TRENDS AND APPLICATIONS

The Future of Analytics:

Leveraging Data Lakes and Data Warehouses

> Dark Data: The Next Enterprise Frontier

How Microsoft and Citrix are Enabling the Future of Remote Work

> QLDB Paves the Way for Blockchain Databases

WWW.DBTA.COM

Data Quality, Easy as 1-2-3!

Bad data is bad for business. Its real cost is three costs: time, money, and your sanity. Leverage Melissa's 35+ years of expertise in data quality, address verification and identity solutions to stop bad data in its tracks.

With our solutions, you can:

1. DENTIFU - Target your ideal customer and prevent fraud

2.VERIFY - Employ real-time address, email, phone and name verification to keep bad data out of your systems



3 A Consolidate duplicate records for a single customer view



Repeat after me: Data Quality, Easy as 1-2-3. Visit Melissa.com for (1) free trials, (2) free developer credits and (3) code samples. It's easy, like counting up to three!



www.Melissa.com 1-800-MELISSA



Volume 34 | Number 5 OCTOBER/NOVEMBER 2020

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

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

Lauree Padgett,

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	Thomas Hogan Jr., Vice President, Market- ing and Business Development
Chairman of the Board	Bill Spence, Vice President,
Mike Flaherty,	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 Martton Pike, Medford, NJ 08055 USA; Phone (609) 654-6206; Fax (609) 654-4309; Internet: infotoday.com. Registered in U.S. Patent & Trademark Office. Periodicals postage paid at Vincentown, NJ, and additional mailing offices.

 $\ensuremath{\mathbb{C}}$ 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.

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@infotday.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 subscriptors 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



FEATURE STORY

4

THE FUTURE OF ANALYTICS: LEVERAGING DATA LAKES AND DATA WAREHOUSES BY JOE MCKENDRICK

DEPARTMENTS

APPLICATIONS

- **9 THE MODERN DATA CENTER: MAKING MIGRATION EASY** BY MICHAEL JACK
- 25 TODAY'S TOP TIPS FOR SECURING WORK-FROM-HOME NETWORKS BY DON BOXIEY.JR.

DBTA LIVE

- **11** THE EVOLUTION OF THE DBA
- 11 HARMONIZING DATABASE MANAGEMENT WITH DEVOPS
- **12** BALANCING DATA INTEGRATION WITH DATA GOVERNANCE
- **12** GOING DEEPER WITH DATA SCIENCE AND MACHINE LEARNING

MULTIVALUE SOLUTIONS

- **13** ROCKET SOFTWARE BEEFS UP SECURITY IN LATEST D3 RELEASE
- **13** ZUMASYS OFFERS TUTORIAL ON HOW TO SUPERCHARGE MV APPLICATIONS

13 BLUEFINITY ENRICHES EVOKE PLATFORM WITH EXTENDED MV INTEGRATIONS

COLUMNS

20	QUEST IUUG UBSERVATIUNS > BY SHELBY KLINGERIVIAN
	IMPORTANT (ORACLE) DBA SKILLS TO HAVE, OBTAIN, AND NURTURE
28	NEXT-GEN DATA MANAGEMENT > BY THOMAS LAROCK
	HOW MICROSOFT AND CITRIX ARE ENABLING THE FUTURE OF REMOTE WORK
29	EMERGING TECHNOLOGIES > BY GUY HARRISON
	QLDB PAVES THE WAY FOR BLOCKCHAIN DATABASES
30	DBA CORNER > BY CRAIG S. MULLINS
	THE CLOUD AND DATABASE ADMINISTRATION
31	SQL SERVER DRILL DOWN > BY KEVIN KLINE
	AUTOMATION IN AZURE USING ELASTIC DATABASE JOBS
32	DATABASE ELABORATIONS > BY TODD SCHRAML
	STAGING VERSUS LANDING AREAS IN
	ANALYTICS ENVIRONMENTS
MEDI	A PARTNER OF THE FOLLOWING USER GROUPS
UG Ideide BH2 Bar Camer	Service State Stat









Dark Data The Next Enterprise Frontier

By Joe McKendrick

his may be the era of the datadriven enterprise, but only a handful of organizations report they are ready for it. There is a growing volume of "dark data" that remains obscure to IT managers and decision makers. This period unfolding before us will be driven by several technology initiatives, from 5G wireless and IoT to AI.

These are the main takeaways of a recent survey of 2,259 IT executives, published by Splunk, which found that organizations are aware of, but woefully unprepared for, the data challenges ahead. These include the prevalence of "dark data," which is growing at least as quickly as their overall data volume. And most of the managers and professionals surveyed have limited understanding about the new wave of data-rich, cutting-edge technologies, such as AI, blockchain, 5G, augmented reality, and more.

For starters, only 14% of respondents to the survey said that their organizations are

said that their organizations are currently preparing for fast-rising data volumes (see Figure 1). "That leaves 53% whose organizations are not prepared or preparing at all," the report noted. In addition, among the 86% whose organizations are not yet pre-

Sixty-six percent of respondents revealed that half or more of their data is dark data—the unquantified and untapped data generated by systems, devices, and interactions.

currently prepared for an imminent wave of new data, whether as projected from their current rate of data growth or as caused by adoption of new technologies. Another 33% pared, only 8% are very confident that their organization will be ready in time.

Executives and professionals recognize data as an asset, but most are struggling

VVV U. U U U 0 00 1 1 0 1 11 1000000 0 1 0

Figure 1: Preparedness for the "Data Age"



to keep up with it, the survey also showed. Two-thirds of respondents said both the value and amount of data in their organizations will increase—expecting the sheer **Edge computing:** Edge computing is among the technologies least understood by respondents, with just 39% understanding it well or at an expert level. "A

The leading challenge to discovering and accessing dark data is the sheer data volume.

quantity to grow nearly 5 times by 2025. In addition, 57% said the rate of data growth is outpacing their organizations' ability to keep up with it, and 47% bluntly said that their organizations will fall behind when faced with rapid data volume growth.

In addition, the survey showed that much of this value has yet to be captured. Sixty-six percent of respondents revealed that half or more of their data is dark data—the unquantified and untapped data generated by systems, devices, and interactions. This is a 10% increase over the previous year. The leading challenge to discovering and accessing this unutilized data is the sheer data volume.

The survey assessed the awareness and adoption of a number of key emerging technologies, including the following:

5G: Many IT and business managers don't have a strong grasp of 5G technology, with just 44% saying they "understand the technology well or at an expert level."

behind-the-scenes tool even within the often-overlooked arena of data storage infrastructure, fully 25% of IT and business managers don't know their organihave adopted AI and machine leading in some form, and 51% said that their organizations will be using it in the future, with an average time-to-use of 3.3 years. Fifty-four percent said they expect AI and machine learning to increase their data volumes, "perhaps by enabling organizations to put to work data they have previously been unable to utilize or to bring in more data they've avoided for want of analytics resources," the report noted.

1

IoT: Forty-five percent of respondents said they understand IoT well or have an expert understanding—higher than the other technologies driving the increase in dark data (though by just one percentage point compared to 5G). Twen-ty-eight percent are currently using IoT technology, and an additional 48% said their organization will be using it in the future. These future users expect to be using IoT devices in 3 years, on average.

Augmented reality (AR) and virtual reality (VR): While 42% of respondents said they understand these well or have an expert understanding, 35% admitted to understanding it only a little or not at all. Twenty-two percent of respondents said that their organizations are currently using AR or VR technology, with

For all the attention on **AI** and **machine learning**, the **percentage** of respondents who understand it well or have an **expert understanding**—**42%**—indicates a significant **knowledge gap**.

zation's intentions for edge computing, while 24% report that their organization has already adopted it," according to the report.

AI and machine learning: For all the attention on AI and machine learning, the percentage of respondents who understand it well or have an expert understanding—42%—indicates a significant knowledge gap. At the same time, 25% of respondents said that their organizations

an additional 47% planning to use it in the future. Half expect AR/VR to increase their organizations' data volumes, and only 12% expect it to worsen their dark data problems, while 23% believe it will solve them (29% said it will solve some, worsen others). "This disconnect reinforces that many IT and business managers are not foreseeing the challenges that come with the technology and growing data volume," the report stated.



0.0

The Future of Analytics: Leveraging Data Lakes and Data Warehouses

By Joe McKendrick

oth data warehouses and data lakes offer robust options for ensuring that data is well-managed and prepped for today's analytics requirements. However, the two environments have distinctly different roles, and data managers need to understand how to leverage the strengths of each to make the most of the data feeding into analytics systems.

Data warehouses are repositories of structured, transformed data configured for specific applications. They serve as central locations for integrated data from one or more disparate sources, said Ryan Wisnesky, co-founder of Conexus. "They store current and historical data and are used for creating trending reports such as annual and quarterly comparisons. A data warehouse is highly transformed and data is not loaded to the data warehouse until the use for it has been defined."

Typically, data warehouses "support functions that are used to create reports, understand trends, and make more tactical decisions that address day-to-day and short- to medium-term business activity," said Sri Raghavan, director of data science and advanced analytics product marketing at Teradata. Data lakes on the other hand, which typically see a lot of analytics activity, are used to investigate, discover insights, and address a more holistic set of business challenges. They usually require data and analytics functions that are not a part of the data warehouse environment, Raghavan noted.

Data lakes draw in data from all sources, whether for defined or unspecified

purposes. They serve as repositories for raw, unprocessed data straight from data sources, and this data may reside in the lakes until needed at a future time. While a data warehouse may be more akin to a city water supply, a data lake "is more like a body of water in its natural state," said At the same time, Kaluba continued, data lakes "offer inexpensive options to traditional database systems." They expedite processing and function as more of a sandbox or investigational environment for data. Since data lakes are rarely managed and supported to the degree of the

Both data lakes and data warehouses can be supported at the same time since it is not so much a question of which product you should use for your data but instead a matter of having purpose and intent regarding how you're going to use it.

Wisnesky. "Data flows from the streams the source systems—to the lake. Users have access to the lake to examine, take samples, or dive in. Data lakes retain all data. All data is loaded from source systems. No data is turned away."

The good news is that both environments can be supported at the same time. "In some cases, enterprises are operating an open data lake right alongside of the data warehouse," said Dave Mariani, co-founder and chief strategy officer at AtScale. The choice, he noted, often depends on the business case at the end of the data funnel. "It's not so much a question of which product you should use for your data; rather, it's a matter of having purpose and intent around how you're going to use your data and being able to do something with it that is the gold standard," agreed Nima Negahban, CTO and co-founder at Kinetica.

DATA-DRIVEN FUNCTIONS

It's important that enterprises understand which functions are applicable to either type of environment. Data warehouses, for one, are traditionally seen as systems of records—implying the data within these environments is well-organized, mapped, supported, and has some level of quality, said Kim Kaluba, senior manager of data management solutions at SAS. Data warehouses best support CRM-, ERP-, EDW-, and MDM-type initiatives which require stable and trusted data for decision-making functions, she said. data warehouse, Kaluba added, "the data functions or business needs they best support include exploratory analytical functions where raw, unrefined, and large data is used to test new algorithms, identify insights, and answer questions."

The bottom line is that data warehouses "are best suited to providing high-performance, ad hoc analytics whereas data lakes are more suitable for use cases where raw data access is required," said Mariani. "Data warehouses are ideal for analytics because the data is usually cleaned and normalized. In addition, data warehouse architectures are optimized for analytics. In contrast, a data lake is best utilized

cal data. Data analysts commonly understand data lakes as a repository for raw data. Processed data can also be deposited into a data lake, allowing it to be more easily combined with other data." That's because "in its native and isolated form, accessing data is difficult." Imagine a new analytics project that needs to work with data from a series of databases containing CRM, ERP, syndicated, and sales-channel data, he noted. "Accessing data in each of these repositories is time-consuming and requires authorization and specific skills. Collocating data all in one place makes it much easier to work with. The data lake serves as the common repository for the various data sources, greatly simplifying the job of transformation."

