Volume 34 Number 2
APRIL/MAY 2020

*The Journal of Information Integration and Management*

# database
## TRENDS AND APPLICATIONS

# The Next Great Frontier: Automating Data and Application Deployments

How Organizations Can Drive Better ROI Through DevOps Testing

Evolving Batch Processing to Keep Pace with Modern DevOps

The Role of the DBA in the Application Development Process

# GET THE ELEPHANT OUT OF THE ROOM

Bad address and contact data that prevents effective engagement with customers via postal mail, email and phone is the elephant in the room for many companies. Melissa's 30+ years of domain experience in address management, patented fuzzy matching and multisourced reference datasets power the global data quality tools you need to keep customer data clean, correct and current. Tell bad data to vamoose, skedaddle, and Get the El out for good!

## Data Quality APIs

- Global Address Verification
- Global Email
- Identity Verification
- Global Phone
- Geocoding
- Demographics/ Firmographics
- U.S. Property
- Matching/ Deduping

## Integrations

talend

Microsoft SQL Server

pentaho

salesforce

Microsoft Dynamics

X

Activate a Demo Account and Get a Free Pair of Elephant Socks!
i.Melissa.com/dbta

melissa

www.Melissa.com | 1-800-MELISSA

The Journal of Information Integration and Management
database
TRENDS AND APPLICATIONS

Volume 34 | Number 2
APRIL/MAY 2020

# 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

Dana J. Stevenson,
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**

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
Mike Flaherty,
CFO
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, 2020 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 $99.95 per year. Delivery outside North America is $153 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** **DEVELOPERS INCREASINGLY SEE DATABASES AS APPLICATIONS THAT NEED DEVOPS**
*BY JOE MCKENDRICK*

FEATURE STORY

**4** **THE NEXT GREAT FRONTIER: AUTOMATING DATA AND APPLICATION DEPLOYMENTS**
*BY JOE MCKENDRICK*

## DEPARTMENTS

TRENDS

**17** **HOW ORGANIZATIONS CAN DRIVE BETTER ROI THROUGH DEVOPS TESTING**
BY MALCOLM ISAACS

**18** **WHAT BUSINESSES NEED TO KNOW ABOUT DATA MIGRATION IN MERGERS AND CARVEOUTS**
BY DEREK OATS

APPLICATIONS

**20** **THE DECADE OF DATA**
BY MIKE LAMBERT

**21** **BATCH TO THE FUTURE: EVOLVING BATCH PROCESSING TO KEEP PACE WITH MODERN DEVOPS**
BY KELLY VOGT

DBTA LIVE

**22** IMPROVING DAILY TASKS FOR DBAS
**22** PREDICTING 2020 TRENDS FOR MODERN DATA ARCHITECTURE
**23** USING AN OBSERVABILITY PLATFORM TO UNDERSTAND MULTI-LAYERED ARCHITECTURES
**23** USING A TIME-SERIES DATABASE TO SEARCH THE SKIES
**23** VIRTUAL VOLUMES FOR ORACLE CAN TRANSFORM STORAGE

MULTIVALUE SOLUTIONS

**24** ROCKET SOFTWARE CREATES MV DEVELOPER'S EDITION FOR AWS
**24** KOURIER INTEGRATOR 4.7 INTRODUCES NEW PERFORMANCE ENHANCEMENTS
**25** INFORMER 5.1.2 RELEASE ENABLES QUERYING EFFICIENCY FOR DATASET MANAGEMENT
**25** JBASE 5.7.4 PLATFORM IS NOW AVAILABLE

## COLUMNS

**26** QUEST IOUG OBSERVATIONS > SHELBY KLINGERMAN
**ORACLE CLOUD'S NEW FREE TIER AND ALWAYS FREE ORACLE AUTONOMOUS DATABASE**

**28** NEXT-GEN DATA MANAGEMENT > BY TOM LAROCK
**MAXIMIZING THE VALUE OF DIGITAL DATA**

**29** EMERGING TECHNOLOGIES > BY GUY HARRISON
**ELASTOS AIMS FOR THE DECENTRALIZED INTERNET**

**30** DBA CORNER > BY CRAIG S. MULLINS
**THE ROLE OF THE DBA IN THE APPLICATION DEVELOPMENT PROCESS**

**31** SQL SERVER DRILL DOWN > BY KEVIN KLINE
**AZURE COSMOS DB ADDS FEATURES AND FREE TIER**

**32** DATABASE ELABORATIONS > BY TODD SCHRAML
**THINK OF SOLUTIONS AS INDEPENDENT ENTERPRISES**

MEDIA PARTNER OF THE FOLLOWING USER GROUPS

# Developers Increasingly See Databases as Applications That Need DevOps

By Joe McKendrick

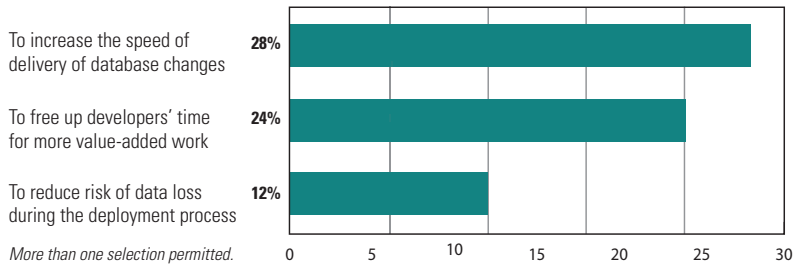
When it comes to DevOps, developers increasingly recognize databases to be code sets that require ongoing integration and deployment. They are "another code deployment which can and should be managed, tested, automated, and improved with the same robust, reliable methodologies applied to application code," according to the authors of a recent survey of 2,000 developers.

The survey, published by Redgate Software, uncovered a rise in the number of DevOps deployments for databases, with almost half of respondents saying they deploy database changes to production weekly or more frequently.

The survey report emphasized the importance of including databases in DevOps, which represents "the union of people, process, and products to enable continuous delivery of value to our end users. The most important word of this definition is value. To deliver value you must deploy all changes, including those to your database."

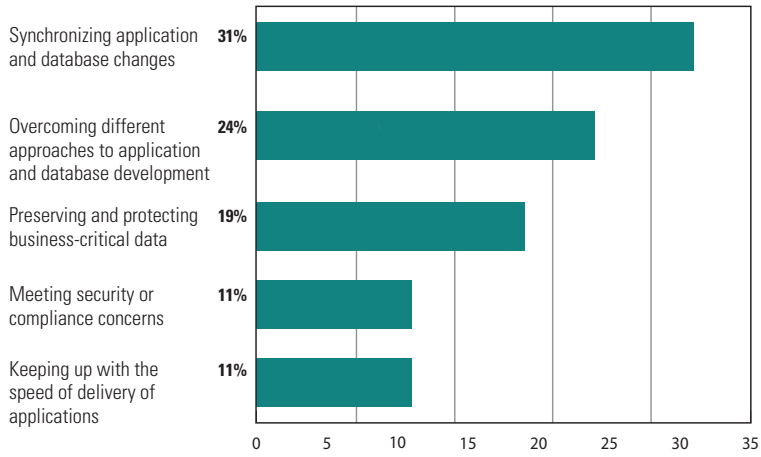
Changes in database technology require a shift in mindset in the way they are managed and deployed within enterprises. "Databases, which were formerly seen as configurable repositories much like file-shares, are now seen as stateful data, con-

**Figure 1: Drivers for Adopting Database DevOps**




*Source: "The 2020 State of Database DevOps" from Redgate Software (www.red-gate.com)*

**Figure 2: Challenges Integrating Database Changes into a DevOps Process**




*Source: "The 2020 State of Database DevOps" from Redgate Software (www.red-gate.com)*

trolled by code. This evolving viewpoint is transformative; changes to databases are recognized in another code deployment which can and should be managed, tested, automated, and improved with the same robust, reliable methodologies applied to application code."

Database DevOps differs from DataOps, which focuses on automating the flow of data through enterprises. DevOps—which concentrates on the flow of application delivery and deployment—will help in ensuring the capabilities and qualities of databases, which essentially are very large, sophisticated applications.

More than two-third of respondents, 68%, are in some phase of DevOps adoption, and another 18% said they have adopted DevOps "across all projects."

There is also significant movement toward Database DevOps. Forty-six percent of respondents are performing some form of database deployment automation. A majority, 60%, believe they will be moving from traditional database deployments to fully automated deployments in a year or less.

***Changes** in **database technology** **require** a **shift** in **mindset** in the way they are **managed** and **deployed** within **enterprises**.*

***Database DevOps** differs from **DataOps**, which **focuses** on **automating** the flow of **data** through **enterprises**.*

Slow release cycles are seen as the main impediment to database development, as cited by 24% of respondents. Another 21% cite the inflexibility of current database environments—an inability to respond quickly to changing business requirements. Accordingly, the main reason developers apply DevOps to databases is to increase the delivery of database changes, the survey found. Developers also see opportunity to free up their time from more mundane database support tasks (see Figure 1).

Bringing databases into the DevOps flow is not without its obstacles. The leading roadblock to successful Database DevOps consists of synchronizing application and database changes, cited by 31% of respondents, and overcoming different approaches to application and database development, as cited by 24% (see Figure 2). Another 18% of respondents expressed concern about increased risk of failed deployments or downtime when introducing changes as an issue with database deployments.

The survey also revealed convergence between developer and DBA roles. Developers in the survey consider databases to be part of their job responsibilities. In the survey, 78% reported they are responsible for both application and database deployment. And database deployments keep increasing in frequency—49% of respondents said they need to deploy database changes weekly or on-demand, up from 37% in a similar survey in 2017.

Database DevOps is delivering results, respondents confirmed. Those who reported that all or nearly all their database deployments take place within DevOps say they have seen lower lead time for changes and lower defect rates. Thirty-seven percent of respondents embracing Database DevOps have seen virtually no code defects which require hotfixes. ■

# The Next Great Frontier: Automating Data and Application Deployments

*By Joe McKendrick*

DevOps, DataOps, AI, and containers all lead to one important innovation for enterprises seeking to be more data-driven—and that is greater automation. Data-driven enterprises cannot function if data resources and applications are in any way being manually administered, deployed, remediated, or upgraded.

The ability to move fast, make decisions in real time, and respond quickly to events requires automated processes for ingesting and managing data. Organizations that fail to effectively leverage and deploy their data assets will find themselves falling behind. Data managers are turning to automation and autonomous databases and platforms, a recent survey of 217 data managers by Unisphere Research, a division of Information Today, Inc., found. According to the research, three in four DBAs feel that applications can be deployed faster with increased database management automation, and seven in 10 expect increased database automation to boost the impact of their roles ("2019 IOUG Autonomous Database Adoption Survey").

Already, database functions such as backup and recovery are highly automated, and plans are underway to automate such day-to-day functions as monitoring, provisioning, and maintenance. Data managers welcome the advance of automation of these tasks and see greater roles for themselves in higher-level business decision making.

## AUTOMATION CHALLENGES

Moving to a high-functioning, automated enterprise is not without its challenges—both at organizational and technological levels. But corporate culture may be the most significant obstacle, industry observers agree. "Many people and teams who do not come from an automation mentality struggle to appreciate the value," said Ben Klang, vice president of business technology at Power Home Remodeling. "This is fundamentally different than manual

processes," Klang said. Enterprise teams have been deployed for a long time, and processes often come from a previous generation. Many safeguards and sign-offs are in place to enforce the wisdom of lessons learned the hard way and to help prevent future mistakes, he noted. "It's easy to believe that those steps are still necessary, though they come into conflict with the goals of automation."

IT departments themselves have historically grown up in cultures emphasizing risk avoidance, added Miha Kralj, managing director of cloud strategy, architecture, and delivery at Accenture. "Data centers, hardware, and software licenses—the foundation of IT until the cloud—were all expensive. To avoid any risk of an investment failing, companies instituted rigorous processes to analyze and justify these investments before making them."

Ben Newton, director of operations analytics at Sumo Logic, agreed that culture change is the greatest inhibitor to enterprise automation, suggesting that many organizations may have difficulty adopting newer technology trends such as microservices and DevOps, which are all "built with small-team, agile structures in mind. Applying those ideas to older, siloed organizations is a recipe for failure and disappointment. Enterprises that want to compete with modern technology giants entering their markets must be willing to rearchitect their organizations, not just their tech stack."

