database
TRENDS AND APPLICATIONS

database
TRENDS AND APPLICATIONS
ONE COMPLETE MARKETING PROGRAM

DATA INTEGRATION
for the Modern Enterprise

Best Practices Series

database
TRENDS AND APPLICATIONS

SEVEN GOOD REASONS for Data Integration Now

Best Practices Series

DATA HAS BECOME the secret sauce to corporate innovation and success. But we're not just talking about the raw bits and bytes that are jamming up storage systems and network feeds. In and of itself, data has little value or use. But when applications or algorithms are applied against this data, it becomes an illuminating beacon, shining lights and opening business minds as never before.

Data is the essential fuel of the data-driven digital enterprise everyone is seeking to build. But it needs context and consistency. In many ways, data integration is the prerequisite for digitization. It's especially important as time is an essential element within the digital enterprise, as time-to-insight and time-to-action are critical competitive differentiators. The demand for quick, easy access to information among decision makers is growing.

As the digital enterprise grows, so do efforts to combine data into meaningful information. A recent survey by Unisphere Research, a division of Information Today, Inc., among members of the Independent Oracle Users Group confirms that the era of data silos is coming to an end. Organizations are increasingly embracing universal approaches to managing data across their entire asset base. However, the survey also found, the tooling to accomplish this has yet to catch up. The survey results indicate customers have or are planning to consolidate databases to gain efficiencies and savings from consolidation. About one-third already run common converged infrastructure or are in the process of consolidating these environments. Another 26% are planning or considering such consolidation ("Consolidate, Virtualize, Automate, and Standardize: 2016 IOUG IT Resources Survey," Unisphere Research, September, 2016).

At the same time, there's a recognition that data is a key component of competitiveness in the new digital economy moving forward. A recent survey from Capgemini found that 64% of respondents agreed that big data is changing traditional business boundaries and enabling non-traditional providers to move into their industries; 53% said data-driven startups are already disrupting their industry spaces.

The challenge, then, is to have all the data needed at the ready to compete in this data-driven environment—and not be slowed down manually prepping needed information. A new generation of technologies is emerging to enable flexible, agile, and efficient data integration approaches—from self-service tools and platforms to cloud-based services and real-time solutions.

Such solutions leverage cloud, collaborative, and analytical platforms and technologies that are designed

to build data ecosystems in which developers, designers, users, and suppliers can exchange ideas and techniques while delivering high-speed processing capabilities. Such solutions and approaches are intended to provide a common service layer that abstracts underlying databases and applications while encouraging the sharing of information. Today's new generation of cloud-born solutions lighten the load even further for organizations, enabling high levels of scalability while relieving enterprise IT and data departments of needing to tend to the plumbing and connectors underneath their separate data stores. Environments such as Hadoop also bring various forms of data together, and these, too, can be hosted in the cloud.

Here's how data integration initiatives can propel the digital enterprise forward:

AUTOMATE AND SIMPLIFY DATA DISCOVERY, CAPTURE, AND ANALYSIS.

Business analysts are challenged with too much data and are often tasked with seeking small nuggets of information amidst oceans of data. They also need to look across data silos to have a better-rounded picture of the business challenge they are addressing. Data integration on the back end—whether in the cloud or on-premises—frees these analysts up to focus on their business problems or opportunities rather than to fiddle with manually scripting solutions to rationalize data.

GET A BETTER GRIP ON THE INTERNET OF THINGS.

IoT is poised to dramatically alter corporate business models. For example, if products are able to communicate their status back to the company that makes them—such as elevators signaling when maintenance is required—organizations can shift from the role of manufacturer to that of 24x7 service provider. The ability to keep lines of communication open with sensors, software, or devices in the field will propel many businesses into this new era.

PROVIDE A WORKING FOUNDATION FOR ARTIFICIAL INTELLIGENCE (AI) AND MACHINE LEARNING.

Data is the fuel of AI and machine learning, which is seen as the force that will drive automated decision making within the digital enterprise. Data needs to be constantly fed and refreshed into AI or machine learning algorithms. As demonstrated by large web companies such as Google and Amazon, large datasets can be employed within continuous feedback loops within online systems to support real-time transactions. Now, mainstream businesses increasingly have access to the same kind of AI and machine learning power and, most importantly, vastly expanding datasets to make such advancements possible. By integrating data sources at the back end, the possibilities are limitless in this bold new field.

