Volume 30 Number 3
JUNE/JULY 2016

The Journal of
Information Integration and Management

database
TRENDS AND APPLICATIONS

COMPANIES THAT MATTER MOST IN DATA

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

Try a little Vitamin Q.

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

ACT NOW

Get a FREE 30-day trial!
www.MelissaData.com/vitaminQ

Solutions for 240+ Countries	10,000+ Customers Worldwide	30+ Years Strong	Data Quality & Mailing Solutions	Cloud • On-Premise • Services

Germany
www.MelissaData.de

United Kingdom
www.MelissaData.co.uk

India
www.MelissaData.in

Australia
www.MelissaData.com.au

MELISSA DATA®
Your Partner in Global Data Quality
www.MelissaData.com | 1-800-MELISSA

The Journal of Information Integration and Management

database
TRENDS AND APPLICATIONS

Volume 30 | Number 3
JUNE/JULY 2016

CONTENTS

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

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

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

Joseph McKendrick,
Contributing Editor; Joseph@dbta.com

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

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

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

Tiffany Chamenko,
Production Manager

Laura Hegyi,
Senior Graphic Designer

Denise M. Erickson,
Cover Designer

Jackie Crawford,
Ad Trafficking Coordinator

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

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

COLUMNISTS

Gerardo Dada, www.solarwinds.com
Guy Harrison, guy.harrison@quest.com
Kevin Kline, Kevin_Kline@dbta.com
Craig S. Mullins, www.CraigSMullins.com
Todd Schraml, TWSchraml@gmail.com

ADVERTISING

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

INFORMATION TODAY, INC. EXECUTIVE MANAGEMENT

Thomas H. Hogan, President and CEO
Roger R. Bilboul,
Chairman of the Board
John C. Yersak,
Vice President and CAO
Thomas Hogan Jr., Vice President,
Marketing and Business Development
Richard T. Kaser, Vice President, Content
Bill Spence, Vice President,
Information Technology

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

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

ONLINE ACCESS Visit our website at www.dbta.com

Searchable archive of all articles with digital document delivery: www.iti-infocentral.com

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

SUBSCRIPTION INFORMATION
Subscriptions are available free to qualified recipients in the U.S. only. Nonqualified subscribers in the U.S. may purchase a subscription for $69.95 per year. Delivery outside North America is $135 via surface mail per year. All rates to be prepaid in U.S. funds. Subscribe online (circulation@dbta.com) 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.

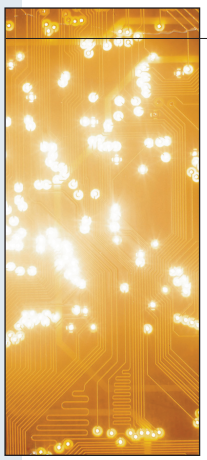
Reprints: For quality reprints of 500 copies or more, call (908) 795-3703 or email reprints@dbta.com.

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

DEPARTMENTS



COLUMNS

MEDIA PARTNER OF THE FOLLOWING USER GROUPS

Database Lifecycle Management Emerges

to Unravel Ever-More Complex Data Sites

By Joe McKendrick

Organizations have directed a lot of attention recently to consolidation, automation, and cloud efforts in their data management environments. This will purportedly result in decreased demand for data managers and the need for fewer DBAs per groups of databases. However, the opposite seems to be occurring. In actuality, there is a growing need for more talent, as well as expertise to manage through growing complexity. A new survey among data executives, managers, and professionals finds that a more challenging data environment is arising due to a confluence of factors.

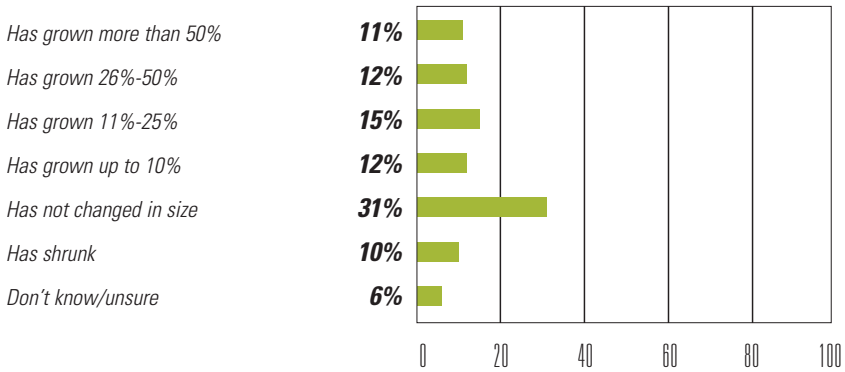
The research, conducted by Unisphere Research, a division of Information Today, Inc., included the responses of more than 300 *DBTA* readers who represent a range of industries and company sizes.

Half of the data shops covered in the survey have grown in size over the course of the last 5 years—some dramatically. Close to one in four respondents reports growth exceeding 25% of their organization's original staff size. By contrast, only 10% of respondents say their staffs have shrunk in size. (See Figure 1.)

What's driving the continuing growth in database staffing? For the most part, companies have been expanding—adding more lines, more services, and increasing transaction volumes. Sixty-one percent of sites experiencing staff growth say the growing volume of business necessitates adding more data managers to their teams. The growth of data itself—exacerbated by big data—is also a contributing factor, cited by half of this group. The rise of new data frameworks, such as Hadoop or data warehouse expansions, is yet another driver among 44% of sites. (See Figure 2.)

In addition, the vast majority, 89%, agree that the complexity of their database environments has increased over the past 5 years. Close to half, 46%, state that their database environments have grown "significantly" or "extremely" more complex during this time.

Figure 1: How has the size of your data management team changed over the past 5 years?



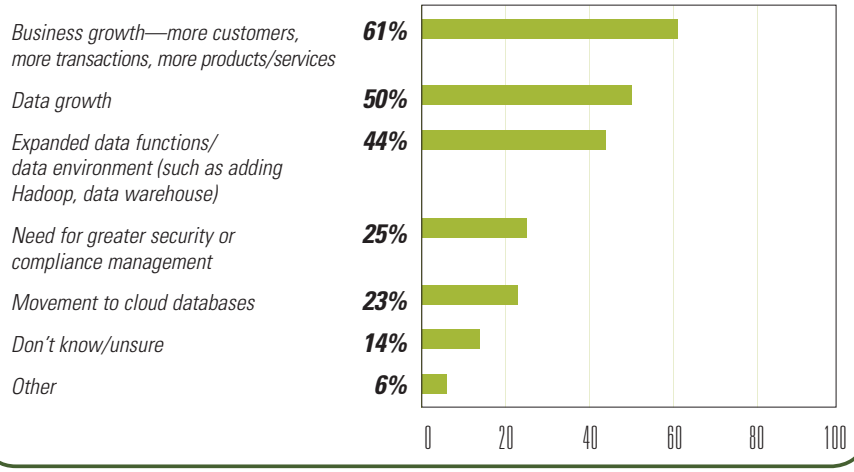
As with the challenges that caused the ongoing growth in database staff sizes, both business growth and data growth are also adding complexity to data environments.

The move to cloud computing, at least for mission-critical enterprise data, will be a slow one. Only 19% of data managers indicate that they intend to move a significant portion of their enterprise data (defined as more than 25% of their total data stores) to a public cloud, while 26% intend to move a significant portion of their data to private or hybrid cloud arrangements.

Are any proactive measures being taken to address this complexity? Virtualization and automation are the top options being adopted by data managers to provide some much-needed simplicity to increasingly heterogeneous environments. The use of management and configuration tools is seen as a way forward for 38% of respondents. About one-third of respondents report they are adopting database lifecycle management (DLM) methodologies to address growing complexity within their data environments. (DLM involves coordinated processes, tools, and people to optimize all aspects of the lifecycle of data, including data architecture and modeling, database design, monitoring, administration, security, storage, and archiving.)

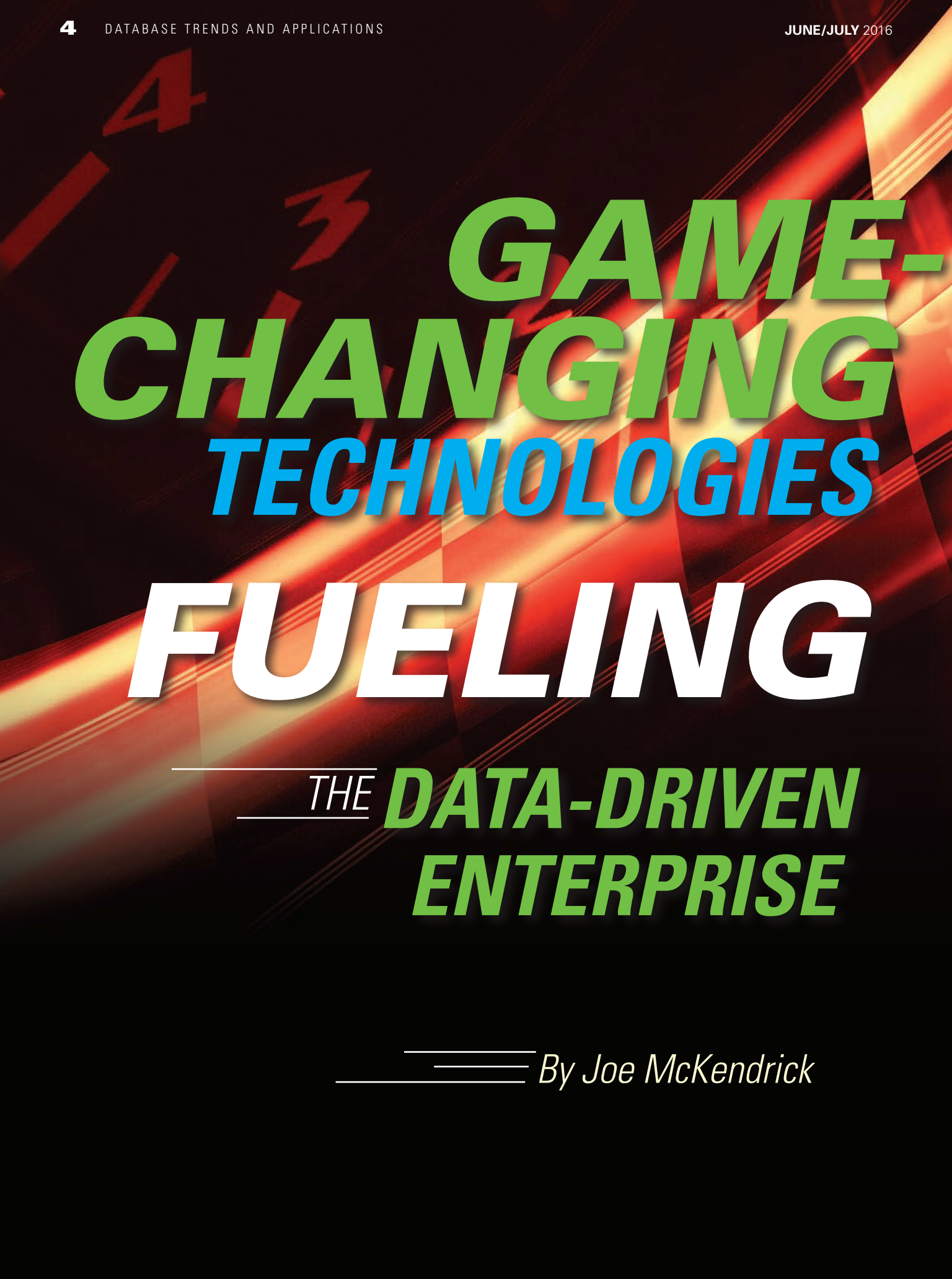
Data managers report a range of tangible business advantages that their organizations are gaining as a result of their DLM efforts. More uptime of data systems is the leading benefit being realized. A majority, 57%, say they have experienced reduced system downtime as a direct result of their DLM engagements. Another 55% of respondents report that their efforts have made data more available to their end users. Confidence in the data itself is also up at 38% of sites.

Figure 2: If your data management team has expanded over the past 5 years, what has driven this growth?



Yet, data managers have also encountered obstacles in their efforts to implement DLM. Half, for example, report that their efforts have been stymied by the need for greater funding or staff time to pursue DLM. At least one-third of respondents to the survey also indicate that their DLM programs do not have as high a priority at other similar initiatives, such as application lifecycle management. A similar percentage of respondents see other impediments, such as a lack of visibility to the issues that may be affecting database performance. ■

*Database lifecycle management methodologies—which involve **coordinated processes, tools, and people**—are emerging to address growing complexity within data environments.*



GAME-CHANGING TECHNOLOGIES FUELING THE DATA-DRIVEN ENTERPRISE

By Joe McKendrick

The data manager now sits in the center of a revolution swirling about enterprises. In today's up-and-down global economy, opportunities and threats are coming in from a number of directions.

Business leaders recognize that the key to success in hyper-competitive markets is the ability to leverage data to draw insights that predict and provide prescriptive action to stay ahead of markets and customer preferences. For that, they need to keep up with the latest solutions and approaches in data management.

What are the key technology developments being seen in data-savvy enterprises this year? There are some familiar ones that still dominate, as well as a new breed of solutions that promise to disrupt the data center as we know it.

12 Key Technologies Gaining Ground

Here are 12 of the key technologies turning heads—or potentially opening enterprise wallets—in today's data centers:

Data Lakes

Data lakes provide a way for organizations to rapidly ingest and store data assets from any and all sources as they arise, without the immediate need for integration or extract, transform, and load solutions. In today's rapidly evolving analytics and business intelligence space, users are getting comfortable with the notion of asking questions of data that have never been pondered before. With a vast set of data now quickly accessible, new forms of querying and reporting can be constructed.

By deploying data lakes, data can be held at the ready for indefinite periods as analytical applications evolve. A survey of 385 data managers, conducted by Unisphere Research, a division of Information Today, Inc., finds that 20% of organizations are implementing data lakes, and another 51% report they are researching the approach ("Data Lake Adoption and Maturity Survey Findings Report," October 2015).

The challenge for data managers will be ensuring adequate governance over the data assets being maintained within the data lake.

Virtual Reality

Virtual reality, or VR, is not a core part of enterprise culture yet, but look for business-centric implementations to begin hitting the market soon. The potential use of VR for a range of functions—from surgery to parts repair to building maintenance—is vast.

Making VR possible at many levels is the real-time deployment of data to add context and meaning to projects. For example, a technician viewing a malfunctioning machine part may be able to view, in 3D, instructions and parts needed to address the problem.

The challenge for data managers will be identifying the areas likely for adoption by the business to enable the intelligence of VR implementations.

Cloud

Cloud computing has been part of the enterprise scene for several years now and will continue to offer a way to scale large datasets on demand. There are various applications that provide immense resources to data shops, such as backup as a service, or database as a service (cited as a separate category, below).

In addition, most analytics and business intelligence tools are now provided through the cloud or software as a service, providing ways to quickly access insights. Salesforce, one of the most common cloud application suites, has accelerated these capabilities with its wave analytics platform, which can provide insights on both Salesforce and non-Salesforce data.

The challenge for data managers is ensuring data security as applications and functions move to the cloud.

Database as a Service

Database as a service, also known as DBaaS, offers a way for enterprises to maintain and update back-end technologies as well as integrate data from multiple, changeable sources without the need to rewrite the applications that depend on them. It also provides an approach for making data readily accessible to end users who need it regardless of the device they are using.

A recent survey of 300 members of the Independent Oracle Users Group (IOUG), conducted by Unisphere Research, finds cloud computing is part of the mainstream of enterprises and, along with it, comes a growing interest in DBaaS as a viable approach to serving their enterprises. The survey finds organizations are employing a range of new strategies and approaches to improve the speed of data delivery and integration ("Database as a Service Enters the Enterprise Mainstream: 2016 IOUG Survey on Cloud and Multitenant Strategies," March 2016).

DBaaS isn't just gaining a foothold in enterprises—it is expected to take off significantly, with adoption nearly tripling over the next 24 months. In the next 2 years, 73% of managers and professionals expect to be using DBaaS within their enterprises, versus 27% who are doing so at present.

The challenge for data managers will be providing a consistent and robust environment from which end users can access any desired data source or application.

Artificial Intelligence and Machine Learning

Artificial intelligence and machine learning are poised to enter the enterprise mainstream, promising to provide self-healing and management capabilities for complex systems, ranging from climate-control to internal IT infrastructures. This is where data analytics plays a critical role, as these systems rely on algorithms that can provide both predictive and prescriptive analytics against data streaming through enterprise systems.

The challenge for data managers will be in applying these capabilities in environments that will best augment the strengths of their employees.

Internet of Things

The ability to read and analyze streams of data coming in from devices, sensors, and applications spread across the globe will change the relationship between organizations and their customers. Estimates of the number of connected devices, sensors, applications, and embedded systems across the globe run as high as 6 billion, with an approximately 5 billion-plus new "things" coming online every day.

The amount of data being generated by these connected endpoints is significant, and it is creating new requirements and ways

of configuring data systems. The Internet of Things (IOT) promises to greatly enhance customer relationships, as enterprises will remain connected and capable of upgrading and maintaining the products they ship, in real time, long after the initial sale.

The challenge for database managers will be to shift from historical reporting and analytics to real-time analysis.

