Volume 33 Number 1
FEBRUARY/MARCH 2019

The Journal of
Information Integration and Management

database
TRENDS AND APPLICATIONS

Automation Takes on the Heavy Lifting of Data Management

4

WWW.DBTA.COM

How DBAs Can Survive and Thrive in a World of Agile Development 27

Four Steps for Ensuring Compliance Post-GDPR 30

The Shift to a Serverless World 31

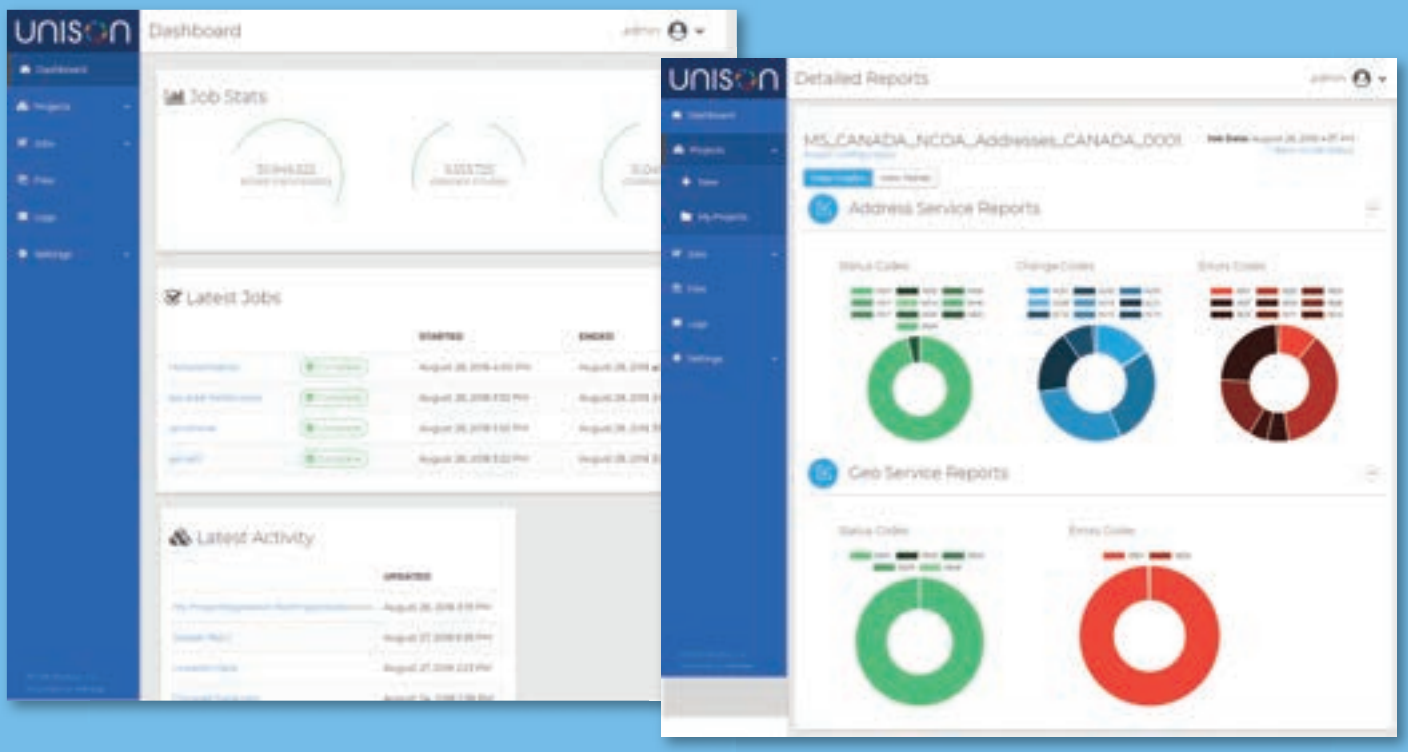
Multiplatform Data Quality Management

Unison – Speedy, Secure, Scalable

Melissa's Unison is a data steward's best friend. Unison is a unique multiplatform solution that establishes and maintains contact data quality at lightning speeds – processing 50 million addresses per hour – while meeting the most stringent security requirements. With Unison, you can design, administer and automate data quality routines that cleanse, validate and enrich even your most sensitive customer information – name, address, phone, email address – as data never leaves your organization.

- **Lightning fast processing (50M records/hour)**
- **Streamline data prep workflows**
- **Reduce analytics busy work**
- **Services: Address, Name, Email, Phone Verification & Geocoding**

UNISON by melissa



Request a Demo

Melissa.com/dbta-unison

1-800-MELISSA

melissa®

The Journal of Information Integration and Management

database
TRENDS AND APPLICATIONS

Volume 33 | Number 1
FEBRUARY/MARCH 2019

CONTENTS

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, Editor-in-Chief
908-795-3704; Joyce@dbta.com

Joseph McKendrick,
Contributing Editor; Joseph@dbta.com

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

Stephanie Simone, Managing Editor
908-795-3520; ssimone@dbta.com

Don Zayacz, Advertising Sales Assistant
908-795-3703; dzayacz@dbta.com

Celeste Peterson-Sloss, Lauree Padgett,
Editorial Services

Tiffany Chamenko,
Production Manager

Lori Rice Flint,
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

Rob Mandeville, www.solarwinds.com
Guy Harrison, guy@tobacapital.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
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, 2019 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 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. If you live outside the USA, request permission from your local Reproduction Rights Organization. (For a list of international agencies, consult www.ifrro.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

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 $79.95 per year. Delivery outside North America is $145 via surface mail per year. All rates to be prepaid in U.S. funds. Subscribe online (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.

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

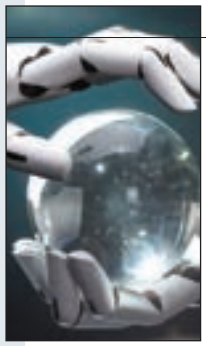
FEATURES

RESEARCH@DBTA

2 **SURVEY: ONE-FOURTH OF DATA ASSETS NOW IN THE CLOUD**
By Joe McKendrick

FEATURE STORY

4 **AUTOMATION TAKES ON THE HEAVY LIFTING OF DATA MANAGEMENT**
By Joe McKendrick



DEPARTMENTS

TRENDS

9 **AI AND MACHINE LEARNING: NINE PREDICTIONS FOR 2019**
By Joyce Wells and Stephanie Simone

27 **HOW DBAS CAN SURVIVE AND THRIVE IN A WORLD OF AGILE DEVELOPMENT**
By Ken Rugg

APPLICATIONS

30 **THE SHIFT TO A SERVERLESS WORLD**
By Nenshad Bardoliwalla and Pakshi Rajan

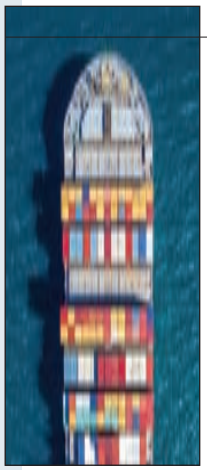
DBTA WEBINARS

33
> **CRAFTING A DATA GOVERNANCE STRATEGY**
> **MONITORING THE PERFORMANCE OF DATABASE SERVERS WITH IDERA**
> **ENTERPRISE MUST-HAVES FOR A MODERN DW**

MULTIVALUE SOLUTIONS > SPECIAL SECTION

22 **MULTIVALUE NEWS**

24 **MV EXPERTS PLAN TO EVOLVE WITH THE MARKET IN 2019**
> Mike Ruane, Revelation Software
> Julianna Cammarano, Rocket Software
> Mark Pick, Pick Cloud
> Doug Leupen, Entrinsik
> Malcolm Carroll, BlueFinity International
> Keith Lambert, Kore Technologies



COLUMNS

34 NEXT-GEN DATA MANAGEMENT > by Rob Mandeville
WHY IS TUNING HARD?

35 EMERGING TECHNOLOGIES > by Guy Harrison
A BIG YEAR AHEAD FOR ETHEREUM

36 SQL SERVER DRILL DOWN > by Kevin Kline
CONTAINERS AND THE WONDERFUL WORLD OF OPERATING-SYSTEM-LEVEL VIRTUALIZATION

38 IOUG OBSERVATIONS > by Seth Miller
CAN HIGH-TECH REALLY BE COMPLETELY AUTOMATED?

39 DBA CORNER > by Craig S. Mullins
WHAT IS A DBA?

40 DATABASE ELABORATIONS > by Todd Schraml
WHEN IS A DIMENSION A FACT?

MEDIA PARTNER OF THE FOLLOWING USER GROUPS

Survey: One-Fourth of Data Assets Now in the Cloud

By Joe McKendrick

Increasingly, cloud services are seen as a vital resource in the data manager's toolkit. There's good reason why cloud is a preferred option: There are simply not enough on-premise resources to keep up with the growth of data management requirements. Organizations keep evolving, business priorities keep shifting, data compliance requirements keep expanding, and user demands keep growing. Already, one-fourth of corporate data is being maintained by cloud providers, and data managers intend to move as much of their data environments into the cloud as soon as they can.

These are some of the findings of a new survey of 202 data managers and professionals, conducted among members of the Independent Oracle Users Group. The survey, fielded by Unisphere Research, a division of Information Today, Inc., in partnership with Amazon Web Services, covered a broad sample of company types and sizes ("2019 IOUG Databases in the Cloud Survey").

While database systems have been part of corporate data centers for decades, and substantial infrastructures have been built around them to ensure their security, resilience, and ability to support mission-critical applications, much of the technology required to keep delivering business value is available in the cloud.

Data in the cloud is still a DBA's responsibility, suggesting that cloud-based systems are not going to replace DBA roles anytime soon. In terms of lines of responsibility, DBAs are most likely to be put in charge of public cloud database projects, as cited by 42% of survey respondents. IT operations teams and outside service providers each took the lead at 18% of enterprises. When it comes to mature cloud data sites, DBAs tend to assume a greater role, the survey shows. Among high cloud adopters in the survey, 53% said their DBAs are in charge, while low cloud adopters tend to divide cloud management between their service providers and DBAs.

In the survey, respondents were asked about their most recent database project—whether it was an upgrade, migration, or addition of new functionality—and whether it was in the cloud or on-premise. They were then asked about their most recent cloud-based database project. In some cases, these were one in the same.

The survey shows that, on average, one in every four bytes of enterprise data is now managed by public cloud providers. This will increase, but hybrid cloud is the model for most (see Figure 1).

Figure 1: Where is your organization's data located and managed? *(Average percentage of respondents' total data stores)*

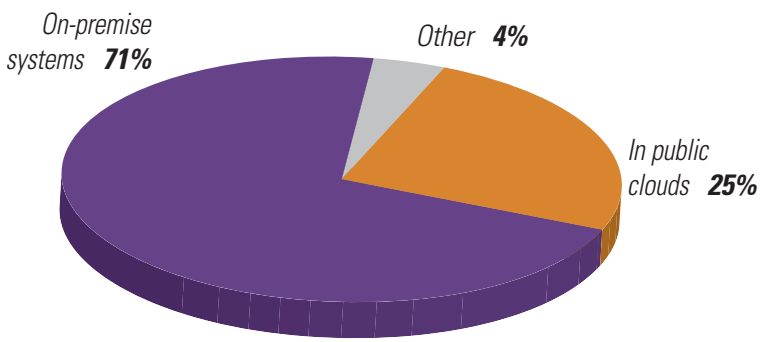
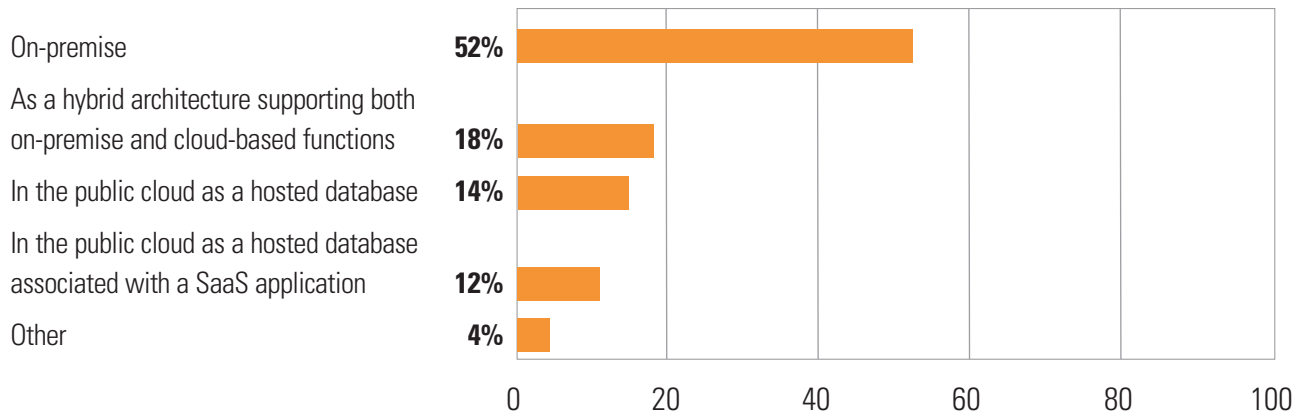


Figure 2: What was the location of your most recent implementation?



There's an overriding reason why data managers are moving many of their assets to the cloud: It can expand as their data stores and access requirements grow. Scalability is the most oft-cited benefit seen in the cloud, the survey shows.

While reduced costs were often the driving force for cloud implementations in the early days, leaders and professionals are now even more likely to see cloud as a way to quickly scale and support their growing application environments and datasets. A majority of data managers that have migrated to cloud, 54%, said they are seeing greater scalability. Cost and lowered administration also factor in as important benefits: 44% of cloud implementers say they have achieved lower costs, while 43% are seeing a reduced need for infrastructure maintenance as a result of contracting with cloud services.

Even for organizations that currently have mainly on-premise databases, hybrid arrangements are viewed as options going forward as they seek to combine the best of what the public cloud has to offer with existing on-premise assets.

For their latest database projects, organizations are close to evenly split between deploying in the cloud or on-premise, with a tilt toward on-premise. A total of 44% deployed their projects in the public cloud or as part of a hybrid architecture split between cloud and on-premise, while 52% indicated their most recent database project involved an on-premise implementation. When it came to cloud, 14% opted to host their entire database in the cloud, while 12% turned to cloud SaaS offerings that include back-end data functionality (see Figure 2).

Cloud-based data functions are no longer at the periphery of enterprises—many are now supporting core enterprise applications. While cloud implementations in the early days were typically for edge-of-enterprise functions—such as organizing sales communications—cloud is now moving closer to the core of enterprises. A large segment of public cloud data projects, 41%, went to directly supporting production applications. Interestingly, for 27% of the group, this was their very first public cloud deployment for a database function. The largest segment, 44%, reported this is among their first five or six cloud-focused projects. ■

Automation Takes on the Heavy Lifting of Data Management

By Joe McKendrick

THESE DAYS, EVERY business is a data business. Every enterprise, regardless of industry or even size, is attempting to manage and draw insights from large volumes of data—now often extending into the petabyte range—surging through and between their systems and applications, both within clouds and on-premise data centers. This all needs to be managed appropriately and securely, as well as backed up and made highly available. ▸▸

Running data-intensive operations requires more and more resources—in budgets, staff, and infrastructure. In addition, the complexity and requirements of managing data—thanks to increasing technology, greater user demands, more far-reaching data regulations, and highly competitive business landscapes—will only grow more onerous.

The key to managing in this data-intensive era is having the right tools and processes to automate as much as possible. Leading vendors such as Oracle recognize this, focusing on autonomous databases that are supported by internal robots to manage security and availability. A plethora of tools and platforms on the market now apply automated processes and capabilities that shape data environments.

A data automation strategy is key to avoiding the costs and inefficiencies of running data operations in a manual or ad hoc way. Data managers and professionals are needed for higher-level tasks, from consulting with business leaders on data-driven strategies to promoting and nurturing innovation within their enterprises.

Many enterprises are turning to cloud and automation, particularly in the area of backup and recovery, according to a recent survey conducted by Unisphere Research, a division of Information Today, Inc., in partnership with VMware. The survey, which covered 260 members of the Independent Oracle Users Group ("2019 IOUG Data Environment Expansion Survey") found that, with the increasing demand for real-time analytical capabilities, there is greater pressure on data administrators to deliver responsive, high-performing systems that can scale accordingly.

As a backdrop to this scenario, there is a looming skills shortage that is adding new urgency to efforts to automate a wider breadth of data center operations. "As baby boomers shift out of the daily workforce, the wealth of knowledge they represent regarding mainframes is at risk of being lost," said Jeff Cherrington, VP of product management, systems at ASG Technologies. "Because Gen Xers joined the workforce as distributed and cloud was taking off, the majority of them focused their careers in those areas. To support this transition, many enterprises expect to use AI and machine learning to capture and transfer knowledge to younger generations, and augment mainframe management by automating time-consuming tasks."