Organizational and data structures both play a role in slowing the drive to automation. "It's challenging for companies to deliver real-time insights and analytics if their infrastructure can't accommodate the data being captured," said Rajan Kohli, president of Wipro Digital. "Data-hungry applications require a growing number of pipelines, which have caused data silos to become disconnected from other pipelines, datasets, and producers."

Anders Wallgren, vice president of technology strategy at CloudBees, went further, saying databases themselves "are one of the last holdouts in the digital transformation." He noted that he frequently sees "organizations that are quite broadly automated and reap the benefits of that automation but still grind to a halt if a push involves schema or database changes, which are still handled manually by DBAs."

The sheer amount of data flowing into and through enterprises will also create challenges for the move to automation. "It's no longer just active transactional data," said Dheeraj Remella, chief product officer of VoltDB. "In the quest to understand their customers, internal processes, supply chain, employees, and systems, telemetry collection points are being deployed. Organizations need to act on the incoming data almost immediately. How do you incorporate machine learning into event stream processing in a continuous manner? Where do you make these decisions—in the cloud or near the edge?" Organizations cannot fall into the trap of assembling too many technologies in the interest of a loosely coupled architecture and give up on the primary business objectives of low latency, accurate decision making driving actions and responses to events, he noted.

*The **ability** to move fast, make **decisions** in **real time**, and **respond quickly** to events requires **automated processes** for **ingesting** and **managing data**.*

## SKILLS AND DATA LITERACY

Klang also pointed to the skill challenges needed to make automation work. "The skills needed for automation are most familiar to those with a software development background," he pointed out. "But the things that need automating are often the domain of non-developer experts, such as DBAs or server administrators. High-functioning automation requires expertise from both."

Along with skills, data literacy is another challenge that needs to be tackled if automation is to succeed. "To apply business context effectively and address changing requirements, you need to ensure that everyone working with this data has also achieved a level of data literacy around data matching and can access that information," said Andrew Lee, vice president of emerging technologies and incubation at Syncsort. "What you need to do next is supplement these foundational skills with the right vocabulary and context so people can confidently appraise the data they're using. The strategies and algorithms used, including criteria to support or reject a match, need to be clear and understood."

## DEVOPS AND DATAOPS TO THE RESCUE, MAYBE

DevOps and DataOps are seen by many as the most effective organizational paths to automating data-driven applications. "The pressure to realize the greatest value from one of their most valuable assets, their corporate data, at an ever-faster rate is driving teams to look to DataOps for improved quality and reduced cycle times," said Gaurav Rishi, head of products at Kasten.

The speed at which developers are containerizing applications to allow for rapid provisioning, continuous integration/continuous delivery automation, monitoring, auto-scaling, and self-healing have made Ops teams "unbelievably more efficient," said Monte Zweben, CEO of Splice Machine. Five or 10 years ago, many of these tasks were manual, error-prone, time-consuming, and expensive—resulting in much lower availability, he said. DevOps and DataOps "make an automated flow of applications and data much more achievable."

These practices need to be baked into corporate culture and processes, however. "You cannot buy DevOps. You must do DevOps," said Rajesh Raheja, senior vice president of engineering for Boomi. In addition, different teams have differ-

ent approaches based on the school of thought and how much of a purist they want to be—such as adopting Scrum versus Kanban. Moreover, some companies have DevOps teams separate from development and operations. "Other teams have Ops take on the system reliability engineer role. Others still have SRE [site reliability engineering] operate independently with a more narrowed charter."

Wallgren also sees mixed results from working deployments. "Whether teams practice DevOps true to form or in name only varies widely," he said. "The most cringe-inducing comment I've heard recently is, 'We don't do DevOps, but we have a team that does that for us.' If you have a DevOps team that may mean you've merely introduced another organizational silo into the mix, which isn't going to improve things very much. Keep in mind that DevOps isn't something you are, it's something you do."

There's often a risk of DevOps being adopted for the sake of DevOps. "In some cases, it's adopted almost religiously, which leads to the adherence to a philosophy without any considerations of implications," said Remella. Data is only as useful as the the user's understanding, according to Remella. For example, Remella said, if one thinks that everything can be done in the application development and infrastructure components—such as databases should only be the responsibility of the operations team and there is no need to develop in the database—the potential result is that the infrastructure gets inefficiently used while not meeting scale and latency SLAs.

Ajay Gandhi, vice president and digital performance evangelist at Dynatrace, sees AI and automation within DevOps as the first key to mastering IT complexity in order to intelligently automate cloud-native software delivery and cloud operations. "However, just any generic AI alone is not enough. It is crucial to implement purpose-built AI that can understand the full context of IT environments, including dependencies between containers, microservices, and applications."

## CONTAINERS AND MICROSERVICES

Containers and microservices are helping enterprises achieve an automated flow of applications and data. "For applications, containers and microservices have standardized the distribution and delivery of code, which has helped simplify DevOps," said Raheja. "The intersection of DevOps, containerization, and microservices has increased cloud adoption and accelerated the building of modern applications. For data, the configuration required can be treated as code, and the environments can be quickly spun up and down for data pipeline analysis and training. To that extent, containerization does help enterprises achieve some level of automated data flow."

***Moving** to a **high-functioning**, **automated enterprise** is not without its **challenges**—both at **organizational** and **technological levels**.*

These approaches are important to increasing confidence. "Microservices break monolithic applications into decoupled, more easily understood pieces that decrease the automation effort by dramatically reducing the what-could-go-wrong surface area," said Zweben. "Containerization helps automation in a similar way, but at the infrastructure level," he said. In addition, containers abstract away the underlying infrastructure, allowing automation to bypass Byzantine platform configurations.

At the same time, underlying data needs to be a component of microservices and container environments. "The challenge for data management isn't the containerization perspective, but the data itself," Raheja cautioned. "Getting data tested in various environments is challenging, but discovering unknown data and cataloging it is an even bigger challenge."

As with DevOps, many enterprises seek to adopt containers and microservices, but execution is still poor. "For example, having the mindset of only going through APIs instead of databases creates a 'just-get-it-done' mentality, which then leads to rigid implementations that are hard to change and difficult to understand," said Robert Castles, principal and chief technology officer for PMG. "Ideas and concepts like exposing back-end data to end users is challenging to users in a B2B context. For example, is the enterprise data searchable? Can it scale to end-user expectations for speed of queries? What are the performance standards of the target system? Is it reliable enough to count on, and how will unavailability impact the application?"

## ADVICE ON AUTOMATION

To achieve a more automated, data-driven environment, data managers need to think about the way data and applications are managed in profound ways. The following measures will help companies on their journey to automated, data-driven processes.

**Move incrementally.** "Each step in data analytics should be evaluated to determine how automation and intelligence could improve the process," said Kohli. "In many enterprises, data managers should seek to implement DataOps for greater agility in their delivery of applications and data. When implementing new processes like DataOps, data managers must align relevant stakeholders to ensure streamlined operation across the enterprise—from engineers and analysts to data scientists and executives. Analysts and developers should also receive a steady stream of feedback so they can move quickly to identify and implement corrective measures."

**Decouple applications from data.** "Work with application teams to, as much as possible, decouple changes to the data from changes to the application," said Wallgren. "If you're able to push data changes ahead of application changes, you can reduce risk

significantly. Of course, this doesn't come free, but the effort pays off as you can often avoid having to roll back data changes if it's necessary to roll back application changes. Organizations that have done the work to decouple their data from their applications are in a better position. For example, architecting an application so you can roll out a schema change before the application code that uses that change reduces quite a bit of the risk involved."

**Leverage the power of AI.** "With the right AI in place, organizations can begin to automatically pinpoint the root cause of any performance problems and take precise steps to remediate them," said Gandhi. "Not only does this level of intelligence help address issues of complexity, but it also lays a foundation upon which organizations can build an automated flow of applications and data across their DevOps toolchain, from continuous delivery pipelines to operations."

**Measure progress.** Progress should be measured in terms of how much you and your team are finishing, said Kevin Kline, principal program manager at SentryOne. "Lots of managers measure how much activity is happening across their teams, but not how much is finished, delivered, and approved by the customers that they service," explained Kline. "They are measuring how busy they are but not how productive they are. To achieve agility, many managers need to change how they are measuring the success of their teams. Once they make that change, other team improvements will quickly follow by detecting and remediating those issues that are holding them back from finishing projects that they've started."

**Encourage collaboration and teamwork.** "The success of adopting automation or containerization depends on the tight working relationship between infrastructure and development professionals," said Klang. "Neither team can unilaterally design a successful system. Teams should share common goals around availability and speed to delivery, and have a strong agreement about how resources (compute/network/storage) will be made available." For data managers, the worst-case scenario is building a system that nobody wants to use. For developers, they lack the access or authority to deploy systems themselves, which can lead them to look externally for solutions, said Klang. "Working together is the only way for both sides to be successful."

*The **sheer amount** of **data flowing** into and through **enterprises** will **create challenges** for the move to **automation**.*

**Reevaluate how teams are structured.** "While IT, developer, and operations teams have traditionally worked in silos, and while these teams have often remained disconnected from the business, today's IT complexity, coupled with customers' demands for amazing digital experiences, requires an entirely new approach," said Gandhi. "It's critical to empower all teams to clearly see how their work is aligned with, and supports, other teams and—most importantly—how their work ties back to business goals." This can be done by sharing insights on IT performance and customer experience that all teams can share, using a common language and common view into how user experience and performance are impacting business outcomes, Gandhi explained.

*Along with **skills**, **data literacy** is another **challenge** that needs to be tackled if **automation** is to **succeed**.*

**Boost data quality.** "Going from on-premise solutions to cloud-based platforms, for example, changes neither the data utilized nor the data produced," said Lee. "If you deliver garbage data to your cloud applications, you still have the same data quality problems as before—and, in the era of AI and machine learning—the consequences of poor data quality can be even more severe. That's because there's a multiplier effect with machine learning—first in the historical data used to train the predictive model, and second in the new data used by that model to make future decisions. Establishing stringent data quality standards and practices that are sustainable and can be adapted for the future is a crucial first step. Data managers need to efficiently cleanse, transform, and enrich data in a scalable manner for accurate analytics. This can be a challenge, especially for large sets of data. Enterprises should look to tools to make the process faster and more efficient."

**Apply culture change to ideas about data and data platforms.** Once corporations have embraced the idea of smaller, cross-functional, agile teams, those teams need to be working from a common understanding of the business and their customers, Newton stated. "That common understanding is expressed in a common data platform. Containerization in particular and microservices in general are generating too much complex data to allow the old silos to exist. A shared data pipeline and data platform will allow enterprises to unleash their data-hungry teams and encourage collaboration."

The greatest wins in automation "come from rethinking the problem being solved," agreed Klang. Just as virtualization replaced metal servers, and containers are replacing virtualization, the organization and the people both have to be part of the change, or it is almost certain to fail, Klang observed. ■

CData
Page 12
DATA PIPELINES AND THE FUTURE OF DATA WAREHOUSING

Vertica
Page 14
THE FUTURE OF DATA WAREHOUSING: VAST, FAST, AND OPEN

database
TRENDS AND APPLICATIONS

# THE FUTURE OF DATA WAREHOUSING

Best Practices Series

database
TRENDS AND APPLICATIONS

# HOW TO PREPARE FOR THE FUTURE OF DATA WAREHOUSING

Best Practices Series

What is the future of data warehousing? This has been a long-running question within the industry, with occasional declarations of the death of the traditional, archival data warehouse. But as each new technology wave washes over, the data warehouse model remains vital by adapting to new realities and business requirements.

A recent survey of 1,388 data scientists, executives, and IT professionals by *Business Application Research Center (BARC)*, in fact, places data warehouse modernization among the top 10 priorities this year in data analytics and BI investments ("Top Business Intelligence Trends 2020: What 2,865 BI Professionals Really Think").

How will data warehouse modernization play out? Data warehousing is evolving in the following ways, by becoming:

**Cloud-enabled:** Perhaps the greatest transformation seen in the data warehousing space—a transformation still underway—is the move to the cloud. Data warehouses are notoriously huge, requiring significant amounts of storage and processing capacity. Today's data warehouses are residing in the cloud. This is usually positioned as data warehouse as a service (DWaaS). A cloud provider—such as Amazon Web Services or Microsoft Azure—provides the infrastructure and storage capacity required for data warehousing. In addition, DWaaS is being delivered by dedicated data warehouse vendors, who have shifted their business models from hardware and software to on-demand delivery.