Real Time

As IoT gains hold, the delivery of real-time data to assess real-time problems or predict imminent complications will soon become commonplace. Real time has become a vital part of the data integration equation. Fifty-seven percent of managers and professionals in a recent Unisphere-IOUG survey state that there is now strong demand for delivery of real-time information within their organizations. Less than one-third of respondents, however, say they are capable of delivering most of their data in real time at this time ("Moving Data at the Speed of Business: 2016 IOUG Survey on Data Delivery Strategies," February 2016). The challenge for data managers will be to work with the business to identify what parts of the enterprise or which datasets should be configured for real-time delivery.

Hadoop and Spark

Hadoop and Spark are part of the open-source wave that continues to offer cost-effective and useful capabilities to enterprises of all sizes. Being open source environments, they provide a cost-effective and highly scalable means to package and deploy large and varied datasets. While there has always been some form of "big data" in existence—"big" is a relative term—organizations have long had networks of sensors, devices, embedded applications, remote systems, and log files providing significant pools of streaming data.

Prior to Hadoop, the capture and analysis of datasets of any kind required proprietary tools, and was an expensive and resource-intensive undertaking. Hadoop—and the open source ecosystem that accompanied it—has made big data analytics an extremely cost-effective option within reach of everyone.

The challenge to data managers will be acquiring the skills needed to build out these open source environments.

NoSQL and Cloud-Based Databases

Databases that can be quickly deployed and accessed for online functions are enabling an agility unseen within many large enterprises, while providing growth platforms for small to medium-sized businesses. A survey of 300 data managers and professionals, conducted by Unisphere Research and sponsored by Dell Software, finds NoSQL technology is being used or being deployed at 21% of enterprises.

Another 19% of respondents say they plan to implement NoSQL within the next 1–2 years. One-third of respondents expect NoSQL to have a significant impact on their database operations in the next 3 years. Hadoop is being used or is being deployed among 20% of respondents' companies ("The Real World of the Database Administrator," March 2015).

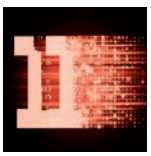
The challenges for data managers are integrating these new forms of databases with existing relational database management system environments and enabling the seamless transfer of data between the two.

Virtualization and Modernization Solutions

The ability to plug legacy systems—and their silos of data—into evolving big data analytics networks will unleash significant amounts of untapped enterprise information. The modernization wave keeps accelerating, with the deployment of virtualization solutions and approaches to enterprise systems that abstract the underlying legacy systems within an accessible service layer.

A recent survey of data managers finds that close to half, 45%, report they have virtualized their data environments. In addition, close to two-thirds, 62%, report that more than half of their mission-critical data is included in their virtualized environments ("2015 IOUG Data Protection and Availability Survey," August 2015).

The challenge for data managers is addressing the performance issues and additional complexity that arises as virtualization expands across their enterprises.



Data Discovery

Data discovery opens up business intelligence and analytics to business end users through visualization and easy-to-use interfaces. The emerging technology reorders and reorganizes the way data is managed, and provides a great deal more flexibility than traditional BI and analytics tools. Data discovery is inherently designed and built for users, enabling them to sift and sort through datasets.

The challenge for data managers is maintaining the integrity and security of data before it is streamed to business users' workstations.

Blockchain

Blockchain—which is essentially a distributed general ledger maintained in a highly distributed fashion across the internet—is most commonly associated with the Bitcoin virtual currency. But ultimately, it can function as a global database for a range of applications and interactions, from all manners of finance to maintenance agreements. Blockchain may be in its nascent stages, but it has the potential to disrupt many aspects of data management and information technology, as it is community maintained and verified.

The challenge for data managers at this point is to learn how blockchain works, and probe vendors to see what progress is being made in making this a mainstream solution.

The Changing Enterprise

Cost management is still a driving concern of data executives, managers, and professionals. However, attention has also turned to being able to deliver new capabilities which reach customers in new ways, and delivering new forms of value to organizations. ■

KMWorld 2016

REGISTRATION OPENS SOON!

NOV. 14–17

JW MARRIOTT

WASHINGTON, D.C.

- **GAIN** Knowledge to Build Strong Organizations
- **COLLABORATE** With Experts & Colleagues
- **EXPLORE** Social Media & Networking for the Enterprise
- **LEARN** About Leading-Edge Practices
- **IMPROVE** Taxonomies & Content Management
- **STREAMLINE** Information & Knowledge Flows

HACKING KM

PEOPLE, PROCESSES, & TECHNOLOGY

Includes Enterprise Search & Discovery 2016 | SharePoint Symposium | taxonomy Boot Camp

Diamond Sponsors

brightstar | coveo | SEARCH TECHNOLOGIES | synaptica knowledge organization systems

Media Sponsors

BESTSEOs | BusinessWire | CMS WiRE | CrowdReviews | InformationToday | KMWorld | Online Searcher | TOPSEOs | Visibility

Organized and Produced by

Information Today, Inc.

Connect with us!

#KMWorld

kmworld.com/conference

DBTA 100

The Companies That Matter Most in Data

By Joyce Wells

The IT landscape is always shifting and being contoured by external market forces and internal industry initiatives.

New requirements for data management and analysis emerge and, in turn, fresh approaches and technologies are developed. Vendors grow and mature and sometimes outstrip others, while new startups take root.

Against this changing backdrop, each year, *DBTA* presents a list of 100 companies that matter in data, compelling us to pause and reflect on the market changes taking place.

For example, with Hadoop marking its 10th anniversary in 2016, it is now clear that while it is still in early stages of adoption, it is a technology that is here to stay and is also evolving beyond its initial role as a platform for batch processing to include a large network of interwined open source projects.

Despite its name, it is also apparent that NoSQL database vendors recognize the value in SQL as a bridge from the RDBMSs that have reigned in the IT world for decades.

And, all the while, relational and MultiValue database systems—upon which many critical applications have been built—continue to adapt and grow.

With a growing variety of platforms to store and leverage data, there is related movement to overcome data silos with integration and data quality initiatives to enable accurate data to be shared more rapidly among a greater swath of users for better decision making. Supporting that goal, cloud platforms are emerging to work alongside on-premise approaches, giving companies the leeway to focus more on their market-differentiating characteristics rather than IT. And, in the wake of this expanded data movement and user access, more sophisticated security techniques are also being advanced to address new threats to enterprise data security and governance.

In all, a lot is changing. This fourth annual *DBTA* 100 list encompasses a wide array of companies that in myriad ways are addressing market demands with hardware, software, and services. Some are long-standing companies with well-established offerings that have evolved over time, while others have sprouted up more recently with new approaches as the combined forces of big data, security, cloud, and analytics have made an impact.

In addition, in this issue, we include "View From the Top" articles penned by company executives explaining how their organizations uniquely approach today's data challenges.

We encourage you to learn more about these companies by visiting their websites. And, to stay on top of the latest news, IT trends, and research, go to www.dbta.com.

DBTA 100 THE COMPANIES THAT MATTER MOST IN DATA

Actian Corp.
www.actian.com

Helping to transform big data into business value for organizations of all sizes, Actian provides an analytics platform that delivers high performance on off-the-shelf hardware to overcome the technical and economic barriers to leveraging big data.

Aerospike, Inc.
www.aerospike.com

Providing a high-performance NoSQL database that delivers speed at scale, Aerospike is purpose-built for the real-time transactional workloads that support mission-critical applications and enable decision making for industries such as financial services, adtech, and ecommerce.

AgilData
www.agildata.com

With a vision of enabling an agile approach to big data processing and data modeling, AgilData delivers a platform that provides high-performance stream processing to allow organizations to quickly and efficiently gather meaningful insights from fast data.

Alphabet
https://abc.xyz

Providing an array of products and services, such as Google Maps, YouTube, Chrome, and Android, Alphabet is a collection of companies, the largest being Google, which in 2015 became a wholly owned subsidiary of Alphabet.

Alpine Data Labs
http://alpinedata.com

Alpine Data provides a platform for advanced analytics to enable organizations to manage the analytic lifecycle in one environment and support people in building, deploying, and consuming analytic applications and insights in an agile and collaborative manner.

Amazon.com, Inc.
www.amazon.com

Through Amazon Web Services, Amazon.com provides compute, storage, database, analytics, application, and deployment services that support a variety of workloads, including web and mobile applications, IoT, game development, data processing and warehousing, storage, and archive.

Attunity Ltd.
www.attunity.com

Attunity offers big data management software solutions, providing capabilities such as data replication, data flow management, test data management, and change data capture to help organizations access, manage, share, and distribute data across heterogeneous enterprise platforms, organizations, and the cloud.

BackOffice Associates/ HiT Software
www.hitsw.com

HiT Software, Inc., a BackOffice Associates, LLC Company, is a provider of change data capture and data replication products for heterogeneous database environments, and BackOffice Associates is a provider of information governance and data modernization solutions.

Basho Technologies, Inc.
www.basho.com

The Basho Data Platform, which integrates Riak KV, Riak TS, and Riak S2 with Apache Spark, Redis, and Apache Solr, simplifies distributed big data and IoT applications to enable high availability, scalability, and operational simplicity.

BDNA
www.bdna.com

BDNA addresses the challenges of IT asset management (ITAM) by collecting detailed information about IT assets from a range of data sources and delivering it to ITAM solutions for reporting and analysis.

THE COMPANIES THAT MATTER MOST IN DATA

DBTA 100

Birst, Inc.
www.birst.com

A provider of cloud BI and analytics for the enterprise, Birst offers a networked BI platform built on top of a multi-tenant cloud architecture to enable BI applications to be transparently connected for local execution with global governance.

BMC Software, Inc.
www.bmc.com

BMC is a provider of software and solutions spanning IT service management, workload automation, IT operations, cloud management, IT automation, and the mainframe.

CA Technologies
www.ca.com

Helping companies seize the opportunities of the application economy, CA Technologies provides IT software and solutions spanning agile management, API management, DevOps, mainframe, and security.

CenturyLink
www.centurylink.com

CenturyLink offers a range of solutions to help keep a business connected, including hosted IT services, such as procurement, installation, monitoring, and ongoing management of hosted solutions regardless of location.

ClearStory Data
www.clearstorydata.com

Committed to bringing data intelligence to everyone, ClearStory Data's solution aims to simplify data access to internal and external sources, automate data harmonization, enable collaborative exploration, and reduce business wait time for insights.

Cloudera, Inc.
www.cloudera.com

Cloudera provides an enterprise level data management and analytics platform built on Apache Hadoop and other open source technologies, along with support, training, and professional services.

Compuware Corp.
www.compuware.com

Compuware helps companies succeed in the digital economy by delivering solutions that enable IT professionals with mainstream skills to manage mainframe applications, data, and platform operations in order to leverage their high-value mainframe investments.

Corvil
http://corvil.com

Corvil helps organizations gain a full picture of customers, business, and IT infrastructure by making sense of the raw data streaming through the network, enriching it with analytics, and making it accessible, searchable, and consumable.

Couchbase, Inc.
www.couchbase.com

Supporting enterprises in the digital economy, the Couchbase NoSQL document database platform spans products and services and offers N1QL, a declarative query language that extends SQL for JSON.

Databricks
https://databricks.com

Founded out of the UC Berkeley AMPLab by the creators of Apache Spark, Databricks offers a hosted data platform powered by Spark that enables organizations to go from data ingest through exploration and production on a single data platform.

DBTA 100 THE COMPANIES THAT MATTER MOST IN DATA

Data Intensity
www.dataintensity.com

Data Intensity is a provider of analytics, business intelligence, and managed cloud services for enterprise application and data management support for the full scope of a customer's enterprise data lifecycle.

DataStax
www.datastax.com

Supporting the performance and availability demands of web, mobile, and IoT applications, DataStax, a provider of database software for cloud applications, offers a platform built on the Apache Cassandra NoSQL database technology.

Datavail
www.datavail.com

Remote database administration services provider Datavail specializes in Oracle, Oracle E-Business Suite, Microsoft SQL Server, MySQL, MongoDB, DB2, and SharePoint and offers managed database services, database design, architecture, and staffing.

Datical
www.datical.com

Datical offers agile database automation solutions to help organizations shorten the time it takes to bring applications to market while eliminating the security vulnerabilities, errors, data loss, and downtime often associated with current database deployment methods.

DBI (Database-Brothers, Inc.)
www.dbisoftware.com

DBI provides DB2 LUW performance solutions for retailers, banks, state governments, insurance companies, healthcare companies, manufacturers, and others with demanding requirements.

Dell Inc.
http://software.dell.com

Dell Inc. and EMC Corp. have signed a definitive agreement under which Dell, together with its owners, Michael S. Dell, chairman and CEO of Dell; MSD Partners; and Silver Lake, is acquiring EMC. The combined company, to be named Dell Technologies, will span consumer and enterprise technologies, including servers, storage, virtualization, and PCs.

Delphix
www.delphix.com

Delphix provides software to help organizations deliver virtualized data across the application lifecycle, spanning development, testing, and reporting environments, and improves developer productivity and data security on premises or in the cloud.

Denodo Technologies
www.denodo.com

Denodo offers data virtualization software to enable faster and easier access to unified business information by providing data integration and data abstraction across a broad range of enterprise, cloud, big data, unstructured data sources, and real-time data services.

Driven, Inc., formerly Concurrent
www.concurrentinc.com

Driven, a provider of application development and performance management solutions, aims to deliver products that makes it simpler for enterprises of all sizes to create, deploy, monitor, and manage their big data applications.

Empolis Information Management GmbH
www.empolis.com

Empolis provides Smart Information Management software to support the creation, management, analysis, processing, and provisioning of relevant information for business processes, regardless of source, format, user, location or device.

THE COMPANIES THAT MATTER MOST IN DATA

DBTA 100

EnterpriseDB Corp.
www.enterprisedb.com

EnterpriseDB provides an RDBMS based on PostgreSQL, including software, services, and training, to support mission-critical enterprise applications.

Entrinsik, Inc.
www.entrinsik.com

Entrinsik offers information management solutions, including a reporting and business intelligence solution that enables organizations to connect data across multiple sources, including Microsoft SQL Server, Informix, Oracle, IBM DB2, MySQL, Microsoft Access, UniData, UniVerse, D3, and Google Docs.

ERwin, Inc.
http://erwin.com

ERwin was spun out of CA Technologies as a standalone entity, funded and supported by Parallax Capital Partners, with the intention of building a big data software company, whose flagship product will be the ERwin data modeling and big data solution.

GridGain Systems
www.gridgain.com

GridGain Systems provides enterprise-grade in-memory data fabric solutions based on Apache Ignite to support high-volume transactions, real-time analytics, and hybrid data processing.

Hortonworks, Inc.
http://hortonworks.com

With a focus on driving innovation in open source communities such as Apache Hadoop, NiFi, and Spark, Hortonworks creates, distributes, and supports enterprise-ready open data platforms and modern data applications.

HPE (Hewlett Packard Enterprise)
www.hpe.com

Following its separation from Hewlett-Packard Co., Hewlett Packard Enterprise made its debut in late 2015 as a provider of enterprise technology infrastructure, software and services businesses, spanning cloud, mobility, big data, and security.

IBM (International Business Machines Corp.)
www.ibm.com

Global technology company IBM offers hardware, software, and services spanning mainframe, cloud, security, mobility, cognitive computing, analytics, and commerce.

Idera, Inc.
www.idera.com

Idera, a provider of database and infrastructure management software, recently acquired Embarcadero Technologies, a provider of software for database management and application development, creating a comprehensive portfolio of database management solutions in the market.

Infobright
www.infobright.com

Focused on enabling the rapid analysis of machine-generated data, Infobright delivers an analytic database platform that allows applications to perform interactive, complex queries for better business decision making.

Informatica Corp.
www.informatica.com

Helping companies fully leverage their information assets whether on-premise, in the cloud, or on the internet, Informatica provides a range of products spanning big data management, data integration, cloud integration, data quality, security, and MDM.

DBTA 100 THE COMPANIES THAT MATTER MOST IN DATA

Innovative Routines International (IRI), The CoSort Co.
www.iri.com

IRI is an independent software vendor specializing in fast data management and data-centric protection that uses state-of-the-art tools and techniques to help organizations gain insights from their data.

Jethrodata
http://jethro.io

Providing an SQL-on-Hadoop engine, Jethro acts as a BI-on-Hadoop acceleration layer to speed up big data query performance for BI tools, such as Tableau, Qlik, and MicroStrategy, from a range of data sources, including Hadoop and Amazon S3.

Jinfonet Software
www.jinfonet.com

Jinfonet provides a Java reporting solution that delivers customizable business intelligence. as well as a suite of professional services to complement its solution.

Kore Technologies
www.koretech.com

Kore provides an enterprise integration and data management suite that helps extend the life and capabilities of MultiValue systems, as well as an ecommerce framework and web solutions built to support large enterprises with high transaction volumes.

Looker
www.looker.com

Bringing data-driven decision making to every level of an enterprise with a data platform that creates a single source of truth, Looker provides a data platform that helps everyone in an organization to find, explore, and understand the data that drives business.

MapR Technologies, Inc.
www.mapr.com

MapR provides a converged data platform that integrates Hadoop and Spark capabilities with global event streaming, real-time database capabilities, and enterprise storage, enabling customers to take advantage of their data.

MarkLogic Corp.
www.marklogic.com

Focused on offering organizations a database platform designed to integrate, store, manage, and search more data than ever before, MarkLogic combines the speed, scale, and flexibility of NoSQL coupled with enterprise features for mission-critical applications.