THE FUTURE OF DATA WAREHOUSES

What is the future of the data warehouse in the emerging real-time, datadriven enterprise? How is its role changing, and how does it fit into the picture? "The need for warehouses hasn't changed much; however, now they are being accessed through the cloud in many instances," said Wisnesky. The problem, he said, is that cloud platforms can create interoperability problems by becoming a new type of silo, "especially given that ELT technologies encourage deferring schema construction."

Data warehouses, which once focused on historical data, are also taking on real-time duty. Machine learning and AI

A new vision of a hybrid environment, called the 'lakehouse,' provides a structured transactional layer to a data lake—allowing many of the use cases that would traditionally have required legacy data warehouses to be accomplished with a data lake alone.

as a landing zone for raw data for use in downstream applications and data warehouses. Data lakes are optimal for data science workloads, where access to granular data is needed."

The low-cost availability of storage enables enterprises to increasingly use data lakes, agreed Chris Bergh, CEO of DataKitchen. "A data lake utilizes simple storage to retain the organization's critimodeling allow data warehouses to operationalize those models so that the gap between an activity, such as a customer purchase, and the response, in the form of product recommendation, is a matter of seconds as opposed to days or weeks, Wisnesky said.

Data warehouses can also handle much larger datasets as they speed through rapid analysis. "Computing power and memory have advanced to the point that data warehouses can process much larger and more complicated datasets," said Mariani.

REAL-TIME AND UNIFIED ANALYSIS

The data warehouse has adapted by moving from on-premise to the cloud, and it will continue to adapt, he noted. However, when it comes to real-time processing, data lakes present a better choice. "For real-time workloads, data warehouses are not ideal because even this new generation of data warehouse requires that data be loaded, thereby introducing latency," Mariani said.

While the traditional data warehouse "focused on the first mile of ingesting and storing data for analysis, a modern data warehouse both ingests and stores data, and analyzes that data in real time as it is received," said Negahban. "Modern data warehouses will deliver real-time analysis on incoming data streams, while incorporating all of an organization's data and applying cutting-edge location intelligence and machine learning-powered predictive analytics." Data warehouses of the future that process data in real time and unify analysis of the data in different formats-such as relational, geospatial, graph, and time series-at scale will benefit from increased accuracy and detail for customers across industries, Negahban noted.

Ultimately, the success of data warehouses going forward comes down to "semantics, semantics, semantics," said Wisnesky. "In 2020 and beyond, the new challenge for data warehouses is how to best internalize the domain semantics in a way that provides the most value to users. For example, a data warehouse that automatically knows that two entities—say Pete and Peter—are actually the same can internalize that fact so that anyone who queries the warehouse will be made aware of the fact that there are two references to the same real-world entity. Similarly, a warehouse that automatically knows what risk is because it has internalized an ontology such as the Financial Industry Business Ontology can provide semantic query capability to users. We see lightweight knowledge graphs-as opposed to decades-old semantic web technology-as being the harbinger of semantics in 2020."

An emerging generation of ETL visualization tools may also increase the value of data warehouses into the future. "The visualized ETL process that is essential to integrate data from multiple source systems, especially the legacy systems, is the technology having the most positive impact on enterprises' ability to compete on data," according to Alex Ough, senior CTO architect at Sungard Availabilences in governance, metadata management, and user experiences across different environments, Raghavan added.

In addition, data governance and quality is another challenge with data lakes. "Appending and modifying data is hard, jobs fail without notification, and keeping historical versions is costly," pointed out Joel Minnick, vice president of product marketing for Databricks.

'Data lakes can be just as suitable as traditional data warehouse systems for analytical processes and data-driven initiatives if they are grounded in a comprehensive data strategy supported by data governance and data management processes.'

ity Services. "Machine learning models, along with the frameworks used to train the models, have improved significantly. These technologies have made it easier for less skilled individuals to train models with high accuracy. However, data engineering is still very complex and time-consuming, as many of the processes need to be done manually, especially when there are multiple sources of truth with duplicated data in legacy systems. In most cases, pre-processing data requires a deep knowledge of SQL or other programming languages to define relationships among the source data, remove duplicates, and clean mistyped data to improve data quality. Having top-notch ML algorithms and frameworks is useless if you cannot prepare quality data."

TIME FOR A DATA 'LAKEHOUSE'?

While data lakes may be more ideally suited for fast-paced, real-time requirements, they can be more cumbersome to manage than data warehouses. For example, it's difficult to automate the way they are used. "If an enterprise relies on the cloud, data ingestion into a cloud data lake is usually a laborious process, given the immutable nature of such systems," said Raghavan. Data workflows need to be built and managed with a view toward smooth orchestration across multiple environments, including multi-cloud and hybrid cloud scenarios, while dealing with some environment-specific differIn addition, it's difficult to handle large metadata catalogs, and consistent performance "is elusive due to small file problems and complicated partitioning." Finally, it's a constant headache to maintain data quality, he added.

Mark Fernandes, managing partner at Sierra Ventures, said he has often seen companies build a data lake and quickly start ingesting large amounts of data. "Soon after, the lake turns into a swamp with a lack of visibility and compromised data quality," he said. "End users don't feel confident in the data, and analytics projects come to a halt. Data lake technology stacks based upon Hadoop can be complicated and challenging to manage, especially when you start migrating to the cloud and integrating the various tools needed for data ingestion, quality, data management, governance, preparation, etc. Lastly, many data lake projects fail because they weren't built with a business use case in mind."

Minnick said a new vision of a hybrid environment, called the "lakehouse," is emerging, which "provides a structured transactional layer to a data lake to add data warehouse—like performance, reliability, quality, and scale. It allows many of the use cases that would traditionally have required legacy data warehouses to be accomplished with a data lake alone." A lakehouse architecture also can support "unstructured data like video, audio, and text, as well as structured data that has traditionally been the domain of legacy systems," Minnick explained.

SUPPORTING DATA-DRIVEN INITIATIVES

Establishing the best environments for supporting data-driven initiatives using AI, machine learning, and IoT is a learning process, industry observers noted. "Data lakes can be just as suitable as traditional data warehouse systems for analytical proEnterprises are constantly looking for ways to use data across the business to build smart analytical applications that drive competitive advantage, said Negahban. "Traditional data warehouses do not address the need to integrate data across all aspects of the business—custom applications, IoT, or analytics dashboards. A modern data warehouse should provide a full set of APIs to embed analytics in applications. By taking an API-first approach to building

Enterprises are constantly looking for ways to use data across the business to build smart analytical applications that drive competitive advantage.

cesses and data-driven initiatives if they are grounded in a comprehensive data strategy supported by data governance and data management processes," said Kaluba. "This ensures that the data inside of the data lakes is reliable for organizational decisioning processes."

It may be more efficient to keep storage and compute separate as well. "Decoupling of storage and compute reduces costs and improves scalability," advised Fernandes. "Data can be stored in a cloud environment like AWS S3, and compute clusters can be spun up as needed to run workloads or queries. This type of elastic provisioning and pay-peruse are key requirements for modern data warehouses. Enterprises are also looking for integrated data governance and self-service data access to support various downstream applications, including artificial intelligence and machine learning use cases."

In addition, it's important not to rush into anything. Companies that move into AI before mastering the fundamentals whether their data is in a data lake or a data warehouse—will end up paralyzed, Mariani cautioned. "Not only do you need to be good at data engineering and business analytics, you also need to embrace advanced automation. Lack of automation means people manually use the keyboard to process pipelines, do ETL, move data, and create downstream assets, which does not scale. All of these activities can and should be automated." data-driven applications, a modern data warehouse is able to present data at any point of user interaction, giving the business the flexibility to use the tools, apps, and platforms it prefers across departments."

EMERGING APPROACHES

An emerging discipline, DataOps, may also help bring greater order to data lake or data warehouse management. "Imagine a 50-person team managing numerous large integrated databases for a big insurance or financial services company," said Bergh. "Their customers—colleagues in a business

When it comes to the best ways to manage data lakes to support data-driven initiatives, "first identify your key use cases, your key business sponsors, and organize your data initiative to achieve use-case success," said Fernandes. "Then create a solid foundation for your data lake by leveraging an agile and flexible DataOps approach to automate processes, standardize governance, and provide self-service access to the data. DataOps optimizes the full data cycle by controlling data sprawl and managing the entire supply chain of data from ingestion to consumption," he noted. "Finally, use augmented data management approaches and a unified platform to enable and govern data lake/store functions, such as cleansing, deduplication, data classification, and gain visibility and insights about the data lake's health and usage."

Organizations should also consider looking into the hybrid lakehouse approach, Minnick advised. "It builds on the best qualities of data warehouses and data lakes to provide a single solution for all major data workloads and supports use cases from streaming analytics to BI, data science, and AI. Historically, companies have been forced to create data silos with legacy data warehouses and data lakes, and use them separately for BI and AI use cases. This results in information inequality, high costs, and slower operations. By combining all the data onto

An emerging discipline, DataOps, may help bring greater order to data lake or data warehouse management.

unit—have lots of questions that drive new and updated analytics, but the data team can't keep up due to heavyweight processes, serialization of tasks, overhead, difficulty in coordination, and lack of automation. They need a way to increase collaboration and streamline the many inefficiencies of their current process without having to abandon their existing tools. DataOps automates the orchestration of data to production and the deployment of new features, both while maintaining impeccable quality. DataOps can be incredibly beneficial to both data lake and data warehouse agility in large data teams." the same open, high-performance, lowcost platform, the entire organization is able to move faster and make better decisions."

For those relying more on data lakes as core enterprise repositories, Wisnesky suggested that enterprise data managers build data models. "A data lake is a data storage device. The data stored in it still has underlying meaning, even if that meaning isn't formalized as a data warehouse schema. Automation is driven by formalization. The best-managed data lakes are actually data warehouses of data warehouses."

APPLICATIONS

The Modern Data Center Baking Nigration Easy

By Michael Jack

"It's yours, free. Just pay shipping and handling!" claims the infomercial host. We've all seen them, the advertisements for home maintenance, gardening equipment, or even cooking products that promise the world but end up costing a lot of money. Whether it's through shipping fees or the need to sign up for long contracts that are hard to break, unfortunately, most deals that seem "too good to be true" usually are not what they seem.

I can think of many examples of offers that appear free but always result in unexpected costs and lost opportunities. The same can be said for outdated network-attached storage (NAS) migration tools. Your organization may have something in place that looked brilliant in the advertisement (and was maybe even free), but now that you have installed it, it's bringing up a whole load of problems: lost data, lack of security, slow processing speeds, and the inability to handle anything remotely complex.

That's because many legacy migration tools were created when data amounts were much smaller and there were fewer major storage companies in existence. This is not the case today, and it has made moving from one platform to another extremely complex, and, as a result, many projects are failing.

Let's look at why legacy NAS migration tools are costing companies a whole lot more than they realize, how to modernize your data center with software that suits you, and where the migration industry is heading.

Cut the Cord on Old Storage

As with many free infomercial products, deploying an outdated legacy migration tool often leads to hidden total cost of ownership and lost opportunity.

Many data center managers and CIOs tend to solely focus on upfront hard costs such as personnel time when contemplating modern storage transition tools. Instead, they ought to consider the fact that a slow and disruptive migration means succumbing to the costs of maintaining old storage and wasted opportunity. Each day it takes to make the move is a lost day of business benefits.

APPLICATIONS ---

Ask the Tough Questions

I recommend asking these three questions about any software being considered:

1. Can it handle scale, speed, and complex configurations?

Modern NAS migration tools should be able to graphically define a migration and dictate what content gets copied between the source and target, as well as the frequency at which that content is resynchronized. interface, they find they can quickly initiate and complete all steps of the migration themselves without being experts on every detail of either the source or target platform. This helps maintain costs and complete all migration tasks quickly.

3. Can it track/validate my files securely?

NAS and object data have become some of the most important data stored by an organization, often including customer data subject to government regulations,

Rather than focusing on upfront hard costs such as personnel time when contemplating modern storage transition tools, data center managers and CIOs should consider the fact that a slow and disruptive migration means succumbing to the costs of maintaining old storage and wasted opportunity.