DATA AUTOMATION FUNDAMENTALS

Some of the most fundamental data management processes are now enabled through automation. Automation—whether via cloud services or through internal systems—is seen as critical to ensuring that complex and sophisticated data systems are delivering to the business. Here, data managers and professionals see the greatest value in automating the backup and recovery processes to ensure their data assets are protected. Automating backup and recovery is most important to the ITI/IOUG survey respondents (74%), followed by business continuity/disaster recovery (62%). Processes such as provisioning (53%) and monitoring (51%) score lower overall as automation priorities.

This is only the start, as increasingly sophisticated aspects of data operations will see automation. "As organizations move more toward cloud and DevOps-style operations, database automation is moving away from regular maintenance operations toward the movement and aggregation of data between disparate database types and locations," said Dave Brunswick, vice president of solutions at Cleo. "Automating those data flows will be critical to future business growth."

Additional data management processes that are increasingly being automated include ETL/data movement/data synchronization, security auditing, database maintenance operations—space reclamation, object management—and performance diagnostic collection and analysis, said Robin Schumacher, SVP for DataStax. "Other automated tasks performed less frequently include database provisioning and removal, and software upgrades done on regular cycles," he noted.

Database release automation helps to break the application release bottleneck created by today's manual database deployment process, said Robert Reeves, co-founder and CTO of Datical. This, in turn, improves productivity and performance, allowing development and testing teams and DBAs to focus on more important projects and initiatives. "Database release automation also helps eliminate unavoidable incidents of human error, while increasing data security and application performance and reliability," Reeves added.

Mathias Golombek, CTO of Exasol, sees automation being employed for "event-stream processing databases that have to instantly react in case of certain events." Golombek cited the example of an alarm monitoring system, where sensors send data to a central entity that has

to react as quickly as possible in case of a fire hazard, or a stock exchange news stream analysis system that can trigger automated buys/sells of certain stocks.

The most common forms of data automation are being seen in databases that are provisioned automatically, said Monte Zweben, CEO of Splice Machine. "It used to take days to months to provision and install a petabyte-scale distributed database. Now through container orchestration in the cloud, you can provision and install in minutes." In addition, automation is leading to databases that now self-heal, he added. "Distributed databases systems now can shard or distribute petabyte-scale data across multiple machines and deliver high availability by automatically monitoring all the deployed containers, perform health checks, kill unhealthy containers, and autoscale new ones—all with zero-human operator involvement."

Another area ripe for data automation is in "analyzing large amounts of data and getting automatic insights out of it," Golombek continued. "There are various applications, from AI algorithms using deep learning methods to win games against human beings to picture recognition software, or applications that automatically survey and optimize a company's production chain."

BUSINESS PROBLEM SOLVING

The technical problems that can be solved by automation are self-evident, then. But what business problems are suitable for data automation? It boils down to making everyone across the enterprise more productive. "People should realize that they are better off with a system that doesn't require a complex task than a system that automates a complex task," said Joe Pasqua, EVP of products at MarkLogic. "Today, the most commonly automated database tasks are in some sense the most prosaic, like deployments, upgrades, backups, and change testing. Those are all important, but are primarily operational rather than providing direct business value."

Schumacher pointed to staffing and skills issues as a prime driver for aiding the business at large, given a "lack of operations headcount or lack of skills and experience at knowing what's needed to maintain sophisticated database deployments." In addition, automating database tasks "increases the productivity of the operations staff and allows them to focus on other things," he noted.

For a business that intends to run on data, this becomes key. "Numerous business decisions have to be taken every moment, 24 hours a day, and database automation helps to solve this problem," said Golombek. "By automating decisions with intelligent algorithms and analyzing comprehensive data across the organization, businesses can reduce the number of mistakes, increase the complexity of decisions made, and minimize the reaction time."

Data management helps people make better decisions by using better data, said Jake Freivald, VP of marketing at Information Builders. "This is especially important for AI solutions, where 'black box' algorithms don't allow businesspeople to see the process that gets them to an answer. Small errors in an AI situation can make business users doubt the results they get." Moreover, Freivald said, businesspeople really "own" the data. "They know how critical specific pieces of data are, and what the consequences are for having bad data. They're responsible for fixing data when it's broken. IT can only help so much with those things, and typically only in generic ways, before the business has to get involved. Without automation, data management involves businesspeople paying close attention to many different data points and looking for anomalies. If we can automate the anomaly-identification process, we can help businesspeople spend less time looking for bad data and more time fixing it."

Chadd Kenney, VP of product marketing and solutions engineering at Pure Storage, sees reduced application time-to-market as an important business driver for data automation. Whether DBaaS, on-premise, or in the cloud, automation "gives you the ability to rapidly develop, test, and deploy new applications at a fraction of the time it would take with legacy systems," he said. "Enterprises can take new applications to market much faster, gain an edge over competitors, and service customers better. In addition, "storage administrators will no longer need to spend time and effort on provisioning storage resources for databases. DBAs will no longer have to wait for database resources to be allocated and provisioned. This frees up time for both teams to focus on more strategic, higher-value activities."

STILL OUT OF REACH

Yet, there are areas of the data center and the business that may be out of the reach of automation for now, for a myriad of reasons.

"Much of database automation is about storing data, and maintaining data, but less about exacting significant business insights out of the data stored in the database," said Ryohei Fujimaki, CEO of dotData. "As we all know, data is only as valuable as the business insights it produces. From that perspective, the current database automation technology needs improvement in this area."

The majority of business tasks are still out of reach for automation, Golombek agreed. This is partially because there is not enough data collected, and automation algorithms—such as machine learning—are simply working to solve way too many problems, Golombek noted. "Databases can address only tasks with relatively simple rules and limited complexity. Compare that with robots that can take over many manual tasks within a car factory, but for the tasks that require soft skills, emotional intelligence, and a far more complex way of communication, robots are still not a fit." ▶▶

Even a seemingly straightforward process such as performance tuning may be difficult to turn over to automation at this time, said Brunswick. "Performance-tuning, particularly for high-transaction-rate systems, is difficult to fully automate because of the complexity and variance between organizations."

Data managers and administrators need to ask "when it's smart to automate, and when it's not," said Sumeet Singh, VP of engineering for Juniper Networks. "Projects that require more strategy, innovation, or creativity are best left to the human touch. For companies implementing automation, it's crucial they have specialists who can help troubleshoot or intervene if a problem arises."

There are still "many situations in which we do things manually based on ideas and information that aren't captured, standardized, or maintained in a way that automated processes can use," said Freivald. "Even AI and machine learning processes are only as good as the data we feed them. It's hard enough to capture all of that data about highly structured IT processes, but data management requires businesspeople and IT to collaborate—so those combined business/IT processes, heuristics, and ideas are even harder to nail down in a way that doesn't leave us open to high risks. We're getting better, but these things take time."

BEST APPROACHES

To achieve successful data automation, data managers and administrators need to take a holistic or strategic approach, industry observers advise. "The key aspect for database automation is the company's data strategy," said Golombek. "All initiatives and applications should be aligned with that strategy."

Cloud computing offers a way to make this transition—37% of respondents to the ITI-IOUG survey said moving to cloud has aided their efforts to automate data operations. "The most direct way to address data center automation—such as auto-healing, predicted failures, and performance degradation—is to decouple the physical mapping between database assets and physical servers," said Kevin McNamara, CTO and co-founder of HiveIO.

"This can be accomplished through cloud computing or software-defined virtualization solutions," said McNamara. Traditionally, there has been resistance to moving to the cloud because performance has been tied to hardware physics but, McNamara said, with the advancement of storage acceleration, such as inline deduplication, compression, and cache, many administrators have now shifted to software-defined processes rather than hardware acceleration. "In doing so, cloud features, such as live migration, erasure coding, and API metrics, offer significant value."

Others advocate a workflow-by-workflow approach to automating database tasks. "Automation matters most when you need to do something fast, like fix something that is broken or roll something out that has value," said Singh. "Organizations ought to identify their key troubleshooting workflows, as those will lend themselves quite naturally to automated outcomes. It's also important to focus on targeting one workflow at a time, diligently building automated outcomes into data center processes."

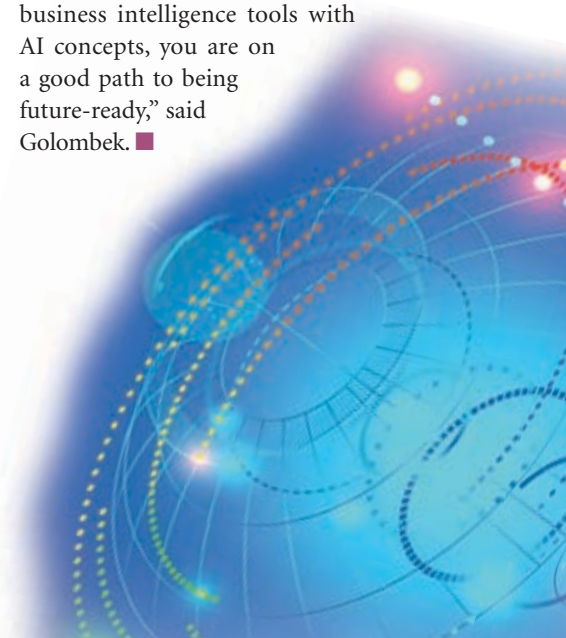
Vendors are building greater automation into their solutions, but this is only part of the story. "Vendors supply some automation today, including the cloud vendors, however those hoping that the cloud will automate the bulk of their larger database operations will be sadly mistaken, and will find both missing functionality and a lack of existing functional depth," said Schumacher. It's important, then, for enterprises to "look at databases in the same way as other DevOps approaches, and also to look beyond the tooling provided by the database vendors," said Brunswick. "Third-party integration providers can lay the foundation for effective automation of processes involving different database vendors and hybrid ground/cloud environments."

ENTER AI

Of course, these days, no discussion about data automation is complete without mentioning the growing role of artificial intelligence (AI). While Schumacher noted current AI is still in its "baby-steps phases" with auto-scaling and best practices enforcement that learns its specific environments, AI will increasingly be applied to application-oriented data and building out recommendation engines.

"Like all AI and machine learning domains, the key is to successfully model the problems you are trying to solve," said Brunswick. "Aspects like future capacity forecasting or log anomaly detection, for example, may be well-suited to an AI approach, whereas others, such as performance tuning, may be much harder to model."

AI and machine learning offer a powerful option for getting new insights or predictions out of data, Golombek noted. "Especially for automation, these algorithms can make a data scientist's job easier." However, he cautioned that organizations have to be aware that it's just another option—another tool. Teams should try to integrate AI and data science seamlessly into their organizational structure, and technology stack and landscape. "If you are able to integrate business intelligence tools with AI concepts, you are on a good path to being future-ready," said Golombek. ■



AI and Machine Learning:

Nine Predictions for 2019

By Joyce Wells and Stephanie Simone

In 2018, there was heightened awareness of AI and machine learning (ML); and in 2019, the emphasis is expected to increase on cognitive computing technologies that can analyze data in ways previously unimaginable and act independently or with limited human intervention. Here, executives of leading IT companies offer nine predictions for what's ahead in 2019.

1 Machine learning projects will move from science projects and innovation labs to full production led by industry disruptors in 2019: Virtually every company has machine learning projects, but most of them are reliant on specialty platforms that cannot access all of the data relevant to business objectives. All of the data is stored in a variety of data warehouses and data lakes, none of which have the ability to run end-to-end machine learning, forcing data movement to the specialty platforms. Only a subset of data is then used to train and score ML models, resulting in limited accuracy. In 2019, current industry disruptors and smart traditional companies will bring machine learning to all its data, instead of moving its data to the ML platforms. These companies will more accurately predict and influence outcomes, including predictive maintenance on medical devices, predictive revenue based on personalized customer behavior analytics, and proactive fraud detection and prevention.

— Joy King, VP, Vertica product marketing, MicroFocus

2 Advanced analytics and AI will be everywhere and in everything, including infrastructure operations: Advanced analytics and AI will continue becoming more highly focused and purpose-built for specific needs, and these capabilities will increasingly be embedded in management tools. This much-anticipated capability will simplify IT operations, improve infrastructure and application robustness, and lower overall costs. Along with this trend, AI and analytics will become embedded in high availability and disaster recovery solutions, as well as cloud service provider offerings to improve service levels. With the ability to quickly, automatically, and accurately understand issues and diagnose problems across complex configurations, the reliability, and thus the availability, of critical services delivered from the cloud will vastly improve.

— Jerry Melnick, president and CEO, SIOS Technology

3 AIOps and the "right type" of monitoring: AI is making huge strides in monitoring things such as signals that are natural to humans (images/video/speech). However, in comparison, the killer applications within IT have not emerged. No company knows yet how to prepare the "right type" of signals and the related feedback to allow for machine learning and produce a strong, meaningful application supporting IT management. Within the ▶▶

TRENDS

next 3 years, companies will figure out the right mix of signals and feedback for machine learning and will create a breakthrough in monitoring strategies. The first to leverage this new strategy after gathering the right data will have the key advantage in the market. Ultimately, these tools will increase team efficiencies by enabling teams, which used to require experts, to operate through generalists providing considerable customer value.

— Gadi Oren, VP of products, LogicMonitor

4 **AI-augmented analytics will be mainstream:** 2019 will be the year when AI-led analytics (known as automated discovery) will become mainstream. Human brains are not wired to evaluate millions of data combinations at sub-second speeds, but machine learning is literally built for this problem and the perfect solution. Business leaders and data analysts are better understanding that AI is not going to replace jobs, but augment them, and I expect that in the next year, the majority of data analysts will have the power of data science at their fingertips without the need to write code.

— Ketan Karkhanis, SVP and GM analytics, Salesforce

5 **AI/data science meets the line of business:** One of AI's biggest obstacles has been the disconnect between data science teams and subject matter experts (SMEs) in the business. SMEs play a critical role, but the complexity of the underlying tech typically requires a lot of data science expertise. Enterprises will put increasing pressure on their teams to close this gap so that they can get more value from their AI initiatives. In addition, as AI becomes embedded in more and more processes, there is an increasing need for transparency in how it works and makes decisions on our behalf. Users will demand real-world, plain English examples and explanations for full transparency and "explainability." This will also make it easier for data scientists and SMEs to collaborate on improving AI's contribution to the business.

— Tom Wilde, CEO and founder, Indico

6 **AI connects employers with their employees:** While companies continue to spend money on AI technologies that improve the customer experience, in 2019, we'll start to see companies delivering similarly compelling experiences for their employees. Conversational AI and development bot platforms are a few examples of technologies that businesses will use to help current and new employees succeed.

— Steve Hamrick, VP of product management, SAP Jam

7 **AI is revolutionizing B2B sales and marketing, but enterprises will struggle in their adoption:** Using AI for revenue optimization (account-based marketing, forecasting, and lead scoring, etc.) is slow going due to weaknesses in internal customer databases. Limited and inaccurate data will be one barrier, and immature AI technology the second barrier. Traction will come when companies take command of their first-party data, clean and augment it with third-party data.

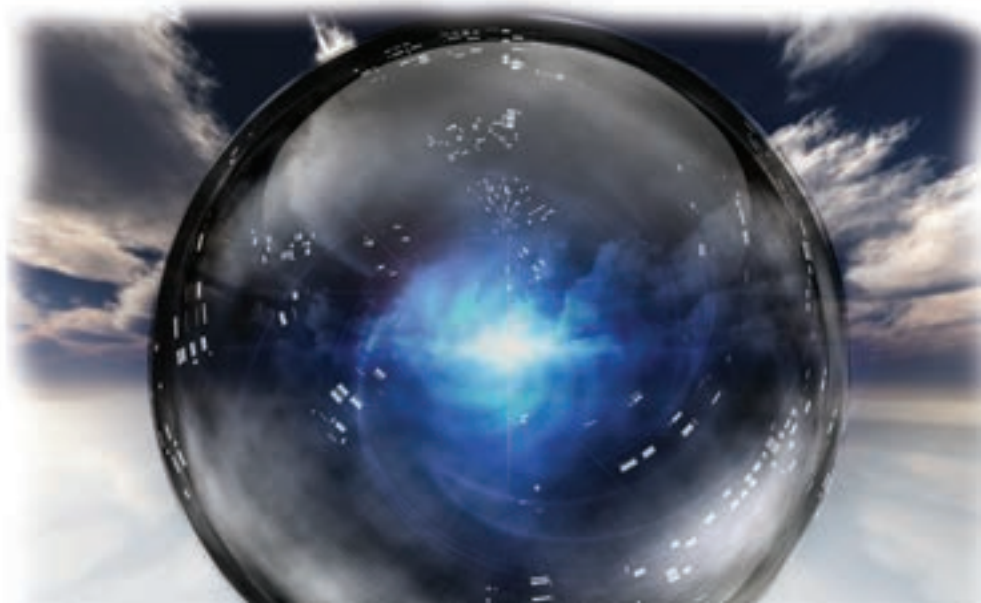
— Umberto Milletti, entrepreneur and sales and marketing leader, InsideView

8 **AI and object storage play tag:** As businesses increase their use of AI to extract greater value from their digital assets, metadata tagging will become an even more critical element of enterprise storage. This will bring more attention to object storage, which is centered on metadata, and the key will be integrating well with AI tools.