**Software-defined:** As they move into the cloud, fewer data warehouses will be defined as large pieces of hardware residing in a corporate data center. Acquiring, upgrading, and expanding the data warehouse will be reduced to clicks on a cloud provider's dashboard.

**Virtualized:** With remaining on-premise warehouses, as well as off-premise DWaaS services, an organization

*Perhaps the greatest transformation seen in the data warehousing space—a transformation still underway—is the move to the cloud.*

may have multiple warehouses within its workspaces. This calls for a virtualization layer that can integrate these various environments or services into a single view.

**Analytical powerhouses:** Today's organizations seek advanced data analytics, AI, and machine learning capabilities that extend well beyond the power of existing infrastructures. The emerging generation of cloud-based data warehouses have backend resources that enable almost unlimited scaling. As any data scientist knows, AI and machine learning algorithms are compute-resource hogs, and also have insatiable appetites for datasets, for training as well as production loads. Data warehouses have always been built to scale, but there's only so much room in on-prem data centers. The data warehouses of today—and tomorrow—take advantage of the massive resources maintained by cloud providers.

**Container-enhanced:** The ability to deploy a range of applications and update them as necessary calls for a high degree of flexibility. With containerized applications—and containerized data—instances can be spun up on demand.

**Automated:** Increasingly, advanced automation and AI are taking on the manual aspects of data warehouses. A new generation of tools is enabling the automation of the previously manually intensive aspects of data warehouses, from ETL scripts to data lifestyle management and security. Data warehouse vendors would point out that a single DBA could manage a warehouse—and now it may only be a small part of a DBA's time.

**Streaming:** More data than ever moves through enterprises on a streaming basis. This may reshape the purpose of data warehouses, which typically have supported on-prem, relational data. While the conventional wisdom may dictate that streaming services will replace data warehouses, it's more likely that the data warehouse will continue to adapt to accommodate the integration of streaming data with pertinent archived data to deliver enhanced real-time insights. Watch this space.

**Agile:** The proliferation of DevOps practices will encompass data warehousing platforms as well, and updating and boosting data warehouse capabilities will be part of a continuous improvement process. Along with DevOps, DataOps will automate the continual delivery of data through the pipelines that lead to and away from the data warehouse. In addition, data warehouses will be backended by serverless technologies in which cloud providers manage all the technical and tuning issues associated with running the environment. Organizations will only need to be concerned with deploying the analytics applications that run on top of the warehouse.

**Part of a multi-tiered data architecture:** Data warehouses will coexist with data lakes, and are well-positioned to serve as readily transformed environments for the wide array of data formats and structures that end up in the data lake. The combination of these data environments will be part of successful business data strategies in the emerging digital economy. In addition, with the proliferation of various databases, as well as emerging disruptive models such as blockchain, the data warehouse of the future will be highly distributed.

## A NEW PURPOSE

Now, in the era of data analytics and AI, data warehouses have gained a new purpose, serving as repositories of well-vetted and standardized datasets essential to any data-driven business. The result is that for the foreseeable future the data warehouse isn't going anywhere. If anything, it's more vital than ever in a world that demands timely, quality data. ■

*Today's organizations seek advanced data analytics, AI, and machine learning capabilities that extend well beyond the power of existing infrastructures.*

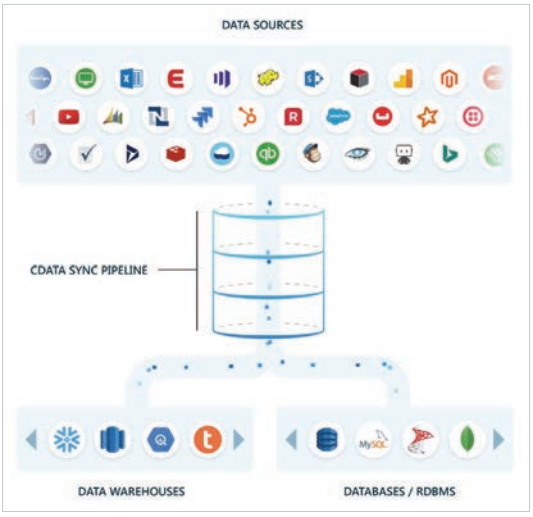
—Joe McKendrick

# Data Pipelines and the Future of Data Warehousing

The popularity of data warehousing continues to surge as tools for business intelligence, artificial intelligence, and machine learning help organizations realize new business value from their enterprise data. Advances in data warehousing and subsequent tooling are enabling businesses to perform data operations on a massive scale.

The benefits are obvious, but aggregating all of your organizational data into a common repository can be a challenge. Your organization is likely using a variety of applications, platforms, and services for everything from payroll and accounting to marketing automation and CRM — each with their own unique APIs or interfaces.

The process of aggregating data has traditionally centered on bulk/batch data movement and ETL processing. However, as the volume and velocity of data increase, conventional ETL processes start to break down. What's worse, these processes often fail silently, resulting in painful troubleshooting and data re-engineering.




## THE MODERN DATA PIPELINE

To get a complete picture from their data, organizations need a solution to bring that data where they want it. In recent years, IT teams and organizations across the world have turned to modern data pipelines to solve these challenges.

A data pipeline is a tool that enables a set of data movement workflows to extract data from one or more data sources and land that data in some form of database or data warehouse. The modern data pipeline includes broad support for data source & destination connectivity, automation, schema flexibility, and data transformation capabilities.

## BROAD-SPECTRUM DATA CONNECTIVITY

The typical enterprise leverages dozens of SaaS applications and disparate data sources in a sprawling, hybrid cloud/on-premise mix. Data pipelines deliver seamless connectivity from all of these mission-critical data sources where organizations generate, gather, analyze, and store data to a central data warehouse for processing.

## AUTOMATION & INCREMENTAL UPDATES

Given the exponential growth in both the volume of data and the applications which organizations require, these data pipelines must provide for automated, efficient processes. Solutions must support the quick setup of scheduled replications that can run in the background AND keep piping data into the repository, without constant troubleshooting of brittle scripts. An incremental approach, which delivers regular imports of new data, will ensure the data pipeline keeps flowing, and the data warehouse provides useful up-to-date information.

## FLEXIBILITY

The modern data pipeline must be flexible enough to accommodate all kinds of data. Applications and systems are continually evolving, and modern data pipelines must be able to handle schema changes without breaking automated data aggregation processes.

## DATA TRANSFORMATION

While there is incredible diversity in the data types and formats that organizations need to process, often the database or data warehouse they use to power their analytics only supports limited schemas. Therefore, the modern data pipeline needs to support both traditional ETL processing as well as rapid ELT-driven transformation, leveraging the processing capabilities of the underlying data warehouse, if available.

## ON-PREMISE OR CLOUD

Security and regulatory compliance can impact which applications or services should have access to data, both in transit and at rest. For the broadest use, data pipelines should support deployment across all established IT environments, including cloud, hybrid cloud, and on-premise.

## A UNIVERSAL PIPELINE FOR DATA WAREHOUSING

At CData Software, we specialize in data connectivity and have developed a data pipeline platform with the diverse needs of enterprise data warehousing in mind.

CData Sync is a flexible, go-anywhere data pipeline that can automate the data ingestion of all your critical data into any backend database or data warehouse..

With a straightforward, point-and-click interface and support for hundreds of SaaS, Big Data, and NoSQL data sources, CData Sync allows you to rapidly configure and customize your data pipeline to fit any data warehousing scenario.

To learn more, visit us online at www.cdata.com/sync.

# The future of data warehousing: Vast, fast, and open

VERTICA

Data warehousing, and the power of analytics that makes all that data storage worthwhile, has become an essential technology for businesses to stay ahead of their competitors. The volume of data that businesses need to collect and store – especially via object storage – is vastly greater than what organizations 20 years ago imagined was possible, or useful. But it's precisely those vast data stores that make machine learning and the promise of AI finally possible, and beneficial.

Consider the sources of today's data. Text, transactional data, videos, geospatial, machine, and sensor data, and more comprise many terabytes of information generated every minute. And that's often for a single organization. S3 or Hadoop data lakes have made storage of all that data relatively affordable. The question is, how to unify various data stores designed for various purposes? Then, how to analyze it all efficiently and cost effectively?

This article will describe how the Vertica Analytics Platform addresses these needs, and more.

## SPEED TO HANDLE TODAY'S MASSIVE DATA VOLUMES

What does it take for a modern business to become data-driven? Handling massive data volumes is one thing; processing that data at blazing speed is another. An analytics platform needs to be fast enough to analyze data rapidly. There are patterns and trends in the data that can lead to new business opportunities, and opportunities don't wait for slow analysis.

The Vertica Analytics Platform enables organizations to go where the data leads them, without waiting hours or days for the results. Vertica is uniquely designed using a memory-and-disk balanced distributed compressed columnar paradigm, which makes it exponentially faster than traditional row-store and other methods, without requiring expensive hardware. It manages data at exabyte scale, and enables an organization to rapidly connect data siloes across multiple cloud and hybrid (cloud + on-premises) environments.

## OPERATIONALIZING MACHINE LEARNING – AT SCALE

Not long ago, data science was limited by the inability to base machine learning models on full data volumes run at extreme scale. More data means better accuracy of predictive models, and because Vertica's in-database approach to machine learning is undaunted by vast data sets, machine learning finally lives up to its promise.

Vertica's in-database machine learning helps analytics teams become productive right out of the box, by addressing every step of the machine learning process in SQL. Plus, Vertica can import models built and trained in other platforms and languages – like TensorFlow, Spark, Python, and SPSS – via the PMML (Predictive Model Markup Language) format.

## ELASTIC SCALABILITY AND OPERATIONAL EFFICIENCY

Some data analytic workloads are predictable, as in manufacturing, where a set number of data inputs need routine analysis to determine time-to-failure, for example. But increasingly, organizations need big data analysis on an *ad hoc* basis, for data science projects or seasonal analysis of, say, consumer activity and buying habits. This sort of variable workload requires an analytics system that can easily scale up and down according to demand.

Vertica in Eon Mode enables teams to scale data storage capacity and the number of nodes independently. As the need scales upward, teams can increase concurrency – number of users accessing compute resources simultaneously – and rapidly expand the number of queries a cluster can handle, all by simply adding more nodes. When those projects are done, teams can spin down the added power, thus saving money, especially when the added resources are billable by their cloud provider.

Businesses can also isolate different workloads via "subclusters," which allow projects to scale data storage capacity and the number of nodes independently, without disruption to other workloads in play.

## OPENING UP TO DATA INFRASTRUCTURE INVESTMENTS

Most organizations embarking on big data analytics already have deployments in the cloud, on-premises, or in Hadoop data lakes where information is being stored. But which cloud provider? What kind of on-prem resources?

Vertica supports major cloud vendors, with more flexibility – including Google Cloud, AWS, and Azure. And it offers a highly flexible array of on-prem capabilities for communal storage, including Apache Hadoop HDFS, MinIO, and Pure Storage FlashBlade technology.

## EMBRACE THE VAST, FAST, AND OPEN

Here's what makes a smart big data analytics investment: A database that can handle data at Exabyte scale, spanning multiple data resources and storage modes; one that can process data at lightning speed, no matter the size of data sets; and one that works with the most number of popular data stores, whether in the cloud or on-premises.

Take a look at Vertica, and see how it compares with other vendors on all three of these criteria.

Learn more about the Vertica Analytics Platform at https://www.vertica.com

And find out more about Vertica release 10 at https://www.vertica.com/blog/vertica-version-10-launches-today/

# The 5 V's of Big Data Warehousing ...

Volume

Variety

Velocity

Veracity

Value

*... need the 6th V to handle them all:*

iri.com/voracity
info@iri.com
www.iri.com
1-321-777-8889

IRI Voracity
An Insatiable Appetite for Data

GET SOCIAL!
Connect with *Database Trends and Applications* online and get industry news, trends, and analysis, plus information on learning opportunities in the field.
**Join our networks today.**
dbta.com/social-media
database
TRENDS AND APPLICATIONS

**dbta.com/subscribe**

BRINGING YOU THE WORLD OF DATA MANAGEMENT

# READY TO MAKE DATA WORK FOR YOU?

## SUBSCRIBE TODAY!

**FOR MORE THAN 20 YEARS,** ***Database Trends and Applications*** **magazine has covered the technologies and processes involved in every aspect of the creation, management, application, and storage of structured and unstructured data.** ***DBTA*****'s content is original, factual, and uniquely valuable—providing clarity, perspective, and objectivity in an increasingly complex and exciting world, where data assets hold the key to organizational competitiveness.**

# How Organizations Can Drive Better ROI Through DevOps Testing

*By Malcolm Isaacs*

The hype around DevOps and its potential to drive greater ROI across a wide range of enterprise operations increased substantially in the last decade. However, as these expectations carry into 2020, organizations will start to take a more sober approach to DevOps implementations. While DevOps was initially seen as a widespread solution to all sorts of enterprise IT issues, the implementation of DevOps approaches is now shaping up to become more strategic and focused, with much of the emphasis on how to maximize the ultimate return on investment.