Just 10 or 15 years ago, a typical NAS migration was in the tens of terabytes. There weren't many storage companies in existence then, so we were typically dealing with transfers between similar systems. This means that the legacy migration tool you may be currently using is likely not able to handle complex or important data transfers. Today, it is not uncommon for a NAS migration to involve thousands of terabytes and billions of files. The storage market is booming, and every company has its own take on how to implement a storage protocol. This provides you with a wide variety of storage choices, but it also makes things much more complex.

2. Is it easy for my team to manage?

The tool should also present administrators with a single pane of glass for easy monitoring and management. Your NAS administrators must be able to understand the ins and outs of the migration so that they can confidently determine cutover event schedules and agree on a tolerable amount of downtime with end users. Many organizations have data center administrators within their IT teams, and when they're able to work with a simple user financial data subject to audit, or research information used for business intelligence. Any loss of this data or disruption to the systems that store it can cause business damage and expense. Yet many However, data corruption can occur without affecting the timestamp and size of the file. Any corruption of the actual content of the file during the migration is not discovered by this rudimentary check. Your software of choice should provide chain of custody by hashing every single file as it is migrated. A file is declared successfully migrated only if the source and target match. A report can be created to show every single hash of every single file, which can be kept for future auditing.

Where the NAS Market Is Trending

Many organizations say they offer enterprise-class software to help organizations transition to new storage technology quickly, but there are very few that were actually designed to safely protect unstructured data on any NAS or cloud. Using a tool designed for NAS migration means that projects of any size are no longer tedious and time-consuming, but fast, easy, and affordable.

Data center managers today are looking for a consultative approach when choosing a NAS solution. They're taking a longterm view rather than jumping from one storage solution to the next. I predict more

The storage market is booming, and every company has its own take on how to implement a storage protocol. This provides you with a wide variety of storage choices, but it also makes things much more complex.

companies are still using migration tools that have no chain of custody integrity checking.

Even external tools designed to check integrity are rarely used because they add steps to the migration process that are slow and painful, require additional scripts to run, and, when errors are found, require even more scripts to be created—leading to an error-prone and long migration. If any integrity-checking is done, it is mostly to verify the timestamp and size of a file to declare it has been copied correctly. NAS solution providers will trend toward a custom, consultative partnership model. And I advise anyone looking to modernize their data center to seek providers that were specifically designed to handle NAS migration of unstructured data for storage and cloud environments.

Michael Jack is co-founder and vice president of global sales at Datadobi (https://datadobi.com), a provider of unstructured data management software.

-- DBTA LIVE

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

The Evolution of the DBA

Major trends in technology are reshaping the role of the DBA at many organizations. The size and complexity of database systems continue to grow with higher data volumes, more workloads, and an increasing rate of database deployments that need to be managed.

What's more, new data types and emerging applications continue to drive the adoption of new database management systems. As a result, DBAs are under



constant pressure in a steadily evolving environment—handling crises to support the business while also navigating the impact of cloud and automation in their daily jobs.

DBTA recently held a roundtable webinar with Krishna Kapa, senior solutions architect, Nutanix; Peter Albert, CISO, InfluxData; and Srinivasa Krishna, global practice lead and director, MySQL Services, Datavail, who discussed how IT decision makers and database professionals can tackle these changes.

Krishna said DBAs face database challenges in agility and innovation, reliability, cost and efficiency, scalability, and security. Cloud is the future, Krishna said, and, as a result, DBAs need to understand the following technologies:

- Cloud ecosystems
- Hybrid cloud and multi-cloud approaches
- IaaS/PaaS/DBaaS/serverless offerings
- Monolithic versus microservices architectures
- Relational, big data, caching, graph, ledger, and time series systems

- How to choose the right database
- Database containerization

DBAs must align goals for the business as well as their future careers, identify skills that need improvement and get training, and master the cloud, Krishna said.

Albert said there are now more databases than ever before. In addition, cloud and the digital supply chain have altered the way DBAs collect and protect information in a database.

Routine day-to-day administrator tasks include provisioning new environments; doing database cloning or refreshing; backup and recovery; performing patching, migrations, and upgrades; and handling tuning and capacity management. It can take hours, days, or weeks to do this with a traditional database provisioning/cloning process, Kapa said. The process involves multiple teams, introduces friction and complexity, lowers business agility, decreases the innovation rate, and increases time-to-market. With a focus on DBaaS, DBAs can gain elasticity, scalability, high availability, security, the ability to migrate to different platforms, and more.

Harmonizing Database Management With DevOps

Even the best DevOps initiatives can fall victim to data bottlenecks. These include challenges in provisioning data quickly and easily, synchronizing application and database changes, and, in some cases, a serious lack of cooperation between developers and database administrators. The speed and flexibility requirements of DevOps and modern applications overall can be difficult to achieve in the world of data management with its rigid schemas, manual processes, and data silos.

DBTA recently held a roundtable webinar with Corey Brune, manager, solution architecture, Delphix; Matthias Zieger, pre-sales engineer, Central Europe, DACH, and Eastern Europe, Digital.ai; and Grant Fritchey, Microsoft Data Platform MVP and DevOps advocate, Redgate, who discussed key technologies and strategies for bridging the gap between data management and DevOps.



Enterprise data drives business and technology initiatives, Brune said. Speed and scale of data delivery is critical in these endeavors and is a recurring problem, he explained. The status quo consists of a patchwork of point solutions and manual processes. However, he noted, there is a better way. The Delphix DataOps platform combines enterprisewide data coverage with data compliance, according to Bruno. The platform provides automated, self-service access to a test data library along with securing data management in the delivery pipeline.

Database DevOps/DB automation is still an obstacle for many teams across the industry, Zieger pointed out. In a lot of cases, databases are still not part of continuous integration/continuous deployment pipelines because databases are stateful by definition, and there are challenges with data integrity, staging data development to production, and test data generation. What is needed is reduced time from development to production through the use of database deployment/release automation. Companies should also align database deployments with app deployments in all stages and validate database changes, Zieger said.

According to Fritchey, to improve database deployments, it is necessary to adopt development methods. Organizations should build a database continuous integration process and put the database into continuous delivery mode, which enables faster deployments, enhanced safety of



production, more testing, scaled support, and better compliance

Balancing Data Integration With Data Governance

The proliferation of data sources, types, and stores is increasing the challenge of combining data into meaningful, valuable information, while the need for faster and smarter data integration capabilities is growing. At the same time, to deliver actual value, people need information they can trust—now more than ever during this COVID-19 pandemic—which makes data governance absolutely essential.

DBTA recently held a webinar with Quinn Lewis, consulting director, Denodo; Keith Lambert, VP, marketing and business development, Kore Technologies; and Ginger Gatling, senior director, SAP, who discussed key technologies and best practices for overcoming data integration and governance challenges.

According to Lewis, current challenges in data integration and governance include faster and more complex demands for decision making, the need to reduce the cost of data infrastructure, new regulations, and requirements for enterprisewide governance and data security.



Companies can balance these challenges with a logical data warehouse (LDW), Lewis said. The LDW is a next-gen data warehouse that uses a multi-engine approach to fulfill conflicting demands. Rather than seeing its components as competing solutions, they are complementary engines. LDWs enable faster and more effective data delivery, increased flexibility, improved decision making, maximized ROI, and greater responsiveness to big data and IoT, Lewis said.

Lambert suggested the top must-haves for overcoming data integration challenges.

The first is adopting a set of best practices which includes defining your data model and naming standards, creating a data flow diagram, building a source-agnostic integration layer, adopting a data warehouse architecture standard, and considering an agile data warehouse methodology.

Additional key capabilities for overcoming data integration challenges are multi-source and database aggregation, point-in-time snapshots, and incremental database updates. The other must-haves are automation with message-based architecture, detailed transaction logging, the ability to analyze and profile data sources, template-driven ETL software that is easy to change, a development environment and tools, and a technology partner with reliable and flexible software.

Advanced analytics and machine learning challenge data management to a new degree, Gatling explained. "Accept that you will have to use multiple tools and processing engines, but be sure to build a strong metadata foundation to ensure data understanding, quality, and standard catalog and glossary," she said.

Going Deeper With Data Science and Machine Learning

Surviving and thriving with data science and machine learning means not only having the right platforms, tools, and skills, but also identifying use cases and implementing processes that can deliver repeatable, scalable business value. However, the hurdles are numerous and include the selection of datasets and data platforms, architecting and optimizing data pipelines, and model training and deployment.

In response, new solutions have emerged to deliver key capabilities in areas including visualization, self-service, and real-time analytics. Along with the rise of DataOps, greater collaboration and automation have been identified as key success factors.

DBTA recently hosted a special roundtable webinar featuring Alyssa Simpson Rochwerger, VP of AI and data, Appen; Doug Freud, SAP platform and technology global center of excellence, VP of data science; and Robert Stanley, senior director, special projects, Melissa Informatics, who discussed new technologies and strategies for expanding data science and machine learning capabilities.

According to a Gartner 2020 survey of



CIOs, "only 20% of AI projects deploy," Rochwerger said. The top problems are staff skills, understanding the benefits and uses of AI, and data scope and quality.

She said businesses need to start out by clarifying a goal so they can understand where the data is coming from. Once organizations know that, they can find and fill in the gaps.

According to "Data 2020: State of Big Data Study" (SAP/Regina Corso Consulting), 86% of companies aren't getting the most out of their data, and they are limited by data, complexity and sprawl, Freud noted. SAP Data Intelligence can meet companies in the middle, Freud said, stating the platform's enterprise AI meets intelligent information management.

Stanley added to the discussion by introducing the concept of data quality fundamentals with AI. Using AI (machine reasoning and machine learning), more efficient methods for identification, extraction, and normalization have been developed. AI on clean data enables pattern recognition, discovery, and intelligent action. Machine reasoning relies on knowledge captured and applied within "ontologies" using graph database technologies, he explained. Machine reasoning can make sense out of incomplete or noisy data, making it possible to answer difficult questions. It also enables highly confident decision making by applying existing knowledge- and ontology-enabled logic to data, Stanley noted.

MV SOLUTIONS

Rocket Software Beefs Up Security in Latest D3 Release

Rocket Software is releasing Rocket D3 10.3.2 with an assortment of security, performance, and integration enhancements. Version 10.3.2 has encryption enhanceAdditionally, encryption at the field level, rather than at the file level, offers a performance benefit, especially for large files. The architecture allows D3 to utilize more than 4GB of memory, taking advantage of greater addressing space and



ments plus SSL connectivity so users can securely share files. For encryption, 10.3.2 provides tools to secure both data at rest and data in transit.

Additional security updates include the following:

- STARTTLS support
- Encryption of hot backup, OSFI, and MVSP API connections
- Encrypt/decrypt functions in BASIC
- SSL connection functions in BASIC for socket applications
- SSH for D3/Windows and host authentication for all D3 platforms to increase the likelihood of passing security audits.

There are additional security features:

- Support for A/D, LDAP, PAM, etc.
- Better password management around strength and expiration
- User lockout after a specified number of failed attempts

Version 10.3.2 contains a variety of enhancements across the data server engine for a better end-user experience. Additionally, there are integrations with Red Hat Linux, Rocket MV Integration Server, and Windows Server 2019.

An MV workstation with access to D3 was also added to allow remote workers to print from their home or field office. With the widespread shift to working from a home environment, this update is very important, said Brian Cram, technical support engineer, Rocket Software. "User- defined functions is also a big update." improved file caching in memory.

D3 10.3.1 allows BASIC developers to leverage the Python Open Source library of hundreds of thousands of solutions for quickly adding new features to improve an application. The ability to call into D3 from Python in version 10.3.2 completes the bidirectional support of Rocket D3 Python for better integration, allowing Python developers to access data and existing business rules that are stored in the D3 database environment.

"We have a very loyal customer base and they've been really appreciative of our MVU sessions," said Chris Rizza, product manager at Rocket Software. Customers that have been gravitating toward these virtual educational workshops were requesting Python integration to attract new developers, added Cram.

Zumasys Offers Tutorial on How to Supercharge MV Applications

In a continued effort to help the Multi-Value community evolve their applications using free industry-standard tools, Zumasys has shared a four-part hands-on video tutorial on the PICK MultiValue GitHub repository (a public, open source community for PICK BASIC and PICK BASIC-related projects).