McAfee, now part of Intel Security
www.mcafee.com

With its Security Connected strategy, hardware-enhanced security, and McAfee Global Threat Intelligence, Intel Security is focused on developing proactive, proven security solutions and services that protect systems, networks, and mobile devices.

Melissa Data Corp.
www.melissadata.com

Melissa Data Corp. provides global address, phone, email, and name identity verification solutions and data enrichments to maximize the effectiveness of business intelligence, big data analytics, and omnichannel marketing initiatives.

MemSQL, Inc.
www.memsql.com

By delivering an in-memory, distributed database with SQL, MemSQL allows customers to capture and query high-value and high-velocity data, enabling quick analysis of transactions, performance, and the ability to deploy and scale in a data center or in the cloud.

THE COMPANIES THAT MATTER MOST IN DATA

DBTA 100

Microsoft Corp.
www.microsoft.com

In the cloud and on-premise, Microsoft offers an array of technologies and solutions for businesses of all sizes, spanning desktop applications, relational database management technology, operating systems, search, and mobile devices.

MicroStrategy, Inc.
www.microstrategy.com

A provider of enterprise software platforms, MicroStrategy's mission is to provide a flexible, powerful, scalable, and user-friendly analytics and identity management platform on premises or in the cloud.

MongoDB, Inc.
www.mongodb.com

MongoDB is a NoSQL database that uses a document data model that is similar to JSON to provide flexibility that allows development teams to evolve the data model rapidly as its application requirements change.

Neo Technology, Inc.
http://neo4j.com

Neo Technology is a provider of NoSQL graph databases and creator of Neo4j, an enterprise database that combines native graph storage, scalable architecture, and ACID compliance for predictability of relationship-based queries.

Ntirety, a Division of HOSTING
http://ntirety.com

Ntirety helps companies manage the entire database governance lifecycle with services, including remote database administration as a service, consulting, DBA on-demand, application and database performance as a service, and cloud services.

NuoDB, Inc.
www.nuodb.com

NuoDB offers NewSQl database technology for the cloud and the modern data center to support distribution of a single, logical database across multiple geographies with multiple master copies and transactional consistency.

Objectivity, Inc.
www.objectivity.com

Bridging the gap between big data and fast data in the Internet of Things landscape, Objectivity provides InfiniteGraph, a distributed graph database, as well as Objectivity/DB, a distributed and scalable object management database.

OpenText Actuate Analytics
www.actuate.com/products

In an increasingly social and mobile world, the OpenText EIM strategy consists of comprehensive and integrated product suites spanning content, process, experience, discovery, information exchange, and analytics to help organizations increase innovation and productivity while minimizing risk.

Oracle Corp.
www.oracle.com

Oracle offers an integrated stack of cloud applications, platform services, and engineered systems that combine hardware and software to provide customers with deployment flexibility and benefits including application integration, advanced security, high availability, scalability, energy efficiency, performance, and low total cost of ownership.

ParStream
www.parstream.com

The ParStream platform, recently acquired by Cisco, allows users to process, analyze, and store data at the edge by connecting streaming data with historic data.

DBTA 100 THE COMPANIES THAT MATTER MOST IN DATA

Paxata
www.paxata.com

Paxata provides a data preparation application that sits on top of an enterprise-scale platform built on Hadoop and powered by Spark, increasing users' analytic productivity on managing data volumes and reducing the risk of data chaos.

Percona, LLC.
www.percona.com

After acquiring Tokutek, including the Tokutek distribution of MongoDB, called TokuMX, Percona offers design, service, support, and remote management for both an ACID-compliant NoSQL database as well as MySQL.

Pick Cloud, Inc.
www.mypickcloud.com

A MultiValue database-as-a-service provider specializing in both public and private cloud solutions, Pick Cloud brings MultiValue environments to the cloud, allowing users to deploy and begin managing applications and workloads.

Pivotal
http://pivotal.io

A supporter of open collaboration for big data, Pivotal provides a big data suite, whose components can be deployed on commodity hardware, pre-certified appliances, virtualized and private cloud instances, and in public clouds.

Predixion Software
www.predixionsoftware.com

Predixion Software offers a predictive analytics solution targeting the Internet of Things that can embed on devices, gateways, or the cloud.

Progress Software Corp.
www.progress.com

With the conviction that standards-based access is key to a simplified process, regardless of the data's format or location, Progress Software provides data connectivity and integration solutions for a range of sources and systems, on-premise and in the cloud.

Protegrity
www.protegrity.com

Protegrity provides data security across big data clusters, cloud environments, databases, mainframes, and components of the heterogeneous enterprise by leveraging the knowledge and experience of leading tokenization specialists.

Pythian
www.pythian.com

As a global IT services company, Pythian specializes in designing, implementing, and managing systems that directly contribute to revenue and business success.

Qlik
www.qlik.com

Allowing users to turn data into insights across all levels of the enterprise, Qlik provides a portfolio of products ranging from self-service visualization to guided and embedded analytics.

Qubole
www.qubole.com

The Qubole Data Service platform is based on open Apache technologies and includes Hive, MapReduce, Pig, Oozie, and Sqoop to enable a complete big data service to provide data analysts with a graphical interface, built-in connectors, and elastic cloud infrastructure.

THE COMPANIES THAT MATTER MOST IN DATA

DBTA 100

Rackspace
www.rackspace.com

With data centers positioned across the globe, Rackspace provides a variety of managed services to support all cloud workloads, including a range of database as a service options.

Raytion
www.raytion.com

Raytion offers advanced, feature-rich, and scalable search solutions that organizations can utilize to securely access information from heterogeneous systems in a unified way.

Redgate Software Ltd.
www.red-gate.com

Providing a range of tools for database comparison, productivity, source control, and troubleshooting for developers and DBAs, Redgate allows users to tap into Microsoft SQL Server, Azure, Oracle, and MySQL.

Red Hat
www.redhat.com

Red Hat uses a community-powered approach to create products and services spanning cloud, Linux, middleware, storage, and virtualization technologies.

RedPoint Global
www.redpoint.net

RedPoint Global provides a comprehensive set of ETL, data quality, and data integration applications that operate in and across both traditional and Hadoop 2.0/YARN environments.

Reltio
www.reltio.com

Offering users with data-driven applications a modern data management platform as a service, Reltio gives users a variety of resources to analyze a breadth of information and store it on-premise or in the cloud.

Revelation Software
www.revelation.com

Revelation provides a suite of development tools, a database, and related services that take full advantage of leading computing architectures and operating environments to ensure the preservation of information systems and application investments by its developer and user community.

Rocket Software, Inc.
www.rocketsoftware.com

Helping customers prevent outages, as well as protect, store, share, and virtualize data, Rocket Software provides solutions and products across areas such as DBMSs and application servers, the mainframe, data migration, network management, virtualization, enterprise performance management, and BI and analytics.

SAP SE
www.sap.com

Known for HANA, its platform for next-generation applications and analytics, SAP SE is a global provider of enterprise application software, whose services are used by more than 291,000 customers.

SAS Institute, Inc.
www.sas.com

Giving customers "the power to know," SAS is a provider of business analytics software and services, with solutions across areas such as advanced analytics, business intelligence, customer intelligence, data management, risk management, fraud and security intelligence, Hadoop, and more.

DBTA 100 THE COMPANIES THAT MATTER MOST IN DATA

Search Technologies
www.searchtechnologies.com

Search Technologies specializes in the design, implementation, and management of enterprise search and big data applications.

Sinequa
www.sinequa.com

Sinequa helps customers make sense of their big data and extract value from it by providing business insights and actionable information with comprehensive and highly integrated real-time big data search and analytics.

Sisense
www.sisense.com

By providing a complete solution for preparing, analyzing, and visualizing big or disparate datasets, Sisense offers business intelligence and analytics software that allows business users to join multiple large datasets, build smart dashboards with data visualizations, and share with thousands of users, without dependence on coding, IT, or data scientists.

SnapLogic, Inc.
www.snaplogic.com

Delivering an elastic integration platform as a service to connect cloud applications and disparate data sources, SnapLogic assists enterprises with making better decisions and provides a single view of customers, products, and services.

Snowflake Computing, Inc.
www.snowflake.net

A cloud data warehousing company, Snowflake Computing aims to reinvent the data warehouse, bringing together all users, all data, and all workloads in a single repository.

Software AG
www.softwareag.com

Software AG provides big data, integration, and business process technologies that enable customers to drive operational efficiency, modernize their systems, and optimize processes for smarter decisions and better service.

SolarWinds
www.solarwinds.com

Striving to eliminate complexity within the IT space, SolarWinds provides products and tools to solve a broad range of IT management challenges across networks, servers, applications, storage, and virtualization.

Splice Machine
www.splicemachine.com

Supporting ACID transactions via Hadoop RDBMS, Splice Machine is designed to scale real-time applications using commodity hardware without application rewrites.

Splunk, Inc.
www.splunk.com

Splunk software and cloud services help organizations to search, monitor, analyze, and visualize machine-generated big data coming from websites, applications, servers, networks, sensors, and mobile devices for real-time operational intelligence.

SQLstream, Inc.
www.sqlstream.com

Providing customers with actionable insights extracted from their machine data in motion, SQLStream provides SQLstream Blaze, a scalable, real-time data hub for operational intelligence and the Internet of Things that is built on a SQL-compliant data stream processing platform.

THE COMPANIES THAT MATTER MOST IN DATA

DBTA 100

SQL Sentry
www.sqlsentry.com

SQL Sentry's software solutions for monitoring and optimizing data performance in Microsoft SQL Server and Windows environments are sold and supported directly and through authorized resellers and partners.

Syncsort, Inc
www.syncsort.com

Syncsort helps organizations collect, integrate, sort, and distribute data using fewer resources by providing solutions that span big iron to big data, including next-gen analytical platforms such as Hadoop, cloud, and Splunk.

Tableau Software, Inc.
www.tableau.com

Making business intelligence faster and easier, Tableau provides a suite of interactive products to help users quickly analyze, visualize, and share information in the office and on-the-go.

Talend
www.talend.com

Talend builds on open source innovation to create enterprise-ready solutions spanning big data, cloud, application integration, and master data management to help customers derive business value more quickly.

Teradata Corp.
www.teradata.com

Teradata has a portfolio of big data analytic solutions, integrated marketing applications, and services across cloud, Hadoop, data warehousing, and analytics to help organizations gain a sustainable competitive advantage with data.

TIBCO Software, Inc.
www.tibco.com

Combining real-time event processing and insightful analytics, the TIBCO platform offers a mix of products and services to help customers capture the right information at the right time and act on it quickly for a competitive advantage.

Trillium Software, a Harte Hanks Company
www.trilliumsoftware.com

Enabling customers to anticipate opportunities, uncover risks, and make better decisions, Trillium's technologies and services include global data profiling, data cleansing, enrichment, and data linking for ebusiness, big data, CRM, data governance, ERP, supply chain management, data warehouse, and other enterprise applications.

Violin Memory
www.violin-memory.com

Violin Memory supports the next-generation data center and provides all-flash storage arrays, recently introducing a new all-flash storage system designed to run all primary storage and active workloads below the cost of traditional disk.

VoltDB
http://voltdb.com

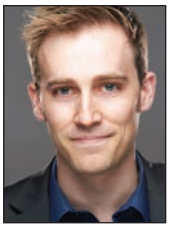
NewSQL database provider VoltDB combines the speed and scale of NoSQL databases with the ACID guarantees, relational data models, and transactional capability of traditional RDBMSs.

Yellowfin International
www.yellowfinbi.com

Delivering BI functionality via a single integrated platform, Yellowfin provides interactive dashboards, support for data exploration, collaboration, data visualization, and mobile device access.

DBTA 100 VIEW FROM THE TOP

AgilData



Dan Lynn, CEO

AgilData provides solutions and services to help firms deliver on the promise of Big Data and scale complex MySQL data infrastructures. AgilData provides seamlessly integrated best-of-breed Big Data solutions for BI analytics, data pipeline orchestration, data wrangling and event processing. We partner with firms to transform their engineering practice, shorten development cycles, reduce their ops load and leverage data assets to build new value streams.

AgilData delivers data strategies to help global brands, Startups and Enterprises develop custom applications and scalable Big Data solutions to drive BI, analytics, machine modeling and custom applications using technologies like Hadoop, Spark, Storm, Kudu, Redshift, Alooma, Chartio, Arcadia Data. As an Agile company with an agile engagement model based on two week iterations and we actively contribute to the open source community.

As a leading MySQL database solution to scale data, AgilData Scalable Cluster for MySQL helps MySQL operators meet the demands of reliability, scale, replication, streaming and performance for business critical applications. Available as Enterprise Support Subscription License and with Remote Managed Service for complete white-glove DBA staff augmentation, AgilData Scalable Cluster for MySQL is the low cost standard MySQL scaling solution.

Developers of applications based on MySQL need a low-cost and reliable solution to immediately scale existing MySQL databases without impacting the application, allowing them to focus on developing application features. Our fully managed Scalable Cluster for MySQL allows developers to delegate ongoing DBA costs while gaining all the benefits from having 24×7 experts running your data infrastructure.

AgilData
www.agildata.com

Attunity

Shimon Alon, Chairman & CEO

Experts in Enterprise Big Data Integration and Management

Big Data is the new gold. But to gain value from it, enterprises must first understand which data to move, where to place it, and how to manage it efficiently. This multi-step process, along with ever-growing data volumes, data types and velocities, inevitably increases complexity. Bringing innovative advancements to these areas is Attunity, a provider of high-performance data management software solutions.

Attunity's solutions enable optimal information availability across enterprise platforms, organizations, and the cloud. With 20+ years of experience, Attunity serves more than 2,000 customers, including half of the Fortune 100, across all industries in need of solutions such as data replication with change data capture (CDC), data warehouse automation, data usage analytics, data connectivity and cloud data delivery.

Through direct selling and partnerships with the top IT vendors in the world, Attunity provides solutions that can assess, move and manage data in real time. The firm simplifies data usage assessment and EDW performance optimization by providing a single management dashboard across heterogeneous environments and eliminating the time-consuming and costly process of scripting code.

Sample uses of Attunity solutions include:

- Modernization: Incorporating Hadoop to the IT infrastructure
- Real-time Big Data access and delivery for:
  - fraud detection
  - BI and predictive/operational analytics
  - 360° view of the customer/business
  - Global data distribution (i.e. one source to many targets)
  - Data consolidation/migration (i.e. many sources to one target)
  - High availability/disaster recovery
- Data re-balancing for system performance optimization
- Charge-back/Show-back

For more information, visit http://www.attunity.com or our blog and join our community on Twitter, Facebook, LinkedIn and YouTube.

Attunity
www.attunity.com

VIEW FROM THE TOP DBTA 100

BackOffice Associates

David Booth, Chairman & CEO

As the amount of data created increases, so do the complex data challenges large enterprises are confronted with. These challenges can have a direct impact on the quality of an organization's data, and that data quality directly impacts business performance and profitability. It is because of these trends that there is so much interest in data management solutions and innovative information governance programs.

At BackOffice Associates, we understand that next-generation information governance is necessary to maximize the value of an enterprise's data assets, as well as improve the efficiency of business processes and increase the overall value of the organization. To do this, you need to embrace the next wave in data stewardship with products and solutions that not only guide but also empower your organization to turn data into a business value.

Our vast experience in working with the largest organizations in complex industries such as manufacturing, life sciences and retail has helped us craft a product strategy that scales to meet the demands of any industry. Coupling our products with SAP solutions for enterprise information management clearly propels our offerings into an industry leadership position for companies seeking information governance and data quality excellence with cost, value and efficiency in mind.

We have and will continue to successfully help these organizations and others utilize strategic data quality initiatives to support their critical business processes and decisions now and well into the future.

BackOffice Associates
www.boaweb.com

Cloudera

Tom Reilly, Chief Executive Officer

Cloudera delivers the modern data management and analytics platform built on Apache Hadoop and the latest open source technologies. The world's leading organizations trust Cloudera to help solve their most challenging business problems with Cloudera Enterprise, the fastest, easiest and most secure data platform available for the modern world. Our customers efficiently capture, store, process and analyze vast amounts of data, empowering them to use advanced analytics to drive business decisions quickly, flexibly and at lower cost than has been possible before. To ensure our customers are successful, we offer comprehensive support, training and professional services. Learn more at cloudera.com.

Cloudera
cloudera.com.

DBTA 100 VIEW FROM THE TOP

Compuware

Chris O'Malley, CEO

Enterprises have invaluable intellectual capital residing on their mainframes in the form of data and business logic. Leveraging this treasure trove of corporate IP is an imperative if enterprises are to serve effectively in the new digital economy. This, however, is proving a formidable challenge as experienced mainframers retire, leaving the stewardship of mainframe data and applications to a new generation of developers, who are unfamiliar with the platform.

Mainframe-hosted applications and data must be continuously adapted to work in concert with a growing plethora of Web and mobile apps, so businesses can improve the digital customer experience in preferred ways. But how can mainframe-inexperienced developers work with unfamiliar, highly involved and evolved mainframe applications and perform data related tasks with confidence and excellence?

Compuware's offerings "mainstream the mainframe," so that any developer can understand and work with mainframe applications and data in ways that are different only in syntax from other platforms. As the world's leading mainframe-dedicated software company, Compuware is laser-focused on enabling mainframe DevOps by delivering new software and integrations every 90 days.