— Jon Toor, CMO, Cloudian

9 **As AI and ML become mainstream, a new breed of security data scientists will emerge in 2019:** AI and ML techniques are data-dependent. Preparing, processing, and interpreting data require data scientists to be polymaths. They need to know computer science, data science, and above all, need to have domain expertise to be able to tell bad data from good data and bad results from good results. What we have already begun seeing is the need for security experts who understand data science and computer science to be able to first make sense of the security data available to us today. Once this data is prepared, processed, and interpreted, it can then be used by AI and ML techniques to automate security in real time.

— Setu Kulkarni, VP of corporate strategy, WhiteHat Security ■



AWS
PAGE 14
AMAZON RDS ON VMWARE

Percona
PAGE 16
MANAGING THE HYBRID FUTURE: FROM DATABASES TO CLOUDS

MariaDB
PAGE 17
THE RISE OF DATABASES AND HYBRID EVERYTHING

SentryOne
PAGE 18
STREAMLINE CLOUD DATABASE MIGRATIONS WITH SENTRYONE

Delphix
PAGE 19
STANDARDIZE ON AN EXTENSIBLE PLATFORM TO MANAGE DATA SOURCE HETEROGENEITY ACROSS HYBRID CLOUD

Couchbase
PAGE 20
THE NOSQL DATABASE ARCHITECTED FOR MULTI-CLOUD AGILITY

Quest
PAGE 21
ALL YOUR PLATFORMS IN ONE CONSOLE: MONITORING HYBRID CLOUD ENVIRONMENTS WITH A SINGLE SOLUTION

database
TRENDS AND APPLICATIONS

database
TRENDS AND APPLICATIONS
ONE COMPLETE MARKETING PROGRAM

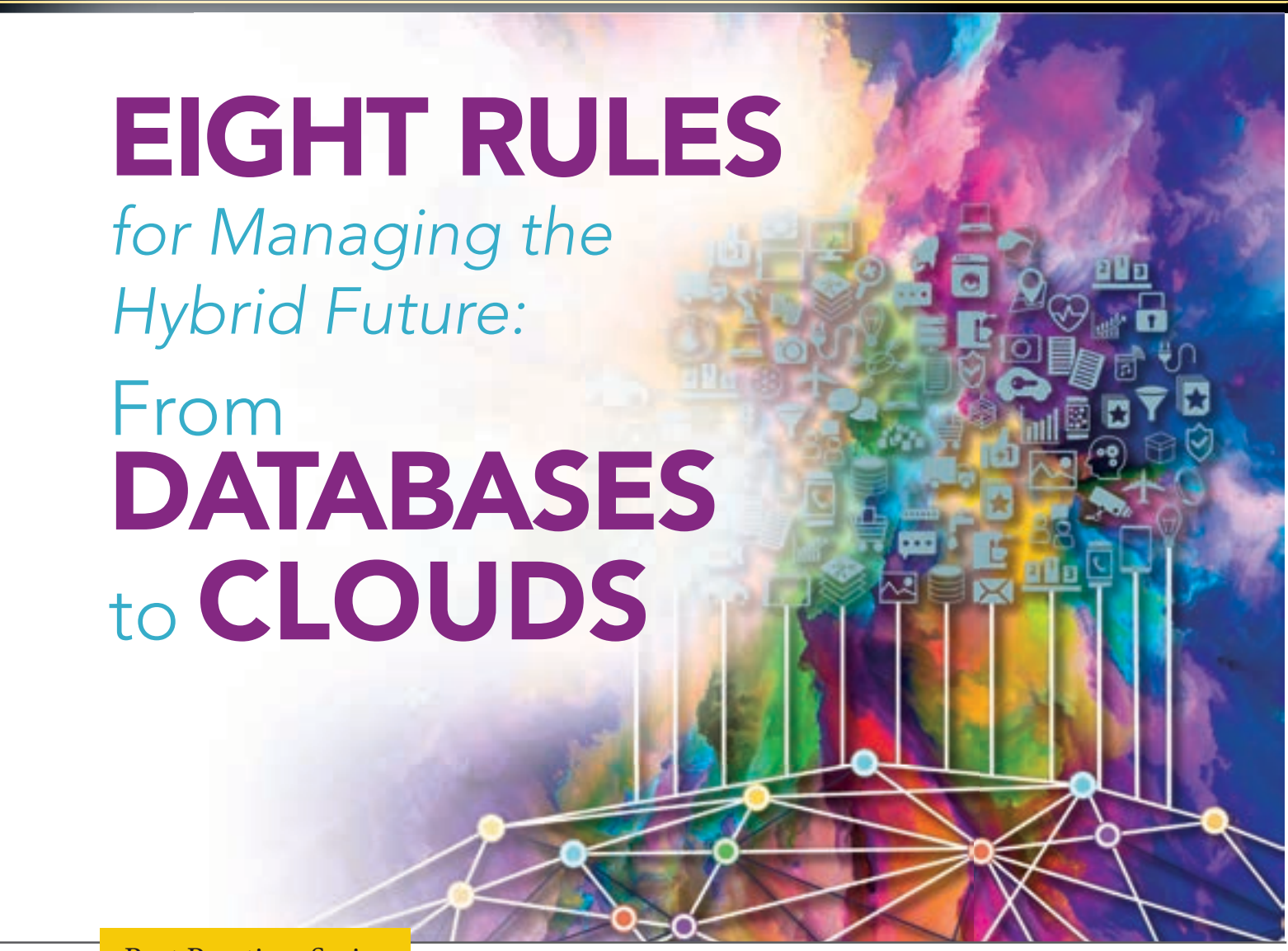
Managing the Hybrid Future: FROM DATABASES TO CLOUDS

Best Practices Series

EIGHT RULES

for Managing the Hybrid Future:

From DATABASES to CLOUDS



Best Practices Series

When it comes to databases, the future is hybrid—and many organizations are assembling wide assortments of on-premise and cloud-based systems. "If you're using any cloud today, it's likely that you're already at hybrid cloud," states a recent Forrester report.

The impending rise of hybrid clouds supporting data environments is the dominant trend seen within the findings of a recent survey 202 data managers and professionals, conducted among members of the Independent Oracle Users Group. The survey, fielded by Unisphere Research, a division of Information Today, Inc., in partnership with Amazon Web Services, calculates that at least one in every four bytes of enterprise data is now managed by public cloud providers. Twenty-four percent of "high cloud adopters" in the survey report having in excess of 35% of their data assets with public clouds. This will increase, as hybrid cloud becomes the model of choice for most enterprises ("2019 IOUG Databases in the Cloud Survey").

While public cloud is gaining traction among most enterprises, respondents made it clear that they have no intention of going all the way with public cloud. Two-thirds said it's likely they will be moving into hybrid cloud arrangements over the next 1 to 2 years. One in five indicated they still have no cloud plans at this time.

Here is what data managers need to know about managing hybrid environments that span both on-premise and cloud settings:

1. Determine how a hybrid strategy fits the requirements of the business.

A hybrid strategy ensures a bridge is put in place for the smooth transition from a mainly on-premise setting to a mainly cloud-based setting. This may occur in a matter of months, or it may take years. Moving to hybrid data arrangements may mean significant shifts in resources, which may soak up budgets and staff time. More often than not, it doesn't make sense to build or expand capabilities when a robust cloud option is readily available on the market. At the same time, the long-term benefits of a major transformation of resources may justify any short-term pain. In addition, the business at large may be seeking greater agility as part

of a transformation to a more digitally enhanced enterprise. Some on-premise systems may be functioning well and continuing to provide value to the business, or they may be an anchor holding back progress. Either way, a hybrid strategy serves to ease the transition as an organization embraces cloud as it moves forward.

2. Consider the implications of data governance.

There are, and will remain, many reasons for keeping portions of data onsite, thus providing ongoing business justification for hybrid arrangements. There may be high security requirements, in which proprietary data must remain within the walls of the organization. There may regulatory or compliance requirements for keeping data out of third-party hands. Geographic requirements may dictate that data remain within national borders. Data managers need to understand and identify the portions of their data environments that absolutely must remain on-premise, and build capabilities accordingly.

3. Enhance business continuity.

Another advantage of a hybrid strategy is that it may help ensure seamless availability of applications so that business users are oblivious to outages or downtime. However, a hybrid cloud strategy needs to ensure that workloads can quickly and readily be moved between on-premise and cloud environments. At the same time, conversely, it may be prudent to maintain copies of data within an on-premise system in the event of cloud outages.

4. Don't outsource security.

Whether data is managed by a cloud provider or is managed on-premise, it's incumbent on the organization to take ownership of data security. While many major cloud providers may be better trained, certified, and prepared for security incidents, the onus should always be on the enterprise customer to perform due diligence when it comes to protecting data. This includes holding vendors' feet to the fire to guarantee that the enterprise's security protocols are met.

5. Ensure the flexibility of data movement.

As the enterprise evolves into a digitally powered, cloud-first business, it's important to be able to move data assets between these environments as well. While it's relatively easy to move datasets between environments, many enterprises have accumulated features or integrations—and cloud providers provide value-added services or templates which ease deployments—that may make it more difficult to migrate data to alternative environments.

6. Embrace the range of tools.

Hybrid cloud means employing a range of tools for different dimensions of the deployment, from traditional tools for on-premise solutions to tools specific to cloud providers. As a result, enterprises may be overwhelmed with tools. Such tools cover backup and recovery, cloud management, API management, performance management, and security management. There may be separate teams working with specific tools as well. Vendor-specific tools and cross-platform tool offerings need to be evaluated. Data managers need to focus on tools that provide the greatest coverage for both on-premise and cloud environments to ensure the greatest leverage. Ultimately, the greater the automation such tools provide, the better it is.

7. Expand the scope of service-level agreements.

In an era of enhanced compliance mandates and scrutiny of data privacy and security, data governance is an extremely critical requirement, regardless of its origins or where it is stored. Data governance does not stop outside the enterprise walls; rather, a governance strategy needs to extend across networks to all data resources. In addition, SLAs will need to be applied and enforced across a range of environments. There may even be pass-through SLAS, in which IT departments contract to provide certain levels of availability and performance, based on their own agreements with backend cloud providers. It's important to remember as well that business users may be oblivious to where agreements are applied—they just want their applications to work as intended.

8. Prepare for new skill sets.

Moving between on-premise systems and data and cloud-based resources requires a recalibration of skills. While on-premise skills—heads-down integration, programming, security—are still a necessity, the emerging cloud side of the equation requires higher-level skills such as architecture and consulting. DBAs and developers will need more frequent refreshes of their skillsets to prepare for this new environment.

Managing data environments that cross over from on-premise to public cloud sites requires different approaches and technologies than either traditional on-premise data environments or fully cloud-based services. Following the eight rules outlined above will help. ■

—Joe McKendrick

Amazon RDS on VMware

Relational databases underpin many of the world's business-critical systems. Provisioning, patching, backing up, cloning, restoring, scaling, and monitoring these databases is tedious, expensive, and risky—any mistake can lead to extended application downtime. For the past nine years, Amazon Relational Database Service (RDS) has managed this undifferentiated heavy lifting while providing high availability, durability, and security for databases in AWS, allowing our customers to focus on higher-value work like optimizing application performance. Amazon RDS manages the largest fleet of relational databases in the world, hardened by billions of database-hours of operational experience. Amazon RDS on VMware brings this same experience to customers' on-premises environments.

Amazon RDS on VMware lets customers deploy managed databases in on-premises VMware environments using the Amazon RDS technology enjoyed by hundreds of thousands of AWS customers. Amazon RDS provides cost-efficient and resizable capacity while automating time-consuming administration tasks including hardware provisioning, database setup, patching, and backups, freeing customers to focus on their applications. RDS on VMware brings these same benefits to on-premises deployments, making it easy to set up, operate, and scale databases in VMware vSphere private data centers, or to migrate them to AWS.

RDS on VMware allows customers to utilize the same simple interface for managing databases in on-premises VMware environments as they would use in AWS. Customers can easily replicate RDS on VMware databases to RDS instances in AWS, enabling low-cost hybrid deployments for disaster recovery, read replica bursting, and optional long-term backup retention in Amazon Simple Storage Service (S3).

RDS on VMware is in preview and will be generally available soon, supporting Microsoft SQL Server, PostgreSQL, MySQL, and MariaDB databases, with Oracle to follow in the future. Registration for the preview can be done via this link: https://pages.awscloud.com/rdsonvmwarepreview.html.

BENEFITS AND FEATURES OF AMAZON RDS ON VMWARE

Fully Managed

Automated Database Management

Amazon RDS on VMware automates administrative tasks for on-premises databases in VMware vSphere environments, including database provisioning, operating system and database patching, backup, point-in-time restore, storage and compute scaling, instance health monitoring, and failover. RDS on VMware will support Microsoft SQL Server, PostgreSQL, MySQL, and MariaDB database engines, with Oracle to follow in the future.

Simple Interface

Customers can create, modify, and manage their databases using the RDS Console, APIs, and Command Line Interface (CLI). The same simple interface can be used to manage RDS databases running on VMware or RDS databases running on AWS.

Scalability and Performance

Scalable Resources

Amazon RDS on VMware allows customers to easily scale the compute, storage, and memory resources in their on-premises database instance, similar to an Amazon RDS instance in AWS. For read-heavy workloads, read replicas make it easy to elastically scale out beyond the capacity constraints of a single database instance. Customers can further leverage the scale of the cloud by bursting read replicas to Amazon RDS instances in AWS.

Performance Monitoring

Amazon RDS on VMware provides Amazon CloudWatch metrics for on-premises databases at no additional charge, including resource utilization, I/O activity, and instance connections. RDS on VMware also provides access to Performance Insights, an easy-to-use tool that helps customers quickly assess the load on their database with in-console visualizations.

Available and Durable

Availability Protection

Amazon RDS on VMware uses health monitoring to detect unhealthy database instances and automatically recover them using the same storage volume.

Simple Backup and Restore

Amazon RDS on VMware allows customers to manually or automatically back up their on-premises databases, whether locally or in AWS. RDS on VMware supports point-in-time restore and allows customers to specify an automated backup retention period per database. With per-database retention poli-

cies, customers can specify how long to keep daily snapshots on-premises, if and when to replicate them to AWS, and how long to retain them in the cloud.

Secure and Compliant

Encryption at Rest and in Transit

Amazon RDS on VMware allows customers to encrypt their databases using keys they manage, including encryption of data at rest, in transit, and when archived. RDS on VMware allows customers to control the actions that their users and groups can take on specific RDS resources.

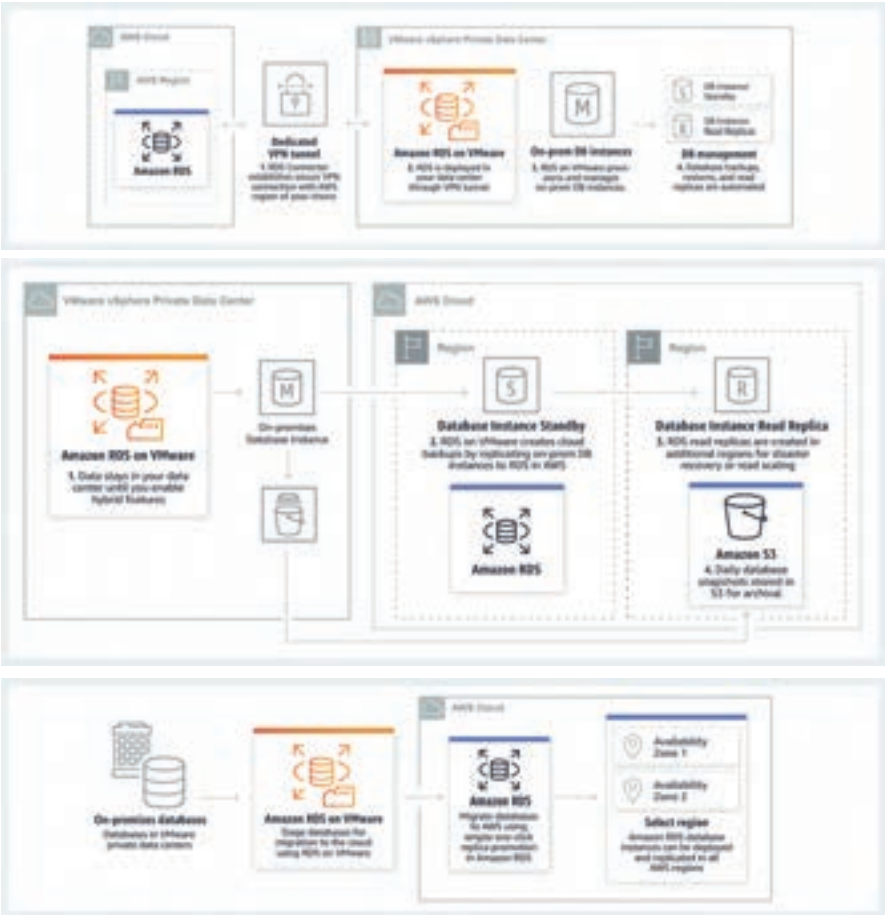
Compliance Readiness

Amazon RDS's default behavior is to keep databases on-premises, allowing customers to run workloads that need to comply with security, privacy, regulatory, and data sovereignty policies. Hybrid cloud features can be enabled at any time, at their discretion.