GROW THE BUSINESS, AND OPEN UP NEW PRODUCT LINES.

The availability of information opens up new ways to better understand markets and customers as well as a two-way interaction for greater innovation and design. No longer are products built and thrown out to the market—the name of the game now is collaboration with customers and partners who play an active role in the initial design and development of the products they consume. This open approach to innovation is data-driven and requires a robust data infrastructure. In addition, with data integration in place, there are opportunities to enable data to be packaged as a product in and of itself.

REDUCE DUPLICATION, MAINTAIN ONE VERSION OF THE TRUTH.

One of the greatest challenges organizations face today is that data on the same customers or transactions is often managed separately within the domains of different departments. Without data integration, different flavors of the same data may be managed and stored in different places across the enterprise. In a digital business, which relies on multiple channels of contact or engagement with customers and partners, a single, real-time view is essential. At the same time, data is put to work where it counts. Different departments need data for different reasons, and this may vary from one moment to the next. Within traditional siloed environments, decision makers were either cut off from data that could help them, or unaware that certain reporting even exists. Or, there may be hard-coded data feeds set up that are no longer used, because business requirements have changed.

LEVERAGE DATA OF ALL TYPES, REGARDLESS OF SOURCE OR FORMAT.

For decision makers, a prime roadblock to gaining an expansive view of their environments was the many forms important data takes—from relational data to social media to spreadsheets. This wide assortment of data and data types needs to be available at a meta layer for enterprise users.

HELP CUT COSTS AND INCREASE EFFICIENCY.

Relying on data to help automate data-driven decision making means services and information are delivered to customers faster and more efficiently. Plus, these systems are capable of scaling to meet new growth and opportunities while avoiding a proportionate increase in costs associated with ramping up staffing or facilities. Data emanating from operations can also provide insights on where corporate processes are not operating at peak performance, or if there are bottlenecks, thus providing managers with informed streamlining opportunities. Such information can also serve to help maintain the efficiency and cost-effectiveness of the data management systems themselves. ■

—Joe McKendrick

Unified Governance Helps Make it Possible to Discover Hidden Business Insights

MANY BUSINESSES TODAY are data driven, with both structured and unstructured data serving as the lifeblood of their daily operations. But in many cases, businesses fail to tap the full value of their data. For example, unstructured data in the form of emails, documents, meeting notes, call center notes, social media data and more represent an enormous store of business intelligence that often lives a short life before passing into an archive. Transactional data contains valuable market insights that may only become apparent when analyzed in the context of other structured and unstructured data.

The siloed nature of enterprise data presents challenges for businesses aiming to extract greater value from their data. With different applications and organizations across the enterprise capturing data, it can be difficult to find all your data, access it, analyze it, and even manage data for compliance purposes.

Unified Governance is an integrated approach that looks at all structured and unstructured data across the enterprise so that you can more effectively discover, enrich, integrate and manage data over its lifetime.

This article examines 4 key questions about Unified Governance:

- How can you identify your data assets?
- How can you make them available within your company?
- How can you harness greater insights to gain more value from your data?
- How can you manage your data according to regulatory requirements?

BREAK DOWN DATA SILOS BY ENABLING THE FINDING, MATCHING, AND TRANSFORMATION OF YOUR DATA FROM A VARIETY OF SOURCES

Organizations need to be able to find their data and render into usable formats so they can make informed business and legal decisions. The first step towards successfully implementing your Unified Governance strategy is to have the right capabilities to discover and integrate all your data assets, from all data sources, including both structured and unstructured data.

Discovering data assets involves more than just finding your data. You must also cleanse, transform, and deliver it into integrated data structures where it can be analyzed. There are other considerations as well. Businesses typically secure and govern access to data, not only to protect those data assets, but to comply with regulatory requirements. Transforming and integrating the data does not eliminate the need to secure it. Also, collecting data into integrated data structures means having a master data management system that provides trusted views of this data.

By performing data discovery for all data sources across the enterprise, Unified Governance can connect and integrate data from different areas of operation. Businesses are then able to synchronize disparate data types, allowing a complete contextual view of what is happening in the business. Business stakeholders can gain a more complete view of information relevant to operational and strategic decision making.