In this "Full Stack with PICK" tutorial, users learn how to supercharge both the front end and back end of an application. Specifically, Zumasys shows PICK development using a modern front-end framework (Vue.js) and RESTful APIs written in PICK (available for D3, jBASE, OpenQM, Uni-Verse, and UniData with MVConnect).

As a bonus, the tutorial includes access to an instance of jBASE in a Docker con-

tainer with a free evaluation license to allow users easy and convenient access to a local development environment.

These lessons are designed to appeal to developers of all types. The videos provide greater understanding of how companies can run jBASE in Docker, build REST APIs in PICK, and build a front-end Vue. js application that can communicate with those APIs.

The following technology components are utilized:

- Docker, a popular and convenient way to run many different environments, all from your local system (for development and even production purposes)
- REST APIs, a universal framework for allowing systems to talk with one another
- Vue.js, a popular modern application framework that marries the best of Angular (led by Google) with the best of React (maintained by Facebook)

Video tutorials cover the following:

- 1. "How to Set Up Demo Data in jBASE"
- 2. "How to Set Up REST API Get End Points"
- 3. "How to Set Up a Simple Vue Environment"
- 4. "How to Display a Database Record With Vue"

BlueFinity Enriches Evoke Platform With Extended MV Integrations

BlueFinity is introducing a series of new functions in its Evoke platform to further help support a wide array of Multi-Value environments. These updates provide access to an extended number of MV databases as well as enable enhanced support for easier integration with MV platforms.

BlueFinity has focused on three primary areas to achieve the high levels of integration required to effectively deploy modern apps across multiple mobile and desktop devices that will work seamlessly with MultiValue databases. These areas are integration with the MultiValue database, regular access to the database, and the assimilation and reuse of existing code in the app.

MV SOLUTIONS >>

Evoke works seamlessly with mv.NET, BlueFinity's own connectivity tool, or, alternatively, any of the available Rocket connection tools, to provide for the interaction with your specific MultiValue database(s).

Evoke will automatically import the data dictionaries from the selected MV database and display the details of the dictionary so that it can be selectively imported into an Evoke app design, resulting in data objects, items, and attributes being fully accessible by Evoke-generated apps.

With Evoke, users can define their data structures within the design in a way that is most suitable and accessible for the proposed app.

Importantly for many users that have both MultiValue and relational databases in their organization, Evoke is also able to support access to both types of database from within the same app.

The level of integration Evoke offers to both MV and SQL environments means that companies that have previously considered a move away from MultiValue can retain their legacy system investment and still provide sophisticated, state-of-the-art apps for their staff, customers, and partners. This puts the emphasis on responding to the requirements of the business rather than limitations of the IT infrastructure. this code to be retained and incorporated into the new app. Evoke provides an infrastructure to do just that.

As part of the CRUD code which is automatically generated by Evoke, Evoke has developed a process of unique BASIC

Importantly for many users that have both MultiValue and relational databases in their organization, Evoke is able to support access to both types of database from within the same app.

Not only that, but if they do decide to migrate to SQL at a later date, the Evoke apps can simply be re-generated for the new environment.

According to BlueFinity, MultiValue applications are generally made up of various types of BASIC routines and logic and the PROC procedure language that have been created over a number of years or even decades. This legacy code is invariably functionally rich and proven software, and it should always be possible for code hooks which allow users to reuse or call any existing or new code (in all its forms, including UniBasic, jBC, Data-Basic, etc.) and PROC from within an app to incorporate it as an integral part of the app.

Evoke is recognized for its flexibility in accessing and managing multiple types of databases; different flavors of SQL Server, Oracle, and Db2; and for ensuring that this technology is fully integrated with all major MultiValue databases.

Enterprise eCommerce & Integration Solutions



Reach the B2B/B2C Market

*Powered by KommerceServer

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



Extend the Enterprise



*Powered by Kourier Integrator

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

Solutions that work. People who care.

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

Quest

Page 18 MODERN DATABASE MANAGEMENT: NAVIGATING OPEN SOURCE, CLOUD AND DEVOPS

Percona

Page 20 HOW IT CAN SCALE AND MANAGE COSTS IN AN ECONOMIC DOWNTURN

Couchbase

Page 21 DATABASE PERFORMANCE TODAY-THE NEED FOR SPEED AND SCALE

SAP

Page 22 COST EFFICIENT IN-MEMORY PERFORMANCE AND SCALE WITH SAP HANA

Rubrik

Page 23 DBAS IN AN EVOLVING AND DEMANDING LANDSCAPE: NAVIGATING NEW CHALLENGES AND EVALUATING SOLUTIONS TO ADDRESS THEM



DATABASE PERFORMANCE TODAY: **THE NEED FOR SPEED & SCALE**



Special October Best Practices Report

Best Practices Series

IMPROVING DATABASE PERFORMANCE FOR THE **GROWING DIGITAL** ECONOMY

What does it take to prepare data environments for the challenges looming in the years ahead? The COVID-19 crisis forced a massive and seemingly overnight shift to digital business. And with it, new requirements have emerged for approaches to data management to support advanced data analytics, AI, IoT, and other ways to compete with data.

The critical role of data as fuel for the growing digital economy is elevating data managers, DBAs, and data analysts into key roles within their organizations. In addition, this rapid change calls for a three-pronged approach that consists of expanding the use of more flexible cloud computing strategies, growing the automation of data environments, and increasing the flow of data and collaboration through strategies such as DevOps and DataOps.

DATA PERFORMANCE INHIBITORS

Recent surveys have shown that data managers at all levels are hampered by

the tedious maintenance requirements of building and managing their environments. This is hurting their businesses overall. Essential database tasks such as collecting, organizing, and storing large quantities of data—much of it in real time—are now beyond the capabilities of human administrators to manage. Areas ripe for automation include operations, repeatability, automated testing, and the release of data.

For starters, database maintenance still takes too much of DBAs' time and this is cutting into organizations' competitiveness, according to a survey of 212 data and IT managers, conducted among members of the Quest IOUG Database and Technology Community by Unisphere Research, a division of Information Today, Inc. ("2020 Quest IOUG Database Priorities Survey").

Even with database automation and cloud resources abundantly available on the market, many database managers still spend substantial amounts of time on low-level tasks. This is holding back progress.

Best Practices Series

Asked which activities are the costliest database management projects from an operational standpoint, two out of three data managers said mundane administrative tasks consume a significant part of their budgets. Maintaining system stability—patching, fixing, and upgrading—is considered the costliest part of their jobs, according to 66%. And 61% indicated that maintaining uptime and availability were also resource-intensive processes. For 49%, security consumes large portions of their time.

Database maintenance also cuts deeply into enterprise competitiveness, the Unisphere survey showed. Data managers in the survey overwhelmingly agreed that the amount of time, money, and resources spent on ongoing data management—versus new projects or initiatives—is affecting the competitiveness of their organizations. More than one-quarter, 26%, said that these expenditures are severely limiting their competitiveness, and another The critical role of data as the fuel for the growing digital economy is elevating data managers, DBAs, and data analysts into key roles wirhin their organizations.

45% said their competitiveness is somewhat inhibited.

CLOUD OPTIONS

While many IT budgets have been on the upswing, cost still is the leading factor in infrastructure purchasing decisions, the Unisphere survey showed. Cloud is seen as the best option for reducing the time and money spent on ongoing database management activities.

There is a notable shift away from on-premise infrastructure. Moving to virtualization or cloud-based solutions tops the list of approaches, cited by 66%. More than half, 59%, indicated they are attempting to address database costs by migrating or updating their databases. More than one-third, 34%, are upgrading or modernizing their hardware and processors. However, adoption of cloud as a backup and recovery environment is still in its infancy.

The shift to cloud-based data and databases has been progressing over the past decade. It's been a gradual move, buffeted by uncertainty over data security and accessibility. An array of choices have opened up to data managers and their organizations—from databases residing in their entirety within thirdparty hosting services, to database-asa-service (DBaaS) solutions offered by both database and general cloud vendors. Within cloud-based data services, underlying functions from provisioning to security are automated at the provider's site, freeing up data managers to focus on higher-level advisory roles within their business. Such cloud-based databases and data support high levels of flexibility and adaptability as organizations grow.

AUTOMATION OPTIONS

Database systems have become complex—too unwieldy to be run manually as they have been in the past. Initiatives critical to corporate growth in the digital economy—AI and analytics require ever-larger volumes of data. Many databases already provide tools that automate tasks, while others still require manual scripting by DBAs.

Those companies that fail to effectively leverage and deploy their data assets will find themselves falling behind. As a result, data managers are turning to automation and autonomous databases. Database functions such as backup and recovery are already highly automated, and plans are underway to automate such day-today functions as monitoring, provisioning, and maintenance. Data managers welcome the rise of automation for these tasks, and see greater roles for themselves in higher-level business decision making.

Support for more automation was also confirmed by a separate survey of 217 database managers and administrators from the Quest IOUG Database and Technology Community conducted by Unisphere Research ("2019 Quest IOUG Autonomous Database Adoption Survey").

A majority of database managers in this survey seek greater automation to assist their application deployment endeavors. Seventy-five percent said automation would help accelerate their efforts, and that applications can be deployed faster with database management automation. More than one in four, 27%, said this acceleration would be significant in speeding up service to their enterprises. Areas that are in need of work include database troubleshooting (only 14% have been able to highly automate this area), along with database tuning (14%).

Within an automated set of processes, a range of database functions can be managed unattended by humans, including testing, security, making changes and updates, and provisioning. Many aspects of database management are now already highly automated. By far, backup and recovery functions are the areas that have seen the greatest levels of automation to date, with close to half of respondents, 48%, citing high levels of automation. Database monitoring is also a highly automated area, cited by 43%.

Enterprises intend to step up their automation of database backup and monitoring processes. Database security will also increasingly be automated as database managers continue to brace for the onslaught of threats that accompany the move to digital. A majority, 77%, believe that turning these tasks over to machines will deliver enhanced productivity and output.

NEW LEVELS OF COLLABORATION

To deliver data quickly and efficiently to where it is needed, a new practice— DataOps—is emerging. DataOps not only syncs the flow of data through enterprises, but also merges the activities of people working with data and enhances this work with automation to ensure the timely delivery of data.

Accordingly, DataOps practices are on the rise across enterprises. A survey of 300 companies by 451 Research found that 72% of respondents have active DataOps efforts underway and the remaining 28% plan to leverage DataOps over the coming year. A majority, 86%, are increasing their spending on DataOps projects over the next 12 months. Most of this spending will go to analytics, self-service data access, data virtualization, and data preparation efforts.

Cloud, automation, and collaboration are supporting organizations' ability to use data more effectively. The ability to deliver data in a rapid and seamless way will pave the way for not only scaling data environments, but also initiatives such as AI and IoT—areas in which the effective deployment of data can have a strategic benefit.

Modern Database Management: Navigating Open Source, Cloud and DevOps

The database landscape has seen significant changes in the past decade, and the next 10 years promises to be equally transformative. Massive data growth continues to present new challenges and opportunities, including how to manage, monitor, protect, analyze and use it. How businesses respond and adapt to these opportunities and challenges will help determine their ultimate success or failure. A strong database management strategy, operated by talented DBAs, is critical.

At the same time, the days of the Oracle or SQL Server shop are over. Open source databases and cloud adoption are on the rise at businesses looking for ways to reduce costs. In addition, the shift toward a DevOps culture is becoming more prevalent. These trends pile onto the already demanding workload of the modern DBA tasked with managing the performance and high availability of an ever-growing spread of diverse systems.

KEEPING AN OPEN MIND ABOUT OPEN-SOURCE DATABASES

The adoption of open-source databases is increasing, as these systems are proving to be low-cost and reliable, and now feature better management functions and support. That being said, there are pros and cons when comparing open-source databases with the commercial players.