Discovery and visualization solution Topaz, for example, enables next-gen mainframe developers to manipulate enterprise data and applications with greater agility, speed and confidence to better meet the demands of mobile and analytics and create a preferred customer experience.

By equipping next-gen developers—the mainframe stewards of tomorrow—with tools that are familiar and intuitive, Compuware is ensuring mainframe applications and data remain essential business assets for generations to come.

Visit us at compuware.com.

Compuware
compuware.com

Couchbase, Inc.

Bob Wiederhold, President and Chief Executive Officer

Couchbase, Inc. delivers the world's highest performing distributed database platform for the Digital Economy. Developers around the world choose Couchbase for its superior flexibility, scalability, performance, and 24x365 availability to build enterprise Web, mobile and IoT applications. The Couchbase platform includes **Couchbase Server**, **Couchbase Lite** — the first mobile NoSQL database, and **Couchbase Sync Gateway**.

Couchbase is designed for global deployments that can **operate at any scale**, with configurable cross data center replication to increase data locality and availability. Couchbase is committed to open development and all products are open source.

Couchbase consistently outperforms competitors in third party industry **benchmarks**, and enables developers to build enterprise applications on NoSQL using **N1QL**, a declarative query language that extends SQL for JSON.

Couchbase Server is a high-performance distributed database with a flexible data model. It scales on commodity hardware to support large data sets with a high number of concurrent reads and writes while maintaining low latency and strong consistency.

Couchbase Mobile is a NoSQL database solution that delivers the full power and flexibility of JSON to mobile. It's engineered to provide fast and consistent access to data, with or without a network connection, removing the network dependency that traditional service-based approaches require.

Leaders from every industry use Couchbase, including AOL, AT&T, Cisco, Comcast, Disney, eBay, General Electric, Marriott, Nordstrom, Neiman Marcus, PayPal, Verizon, Wells Fargo and hundreds more.

To learn more, please visit us at **www.couchbase.com**.

Couchbase, Inc.
www.couchbase.com

VIEW FROM THE TOP DBTA 100

Datavail

Scott Frock, COO

What does "DevOps" really mean? Can its scope and application be confined to a single definition that will last through the years? Or, like the very industry in which it's applied, does it evolve and change over time?

In my last visit to the Cloud Expo/DevOps Conference, it became apparent that DBA life is continuing to evolve and "DevOps" has yet to establish a foothold as a formally adopted methodology. Despite this, the importance of implementing a carefully crafted DevOps program in your enterprise is essential to the evolution and maturation of your application lifecycle approach.

A critical aspect of this evolution is an investment in automation. When standardized and repetitive operations tasks like testing and release management are automated, consistent and high-quality execution is a given. This means that DBA life as we knew it has expanded—in large part due to DevOps. The end result is synchronicity between development and deployment that empowers the developer, delivers software faster, and diminishes the opportunity for error.

As the state of software development evolves and new technology changes the way we approach our industry, your organization's IT strategy should follow suit. Datavail is on the cusp of this rapidly emerging practice, offering 24x7 DevOps consultation and support that encompasses every aspect of the convergence of database administration and software development. Visit Datavail.com to see how we can help you implement DevOps and achieve automation in your organization today.

Datavail
www.datavail .com

DBI Software

Scott Hayes, President & Founder

Scott Hayes is an IBM DB2 LUW Performance Expert, IBM DB2 GOLD Consultant, IBM Information Management Champion, US patent inventor, published author, blogger on DB2 LUW performance topics, and popular frequent speaker at IBM IOD and IDUG Conferences. He started DBI Software in July 2005 with one simple mission: "Help People!" Ten years later, this simple mission is still DBI's #1 core value, though Mr. Hayes admits, "We are better at helping people with DB2 than their marriages or cars."

To that end, DBI boasts many achievements such as saving a health care company $250,000 in one afternoon, solving a performance problem in two hours that two senior DBAs were unable to solve in six months, reducing SAP reporting times from twelve hours to two, and helping top brand name retailers achieve top ranked e-commerce performance on Black Friday multiple years in a row.

DBI's secret sauce is patented technology combined with a bullet-proof methodology that delivers real, positively impactful, measurable ROI within days. DBI's PureFeat™ Performance Management Suite for DB2 LUW brings performance inefficiencies, issues, and solutions to light that other tools and scripts don't find, usually within five mouse clicks!

DBI Software is your trusted partner for Breakthrough IBM DB2 LUW Performance Solutions that DELIVER INVALUABLE RESULTS for Organizations having the most Demanding Requirements and Discriminating Preferences. Achieve improved performance, lower hardware and database license costs, lower energy costs, lower cloud costs, accelerated productivity, higher profits, and ROI in DAYS!

DBI Software
Please visit www.DBISoftware.com to learn more.

DBTA 100 VIEW FROM THE TOP

Denodo

Alberto Pan, Chief Technology Officer

Traditional data integration architectures rely on replicating all of the relevant data and storing it in a central repository. But replicating big data is impossible, and trying to centralize the cloud is like trying to boil the ocean. For today's Fortune 500 companies seeking to monetize big data, traditional data integration architectures lead to very slow development and slow decision making, when they don't lead directly to "big data paralysis," which is an abundance of data without the knowledge of how to benefit from it. Companies need data integration architectures that can support a fast data strategy, and data virtualization provides such an architecture.

Data virtualization combines the data from any source, be it on-premises or in the cloud, in a consistent, business-oriented view. It then publishes the data for each required application in its preferred format, enabling self-service consumption by business users.

Unlike traditional architectures, data virtualization can deliver:

1. Fast, distributed queries across distributed systems, avoiding data replication
2. Abstraction, to insulate the business users and consuming applications from changes in the underlying data processing infrastructure
3. A single entry point from which to apply unified security and data governance policies across several heterogeneous systems

Denodo is the leader in data virtualization. The award-winning Denodo Platform accelerates your fast data strategy by delivering breakthrough performance in big data, logical data warehouses, data lakes, and operational scenarios. It also expedites the use of data by business users with self-service data discovery and search. Denodo customers experience very high ROI from faster decision-making, the shortest time-to-data, and improved business-user productivity.

Denodo
www.denodo.com

Empolis Information Management GmbH

Dr. Stefan Wess, CEO

Big Data has set the stage for the comeback of Artificial Intelligence and many international companies have put it on the top of their agendas. They find it increasingly challenging to make ideal and purposeful use of the exponentially growing amount of unstructured data—such as texts, business and PDF documents, web contents, social media, videos, audio and graphic sources—in order to meet their business-critical processes.

In doing so, companies and their executives face two big challenges: a Quantity Challenge—to analyze a large amount of information in real time, and a Quality Challenge—to deliver only the precise information required.

Empolis is the leading provider of Smart Information Management (SIM) software and a pioneer in artificial intelligence technologies, especially where extremely powerful scalability and the exponentially growing volume of digital information needs managing. Artificial Intelligence, case-based reasoning, semantic and language processing methodologies are applied with which relevant information is extracted from unstructured data streams and text, structure and context is identified, and connections between data, as well as their data sources, are established.

Hence, SIM represents the comprehensive creation, management, analysis, intelligent processing and provisioning of all information relevant to a company's business processes, regardless of source, format, user, location or device.

Therefore, Artificial Intelligence technologies are more prevalent than ever. Empolis technologies help companies by making them "smarter" day by day and enabling them to optimize and better understand and recognize emerging developments and issues, in order to be able to react correctly and in time.

Empolis Information Management GmbH
www.empolis.com

VIEW FROM THE TOP DBTA 100

ERwin, Inc.

Adam Famularo, CEO

ERwin Inc.: Big Data, Big Results

ERwin, Inc. is the maker of the award-winning and industry leading data modeling solution used by more than 50,000 data management professionals in 60+ countries. The gold standard in data management, ERwin enables midsize and enterprise companies to model and govern their most valuable resource — their data — and acts as an information hub to share data definitions, rules and structures across the organization. ERwin has a rich history, but what's most exciting is the role that ERwin is coming to play as the foundation for global Big Data initiatives.

At ERwin, we get the hard stuff right. From enterprise data standards, to data governance and control, to flexibility and customization, and web-based publication and reporting. We know how to help you deliver **big results** from your **big data** initiatives.

Here are the top reasons to deploy ERwin in your enterprise:

- Get a single view of all data sources, based on a common repository of standard data assets
- Leverage a customizable, web interface that allows users across the organization to easily but securely visualize their most important data
- Build better data governance through enterprise standards — including naming standards, data type standards, model templates, and more.
- Works the way you work: ERwin supports the development of big data structures that fit in *your* business and not the other way around
- Open, extensible and easy to integrate: pre-built connectors to more than 130 tools and APIs to build the warehouse of your dreams
- Support for a broad range of data sources, including relational, big data and cloud

Join the ERwin movement. Subscribe to our LinkedIn community and follow us on Twitter @ERwinModeling.

ERwin, Inc.
www.erwin.com

GridGain Systems

Abe Kleinfeld, President & CEO

GridGain Systems uses in-memory computing to enable data-intensive applications to run 1,000x faster and massively scale out without ripping and replacing your existing databases or rewriting your applications. The GridGain cluster can scale out to support hundreds of TBs of in-memory data. The GridGain in-memory performance layer sits between your application and data layers and can be deployed on premise, in the cloud, or in hybrid environments.

GridGain is used by major financial services, telecom, retail, online services, IoT, and pharma companies in mission-critical applications. It dramatically improves the performance of existing systems for high-volume transactions, real-time analytics and hybrid transactional/analytical processing.

Built on the open source Apache Ignite™ project, GridGain includes a compute grid, data grid, service grid, as well as in-memory streaming, messaging and Hadoop acceleration in a single converged data platform. GridGain has a Unified API which connects seamlessly to business-critical applications and a data abstraction layer which allows it to connect to any RDBMS, NoSQL and/or Hadoop database. Data is pulled into the GridGain in-memory computing cluster, increasing processing speeds by 1,000x or more versus disk-based databases. The GridGain cluster is easily expanded by adding commodity nodes, which allows your applications to scale out and also run faster. GridGain offers full ACID transactions and supports all key types of applications (Java, .NET, C++) with ANSI SQL-99 compliance.

Download a 30-day free trial of GridGain at www.gridgain.com/resources/download/. For more information, visit www.gridgain.com. GridGain is headquartered in Foster City, California.

GridGain Systems
www.gridgain.com

DBTA 100 VIEW FROM THE TOP

IDERA

Randy Jacops, CEO

IT'S INCREASINGLY CLEAR that IT doesn't run on the network – it runs on the data and databases that power your business. Research shows that 88 percent of application managers cite the database as the primary challenge with application performance. As the foundation on which all businesses operate, the database is the most critical element of IT infrastructure, requiring a consistent, reliable management approach.

That's why IDERA designs products with the database as the nucleus of the IT universe. Our database lifecycle management (DLM) solutions allow IT professionals to design, monitor and manage data systems with complete confidence, whether in the cloud or on-premises. IDERA is now the only third-party vendor to provide cross-platform DLM tools that meet the needs of application managers, IT admins, business architects and database developers and administrators.

IDERA provides IT pros with the tools they need to address issues before they impact an organization. Our customers face tremendous pressure each day to ensure their IT systems operate at the highest levels. We're dedicated to making database management easier by delivering support at every stage of the database journey.

We aspire to maintain multi-decade relationships with our customers and strive to achieve that goal by focusing solutions on customer value and success. Innovation never stops and we never stop improving our products and advancing the customer value equation. Additionally, we believe great products deliver amazing ease of use, exceptional quality, and unmatched scalability. We call these our core product values, and our success reflects our ability to jointly deliver continuous innovation aligned with these values.

IDERA
www.idera.com

Kore Technologies

Ken Dickinson, Co-Founder & Managing Partner

I REMEMBER THE 80'S AND 90'S when there was a plethora of ERP companies all competing for your business, each with their own unique brand and specialized vertical. But just like many other industries such as medical insurance, pharmaceutical, and financial institutions, only a handful of ERP companies remain today. As these behemoths grow bigger through acquisition, they become more bureaucratic and less responsive to the needs of their customers. Often times the support and maintenance dollars you spend are being invested in the "next generation platform" rather than the legacy ERP platform that you are running.

Well the good news is that this trend may be reversing as we see new ERP and other software startups emerging that are open-source, cloud-based platforms with an emphasis on eCommerce from the start. However, the cost of migrating to any new ERP platform may be cost prohibitive or worrisome given the the new economic norms of low-growth and hidden inflation. So what can you do to modernize and streamline operations to remain competitive in the marketplace without breaking the bank? Glad you asked.

Kore Technologies has been doing business for 18 years and specializes in breathing new life into your aging ERP system. With our award-winning Kourier Integrator enterprise integration and data warehouse solution we can help you integrate with best-in-class third-party applications via real-time RESTful Web Services or with a traditional message-based architecture. Add our SQL Accelerator to take advantage of the numerous software products focused on business analytics reporting and dashboard development. Need to fully embrace eCommerce? Integrate to our trend-setting KommerceServer eCommerce Suite, which includes a B2B/B2C Storefront and Customer Portal. Already have a non-integrated storefront? We can help you integrate that with your back-office system too.

Kore Technologies
Solutions that work. People who care.

Kore Technologies
www.koretech.com
info@koretech.com
866-763-KORE

VIEW FROM THE TOP DBTA 100

Ntirety, a Division of HOSTING

Michael Corey, President, Ntirety

My first foray into the realm of Oracle occurred over 30 years ago with Oracle version 3. Today, just as then, there is a critical need for the technologist to ensure that the database runs optimally and that the information held within the database is safe and secure. Information technology's antiquity limited the companies to directly hired staff or expensive consultants when considering options for management of the database. Managed Database Services companies such as Ntirety had yet to be conceived.

Today, hourly consultants still thrive. However, a more optimal option to consider may be that of partnering with Ntirety, which, through our "On Demand" services, allows you to have access to the right technologist, empowered with the right technology, at the right time. You can use expert DBA services for a short-term, simple task or an extended and long-term project. You can also choose from an array of fixed-price Database Administration as a Service® plans, or purchase "On-Demand" services "by the drink" to fit any needs or budget.

Partnering with Microsoft, our parent company, HOSTING, now offers "Database-as-a-Service," or a portion of a "Database-as-a-Service" for a little or as long as you need. So, if you just want the database, and not the management option, we offer that, too.

At Ntirety, we discovered that by maintaining a U.S.-based DBA team and leveraging technology, we could empower our DBAs to deliver more efficiently, thereby offsetting the apparent competitive disadvantage in required compensation. Those efforts have proven wildly successful. Using Ntirety's award-winning technologies, such as Nspect™ and Ntrust®, 55% of the challenges that required human intervention are now resolved automatically through our methodology of "self-healing technology." This approach has resulted in quicker response times for our clients and, ultimately, a higher quality service. By discovering the more effective path and taking "the raised floor less traveled" we are able to offer a very competitive price with extremely high quality results.

As Ntirety moves forward, we will continue to enhance our proprietary technology to ensure our clients' databases run optimally and their data is always safe and secure. We will continue to embrace the concept of "On-Demand" services so our customers are able to acquire the precise services they need from us when, and only when, they need those services.

Ntirety, a Division of HOSTING
www.ntirety.com

Pythian

Robin McIntyre, VP Managed Services

Pythian excels at helping companies adopt and manage disruptive technologies to better compete.

The cloud promises agility, cost flexibility, availability, and elasticity, but moving your applications to the cloud is not always straightforward. Most enterprise applications rely on monolithic architecture and have certain characteristics that make them difficult to migrate to the cloud. These features also make it challenging to take advantage of the benefits touted by cloud vendors.

Pythian has more than 19 years' experience planning, implementing, and managing IT infrastructures and we understand the challenges of moving enterprise systems to the cloud. We have helped organizations migrate applications to the cloud and achieve an optimized IT infrastructure that is agile, cost-effective, scalable, flexible, and elastic. Our cloud transformation services include:

- Strategy development that includes assessing your current architecture, business needs, and future growth requirements
- Application migration from legacy to cloud environments
- Application re-platforming
- Application lifecycle optimization
- Ongoing operations management of the application, data, and infrastructure in the cloud

We recently developed a cloud architecture for a major gaming company, enabling them to:

- Reduce release rates from monthly to twice daily
- Reduce mobile load time by 90%, for huge performance improvements
- Cut marginal cost per user per month by 30%

Our expertise in extreme scale, high availability, and advanced automation lowers your deployment risk and decreases your time to market. Our experts are certified in the three major public cloud platforms. As top-level partners with Amazon Web Services, Google Cloud Platform, and Microsoft Azure, we guarantee we'll find the best solution for your business.

Before making changes to your organization's IT strategies and adopting new deployment and maintenance models, contact us.

Pythian
www.pythian.com

DBTA 100 VIEW FROM THE TOP

RedPoint Global

Dale Renner, CEO & Founder

We are reaching a tipping point in which nearly every business decision will be driven by data. As a result, every organization will need to pay particular attention to the quality of their data. Without data quality, you can't make truly informed business decisions.

RedPoint Data Management™ simplifies the complexity of harnessing data to drive business growth. With RedPoint, you can capture, transform, cleanse, integrate and link all of your data, no matter where it resides or where it needs to go. Data Management works across traditional databases and emerging big data frameworks, including Hadoop, without needing any specialized programming skills.