MAKE BOTH STRUCTURED AND UNSTRUCTURED DATA AVAILABLE THROUGH A SELF-SERVICE MODEL

Once you have discovered your data, you need to turn it into trusted information. The real value comes from your business users being able to access and analyze it. The most effective Unified Governance solutions can deliver a self-service data access model that provides ways to search and find the data you need.

Unified Governance enables you to define policies that make the data available to those who need it, while helping to ensure it is not misused or its sensitive information exposed. In this way, users can access data that is designated for them while gaining an understanding of the legal and regulatory restrictions that come with access.

Unified Governance can deliver this kind of user experience for data assets, both structured and unstructured, through dictionary style catalogs that include taxonomy, terminology, rules and other customizable metadata associated with the data. It helps build trust in data by cleansing and customizing data in near real-time, and it harnesses the power of integrated metadata. These capabilities make it easier to search for, compile, and analyze data, which can make analytical insights deeper and faster.

TURN COMPLEX BUSINESS DATA INTO BUSINESS VALUE BY DISCOVERING HIDDEN INSIGHTS

Through 2021, 80% of successful CDOs will have value creation or revenue generation as their #1 priority, up from <50% in 2016. *(Source: Gartner*

Webinar Presentation, Ten Ways CDOs Fail and How to Avoid the Traps, Doug Laney, October 2016). With the threat of data disruption from competitors on the rise, companies need to learn how to use their data effectively.

For example, analyzing data can contribute to more efficient business operations. This might come from an analysis of customer comments left on social media in the context of actual sales activity and customer service logs. It may also come from comparing regional sales data to other demographic data for identifying new sales opportunities.

Another way to monetize data is to leverage your data into new products and service offerings, or even change business models. For example, you might be an equipment manufacturer who captures data from sensors built into your products for purposes of streamlining maintenance. However, deeper analysis of that data, or combining that data with other business data, may enable you to offer more profitable contracts that open untapped markets, such as contracts that sell the productive output of your equipment rather than selling the equipment itself.

Unified Governance solutions can empower your organization to be more agile and innovative in uncovering powerful business insights and converting this data into currency.

BE PROACTIVE IN THE FACE OF AN EVER-CHANGING REGULATORY ENVIRONMENT

As you collect more data, and as more users in your company share that data in collaborative business initiatives, challenges associated with regulatory compliance grows. At the same time, the regulatory environment is constantly changing, requiring rapid, enterprise-wide responses to new regulatory demands such as the European Union's General Data Protection Regulation (GDPR).

According to a recent PwC GDPR preparedness survey, GDPR compliance is a top data protection priority for 92% of US organizations in 2017. According to survey respondents, over three in four (77%) companies plan to allocate $1 million or more on GDPR readiness and compliance efforts—with 68% saying they will invest between $1 million and $10 million, with 9% expecting to spend over $10 million to address GDPR obligations.

To prepare for GDPR, companies should assess their current data. Unified Governance solutions provide a central way to find, manage and govern your data assets, whether on-premises or in the cloud, structured or unstructured. They enable you to share trusted, high quality data from across the enterprise in a reliable and controlled manner. Through cognitive computing and machine learning strategies, modern Unified Governance can enable you to proactively adapt to regulatory changes, reduce your risk exposure, and protect your brand value. It helps simplify readiness even as you expand your data collection and use of personal data in an increasingly regulated environment.

YOUR DATA EMPOWERED

Structured data is giving way to unstructured data. In fact, as much as 80% of today's business data is comprised of text and images—dark data that escapes analysis. Yet unstructured data presents a rich source of business intelligence.

Unified Governance is an approach to data governance that is geared towards discovery and transformation of all data, structured and unstructured, on-premises or in the cloud, such that it can be made easily available to anyone in the organization. This means data users across the enterprise have access to all data assets. This is true whether they are business users, data scientists and developers, compliance officers and auditors, data engineers, or records managers.

Unified Governance provides you with complete information management and governance solutions for analytical insights to create business value through data while helping you ensure compliance thus lowering cost and risk. It empowers all your information stakeholders with the ability to discover, understand, integrate, analyze and govern data, whether structured or unstructured, on cloud, on premises, and hybrid with a potential to accommodate future growth.