The tools open-source vendors provide usually pale in comparison to those of the larger commercial vendors, meaning there will be certain tasks that DBAs will need to be skilled enough to perform. Open-source databases have improved to the point that they can handle some missioncritical applications, but this will not be the case in every scenario. Commercial players typically have the resources to drive innovation at a rate that open-source providers can't match. But open-source options can be considerably cheaper.

Often, a multi-platform approach—in which open source is included—allows an organization to capitalize on all the various pros and overcome some of the cons. However, implementing a multi-platform system can be time-consuming. The software programming has to be able to function consistently on different systems, which takes some work, but that also means you'll be able to reach a wider range of customers once the programming is completed.

The decision to invest in open source will be determined by several factors unique to your organization. You will need to consider the tasks you're performing, the consistency or usability—of the data you're working with, and the performance that your customers need and demand.

GETTING YOUR HEAD IN THE CLOUD

While on-premises database deployments will never fully disappear, the cloud is continuing to take on increasing prominence. The cloud can provide peace of mind to those concerned with database storage and ensuring databases are running on the most up-to-date software. The cloud is also an important backup plan for production databases that support mission-critical applications, high availability needs or disaster recovery.

While the cloud offers many advantages, there are critical decisions you'll have to make to ensure you capitalize on them effectively. Once you have decided to move into the cloud, you will need to consider whether to implement a public or private model—or a hybrid of the two. You will also have to determine whether a Database-as-a-Service model or Infrastructure-as-a-Service model best fits your needs. Are you looking for an outside vendor to take on the bulk of the tasks typically reserved by DBAs or keep activities like upgrades, backups, and monitoring in-house?

There's also the nitty-gritty of deciding whether to maintain your current service-level agreements (SLAs) with cloud vendors and other third-party vendors or create new ones. With new implementations come new moving parts and more layered considerations. Organizations will have to determine for themselves whether they can fulfill their existing SLAs or whether changing circumstances will require changing agreements.

It will also be important to ensure you manage cloud spend, since database workloads typically behave differently than they do in your data center. Make sure you independently assess

Quest

what service tier is the most suitable to balance workload performance with cost to ensure your monthly subscription bills are within budget.

Hanging over all these decisions is the big one: which cloud vendor to choose. There are numerous options out there so organizations have to decide what's most important to them. Reputation might be more important to one company while the offering's alignment with the infrastructure already in place might be the top priority for another. Ultimately, it will be essential to determine the most cost-efficient way for your business to balance cloud with on-premises deployments, and what information and data is stored where.

DEVOPS CONVERGENCE

The development team and the operations team have historically been in conflict on some level. This is because they have different but equally important priorities that require striking a delicate balance. Dev team members build applications and write code, among other tasks, while Ops team members oversee the infrastructure to ensure it is fundamentally sound and to ensure the integrity, security and availability of data to the business. However, many businesses have recognized that they need to reorganize and merge the two functions in order to operate more efficiently so that change can be delivered to the business faster. The convergence of teams creates new challenges to navigate and decisions to make when considering the bottom line.

Application developers and infrastructure teams have clashing priorities when it comes to stability and performance issues. The Dev side is accustomed to being measured by how quickly and reliably they can get code changes into production. They're used to developing code rapidly and getting it into production as quickly as possible, while also performing tasks that require continuous attention, such as deployment, integration, and testing.

The operations team, on the other hand, is devoted to making sure processes run smoothly by constant monitoring their systems and reacting to problems effectively. They're not concerned with fast and continuous change. Instead, the Ops team's main task is maintaining database stability through diagnostics, tuning, and administration. This is vital to your organization because database changes aren't sequential, and certain changes can't be undone, like buggy code, because the data is changing constantly.

At the convergence of Dev and Ops-DevOps-is a balance. As DevOps becomes more widespread, companies are learning the importance of including database teams in these processes to support communication and collaboration. But they're also starting to see the advantages a DevOps approach offers where established automated processes such as continuous integration (CI) and continuous delivery (CD) ensure rapid movement while ensuring the necessary checks and balances and operational oversight still exist. A sound implementation of DevOps CI/CD for database changes ensures your business data and processes are available and consistent throughout your organization, enabling you to make intelligent, cost-effective business decisions. Moreover, a smartly run DevOps strategy that includes the database team will alleviate bottlenecks, so your company can operate smoothly.

ACHIEVING SUCCESS IN THE MODERN DATABASE WORLD

The rate of change in IT continues to accelerate. By anticipating, understanding, and embracing changes in database operations, you'll achieve success. But you don't have to go it alone. With powerful solutions from Quest, you'll maximize your cloud, open-source and DevOps initiatives. A toolset like Foglight for Databases provides unrivaled visibility across all your database platforms, empowering you to easily and proactively ensure peak database performance on legacy databases and new cloud and open-source platforms. And you can do it all from a single, intuitive console. Tools like Toad and ApexSQL enable developers to write high quality code faster, reducing defects and orchestrate their development tasks for CI/CD automation, enabling companies to realize their DevOps ambitions.

As businesses strive to reduce costs, DBAs must find new ways to adapt and thrive in the face of significant IT change. Expanding your knowledge by attending webcasts led by top industry experts, reading expert guides and arming yourself with easy-to-use solutions that improve database performance across a wide range of platforms will empower you to achieve cloud, open-source and DevOps success.

To learn more about how Quest can help, visit: https://www.quest.com/solutions/database-management/

How IT Can Scale and OPERCONA Manage Costs in an Economic Downturn

During economic downturns, companies must start to look for ways to reduce costs, slim budgets, and make the difficult decisions required to minimize the impact on their business.

In this article, we'll talk about a few things businesses can start doing now to help reduce infrastructure and database costs, but still keep things running smoothly. Some you may already be considering, while others may be options to consider should the situation continue to demand action.

FIVE EASY WAYS TO CUT COSTS QUICKLY

1. Scale Down

In the recent past, talk of scaling databases and systems was the hot topic—making sure you could respond to quick growth and customer demands on your business. However, when talking about scaling, businesses either didn't want to consider or simply forgot to consider procedures and options to scale down should business demands drop.

The great promise of the cloud, as a managed service, was to provide elasticity to increase and decrease capacity as demand increased and decreased. The problem is a lot of people take advantage of the flexibility to scale up, but they don't often act quickly to scale down. We have seen customers be able to cut a third from hosting costs simply by "right-sizing" and taking a few key steps:

- Audit Your Systems: Reduce cloud spend by eliminating extra servers that aren't in use. If needed, spin up a backup.
- Shrink Your Footprint: Move to smaller instances to match the reduced workload. Reducing the amount of data you store can lead to big savings. Hoarding digital data is expensive.
- **Consolidate Workloads:** In the past, there would be a desire to split out additional slaves or cluster nodes to process specific types of workloads (i.e., reporting). However, these may not be needed as demand decreases, and can always be added later.
- Consider Alternative Services or Components: Many providers offer long-term storage, serverless, or component options for specific needs with vastly different pricing models. Evaluating whether you have all of the parts of your application in the right setup may yield some positive cost results.

Taking full advantage of the tools, systems, and services you have in place is critical for your business at any time, but especially now.

2. Improve Performance to Reduce Workloads

Another tip, especially if you are using cloud resources, is spending some time tuning and optimizing. Today, many

companies have chosen to scale and tune by credit card instead of optimizing code and tuning underlying infrastructure. Taking a few days or weeks to reduce the workload on your servers can dramatically reduce your overall costs. With proper database management, we have seen some companies cut their bills in half. Calculate your savings with our ROI Calculators.

3. Cross-Train

A good way to reduce costs, and possibly provide longerterm impact, is to invest in your existing employees by helping them expand their skill sets. For example, DBAs helping out with development or system administration duties not only allows those employees to be more well-rounded but also enables them to gain insight and empathy into the challenges and processes that other groups face. In the long run, this can support your organization by creating employees who can step into site reliability engineering (SRE) roles. In the short term, the company delays adding additional costs and, at the same time, is building a stronger team that better understands the touchpoints between business units.

4. Audit Your Toolbox

In the age of Software as a service (SaaS) subscriptions, we all have something we have paid for but don't use as often as we thought we would. These smaller expenses can add up quickly—\$99 per month, per user, with a group of 100 users, that's over \$120k a year! Sometimes it's not a question of whether you need this software, as much as, whether all of these people need access to it. Many companies overprovision the number of users who have access to the software and cutting back on licenses can often cut costs quickly.

5. Consider Flexible Staffing

With the COVID-19 outbreak, no one is sure how long the economy will be impacted, or what impact it will have on different types of businesses. Having the ability to add or remove a workforce as demands or projects come and go can help keep things cost-effective without having a negative impact on your full-time employees. Using flexible staffing during uncertain times has always been an option businesses have used to avoid committing to long-term costs, while still supporting their immediate needs.

Our dedication to the open-source community has not wavered over the past 14 years. No matter your need, Percona can help you configure and tune environments to address all your application and business solutions. Contact us today!

Database Couchbase NOEQUAL Performance Today-The Need for Speed and Scale

AS BUSINESS MOVES ONLINE, NOSQL IS UNMATCHED FOR SPEED AND SCALABILITY

Today's enterprises are interacting digitally with their customers, employees, partners, vendors, and even their products at an unprecedented scale. At the heart of these interactions area company's cloud, mobile, social media, big data, and IoT applications. Because most of the data used by these applications is unstructured, a flexible and schemaless data model has become increasingly critical as companies continue to move more of their business online every day.

The traditional relational data model relies on rigid adherence to a database schema and normalization of data both fatal limitations for today's dynamic applications. NoSQL however, is expressly designed to meet the requirements of modern applications. As a result, enterprises that move to NoSQL can be rewarded with better performance and availability, easier scalability, greater agility, faster time to market, and lower costs.

Because NoSQL databases are specifically designed for unstructured data, performing read or write operations is faster. NoSQL supports the JSON data model, which developers find far more natural than a rigidly defined relational schema. And operations engineers love the ease of elastically scaling a NoSQL database without skyrocketing costs or all the headaches of manual sharding.

COUCHBASE IS NOSQL WITH NOEQUAL

For modern applications, Couchbase isn't just better than relational databases—it's also designed to be better than other NoSQL databases. Node for node, Couchbase can offer 2x to 20x faster workload performance over its competitors.

Architected like no other database

- Shared-nothing, asynchronous, elastic architecture
- Consistent performance at any scale
- · Always on, globally distributed, edge to cloud
- · Location and deployment agnostic
- Schema flexibility plus SQL
- Built-in replication
- · Workload isolation with multi-dimensional scaling

Unlike other NoSQL databases, Couchbase provides an enterprise-class, multicloud-to-edge database that offers the robust capabilities required for business-critical applications on a highly scalable and available platform. As a distributed cloud-native database, Couchbase runs in modern dynamic environments and on any cloud, either customer-managed or fully managed as a service. Couchbase is built on open standards, combining the best of NoSQL with the power and familiarity of SQL, to simplify the transition from mainframe and relational databases.

ACCELERATE YOUR MOVE TO THE CLOUD WITH A NOSQL DBAAS

As a wide variety of business transactions increasingly move online, many organizations are shifting their IT environments to the cloud in order to accommodate more users and cut costs. Couchbase Cloud[™] accelerates this transition by providing all the speed and scalability of NoSQL as a fully managed, automated, and secure Database-as-a-Service (DBaaS). Couchbase lets you operate across your hybrid cloud environments with just a few clicks. And because you deploy clusters in your own virtual private cloud, you maintain control of your security and access policies.

Couchbase Cloud lets you choose from numerous instance types in order to optimize performance. You get unified cluster management, monitoring, and alerting from a single console. The DBaaS automates deployment, scaling, failure recovery, and upgrades. It provides hybrid cloud migration and backup, plus multi-region support. And the high performance and low TCO of Couchbase Cloud come with transparent licensing options and policies.

Couchbase is everywhere you work, shop, and play

Couchbase customers include industry leaders Amadeus, American Express, Carrefour, Cisco, Comcast/Sky, Disney, eBay, LinkedIn, Marriott, Tesco, Tommy Hilfiger, United, Verizon0, and hundreds more household names.