This can be a difficult task for organizations as return on investment (ROI) is not always clear when it comes to DevOps. This is due to the lack of usable data, making feedback difficult to assess. For organizations looking to drive greater ROI of DevOps implementations, the focus should be on efficient, high-quality, and secure delivery, which is largely enabled through the implementation of automated testing. Through automated testing, organizations can obtain actionable data to analyze and derive insights that will ultimately be used to improve their DevOps implementation.

## Delivering Results Quickly

As systems become more complex, the number of tasks that require testing is only going to continue to grow. Speedy and reliable delivery is too often hampered by manual testing, which can be time-consuming and, in some cases, susceptible to mistakes. Although manual testing is inevitable, automation offers a way to alleviate some of the problems that arise with it. Once a test has been automated, it can be run quickly and consistently every time a code change is made, increasing returns and improving the end product.

Although automation has vast benefits, this alone doesn't mean you can take a hands-off approach to testing. There are several issues that still need to be considered to ensure the best outcome. These include how much time the testing systems are idle, reducing the time it takes to provision testing environments how long it takes to recover from a failure, and what needs to be done to reduce the downtime lost while fixing that failure.

Because updates are—by the nature of DevOps—continuously introduced into production, businesses can track the impact a given update is having on their customers and point to the specific reasons affecting any given production chain. These insights can prove highly valuable and should be incorporated into future business strategies. In order to maximize ROI, executives need to look closely at this data and incorporate the insights gleaned from it into long-term business operations.

## Measuring Success

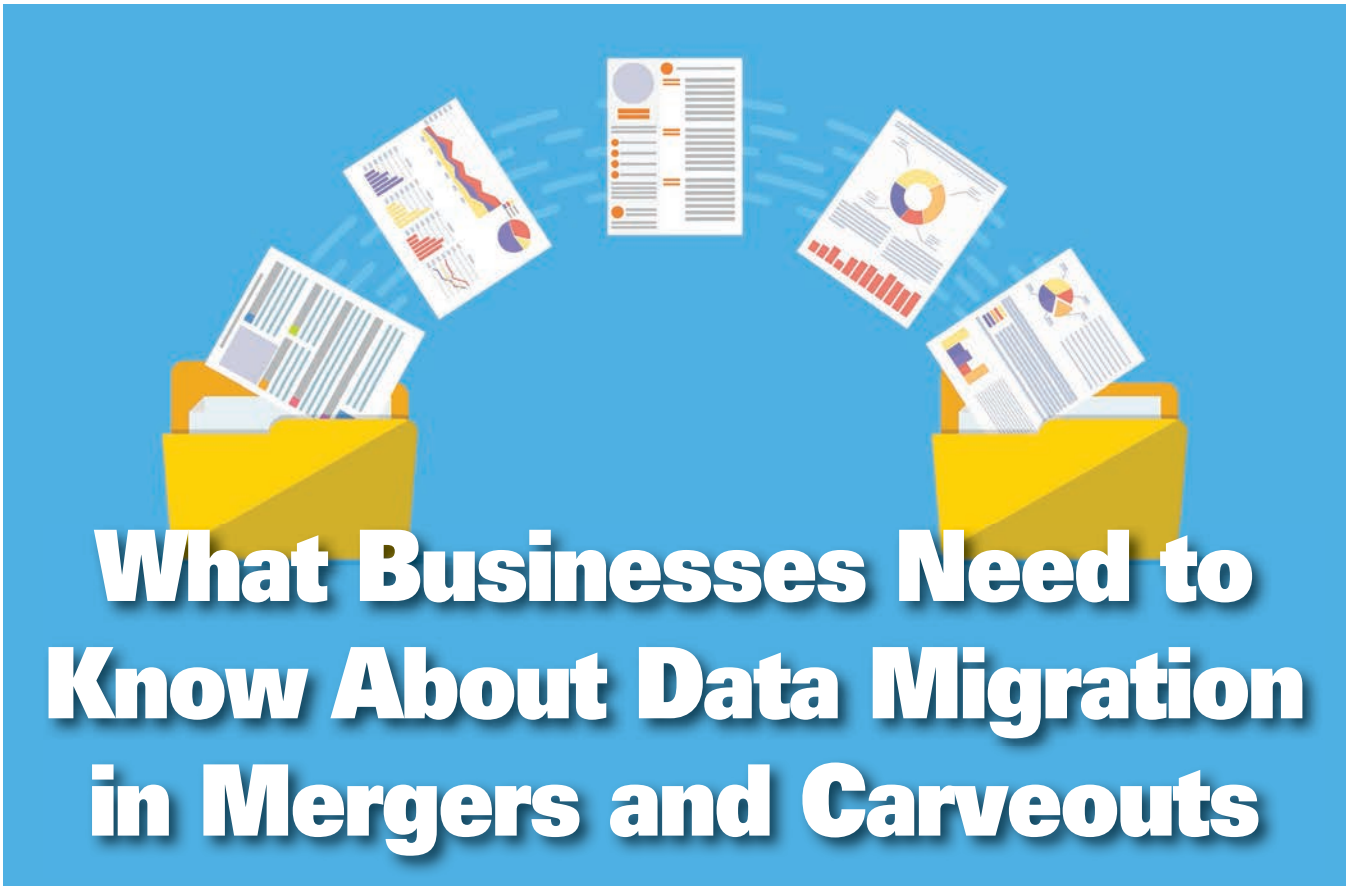
Just because DevOps has been adopted for software delivery, it doesn't mean that the value it generates is ultimately worth the cost. In order to make DevOps worthy of the cost and to avoid wasting countless hours of development, organizations should keep end users involved throughout the process. It's important for organizations to not simply get something up and running and then forget about it, but to continuously monitor the results, push for feedback, and seek out areas of improvement. Staying in tune with the needs of end users will ensure that no features go unnoticed, and the potential for maximum ROI is recognized.

Measuring success comes down to looking at a few critical aspects of the testing process. One valuable approach is to focus testing efforts on the areas of software that see the heaviest usage. Ensuring consistency and quality can go a long way to improving results. Another consideration is how many defects are going unnoticed throughout production and how those are affecting the end result. By utilizing this sort of information, developers can begin to identify problems earlier in the production process and stop them before any serious damage is done.

## A New Approach to DevOps

As enterprises get deeper into 2020, leaders will continue to look to DevOps as a means to improve operations and increase overall return on investment. And, as the hype around DevOps continues, the focus will shift toward the most important aspects of the technology—specifically, the adoption of automated testing as a means to improve not only speed of delivery but also the quality of outcomes. This kind of testing will not only improve operations but will give businesses a tangible method of measuring their success with quantifiable data and metrics. ■

**Malcolm Isaacs** is head of ADM solutions marketing at Micro Focus (www.microfocus.com).



# What Businesses Need to Know About Data Migration in Mergers and Carveouts

*By Derek Oats*

With $3.6 trillion in mergers and acquisitions completed in 2019 alone, M&A activity has been booming. However, a merger or acquisition isn't just a business decision and a business process. It's also a massive undertaking on the IT side, as you figure out how to migrate and integrate business applications and business data.

There's a lot of risk involved in M&A transactions, with many opportunities for error along the way. One of the biggest risks, of course, is that the post-M&A entity doesn't accomplish what the merger was designed to do. Instead, unsuccessful data migration and lack of harmonization may mean that you end up with additional technical debt. If the task of separating the data in a complex divested system wasn't evaluated correctly, the carveout project is also likely to require additional effort, leading to dreaded TSA (transition service agreement) penalties. TSAs are a significant driver in M&A projects—the terms need to be emphasized prominently in the planning and scoping.

Setting clear M&A goals, creating alignment between the business and the IT implementation, and intertwining those processes can help you find clarity and success throughout, while avoiding costly pitfalls. Here's what you need to know about data integration and how to start the M&A process.

> ***M&A projects** are incredibly complex by **nature**, and **executing data migration** is often the most **daunting aspect**.*

## Create a Clear Plan of Action

M&A projects are incredibly complex by nature, and executing data migration is often the most daunting aspect. In order to figure out what you're going to do with your data, design a plan that takes you through the M&A process from start to finish, and visualize the ideal end goal. Keep the business case for the acquisition in mind as a guiding principle as you assess the technical challenges and opportunities of the deal.

To help guide you as you design the M&A process, consider questions about your company's data needs. To start, do you need to integrate the company you just bought into your system? Or, if you sold off a business unit, do you need to separate data for the divested units from your current system? Does the company you're acquiring have the same ERP system? If not, what ERP system is it using? What regulatory issues might come up

that could delay, lengthen, or even halt the process? Think about how quickly the data migration needs to be completed, as well as your maximum budget for the project.

Answering these questions will help clarify your data needs and project goals so you can create a customized plan of action for your M&A. By creating a flexible strategy that takes into account the framework and business goals of your M&A, you'll give the process the highest chance of success and maintain a competitive edge. Otherwise, you may end up wasting time and resources by completing initiatives that don't fulfill your end goals.

### IT and Business Alignment

All of that strategizing could go to waste, however, without cooperation between your IT department and other departments, such as the office of the CFO and corporate development. Companies that align their business strategy and IT implementation see the most success from M&A transformations due to the greater understanding gained and the flexibility enabled.

To that end, the IT department should be involved in the transformation project from the starting line. Unfortunately, that only happens in the rarest of cases, and usually IT is only involved within the narrow scope of carrying out due diligence. But, that's a big mistake. When you wait to involve IT until the ball is already rolling, it has less time to prepare for the project. And, importantly, IT may not necessarily have all the information it needs to plan the project road map in its entirety. This can lead to solutions that, while quick, aren't properly thought through. Quick solutions also may lead to challenges that could have been avoided.

Of course, in order to be useful throughout the planning process, the IT department needs to understand business goals and be familiar with business processes. This deep level of insight into the business gives IT a superpower: the ability to implement necessary changes with more flexibility and less friction.

Flexibility is critical to any successful transformation project; there are bound to be unexpected challenges and twists in the road. But, having the flexibility to adapt to those changes and adjust IT processes or decisions can be the deciding factor between a well-executed M&A project and a project ridden with delays and additional costs, such as TSA penalties. That's why alignment of business and IT—particularly, the early involvement of IT—is a major factor for the project's success.

***Data integration*** *involves many* ***labor-intensive*** *and* ***repetitive tasks****, and each decision you make has a far-reaching impact across the* ***preservation*** *of history, master data* ***relevance****,* ***security****, and* ***regulatory*** *compliance.*

### The Role of Automation

Data integration involves many labor-intensive, repetitive tasks, and each decision you make has a far-reaching impact across the preservation of history, master data relevance, security, and regulatory compliance. The tedious processes that plague most carveouts can be automated, ensuring fast, secure results, with the added benefit of automated audit trails.

Automation plays an important role in the M&A process, enabling IT to support even the most challenging business carveout plans. It also helps speed up post-merger integration projects, providing a level of insight and control over the process that can't be achieved with traditional approaches.