Discover how to jumpstart your governance projects by test driving IBM InfoSphere Information Governance Catalog. Visit http://bit.ly/IGC-Trial for a free trial. ■

IBM
www.ibm.com

Notice: Clients are responsible for ensuring their own compliance with various laws and regulations, including the European Union General Data Protection Regulation. Clients are solely responsible for obtaining advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulations that may affect the clients' business and any actions the clients may need to take to comply with such laws and regulations. The products, services, and other capabilities described herein are not suitable for all client situations and may have restricted availability. IBM does not provide legal, accounting or auditing advice or represent or warrant that its services or products will ensure that clients are in compliance with any law or regulation.

Pythian

How cloud answers the call for high-performance data integration

Lynda Partner, Vice President of Marketing

To keep pace with the speed of today's business, modern IT departments are being pushed to reduce the time to answer for consumers of their data, while integrating more data types from more sources. But traditional data warehouse technology is simply not up to new demands for complex integrations that involve exploding volumes of structured and unstructured data, and ever-expanding data sources.

Bringing together unparalleled agility, scalability and cost effectiveness, cloud platforms are proving themselves by providing automation features to speed the integration of large data sets. And, by providing a means for more users to access more data through self-service, cloud can play a key role in making data part of a company's DNA.

Here are just a few of the ways cloud is answering the call for high-performance data integration:

CLOUD IS BUILT FOR BIG DATA INTEGRATION

Cloud platforms can make ongoing management of your data much easier, particularly if you use PaaS (platform as a service) services such as DBaaS (database as a service). With unlimited processing power and low storage costs, cloud can bring streaming data and batch data together on a single platform, making it possible to integrate much larger data sets than ever. With API-driven automated processes, cloud makes previously onerous functions like creating data pipelines for data ingestion more agile. Adding to this agility are connectors that are built into the leading cloud platforms, making it easier to quickly connect to common data sources such as your CRM, financial, ERP and other systems.

CLOUD OFFERS A MODULAR ARCHITECTURE FOR FLEXIBILITY AND AGILITY

Cloud platforms don't tie you to particular software or hardware technologies, as with traditional data systems. Their modular architecture allows you to decouple functions like extract, transform and load (ETL) to give you more flexibility and agility. This means that as technology advances and your business evolves, you can take advantage of the best mix of cloud services, software and tools for your needs.

CLOUD SAVES ON COSTS

Because cloud vendors often provide a "pay as you go" model, you can move away from the capital expenditure of an on-prem system, to a model where you get capacity on-demand. Cloud platforms separate the charges for data storage and compute processing, so you can optimize costs across varying workloads. The flexibility of cloud allows you to use lower-cost open source technologies when and where you need them. For example, you can use open source tools like Apache Spark and NiFi for some or all of your ETL functions.

CLOUD ENABLES BETTER INSIGHTS AND DATA SCIENCE EXPLORATION

More and more, data integration needs to be driven by business requirements. Cloud platforms let you do just that, in a much more agile way. Unlike traditional enterprise data warehouses where the data model must be predefined, in the cloud, you can easily define new data models after data has been ingested. And with its ability to apply a schema on read, you can continually develop models to support new use cases and applications, including all of your historical data.

In the cloud, the same data that feeds your data warehouse can also be available for machine learning or data science exploration. While providing a cloud-based data warehouse to perform analysis through standard SQL queries, a cloud environment lets you maintain peripheral data for unforeseen future uses, or for data science exploration. Cloud-based sandboxes provide environments that support a variety of resources and APIs so data scientists can access compute, storage and all enterprise data to use with their preferred tools (e.g., Python, Scala, R and Java).

Cloud also enables the processing scale for real-time analytics, which is required for applications like IoT. Achieving this within a traditional data warehouse environment is much more difficult and costly.