Use cases

Automate Management of On-Premises Databases

Amazon RDS on VMware reduces operational overhead for database management in on-premises VMware data centers, by automating administrative tasks including software installation, patching, monitoring, and backups. The RDS Connector, a software appliance for the VMware vSphere environment, packages RDS technologies that provide highly available, scalable, and durable database management, enabled through a dedicated VPN tunnel.



Hybrid Cloud Backups and Scaling

Amazon RDS on VMware keeps all of a customer's data and backups in their data center by default, but it also makes it easy to take advantage of AWS scale for disaster recovery, cloud backups, and read replica scaling. The same RDS console, APIs, or CLI can be used to manage RDS databases running on VMware or RDS databases running on AWS.

Migrate Databases to AWS

Through simple one-click replication, Amazon RDS on VMware database instances can be migrated to Amazon RDS database instances in AWS. Databases can be migrated with no impact to uptime, giving customers the ability to rapidly deploy databases in all AWS regions without interrupting the customer experience. ■

AMAZON WEB SERVICES
https://aws.amazon.com/rds/vmware

PERCONA

Managing the Hybrid Future: From Databases to Clouds

DATABASE ADMINISTRATORS IN THE AGE OF DBAAS

Many companies are using or considering database as a service (DBaaS) options as part of their database strategy. One of the alluring attributes of DBaaS is that many of the concerns about running and managing a database are offloaded to the service provider. The provider guarantees uptime and access, has built in redundancies to ensure data availability, and in many cases handles backups and scaling.

The requirements for managing and running a database in a modern enterprise have evolved over the past ten years. Those in charge of running enterprise databases have seen their focus shift from ensuring access and availability, to architecture, design and scalability responsibilities. This means that the role of the DBA is changing as well.

Online enterprises pioneered this change by focusing on ensuring that the company's main revenue engine not only stayed up, but could scale to wherever the business needed to go. This is a far cry from the classic enterprise DBA's top responsibilities: keep it up, backed up, and then react to issues as they present themselves.

When companies move their database to the cloud, the cloud provider takes care of many of the operational, mundane day-to-day tasks. But this does not eliminate the need for database expertise: it moves the function closer to the design and development side of the application. Someone needs to not only design and tune the database to support the application, but also has to understand how to build the modular pieces available in the cloud into a cohesive scalable unit that meets the needs of the application and the company.

WHEN SCALING UP IS NOT THE ANSWER

Cloud providers are happy to scale your environment up to meet growth needs. After all, they make more money and you get continued access to your data. Everyone is happy! At some point, however, you should start questioning the cost of unfettered growth and look for solutions that accommodate growth without more cloud spend. For many of our customers, this is the point where they reach out to us for assistance.

If your page load time increases from one second to three seconds, it is likely that users will notice the delay. A load time increase from 50 milliseconds to 300 milliseconds probably will go unnoticed. Even though this represents a six-fold increase, pages still load quickly.

Under heavy traffic scenarios (e.g., Black Friday), such a seemingly minor issue can massively affect your environment.

From the DBaaS provider's point of view, the obvious solution is to scale up. More hardware means better response time for everyone. This can work well if the problem is with reads, but what if your database is write constrained? There are solutions to this, like implementing sharding, but no one wants to be making that sort of change during a Black Friday type of event. Foresight and foreknowledge are key to managing your environment, containing costs, and keeping the company and users happy.

DATABASES DESIGNED FOR THE FUTURE

Today, application outages mean lost revenue. Worse, it leads to customer churn and gives your competitors an opening. To keep revenue flowing, every part of a company's critical infrastructure needs to be planned out: redundancy should be built-in, and a future-proof architecture should be built to accommodate scale.

The more issues you can design out before launch, the less chance of a catastrophic outage later. This means you want your DBAs and database engineers architecting a database that avoids problems at scale, and you want them working with your developers to write better, more efficient database calls.

This has even greater impact if a company changes direction. If there are significant changes to the application, those alterations can have severe impact on the database. If the corporate direction shifts, that may signal a need to correspondingly redirect the database. Without someone with knowledge and understanding of both sides of the equation, it is easy to get out of balance. What may have worked well in the past may not be ideal as your company's goals and direction change.

CONCLUSION

In the rush to get products out the door, meet deadlines and boost sales, companies often overlook the need to proactively manage and engineer their database infrastructure on-prem and in the cloud. By committing to regular reviews of your infrastructure and application, as well as a regimen of proactive activities, you can avoid wasting money, improve your ROI and keep your business running smoothly. Percona support, managed services, consulting and open source software can help you keep your database environment—on-premises or in the cloud—suited to meet your application workloads. ■

PERCONA
https://www.percona.com/

The Rise of Databases and Hybrid Everything

A MODERN DATABASE needs to support hybrid cloud infrastructure, hybrid transactional/analytical workloads, and hybrid structured/semi-structured data models if it is to meet evolving application development and operational requirements. Whether directly or indirectly, applications require greater analytics, and like agile development before it, agile operations is becoming a requirement in and of itself. As a result, modern databases need to support hybrid everything: hybrid workloads, hybrid infrastructure, and hybrid data.

HYBRID WORLOADS

Today's transactional applications require full analytics and access to years' worth of historical data in order to deliver actionable insight, find compelling opportunities, and create personalized experiences. However, few transactional databases are capable of processing analytics at the scale needed to create competitive advantages and improve customer engagement.

While data warehouses can scale analytical workloads, they're not well-suited to powering interactive, customer-facing applications. Rather, they're used to support offline reporting and analytics—often relying on batch processes and ETL jobs to import data once a day, once a week, or once a month. In time, the results may be used by transactional applications, but not fast enough. As a result, many organizations are turning to hybrid transactional/analytical databases, databases capable of processing transactions and analytics equally as well. MariaDB Platform is such a database, with both row and columnar storage, replicated or distributed, it is engineered to support hybrid transactional/analytical processing at scale.

HYBRID CLOUD

When it comes to the cloud, many organizations are looking to hybrid cloud as a means to introduce cloud infrastructure without incurring greater security and privacy risks, significant development and operational disruption, or increased vendor lock-in. By supplementing on-premises infrastructure, they can gain the operational benefits from cloud infrastructure, including reduced costs and administration overhead, without having to rewrite or migrate existing applications—or operational processes.

There are a number of ways to use hybrid cloud infrastructure for databases. The first, and most common, is to support high availability and disaster recovery by deploying primary databases on premises and secondary databases in the cloud. The next is to support data locality or "bursting" by deploying both primary and secondary databases on premises as well as additional secondary databases, permanently or temporarily, in the cloud to increase read scalability and performance as needed. Conversely, if separate primary databases are running in the cloud, and in multiple regions, their data can be consolidated in a secondary database on premises.

However, innovative organizations are using hybrid transactional/analytical processing to both meet the increasing analytical requirements of customer-facing applications as well as ignite hybrid cloud strategies. MariaDB Platform separates and isolates different workloads. While some database instances process transactions, others process analytics. This not only prevents resource contention, but allows different workloads to run on different infrastructure. For example, to process transactions on premises and with small, but fast SSDs and analytics in the cloud with big, multi-core processors. This approach allows organizations to implement hybrid cloud strategies by continuing to use on-premises deployments for transactions and extending them with cloud infrastructure to support scalable analytics.

HYBRID DATA

Finally, while most transactional applications read and write structured data, there is a place for semi-structured in transactional databases. For example, by extending a relational database schema to include JSON documents, applications can be provided with greater schema flexibility. It should never be relational or JSON, but rather relational plus JSON. In fact, while most data is best modeled as relations, some data is best modeled as JSON documents, and it doesn't require a NoSQL database. MariaDB Platform includes a comprehensive set of SQL functions for querying JSON documents so database administrators can extend relational models with JSON documents—and application developers can take advantage of it.

MARIADB PLATFORM

With MariaDB Platform, there is no need to choose between on-premises and public cloud infrastructure, transactional and analytical workloads, or structured and semi-structured data. Both application developers and database administrators can get the best of all worlds without sacrificing mission-critical database features, including both scalability and performance as well transactions and consistency. It is a database built for a hybrid everything future, providing nothing less than total flexibility and control. ■

Learn more about MariaDB Platform:
https://mariadb.com/products
/mariadb-platform/

Learn more about the MariaDB Platform Managed Service:
https://mariadb.com/products
/mariadb-platform-managed-service/

MARIADB
www.mariadb.com

SentryOne

Streamline Cloud Database Migrations with SentryOne

Cloud data platforms offer agility, scalability, and elasticity; business process efficiency; and a lower total cost of ownership (TCO). But how do you smoothly migrate on-premises and virtualized databases to the cloud and ensure optimal performance post-migration? Cloud database migration might seem daunting, but the proper process and tools can eliminate many of the potential risks.

DEFINE YOUR CLOUD MIGRATION PROCESS

A defined process is critical when tackling a cloud database migration. Your process must align with your business objectives and environment. The primary phases of the cloud database migration process are as follows:

1. Discover
2. Optimize and Test
3. Migrate
4. Test, Monitor, and Optimize

You should expect to modify your plan and scope throughout the process based on feedback and reality—but try to avoid scope creep.

Let's take a closer look at each phase of this process and how SentryOne can help simplify cloud database migration.

DISCOVER

You need to understand the size and complexity of your migration, which can be difficult and time-consuming if you don't already have detailed documentation of your data estate, its properties, and interdependencies. SentryOne DOC xPress automates metadata-based documentation and data lineage tracking for SQL Server, SSIS, SSAS, SSRS, Power BI, and Tableau. Although documentation is an arduous task, you can't underestimate the complexity of data lineage and dependency tracking. Automation relieves a considerable burden and reduces human error. What's the impact of moving a specific database? DOC xPress provides this information. This visibility is critical and delivers incalculable value as you confirm the scope of your migration and adjust your plan.

Understanding your end users' performance expectations is also crucial to ensuring you can provide a similar performance experience post-migration without overpaying for cloud resources. You can use SentryOne SQL Sentry, which provides breakthrough Microsoft Data Platform performance monitoring, to capture baselines on your current databases for future, post-migration comparison. SQL Sentry also provides visibility and metrics to justify decisions about the cloud technologies selected.

OPTIMIZE AND TEST

Bloated and poorly performing databases won't improve in the cloud and will cost more. Prior to migration, you'll want to optimize your on-premises and virtualized workloads. Use SentryOne SQL Sentry, BI Sentry, DBA xPress, and BI xPress to optimize database performance and reduce disk space requirements across all SQL Server workloads. Doing so can significantly reduce your cloud footprint—and your cloud spend.

Before migrating SSIS into IaaS or Azure Data Factory, standardize your architecture with SentryOne Task Factory, which includes more than 60 data connectors and SSIS components.

After you make a change to your databases, you need to confirm everything is working as expected. If you already have data testing in place, your current tests can be revised to validate migrated data. But if you aren't already testing your data, you're open to significant liability. SentryOne Test and LegiTest are automated data testing platforms built on industry standard frameworks that enable extensive data validation. Build data validation tests that can be used to assure stakeholders that migrated data is intact.

MIGRATE

The actual migration phase is often the smallest part of the process. There are many options available to move data into the cloud, including backup and restore, detach and reattach, replication, log shipping, and data migration services. The key is ensuring you have a series of decision points set so that if your migration isn't progressing as planned, you can roll back.

TEST, MONITOR, AND OPTIMIZE

As your data lands in the cloud, you will want to perform your planned data tests with SentryOne Test and LegiTest to make sure your actual results align with your expected results. As tests are completed, bring your cloud databases into service. Use SQL Sentry (or BI Sentry and DB Sentry as appropriate) to capture new performance baselines of cloud database instances and compare them with your pre-migration baselines. Then, set up alerting and other automations and continue to monitor and optimize database performance post-migration.

SUMMARY

Successful cloud database migrations are possible with the right process and tools—and when risk is managed. From data documentation, data testing, and monitoring to cloud database workload optimization, SentryOne helps you maintain data integrity and ensure peak performance before and after you transition to the cloud.

ABOUT SENTRYONE

SentryOne helps Microsoft data professionals build, test, document, and monitor data-centric applications on SQL Server, Azure SQL Database, and the Microsoft Data Platform. ■

SENTRYONE
https://www.sentryone.com/

DELPHIX

Standardize on an Extensible Platform to Manage Data Source Heterogeneity Across Hybrid Cloud

Data lives in many different places within organizations, across on-premise and public cloud. Enterprise software teams use a diverse set of data sources tailored to various applications running in these locations. Common data sources include Oracle, SQL Server, PostgreSQL, MongoDB, AWS, RDS, AWS Aurora, AzureSQL, among many others. What's changed today is that organizations are no longer compromising on features, functionality or delivery schedules of their applications to adjust to a single data source capability. This has resulted in an explosion of multiple data sources within a single organization spread across on-premise and public cloud. As a result of data source heterogeneity, the complexity of managing the flow from multiple data sources to different teams, like AppDev, QA/test, and analytics, increases over time. Furthermore, the associated support, maintenance, and personnel also vary for each data source. Instead, organizations need to align their data management policies and tools to modern CI/CD, DevOps and cloud infrastructure tooling to accelerate time-to-market for the applications that drive their competitive advantage.

DATA SOURCE HETEROGENEITY LEADS TO DATA MANAGEMENT PROBLEMS

Standing in the way of executing application development projects faster is data access from different data sources customized to the various requirements as each data source has its unique data management features and limitations. Both data consumers—such as developers, testers, data scientists—and data managers—including database administrators, storage admins, infosec—spend additional time learning data-related tasks tied to each data source for every application to support development, testing, and analytics. For data managers, this means building custom workflows with trade-offs for regular data management tasks, such as provisioning, refresh, integration, and recovery tasks. The entire process may be done through a cumbersome request-fulfill model fine-tuned to that data source. By adding coordination and approvals, each request can balloon into a multi-team and multi-week exercise that can cripple data delivery and compromise data quality, ultimately putting projects at risk. On the other hand, data consumers are being asked to adapt and incorporate individual data source limitations and dependencies into their workflow, which can delay and restrict the number of software releases.

SUCCESS STRATEGY

To be successful, organizations need to adopt a platform-based DataOps approach that is extensible to heterogeneous data sources and can work seamlessly across on-premise and cloud. A DataOps platform enables the rapid, automated, and secure management of data. It reduces data friction by providing a collaborative environment for data managers and data consumers. Further, it ensures that sensitive data is secured and the right data is made available to the right people, when and where they need it.

For an environment with heterogeneous data sources, a DataOps platform provides the same experience across multiple data sources. For data managers, the platform virtualizes data sources to create lightweight virtual copies that are continuously synced with source data to deliver data to key stakeholders. It non-disruptively collects data from databases, applications, and file systems. After compressing the data, the platform stays synchronized with sources by recording all changes over time. It automates tasks by eliminating manual effort and coordination challenges, thus removing critical bottlenecks to provision data and reduces the wait time and receiving times of data environments. The platform provides the ability to view and manage data environments across users while driving role-based accessibility of data environments. Data delivery is done securely from production to non-production environments on a recurring basis with access control policies implemented in non-production. The platform spins up and tears down data environments as a part of development or the CI/CD pipeline seamlessly and integrates with existing built-in workflow tools, like Jenkins. Data consumers get personal data environments to manipulate at will with Dynamic Data Controls. They can quickly replace the data within virtual data environments through refresh, branches, bookmarks, sharing, and reset provided as self-service data version controls.

ABOUT DELPHIX:

Delphix transforms how organizations address the data source heterogeneity challenges by delivering data for modern SDLC, AppDev or QA teams whether it's on-premise or on the cloud. The Delphix platform enables enterprises to provide developers and testers with 24/7 access to all the data they need while meeting the requirements of their regulatory and compliance framework. With Delphix, teams can stand up a complete development environment in minutes, deliver data that is full and has faithful copies of production, and automated testing to achieve CI/CD targets. Delphix makes this possible by integrating data masking that improves, rather than threatens, data security requirements.