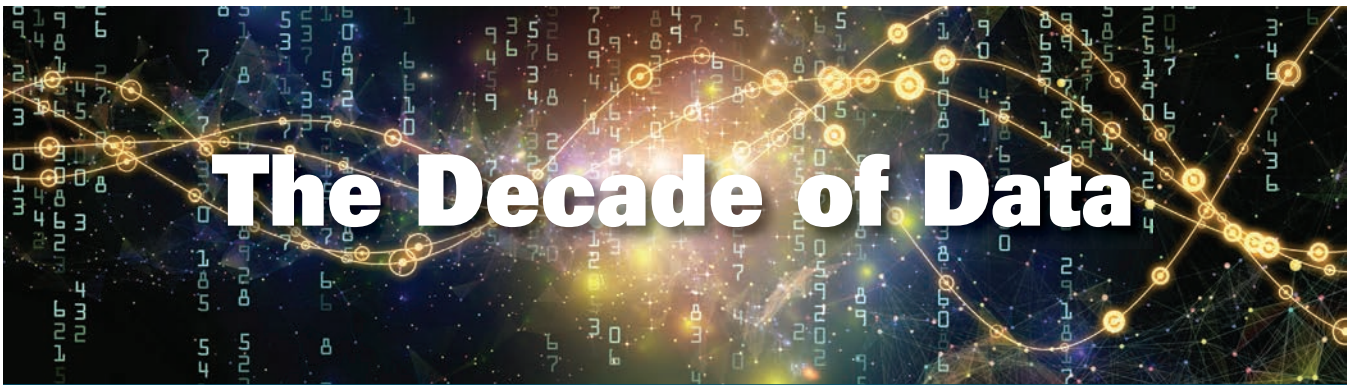
### Executing With Minimal Business Disruption

Now that you have a plan in place that incorporates automation, think about what you need to go live. Can your business handle a disruption lasting longer than a weekend? Or do you need to execute as quickly as possible in order to avoid interrupting business operations? Can you afford to take 9 months to complete the data integration project? Or does it really need to be finalized even faster, say, within a 3- to 6-month window?

For example, Jones Lang LaSalle (JLL), a financial and professional services firm specializing in commercial real estate services and investment management, needed to integrate large sets of financial data from significant acquisitions. To add to the complexity, it also needed to convert the data onto its ERP system in a way that would ensure the accessibility and availability of historical and master data. The massive historical data conversion needed to be completed almost as soon as the acquisition was finalized, so JLL could provide uninterrupted service to new clients. By partnering with an automation-driven transformation company, JLL successfully completed the project in just 3 months without interrupting business operations. But regardless of the timeline, companies need a software-based, predictable process for M&A.

Here's one way to think about it: If you put your house on the market, you'd want the selling and moving process to be as smooth as possible so you could get into your new home quickly and without any problems. That's why it's so important to close the gap between business decisions on the CEO/CFO side and IT execution on the CTO side when you start the M&A process. ■

**Derek Oats** is CEO of SNP Group (www.snpgroup.com).



# The Decade of Data

*By Mike Lambert*

To get a full appreciation for the incredible pace of change in business technology, look at the past 6 years.

In 2014, IDC published a report that said that, by 2020, the digital universe would contain nearly as many digital bits as there are stars in the universe, and the data we create and copy annually would reach 44 zettabytes, or 44 trillion gigabytes. Guess what? It's 2020. And it turns out IDC was correct in assuming that we were about to endure a data deluge.

## Hybrid Default

Today, data is similar to a snowball going down a never-ending hill. There are 2.5 quintillion bytes of data created daily by consumers and professionals globally at our current pace. Think about it: We are all producing data—both in our personal lives and in the business world—on the internet, social media, through texts, emails, digital photos—and don't forget all of the data generated by the Internet of Things (IoT). Ninety percent of the data in the world today was generated just in the last few years.

A more recent IDC report has predicted that by 2025, worldwide data will grow 61% to 175 zettabytes, with as much of the data residing in the cloud as in data centers. Because of this, industry experts are predicting that in the 2020s, traditional on-premise deployment of servers and applications—not cloud—will require justification. This complete changeover from the trends we've seen to date will introduce the new world of the "Hybrid Default."

In many ways, data has always been taken for granted, and the folks who managed it—the DBAs—were doing all of the work behind the scenes. Today, data is quickly taking on increasing importance in the business realm, and it's even more apparent what a critical role DBAs play in an organization. They are—and will increasingly be—the unsung heroes in every company, especially if the data growth predictions hold true. In many organizations, data is the new currency, and DBAs hold the keys to the vault. Their importance to the organizations they service cannot be underestimated.

## Getting 'Cloud Ready'

The 2020s certainly are shaping up to be the decade of data. CIOs' big strategy will be making sure they can get their data ready to move the cloud. But what does "ready for the cloud" really mean?

Many enterprises are in such a rush to move as much of their business as possible to the cloud, they overlook the importance of data workload readiness. In increasingly complex DataOps environments, IT and data managers need to be able to deliver business-critical data without disruption and to mitigate the risk of distributed denial of service (DDoS) and similar attacks that impact cloud services. Smart enterprises are methodically migrating data workloads to the cloud, starting with less mission-critical workloads, and making sure they optimize their data for cloud applications.

We have seen the introduction and adoption of DevOps and DataOps, the new roles, processes, and technologies to enable the velocity of development and release cycles for data-centric apps, especially in the cloud. In this decade of data, we will see DataOps mature within the organization to assure value is extracted from growing data assets.

## A New Era of Cooperation

As the decade progresses and organizations continue relying on multi-cloud and hybrid environments to drive their business, we'll see CIOs take an increasing interest, and assume a larger role, in assuring that, cloud-based data gets to the business users who need it. Data will no longer be the "second-class citizen" it's traditionally been. Business and IT will work hand in glove and jointly develop a clear vision and strategy for their cloud-based workloads. As IT becomes more agile with the data it is charged with managing, protecting, and optimizing, it will enable organizations to accelerate the continuous improvements to those services and applications that users want and need.

The 1920s were about people on the move—cars for the every person and new roads. Among the many byproducts of these advancements were much greater speed and agility in personal and business travel. Similarly, the 2020s will be about data on the move—from on-prem to the cloud. The most successful enterprise CIOs will work with their IT and data management teams to figure out how to seamlessly move data to the cloud, securely and without disruption, giving business users faster and more agile access to data where and when they need it. Hang onto your hats—this ride is going to be an exciting one! ■

**Mike Lambert** is the president of field operations for SentryOne (www.sentryone.com), where he oversees revenue growth for the company through the management of sales, channel, support, customer success, and marketing.

# Batch to the Future:
## Evolving Batch Processing to Keep Pace With Modern DevOps

*By Kelly Vogt*

We live in a business world that craves instantaneous response times, especially in terms of the responsiveness of modern transactional applications. Customers expect to make a money transfer, pay a bill, or buy a product online in a matter of seconds. This insatiable desire for *speed to customer* parallels the need to increase *speed to market* for these modern apps. With development activity accelerating through DevOps, do we ever stop to consider the increased demands this is putting on batch processing?

### What Is Batch Processing?

Batch processing simply refers to the processing of transactions in a group or batch. This is different from real-time transaction processing, which happens one transaction at a time. Consider the example of an ecommerce purchase—behind the customer's real-time transaction is a series of ancillary tasks associated with orders being placed, which might not happen right away.

These tasks may include inventory replenishment that may not occur immediately, especially for products that are purchased early in the day and/or sell by the tens or hundreds of thousands daily. Other examples include order boxing and shipping, which benefit from assembling orders. There are numerous reasons—often related to cost and warehouse efficiencies—that these processes occur in bulk as part of batch processing jobs once orders reach a certain threshold.

Batch processing is the key to many of these supporting back-end processes. But at precisely the time the proliferation of modern apps is placing greater demands on batch processing, the industry is also facing a retiring workforce of batch operators. It is imperative for batch processing to modernize alongside the real-time software products that are emerging from DevOps. The characteristics of batch modernization include the following considerations:

**Automation:** Manual system operation of batch processing must evolve to automation via policy-based controls. (Organizations can assess their current operation using the Mainframe Batch Maturity Model at https://resources.compuware.com/batch-maturity_level_lp). SLAs should dictate job initiation and prioritize CPU access—in other words, job queues should be automatically organized such that the most urgent job is the next job to initiate. Prioritized access to datasets and readiness of job resource requirements available for processing should also be automatic, all helping to prevent unnecessary time delays.

**Cost Awareness:** The costs associated with larger and more frequent batch workloads can send monthly license charges (MLC) skyrocketing if organizations aren't careful. MLC costs are based on Peak 4-Hour Rolling Average (4HRA)—the highest utilization determined from the sum of the Peak 4HRAs for all logical partitions (LPARs) in which a particular product runs. The key to reining in these costs is to execute jobs in a cost-aware manner that deliberately reduces the Peak 4HRA. So, in addition to factoring which jobs are the most time-sensitive, job execution must be automatically manipulated in a manner that exploits savings opportunities while sustaining maximum throughput.

**Continuous Improvement:** Once job execution is automated and operational costs are minimized without detriment to SLAs, operators have an opportunity to focus on continuous performance management and improvement. Urgent and critical jobs should be monitored and compared with historical performance levels. Any deviations from normal wall clock time or excessive resource consumption should stimulate automation to capture symptom data and ticket the incident. This creates a proactive opportunity for operators to make tweaks that help optimize performance of the system while also avoiding errors and failed jobs that could put the business at risk.

**Customer Service-Focused:** When batch processing is automated and optimized, from both a cost and job performance perspective, an improved level of customer service emerges as the natural byproduct. Per the ecommerce example, jobs are orchestrated in a manner that leads to satisfied customers through products that are readily available and delivered on time and cost-effectively.

### The Evolution of Batch Processing

Customer demand to engage with services and products through mobile apps and the web, whenever and wherever they wish, is dramatically accelerating the speed of software development and delivery. This increased activity on the front end ultimately triggers more frequent, more compute-intensive, and more critical and complex mainframe batch processing work on the back end, including inventory management, warehouse and shipping management, customer billing, and more.

A batch approach is often optimal for these scenarios, which means batch can no longer be simply dismissed as old-school overnight processing that can be ignored or deprioritized. As mainframe-based batch processing continues to grow in volume and complexity, it must evolve. Organizations that fail to recognize this as part and parcel of their DevOps approaches are ultimately putting at jeopardy their ability to compete.

**Kelly Vogt** is a solution consultant at Compuware (www.compuware.com).

*A list of upcoming and archived webinars is available at www.dbta.com/Webinars.*

## Improving Daily Tasks for DBAs

DBAs juggle multiple responsibilities all day, and that load is only getting heavier.

*DBTA* recently held a webinar with Chris Paap, solutions marketing manager, Nutanix, and David Teague, technical marketing engineer, Nutanix, who discussed a typical day in the life of a DBA and the daily challenges they conquer with the Nutanix Database Solution.

In this webinar, Paap and Teague explained how to tackle zero-day database vulnerabilities without throwing a wrench in the entire day or week, and how to increase database backup frequency without affecting performance and while using less storage than traditional methods.

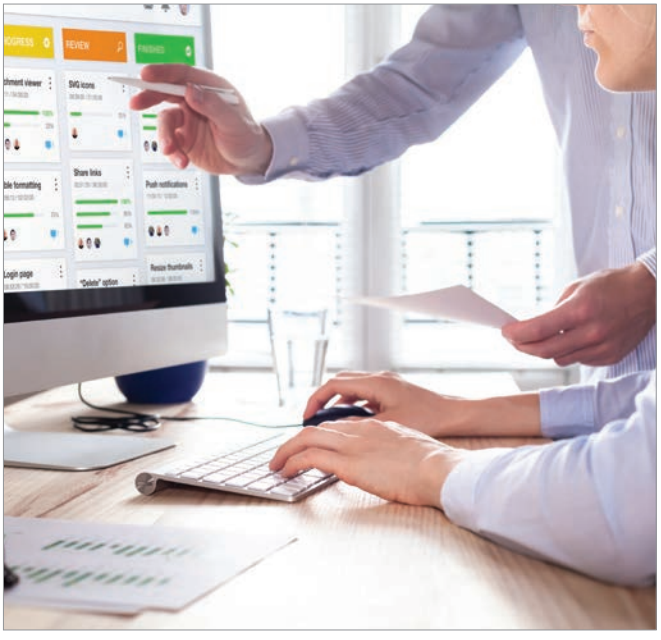
DBAs deal with a variety of challenges on a daily basis. These include ensuring database security and availability; dealing with daily business demands across production, development, and testing; building database instances and provisioning; and performance management and troubleshooting.

The complicated database provisioning process can take hours, days, or weeks, Paap and Teague explained. It also involves multiple teams and introduces friction through a complex process. Concerns around vendor lock-in, staffing challenges, and evolving pricing models are forcing companies to evaluate alternatives.

