoft-sql-server-developer-edition-is-now-free>.

Have questions? Drop me a note!

Cheers,

—Kevin ■



<p>SEPT 2016</p>	<p><b>Best Practices Series</b></p> <p>The Evolution of Big Data: NoSQL, Hadoop, Spark &amp; Beyond</p> <p>For sponsorship details contact Stephen Faig, <a href="mailto:stephen@dbta.com">stephen@dbta.com</a>, or 908-795-3702.</p>
------------------	---





TODD SCHRAML

Todd Schraml has more than 20 years of IT management, project development, business analysis, and database design experience across many industries from telecommunications to healthcare. He can be reached at TWSchraml@gmail.com.

## Before We Get to Normal

■ AS ONE WORKS THROUGH the normal forms, be it a journey to the placid shores of third normal, or the more arid climes of Boyce-Codd normal form, or even the cloudy peaks of fourth normal and beyond—and before one starts thinking about normalizing the design—the database designer has covered a lot of ground work already. Before thinking of normalizing, one needs to have conceptualized the relations that might be within the solution's scope. One should be able to name these relations, whether they be CUSTOMER, ORDER, or PRODUCT, for example. As a general rule of thumb, most modelers try to keep the names of these relations singular rather than plural, i.e., CUSTOMER instead of CUSTOMERS. If the relation can be named, it should also be defined, with some explanation of why it exists and the kinds of data to be contained within it.

Once the idea of some relations exists, then one can start thinking about what attributes might go within them. The attributes need names, too. The names should be clear; they should follow within the organization's naming standards. (Hopefully, this means each name ends with a class-word or other domain designation.) And, if one can name the attribute, then one should be able to define the attribute. By definition, this does not mean taking the abbreviated name and simply providing a fully fleshed-out English version of the attribute's name, such as defining "CustCd" as "Customer Code." In point of fact, at this level of the design activities, it would be more appropriate to name the attribute "Customer Code" with an understanding that by using the organization's standard abbreviations, any implemented column would be titled "CustCd" or "Cust\_Cd," for example. A definition is not a trivial rehashing of the name of an attribute; rather, it is a small excursion into why the column exists. There are plenty of web-

sites offering advice on good versus bad definitions, so that discussion does not need to be surveyed here.

Next, each of these relations needs to have an identified primary key. These days, it seems that far too often a generated key is mindlessly plugged in solely to provide functionality as a primary key, but that approach actually is cheating. The designer should look for the natural candidate keys on each relation inside the nascent model. Composite keys can be fun. If it happens that a relation has multiple candidate keys, choosing which one to use as the primary key can be entertaining. On the other hand, if no candidate key arises, then one needs to dig deeper for a better understanding of what is happening. Every relation should have a primary key, something that allows each row to be uniquely identified;

it goes to the heart of what an entity means. If a relation does not have anything that is useful as a primary key, then how can that relation have a meaning? And, if the relation has no meaning, it should likely not be in your data model.

Before working through the process of normalization, we seem to have quite a list. An organization has a start on naming standards for entities and attributes, lists of standard abbreviations, lists of agreed upon class-words and domains. For the solution in focus, a listing is created of possible relations and their meaning, each relation's attributes, and those attributes' meaning. Each relation has an identified primary key.

Upon reaching this point, one may finally start normalizing those structures into a solution database design. Certainly, the intention is not to say that all relations and all attributes need to be defined first, but enough of them should be identified and defined to aid in the start of fleshing things out. Data modeling is a process, new attributes and relations will arise, and each element should be stepped through the process as one completes a design. ■



### AUG/SEP 2016

#### Ad Index

Kore Technologies .....	26
Melissa Data .....	Cover 2
Revelation Software .....	Cover 4

### Best Practices Sponsors

Cask .....	25
Denodo .....	23
Kore Technologies .....	22
Oracle .....	20
SnapLogic .....	24

### Readers' Choice Awards

#### Winners Circle

Amazon Web Services .....	15
Attunity .....	16
BackOffice Associates .....	15
erwin .....	14
MapR .....	13
Navicat .....	14
Progress Software .....	16
SolarWinds .....	13

# SharePoint

## SYMPOSIUM

Registration Open!  
Use code  
**DBTA** to save!

**JW MARRIOTT**  
WASHINGTON DC  
**NOVEMBER**  
**16-17, 2016**

**EVALUATING  
WHERE, WHEN,  
AND HOW  
TO USE IT IN THE  
ENTERPRISE**

Co-located with

**KMWorld**<sup>2016</sup>

*Enterprise Search  
& Discovery 2016*

**taxonomy**  
BOOT CAMP >>>>

[SHAREPOINTSYP0SIUM.COM](http://SHAREPOINTSYP0SIUM.COM)

Diamond Sponsors



Gold Sponsor



Media Sponsors



Organized and Produced by



**REVELATION**  
SOFTWARE

Software architects need database development tools that evolve with their rapidly changing business landscape. We are Revelation Software, creators of the OpenInsight Development Suite, bringing you one of the best browser-based, mobile computing and robust reporting toolkits on the market. Go to **revelation.com** and start inventing your next great software solution today.



SCALABLE

MORE DAZZLING

FASTER

SYNERGISTIC