Organizations can now manage data distribution and access with the speed, simplicity and levels of security required to drive their digital transformation with the Delphix platform. ■

DELPHIX
www.delphix.com

Couchbase

The NoSQL Database Architected for Multi-Cloud Agility

With technology evolving at a record pace, today's customers and employees expect their digital experiences to keep improving just as often and just as fast. To meet these constantly rising expectations, businesses are continuously looking for new ways to become more agile. At the same time, IT departments are looking for new ways to rein in costs even as their technical requirements grow more demanding day by day.

WHY HYBRID?

Hybrid clouds address both challenges—agility and cost—and have become a key component of both IT and business strategy for most enterprises. A hybrid cloud simply refers to a mixed environment of on-premises IT infrastructure, private cloud, and public cloud. A hybrid strategy can be more cost-effective because the public cloud hardware, software, and supporting infrastructure is owned and managed by the cloud provider. You pay only for the resources you use, and only when you use them.

Managed correctly, a hybrid cloud strategy can also provide increased agility from many angles. Business continuity is assured because your mission-critical data is replicated and available even if you experience an on-premises failure or disaster. Innovation is easier because additional infrastructure is always available on demand. Scalability is virtually unlimited, and you can scale quickly and easily at times of peak demand. And you can speed up time to market by quickly setting up environments to test, prototype, and launch new products or services.

As attractive as the upside may be, hybrid clouds don't come without some challenges of their own. Fortunately, NoSQL is especially well-suited to help you overcome them.

OBSTACLES TO CLOUD ADOPTION

Cloud portability: Today's enterprises depend on a mix of multiple private and public clouds. To leverage the full benefits of these clouds, enterprises need frictionless data portability across multiple cloud platforms. An agile business can't afford to be locked into a particular cloud provider or its proprietary database.

Business continuity: Organizations spend substantial time planning for unpredictable peak workloads and scaling their applications to accommodate new geographies. It's already a struggle to meet fluctuating demand in a cost-effective manner using traditional methods. Adding cloud into the mix can make things even more complicated.

Data governance: "Anytime, anywhere" access to data in the cloud can create major challenges for businesses that are required to enforce authorized access while complying with the laws, rules, and regulations of each country as the data crosses numerous geographic boundaries.

COUCHBASE IS CLOUD NATIVE

Couchbase is architected to specifically address the challenges of both cloud and on-premises solutions—making it ideal for hybrid scenarios requiring maximum deployment flexibility.

Couchbase is a geo-distributed cloud-native NoSQL database that provides unique data model flexibility with native support for JSON, elastic scalability with a shared-nothing architecture, high performance through built-in cache, and always-on availability through inter- and intra-cluster data replication.

Elastic scalability and global data distribution: Couchbase's cloud-native architecture provides performance at scale, on-demand scaling to handle unpredictable and spiky workloads, and the ability to easily distribute data globally.

Governance and security: With built-in Role-Based Access Control (RBAC), enterprise-grade encryption from client applications to the cloud, and data filtering capability in the replication protocol, you can enforce security standards and also define geo-boundaries where data should reside based on data governance rules.

Containers and orchestration: With the Couchbase Autonomous Operator for Kubernetes, customers can now run Couchbase as a service on their private clouds. Operational complexity can be reduced up to 95% by implementing operational best practices and running Couchbase as an autonomous, fully managed stateful database application next to other microservices applications, all on the same Kubernetes platform.

Public cloud partners: Couchbase can be deployed in just minutes on the most popular public clouds, including Amazon Web Services, Microsoft Azure, and Google Cloud Platform.

FAST TRACK TO CLOUD ADOPTION

Couchbase is considered the leading NoSQL database technology in the cloud for businesses focused on digital innovation, and is the first and only NoSQL vendor offering native integration of Kubernetes. Now, Couchbase Managed Cloud allows you to simplify your cloud adoption even further with an always-on 24/7/365 support model that eliminates the operational burden of deploying and managing your data platform infrastructure in the cloud. ■

COUCHBASE
www.couchbase.com

Quest

All Your Platforms in One Console: Monitoring Hybrid Cloud Environments with a Single Solution

Cloud computing is having a bigger impact on database administration than any other emerging trend, according to the majority of DBAs surveyed by Unisphere Research.1 As your database landscape shifts from relational databases offered by commercial vendors like SQL Server, Oracle and DB2 to include open source systems like MySQL and PostgreSQL, and non-relational systems like MongoDB and Cassandra, you not only have to administer more and more databases, but you also have to learn and deploy a much broader variety of platforms and database instances.

LARGE, CROSS-PLATFORM ENVIRONMENTS POSE SERIOUS CHALLENGES

As a DBA, database performance monitoring is one of your top responsibilities and system uptime is likely the key metric by which you are measured. But when you're juggling new technologies and growing responsibilities, it's hard to keep up. Adding tools for each platform may seem like a helpful approach; however, this comes at a high price, adds complexity and reduces productivity, as you're forced to switch from tool to tool. But the most serious challenge with this approach is that it doesn't give you an overall understanding of the health and performance of your environment as a whole, which leaves you vulnerable to risk.

A SAFER, EASIER STRATEGY FOR MANAGING HYBRID ENVIRONMENTS

What if you could use a single, intuitive interface to monitor and manage all the diverse platforms in your environment? Standardizing on one powerful solution would reduce licensing costs while giving you cross-platform visibility. It would also shorten the learning curve for taking on new platforms, which would increase your productivity and make it easier for you to avoid risk. All of this means you'd cut costs while gaining the ability to proactively ensure peak database performance—across your entire database environment.

THE SINGLE INTERFACE FOR EVERY CROSS-PLATFORM CHALLENGE

With Foglight® for Databases, you gain unprecedented visibility across all your database platforms. From one simple-to-use console, you can quickly diagnose and resolve emerging issues to prevent business interruption. And because Foglight supports a wide range of platforms, including on-premises and cloud databases, you get cost-effective database monitoring software for the platforms you use today, and the ones you'll use in the future.

A WEALTH OF FEATURES FOR ALL YOUR DATABASE PLATFORMS

Unlike database-specific point solutions, Foglight for Databases delivers a global view and powerful features to monitor and manage your entire database environment, including:

- Centralized database monitoring across on-premises, hosted systems and cloud databases
- Unrivaled cross-platform visibility
- Proactive issue resolution
- Alerts for deviations from normal metrics
- Diagnostics and alarm data viewable from any drilldown screen
- Custom views and reports

ADD PLATFORMS—SUBTRACT STRESS

The success of your business hinges on database availability. You can easily ensure uptime across your on-premises and cloud databases with Foglight for Databases. ■

WANT TO LEARN MORE?

www.quest.com/foglight-for-cross-platform-databases

1 Elliot King, "The Real World of the Database Administrator," Unisphere Research, March 2015, https://www.quest.com/whitepaper/ the-real-world-of-the-database-administrator-875469/

Monitor all your databases through one console with Foglight for Databases.

MV SOLUTIONS

ZUMASYS ACQUIRES OPENQM AND HIRES MV EXPERT

Zumasys, a provider of software, databases and tools for business-critical Pick applications, has acquired OpenQM, a high-performance, self-tuning MultiValue database, from U.K.-based Ladybridge Systems.

According to Zumasys, as the exclusive global distributor of OpenQM since 2015, Zumasys more than doubled worldwide sales of OpenQM for 3 consecutive years by adding 24×7 telephone support, professional services, and an automated conversion process to help customers migrate their existing applications to OpenQM.

Since the first release of the OpenQM product in 1993 as an in-house embedded database, its creator, Martin Phillips, has honed and refined the product to include a variety of modern features that help customers cut development time and enables the utilization of big data, virtualization, and cloud-based systems, said Zumasys.

OpenQM provides customers a scalable MultiValue database solution that's compatible with existing applications and other major MultiValue products.

"It is the right moment for Zumasys to take on ownership of OpenQM," said Phillips. "We are excited by the opportunities that this brings for an increased development team to add further new features to this well-established product. Continuity will be ensured as I will continue to work on OpenQM as a Zumasys employee."

Separately, Zumasys also announced that Josh Camacho, former CEO of Chicago-based OPTO International, is joining the company. Camacho, who will become senior manager of software development, brings more than 15 years of hands-on development experience in jBASE, which was used to run every aspect of his business.

Camacho has been a MultiValue software development contractor with Zumasys since March 2018. In his new role, Camacho will provide development services to customers and have direct input into the Zumasys MultiValue product roadmap and strategy for the future.

"Josh was once a jBASE customer himself, so he truly understands firsthand the challenge that our customers face with their applications," said Zumasys president Paul Giobbi. "Customers want to modernize their apps without losing decades of development and innovation. Josh used the company's jBASE application and fundamentally transformed the way they do business. We want to share that success and replicate it with our Pick MultiValue customers around the world."
www.zumasys.com

Jigsaw24 Continues to Grow With Rocket Software

Rocket Software is continuing to help Jigsaw 24 to automate and extend business processes to scale and enable continued growth. Founded in 1992, Jigsaw24 is a U.K.-based provider of technology solutions, managed services, and strategic consulting that is focused on Apple products. Its portfolio includes managed and productized services and solutions, as well as specialized infrastructure for media and end-user creative technologies used for design and print.

In the late 1990s, it became increasingly apparent that the slow and disconnected systems used for order processing, inventory management, picklist assembly, and shipping were leading to an increase in errors throughout the business, inhibiting further growth.

The Jigsaw24 team understood it was critically important to install an enterprise resource planning (ERP) solution that would improve their end-to-end processes and drive long-term growth, even though the projected investment was substantial.

After evaluating a number of options, in 1999 Jigsaw24 deployed an ERP system from Masterpack, a solution built on the Rocket UniVerse MultiValue Application Platform.

According to Neil Box, DevOps and information systems and technology director, Jigsaw24, the goal was to deploy a system that could function as an "enabler" and not a "blocker," delivering robust, extensible functionality for the business.

In 2001, Jigsaw24 transitioned away from printed catalogs in favor of an ecommerce-enabled website, implementing the Rocket U2 Web Development Environment to provide dynamic integration with Masterpack. This resulted in an improved experience for both customers and employees. Today, Jigsaw24 also integrates Masterpack with Salesforce for marketing, CRM, and service management, providing an all-encompassing customer-centric solution.

In addition to 8X growth in annual revenue, Jigsaw24 has attracted new investment partners and expanded from 50 employees to 250, with additional offices in London, Glasgow, and Cardiff.

Thanks to the efficiencies delivered by the UniVerse-based ERP system, the information systems team is still able to deliver across all functions with only four people.

In 2007, Jigsaw24 began to develop custom Masterpack-integrated portals for its customers at no charge. "UniVerse enables our team to measure, deliver, and constantly improve our offerings on a truly agile basis. Simply put, we are able to quickly meet the demands of the business as it continues to grow," Box said.
www.rocketsoftware.com

MV SOLUTIONS

Axiz to Distribute Evoke in South African Markets

South African technology distributor Axiz is partnering with BlueFinity for the distribution of Evoke. The partnership will enable Axiz to distribute Evoke to customers across the South African continent as well as provide for the support, development, and implementation services associated with the development of business apps. Axiz will also leverage Evoke to design and develop its own apps for the benefit of its customers.

For BlueFinity, the partnership brings with it the support of Axiz' experienced technology, development, and sales teams, which will allow it to expand the already rapidly growing, Evoke customer base into new markets. In addition, Evoke's ability to generate solutions that can integrate both SQL and MultiValue databases into the same desktop and mobile app offers great possibilities to Axiz' clients.

"Axiz is a dynamic company and a major player in the technology market in South and East Africa, with an excellent reputation. We're delighted that Evoke will become their flagship app development solution," said Malcolm Carroll, director at BlueFinity International. www.bluefinity.com

Pick Cloud Offers Path for MV Customers to Enter the Cloud

Pick Cloud is introducing another cloud hosting option that bundles printing and reporting as part of its Google Cloud offering. "It is now easier than ever for MultiValue customers to enter the cloud. Those wanting to start their journey securely and affordably can now do so with our cloud starter pack," said Mark Pick, CEO of Pick Cloud.

The new approach leverages the Google Cloud Platform to host infrastructure; CirrusPrint, a seamless way to print from the cloud to any on-premise printer; and phiReport cloud edition, a reporting tool that allows any Windows user with a current AccuTerm terminal emulator to define and generate attractive Excel worksheets from a graphical user interface.

"We are maximizing the relationships with our SaaS vendors and our decades of Pick/MultiValue experience to offer the community a way to stay on top of the latest technology and extend their MultiValue investment," said Pick. www.mypickcloud.com ■

MV EXPERTS PLAN TO EVOLVE WITH THE MARKET IN 2019

New challenges and opportunities in data management and analytics are constantly evolving. With a new year just beginning, MV companies are renewing their focus on the next big thing.

As we enter 2019, compliance with the EU's General Data Protection Regulation (GDPR), California's Consumer Privacy Act, increasing security concerns, the need for integration with additional data management frameworks, and cloud platforms are just some of issues that are trending for many companies.

In the annual "MultiValue Special Report," this year, *DBTA* asks: What are your predictions for 2019 and how is your MV platform addressing them? ■

—Stephanie Simone

Mike Ruane, President and CEO, Revelation Software

As we continue development on OpenInsight v10 in 2019, we will be providing developers with tools that concentrate on the end-user experience. Our focus on user interface tools allows OpenInsight to consume data from a variety of back ends, and "mashup" the data in a variety of front-ends (e.g., the browser, desktop, mobile, and office applications).

Among the available back-end databases we can interface with is, Couchbase. We anticipate growing the stable of NoSQL cloud database sources throughout the coming year.

To address growing security concerns, our Universal Driver data server now supports "end to end" encrypted communication as well as an optional "verified client." Our strategy is to enable our customers to use best-of-breed proven security tools.

We also see growth in our ability to offer management tools for maintaining optimum system performance. While OpenInsight v10 includes a management console that provides database administrators with basic tools to monitor and measure their systems, in 2019 we see continued evolution to extend our OpenInsight toolset to accommodate increasing deployment complexity.

Julianna Cammarano, Director, MultiValue and Business Intelligence Product Marketing, Rocket Software

2019 will be all about modernizing your application by leveraging modern development architectures and technology.

The Rocket R&D organization has expanded its global footprint, enabling 24x7 development with MultiValue teams around the world from Denver to China and India. Our commitment to the MV platform means you can expect significant advancements in methods to enrich your application by opening access to your application logic and data.

We'll be announcing an API gateway based on exposure through RESTful services as well as native APIs, allowing connectivity through Node.js, microservices, Python, and native Rocket UO libraries. This is just the beginning; the MV gateway will be the foundation for continuous service monitoring, built-in resiliency, load-balancing, clustering, containerization, and easy deployment.

We'll also be working with customers as they navigate through cloud options: Is it the right move? What are the advantages and disadvantages? Which cloud provider is most aligned with their business needs?

We've started the MV Cloud Accreditation Program which will provide MV customers with a suggested list of vendors that can help them host their data and applications. Customers can then select the best accredited partner to work with on their needs and application architecture requirements.

We see 2019 as a year of application enrichment through modernization so that our customers can expand their market share.

SPECIAL SECTION

MV SOLUTIONS

Mark Pick, President and CEO, Pick Cloud

We see MultiValue as integrating with new and emerging technologies and 2019 will be no different.

MultiValue vendors such as Rocket Software and Zumasys are bringing new technical talent into the market. This helps keep MultiValue innovative.

We built Pick Cloud around the cloud and SaaS model. We help MultiValue clients of all sizes take advantage of everything the cloud has to offer. Additionally, we provide OpenQM DBaaS and phiReport cloud edition. These are both offered for a small predictable monthly fee.

OpenQM DBaaS allows developers and VARs to offer their software at a competitive price point to that of non-MultiValue providers with all the infrastructure and licenses included (OpenQM and AccuTerm). This is a departure from the traditional per-seat model where clients had to come up with a large sum of money in advance for both software and hardware.

For security, Google data centers are manned and monitored 24/7 by more than 900 security experts. They adhere to the major data handling mandates including GDPR, HIPAA, GLBA, PCI DSS, and many more. Data is encrypted at rest and in transit automatically.

Technology is moving more and more toward cloud and SaaS-based models and MultiValue should do the same.

Doug Leupen, President and CEO, Entrinsik

Working with MultiValue databases since 1984, Entrinsik believes MV databases will continue to be excellent resources that deliver tremendous flexibility to both enterprise and SMB organizations.