Nutanix can deliver database and application high availability along with one-click database operations, Paap and Teague said. With Nutanix, they noted, DBAs can deliver value for all databases because the platform offers enhanced time-to-value and business agility, maximizes uptime and security, and enables operational efficiency and reduced costs.

## Predicting 2020 Trends for Modern Data Architecture

From AI and machine learning to data discovery and real-time analytics, a strong data architecture strategy is critical to supporting an organization's data-driven goals.



Greater speed, flexibility, and scalability are common wish-list items, alongside smarter data governance and security capabilities.

During a recent *DBTA* roundtable webinar, Jeff Bayntun, manager, customer success, at Simba Technologies; Suphatra Rufo, principal product marketing manager at Couchbase; and Ali LeClerc, director of product marketing at Alluxio, discussed the top trends in modern data architecture for 2020.

Right now, there is market fragmentation because there are so many different types of databases and database offerings to choose from, Bayntun explained. Companies rely on many databases, with different databases used to solve different problems. This leads to an accumulation of data that is "locked" away. There are also limitations with the SQL ecosystem, and there is often ad hoc use of tools supporting non-SQL access, Bayntun said.

He suggested using the Simba platform to connect all data within the enterprise. The platform now has tooling and integrations with Apache Hive, Spark, Hadoop distributions, Databricks, and more. Simba partnered with Couchbase in 2014 to build the first Couchbase ODBC/JDBC drivers. Simba also builds the official DataStax, Cosmos DB, and Dynamo DB ODBC/JDBC drivers.

Rufo cited the evolving growth of the cloud as a big factor in upcoming trends this year. According to Rufo, the industry is seeing the death of monolithic legacy database systems, which are expensive, proprietary, and can result in vendor lock-in.

Couchbase can offer what customers want, she said. Couchbase is a NoSQL database that supports interactions online, in store, or on-the-go. Moreover, Couchbase Mobile extends Couchbase Server to the edge with an embedded NoSQL database (Couchbase Lite) and a web gateway (Sync Gateway).

LeClerc said people are moving from mainframes to big data platforms. Companies are moving from tightly integrated to loosely integrated architectures. Cloud and the expanding influx of data are changing the marketplace.

Alluxio's goal is to orchestrate data for the cloud for data-driven apps such as big data analytics, machine learning, and AI. The platform can accelerate modern app frameworks running on Hadoop Distributed File System/Amazon Simple Storage Services-based data lakes or warehouses, LeClerc explained.

## Using an Observability Platform to Understand Multi-Layered Architectures

Easy-to-understand monoliths are giving way to distributed systems: microservices, serverless, meshes, proxies, and every possible combination. These systems offer developers the freedom to build new features and technology faster, as they are no longer beholden to the elaborate release processes associated with monolithic architecture. But, as with all good things, there is a price: Distributed systems are inherently difficult to operate and maintain.

*DBTA* recently presented a webinar featuring Austin Parker, principal developer advocate of LightStep, who discussed how observability can help developers understand multi-layered architectures.

According to Parker, traditional tools weren't built for deep systems. Users may be scrolling through dashboards that don't reveal root causes of problems or spend time sifting through logs for an expensive needle in a haystack.

What these organizations need is visibility across services and dependencies, said Parker.

The right observability solution should provide the following elements:

- Make it easy to see and understand your highly distributed environment
- Help to immediately pinpoint performance regression
- Enable developers to ship with confidence
- Enable easy sharing of information
- Provide context
- Enable insights

LightStep offers an observability platform with context, said Parker. The platform includes the following capabilities:

- Automatic deployment and regression detection
- System and service diagrams
- Real-time and historical root cause analysis
- Correlations
- Custom alerting
- Easy setup with no vendor lock-in
- No cardinality limitations

## Using a Time Series Database to Search the Skies

The Vera C. Rubin Observatory, currently under construction in Chile, will conduct a vast astronomical survey of the universe starting in 2022.

The astronomical observatory plans to collect 500PB of image data by studying the skies continuously for 10 years and producing nearly instant alerts for objects that change in position or brightness every night. In addition to astronomical data, its dataset will include DevOps, IoT, and real-time monitoring of data.

*DBTA* recently held a webinar featuring Angelo Fausti, software engineer, Vera C. Rubin Observatory, who discussed how a time series database has the versatility to address the observatory's needs.

Because relational databases are not optimized for time series data, Fausti explained, the observatory chose the InfluxDB time series relational database to time-stamp data and analyze it.

The platform can ping scientists when there is an update via Slack, which keeps the observatory abreast of any developing situations, according to Fausti.

InfluxDB is a complete solution, he said. It offers the following elements:

- A time series database, a language similar to SQL, and an HTTP API
- A Chronograf user interface for time series visualization written in Go and React.js
- A Kapacitor framework written in Go for processing, monitoring, and alerting on time series data

"Astronomers working on telescopes have to monitor different things, depending on what their assignments are. Chronograf gives us the flexibility to create dashboards that suit our needs in different situations," said Tiago Ribeiro, scheduler scientist at the Vera C. Rubin Observatory.

In addition, the scientists were able to have sensors publishing information in many streams and from multiple physical locations, according to Fausti. Kapacitor has been critical in sounding the alarm when the system is not behaving correctly.

## Virtual Volumes for Oracle Can Transform Storage

VMware Virtual Volumes has gained increasing adoption from storage vendors since its release. Virtual Volumes can help provide many distinct capabilities for virtualized Tier 1 business-critical Oracle workloads that have not been available with traditional storage. Running business-critical databases on software-defined data center requires the use of policy-based storage leveraging Virtual Volumes capabilities.

*DBTA* recently held a webinar with Sudhir Balasubramanian, staff solution architect—Oracle technology lead at VMware, and Cody Hosterman, technical director, VMware Solutions, at Pure Storage, who discussed VMware Virtual Volumes and how it can help with Oracle deployments.

Virtual Volumes for Oracle can allow users to achieve scale, simplicity, and performance, Hosterman said. Right now, organizations are struggling with integration sprawl and are not using arrays as the technology was designed to be used.

To meet database performance demands, companies need to understand database requirements, Balasubramanian explained. While provisioning storage for an Oracle database, enterprises also need to understand current infrastructure constraints.

Virtual Volumes is a game-changer for virtualized databases and applications, according to Balasubramanian. The tool is trustworthy and offers seamless backup and recovery. It provides simplified cloning and refresh operations and also provides effective and consistent storage-based policy management. ■

### Rocket Software Creates MV Developer's Edition for AWS

Rocket Software has introduced the Rocket MultiValue Application Platform-Developer Edition, which targets new and existing users who want to learn about MultiValue development.

"It allows them to bridge the resource gap that is currently being seen across all industries," said Zain Master, senior product manager in MV, Rocket Software. The product creates a bridge between how old applications work and newer frameworks and architectures.

The Rocket MultiValue Application Platform-Developer Edition includes the Rocket UniVerse 12.1.1 database server and development platform and the Rocket MultiValue Integration Server 1.2.1—Rocket's newest gateway offering.

The solution is cloud-friendly and is easy to absorb, Master said. It allows users to clearly understand how to start, offers examples and samples included with the product, and can be used to develop mobile or web applications on top of an existing server architecture.

As the platform expands, Rocket will include new versions of servers and tools as they are released along with new examples and samples.

"We want to grow Rocket across the world," said Master. "We are now incorporating and modernizing our applications not only for the new frameworks coming out, but the new developers graduating from college."

As the company becomes more agile, Rocket hopes to hear from more customers about what they want and need.

"We started hearing from customers that there was this need for something like this so that not only the seasoned users could start using and playing with it but also novice users who are just starting in their careers. This is why we came up with this product offering," Master said.

### Kourier Integrator 4.7 Introduces New Performance Enhancements

Kore Technologies is releasing version 4.7 of the Kourier Integrator platform, which increases performance across the board. "A lot of our releases are customer-driven and 4.7 is a perfect example of customer-driven improvements and enhancements to the product," said Mark Dobransky, co-founder and managing partner, Kore Technologies.

This latest release includes the Package Execution Manager (PEM), which heightens the overall performance, execution, and transparency of Kourier-generated Microsoft SSIS (SQL Server Integration Services) packages and provides a stable and fault-tolerant environment in which to run them.

Based on new information from Rocket Software, Kourier was re-engineered to take advantage of Rocket's internal software changes and revised architecture to maximize Kourier's performance.

> *With **Kourier Integrator 4.7**, the net result for all of **Kore's** customers in all **environments** is that they are going to see an **improvement** in speed.*

Now, Kourier's overall performance is up to twice as fast compared to previous versions of Kourier when extracting delimited data from Rocket Software U2 files. The actual increase in performance will vary depending on the installed versions of the UniData/UniVerse database and operating system. Kore recommends that existing clients upgrade to this new release as soon as possible.

The net result for all of Kore's customers in all environments is that they are going to see an improvement in speed, Dobransky said. "We are working closely with our Rocket partner to improve performance for all our customers, not just our biggest ones."

Other additions include the following:

- Improvements to the dequeuing logic of Kourier to process the files moved from U2 to SQL and run their associated SSIS packages
- Enhanced output created by Kourier-generated SSIS packages to include additional information related to the file being processed and the SSIS package that is running
- Enhancements to the schedule listing page to sort the list of schedules by schedule description

"It's all about performance and reliability," said Dobransky. The next release on the horizon will make integrations easier, he added. "More people are trying to interface with applications in the cloud and we're going to make that a reality for them."

### Informer 5.1.2 Release Enables Querying Efficiency for Dataset Management

Entrinsik has announced the general availability of Informer 5.1.2, which introduces a bundling module to foster better sharing and accelerate the spread of best practices for data throughout the organization, while eliminating the requirement for complex IT support and expertise.

"Over the past decade, we have built an incredible foundation as a leading business intelligence solution, connecting thousands of people with insights from their data to make informed decisions that solve business problems. Working with our customers, we were able to transform how they scale their business, allowing data-driven decisions to reach more people, while maintaining an intuitive experience for our users," said Doug Leupen, founder and CEO of Entrinsik. "Promoting collaboration through bundling will accelerate the pace of adoption and lead to better outcomes for our customers."

According to Entrinsik, organizations are seeking to become more data-driven, but are limited when it comes to scaling to new departments, as installation, implementation, and replication of new instances can be labor- and time-intensive. The new bundling feature with Informer addresses this issue by providing users a consistent, easy way to share instances. A replication that took days can now be completed in minutes.

A key advantage of Informer bundling is the ability to quickly and easily update multiple contexts of a single multi-tenant Informer environment. Partners can disseminate content or share best practices from individual users to the entire ecosystem, providing continuous improvement for all.

Outside of Informer, queries can take hours or even days when processing millions or billions of records, but using Informer Datasets allows organizations to curate the exact desired data with the respective governance rules, workflows, and calculations, which are indexed into Elasticsearch for quick sorting and filtering. With the options of append and upsert, Informer will only add the new and updated records. This is particularly useful for larger organizations needing near-real-time reports on millions of transactions. A user can set Informer to refresh a dataset periodically throughout the day and only add new or modify changed records.

Additional new features from Informer 5.1.2 provide the following capabilities:

- **User-defined fields** provide a flexible data-sharing design that allows users to show different datasets to different sets of users and enables managers to create a master report, with individuals only seeing the data that is relevant to their role.
- **Private filters** enable users working with filters saved to a private folder to run reports with confidence since no one else has access to the filters.
- **SQL savelists** enable savelists for customers that have moved from U2 to SQL.

Other features and improvements, including enhancements to Pivot Tables, and the ability to apply selection criteria to normalized values and bug fixes across the system can be found in the release notes on the Informer Help Center.

### jBASE 5.7.4 Platform Is Now Available

Zumasys is introducing jBASE 5.7.4, which now includes object-oriented programming, JSON support, enhanced security, and more.

Object and JSON support updates allow users to parse JSON strings into jBASE, debug applications quicker, and enhance support for applications with dynamic objects.

In addition, enhanced encryption also allows users to secure the valuable data in their Pick system with AES-256 encryption and the latest OpenSSL libraries.

According to Zumasys, massive file support enables users to combine 64-bit operating systems, allowing files to be up to 16TB on Windows (64TB on Linux), and dynamic files perform up to three times faster.

System user profiles can now also simplify a user's system configuration and are stored as a JSON string, while jBASE profiles require zero scripting skills.