A LEADING CHOICE ACROSS INDUSTRIES AND USE CASES

Because of its exceptional flexibility, performance, scalability, and high availability, Couchbase is the database of choice across industries and for a wide variety of use cases, including:

Customer 360—Comcast, Equifax, and LinkedIn use Couchbase for user profile, session store, and data aggregation.

Catalog and inventory management—Amadeus, Marriott, and Sky count on Couchbase for media/content catalogs and product/pricing recommendations.

Field service—GE, PG&E, and United selected Couchbase to run their applications for work order management and asset tracking.

IoT data management—BD, GE, and SyncThink use Couchbase for operational dashboarding and device/endpoint management.

To learn more, visit couchbase.com

COUCHBASE

Cost Efficient In-memory Performance and Scale with SAP HANA



By Ryan Champlin, Senior Director, SAP HANA Product Marketing

In these challenging times every organization is looking to see how they can do more with the existing investments they've made, but also continue to innovate by cautiously investing when it will help their business grow. This applies to the world of data management and analytics which underpins the critical business applications that keep the lights on and the decision support systems used to gain business insight to keep a competitive edge.

Today, as the volume of data that businesses manage grows so does the myriad of data silos and different technology platforms that support it. It started with traditional databases, then Hadoop,

The best-in-class performance of SAP HANA is well known, but what's less known is that it's also incredibly cost-efficient.

and continues today with various flavors of cloud storage. This has significantly increased complexity with data scattered across multiple data centers as well as technology and application boundaries—ultimately slowing down a business's ability to derive value from their data and improve how their business runs.

It's important that data management solutions equip IT to quickly adapt to business needs without disrupting current business operations. Leveraging cloud solutions for new projects is common but each organization is on a different journey to the cloud. Critical data may move into new applications or different data stores in the cloud over time. At SAP, we looked for ways to break down some of these data silos and simplify things for our customers, and 10 years ago we introduced SAP HANAO to do just that. SAP HANA delivers an innovative in-memory first architecture which breaks down decades-old data silos that existed because of limitations in legacy hardware and database management systems. SAP HANA enables organizations to renovate IT landscapes at their own pace but without disrupting the data landscape by virtualizing data access. Abstracting complexity from data consumers with a single point of data access enables data to be combined and transformed from anywhere in the enterprise in real-time.s

This single platform, or service, approach helps organizations simplify IT landscapes and reduce the data management burden, lowering overall TCO. SAP HANA provides a seamless way to dramatically scale while optimizing performance and cost using advanced in-memory and data tiering capabilities to leverage memory, disk, cloud, and data lake storage and compute. Organizations can choose to use their preferred infrastructure or leverage a fully managed database (SAP HANA Cloud) on almost any cloud.

Challenging events, such as what we're all experiencing now, create new opportunities and to take advantage of them organizations need to be able innovate without rigid boundaries. SAP HANA's in-memory performance enables SAP, partner and customer applications to support complex analytical workloads directly within operational systems, enabling access to analytics right when it's needed directly within the applications. This includes advanced analytics, such as machine learning (ML), embedded within the business processes enabling new levels of automation and insight without duplicating data and the need for separate ML tool sets. New innovation commonly requires leveraging new types of data beyond relational such as graph, spatial, and document structures. SAP HANA enables easily combining all of these into simple solutions with built-in multimodel capabilities with advanced data types and algorithms.

SAP HANA has enabled many organizations to realize dramatic efficiency gains with improvements in agility, while delivering new innovations to the business.

One such company is Canadian Pacific Railway Limited (CP), a leading freight carrier in North America. They use SAP HANA across their business-critical SAP applications as well as numerous custom applications.

Dean Stoffel, IT Director at CP, states, "The best-in-class performance of SAP HANA is well known, but what's less known is that it's also incredibly cost-efficient. With a 12 to 1 data compression ratio compared to a leading competitive database we decreased our storage costs by 20%."

On top of this, they recently took advantage of SAP HANA being the first major database optimized for IntelÒ OptaneÔ persistent memory by adopting new Cisco infrastructure with Intel's innovative memory technology.

Combined, they have been able to realize the benefits of a 70% consolidation of SAP application instances and through simplifying their IT landscape they have been able to increase their ability to deliver business innovation by 8x.

To learn how SAP HANA can help put the power of real-time data at your fingertips, visit www.sap.com/hana

SAP www.sap.com

DBAs in an Evolving and Demanding Landscape: Navigating New Challenges and Evaluating Solutions to Address Them

DBAs are known for their masterful ability to untangle complex systems—but that's not to say that the massive growth in database types and data and the ever-increasing need for business agility haven't made their jobs more challenging. Given the range of applications DBAs must now support and the acceleration of database and data growth, today's DBAs are managing terabytes, if not petabytes, of data across on-premises and cloud—and these numbers will continue to grow. In fact, IDC analysts estimate that data volume will more than double every two years.

WHAT ARE THE KEY CHALLENGES FACING DBAS?

- Maintaining and managing existing backups: DBAs may be comfortable with building impeccable scripts, but the problem lies in maintaining and managing those across multiple databases, adapting to changes in topology, and ensuring compliance needs of the business are met indefinitely. Thus, DBAs would prefer to offload database backups to a central backup system so that they can focus on running the database to meet the business application needs.
- Managing database sprawl: DBAs are now expected to manage a range of databases with different types of data and varying data sets. With multiple databases to check on regularly, DBAs are at higher risk of making preventable errors.
- Accounting for cloud: Modern DBAs must be experts in managing cloud-based databases, as well as understand how to integrate the cloud with existing onsite operations and data management systems. For example, using a cloud-based system makes the process of backing up a database simpler than creating physical backups and storing them onsite. However, DBAs still have to plan for these backups and implement automation to ensure they happen regularly.
- Serving the needs of secondary users: A typical day for a DBA doesn't go by without a steady stream of requests from secondary users (developers, quality assurance teams, and analytics departments) who are looking to gain access to copies of production data for use cases such as testing, development, and ETL (extract, transform, load) workflows. The DBA needs to juggle managing the needs of different departments with protecting mission-critical databases serving the business.
- Backup admin and DBA divide: Backup administrators often handle the complete protection lifecycle of an organization's applications from the backup and restore to the compliance and governance policies set forth by the business. While a DBA may be open to delegating backup tasks to the backup administrator, the responsibility still falls on the DBA to build and maintain scripts as the database environment changes.

There is a high chance of human error due to multiple stakeholders being involved.

WHAT SHOULD DBAS CONSIDER WHEN EVALUATING A SOLUTION?

Third-party data protection solutions can save DBAs time, mitigate risks, and empower them with control they need.

When evaluating data protection services, DBAs should consider the following:

DOES THE SOLUTION OFFER AUTOMATED DISCOVERY AND PROTECTION?

For organizations looking for a hands-off approach to database protection across their IT environment, they need automated discovery and protection of databases through a backup service and a systematic approach to supporting compliance with SLAs. After an SLA policy has been mapped, databases will be automatically protected, replicated, and archived as they are discovered, freeing up management time to focus on other strategic initiatives.

DOES THE SOLUTION OFFER FLEXIBLE RECOVERY OPTIONS?

When it comes to backup and recovery, DBAs often take great interest in managing recoveries, which can be a delicate process. What are the range of options the solution offers in the event of a database failure? Do those align well with the organization's needs and business model? Ideally, in a recovery situation, DBAs are freed to devote their precious time to other demanding tasks without worrying about database backups or relinquishing restore control.

DOES THE SOLUTION OFFER THE REQUIRED LEVEL OF ACCESS TO SELF-SERVICE?

Does the solution enable access to complete a variety of tasks, such as instant recovery, testing a patch or an upgrade, verifying data recoverability, running point-in-time queries and historical reports, and even just meeting ad hoc requests?

With the growing complexities DBAs face, equipping them with the tools they need to succeed is more critical now than ever.

About the Authors:

Justin Ruiz (Senior Product Marketing Manager) and Sarah Tarter (Global Campaigns Specialist) are with Rubrik, a company just named a leader in Gartner's Magic Quadrant for Data Center Backups and Recovery Solutions.



info@iri.com iri.com/voracity 1-800-333-SORT



BRINGING YOU THE WORLD OF DATA MANAGEMENT

An Insatiable Appetite for Data

READY TO MAKE WORK FOR YOU?

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.

SUBSCRIBE TODAY! dbta.com/subscribe

121

266

APPLICATIONS

Today's Top Tips for Securing Work-From-Home Networks

By Don Boxley Jr.

COVID-19 is far from over, and consequently, many people find themselves facing extended work-from-home (WFH) timelines. Tech companies, such as Twitter and Square, which in May announced the "work from home forever" option—and Facebook, which CEO Mark Zuckerberg has stated may have as many as half of the company's employees working remotely in the next 5–10 years—are leading the WFH charge.

While working from home has its share of detractors, WFH offers benefits besides health safety-both from a personal standpoint for individual employees and a business bottom-line standpoint. From the perspective of workers, being able to get their job done remotely allows for a better worklife blend, allowing them to take care of family matters while still ensuring their job gets done, which is particularly important given the current pandemic. From the employer's perspective, WFH has actually been shown to boost productivity. One study found that remote employees pump out three additional weeks of work per year compared with office-based workers. More productive workers translate to a better bottom line for companies.

WFH Challenges

WFH certainly presents challenges as well, which, in addition to a lack of in-person collaboration and the potential for some employees to feel out of the loop, include the possibility that WFH employees could inadvertently unlock the door to full-on corporate network attacks. But with the right security solution, you can avoid this latter challenge by eliminating your network attack surface, thereby protecting both home and corporate office systems and data.

But what is the right security solution to effectively accomplish these goals? One of the more long-standing traditional solutions that attempts to protect data and network security is a virtual private network (VPN). Yet VPNs have a number of disadvantages in the new world of widespread working from home. While the threat of opening your home network, computers, and smart devices to cybercriminals is a terrifying thought, the idea that WFH employees are increasing the risk of corporate network attacks through VPNs is keeping more than one IT manager and chief information security officer up at night.

VPN Solutions

VPNs may still be popular in some companies, but that is based on what they achieved in a world of physical servers and virtual machines, which is not the current model. VPNs worked just fine in the past for on-premise, non-cloud environments, serving up reliable traditional perimeter security. Today's IT reality, however, features more of a hybrid cloud setting, with on-premise mixed with multiple (public and private) clouds. Each additional IT setting increases the risk of data exposure and security breaches, as does having many people working from home and creating additional security vulnerabilities, as homebased users are all treated to a slice of the network, so to speak, and an unprotected attack surface is created.

Security Risks

In fact, that's by far the biggest security problem related to VPN solutions: They create a massive attack surface, as well as a nightmare for administrators to manage. Also, because every third party generally has different networking gear, companies are stuck managing multiple types of VPN connections. VPNs not only bring added maintenance headaches to IT when people are working from home, they also ensure additional security risks will be in the mix, since they expose the entire network to lateral movement.

Alternative to VPNs

There's an updated alternative to VPNs, though: Employing networking software that specifically eliminates these problems. That software is known as SDP, or software-defined perimeter.

The way that SDP solves VPN security issues is by allowing connectivity across multiple clouds, sites, and domains to distributed apps and clients—which is exactly what's needed when practically everyone is working remotely at least some of the time. The result is a truly secure approach that involves application-level access, moving beyond network-level access. This decreases lateral attacks and facilitates an environment that's "secure by default."

SDP solutions set strong limits on remote users, giving them access only to specific services that each user really needs to get their job done. As a result, lateral network attacks become a thing of the past, relegated to VPN solutions.

While WFH is necessary in many cases to help keep workers safe and also offers the advantage of flexibility for employees, its growth requires new IT solutions that help safeguard the company's data. Enterprise VPNs and firewalls may still stick around a while longer, but that doesn't mean they're the right choice for today's work-from-home world. By eliminating your company's network attack surface, you can protect your corporate office systems and data—as well as personal data—for a win-win that employers and employees alike can feel relieved about.

Don Boxley Jr. is co-founder and CEO of DH2i (www.dh2i.com). Prior to DH2i, Boxley spent more than 20 years in management positions for leading technology companies.