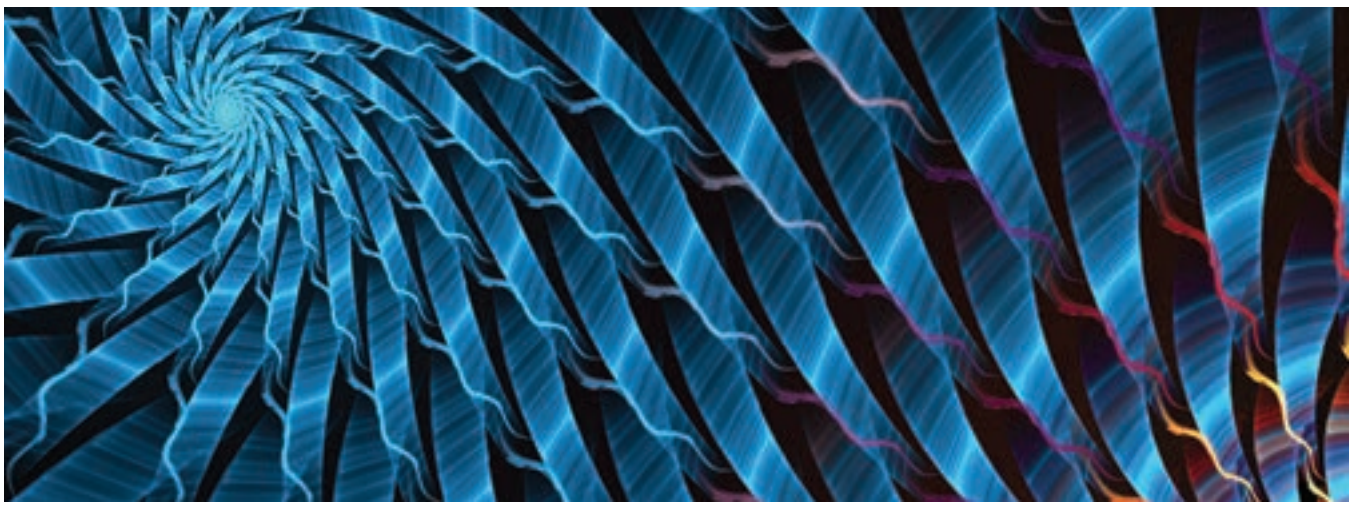
While our two products offer two different solutions, MV databases play a critical role within these applications. For example:

Informer 5, our business intelligence product, pulls data from multiple data sources into an integrated data management framework to provide intuitive charts and reports for end users.

Because of our in-depth knowledge of MV, we have effectively modernized reporting from MV databases by making them significantly easier to work with, enabling organizations to easily access MV data, blend this data with data from other data sources, conduct detailed analysis, and produce highly informative reports.

Enrole, our cloud-based product for universities and corporations needing to schedule and manage events, provides an enterprise resource planning (ERP) solution that uses Rocket UniVerse as its backbone. To provide an easy-to-navigate user interface, we enabled connections to UniVerse via methods used by Java.

Looking forward to 2019, it will be increasingly important for MV database providers to continue to improve the performance of their respective databases and to provide programmatically addressable structured result sets to a database query.



MV SOLUTIONS ▸▸ SPECIAL SECTION

Malcolm Carroll, Director, BlueFinity International

This last year has seen significant growth in the use of apps in the MultiValue environment. Customers are becoming more aware of the potential of creating apps through Evoke and consequently requirements are becoming more sophisticated.

One is the access to data from multiple sources. Many MV customers also have SQL-based systems. Increasingly, there are requirements to include data from multiple databases and multiple types of databases in the same app.

The second is being able to include custom code from established MV systems into new apps as an integral part of the app (while maintaining the integrity of the rapid application development platform so that changes to the app can still be made through the platform).

The third is having the flexibility to migrate from web apps to hybrid or native apps from the same design.

With regard to businesses, there are several areas where we see greater focus. The deployment of "citizen developers" using RADP is of growing importance.

We see the introduction of integral data visualization (graphics); the use of chatbots for all sizes of companies gaining traction; and the deployment of apps across multiple devices as ever-increasing requirements in 2019.

Keith Lambert, VP, Marketing and Business Development, Kore Technologies

As in the past, Kore's clients use a multitude of ERP applications, but those clients are buying other third party apps to supplement their ERP apps.

These third-party applications need to be bidirectionally integrated to the MultiValue ERP application. What Kore is doing is providing these companies with tools that simplify the import and export of data to these third-party applications and the ERP application.

As batch-oriented integration evolves into real-time integration, Kore's Kourier REST platform continues to add features and updates for RESTful web services to improve access to the ERP data from CRM and other third-party apps. This REST-as-a-server approach has been supported by Kourier for several years.

Companies now want their ERP system to talk directly to third-party applications by calling REST endpoints. Kore has created application connectors, which will be part of a library, and are used to send and receive data through the application using a REST-as-a-client approach. Kourier acts as the middleware piece for the MultiValue applications.

Kore will continue to support hybrid integrations that use multiple approaches to connect MultiValue systems with other applications as the market moves toward supporting real-time as the standard. ■

How DBAs Can Survive and Thrive in a World of Agile Development

It's all about how fast your business can search and query its data.

By Ken Rugg

In the days of cloud computing and agile development, it might seem that being a DBA is somewhat less appealing as a career choice. However, the role of DBA is changing; it is not going away. In fact, it is becoming even more important. That said, if DBAs are to explain their value to the rest of the business, they must be seen as more than the company's data librarians. Rather than belonging in a musty old library, DBAs should be responsible for creating the foundations to turn their organizations into high-performance businesses fit for the digital era. This means addressing a number of technical requirements, but above all, it boils down to how fast an organization can use its data to gain business-impacting insights. To achieve this, DBAs must retool their query optimizing skills to become data performance pros!

Database Complexity

This is no small challenge. The pressure to deliver applications at an ever-faster rate has given developers the freedom to "do whatever it takes" to build at speed and ignore the constraints of traditional corporate IT processes. This, in turn, has given rise to a whole new variety of database architectures such as document databases, which allow developers to compile data more quickly.

The document model works well when there are many variables that the developer wants to address. For example, consider a customer that is taking his or her car to the garage for service. The customer record will contain information that is structured, such as billing information and the make and model of the car. Some of it may not be, because the mechanic may have taken ad hoc notes or taken pictures of the repair work to explain what has been done to fix it. This could sit in a document database, enabling the developer to have all the information about the customer represented in a simple way, and information could be added to the database very

TRENDS ▸▸

quickly as the status of the car changes. The same approach could be used for medical records, where the information about the patient is always accessed through the record, but lots of different data types, such as diagnosis information, X-ray material, courses of treatment being recommended, and general notes from the physicians, are continuously added.

All this sounds great if you are an application developer, but it does make it harder to search and analyze data. Compare this to the old days when DBAs controlled the database structure to get the best out of it. The DBA was the data expert, who understood there are trade-offs and how to optimize systems to get the best performance overall. The DBA would make sure the database used the right indexes, normalized the schema, and optimized the overall performance of queries. For instance, while spreading the information about a customer or patient across several tables in a normalized schema may make the application developer's job less straightforward when creating the record, a query to find everyone with an overdue bill could be highly optimized.

Databases Demand a Hybrid, Multi-Model Approach

Today, with so many types of data modeling options, it is harder to optimize data access across the different formats and access patterns. It can be very convenient to use document models to organize all the information relating to a particular entity in one place, to avoid storing unnecessary information, and to be able to attach ad hoc information directly to that record. At the same time, the lack of consistent structure and content across all entities of a given type can make it harder for the system to optimize standard query performance.

The solution for this challenge is to adopt a hybrid or multi-model approach to represent and search data. For instance, this is why PostgreSQL provides native support for storing JSON documents alongside its traditional relational table structures. This way, data that is highly variable can be represented as a JSON document and can be easily integrated with more structured data in the PostgreSQL database.

*There is a **great opportunity** for DBAs to be integral to the performance of applications in the digital business era, but it does require an understanding of the **whole toolbox** of optimization techniques available.*

This approach gives DBAs greater flexibility to move data between document and relational models. It also reflects situations where objects can have specific immutable features, such as size, price, and what type of object it is. Returning to the garage example, the standard customer record could be in a relational table while the log of work history for a particular vehicle could be a document which includes the mechanic's notes and photos.

Another benefit of this approach is that you can start with a very ad hoc, document-based model and evolve it over time. Once you have gained enough experience on what data is required and how the data is most frequently accessed, you can then clean up the database, moving data from the document model to a relational one to provide a more structured view of the data.

Of course, document and relational models aren't the only options provided by data management systems today. For some use cases, it is convenient to be able to search using time-series, geo-based representations. Time-series approaches allow the database to run queries based on when certain events have happened. For example, the mechanic could search for how often customers replace their tires within months of having their oil changed. That may seem silly, but as we move closer to autonomous vehicles, it will enable the mechanic to analyze information in real time as event stream data. Likewise, using a geographic index to analyze where these events occur could also yield valuable insight. These same approaches could be used to analyze high-volume transactions in stock training or assessing IoT data to identify threats or peak performance demands. Even simple key-value lookups can be useful to optimize certain types of data access, for example, when caching information for frequent access.

A True Pro Knows When Not to Optimize

It should be noted, however, that being a true database performance expert is not just be about data modeling and optimizing queries. It is also about knowing when to apply those skills. There are many situations in which it is a waste of time to optimize queries. For example, the garage may only need to run a query about its office stationery requirements once each quarter, as it may only entail searching a small amount of data requiring a limited timeframe to complete the command. This type of optimization tiers what data an organization needs to query and when it is best to run the query.

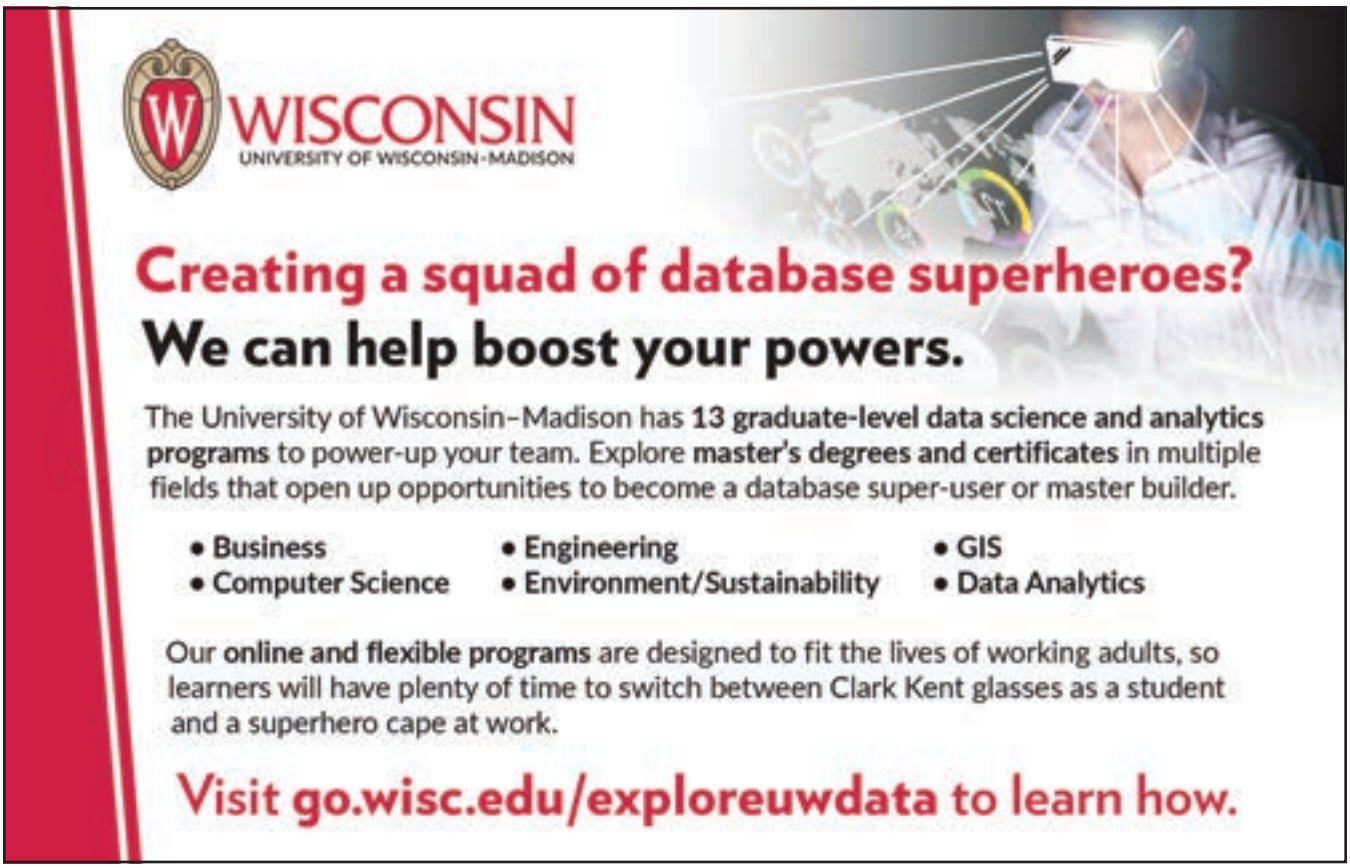
The advent of cloud computing adds another dimension. If a database is hosted in a public cloud where resources

are flexible, it may not make sense to optimize queries that are quite complex and require a lot of resources. For example, if a report is only run at the end of the quarter, perhaps the DBA can simply scale up the hardware to support this demand. The cost of a few hours of additional capacity would be dramatically lower than the cost to rework the queries, much less to re-architect the data model to squeeze out a bit more performance.

Ultimately, all of this knowledge about query optimization lies with the DBA, because that person's remit has always been about the data. It presents a great opportunity for DBAs to be integral to the performance of applications in the digital business era, but it does require an understanding of the whole toolbox of optimization techniques available. The DBA's role is broadening too, because it is not just about optimizing queries as in days gone by; it is also about what that optimization means for the business. What are you optimizing for? Are you optimizing for the best experience for the end user? Technically, this translates into ensuring a user can run a query in 1 second not 30 seconds, which requires an in-depth understanding of the impact on latency of running a particular search.

Application developers are going to continue to choose the database model that is most convenient for them, so it is incumbent on the DBA to optimize access to the data for users. That is more easily achieved if the DBA is using a multi-modal database that can integrate different data types, but also can be tuned for structured queries. It allows for far greater management and optimization of the data in a single location. The DBAs that become obsessive about such performance will be able to demonstrate the visible impact they are having on the business needs of their organizations. By its simplest measure, users will experience far more responsive, efficient, and accurate data searches, but more importantly, they will be able to do their jobs more effectively, collaborate more closely, and understand their customers more deeply as the data they are using is more complete. The DBAs that can meet this challenge will demonstrate how strategic their roles are to creating high-performance IT infrastructures for the digital business era. ■

Ken Rugg is chief product and strategy officer for EnterpriseDB (www.enterprisedb.com) and is charged with leading the company's product and strategic vision.



APPLICATIONS ▸▸

The Shift to a Serverless World

By Nenshad Bardoliwalla and Pakshi Rajan

With very few exceptions, today's IT modernization strategies include moving workloads from traditional data centers into the cloud. In fact, a recent survey on cloud adoption by LogicMonitor indicates enterprises will run 83% of their workloads this way. Topping the list of business initiatives that are driving this trend is digital transformation, followed by IT agility.

The shift to cloud eliminates the challenges of running a physical data center. However, organizations still continue to face the software deployment management challenge in terms of ensuring that their applications are always running and running in an optimal manner (i.e., they consume the right level of resources).

Cloud service providers have long understood the key to providing these shared compute resources is to be able to have massive resource pools that are then virtualized or containerized into smaller, multi-tenant storage or compute units which are then charged out. But, while the abstraction of the storage and compute aspects has been achieved, the management of applications all the way, from the operating system (OS) on up, and taking care of variations at each layer have remained a major challenge.

Over the past 5 years, we have seen the emergence of new technologies such as Docker that enabled organizations to bring the same abstraction and scaling to the applications side. In short, Docker provides a way for a software developer to put the entire software stack into a container, i.e., a Docker image. The developer then tests and ships their software as one or more Docker images, knowing that it can be deployed anywhere using Docker run-time and it will work exactly as they intended and tested it. The developer now controls the entire software deployment stack without having to cater to every nuance in the target environment.

Most enterprise software these days consists of multiple smaller components or services. Gone are the days of large monolithic software. When software consists of multiple components, Docker Compose provides a way for a developer to specify how these services should coordinate. But making sure that these individual processes continue to run and coordinate as originally designed remains a challenge at many organizations.