Docker OS-level virtualization also delivers software in packages in containers. Docker is supported in development systems only but will be certified for production in jBASE 5.8. ■

SHELBY KLINGERMAN

**Shelby Klingerman** is a content curation specialist at Quest Oracle Community. She works to produce quality educational content for Quest's JD Edwards, PeopleSoft, Cloud, and Database & Technology communities through articles and customer stories. You can contact her at shelby.klingerman@questoraclecommunity.org.

# Oracle Cloud's New Free Tier and Always Free Oracle Autonomous Database

At Oracle OpenWorld 2019, Oracle introduced the Cloud Free Tier with Always Free services so anyone can try the world's first self-driving database and Oracle Cloud Infrastructure for an unlimited time. It provides customers with the opportunity to build, learn, and explore the full functionality of Oracle Autonomous Database and Oracle Cloud Infrastructure, including Compute VMs, Block and Object Storage, and Load Balancer. Many developers are excited but will likely have lots of questions such as these:

- How do you get started?
- What do you get for free?
- What happens if you want to upgrade?

As part of Quest Experience Week (QXW) 2019, Todd Bottger, product management director at Oracle, gave a first look at the new free tier, which includes Always Free services such as Oracle Autonomous Database with Oracle Application Express (APEX).

Bottger gave a brief overview of the opportunity and then dove into a list of answers for these FAQs.

## Overview of Cloud Free Tier and Always Free Services

Oracle offered its Autonomous Database with APEX in response to an outpouring of requests from the developer community to have space available for development, experimenting, learning, and demo-ing. The developers wanted access to always be free and not limited in time. Oracle answered with the creation and launch of Oracle Cloud Free Tier.

This free tier is made up of "always free" services that can be used for an unlimited amount of time. Bottger explained what is included in Always Free, which is basically all of the essentials needed to build complete applications on Oracle Cloud.

- **Oracle Database 18c Express Edition (XE):** This is a full version of the Oracle Database with all of the core capabilities and very few restrictions. It is the same powerful Oracle Database that enterprises rely on worldwide and can be downloaded and installed on a laptop, workstation, or cloud. You get an Oracle Database to use in any environment, plus the ability to embed and redistribute for free.
- **Database Client Technology:** Oracle offers database client technology for free—including SQL Developer, SQL Developer Data Modeler, language drivers, and plugins. It is an integrated development environment that simplifies the management of Oracle Database in both traditional and cloud deployments. Oracle SQL Developer Data Modeler is a data modeling and database design tool that provides an environment for capturing, modeling, managing, and exploiting metadata.
- **Multiple Languages:** There are numerous languages—C, C++, Java, COBOL, PL/SQL, Visual Basic, and C#—available for developing applications. Developers can choose the languages in which they are most proficient or is most suitable for a specific task. Plugins can be used to accommodate popular third-party IDEs.
- **REST Data Services and SQL Developer Web:** Two important-to-know database technologies, REST Data Services and SQL Developer Web, are also included. REST Data Services allow users to set up database access to REST endpoints. SQL Developer Web is a web-based version of SQL Developer IDE so developers can get a competent IDE directly in a browser interface for free.
- **APEX:** Low-code application development is possible through the very popular Oracle low-code APEX. This enables the quick and easy setup of applications that look great in both mobile and desktop browsers right on the Oracle Database.
- **Hosted/Cloud Services:** Hosted/cloud services are also available for developers who can very easily sign up a schema running on a hosted environment with APEX workspaces already preconfigured and ready to go.
- **LiveSQL:** This works as a sort of scratch pad where developers can develop SQL to refine land share SQL statements.

In addition to the Always Free services, Oracle is offering a free trial with $300 in free credits for 30 days. The free trial is great for looking at and experiencing the functionality, performance, and scalability of the Oracle Database. During this time, users can access a wide range of Oracle Cloud services including Databases, Analytics, Compute, and Container Engine for Kubernetes. Free credits can

be used to provision multiple OCPUs (Oracle CPUs) with larger storage, more bandwidth, and greater network consumption with a full, rich functionality experience. Single sign-up provides access to both Always Free and Free Trial with a unified experience including a common console. For more information, visit http://oracle.com/cloud/free.

### Free Tier and Always Free FAQs

### What exactly is included? What are the specs?

In terms of databases, you have your choice of Autonomous Transaction Processing or Autonomous Data Warehouse, each with 1 OCPU and 10GB Object and 10GB Archive Storage prepackaged with APEX. For computing, the new free tier includes two virtual machines with 1/8 OCPU and 1GB memory each. For storage, the free tier includes two oVolumes (100GB total), 10GB Object Storage, 10GB Archive Storage. For load balancing, there is one instance, 10 Mbps bandwidth. For monitoring, 500 million ingestion datapoints, and 1 billion retrieval datapoints are included. For notifications, the free tier allows 1 million notifications to be sent through HTTPS per month, and 1,000 sent through email per month.

### What can I do with this?

When you log into your Oracle Cloud Counsel, you will see quick actions and solutions. Both provide a good overview of what's available. The homepage links to a live application running multiple Always Free services.

*Oracle introduced its Autonomous Database with APEX in response to an outpouring of requests from the developer community to have space available for development, experimenting, learning, and demoing.*

### Where is it available?

Availability across the globe is quite good for the Always Free services, according to Bottger. He recommends surveying the current Data Regions list at www.oracle.com/cloud/data/regions.html.

### How do I get it?

It's easy. Start at www.oracle.com/cloud/free.

### Is a credit card always required to sign up?

Many email addresses are already recognized by Oracle and won't require credit card entry. Recognized addresses include those from event attendees at OpenWorld and Code One, from students and educators participating in Oracle Academy, and from customers engaged through Oracle sales. In general, though, credit card entry is required by default for identification purposes.

### What happens at the end of the trial period?

If an account is upgraded to paid services, the remaining free trial credits are carried forward. Oracle will bill your credit card if you opt for paid-level services, but you will not be billed for the Always Free resources. If an account is not upgraded to pay level, the core class account will remain stable and usable indefinitely. Existing paid resources though are soon reclaimed and terminated. The existing Always Free resources continue running and can be provisioned up to the allowed maximums at no cost.

### How do I get Oracle support?

During the trial period, you can access My Oracle Support for free to create service requests. If you remain on Always Free beyond the trial period, you can use Live Chat via the Cloud Counsel on the user interface or you can use free public forums on Oracle Cloud Customer Connect.

### What database workload types are supported?

You can have one of each of the following:

- A free Autonomous Database for autonomous transaction processing for transactional and mixed workloads
- An autonomous data warehouse for data warehousing and analytics

### What can I do with it?

You can use the free Autonomous Database to store and manage diverse data in one place. Data scientists can use the built-in machine learning notebook to connect from third-party BI tools. Developers have access to preconfigured, fully prepackaged functionality, including APEX, REST data access endpoints, language drivers, power utilities, IDEs, and plugins.

### How do I get started with free Autonomous Databases?

Create an Always Free Oracle Autonomous Database—either transaction processing or data warehouse—to see the possibilities. Simply go to the menu and jump in. From there, use SQL Developer Web to interact with the database via SQL, to learn APEX, to experiment and create an APEX app on the database, or to download a client wallet to connect to the database form power utilities such as SQL and SQL*Plus.

### How do I upgrade to a paid database?

You can upgrade during the trial period. Go to the Actions Menu for free and see the option for "Upgrade Instance to Paid." Upon upgrading, the Always Free database is cloned into a new paid database with a different SKU but with stored data preserved.

### Are there differences between free and paid databases?

Yes, Always Free databases do not accommodate scale beyond the fixed resource limits. There is no built-in full database backup or restoring. There is no preview mode to experiment with future database releases. Also, databases may be paused at 7 days of non-use. ■

BY TOM LAROCK

Tom LaRock, head geek at SolarWinds (www.solarwinds.com), has more than 20 years of IT experience and is a Microsoft Certified Master, VMware vExpert, Microsoft Certified Trainer, and a 10-time Microsoft Data Platform MVP.

# Maximizing the Value of Digital Data

After years of new technologies being introduced into IT departments and new roles being created to manage them, most enterprises now operate in siloed worlds. At the same time, we've seen a rise in the value of data in powering today's businesses.

This reality has created a perfect storm in which siloed IT departments make it increasingly difficult to maximize the value of data. Today's technology professionals are left trying to navigate data optimization across complex environments.

*Even before the **IT elements** of **data optimization** begin, aligning organizational culture around a **data-driven** mindset will be a **major challenge**. Making the case for **data optimization** is important.*

Compounding this challenge is a simple fact: The larger the enterprise, the more opportunities for data silos to exist. These silos often collect, curate, and store datasets, leading to overlapping and duplicate data across the IT environment. To optimize data, these inherent issues need to be addressed once and for all.

## How to Achieve Data Optimization

Achieving data optimization isn't just about removing silos and putting tools in place; it's about instituting a data-driven mindset. The following are some best practices you should consider in order to achieve data optimization at the IT level and throughout your business:

- **Shift the organizational mindset:** Even before the IT elements of data optimization begin, aligning organizational culture around a data-driven mindset will be a major challenge. Getting employees across the enterprise to understand the benefits and making the case for data optimization will be difficult, but it's important.
- **Prioritize a governance strategy:** When thinking about maximizing the value of data, it's critical to implement a proper data governance strategy. The first step in this strategy is eliminating redundant data feeds. Eliminating these silos will lead to a reduction in storage, processing, and the total cost of ownership. This won't be an easy task—the volume and velocity of data increase every day—but building such a strategy helps expose the costs, benefits, and risks associated with your mission-critical data.
- **Optimize across the tech stack:** Once leaders understand the foundational elements of a data governance strategy, implement data optimization procedures and processes extending down into the application and database layers of the infrastructure.
- **Optimize data integrity:** Digital ledger technologies can help optimize the integrity of your data, ensuring your data is a source of truth and helping to make sure it hasn't been tampered with in an unauthorized fashion. Historically, this has been called master data management. These days, it's the same thing with a distribution of decentralized database nodes.

*The first step in a **data governance strategy** is eliminating **redundant data feeds**. Eliminating these silos will lead to a **reduction** in **storage**, **processing**, and the **total cost** of **ownership**.*

IT leaders need to bring their staff up-to-speed on the latest data optimization technologies and practices. The first step is getting everyone to understand the benefits, which means getting them to buy in to being truly data-driven. The next element is understanding how optimization reduces the total cost of ownership by removing the redundancies introduced by siloed IT departments. If tech pros can succeed in these areas, they can maximize the value of their data—not just today but into the future. ■

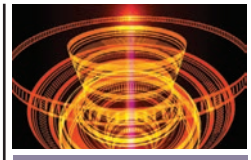
GUY HARRISON

**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.

# Elastos Aims for the Decentralized Internet

In the HBO series *Silicon Valley*, Pied Piper CEO Richard Hendricks pivots his startup toward the creation of a new, decentralized internet.

The series is a parody, but the concept is real. Before the internet, proprietary networks were fragmented, expensive, and private. The internet changed the world by creating a universal wide area network without any centralized ownership that is shared by all. The internet almost immediately replaced existing public networks and eventually became the most powerful communications infrastructure of all time.



EMERGING TECHNOLOGIES

The internet created a decentralized and democratized communications layer that allowed for enormous innovation. However, the applications built on the internet became centralized and monopolized by massive corporations such as Facebook and Google.

These companies own your data and, to some extent, have control over your digital life.

In *Silicon Valley*, CEO Hendricks attempted to return power to the people by creating a peer-to-peer "new internet" in which data and communications are distributed across all the devices on the network with no central point of control.

In the real world, several technology companies are trying to do the same thing, and blockchain is the critical enabling technology. Elastos is one of these next-generation blockchain projects.

Elastos describes itself as "the internet of the future" or a "blockchain-powered internet." Similar to Ethereum, Elastos supports the concept of "dApps"—distributed applications. But while dApps on Ethereum run entirely on the blockchain, Elastos dApps run within the Elastos runtime. The runtime can run on a desktop computer, servers, or smartphones. Offloading applications from the blockchain itself results in reducing congestion and a better end-user experience. dApps use the blockchain only for authentication and critical transactions.

Elastos also supports sidechains that further offload processing from the main blockchain. These sidechains can perform computation and data management and integrate with the main blockchain for critical transactions and authentication. Sidechains can be created for specific purposes but, within the core Elastos platform, sidechains provide identity and token services. There is also an Ethereum sidechain that allows Elastos to process dApps written in the Ethereum smart contract language and a similar NEO-compatible sidechain.