KIINGERMAN

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

Important (Oracle) DBA Skills to Have, Obtain, and Nurture

During Quest Forum Digital Event: Innovation Week, Markus Michalewicz, senior director of product management at Oracle, outlined the critical skills that DBAs need to have in the ever-changing IT industry. Employers today look for a broader and more adaptive skill set in their DBAs. Michalewicz's session took a look at skills that any DBA should have and also those that a DBA should look to obtain and nurture, regardless of which new technology is entering the hype cycle.

Soft Skills

Michalewicz started with what may be the most important soft skill of them all-communication. People communicate every day. Whether you are talking with colleagues, writing emails or texts, giving a presentation, applying for a job, etc., it's important to communicate effectively. Michalewicz shared a LinkedIn report, "Skills Companies Need Most," that



cited communication and persuasion as being among the top three soft skills for the next 3 years.

Along with communication and persuasion, some of the other notable soft skills cited in the LinkedIn report were leadership, collaboration, time management, creativity, adaptability, and emotional intelligence.

DevOps

In addition to these important soft skills, DevOps is taking over parts of the DBA work. DBAs have always been in the middle between storage administrators, network administrators, and app developers. Demarcation lines are fading with the expansion of DevOps processes, and DBA roles need to adapt to become part of the DevOps strategy.

If you are an operational DBA, you probably spend a great deal of your time in the areas of maintenance, security, and reliability. Roughly 72% of IT budgets are spent on generic maintenance tasks versus innovation. Additionally, 85% of security breaches occur after the CVE (Common Vulnerabilities and Exposures list) was published, and 91% experience unplanned data center outages.

Autonomous Database

The Autonomous Database means the removal of tactical drudgery and more time for strategic contribution for DBAs. Now, DBAs can focus on architecture, planning, data modeling, and strategy. Efforts can also shift to focus on data security and data lifecycle management. There is also application-related tuning and end-to-end service level management.

Converged Databases

Converged databases are another extraordinary evolution. Phone calls, messaging, cameras, calendars, and music all originally required separate products, but now they are largely features of smartphones. Similarly, key-value, in-memory, analytics, machine learning, JSON, and sharding originally required separate products, but now they are becoming features of a converged database.

There are numerous benefits to a converged database. It supports mixing of workloads, data types, and algorithms. A converged database also enables SQL and transactions across any data type and is simpler, lower cost, and more reliable. In addition, it provides unified security and management across all data and prevents data fragmentation and copy contagion. It also removes initial and recurring integration costs and enables powerful synergies across

features.

Synergy across features makes the whole better than the sum of its parts. Real-time fraud is detectable by converging analytics with online transaction processing (OLTP). OLTP adds flexible data models by converging with document types. Data warehouses achieve high concurrency and security by converging with OLTP, and document stores gain analytics by converging with warehouses.

An Oracle converged database enables simplicity and data synergy and supports all workloads. For DBAs, a converged database means a "transferrable technical skill."

Technical DBA Skills

There are three essential technical skills that Michalewicz emphasized during his presentation: cloud tool skills, networking skills, and security skills.

DBAs need cloud skills because, unlike some other technologies, the cloud is here to stay. You will need to learn new tools and cloud architectures because different cloud infrastructures come with their own tooling and architecture. In addition, DBAs must understand networking because the only way to reach a database in the cloud is the network. Finally, DBAs need to be able to secure the data at all times. To help obtain those skills, Michalewicz recommended looking into Oracle Autonomous Database and cloud skills through data modeling, data security, and application tuning.

Some transferrable on-premise skills that Michalewicz cited as beneficial include Oracle Database, Real Application Clusters (RAC), Data Guard, sharding, and Maximum Availability Architecture (MAA) skills. Infrastructure- and database-tuning skills are also very useful.

In addition to soft skills such as communication and persuasion, three technical skills that DBAs need are cloud tool skills, networking skills, and security skills.

Obtaining DBA Skills

Obtaining a set of strong skills and identifying your strengths are vital to your success. Build on what you can do best, and keep in mind that very few people can do everything. Extend your skill set into areas you are interested in. The easiest way to do this is by looking into skills related to those that you already have. Remember that obtaining a completely new set of skills gets harder later in life.

Michalewicz shared what Jack Ma, co-founder of Alibaba Group, categorizes as the five career stages:

- In your 20s (20+): Join a good company to learn from a good boss
- **In your 30s (30+):** If you want to try to be an entrepreneur, try it!
- In your 40s (40+): Do something that you are good at.
- In your 50s (50+): Spend time training and developing young people.
- **In your 60s (60+):** Harvest the fruits of your labor and play with your grandkids.

Ma says to plan 10 years ahead and adjust the timeframe as needed.

Nurturing DBA Skills

After obtaining your set of skills, you should nurture them. Find areas that interest you and stick with them. However, this does not mean that you must remain with only one thing. Doing what you like will motivate you, make you focus on details, and make you enjoy your work. Consequently, loving what you do will make you better at what you do.

Some general direction and next steps that Michalewicz shared include the following:

- Increase your value as a DBA
- Explore new technologies
- Change your job
 - Consider changing into an emerging job (per country)
 - Know which skills are currently in demand
- Change your field of expertise

- Decide on a new field
- Consider "your stage of life" and potentially adapt that field accordingly

Michalewicz shared a Gartner research report about "hype cycles" that will help DBAs distinguish between hype and what's commercially viable when looking to change their field of expertise. Gartner provided a graphic representation of the maturity and adoption of technologies and applications and how they are potentially relevant to solving real business problems and exploiting new opportunities. The Gartner Hype Cycle methodology gives you a view of how a technology or application will evolve over time, providing a sound source of insight to manage its deployment within the context of your specific business goals.

The hype cycle graph walks through the five different phases of a technology's life cycle:

- **Innovation Trigger:** A potential technology breakthrough kicks things off. Early proof-of-concept stories and media interest trigger significant publicity. Often no usable products exist, and commercial viability is unproven.
- **Peak of Inflated Expectations:** Early publicity produces a number of success stories—often accompanied by scores of failures. Some companies take action; many do not.
- **Trough of Disillusionment:** Interest wanes as experiments and implementations fail to deliver. Producers of the technology shake out or fail. Investments continue only if the surviving providers improve their products to the satisfaction of early adopters.
- **Slope of Enlightenment:** More instances of how the technology can benefit the enterprise start to crystallize and become more widely understood. Second- and third-generation products appear from technology providers. More enterprises fund pilots; conservative companies remain cautious.
- Plateau of Productivity: Mainstream adoption starts to take off. Criteria for assessing provider viability are more clearly defined. The technology's broad market applicability and relevance are clearly paying off.

Should you make an early move? If you are willing to combine risk-taking with an understanding that risky investments don't always pay off, you could reap the rewards of early adoption.

Is a moderate approach appropriate? Executives who are more moderate understand the argument for an early investment but will also insist on a sound cost/benefit analysis when new ways of doing things are not yet fully proven.

Should you wait for further maturation? If there are too many unanswered questions around the commercial viability of an emerging technology, it may be better to wait until others have been able to deliver tangible value.

To learn more about important DBA skills, check out Michalewicz's Quest Forum Digital Event: Innovation Week presentation at https://bit.ly/3gYy6QV. ■



more than 20 years of IT experience and is a Microsoft Certified Master, VMware vExpert, Microsoft Certified Trainer, and a 10-time How Microsoft and Citrix Are Microsoft Data Platform MVP. Enabling the Future of Remote Work

Citrix and Microsoft have a complicated history, but they've recently taken their relationship to new heights by announcing an extension of their partnership to help businesses better manage a remote workforce.

Since the pandemic began, CIOs have been looking for ways to bring the office experience to the home and in the process put their enterprises anywhere. At the same time, CFOs are wondering just how to do so with the evolution of OpEx subscription options.

The recently announced collaboration between Microsoft and Citrix is meeting those needs.

Reducing Complexity Across Remote Workspaces

The pandemic has caused entire workforces to go remote. It has left IT teams scrambling to support the businesses while ensuring the newly formed remote workforces have consistent

The Microsoft-Citrix partnership and entry into the Azure marketplace exemplifies a transformation in how SaaS is becoming part of this new normal.

and secure ways to access desktop environments.

The new partnership aims to offer better desktop standardization for remote user endpoints, something more critical now than ever. Now each remote worker has an independently unique network experience-something which is traditionally not easy to connect to in remote environments.

When an employee, customer, or supply chain partner is using technology supplied by their employer, IT teams have some control over the commonality of CPU, memory, and other aspects affecting the performance of applications. However, with the rapid transformation to remote work, many employees are using machines that IT teams can't directly service.

Citrix specializes in normalizing the interface, creating a consistent service abstraction from endpoints, and reducing the complexity of IT management by enabling solutions to offer the same, familiar desktop across a number of locations.

But while reducing complexity across IT endpoints is a critical part of our new normal, so too is adapting to increased security concerns.

Meeting New Security Needs

The partnership between Microsoft and Citrix will also improve security governance. Microsoft is selecting Citrix Workspace as a preferred digital workspace, and Citrix is choosing Mic-

> rosoft Azure as a preferred cloud platform. This will help standardize desktop configuration and policies for improved security governance while also reducing the cost of analysis and configuration.

> Additionally, IT teams need established and connected road maps for consistent, flexible, and secure work experiences. Citrix and Microsoft are creating a road map to allow organizations to deliver and manage a larger set of services,

including Citrix Workspace, Citrix SD-WAN, Microsoft Azure, and Microsoft 365. These road maps also launch a fresh copy of the OS, making it harder for viruses and ransomware to gain a foothold.

Monitoring for the Future of Remote Work

For companies migrating to Azure, this partnership should be music to their ears. It makes it incredibly easy to attach Citrix delivery costs to existing Azure invoices, and it provides added scale-out flexibility. Smaller businesses may also be able to use virtual desktops, depending on the complexity of their services and how they match up to what Azure provides. Regardless, it will be imperative to monitor your Citrix environment and, with Azure delivery, equally critical to configure cloud resource monitoring in your dashboards. And there's no reason not to do so; it's easier now than ever before to integrate accurate cloud infrastructure monitoring with mature solutions, making the transition easy. With cloud adoption experiencing exponential growth, it will be important for IT teams to use monitoring tools to better manage costs. Consumption-based billing can shed light on unused resources, sprawl, and the resource-sizing guesswork that often creates unexpected invoice consequences. Organizations need broad monitoring to ensure businesses are only paying for what they're using. Monitoring will help businesses accurately measure how they're using custom apps and deliver better insight into whether they're delivering acceptable user experiences with as few resources as possible.

The "new normal" has IT teams assessing virtual desktop and applications delivery options as never before. The Microsoft-Citrix partnership and entry into the Azure marketplace exemplify a transformation in how SaaS is becoming part of this new normal.







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.

OLDB Paves the Way for Blockchain Databases

For the past few years, database vendors have been busily enhancing their cloud offerings and consolidating the inno-

vations that arose more than 10 years ago from the big data and NoSQL movements. While both NoSQL and big data were enormously influential for database technology, it remains true that the vast majority of databases are running on architectures that are positively ancient in computer science terms.

The most popular databases in use today are Oracle, SQL Server, MySQL, Postgres, and MongoDB. Of these five, four are based on technologies

established in the 1980s—the triumvirate of SQL, ACID transactions, and the relational data model. In reality, very little has changed in database management over the past 30 years.

For about 20 years now, I've kept watch for impending changes in database technology. At this point in 2020, I see three possible paradigm shifts ahead. First, quantum computing stands to turn almost every aspect of computer science on its head—but



of security surrounding enterprise databases, data falsification is still a big problem

Blockchain solves this problem by a clever combination of cryptography and game theory. Participants around the globe are motivated to participate in the blockchain network by a reward system. Entries written to the blockchain are cryptographically signed in a manner that is virtually impossible to undo once written. In the 10-year history of the Bitcoin blockchain, no one has managed to falsify a Bitcoin entry despite the billions of dollars in rewards that a successful falsification could unlock.

Alas, blockchains are not database replacements and so can't be used as the building blocks for applications. Amazon's QLDB (Quantum Ledger Database) is an example of a new breed of "hybrid" systems that attempts to combine the best features of blockchain and database.