And then there is the need for scaling the applications up and down in response to usage or demand and making sure that the user experience is pleasant and at the same time it is not achieved by over-paying for infrastructure than what is absolutely necessary. And this requires ▸▸

Enhancing ETL with Automated Data Quality

DATA IS FLOWING into today's organizations from an almost unimaginable array of sources. To make that data usable, companies often turn to ETL (extract, transform and load) programs. These programs help to select the data needed, transform the data by cleaning it and adding information needed for business, and then rapidly move it into that system for data analysis, marketing campaigns and other uses.

Bad Data Has a Cost

Good ETL products offer the ability to chain and automate processes, schedule jobs and support the reporting and sharing of knowledge bases. But, too often, the accuracy of that data fails the company for a number of reasons. That's why it is important for ETL processes to incorporate data cleansing.

The risk to companies is real. In fact, 91% of businesses are hurt by common data errors; including incorrect and inaccurate data, outdated data, missing information and duplicate data. Without accurate data, it is difficult to identify new customers, understand customer needs and increase the value of each customer.

Dirty data also prevents a single view of your customer. A particular individual or family can be in the data many times with different typos and filled in values. Moreover, bad data costs U.S. companies $3.1 trillion annually, according to a report in the Harvard Business Review.

That's why many organizations, as well as leading data integration software vendors, choose data quality expert Melissa to help them chain in data quality as an integral part of the ETL process—improving the efficiency and accuracy of every project.

Melissa offers a complete data quality platform, as well as plug-ins for leading data integration vendors such as Talend, Microsoft SSIS and Pentaho.

Build or Buy

Melissa enables customers and ETL partners to easily incorporate data quality components such as Personator Consumer, a contact cleansing tool that corrects and standardizes all your U.S. and Canadian contact data pieces in real-time. It can also verify that the data inputs match (name matches the address), append missing pieces of data, and enrich the contact with demographics such as household income, occupation, school district, country FIPS information, date of birth and deceased information and much more.

Melissa also offers global address verification to clean, standardize, transliterate and verify deliverability of addresses in more than 240 countries. The combined solution brings agility in the data pipeline as data processing and data quality controls can be applied inflight, in the cloud or on-premise.

All this enables organizations to do more with their data projects without the overhead and risks associated with attempting to build rules-based data quality routines from scratch. Melissa taps into more than 38 years of contact data quality expertise to create out-of-the-box solutions that work seamlessly with your platforms. Our solutions leverage the most up-to-date, multi-sourced reference data so you get the most accurate view of your customer.

Data Quality Steps in Unison

For DBAs and data scientists who don't have a process in place yet and are looking for a user-friendly way to jumpstart their data management and data quality routines, Melissa also now offers another option—Unison. Unison is also ideal for users that don't have the time or resources to invest in one of the sophisticated ETL programs that incorporate data quality.

Unison is a new software solution that offers a centralized way to establish and maintain data quality at higher speeds—processing 50 million addresses per hour—while still meeting the most stringent security requirements. Unison offers advanced data quality capabilities in a single on-premise platform.

Unison is a lightweight, multiplatform solution for data quality and data management. It was designed to be easy-to-use so that anyone in the organization can create and deploy a data quality project. With features like automated scheduling, project sharing and reporting, Unison is the perfect fit to address any and all current and future data quality issues.

Security and Compatibility

Beyond ease-of-use, Unison offers two more important advantages—security safeguards and compatibility with popular data platforms.

With Unison, data never leaves an organization and only authorized users can access the application. Administrators can choose to integrate Unison within their company's LDAP system for pre-existing logins. Conversely, administrators can create account logins with role-based capabilities and configurable user rights through the Unison Authentication System.

Unison supports Oracle, SQL Server, MySQL and a variety of delimited flat files (more connectors are coming in the near future) to provide contact data verification and geocoding for U.S. and Canadian records. The product roadmap also includes plans to expand Unison's capabilities beyond contact data with additional data enrichment service for all types of data. ■

Melissa ■ 22382 Avenida Empresa ■ Rancho Santa Margarita, CA 92688-2112
Phone: 949-858-3000 ■ Fax: 949-589-5211 ■ Toll Free: 1-800-635-4772 ■ www.melissa.com

APPLICATIONS ▸▸

some form of dynamic scaling or what is often referred to as auto-scaling since load or demand is never constant.

In the world of data preparation, this is of critical importance. Not only are workloads complex with engines such as Spark, data ingest pipelines, data profiling, and storage, the prep also often requires expanding the compute power to manage all these processes and then contract again when the jobs are completed.

The emergence of Kubernetes (k8s), originally developed by Google, is now giving organizations the means to manage the deployment of these Docker containers and their scheduling as well as the ability to restart and scale out when needed.

It is the combination of Docker and Kubernetes that is creating a tipping point that will accelerate the industry toward a serverless world capable of powering business agility, lowering administration and management, and disrupting costs. The tipping point is caused by three factors:

1. With Kubernetes, an organization does not have to actively babysit its applications and make sure they are running, scaling, and coordinating. That heavy lifting is done by the framework. This changes traditional Ops in a fundamental way.
2. The framework is available on all major cloud providers. Some of them, such as Azure, have made their k8s offering (i.e., control plane) free. This means you only have to pay for the VMs (host nodes) you consume. The orchestrator is available at no cost.
3. Low cost of entry. If you can ship your application as a Docker image, then you are just a few steps away from making it available on k8s.

*It is the combination of **Docker and Kubernetes** that is creating a tipping point that will accelerate the industry toward a serverless world capable of powering business agility, lowering administration and management, and **disrupting costs**.*

This is of vital importance in the world of analytics and data science, and preparing the data for these critical business initiatives holds tremendous potential for all parties, including these:

- Supply-side: On the supply side, software developers are in a unique position to lower their cost by developing and supporting applications built in Docker containers. This approach is similar to large retail chains such as Safeway or Walmart that demand suppliers not only deliver their products but also take care of how they are displayed and put into shelves inside the retailer. Running complex applications like Spark and analytical engines are now easier to compose into a single isolated container that can run anywhere. Interestingly, many organizations are beginning to set standards to only work with software that is in Docker containers.
- Demand side: With Kubernetes as the orchestrator to deploy and manage these containers across very complex, hybrid cloud environments, customers now get the benefit of improved agility to deploy technologies in support of business initiatives. In addition, Kubernetes will enable auto scaling of workloads and ensure that these resources are available with high performance and availability. Today, several large sophisticated IT buyers are hesitating to even license applications that would not run on their k8s clusters. The buyer checklist in many evaluations now includes the question "Do you ship a deployment YAML [YAML Ain't Markup Language] file for your application along with the necessary images?" Customers want their applications to keep running as they expect in a lights-out scenario.

While the combination of Kubernetes and Docker delivers some great advances for software development and deployment in cloud environments, it is increasingly considered in on-premise settings. In fact, the LogicMonitor study indicated that by 2020, some 27% of workloads will remain on-premise.

This may be because organizations that have vast data center resources understand that these on-premise workloads are not going to go away quickly. In many countries and industries, security concerns remain a barrier to cloud entry. The adoption of Kubernetes orchestration layers on-premise promises to offer some of the same benefits that can be achieved in the cloud. And it opens the door for much easier shifting of workloads between on-premise and cloud should the need or opportunity arise. ■

Nenshad Bardoliwalla is co-founder and chief product officer, and **Pakshi Rajan** is senior director, product management, at Paxata (www.paxata.com), a provider of enterprise-grade self-service data preparation for analytics.

DBTA WEBINARS

DBTA hosts live webinars, with archived on-demand replays made available at www.dbta.com/webinars.

Crafting a Data Governance Strategy

Data governance works behind the scenes to deliver the trust and compliance every data citizen needs to help drive their business forward. But for too many organizations, data governance still just means control when it actually can open the door to a whole lot more.

DBTA recently held a webinar with Stan Christiaens, co-founder and CTO of data governance company Collibra, who discussed why industry leaders are changing their approach to data governance.

Digital disruption is occurring faster than ever, Christiaens explained. The goal of the organization should be to reach the needs of all data citizens. To close the gap between data and the creation of a solid data governance strategy, there are several key points to invest in, including analytics, visualization/dashboards, models, AI/machine learning, cataloging, data management, MDM/ETL, and data lakes, he said.

Structuring a data governance strategy requires a balance of offense and defense spanning finding, understanding, and trust, Christiaens said.

Collibra offers data engagement with governance, blending data experience with cataloging, governance, privacy, and data management. This allows data consumers to easily find the right data, quickly discover what the data means, trust the data because its entire context is known, and advance data privacy in a changing regulatory environment. The approach, he noted, is business user-driven, with collaboration between business and IT, and is adaptable across industries/processes.

Enterprise Must-Haves for a Modern DW

As the tech world continues to change, so, too, is data warehousing. The goal of modern data warehousing is not only to deliver insights faster to more users, but to provide a richer picture of your operations afforded by a greater volume and variety of data for analysis.

Increasingly, organizations are seeking to reduce the cost of legacy systems. They want flexibility to reconfigure, scale up, or scale down to meet changing requirements to enable real-time analytics at scale.

In a recent *DBTA* roundtable webcast, Kevin Petrie, senior director and technology evangelist, Attunity, and Gary Orenstein, chief product and marketing officer, Yellowbrick, discussed novel solutions to chronic data warehousing challenges.

Attunity offers modern data integration, and its Replicate solution provides automation and real-time data delivery and is universal, said Petrie. In addition, Attunity Compose accelerates the creation of analytics-ready data structures, and Attunity Enterprise Manager enables intelligent management, metadata, and control. The platform can improve efficiency and flexibility for real-time, advanced analytics.

For the modern data warehouse, enterprises must support always-on and available capabilities; ad hoc SQL; correct answers on any schema; terabytes-to-petabytes of data; mixed real-time inserts, ETL, batch, interactive workloads; and thousands of concurrent users. According to Orenstein, the Yellowbrick flash-based data warehouse, which is a modular built appliance, offers capacity that can hold tens of terabytes to petabytes of data and can give businesses all those qualities and more.

Monitoring the Performance of Database Servers With IDERA

It is essential to have access to detailed, real-time information on the performance of database servers and databases. But, database performance problems affect the responsiveness of the business-critical applications that they support. For this reason, it is important to use diagnostic tools to quickly locate and resolve performance issues in the early stages. Monitoring and improving the performance of database environments is much more straightforward with quality monitoring applications and the right alerts in place.

DBTA recently held a webinar with Patrick O'Halloran, senior sales engineer, IDERA, who discussed how to continuously monitor and analyze MySQL and MariaDB to ensure application responsiveness and database security. SQL Diagnostic Manager (DM) for MySQL is an IDERA tool that allows organizations to monitor Oracle MySQL and MariaDB databases, O'Halloran explained. It is a low-overhead, low maintenance tool that helps users pinpoint the instances, database, tables, and other activity in their environment that needs attention.

DM for MySQL is a Linux or Windows application that runs on a server. It is entirely self-contained and does not require any client agents, web servers, or any other software installed, O'Halloran said. It connects to MySQL instances across the network and uses low-overhead and well-documented commands to gather information about the sessions, activity, and SQL statements that MySQL is executing, as well as server configuration and various logs. The information gathered by DM for MySQL is stored in an internal SQLite database. ■

ROB MANDEVILLE

Rob Mandeville is a senior DBA at SolarWinds, experienced with major RDBMS vendor engines.

Why Is Tuning Hard?

There are many reasons we all get tuning wrong at times. Here is a list of some I can think of (but is by all means not exhaustive):

- We tune based on object definitions and how they relate to each other, not based on actual workload or how those objects are accessed.
- Once we tune, we set and forget—assuming all is good until someone complains again.
- We tune a query without regard for broader impact (such as hinting an index or creating an index)—things that can have negative impacts down the road.
- We tend to look at queries and indexes when the issue may be with the data model, optimizer, or statistics.
- We tune in development or test where datasets and workload don't represent production.
- We lack the skills to read a plan correctly to know what to tune.
- "Good" can be ambiguous—compared to what? Do we have a baseline? What is acceptable?

Tuning Headwinds

There are a lot of folks out there who make a living (many of them quite a good one) doing database performance tuning. Why? Tuning requires a high degree of knowledge and performance skill, is time-consuming, and means knowing the right diagnostics to collect when performance hits occur. Because of these and many other reasons, database performance experts thrive—and let's not forget about job security. No one that I have ever talked to has the privilege of working in a static environment. Things that are constantly changing include the following:

- New application releases or updates
- Database upgrades or patches
- OS upgrades or patches
- Migrating/motion to new physical nodes (if virtualized) with different workload, available resources, and resource consumption patterns
- Seasonal patterns occur (end of month, end of quarter, holidays, etc.)
- Data grows, statistics change (hopefully)
- User needs change (i.e., a transactional system is now needed for my real-time reports)
- Business needs change (i.e., BI and reporting)

Another headwind to tuning is this question: Why bother? As mentioned, as soon as we tune, things start changing immediately with the potential to invalidate any good work we've done. Also, the availability of resources is much improved. Hardware costs have come down significantly per unit while scalability has gone way up. In addition, the time to provision in a virtualized environment can now be measured in minutes or hours instead of weeks or months. So, regarding tuning, some take the stance of "Let's just feed the beast." And yes, that is a strategy—one that can mask a lot of inefficiencies in SQLs, indexes, and data models. I would like to go on record that I feel this is *not* a good strategy. Especially if any workload is running in or moving to the cloud where we trade green dollars for more resources.

Introducing Table Tuning

I'd like to propose a new way to look at optimizing. I'm going to call it "table tuning." Here's the reason: You don't index a query, you index a table. However, you index a table so that queries can access data more efficiently. When we do index the table, there are two significant considerations beyond the query improvement that we should be looking at:

- What is the cost of maintaining the index? Meaning with data churn (inserts, updates, and deletes), there will be a cost (overhead) for maintaining the index as data changes in the table.
- What other queries might be impacted? This could be either negatively or positively. An index could help more than one query to access required data more efficiently.

The benefit to this approach to tuning is that we can look for any inefficient data access against a table regardless of the reason (of which there can be many), meaning that we should look for low ROI queries—queries that are asking the database engine to do a lot of work when it really doesn't care about most of the data read. This will ensure that we find impactful, inefficient workloads. Some causes include:

- Missing indexes
- Data type conversions (either implicit or explicit)
- Functions in the WHERE clause
- Spools
- Lookups
- Stale statistics

Conclusion

In summary, tuning is difficult because inefficiencies can be caused by many things. At times, we need to get into the detail of what needs to be tuned by parsing the plan. It is practically impossible to look at everything—you need to have some system or method that highlights the most impactful, inefficient workload; aggregate that workload at the table level; and make better decisions with that higher viewpoint. And, because things are always changing, we need to continually watch for new things popping up. Easy, right? ■

GUY HARRISON

Guy Harrison, a software professional with more than 20 years of experience, is a partner at Toba Capital and the author of *Next Generation Databases* (Apress). Contact him at guy@tobacapital.com.

A Big Year Ahead for Ethereum

EMERGING TECHNOLOGIES

Interest in and enthusiasm for blockchain waxes and wanes with the market capitalization of cryptocurrencies. During 2018, the market value of all cryptocurrency assets dropped from more than $700 trillion to slightly more than $110 trillion, resulting in pessimism around cryptocurrencies and blockchain technologies.

Nevertheless, virtually all analyst firms, industry experts and technology authorities still believe that blockchain is a key innovation in computer science and expect blockchain to have a massive impact across multiple industries over the next 10 years. The same authorities also agree that the ultimate success of blockchain technology depends on significant advancements to overcome the limitations inherent in the current implementations. In particular, the environmental impact of the public blockchain is too high, and the throughput of the public blockchain is too low.

Bitcoin remains the most established and widely recognized blockchain and has a deserved reputation for being highly secure and stable. But this stability comes at a cost—innovations in the Bitcoin blockchain come very slowly. By contrast, Ethereum takes more of a "move fast and break things" philosophy. Consequently, it is in the Ethereum blockchain that we expect to see the more interesting innovations in 2019.

Ethereum already expanded on the Bitcoin concept by providing a Turing complete programming language that allowed for the creation of "smart contracts." Smart contracts in turn allowed for the creation of "tokens": special-purpose coins that can be created and managed on the Ethereum blockchain. These tokens powered the ICO (initial coin offering) craze of 2018. Although the ICO craze was in the main an irrational investment bubble, ICOs provided a glimpse of what might be possible from smart contract technology.

In 2019, two major innovations are expected in the Ethereum architecture: *Proof of Stake* and *Sharding*.

Proof of Stake is an alternative to the *Proof of Work* algorithm pioneered by Bitcoin and implemented in the existing Ethereum blockchain. In Proof of Work, "miners" or "validators" compete to add the next block on the blockchain by attempting to find a sort of magic number derived from a complex mathematical calculation. The first validator to solve the mathematical puzzle is rewarded with Bitcoin. The electricity consumed by the network attempting to find these magic numbers is massive: By some estimates, the Bitcoin network burns more electricity than Ireland.