Data-driven organizations understand the value of deploying a single data management tool that integrates all these capabilities through a graphical drag / drop UI. It enables business users, those most familiar with data use cases, to play active and, even, self-service roles in preparing aggregations for consumption by analytics and legacy applications. RedPoint Data Management delivers unique value in loading big data, integrating it with primary data, and achieving high quality in probabilistic matching and master key management.

RedPoint customers are maximizing value from their data, as demonstrated in Gartner's latest Critical Capabilities for Data Quality Tools report, where RedPoint received the highest score in both the Data Integration and Operational/Transactional Data Quality Use Case categories. RedPoint also received the second-highest scores in Data Migration, Big Data & Analytics and Master Data Management; and the third highest score in Information Governance Initiatives.

For more information, visit www.redpoint.net or call 781-725-0250.

RedPoint Global
www.redpoint.net

Reltio

Ramon Chen, Chief Marketing Officer

Modern Data Management Powers a New Generation of Data-driven Applications

Today's data management and analytics landscape is more complex than ever. With so much cloud technology available, at a fraction of the cost of heavyweight on premise tools, you would think life would be easier for IT, and business would be able to stop resorting to self-service ways of getting the information they need. Yet a disconnect remains between IT efforts and business agility.

Fortunately a new wave of modern data management Platform as a Service (PaaS) offerings are now available (Read Constellation Research's latest report). These platforms start with a reliable data foundation with built-in Master Data Management (MDM). No longer a separate expensive and siloed endeavor, todays modern PaaS deliver MDM as a seamless, integral part of an overall platform that leads to relevant, contextual analytics, and ultimately business-user facing recommended actions through data-driven applications.

The essential characteristics of a modern PaaS include:

- MDM as a foundational core with fast time to value and personalized context (Read Forrester's MDM Q1 2016 Wave)
- Multi-tenant Cloud with low TCO, no impact upgrades, with immediate access to new technology, 3+ times a year
- Integrated Data as a Service (DaaS), with third party data pre-loaded, available on demand
- Big data scale, with a commercial graph foundation (Read Forrester's report "To Graph or not to Graph")

Ultimately a modern PaaS needs to deliver business applications with the ease-of-use, and role-based guided assistance to users, much like how Facebook, LinkedIn and even Amazon is tailored to your needs. Evolving to the next generation PaaS and data-driven apps is easier than you think. Visit reltio.com/modernize to take this quick assessment to see where you stand.

Reltio
www.reltio.com

10 Emerging Companies to Watch in Data

New and emerging vendors offer fresh ways of dealing with data management and analytics challenges. Here, *DBTA* looks at the 10 companies whose approaches we think are worth watching.

AtScale

www.atscale.com

Founded with the philosophy of "playing the data where it lays," AtScale provides a platform that allows business users to use tools such as Excel, Tableau, and QlikView for BI on Hadoop.

Avi Networks

www.avinetworks.com

Aiming to deliver public cloud-like agility for application services, Avi Networks offers the Avi Vantage Platform, which provides elastic, software-defined application services on commodity x86 servers, VMs, or containers.

Cask Data, Inc.

www.cask.co

Focused on developers, Cask is an open source, big data software company whose flagship offering aims to make it possible to quickly and easily develop and deploy more powerful applications for Hadoop.

CloudBerry Lab

www.cloudberrylab.com

Established by a group of experienced IT professionals, CloudBerry Lab provides cloud-based backup and file management services and allows customers to store their backup data with more than 20 online storage providers, including Amazon S3, Microsoft Azure, and Google Cloud.

THE COMPANIES THAT MATTER MOST IN DATA

Glassbeam, Inc.
www.glassbeam.com

Machine data analytics company Glassbeam was founded with the mission of "helping product companies make sense of complex machine data collected from connected devices."

Graylog, Inc.
www.graylog.org

Graylog provides service and support for the open source Graylog solution, which stores, searches, and analyzes machine data collected from IT infrastructures and applications, to help organizations improve IT operations efficiency and security, as well as reduce IT cost.

Greenwave Systems
www.greenwavesystems.com

Greenwave Systems, a provider of IoT software and managed services, seeks to drive mass adoption of IoT by enabling leading brands to use technology in ways that foster deeper customer relationships.

JFrog Ltd.
www.jfrog.com

With the goal of changing the way companies and individuals develop, release, and distribute software, JFrog products provide developers and DevOps engineers with centralized control, management, and monitoring for all enterprise artifact assets globally.

Kyvos Insights
www.kyvosinsights.com

Founded by a team of veterans from Yahoo, Impetus, and Intellicus Technologies, Kyvos Insights, a big data analytics company, emerged from stealth mode in 2015 to introduce Kyvos, an online analytical processing (OLAP) solution for Hadoop designed for big data needs.

SnappyData
www.snappydata.io

SnappyData was incubated at Pivotal and launched in 2016 as a standalone entity providing an in-memory hybrid transactional analytics database built on Apache Spark.

10 Emerging Companies to Watch in Data

THE NEW WORLD OF DATABASE TECHNOLOGIES

database
TRENDS AND APPLICATIONS

Best Practices Series

10 Ways to Ride THE NEW WORLD OF DATABASE TECHNOLOGIES

Best Practices Series

In the past 5 years, there has been nothing short of an earthquake in the database space. Many of the edifices of data management thinking have come down. Yet, there are tried-and-true rules of data that have stood the test of time.

Consider the paradoxes seen as new technologies and approaches have come on the scene. The new data world emerging is one that is increasingly being built on cloud but still supports on-premises assets. It runs on open source code and frameworks, yet offers more commercial choices than ever before. In today's data world, analytics is being embedded behind many processes and systems, yet is becoming more visible through smarter, more user-friendly business intelligence tools. Data is accessible through terminals and PCs, but also can be viewed through devices such as smartphones and tablets. There are many new job titles emerging in this space, yet there remains critical demand for tried-and-true database administration and management skills. Startup businesses are employing data to disrupt or create entire new markets, but established enterprises are learning to leverage their immense data assets to better serve customers.

While the seeming contradictions introduced by today's database technology may be perplexing, one thing is clear: The need for faster, more effective information delivery is growing more urgent every day. Data analytics is now the foundation of many high-level business strategies and is a major competitive differentiator in today's hectic global markets. A majority, 57%, of data managers and professionals in a recent Unisphere Research survey report that analytics is part of the day-to-day decision making within their organizations. However, an inability to get at needed information is the major inhibitor for decision-making in today's organizations. A majority of the more than 300 managers and professionals responding to the survey say they frequently encounter a lack of complete information, as well as delays in getting the information they need ("Moving Data at The Speed of Business: 2016 IOUG Survey on Data Delivery Strategies").

Providing this information when it is needed, and in the right context, to decision makers—which may be at real-time speed—is the challenge of data delivery in this era. This demands a rethinking of how data managers need to respond to their organizations' insatiable information requirements.

The following are 10 key best practices for preparing for the new world of database technologies.

1. THINK MORE BROADLY ABOUT WHAT CONSTITUTES A 'DATABASE' BECAUSE IT'S NO LONGER A SELF-CONTAINED SYSTEM.

It's time to rethink our notion of what constitutes a database. They are no longer just rows and volumes of data stored on a server somewhere in someone's data center, only visible in reports. They encompass a broader world that links into the internet at large, the Internet of Things, and clouds. For example, emerging blockchain technology is designed to store data, general-ledger style, in a highly distributed fashion across the wider internet. The NoSQL world offers a variety of structures that open up new types of data sources and provide ways to tap into the institutional knowledge long bottled up in PCs and departmental silos. The big data revolution—supported on frameworks such as Hadoop and Apache Spark—is wiping away the distinctions between data types, and treats all data types equally. Never before has there been such an opportune time to assimilate all this data into information, at a level in which decision makers no longer need to be limited by data types.

2. REORIENT YOUR SKILLS PROFILE TO THE DIGITAL AND DATA ANALYTICS ECONOMY BECAUSE SKILLS ARE SCARCE.

Many organizations are scrambling—and competing with one another—for data management and analysis skills. Such skills don't just come directly out of colleges. A data-driven culture requires ongoing education and training for current employees. There is a new classification of jobs and skills that are emerging as this evolution progresses. For

example, the data scientist—one of the hottest job categories—didn't even exist a decade ago. Skills that are coming to the fore not only include data scientists but also titles such as cloud architects and knowledge management specialists.

3. INTRODUCE CHANGE MANAGEMENT STRATEGIES BECAUSE PEOPLE MATTER MORE THAN TECHNOLOGY.

The human element is a key part of moving to a digital and data-driven enterprise. Data analytics and management is moving from back-office settings to become part of everyone's jobs. For success at data-driven transformation, it's important to build these capabilities and new responsibilities into all jobs. Employees—technical and non-technical—need to be brought into the transformation process and develop comfort with data-driven decisions. Just as importantly, there needs to be a spirit of collaboration encouraged between business end users and data analytics and management teams.

4. WORK CLOSELY WITH THE BUSINESS BECAUSE DATA IS THE BUSINESS.

A data-driven culture requires strong analytical links to business performance. These can be measured and tracked through the adoption of key performance indicators. Data will play an increasingly vital role in monitoring the health, performance, and growth potential of the business.

5. OPEN SOURCE IS OK AND MAY EVEN BE PREFERRED BECAUSE OF ITS COMPELLING VALUE PROPOSITION.

Open source software has evolved dramatically, from the corners of development workshops to becoming the foundation of many data environments. If you are using cloud-based data services, chances are, the cloud provider is deploying many applications on open source technology. Likewise, open source solutions provide highly valuable and scalable options for on-premises data centers

6. INTRODUCE AND ENCOURAGE A SELF-SERVICE ETHIC BECAUSE POWER IS SHIFTING TO USERS.

There are a number of self-service data discovery tools on the market, but self-service means more than simply working with tools. The locus of analytical power has shifted to users, and with it, users are increasingly finding their own ways to discover, integrate, and incorporate data into their day-to-day decision making. This includes the ability to build lightweight front-end applications to access data. The future belongs to users or decision makers who are empowered to take their own initiative with data.

Providing information when it is needed, and in the right context, to decision makers—which may be at real-time speed—is the challenge of data delivery today.

7. MOVE TOWARD REAL TIME, BECAUSE BUSINESS SITUATIONAL AWARENESS IS CRUCIAL.

With the emerging Internet of Things (IoT), organizations will increasingly be acting on streams of data coming in from sensors and devices associated with products and services. The good news is that real-time-enabling technology, once requiring expensive hardware and software, is now available via cloud services and APIs.

8. RECOGNIZE THAT EVERY ORGANIZATION IS EVOLVING INTO A 24X7 DATA CENTER OPERATOR AND REMAINS CONNECTED TO CUSTOMERS LONG AFTER THE SALE.

With the rise of real-time and IoT-generated data, organizations will need to maintain constant connectivity with customers and the products or services they are using. Organizations will either need to contract with cloud or hosted application providers to provide real-time data center capabilities, or beef up their own on-premises facilities to meet this need.

9. EMPHASIZE VISIBILITY AND TRANSPARENCY BECAUSE DATA NEEDS TO BE TRUSTED.

By relying on data, decision makers not only need access to data but also must be able to understand its lineage, timeliness, and context. Data is not information until its value and meaning are understood to decision makers. Today's data management tools provide the log data and metadata that demonstrate where data came from, adding needed context and assurance. Trust is the most valuable currency within a digital-driven enterprise, and visibility and transparency provide that needed trust.

10. OPERATE LIGHT AND FAST BECAUSE AGILITY DELIVERS.

Organizations need to operate efficiently and quickly on information and cannot be tied down with long, expensive, and cumbersome processes for managing the data. The future will go to the tools, interfaces, and programming or scripting languages that enable users to rapidly squeeze the value out of data when and where it is needed. The more highly automated a data operation is, the more responsive it will be to changes in the enterprise. The software-defined data center enables the deployment or realignment of new tools and processes without the need to uproot systems or hardwired methods.

The future of data is already here, and it is highly distributed, highly accessible, and vital to enterprise success. In today's database world, there is no one right way to do things—there are multiple options now available to enterprise data managers and professionals. Data-driven capabilities can be supported in the cloud or on-premises and by open source or commercial solutions. We need new types of professionals, such as data scientists, to succeed with data, but we also need retraining for seasoned data staffs. Most importantly, today's decision makers need to be comfortable and assured that the data they are working with is of the highest quality and caliber possible. ■

—Joe McKendrick

Case Study

Balluff delivers high-speed information with IBM DB2 with BLU Acceleration in a private cloud

Family-owned Balluff GmbH, headquartered in Neuhausen, Germany, is a world-leading manufacturer of sensor solutions, employing around 2,750 people in 50 countries. Offering sophisticated technology, cutting-edge electronics products, application-specific customer solutions and highly trained service staff, Balluff's annual sales exceed EUR 300 million.

Over a period of rapid business growth, growing data volumes created practical difficulties; sales and business development reports were taking longer to produce, and processing tasks normally scheduled to run overnight were almost running into business hours. Switching to an environment that would provide real-time information would help Balluff achieve faster insight into markets and customers, and allow business users to explore data interactively to help drive the business forward.

As a first step, Balluff wanted to optimize its entire IT environment, including software, servers and storage, to increase the performance of its transactional SAP ERP environment and boost productivity of its business users across all departments. To achieve this, Balluff looked at innovative storage technology and a general server upgrade.

In a second step, the company aimed to improve decision-making with faster, more accurate business analytics capabilities for its data warehouse by leveraging the latest technology of IBM DB2 for Linux, UNIX, and Windows and its BLU Acceleration feature. Bernhard Herzog, Team Manager Information Technology SAP at Balluff, explains, "Balluff relies on SAP solutions to manage the business, including SAP Business Warehouse for data analysis and reporting. As the business grew, data delivery slowed down. Without timely, accurate information available, we risked making poor investment decisions and being unable to deliver the best possible service to our customers."

By improving internal processes, Balluff also sought to implement a more flexible approach to sales management and ensure excellent customer service.

"To realize our true business potential, we wanted to transform the speed and reliability of the enterprise information management, and turned to IBM for help," adds Herzog.

TECHNICAL SOLUTION

Balluff's IT team was not interested in incremental changes; it wanted game-changing improvements for its private cloud environment in information delivery across its transactional (OLTP) and analytics (OLAP) business applications. To achieve this, the company upgraded in a first step its SAP ERP applications and the underlying IBM server and storage infrastructure before deploying IBM DB2 with BLU Acceleration software in a second step to speed up its SAP Business Warehouse application.

FIRST STEP: RUNNING SAP ERP IN A PRIVATE CLOUD WITH IBM INFRASTRUCTURE

Server architecture

To optimize the performance of its SAP ERP applications with IBM DB2 for Linux, UNIX, and Windows databases and to ensure fast and reliable transaction processing, Balluff upgraded to two new IBM Power 750 servers.

Storage architecture

Following the server upgrade, Balluff optimized its storage architecture and added two all-flash IBM FlashSystem 840 storage devices.

Backup architecture

As the value of business data increases, it is particularly important to fully protect the database and other applications such as email and file servers with automated, reliable backup processes to ensure that data can be recovered quickly, should a disaster strike.

To meet these needs, Balluff teamed up with IBM Business Partner PROFI Engineering Systems and implemented IBM Tivoli® Storage Manager.

Project achievements and performance improvements

In a collaborative effort based on standardized best practices, the combined Balluff and IBM team migrated up to five SAP applications per day in parallel with a runtime of about three hours per system. IT staff were surprised to see how rapidly IBM had moved its systems over to the new servers and upgraded to the latest SAP and IBM DB2 releases. Herzog remarks, "We could never have completed this upgrade so quickly without professional help from experienced IBM specialists."

As a result of the IBM technology upgrade, Balluff cut response times of its SAP ERP applications powered by IBM solutions by 50 percent while also improving information delivery.

"Our business users noticed a real performance boost after we deployed IBM FlashSystem solutions for SAP ERP with IBM DB2 databases. One manager said that his team can now work more efficiently than before. As users complete their tasks more quickly, they have more time to provide better customer service," said Herzog.

By using IBM DB2 compression technology, the database is about 85 percent smaller than when using

uncompressed data, reducing the storage needs for backups and further speeding up the backup process.

The Balluff team is dedicated to the continuous optimization of its IT infrastructure and database architecture to support its broader business ambitions. By maximizing technical performance, Balluff can avoid over-investment and cut operational costs while delivering enhanced, near real-time information to executives.

SECOND STEP: BOOSTING SAP BUSINESS WAREHOUSE PERFORMANCE WITH BLU ACCELERATION

The company's SAP Business Warehouse application serves around 500 concurrent users, and the aim was to transform personal productivity and enable real-time enterprise-wide information insight.

The team evaluated possible options, including SAP HANA, SAP Business Warehouse Accelerator, and IBM DB2 for Linux, UNIX and Windows. The Balluff team looked at price, complexity and predicted system performance. Balluff selected IBM DB2, with the intention to leverage BLU Acceleration, as the most effective way to achieve its business objectives.

Bernd Dümmel, Team Manager SAP BW and Application Development at Balluff, says, "IBM DB2 with BLU Acceleration is the ideal solution for us as it enables rapid new insights into business data. Deploying BLU Acceleration was a low-risk project - implementation was quick, easy and did not affect availability. The BLU Acceleration feature is already included in our IBM DB2 licenses, and we can operate the solution without additional training. As a further bonus, investment and operating costs are much lower than other in-memory solutions."