In QLDB, data is represented as tables. Each row in the table is cryptographically linked to the previous row, just as in a blockchain,

Blockchain addresses a question in database reliability that we've had no answer for since the emergence of digital storage almost 70 years ago: How do we know that what has been written to a digital device has not been maliciously overwritten?

doesn't look likely to change anything in the near future. Second, if anyone ever develops a persistent storage medium as cheap as disk and as fast as RAM, then every database architecture we have today would need to be revisited—but again, there's no sign of anything imminent. Finally, blockchain represents a significant shift in the way we store data—and it is here now.

Blockchain addresses a question in database reliability that we've had no answer for since the emergence of digital storage almost 70 years ago: How do we know that what has been written to a digital device has not been maliciously overwritten?

Pre-digital storage technologies—paper for instance—had many limitations, but it was hard to alter a record without leaving some sort of forensic evidence. For many years, paper ledgers represented a good enough form of proof for financial records because attempts to overwrite individual entries would leave some trace. With digital technologies, an overwritten record is indistinguishable from the original, and despite the many layers each block is linked to the previous block. Unlike a blockchain, though, the ultimate provenance of any record is guaranteed by Amazon, not by a distributed network. You can think of Amazon as being the witness or notary for the transactions.

However, despite the lack of multi-party consensus, given Amazon's strong market presence, QLDB is gaining traction and validating the general concept of a merger between blockchain and databases. Oracle has introduced a similar blockchain table type in its latest release, and there are some innovative startups offering blockchain databases services—Fluree, BigChainDB, and ProvenDB (my company).

As a true believer in the power of blockchain to revolutionize database technology, I'm hardly unbiased. But with major players such as Amazon and Oracle deploying blockchain databases, and with increasing innovation from startups, it would seem that this is one database technology trend that is not in the very distant future.



The Cloud and Database Administration

Database administration includes responsibility for ensuring the ongoing operational functionality and efficiency of an organization's databases and the applications that access those databases. Traditionally, for the DBA, this has meant becoming an expert about "all things" related to a particular database management system (DBMS), such as Microsoft SQL Server, Oracle, IBM Db2, and others, and that the DBMS was installed on-premise on physical servers. However, cloud computing is changing the traditional role of the DBA as cloud providers deliver some of the management services heretofore delivered by DBAs.

The general idea behind cloud computing is that IT resources

such as computing power, database services, applications, and storage are delivered by a service provider over the internet with pay-as-you-use pricing. Organizations are shifting to the cloud for many reasons, but the most prominent are the promise of cost reduction and improved scalability.

What Will Change

With database-as-a-service (DBaaS), some traditional DBA tasks will be managed by the cloud

provider. DBaaS is a cloud computing managed-service offering that provides access to a database without requiring the set-up of physical hardware, the installation of software, or the configuration of the database. With DBaaS, many maintenance and administrative tasks are overseen by the service provider, freeing up users to quickly benefit from using the database.

For example, most DBaaS providers deliver backup services for cloud databases. This means that some of the manual backup procedures heretofore handled by DBAs are controlled by the cloud service provider. Of course, DBAs are still required to know their business recovery time objectives, understand where failures can occur (both on-premise and in the cloud), and ensure that the backup plans are adequate to meet their company's requirements.

With cloud databases, availability also can become less of a DBA concern because the cloud can replicate databases across multiple geographical locations. Keep in mind, too, that cloud DBaaS providers manage the maintenance, the upgrading, and the application of fix packs to the DBMS, a particularly time-consuming process most DBAs dislike. Nevertheless, DBAs still need to communicate with the DBaaS provider to know when maintenance is being applied and if there could be any impact that would need to be addressed.

Furthermore, the DBA team must comprehend the exact cloud services being provided and verify that they conform to its

usage needs. For example, DBAs need to understand issues such as latency problems, size limitations, and any other considerations that change the way that an organization works with the database.

With these differences, the role of the DBA will likely shift from being more implementation-focused to being more strategic when DBaaS is deployed because many of the routine, day-to-day management tasks can be handled by the cloud provider.

New Skills Required

That said, implementing cloud databases will require that new

skills be mastered by the DBA team. For example, loading a cloud database can be a challenge because of latency issues. But it is not just in the initial loading that latency can be problematic; any data access over the cloud can be delayed due to latency issues. DBAs need to build such expectations into their service-level agreements—both with the service provider and end users.

And let's not imagine that moving databases to the cloud will change the DBA's responsibility for

performance management, security, database change management, and many other typical DBA duties.

There are also many new DBA duties that will arise as organizations embrace cloud computing. One such responsibility is acquiring and maintaining deep knowledge of the cloud architecture. This includes knowing what is in the cloud and what is on-premise—not just for data, but for the entire compute stack as well as how to utilize the components that are in the cloud from an administrative and development perspective.

Working in the New Hybrid Environment

The DBA must also become a vital cog in the budget management process for the cloud database. Just because it is easy to scale up using cloud database services does not mean there is no cost. As a result, the organization needs an expert who understands the impact of additional workloads on the DBaaS contract. Moreover, the DBA must have knowledge of the impact of database and application design and architecture decisions and how they affect costs.

Finally, it is improbable that all of an organization's databases will move to the cloud. That means DBAs must be able to work in a hybrid environment, maintaining their traditional skill set for on-premise data, while embracing and extending their capabilities to manage the cloud data.



Automation in Azure Using Elastic Database Jobs

If you are an old hand at Microsoft SQL Server, you have probably used the SQL Server Agent (i.e., SQL Agent) service and MSDB database for the last 20 years. With a hum-

ble and largely unchanged feature set since it was first introduced, SQL Agent is one of those incredibly useful and cost-saving features we have come to expect from the Microsoft data platform. However, SQL Agent faces one important shortcoming: It can run on Azure VMs and on Azure SQL Managed Instances, but it is not available on Azure SQL Database. As a result, we need a substitute.

I recently had the pleasure of discussing Azure

Elastic Database Jobs (i.e., Elastic Jobs) with Kate Smith and Srini Acharya, program managers in the Microsoft SQL Database product group, specifically for Azure SQL. Elastic Jobs is a new feature set currently in public preview, with an expected general availability date sometime in the first half of 2021. Kate and Srini

Although Elastic Jobs is currently in public preview, the feature set is essentially finished with only incremental changes to come.

are excited about Elastic Jobs and the cool things it can do, such as automatically detecting a newly added or dropped database at execution time and responding accordingly. SQL Agent and other automation products usually expect a fixed set of target databases, and the list needs an explicit refresh through app logic. Elastic Jobs provides this dynamic enumeration at runtime to address such scenarios.

Here's a bit more good news: Public preview might sound as if you have a long time to wait, but it actually means that the feature set is essentially finished with only incremental changes to come. Many Microsoft customers are currently using Elastic Jobs in production and are fully supported through their Microsoft Support contracts. With that behind us, let's dive into the features.

All About Azure Elastic Database Jobs

As a starting point, let's assume you know at least the basics of SQL Agent so we will simply compare it to Elastic Jobs. Many



fully customize the execution sequence of a series of actions. As another example, both Elastic Jobs and SQL Agent can be managed through a set of stored procedures to add, drop, and manage your jobs.

However, there are some key differences between Elastic Jobs and SQL Agent to keep in mind. First, Elastic Jobs supports T-SQL, PowerShell, and REST APIs, as well as API interfaces to the Azure Portal and the Azure Resource Manager. SQL Agent, in contrast, supports only

T-SQL and the SQL Server Management Studio API and SQL Server Management Objects. Second, Elastic Jobs has a broader scope in Azure, running jobs on multiple Azure SQL databases and/or data warehouses as long as they're in the same cloud as the job agent. The targets can be on different servers (including Hyperscale and serverless databases), subscriptions (including Azure SQL elastic pools), shard maps, or regions. SQL Agent runs jobs on a local database server or-using a master serveron multiple target servers.

In terms of its other key features, Elastic Jobs can dynamically enumerate the database in a target group, including or excluding databases that were added or dropped without manual intervention. Target groups can also be customized to include/exclude specific databases, Elastic pools, and/or servers. This is a key differentiator from other solutions in this area, especially in SaaS or other dynamic settings, where additional databases are added to a server/pool frequently.

Elastic Jobs can also support a limit on the number of databases a job runs in parallel for more efficient resource consumption and support APIs and stored procedures so that you can build additional GUI functionality, and it is a fully integrated Azure service.

A Panoply of Options

You also have other options for automation when moving from on-premise SQL Server to Azure SQL Database. For example, depending on your situation, you could use Azure Automation Hub, Azure Runbooks, Azure ARM templates, Azure Functions, and Azure Automation Desired State Configuration (DSC). As mentioned earlier, you can still use SQL Agent on Azure VMs and, with some limitations, on Azure SQL Managed Instances.

With all of those options competing for your attention, just remember that Azure Elastic Database Jobs are designed from the ground up to work with and support Azure SQL Database. They're a natural fit. However, pricing has not yet been set for Elastic Jobs.







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.



Staging Versus Landing Areas in Analytics Environments

Plenty of analytics environments have landing areas, plenty have staging areas, and some have both. So, are landing and staging just synonyms for the same thing? A survey of usage

would show there is much overlap in implementations, and even some confusion.

As the words themselves imply, "land" is a place and "stage" is part of progression. Landing databases serve as the initial dropping off point for incoming data—although, as noted, there are those who choose to blur the lines between a landing area and a staging area.

Purpose of the Database

Usually, structures within the landing database are one-for-one matches of the received sources. When these landing structures are extended beyond the incoming content, such extensions consist of auditing information. Additions may be a column for when the data was loaded, the name of the loading program, or a surrogate key in case the source is remiss in row uniqueness. Perhaps metadata tables are defined to track when cycles ran and the success or failure of loads.

Staging databases are intended for preparing data for further processing. A staging process can have few or many pieces to it, depending on how many steps are needed to get from the incoming "here" to the ready-to-go-out "there." In this preparation role, a variety of table structures may be necessary to hold the varying forms of in-process work. Some tables may hold versions of data to be used for a future cycle's comparison; other tables, being purely intermediary, are truncated and repopulated each cycle.

Variety Within a Framework

Building up an analytics environment involves choices based on a combination of the needs within a specific circumstance and the desires of the architects involved. This last aspect means that not everything is a science. Preferences, individualism, and comfort levels play a part in arriving at a solution to be implemented. However, this is not the Wild West, and every architect is not free to establish brand-new paradigms. Variety within a fairly standard framework is expected as one explores the design components

established across organizations.

Of these choices, decisions about having a landing area and a staging area may arise. If a business has no need for staging processes because the data, as it comes in, is ready for use, then only a landing area may be defined. If everything goes through significant processing—say every data store must be compared to a

previous copy to determine deltas—then a distinct landing area may seem moot. In this case, only a staging area may be defined. It is possible that if a staging area is forgone, in its place, an ETL area may be defined for engineers to use as a sandbox for their processing needs. Another possibility is that an architect-in-charge, who chooses to see no distinction, may lump everything together and simply name things either landing or staging based on the word he or she likes better.

Agreement Is Crucial

If one is building a data warehouse, a data hub, or a data lake, one might have a landing area, a staging area, or both. The chosen preference of defined areas is not a harbinger of design quality. As long as needed functionality exists and has a designed place to live, agreement is the crucial key to making things work. Whether an organization chooses to squish landing into staging, or staging into landing, or—similar to the compartments on the microwave dinners separating the peas and the carrots—have both a functioning landing and staging area, the selection is OK.

The only real requirement is that everyone knows what has been chosen as the design option. Equally important, folks should also acknowledge an understanding of these subtle industry distinctions so that when the team is augmented by newer members who think differently, those new members may be set right without being scolded for thinking "wrong."



YOUR
CONNECTION
CONNECTION
TO THE
NOUSTRY

Database Trends and Applications produces 10 original email newsletters, each with targeted content on specific industry topics. Subscribe today to receive concise reports on what's happening in the data world.

dbta.com/newsletters



database

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 browserbased, 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

GIST