One of the challenges for new blockchains is the issue of assembling enough nodes on the network to create a sustainable and secure blockchain. Blockchains with relatively small numbers of computing nodes are vulnerable to "51%" attacks, in which an infiltrator assembles sufficient computing power to overwhelm the network.

Elastos avoids this issue by integrating with the Bitcoin blockchain. The Elastos mainchain is effectively a sidechain to the Bitcoin blockchain, which allows Elastos to provide a highly trusted consensus mechanism regardless of the computing power dedicated to the Elastos mainchain.

*In Silicon Valley, Pied Piper CEO Richard Hendricks attempted to return power to the people by creating a peer-to-peer 'new internet' in which data and communications are distributed across all the devices on the network with no central point of control. In the real world, several technology companies are trying to do the same thing, and blockchain is the enabling technology.*

The Elastos vision is ambitious, and to succeed, the Elastos foundation must surmount significant engineering challenges while simultaneously nurturing a developer community and a token economy. Furthermore, Elastos is not the only next-generation blockchain in development—EOS and Hyperledger, in particular, have similar ambitions.

Nevertheless, the Elastos vision is compelling, the implementation is relatively mature in blockchain terms, and it has already established an active community—particularly in Asia, where Elastos seems well-positioned to exploit the Chinese government's commitment to blockchain development. ■

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

# The Role of the DBA in the Application Development Process

Modern software development teams have adopted a continuous delivery approach based upon DevOps and agile development techniques. The small and frequent code changes that result from such methodologies can deliver significant benefit in terms of reduced lead time for changes, a lower failure rate, and a reduced mean time to recovery when errors are encountered. Indeed, today's developers migrate more frequent changes into the production environment than ever before.

This would seem to be a good thing—and it can be. However, when you add database change into the mix, things get a bit more challenging. More focus is needed on coupling software/database change both at a technology level and at a personnel level. DBA-level expertise is needed on development teams to help guide and implement changes appropriately. Automated software solutions that combine and coordinate application code and database changes are also needed.

This means that application and database changes must be tightly coupled and orchestrated in tandem for successful DevOps. This brings us to the question posed by the title of this month's column: What is the role of the DBA in the application development process?

This is a somewhat loaded question. Application folks generally don't want to turn over too much control of their applications to the DBA. And many DBAs don't want to get too involved in writing program code. But the two functions need to collaborate in order to produce efficient database applications. And this can be done without stepping on the toes of either the DBA or the application developer.

***DBAs** need to be **involved** in **application design** to ensure efficient **database usage**—and **therefore efficient applications**.*

DBAs must understand and participate in application design, not necessarily the actual coding though. Application design includes database concerns such as interfacing SQL with traditional programming languages and the type of SQL to use. But every aspect of program coding will affect the usability and effectiveness of the application. Furthermore, each application program must be designed to ensure the integrity of the data it modifies.

Designing a proper database application system is a complex and time-consuming task, and the choices made during that process will impact the usefulness of the final delivered application. An improperly designed and coded application may need to be redesigned and re-coded from scratch if it is inefficient, ineffective, or not easy to use. To properly design an application that relies on databases for persistent data storage, the system designer must match the application development languages and tools to the physical database design and the functionality of the DBMS being used. The first thing to be mastered, though, must be a sound understanding of SQL. And, unfortunately, many application folks just do not have enough SQL knowledge to formulate the most efficient query in every situation. SQL is changing and growing, and the DBA will usually have current and in-depth knowledge of what is possible and what SQL formulations perform best. The application developer and the DBA need to work as a team to make sure that this knowledge gets built into each and every DB2 application.

Access paths, the internal instructions used by the DBMS to actually access data, are another critical area where DBA guidance can help. The database optimizer converts SQL into access paths, and any time the SQL changes, the access path for the statement may also change. And that can dramatically alter the performance of the application. Oversight is crucial. As such, some organizations require the DBA group to review and approve all SQL changes to identify potential access path problems. DBA review of all SQL changes before they are migrated to production is a reasonable precaution, especially for organizations with developers that are not skilled at SQL coding and access path interpretation. Automated tools that examine and report on changes and proffer guidance are useful here, particularly in an automated DevOps culture.

Of course, there are other issues, such as the interface to the programming language, drivers, procedures for compilation, and so on. But let's just summarize by saying the DBA needs to be involved in application design to ensure efficient database usage—and therefore efficient applications.

And be sure to involve the DBAs in all of the design reviews for the code too. ■

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.

# Azure Cosmos DB Adds Features and Free Tier

I've kept you informed over the years about not just the latest news and features in SQL Server, but across the entire Microsoft Data Platform, both on-premise and in the cloud. This includes additional products such as SQL Server Analysis Services (SSAS), SQL Server Integration Services (SSIS), and their cloud siblings, Azure Synapse Analytics, and Azure Data Factory (ADF). Another important and powerful sibling product is Azure Cosmos DB, previously known as Document DB, which I first wrote about here many years ago.

## What Is Cosmos DB?

Azure Cosmos DB is a global NoSQL data management system that offers fast read and write speeds with automatic scaling, high availability, and a very flexible set of options for consistency, availability, and "partitionability." While Cosmos DB is its own data management platform, you can use it to emulate or even directly substitute for popular NoSQL alternatives such as MongoDB or Cassandra. Cosmos DB also provides a variety of data models for storing data and APIs so that you can access your data as if it were SQL, Cassandra, MongoDB, Gremlin, Spark, etcd, or Table (of the wide-column family popularized by Cassandra and HBase).

SQL SERVER DRILL DOWN

I like how my colleague and fellow *DBTA* writer, Guy Harrison, describes Cosmos DB (https://bit.ly/2WdCWTn):

"At the heart of Cosmos DB is a geo-distributed and replicated database. Data is stored in containers that can be partitioned across regions. The partitioning can be adjusted dynamically by the system to optimize throughput. The partitioning can also be tuned to provide geographic optimization: Data specific to Europe can be located primarily in European data centers, for instance. Cosmos DB offers uniquely configurable options across consistency models, data models, and APIs."

## New Features Every Month

Like many of Microsoft's Azure products, Cosmos DB has a rapid release calendar with lots of new features and capabilities releasing at a regular cadence. For example, in January, the team released new APIs to support MongoDB server version 3.6, as well as a new Autopilot feature (in preview) to automatically manage and elastically scale the request units (RUs) consumed by your instance of Cosmos DB. And in February, the team rolled out the new SDK 2.2.2 release of the SQL API that supports Spring Data, also incorporating new diagnostic information, such as the RequestCharge API with better handling for user requests (https://bit.ly/3cVkQuY).

## Announcing the New Free Tier for Azure Cosmos DB

The new free tier for Azure Cosmos DB offers new accounts 400 RUs and 5GB of storage free each month. This amounts to about $25/month in free services. If you've built apps on Cosmos DB that run on less than 400 RUs and 5GB of storage, then you won't have to pay anything. Conversely, however, if you know your application will exceed that usage, then you get to apply the discount once you consume more than those resources.

For users who need lots of collections, up to 25, you will now be able to run those in a shared throughput dataset where they all share the free 400 RUs/month so apps that previously would have cost a bunch of RUs can now run for free with the new free tier. Just because this offering is "free" doesn't reduce its value to you. The free tier still includes the same 99.99% availability and the same <10ms read/write latency normal to all Cosmos DB accounts. However, Microsoft does not offer an SLA on the free tier of resources consumed. You can have up to one free tier per Azure Cosmos DB account per Azure subscription.

Please note that if you create a database or container that's replicated in multiple regions, then the amount of RUs consumed is multiplied by the number of regions. So if you have a three-region Cosmos DB database and you use 150 RUs in a month, that will actually total up to 450 RUs across all three regions, thus exceeding the free tier limits. However, you would receive a bill for only 50 RUs (450 RUs used – 400 RUs free = 50 RUs billed).

## Get Started

Start by creating a new Azure Cosmos DB account (https://aka.ms/create-cosmos-account) and then activate the free tier (https://aka.ms/cosmos-free-tier) during account creation. Remember that you are limited to one free tier account per Azure subscription. Check out the Microsoft quickstarts and tutorials (https://docs.microsoft.com/azure/cosmos-db) for traditional lessons. Or try out a sample Azure Notebook (https://docs.microsoft.com/azure/cosmos-db/create-notebook-visualize-data) to see what you can do with Azure Cosmos DB right away. ■

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.

# Think of Solutions as Independent Enterprises

As we have moved forward with APIs and microservices, every organization has even more data stores to manage and more sources of data to consider. Sorting through data structures for operational solutions can become mind-numbing due to the variety, or even frustrating due to a lack of detail from many vendors. Source systems are no longer the monoliths they once were. Cloud-based solutions that are proffered for enterprises to use may very well have multiple physical data stores across their internal microservices. This can preclude a simple download from a centralized database to a customer organization for ingestion on their analytics platform. It is becoming more normal for developers with cloud-based solutions to respond as if someone is speaking an unknown language when customers ask for a "dump" of their database. It is entirely possible that said cloud-based solution may not have a database to dump. Data persisted could be JSON documents that may or may not be stored within some DBMS tool. Or, even more likely, the components or services comprising the solution may each have their own separate databases, with each microservice using a completely different tool. Therefore, part of the data may be in a document store, another stored in key-value pairs, and a third in a graph database.

DATABASE ELABORATIONS

Presenting an API for logically unloading data may be the only option one may have for obtaining data for use or integration into an organization's analytics platform. Therefore, an organization should vet how such a data acquisition process is to be handled. Ignoring the need for this function until after a cloud-based tool is deployed can lead to very unpleasant surprises that come far too late to avoid. As mentioned, some vendors have anticipated their customers' needs and have a clean API in place for providing data. However, not every vendor is so forward-thinking. Some may only provide an inflexible one-time snapshot, with follow-on delta files. This means that, for the customer, delta logic must be coded for updating and storing history within one's data warehouse or data lake. Problems may arise if additional one-time snapshots are needed. Other vendors may take a stance that customers must pay as they go, meaning that if a customer wishes to actually have a copy of his or her data, the customer must pay a fee for developing and for each execution of such an extraction process. And yet others may provide nothing but tap dancing and double-talk as they try to keep customers confused until their internal development advances to a point of addressing such questions. Exposing this aspect of a possible tool is vital before an organization commits itself to a solution.

What this signifies is that, more than ever, solutions really need clarity as to the meaning of the data shepherded within them. If the solution were to be compared to an independent enterprise, there would be a screaming-loud need for an "enterprise data model" for the solution—meaning a solution data model. In this fashion, clarity and understanding would be enhanced, and conversations with solution data users might be held that are not frustrating to everyone involved. The data model and the understanding of the model's composition allow for a framework in defining the breakdown of extraction API elements. Having this clarity exposes where the rubber meets the road, because physical structures are not logical structures. Industry-wide, this situation has always been a conceptualizing problem for many people. Data architects supporting this kind of an environment must not suffer from ambiguity in comprehending the physical versus logical differences. Even when differing data structures and DBMS platforms are used for separate microservices, those implementations should be able to be rationalized against the master solution logical data model. Confusing logical/physical distinctions will result in data models that will be much less helpful than they might be otherwise. The bottom line is that data modeling practices are more necessary now than ever before. ■

**APRIL/MAY 2020 Ad Index**

Melissa . . . . . Cover 2
SHARE . . . . . Cover 3
Revelation . . . . . Cover 4
IRI . . . . . 14
Kore Technologies . . . . . 25

**Best Practices Sponsors**

CData . . . . . 12
Vertica . . . . . 13

#SHAREbos

# Join us at SHARE Boston 2020

Hynes Convention Center | Boston, MA

## August 2-7, 2020

**SHARE Boston is a industry-leading educational event that will focus on IT hot topics such as:**

- Open Source for Z
- DevOps
- Data Privacy
- Women in IT
- Security and Compliance
- IBM Z®
- Cloud Technology

"SHARE is North America's premier association and event for people who work in the largest and most critical IT environments to ***meet***, ***network***, ***learn***, and ***give guidance***. There is nothing else that even comes close to its importance."

*–SHARE Pittsburgh 2019 attendee*

**Visit event.share.org to learn more and register by June 19 to save!**

REVELATION
SOFTWARE

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

SCALABLE

MORE DAZZLING

FASTER

SYNERGISTIC