The IBM team worked closely with Balluff to analyze the company's SAP Business Warehouse applications and identified about 1,500 tables in the 1.5 TB database that would benefit from the new BLU Acceleration technology. IBM specialists helped Balluff migrate these 1,500 tables in approximately 17 hours without downtime for the production systems. The table migrations were parallelized, enabling the migrations to be completed during operational hours and without the need for scheduled downtime.

Deploying truly innovative database technology

Herzog summarizes, "IBM DB2 with BLU Acceleration combines innovative features, such as columnar tables, with leading data compression technology and outstanding ease of use for the best benefit of the business." By bringing together a wide range of performance features, IBM DB2 with BLU Acceleration perfectly supports a data-centric approach where business applications such as SAP Business Warehouse push compute-intensive data processing workloads down from the application layer to the database layer. This application development paradigm complements recent advances in database technology and helps deliver faster response times because algorithms can be executed faster when they are implemented closer to the data they operate on.

Project achievements and performance improvements

Building on existing IBM DB2 software and skills, the biggest database implementation was completed online and in just one night. Dümmel, says, "The deployment of BLU Acceleration went very smoothly. No customizations to the database tables or our SAP Business Warehouse application were required. We could even move our data online without disrupting our business users."

Dümmel states, "Our nightly batch jobs were in serious danger of over-running into the next day's operations. Thanks to BLU Acceleration, we can now ensure that all the data is processed and made available on time. This innovative solution allows us to cover future business growth easily, even as we deploy new information services."

Balluff staff now enjoy 98 percent faster access to relevant business data. Dümmel elaborates, "For example, since we moved to IBM DB2 with BLU Acceleration we have cut database run time of complex reports by up to 98 percent. Across all business reports, we achieved an average saving of 30 percent. As a result, staff can access up-to-date information more efficiently and the SAP Business Warehouse application feels more responsive than in the past. Data management tasks are also performed much more quickly, saving us significant time in database administration, data loading and pre-processing."

Herzog continues, "Today, sales staff have instant access to real-time information on customer turnover and other important indicators. With faster access to key business data, sales managers at Balluff have a better overview, sales reps can improve customer service, and the company can increase sales."

Herzog adds, "IBM DB2 with BLU Acceleration has helped Balluff transform the speed and reliability of its enterprise information delivery, giving executives fast insight into critical information and helping us seize new business opportunities."

Dümmel concludes, "Based on our success, Balluff is working with IBM to optimize the database environment further. We are already looking forward to the next release of IBM DB2 with BLU Acceleration, when even more SAP applications and database objects will be supported to speed up additional systems." ■

To learn more about this success story and others like it, please visit **ibm.com/db2/luw/**.

IBM
www.ibm.com

Five Steps for Evaluating Fast Data Choices

Fast data is data in motion, streaming into systems from hundreds of thousands to millions of endpoints – mobile devices, sensor networks, social networks, financial transactions, logs, retail systems, telco call routing and authorization systems, and more.

Businesses are tapping fast data to be more relevant and valuable to their customers and using it to gain competitive advantage with more efficient and agile operations. Smarter customer engagement with real-time, context-aware applications has direct impact on business results. According to Emagine International's CEO David Peters, **"Subscribers receiving tailored real-time offers bought 253% more services."** Fast data systems operationalize the learning and insights that companies derive from "big data".

Telecommunications, mobile, advertising technology, financial services, and IoT sensor management are sectors benefitting from adoption of fast data strategies. But how can you move beyond strategies to results? Take a tip from management guru Peter Drucker, who said, "Strategy is a commodity, execution is an art."

In this new world of database technologies, how do you effectively execute the next stage of information management to handle the influx of fast data? Consider these five important steps:

1. Identify your fast data opportunity
2. Assess and leverage your existing infrastructure
3. Get agreement on success criteria for the project
4. Understand the technical implications
5. Prototype, pilot, refine

IDENTIFY YOUR FAST DATA OPPORTUNITY

Fast data challenges have common characteristics that make them easy to identify in the wild. Most important is the need for four main capabilities:

- Ingesting unbounded streams of data
- Analyzing the data stream in real-time
- Taking some action on the live data stream
- Exporting original data, as fast as it comes in to a data lake or big data analytics engine

ASSESS AND LEVERAGE YOUR EXISTING INFRASTRUCTURE

Start by looking at your existing infrastructure and evaluating what can be adapted to create business advantage with fast data.

Complexity is a drain, on finances as well as on time and resources. A simpler solution is easier to implement, easier to maintain, easier to live with. But make sure you don't try to simplify too much — you can lose performance or functionality if you do. Swiss Army knives do many things but often not as well as single, stand-alone tools. They represent a compromise — you give up the full functionality of each of the single tools for the convenience of being able to carry them all around in one compact tool.

GET AGREEMENT ON SUCCESS CRITERIA FOR PROJECT

Chances are good, especially if you're in a data-driven business, that your developers and architects are already thinking about — or implementing — a fast data approach. But do they know the scope of the problems the enterprise is trying to solve? Have they gathered user requirements and business constraints from all stakeholders? And how will you and they, together as a team, know when success is achieved?

Follow these steps:

1. Describe the project's purpose
2. Scope the project
3. Build a timeline
4. Define success
5. Document fully

UNDERSTAND THE TECHNICAL IMPLICATIONS

You've identified a fast data opportunity and scoped out the success metrics. Now it's time to choose technologies and consider the implications of the alternatives. In our experience working with hundreds of developers on many different fast data use cases, we've learned key questions and considerations to help identify fast data project(s) that will be successful:

- Are you addressing an analytics problem, an operations problem, or a combination?
- Is it real-time or batch?
- Is data integrity important?

PROTOTYPE, PILOT, REFINE

Working with the product in a real scenario, e.g. a prototype, will expose the "warts" that are often only discovered after living with the product for a while. Be realistic about timeframe and resources. Many small companies have few resources to dedicate to prototyping, especially prototyping with several alternative products. Unless you have multiple teams to put on this, it may be completely impractical to play with more than one candidate product at a time.

Download software and test against your use case.

CONCLUSION

Big Data, Cloud, streaming — these technology paradigms have driven massive investment, although quantitative benefits have been hard to document. Now that you know the problem you're trying to solve, you'll be positioned to choose the best fast data solution for your use case. ■

To learn more, check out: Fast Data Choices: Five Steps for Evaluating Alternatives Business and Technology Options

VOLTDB
www.voltdb.com

MEMSQL

Rethinking Lambda Architecture for Real-Time Analytics

Big data, as a concept and practice, has been around for quite some time now. Most companies have responded to the influx of data by adapting their data management strategy. However, managing data in real time still poses a challenge for many enterprises. Some have successfully incorporated streaming or processing tools that provide instant access to real-time data, but most traditional enterprises are still exploring options. Complicating the matter further, most enterprises need access to both historical and real-time data, which require distinct considerations and solutions.

Of the many approaches to managing real-time and historical data concurrently, the Lambda Architecture is the most talked about today. Like the physical aspect of the Greek letter it is named for, the Lambda architecture forks into two paths: one is a streaming (real-time) path, the other a batch path. Thus, it accommodates real-time high-speed data service along with an immutable data lake. Oftentimes a serving layer sits on top of the streaming path to power applications or dashboards.

SIMPLIFYING THE LAMBDA ARCHITECTURE WITH MEMSQL

MemSQL combines database and data warehouse workloads, enabling both transactional processing and analytics. Gartner refers to this as HTAP or hybrid transaction/analytical processing. MemSQL often fulfills the speed layer of the Lambda Architecture, providing in-memory performance to ingest and process streaming data. In our experience with Lambda implementations, we find that most organizations get hung up on details of the Lambda Architecture, introducing unnecessary technologies and workarounds to fit within the model.

USE CASES

Real-Time Analytics

We are in an era of ubiquitous big data, where it's not enough for companies to merely process data. Analyzing that data to detect patterns, which can be immediately applied to maximizing operational efficiency, is the real driver of business value. MemSQL delivers real-time analytics on a rapidly changing data set, making it an ideal match for the characteristics of the Lambda Architecture speed service. Other data stores have limitations that inhibit high-speed data ingestion, lack analytical capabilities, or cannot scale affordably. MemSQL delivers a complete solution: the ability to handle millions of transactions per second while simultaneously performing complex multi-table join queries. Let's dig into some of the features that make MemSQL a great solution for implementing the Lambda architecture.

Scalability

MemSQL uses a distributed shared nothing architecture that scales on commodity hardware and local storage, supporting petabytes of data. MemSQL is a memory-first, relational database that also offers a disk-based columnstore. In-memory optimization delivers high-speed data ingestion while simultaneously delivering analytics on the changing data set. The disk-based columnstore provides historical data management and access to historical data trends to leverage in combination with the "hot" data to deliver real-time analytics.

Multi-model, Multi-mode

MemSQL supports the ingestion of unstructured, structured and semi-structured data. Flexibility to align a structure to data in support of analytics meets the business requirements of the operation. Real-time analytics requires a real-time data structure, which MemSQL supports through a fully relational model. Furthermore, MemSQL supports the ingestion of unstructured and semi-structured (JSON) data into key-value pairs. Full ANSI SQL support makes MemSQL readily accessible to data analysts, business analysts and data scientists reducing application code requirements. Plugging data visualization and query tools into the analytics architecture delivers immediate value from data to the business. MemSQL also has extended SQL including JSON support. Traversing a JSON document is similar to SQL with extensions to traverse the key-value pairs.

Open Source Connectors

MemSQL offers users several connectors for smooth integration with other data sources. One example is MemSQL Streamliner: a fully integrated Apache Spark solution. Streamliner provides easy deployment of Spark — a critical component for building real-time data pipelines that delivers advanced data enrichment and transformation. MemSQL also has native integration with HDFS, Amazon S3 and MySQL, making it easy to important data from HDFS, as well as import and synchronize data from Amazon S3. ■

About MemSQL

MemSQL delivers the leading database platform for real-time analytics. Global enterprises use MemSQL to achieve peak performance and optimize data efficiency. With the combined power of database, data warehouse, and streaming workloads in one system, MemSQL helps companies anticipate problems before they occur, turn insights into actions, and stay relevant in a rapidly changing world. Download MemSQL at **memsql.com/download** or follow us @memsql.

Couchbase

NoSQL for Retail and eCommerce

From giants like Walmart and Tesco, to specialty brands like Fanatics and Nu Skin, retail and eCommerce leaders are adopting NoSQL database technology to build modern web and mobile applications, manage peak-load traffic, reduce costs, and drive revenue. For example:

Dixons Carphone – Saw a 56% increase in Black Friday online sales in 2015, leveraging Couchbase's NoSQL database to meet peak demand.

eBay – Leverages Couchbase's NoSQL database for multiple uses, including caching 100M+ session tokens per day.

Office Depot – Drives $4B+ in online commerce, relying on Couchbase's NoSQL database to serve rich product information with blazing speed.

TOP NoSQL USE CASES FOR RETAIL & ECOMMERCE

Companies are selecting NoSQL to support many use cases, including:

Shopping Cart – Numerous companies use Couchbase to power their shopping cart service, relying on Couchbase's built-in cache for snappy performance at scale.

Product & Pricing Catalogs – Online sellers love the flexibility of Couchbase's JSON data model and its blazing fast memory-first architecture to power services like product and pricing catalogs.

Profile Management – Many of the world's largest shopping sites rely on Couchbase to deliver fast and seamless login, authentication, and profile management.

Personalization – To deliver great personalized experiences, retail and eCommerce leaders are deploying Couchbase to store, update, and access consumer data and business rules with sub-millisecond latency.

360 Customer View & Loyalty – Companies are leveraging the flexibility and speed of Couchbase to aggregate and run near-real-time analytics on customer data from multiple channels – websites, mobile apps, call centers, chat sessions, social media, and in-store – and to manage loyalty programs.

6 REASONS RETAIL & ECOMMERCE LEADERS CHOOSE NoSQL

Retail and eCommerce leaders are choosing NoSQL to overcome limitations of legacy databases for several key reasons:

1. More modern architecture – Due to its flexibility, distributed design, and lower cost, NoSQL has become the de facto standard for modern, interactive eCommerce applications.

2. Easier, less costly scaling – NoSQL products like Couchbase with distributed architectures are expressly designed to effortlessly and elastically scale on low-cost commodity hardware – just add a node, rebalance, and you're done.

3. Better flexibility and agility – A NoSQL document database like Couchbase shares the same flexible data format (JSON) as most modern web and mobile applications – removing the "impedance mismatch" that slows developers down when having to use relational.

4. Higher availability and reliability – With a shared-nothing distributed architecture, along with features like built-in cross datacenter replication and auto failover, a NoSQL database like Couchbase delivers bulletproof availability and reliability.

5. 50-80% lower costs – Choosing an open source NoSQL database like Couchbase – designed to leverage inexpensive commodity hardware – typically reduces your database licensing and operating costs by 50-80% over what you spend on proprietary software.

6. Less complexity – Running high performance applications on a relational database requires numerous components and add-ons – such as separate caching and replication layers. With Couchbase – which integrates caching, storage, and replication in a single product – you have a much simpler solution.

WHY COUCHBASE IS A GREAT CHOICE FOR RETAIL & ECOMMERCE

Companies have chosen Couchbase for several key advantages:

Memory-centric architecture – Couchbase takes full advantage of all available memory to give your application the sub-millisecond responsiveness that today's shoppers expect.

Integrated cache – While other NoSQL databases like MongoDB require a third-party cache – adding to both cost and complexity – Couchbase has a fully integrated cache that delivers blazing performance. No need for a separate product to install and manage.

Powerful, SQL-based query language – Unique among all NoSQL document databases, Couchbase provides N1QL ("nickel") – a powerful, full-featured query language that lets developers easily query JSON documents using a familiar, standards-based syntax.

Built-in high availability and disaster recovery – Couchbase comes with high availability within a cluster and provides market-leading cross datacenter replication (XDCR) capabilities to support DR and data locality requirements with multi-master, multi-site deployments. No need for complicated third-party systems.

Always-on mobile support – Couchbase Mobile is a complete NoSQL solution for mobile application support. It includes Couchbase Lite – an embedded JSON database for devices – and Sync Gateway, a pre-built solution that syncs the device with the cloud. Couchbase Mobile lets you easily support use cases such as in-store personalized apps, point of sales systems, and mobile-optimized digital catalogs.

NoSQL – A BETTER FIT FOR RETAIL & ECOMMERCE APPLICATIONS

If you're running into scalability, performance, or availability challenges with your existing relational database – and you're looking to speed development and innovation – it's probably time to consider NoSQL technology. ■

COUCHBASE
www.couchbase.com

SharePoint SYMPOSIUM

JW MARRIOTT
WASHINGTON DC
NOVEMBER
16–17, 2016

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

Co-located with

KMWorld 2016

Enterprise Search & Discovery 2016

taxonomy BOOT CAMP >>>>

SHAREPOINTSYMPOSIUM.COM

Diamond Sponsors

brightstarr · coveo · SEARCH TECHNOLOGIES · synaptica knowledge organization systems

Gold Sponsor

EZONES

Media Sponsors

BESTSEOs comprehensive research & analysis · Business Wire A Berkshire Hathaway Company · CMS WiRE · CrowdReviews Buyers Guide Based On Client Reviews · InformationToday · KMWorld · TOPSEOs Independent Authority on Search Vendors · Visibility

Organized and Produced by

Information Today, Inc.

CONNECT #SharePointSym

APPLICATIONS ▸▸

The Seven Deadly Sins That Get Companies and Their Users Hacked

By Deepen Desai

It's rare now to read about breaches that resulted from direct attacks on the data center, as organizations have done a great job at securing their network infrastructures. But, with business applications moving to the cloud and mobile users connected all the time, traditional appliance-based security measures can no longer protect the modern business. The glass box is broken, the perimeter has vanished, and attacks have shifted to the weakest link: the user.

Users are routinely putting their organizations at risk by downloading unverified content, clicking on malicious links, bypassing security controls, or simply reusing old passwords. Yet, while it's easy to point the finger at the users, it's still our job as security professionals to be proactive and protect our users to the best of our ability. Yet over time, bad habits often slip into our security practices and we find ourselves taking shortcuts or making assumptions that we shouldn't.

Here's a reminder of some of these practices to avoid—the seven deadly sins that will get your users hacked.

Sin No. 1—Sloth

As it relates to information security, many just don't have the belief that they can stop the bad guys. Many organizations have adopted the mindset that they will be breached, which often leads to complacency and a willingness to accept increased risk. For example, did you know that most organizations don't inspect SSL traffic, despite the fact that that is where more than half of advanced threats hide? Well, it's true, according to Zscaler's ThreatLabz Research. Don't be one of those organizations.

*It's easy to point the finger at the users, but it is the job of security professionals to be **proactive**.*

Continuing down the path of complacency will not have a fairy-tale ending. Organizations need to think outside the box and to the cloud for the highest level of security and compliance.