Proof of Stake avoids the environmental cost and, theoretically, allows for a faster system. Instead of requiring each validator to prove that they have expended computing power, each validator is required to prove an economic "stake" in the network. Validators take turns in voting on the next block in the chain, and the size of the vote is proportional to the amount of cryptocurrency (in this case, Ether) that they have deposited (e.g., "staked"). Validators get a share of the profits involved in validating blocks, providing they vote for blocks that are accepted as valid by the rest of the network. Validators that attempt to process bad transactions can be penalized by losing their stake.

The core concepts of Proof of Stake are relatively easy to understand. But in practice there are many obstacles to overcome. While Proof of Work has proven to be resistant to network attacks, Proof of Stake is at least theoretically susceptible to a variety of vulnerabilities.

Vitalik Buterin, inventor of Ethereum, and his team have worked through these problems over the last 4 years and have unveiled an architecture named "Casper" in honor of the friendly ghost of the children's cartoon—that they believe can solve these problems. When unveiled this year, Casper will provide our first real-world test of Proof of Stake.

Casper will be implemented in 2019 in combination with the *sharding* scalability enhancement. If successful, these two innovations may catapult Ethereum into the position of the dominant public blockchain. At "stake" is not just the future of the Ethereum network, but the possibility of a viable, performant, and ecologically responsible blockchain. ■

Best Practices Series

The Future of Data Warehousing

For sponsorship details contact Stephen Faig, stephen@dbta.com, or 908-795-3702.

APR/MAY 2019

database TRENDS AND APPLICATIONS

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 https://blogs.sentryone.com/kevinkline.

Containers and the Wonderful World of Operating-System-Level Virtualization

SQL SERVER DRILL DOWN

If you haven't been paying a lot of attention, you might have missed the rise of a new form of virtualization, based at the operating system level, called "containers." While there are many specific software programs, by far the most prominent is Docker. So, from here on out, consider "Docker" and "container" to be synonymous.

Docker containers, as I mentioned, virtualize at the operating system level. Within Docker, you can include an entire application, tools, libraries, and configuration files. For example, you might build out a Docker container that "contains" a SQL Server database engine, IIS, SQL Server Reporting Services (SSRS), and all of the reports that run off of that container. Other examples you can check out right now are many of the machine learning images posted by Microsoft at https://gallery.azure.ai. You could send SQL and T-SQL commands to it similar to a regular hardware- or VM-based instance of SQL Server.

But because it doesn't need to instantiate the lower-level software the way a VM does, it is much smaller and a great deal faster to spin up or tear down. And, since it's so lightweight, you can run many more Docker nodes than you can VMs. In my own case, I don't run more than two VMs at a time on my laptop. But I've gone as high as six Docker containers, and with lower overhead, too!

Docker in Relation to Linux, Windows, and SQL Server

Docker is primarily a product for Linux, although it does run on Windows and MacOS. But this complicates things for the DBA who wants to run SQL Server in Docker containers. First, if you are willing and have the skills to run SQL Server on Linux, go for it. Docker on Linux is fully supported by Microsoft and has all the features you could want.

However, Docker for Windows (as released by Microsoft) has a couple of important limitations:

- First, and it's a big one, it lacks support for Active Directory and Windows Authentication. There are workarounds using managed service accounts, but they're somewhat of a kludge in my opinion. Read more at https://docs.microsoft.com/en-us/virtualization/windowscontainers/manage-containers/manage-serviceaccounts.
- No support for RDP or a GUI. It's pretty much PowerShell and command-line interfaces or nothing.
- Since Microsoft's implementation of Windows-based Docker containers is rather quirky, here are a couple other notes to be aware of:

- It puts more of the operating system footprint (based on Windows Nano Server) into the container, thus making most images greater than 10GB. That's a big drawback since a Linux-based container with the same functionality might weigh in at about 2GB.
- It must be rebuilt whenever a Windows update occurs.
- It breaks compatibility with several important helper apps for SQL Server, such as SQL Writer and Volume Shadow Copy Service.

While the Linux-based offering from Microsoft sees plenty of attention from development teams in Redmond, I haven't seen anything new for the Windows-based offering in at least a year. Based on those quibbles, I'd suggest if you want top-tier Docker support that you either get comfortable with Linux (and PowerShell) or you investigate alternative third-party tools. My personal recommendation for a third-party tool is Windocks. The Windocks team supports not only SQL Server images, but also recently added support for SSRS images as well as expanded database cloning from enterprise storage arrays. Check out https://Windocks.com for more information.

Kubernetes is not something you'll be using on existing IT systems. It is entirely designed to support hyper-scale greenfield micro-service architecture-based SaaS applications.

Orchestration—The Way to Deal With Lots of Docker Containers

Let's say that you're comfortably on your way forward with Docker containers, not only using them to run complete databases, applications, and micro-services but also automating their workflows en masse. How do you keep all of them working well together? Enter the "orchestration" category of software. Again, just as there are more suppliers of containers than Docker, there are multiple suppliers of orchestration software. But for the purposes of this article, I'm primarily talking about Kubernetes (frequently abbreviated as "k8s").

Kubernetes is a Google creation, and the company manages all of its worldwide data centers using Kubernetes in an elegant ballet of automation. I am not going to go into as much detail about Kubernetes, which means "helmsman" in Greek, as I did with Docker because it really is overkill for all but large estates. Furthermore, Kubernetes is not something you'll be using on existing IT systems. It is entirely designed to support hyper-scale greenfield micro-service architecture-based SaaS applications. Using Kubernetes, you can scale up your infrastructure to a vast number of nodes in distributed clusters to manage things such as load balancing, spinning up and down services as needed across the cluster.

Are You Using Docker?

Are you using SQL Server inside of Docker containers? Then I want to know more about your experiences! Feel free to contact me. ■

SETH MILLER

Seth Miller, senior principal software engineer with Veritas Technologies, has been working with Oracle technologies since 2005 and specializes in database administration, and solutions integration. He currently serves as the director of communications for the IOUG.

Can High-Tech Really Be Completely Automated?

Much consternation is expressed these days on conference calls and in convention center hallways among technology professionals who are worried that automation in high tech will push them out of their jobs. Taglines such as "Automatic upgrades, automatic patching, and self-tuning eliminate human labor" make that fear not completely unfounded. But in real-world scenarios, there is a big difference between Utopian visions of the humanless data center and a realistic view of automation in information technology.

Our team was recently working with a longtime Oracle customer on its Exadata quarterly patching. The amount of effort and time that Oracle customers put into preparation for each patching cycle is not insignificant. Similar to most Exadata customers, the client uses Platinum Services to implement its Exadata patches.

The latest patching implementation did not go well. There were some errors during the patching that the support technicians either missed or ignored, and subsequently had to be fixed by the customer after the patching was complete. The problems with patching from Platinum Services got so bad with the client that the organization decided to no longer use the services for its patching cycle.

During the conversation that led up to the client's decision to take over the quarterly patching, this question was posed: "Are the problems we are having caused by too much reliance on automation or humans?"

This may help to explain why there is so much focus on machine learning and AI. Scripts written with a narrowly defined set of rules will break and often exacerbate problems caused by anomalies if programming is not able to recognize those specific patterns. Without the ability for an automated process to "expect" something to go wrong, automation can become an enemy of productivity.

Our team recently deployed two different hyper-converged data protection solutions in Amazon Web Services (AWS). Both have a version of their products specifically designed for deployment in AWS and Microsoft Azure. A key benefit of hyper-convergence is that the software is highly integrated with and tuned to the platform on which it runs. The downside of the hyper-converged methodology becomes pretty clear when one or more of the assumptions of the platform isn't met.

The deployment of both products is highly automated and takes advantage of the service integration in both clouds. When we deployed the products, following as closely as possible the step-by-step instructions, the deployments broke somewhere in the middle. The lack of assumption that deployments may not run perfectly from start to finish in a foreign environment forced us to comb through logs to find the error.

We found the error and added a work-around that allowed us to finish that and many more successful deployments. A byproduct of our investigation into the logs revealed that the error we hit was not the only error in the deployment. In fact, there were myriad exceptions found that were mitigated automatically by the scripts. Moreover, the vendor had a process in place to capture the error and resolution, and quickly incorporate it into their next patch cycle.

This example would seem to make the case that automation may very well be reducing the need for humans to be involved in tasks executed by machines. However, rather than being eliminated completely, the need for human intervention has shifted, in this case to helping those machines learn what they did wrong so they can anticipate and resolve more anomalies the next time they execute the same task.

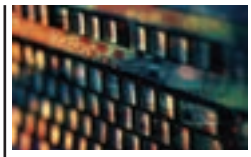
The next time your YouTube video is interrupted by an enterprise technology company telling you that humans are expensive and we have the solutions to get rid of them, take it with a grain of salt, but dismiss it at your peril. The last 100 years of history can teach us a lot about the evolution of technology and the impact it has on different industries. Disruptive technologies such as electricity and the automobile absolutely eliminated the need for human beings—whether they worked in gas lamp manufacturing, as coachmen for horse-drawn carriages, or in other narrowly defined areas automated by those products. Every example not only shifted humans to different jobs but also created new additional opportunities that hadn't existed before.

If this same effect can be expected from the automation of high tech, where do the humans go that are replaced by machines? For the savvy technology professional, the better question might be: What are those new opportunities that don't exist yet and how do I best position myself to take advantage of them? ■

CRAIG S. MULLINS

Craig S. Mullins is president of Mullins Consulting, Inc. He's an IBM Gold Consultant, IBM Champion for Analytics, and the author of three best-selling books, *DB2 Developer's Guide, Database Administration: The Complete Guide to DBA Practices & Procedures,* and *A Guide to Db2 Application Performance for Developers.*
Website: www.mullinsconsulting.com

What Is a DBA?



DBA CORNER

Every organization that manages data using a DBMS requires a database administration group to ensure the effective use and deployment of the company's databases. And since most modern organizations rely on database systems, they also need DBAs. That said, the discipline of database administration is not well-understood, nor is it universally practiced in a coherent and easily replicated manner.

Just walk around your organization and ask the question, "What does a DBA do?" You'll probably hear many different answers, such as, "Design databases," "Install the database system," "Monitor the system," "Fix problems," "Help with queries," "Optimize the database," "Keep things backed up," or even, "I don't know what they do all day!"

This is a sad—but too often true—commentary on the state of DBA in many organizations. Frequently, the DBA is viewed as a guru who uses tricks to make databases and systems operate efficiently. Database systems and applications are increasingly complex and very few people understand more than just the basics, such as SQL ... and even then probably not in any reasonable detail. DBAs understand the intricacies of the DBMS at all levels—or at least they should.

A frequent criticism of DBAs is that they can be difficult to deal with. Sometimes viewed as prima donnas, DBAs can be curmudgeons who have vast technical knowledge but limited people skills. Just about every programmer has their favorite DBA story. You know, those famous anecdotes that begin with, "I had a problem ..." and end with "and then he told me to shut up and read the manual." DBAs do not have a "warm and fuzzy" image. This probably has more to do with the nature and scope of the job than anything else. The DBMS spans the enterprise, effectively placing the DBA on call for the applications of the entire organization.

DBAs are frequently required to sit down and work things through on their own. Many database problems require periods of quiet reflection and analysis to resolve, so DBAs do not generally like to be disturbed. But even though many problems will require solitude, there are many other problems that require a whole team to resolve. And due to the vast knowledge most DBAs possess, their quiet time is usually less than quiet; constant interruptions to answer questions and solve problems is a daily fact of life. Adding to all of these issues is the fact that most organizations deploy multiple database systems, for example Oracle for some applications and Db2 for others—and frequently one DBA must become an expert on more than one DBMS.

Increasingly, organizations are running analytics and models on large sets of data, often stored in non-relational systems. The DBA is usually tasked with supporting the systems used by the data scientists (whether Hadoop or NoSQL) in addition to the operational database systems, further adding to the expertise required to be a good DBA circa 2019.

The DBA is at the center of all things related to data—development, analysis, protection, backup/recovery, and so on. The DBA is tasked with ensuring efficient, accurate access to the corporation's data. This means that DBAs frequently interface with many different types of people: technicians, programmers, auditors, end users, statisticians, customers, and executives. DBAs must ensure that they do not get so caught up in the minutiae of the inner-workings of the DBMS that they never develop the required people skills to interact with all their constituents. Good interpersonal skills are also required to be a good DBA.

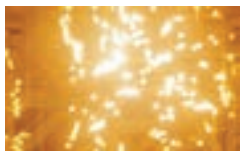
But we have not yet answered the question that is the title of this month's column: *What is a DBA?* The short answer to that question is this: a DBA is the information technician responsible for ensuring the ongoing operational functionality and efficiency of an organization's databases and the applications that access those databases. But the long answer to that question requires a book to answer. I have written such a book, titled *Database Administration: The Complete Guide to DBA Practices and Procedures*, 2nd edition, published by Addison Wesley (ISBN: 978-0321822949).

It can be an arduous task for a DBA to keep up-to-date and educated on all things data, but it is a requirement for doing the job well—so don't forget to keep reading this column and all of *Database Trends and Applications* for more details on data management and the discipline of database administration. ■

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.

When Is a Dimension a Fact?



In dimensional modeling, business events are typically designated as facts while descriptive information elements are dimensions. However, events (or information about them) occasionally serve as dimensions as well as facts. Metaphorically, this event information is somewhat similar to a gerund. Sometimes a verb is a verb, sometimes a verb can be used as a noun—as in the distinction between "gathering wood" versus being at "the gathering." A good data architect must watch their p's and q's and be certain when it is appropriate for a fact to also serve as a dimension—or when the dual function is not appropriate. Data modelers too often normalize dimensional components. Frequently, two distinct facts are associated into parent-child relationships, such as Order and Order Line—which is a normal design thing, not a dimensional design thing. It is only to be expected that subtler issues can prove even more confusing.

Sometimes, a business event may have both numeric elements and non-numeric elements; for example, Order will have a numeric Order-Total-Purchase-Amount, a non-numeric such as Order-Status-Code, and an Order-Number. The numbers to be summed are obviously grist for a fact table. The non-numeric items are supposed to be items for one or more dimensions. However, going through a two-step conversion process may suit everyone better. Conceptually, having an Order-Total-Purchase-Amount, an Order-Status-Code, and an Order-Number could be viewed as having both an Order Fact and an Order Dimension. The Order Fact will contain the Order-Total-Purchase-Amount, and the Order Dimension can have the Order-Number and the Order-Status-Code. Having these two separate objects at the logical level keeps the ideas distinct and manageable; this is the first step.

Step two involves making a big decision. Does the Order Dimension physically stand alone as an isolated table, or is there a good reason to merge it into the Order Fact as a degenerated dimension? The reason for this merging might be that any such Order Dimension is going to have the same number of rows as the Order Fact; therefore, having two large tables to join for virtually every query could result in negative performance. However, on the other side of the coin, a good reason to keep things separated might be justified by many updates of the Order-Status-Code. If the status changes often, it implies the need to update one's fact often, which is never a great idea. If one wishes to degenerate the Order Dimension and avoid updating the Order Fact, then one needs to think through the ETL processing necessary to maintain the table. A bookkeeping approach can be used to maintain the Order Fact. The bookkeeping approach would require that each time an order's status value is changed, then TWO rows are inserted into the Order Fact. One row is an exact negation of the first row; so, if the original row had a $100 Order-Total-Purchase-Amount value, the negation would be -$100. In this way these two rows, when added, would sum to $0. The second row added would have the new Order-Status-Code value and all the normal current values. If nothing had changed, the new Order-Total-Purchase-Amount value would be $100. When all three rows are added together, the total amount result is $100, the desired value.

The bookkeeping approach is effective in identifying changes over time; the downside is that it has more rows, and queries against the fact must always be summations. But on the positive side, as mentioned above, the fact is only ever inserted into, no updates are processed. The decision on whether to degenerate the Order Dimension is not a stylistic choice but an architectural decision that has great impact on the flow of data into the Order Fact. The data architect should always be involved in such decisions, carefully deliberating the options. ■

FEBRUARY/MARCH 2019 Ad Index

Collaborate ..... Cover 3
Kore Technologies 23
Melissa ..... Cover 2, 31
Revelation Software ..... Cover 4
Wisconsin-Madison 29

Best Practices Sponsors

Amazon Web Services 14
Couchbase 20
Delphix 19
MariaDB 17
Percona 16
Quest 21
SentryOne 18

COLLABORATE19

TECHNOLOGY AND APPLICATIONS FORUM
FOR THE ORACLE COMMUNITY

Register Today!

attendcollaborate.com

APRIL 7–11, 2019

SAN ANTONIO

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.
OpenInsight 10
SCALABLE
MORE DAZZLING
FASTER
SYNERGISTIC