Sin No. 2—Gluttony

In many ways, we in IT security are gluttons for punishment. We do risk assessments, which reveal security findings that drive us to a new piece of software or hardware that will work as a patchwork bandage for said findings, which then, ultimately, have us presenting a business case to buy more. The next quarter or year we do another risk assessment and the cycle repeats itself, with no end in sight to the smorgasbord of findings and bandages. In turn, this ecosystem of numerous disjointed security appliances makes it hard to maintain an effective security posture or scope of visibility, especially when insider threats, often by virtue of human error, are the top cause of data breaches.

It's imperative for organizations to break free from this vicious cycle and look for a platform that offers real-time correlation across threat prevention techniques so enabling new services is as simple as clicking a button.

Sin No. 3—Lust

Mobility is a way of life, and enterprises need to contend with every employee's yearning for fast and simple internet access. The need for speed and simplicity, fostered by the likes of software as a service (SaaS) solutions, has rendered the virtual private network useless. It's slow and often deemed an unnecessary hurdle to accessing the web. As a result, users go directly to the internet and, in the process, bypass security controls. Mobile device management and enterprise mobility management

solutions have surfaced in response, but although they're a great way to manage mobile devices, they don't protect mobile users from threats.

Your security is only a good as your weakest link, and having mobile devices bypass security controls is leaving you exposed. Proxy auto-configuration files and mobile agents are great ways to enforce security and compliance for mobile devices and provide a seamless user experience.

Sin No. 4—Pride

Ask just about any security architect within the top Fortune 500 ranks if they are proud of their design and they will surely say yes. But just months later, they're on the wall of shame facing a security breach. According to a recent KPMG survey, 93% of organizations responded that they had experienced infected computers with command and control servers; 52% observed malware coming into their network that was new and previously unknown to antivirus vendors; and 79% of organizations experienced some form of data leak. But when asked why, they said that it's not the architecture: It's the security analysts, who didn't see the indicators as they came through multiple-point solutions or the security program overall, for not providing the right level of staffing—anything but their architecture.

Security appliances in the data center were great at protecting your servers in the data center. But, now that attacks have shifted to your users, your security architectures need to adapt. Effective threat protection requires you to inspect every byte of traffic and use techniques such as behavioral analysis and machine learning to protect against zero-day and emerging threats.

Sin No. 5—Envy

It's OK to be more than a little bit envious of a company that has figured out how to address the majority of its security challenges by leveraging the cloud. It's better than buying piecemeal solutions just because another organization bought it too. In a world with boundless solutions, organizations need to make compromises between how much money they have and their level of security. And, quite frankly, we should all be modeling ourselves to be more similar to the organizations that can truly block those zero-day phishing attacks that many users will ultimately click on anyway, whether it occurs in the office, at home, or over SSL.

Cybersecurity can make or break a business—avoid the repercussions of a breach by putting the right security measures in place.

Fortunately, in a SaaS world, envy is a thing of the past. Salesforce, for example, is used far beyond the scope of the Fortune 100. While cloud security is quickly being adopted by some of the largest enterprises, it's also within reach of SMBs. With cloud security, even the smallest company can afford the same level of security as organizations with billion-dollar IT budgets.

Sin No. 6—Rage

Often, IT managers blame users when they get infected with malware or fall victim to the latest phishing attack. But it's not practical to blame a user for being infected while visiting a top-100 website that's been compromised. The real issue to focus on is how the malware got through the current security controls. It's the job of security professionals to keep users safe.

Blocking threats by matching signatures and destinations is no longer sufficient. To prevent attacks, security solutions need to inspect all traffic for malicious objects, JavaScript, code obfuscation, Zero-Pixel iFrames, images, and more.

Sin No. 7—Greed

Everyone wants more products, more responsibility, believing this equates to a better security stance. But there are simply too many interfaces, tools, and siloed programs to deal with, each one working in isolation, re-creating the wheel a hundred times instead of working through a consolidated front.

Thankfully, leading organizations are figuring out how to best leverage the cloud to solve these problems, elevating the entire security IT organization process to new heights. This allows security professionals to grow and scale their own capabilities far beyond what was previously believed possible.

Protect Yourself From the Seven Deadly Sins

In all, the seven deadly sins can be curbed with corresponding measures. Securely enable your business for cloud and mobility, block threats inline, distrust all traffic (inspect every byte), and eliminate information silos. Cybersecurity can make or break a business—avoid the repercussions of a breach by putting the right security measures in place. ■

Deepen Desai is the director of security research at Zscaler, a cloud security provider.

APPLICATIONS ▸▸

The Benefits of Scale-Out NAS for the Hybrid Cloud

By Stefan Bernbo

The world's data has doubled in 18 months' time. The industry estimates that the global amount of storage will reach 40 zettabytes by 2020. Historically, storage architectures were built on solutions that could only scale vertically. This legacy approach to storage presents significant challenges to being able to store the tremendous quantities of data being created today in a way that is cost-effective and maintains high levels of performance. Today, most of the world's data centers are still using vertical scaling solutions for storage, which means that organizations are seeking alternatives that allow them to scale cheaply and efficiently in order to remain competitive. And now, with software-defined storage moving forward, we see the use of more scale-out storage solutions in data centers.

Hybrid cloud presents a way for organizations to gain the maximum amount of business flexibility from cloud architectures, which helps maximize budget efficiency and performance goals at the same time. In a nutshell, hybrid cloud is a cloud computing environment that uses a mix of on-premises, private cloud, and public cloud services, with orchestration between the two platforms.

Since hybrid cloud architectures are so new, many are still learning about the benefits and challenges associated with deploying a hybrid cloud approach. In this article, we go through some design elements you can use to ensure your hybrid cloud delivers the performance, flexibility, and scalability you need.

Scale-Out NAS Is Critical

The cornerstone of this hybrid-cloud storage solution needs to be a scale-out network-attached storage (NAS). Since hybrid cloud architectures are relatively new to the market—and even newer in full-scale deployment—many organizations are unaware of the importance of consistency in a scale-out NAS. Many environments are eventually consistent, meaning that files that you write to one node are not immediately accessible from other nodes. This can be caused by not having a proper implementation of the protocols, or not tight enough integration with the virtual file system. The opposite of that is being strictly consistent: Files are accessible from all nodes at the same time. Compliant protocol implementations and tight integration with the virtual file system is a good recipe for success.

An ideal hybrid cloud architecture that incorporates a scale-out NAS approach should be based on three layers. Each server in the cluster will run a software stack based on these layers.

- The first layer is the persistent storage layer. We base this layer on an object store, which provides advantages such as extreme scalability. However, the layer must be strictly consistent in itself.
- The virtual file system is the heart of any scale-out NAS. It is in this second layer that features such as caching, locking, tiering, quota, and snapshots are handled.
- The third layer contains the protocols such as SMB and NFS but also integration points for hypervisors, for example.

It is very important to keep the architecture symmetrical and clean. If you manage to do that, many future architectural challenges will be much easier to solve.

Now, let's take a closer look at the storage layer. Since the storage layer is based on an object store, we can now easily scale our storage solution. With a clean and symmetrical architecture, we can reach exabytes of data and trillions of files.

*Hybrid cloud presents a way for organizations to gain the maximum amount of business flexibility from cloud architectures, helping to **maximize budget efficiency** and **performance goals** at the same time.*

Ensuring redundancy is one of the responsibilities of the storage layer, so a fast and effective self-healing mechanism is needed. To keep the data footprint low in the data center, the storage layer needs to support different file encodings. Some are good for performance and some for reducing the footprint.

Metadata: What Is It and Where to Store It?

A very important piece of the virtual file system is metadata. In a virtual file system, metadata are pieces of information that describe the structure of the file system. For example, one metadata file can contain information about what files and folders are contained in a single folder in the file system. That means that we will have one metadata file for each folder in our virtual file system. As the virtual file system grows, we will get more and more metadata files.

Some users prefer centralized storage of metadata. For smaller setups, this might be a good solution, but here we are talking about scale-out, so let's look at where *not* to store metadata. Storing metadata in a single server can cause poor scalability, poor performance and poor availability. Since our storage layer is based on an object store, a better place to store all our metadata is in the object store—particularly when we are talking about high quantities of metadata. This will ensure good scalability, good performance, and good availability.

Cache: Increasing Performance

Software-defined storage solutions need caching devices to increase performance. From a storage solution perspective, both speed and size matter—as well as price. Finding the sweet spot is important. For an SDS solution, it is also crucial to protect the data at a higher level by replicating the data to another node before de-staging the data to the storage layer.

Supporting multiple file systems and domains becomes more important as the storage solution grows in both capacity and features, particularly in virtual or cloud environments. Supporting multiple file systems is also very important. Different applications and use cases prefer different protocols. And sometimes, it is also necessary to be able to access the same data across different protocols.

Virtual Environments

The cloud element of the hybrid cloud naturally requires support for hypervisors. Therefore, the scale-out NAS needs to be able to run as hyper-converged as well. Being software-defined makes sense here.

If we have a flat architecture with no external storage systems, then the scale-out NAS must be able to run as a virtual machine and make use of the hypervisor host's physical resources. The guest virtual machine's (VM) own images and data will be stored in the virtual file system that the scale-out NAS provides. The guest VMs can use this file system to share files between them, making it perfect for virtual desktop infrastructure environments, as well.

Now, why is it important to support many protocols? Well, in a virtual environment, there are many different applications running, having different needs for protocols. By supporting many protocols, we keep the architecture flat, and we have the ability to share data between applications that speak different protocols, to some extent.

Being software-defined, supporting both fast and energy-efficient hardware, having an architecture that allows us to start small and scale up, supporting bare-metal as well as virtual environments, and having support for all major protocols make a very flexible and useful storage solution.

Hybrid Cloud: Features

Each site has its own independent file system. A likely scenario is that different offices have a need for both a private area and an area that they share with other branches. Only parts of the file system will be shared with others.

By selecting a section of a file system and letting others mount it at any given point in the other file systems provides the flexibility needed to scale the file system outside the four walls of the office. It is important to make sure that the synchronization is made at the file system level in order to have a consistent view of the file system across sites. Being able to specify different file encodings at different sites is useful, for example, if one site is used as a backup target.

Looking Ahead

This approach creates a next-generation hybrid cloud system. The result? Clean, efficient, and linear scaling up to exabytes of data. A single file system spans all servers, which offers multiple entry points and removes potential performance bottlenecks. This solution includes native support of protocols, flash support for high performance and flexible scale-out by adding nodes. With a scale-out NAS, you will have much better control over your investments and expansion in your data center. ■

Stefan Bernbo is the founder and CEO of Compuverde.

Supporting multiple file systems and domains becomes more important as the storage solution grows in both capacity and features, particularly in virtual or cloud environments.

MV SOLUTIONS

Revelation Software Hosts Enlightening User Conference

Revelation Software recently hosted its user conference in Orlando, Fla., for developers who wanted to learn more about all aspects of the company's technology, including the much-anticipated OpenInsight 10 release.

"It was a great place to not only work hard but, for our guests and employees, a place to unwind," said Robert Catalano, director of sales and marketing, Revelation Software. "We have customers all over the globe, and it's very nice that we're able to engage with them when we hold our conferences."

The main focus of the conference was driven by questions about the release of OpenInsight 10, which has been delayed due to an overhaul of the platform. "We went straight into talking about OpenInsight 10," Catalano said.

However, Revelation was able to provide attendees with an alpha version of the platform, allowing the company to garner feedback for further enhancements to the solution.

"We were able to distribute to each attendee an alpha version of the product so they can get a good feel for what's inside the product, although not 100% functional, because so much has changed. They can navigate through the product to see what's new, exciting, and what makes sense," Catalano said.

Revelation held informative sessions and roundtable discussions during the event that focused on remote desktop services and web form designing along with introducing tuning tools for apps and databases. "As our clients have big data, speed and efficiency are very important to their applications," Catalano said.

Feedback from attendees was overwhelmingly positive, according to Catalano. "They were pleased we were able to give them an alpha copy of OpenInsight to take home and attendees were thankful because they got to engage with their peers and that camaraderie may translate into business opportunities as well," Catalano said.

The next event will take place in about 18 months, Catalano added. "It will be somewhere between 18 months and 2 years from now," Catalano said. "It's a huge trip, it's a huge expense, and we understand that."

OpenInsight 10 was the main focus of the Revelation conference, which will be followed by another user event in about 18 months.

Rocket Software Releases Updated Version of Its D3 Linux Platform

Rocket Software is releasing an update to its D3 platform, providing users with more flexibility.

D3 Linux, version 10.2.0, includes a new capability for D3 administrators to pause and resume database write activity on disk, with a defined timeout, after which write activity will automatically resume. While database write activity is paused, user activity is not interrupted, and any user changes are cached in memory. While paused, the database is unchanging on disk, allowing for a database snapshot and improved backup performance.

Administrators can also use this state to take a snapshot of the database to be used for testing or other purposes. This new feature provides improved backup performance and gives administrators greater flexibility on how and when they take an image of the database. Once the administrator resumes database write activity, cached user changes are flushed to the disk, and activity resumes normally.

Entrinsik Expands Its Operations in Europe

Entrinsik Inc., developer of information management software, is expanding its international presence with the opening of a European office, meeting a growing need for reporting and data analysis solutions.

"With strong credentials across multiple industries and an international client base, Entrinsik can now better provide European organizations with the tools necessary to help them succeed and stay competitive in a highly dynamic business environment," said Doug Leupen, president/CEO at Entrinsik.

Entrinsik opened an office in Europe with the help of Eurodev, a Netherlands-based business development company.

This global expansion will allow Entrinsik to better support its existing client base in Europe, while continuing to aggressively grow the business internationally.

*Entrinsik's new European office will help strengthen its **global presence** in the business intelligence market.*

The company currently works with several distributors of its Informer software in Europe, Australia, and South America, and with the opening of Entrinsik Europe the company is building a team to consolidate all European sales and marketing efforts locally.

With its Informer platform, organizations across Europe can access key information, identify trends, uncover opportunities, and make data-driven decisions with confidence.

Entrinsik Informer business intelligence software enables organizations to perform ad hoc reporting and analysis, blending data from multiple sources to create interactive reports and visualizations. ■

BDQ

BIG DATA QUARTERLY

THE NEW PUBLICATION FOR THE ERA OF

BIG DATA

Brought to you by the editors of *Database Trends and Applications* magazine, *Big Data Quarterly* is for information management and business professionals who are looking to leverage big data in organizations of all kinds. Subscribe today and stay informed on big data, data science, and the technologies and business strategies surrounding them.

This is a must-read publication for data scientists, CIOs, and other professionals involved with big data projects.

LIMITED-TIME FREE OFFER!*
SUBSCRIBE NOW.

dbta.com/BDQ/Subscribe

*Free to qualified U.S. subscribers.
Regular subscription rate is $95 per year.

GERARDO DADA

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

Seven Pathways to the Cloud

According to a recent SolarWinds survey, almost all organizations have migrated at least some infrastructure to the cloud. In fact, just 9% of IT departments have not migrated anything. Furthermore, databases rank in the top three for infrastructure already migrated to the cloud and for infrastructure with the highest priority for future migration.

If you're one of the few who have not migrated any workloads to the cloud, hopefully, this helps you realize that the cloud is in your future. And if you're one of the many who have already migrated some of your workloads offsite, know that more cloud is in your future (though, surely not all of your infrastructure will find its way to the cloud—hello hybrid IT).

The key question is really how to embrace the cloud: What are the options to transform applications to take advantage of cloud services? Let's explore the seven ways you can progress down the cloud pathway.

First up—do nothing. Even if you "do nothing," you will still likely have a hybrid IT environment, either now or in the very near future. The reality is that nearly every organization uses software as a service applications already. In fact, most marketing departments alone use 30 to 70 cloud-based applications, and most of these applications require integrating with data sources hosted on-premises in some way.

To ensure success with any type of cloud deployment, it's important to develop ***a cloud/hybrid IT mindset****.*

Second, most organizations will eventually transform their internal IT departments to look awfully similar to a private cloud—an elastic, software-provisioned, and measured service. Some companies will go the virtualization route, while others will implement an actual private cloud platform such as OpenStack.

Third, do your testing and development in the cloud. Test/dev environments are notorious for their infrastructure requirements and ephemeral nature, which make them a natural fit for the cloud.

Fourth, "cloudify" existing applications. This entails moving components of existing applications or database workloads to the cloud. For example, most web applications should be storing images, large files, and videos in the cloud, where they can enjoy the benefits of a content delivery network (CDN) and get the load off of web servers. The primary databases may still be hosted on-premises, but you could store backups or archives in the cloud.

Fifth, "lift and shift" your existing applications to the cloud. Once an application has been developed and tested in the cloud, and perhaps its disaster recovery site also hosted, it might make sense to move the application itself to the cloud too. The term "lift and shift" implies that there are no major architectural changes and that the cloud is being used almost as an extension of hosting. During this process, it's not uncommon to also cloudify some components of a workload, using a CDN, as explained before, or using a database as a service (DBaaS) instead of a full database instance on a dedicated cloud server.

Sixth, when replacing an application with an altogether new one, consider starting the new application in the cloud. Let's say an IT department is replacing an internal portal and collaboration tool with the latest version of SharePoint. It would be smart to evaluate how such a workload could run in the cloud. These projects are usually line-of-business applications that also maintain their traditional architecture, aside from perhaps taking advantage of a DBaaS or database mirroring.

Finally, develop a cloud-first application. Cloud-first applications are developed specifically to run in the cloud and architected to take advantage of cloud services. They embrace service orientation, APIs, software-provisioned hardware, and built-in redundancy. Everything is software—databases, encryption, mirroring, and backups are a service. They often use "modern" languages and frameworks, such as JavaScript/HTML5, NGINX, and Perl, and open source components, including NoSQL databases, such as Cassandra or MongoDB.

The reality is that your IT department will likely end up with a mix of these seven cloud deployment models. To ensure success with any of them, it's important to develop a cloud/hybrid IT mindset: Build a cloud adoption road map based on a workload-by-workload evaluation that considers requirements, potential upside, costs, and urgency. And lastly, implement a monitoring and management toolset that surfaces a "single point of truth" across on-premises and cloud workloads. This normalization of metrics, alerts, and other collected data from applications, regardless of their location, will enable a more efficient approach to remediation, troubleshooting, and optimization. ■

MARIA ANDERSON

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

Digital Leadership and the Strategic Importance of Technology

Digital transformation is taking place at an accelerated rate in the world today. CIOs are struggling to maintain legacy systems, while the world around them continues to transform at a rapid pace. At the same time, clients are expecting IT professionals to support the latest technology that becomes available. As IT continues to lag behind the digital wave, clients are becoming impatient and moving toward cloud technologies such as PaaS, IaaS, SaaS, and DBaaS to address their business requirements.

What can be done about this gap that continues to widen? In my opinion, two events must occur: First, IT professionals must learn to embrace the rapid change and become resilient, but, more importantly, business clients, customers, and partners must recognize that they need savvy IT professionals at the leadership table to advise them of this new world full of technology potential. The reality is that most organizations do not have a CIO (or equivalent) sitting at the executive leadership table. The CIO typically reports to another EVP in the organization. With this type of reporting structure, it will become more challenging for organizations to make good decisions as the technology becomes more complex and integrated.

Why this discrepancy? In my experience, IT has not been given the respect it deserves as a profession. There is little understanding about how powerful it can be for organizations to take advantage of technology with the help of IT professionals. In a 2012 research initiative at the MIT Sloan School of Management, organizations that embraced digital transformation in an innovative, yet responsible, manner were 9% to 26% more profitable than their competitors ("The Digital Advantage: How Digital Leaders Outperform Their Peers in Every Industry"). There are different types of digital maturity, according to this study, and not all of them are immediately profitable. Digital transformation on a large scale requires an investment, a solid vision for how technology can improve business process, governance, and a willingness to stay the course. This type of transformation cannot possibly be successful without IT representation at the executive level.

Business clients, customers, and partners must recognize that they need ***savvy IT professionals at the leadership table*** *to advise them of this new world full of technology potential.*

In a recent *strategy + business* article titled "Your Next Board Member Should Be a Geek," not having a technologist involved at the executive level of an organization was described as being equivalent to a company not having any directors with financial expertise on its board. Yet, that is precisely what happens when it comes to IT. "Boards can no longer duck the responsibility for the company's digital transformation," the article noted. "They must take real ownership by ensuring that they are equipped to fully understand this part of the board agenda." If this is indeed true, and I believe it is, inviting IT to the table is crucial.

The other side of this story is that CIOs and IT professionals must be willing to let go of what IT looked like in the past. The nostalgic world of IT from 5 or 10 years ago is long gone, and, if there is any hope of catching that digital transformation wave, we need to step up and take advantage of opportunities to learn about what these new technologies mean in terms of security, governance, data visualization, predictive analytics, and mobility.

There is tremendous opportunity to not only sharpen technical and architectural skills, but to consider leadership skills (regardless of one's title) to be equally important, if not more so, in many situations. This is where connecting with the larger IT community through users groups such as the IOUG can help. The wave is here—grab your surfboard and enjoy the ride! ■

KEVIN KLINE

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

SQL Server on Linux—It's No Joke

On March 7, 2016, Microsoft announced the beta release of SQL Server on Linux with the intention of shipping a full release of the product by April 2017. Many thought it was an early April Fool's Day prank.

But this is no joke. SQL Server is already running on Linux in Azure, Microsoft's cloud service, and is undergoing beta tests by MVPs and members of Microsoft's TAP program.

Linux Equals Open Source, Right?

Under former Microsoft CEO Steve Balmer, Microsoft became "Fortress Microsoft," as the company sought to crush or contain any rival technologies of significant threat. Balmer likened Linux, and its open source origins, to "cancer" and "communism." Since it was a threat to Microsoft's very lucrative server-based Windows OS, it had to be contained wherever possible.

SQL SERVER DRILL DOWN

But Microsoft has made an enormous bet on Azure, its cloud platform. There, customers pay for performance, not necessarily for specific OS licenses. That means Microsoft doesn't have to defend that turf as ferociously as in years past. Plus, adding Linux support opens up a new world of opportunity that has been a strong point for Linux, such as containerized applications running under Docker and big data applications harnessing the power of the Hadoop-HBase-Hive stack.

*Under the new Linux-inclusive strategy, **everyone has a pathway** into the Microsoft ecosystem.*

Microsoft is sticking with solid commercial vendors such as Ubuntu and Hortonworks, as well as Canonical and Red Hat. Still, many have asked, "Does this mean Microsoft is friendlier to open source?" The signs are definitely in favor of a new era of détente between Microsoft and the community-centric open source.

For example, Microsoft has already open-sourced the ASP.NET programming code and moved most of its code off of its proprietary CodePlex code sharing website to the GitHub website. I speculate that its Windows products, such as Active Directory and the network stack, are likely to follow soon.

That doesn't mean we'll be getting open source versions of SQL Server or Windows anytime soon. But Microsoft no longer discounts the Linux and open source world as rivals to be conquered. Instead, it now considers them valid platforms to support.

Transforming From Fortress Microsoft

Under new CEO Satya Nadella, Microsoft has made a variety of overtures to markets where it was previously hostile, such as opening up the Office product line to iOS and Android devices. A notable characteristic of the iOS and Android markets is their vociferous and active communities.

The Linux community, especially developers who want to write code on a Linux-based stack, is even more active. But, by offering a solid database to compete against MySQL and PostgreSQL (two database platforms I write about in my book *SQL in a Nutshell*), Microsoft is much better positioned to gain mind share, market share, and, most importantly, a solid footing in large enterprises that previously used only open source products.

This is a great strategic plan. Microsoft tried and failed with several aggressive strategies, such as "Windows Everywhere" and the introduction of a broader range of Windows smartphone devices with the disastrously expensive purchase of Nokia. Meanwhile, cloud computing was experiencing rapid growth, without any need for Windows Servers. Microsoft clearly has identified the need to secure a position as a cross-platform enterprise computing company. That path, now more than ever, includes a trip through the Linux world (http://blogs.microsoft.com/blog/2016/03/07/announcing-sql-server-on-linux/#sm.0009dqofs14t8en3pv31brjjap10l).

Now, All Roads Lead to Rome, er, Microsoft

The level of transformation happening at Microsoft is colossal. When Microsoft implemented its plans to be a "cloud-first" company, I doubt that the executives realized that their new direction would also alter the company's culture and its perception in the wider market. Under the new Linux-inclusive strategy, everyone has a pathway into the Microsoft ecosystem. Developers used to grumble about the evil "lock in" that came from anything Microsoft built. Now, most Linux developers are excited to see where this is heading. If Microsoft continues to consistently add these small, incremental wins, it is conceivable that those same grumpy developers will see Microsoft as a valid addition to their Linux-based data centers. ■

A longer version of this article appears at www.dbta.com/Columns/SQL-Server-Drill-Down.

CRAIG S. MULLINS

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

The Importance of Keeping Your DBMS Up-to-Date

Keeping your DBMS software up-to-date can be a significant job. The typical release cycle for DBMS software is every 18 to 36 months for major releases, with constant bug fixes and maintenance updates delivered in between those major releases.

In a complex, heterogeneous, distributed database environment, a coherent upgrade strategy is essential. Failure to plan your upgrade can result in improper and inefficient adoption of new features, performance degradation of new and existing applications, and downtime.

An effective DBMS upgrade strategy must balance the benefits against the risks of upgrading to arrive at the best timeline for migrating to a new DBMS version or release—and there are risks when upgrading to a new DBMS release.

A DBMS upgrade can cause disruption to business operations. At a minimum, databases may not be available while the DBMS is being upgraded. This can result in downtime and lost business opportunities if the upgrade occurs during normal business hours (or if there is no planned downtime). Clustered database implementations may permit some database availability while individual database clusters are migrated to the new version. Other disruptions can also occur, such as having to convert database structures or when features are deprecated. Delays to application implementation timelines are another possibility.

The expense of an upgrade can be a further barrier to release migration. The cost of the new version or release must be budgeted. DBMS vendors typically increase the price of a new DBMS version by as much as 10–25%. The upgrade budget must also factor in the costs of planning, installing, testing, and deploying not just the DBMS but also any applications using databases. Finally, be sure to include the cost of any new resources required to use the updated features delivered by the enhanced DBMS version.

There are many rewards that can be gained by upgrading to a new DBMS release, though. Often, developers can avail themselves of new features and functionality delivered only in the updated release. Quick ROI for upgrading may be achievable when program development time can be reduced or made more cost-effective. And new DBMS releases usually deliver enhanced performance and availability features that can optimize existing applications.

DBMS vendors provide better support and respond to problems faster for a new release of their software — so maintainability improves by upgrading.

There may also be cost-savings that accrue by upgrading to a new DBMS release. Older versions that have reached their "end of support" date can be expensive to manage. Sometimes, extended support can be purchased for "unsupported" versions but usually at a very high price.

After weighing the benefits of upgrading against the risks of a new DB2 release, the DBA group must create an upgrade plan that works for the organization. Sometimes, the decision will be to upgrade immediately upon availability, but often there is a lag between the general availability of a new release and its widespread adoption. Your organization style will come into play when deciding on an upgrade timeline.

Industry analysts at Gartner developed an organizational ranking system with three distinct groups: Types A, B, and C. Type A is technology-driven and will be more likely to risk using new and unproven technologies to try to gain a competitive advantage. Type B is less willing to take risks but will adopt new technologies once others have shaken out the bugs. Type C is conscious of cost and averse to risk so it will lag behind when it comes to technology migration.

Only type-A organizations should plan on moving aggressively to new DBMS releases immediately upon availability if the new features of the release will deliver advantages to the company. Type-C enterprises should adopt a conservative strategy to ensure that the DBMS release is stable and well-tested by types A and B first. Type-B organizations fall somewhere between types A and C. Rarely upgrading immediately, the type-B company will adopt the new release after the earliest users have shaken out the biggest problems but well before type-C enterprises.

Some type-C (and, perhaps, type-B) organizations try to get by running DBMS versions that are no longer supported by the vendor. This might seem to be a less risky strategy than upgrading to a new version, but it actually can be more troublesome. Unless you negotiate a high-cost, limited-support contract from the vendor, any bugs or errors you encounter can bring your organization's applications to a standstill. Running unsupported DBMS software is not a wise course of action if your organization is conducting mission-critical services using that DBMS. ■

TODD SCHRAML

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

Relational and Unaware

Many of the NoSQL tools out there, such as MongoDB, Couchbase, Hadoop, and others, purport to be leading a revolution and breaking the bonds of servitude to the restrictive, inflexible, established, relational market. They claim users need more, users need better … and they are there to help. Of course, when speaking about those relational flaws, the comments always focus on problematic aspects of a DBMS' physical implementation. One vendor lauds striping data content across storage devices as being a NoSQL improvement, completely ignoring the fact that several relational DBMS vendors employ that tactic as well. Are these tools offering any theoretical foundation, some version of logic behind them, to provide some substance to the case for usurping set theory or predicate calculus? No. OK, so there is graph theory behind the graphical databases, but that graph theory already existed behind hierarchical databases when relational databases replaced most of the hierarchical databases back in the day.

Many of these NoSQL rebels are also retrofitting SQL or some variant into their solutions in order to reassure users that they have a familiar and comfortable interface. Admittedly, SQL has always had some controversy surrounding how well it really supports relational theory or even how tight a language SQL may or may not be. But those thoughts aside, as these NoSQL vendors keep harping on the performance of their physical implementations over traditional relational databases' physical implementations, are they really doing anything more substantial than rehashing old arguments over C# versus Visual Basic? At times, it seems the vendors are simply caught up in their own rhetoric. If any of these sales teams truly understood relational theory, they would understand that relational theory is *not* about any physical implementation. Relational theory is about the users of the data. Users need to be able to consider the data as if it were tables with columns and rows. There is no requirement that the internals be implemented as columns and rows. Key-value pairs, documents, columnar, graphs, or hierarchies may or may not exist; relational theory does not care, because any physical implementation is fine. As each of these tools incorporates a SQL-type interface, they are allowing users to start conceiving these data stores as tables with identifiable rows and columns, which means these tools are relational, at least as far as relational theory is concerned.

Does this mean that these tools are implemented in the same restrictive, logical-must-be-exactly-physical way that most SQL DBMS vendors have implemented? Thank goodness, no, and viva la difference! The relational database world has been held hostage by too many vendors who could see only a world where logical and physical must be one and the same thing. In embracing that coalescent logical-physical view, the industry has been denied the possibility of what might otherwise have been. Long term, I believe many of these new tools will merge with, and into, the existing SQL DBMS market, and we all will have much stronger and flexible relational tools as a result. Relational theory says that the content of a unique column within a unique row must be "atomic." But atomic is one of those fuzzy words. Similar to the way that one man's ceiling is another man's floor, one person's atomic value can be another man's universe. Atomic does not necessarily mean a single string, either numeric or Boolean. Atomic could be interpreted to mean that a value needs to be singular from some perspective, not atomic from any and all perspectives. Users do need more, and users do need better. The established SQL DBMS vendors have grown soft by focusing on rationalizing their own shortcomings, and in so doing, they have lost their innovative edge. The NoSQL/big data efforts have provided a seriously needed shot across the bow; we can only hope that the established vendors use this call to arms appropriately. ■

Dealing With Big Data's Trough of Disillusionment

GUY HARRISON

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

I have an increasing sense of a building "trough of disillusionment" around big data. For those who haven't encountered the term, the "trough of disillusionment" is a standard phase within the Gartner hype cycle. New technologies are expected to pass from a "peak of inflated expectations" through the trough of disillusionment before eventually reaching the "plateau of productivity."

Most new technologies are expected to go through this trough, so it's hardly surprising to find big data entering this phase. And, of course, the bigger they are, the harder they fall—the expectations of big data have been incredible, so the potential for disappointment is all the greater. But, in the case of big data, it has been obvious for a long time that many early adopters were being set up for a disappointment. At its core, big data projects have two critical success factors: one, establishing the mechanisms for acquiring, storing, and processing massive data, and two, developing the algorithms to more effectively leverage that data for competitive or other advantages.

BIG DATA NOTES

Prerequisite No. 1 is not trivial, but technologies such as Hadoop and Spark at least provide a fairly accessible recipe for success. However, the second objective—which is directly tied to the overall payback for a big data project—is much harder to achieve.

To successfully develop the sort of algorithmic breakthroughs that big data projects promise requires at least three distinct ingredients: strong statistical and data mining expertise, the ability to create software that can implement these algorithms, and the business savvy to identify the problems that these algorithms should solve. Early successful big data companies such as Google and Amazon were able to hire rock star data scientists who combined these three success criteria. However, there simply aren't enough of these rock star data scientists to go around. In addition, while universities are graduating an increasing number of appropriately trained professionals, a truly successful data scientist requires years of experience to develop the sort of judgment and imagination required to create innovative big data solutions.

In my opinion, a big part of the problem is that data science is still bogged down in the minutiae of specific mathematical algorithms. Open any book on data science, and you are likely to begin with a long discussion of some seriously complex mathematical techniques. You need to learn the difference between dozens of algorithms such as K-Nearest Neighbors, Support Vector Machines, logistic regression, and K-Means before you can aspire to develop machine learning solutions.

However, the exact algorithm employed in a data science project is not necessarily decisive. Generally, data science problems involve a few high-level techniques—extrapolation, clustering, and classification. We need software packages that work at a higher level of abstraction, hiding the details of algorithms while allowing data scientists to work at the solution level.

KeystoneML—one of the latest offerings from the AMPLab group, which gave us Spark—attempts to provide such a high-level tool. KeystoneML allows data science problems to be specified in general terms, with the selection of algorithms left to the discretion of the framework. KeystoneML is analogous to the database query optimizer that is ubiquitous within relational database systems. In an SQL statement, you specify the results you want without having to specify the exact access path you expect the database to use. The optimizer determines the most effective path using sophisticated algorithms. In the same way, KeystoneML does not require you to specify the exact algorithms to be used to solve your machine learning problem; the KeystoneML engine will determine the best algorithms to achieve that goal.

Increasing the productivity of data scientists is going to be essential if big data projects can pay off. Initiatives such as KeystoneML are a step in the right direction. ■

Join Us at the **Hottest** Enterprise IT Event This Summer

SHARE Atlanta 2016

July 31–August 5
Atlanta Marriott Marquis

Looking to grow your career?

SHARE events are designed to propel you forward through continued education and professional development. This summer, attendees have access to:

- **500+** Technical Sessions, including hands-on labs and expert panels
- **New! Hot Topic** New and Innovative Workloads for z Systems
- New to z Systems
- SHARE ACADEMY

Save $200 on registration through June 17, 2016.
www.SHARE.org/Atlanta