GET CLOUD EXPERTISE ON YOUR SIDE

Pythian uses a proven methodology to integrate multi-source data on the cloud so you can be sure your system is future-proof and maintainable. Our approach helps you plan, implement, manage and grow your cloud data environment. Find out more about how our Cloud Services remove the effort and risk of integrating data on the cloud at www.pythian.com/cloud-services. ■

PYTHIAN
www.pythian.com

Data Virtualization—Any Data, Any Time, Any Place

As organizations are generating and collecting more and more data, they are also trying to find ways in which they can accelerate the time-to-value of this data. Many of these organizations are moving applications, data, and processing to the Cloud to take advantage of the agility, flexibility, and scalability that it provides. They can scale up and scale down their data storage and compute capacity as and when needed. This provides incredible flexibility to new or dynamic workloads and takes advantage of new data types or new business opportunities without making huge commitments on infrastructure.

Some examples of the drivers for organizations to move to the Cloud are:

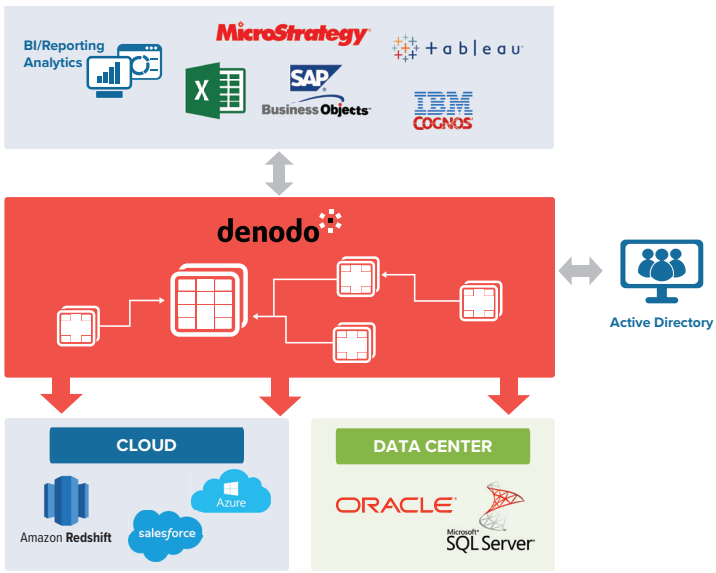
- **Modernizing applications**—migrating from monolithic application suites to a number of SaaS applications each with a specific focus. For example, moving from Oracle PeopleSoft suite to SaaS applications, such as Workday, Taleo, 401K.com, ADP, etc.
- **Moving analytics to the Cloud**—moving from on-premises data warehouses to Cloud-based ones, such as Redshift or Snowflake. Also, taking advantage of the dynamic and elastic analytical processing capabilities of Hadoop in the Cloud—Amazon Elastic MapReduce, etc.
- **Moving databases and custom applications to the Cloud**—to take advantage of the flexibility and agility to be able to scale up and scale down capacity as needed.

However, not all applications and data sources will migrate to the Cloud, at least for many years to come. It will be too disruptive to move some data to the Cloud as some, such as Mainframe data and applications, cannot move. Therefore, the applications and data sources that move to the Cloud are still a part of a larger ecosystem that encompasses the applications and data sources that remain in the data center (or on departmental servers and desktops). All of these applications and data sources need to be integrated to deliver the best insights to the business.

A DATA INTEGRATION PLATFORM THAT CAN STRADDLE THE DIVIDE

Because an organization's data and applications will straddle the Cloud, the Edge, and on-premises Data Center divide, you need a data integration platform that will similarly straddle the divide. That is, a data integration platform that will support a hybrid deployment of on-premises and in the Cloud. Additionally, this platform needs to be intelligent so that it can move the processing to where the data resides and not the other way around. It must move the processing to the data—the days of copying masses of data from one system to another for integration and processing are long gone. The volume of data involved in modern analytics is too large for this old style of thinking—and moving data in and out of the Cloud for integration and processing just exacerbates this problem. The data integration platform—in its hybrid configuration—needs the on-premises and in the Cloud parts for tandem processing of the data where it resides.

This is where Data Virtualization is in a class of its own. Obviously, a Data Virtualization Platform can be deployed on-premises—in the data center, on a departmental server, etc.—or at the edge, but the leading platforms can also be deployed in the Cloud to create a hybrid data integration platform which is intelligent enough to push the processing as close as possible to the data that is being integrated, and only move the minimum amount of data back to the consuming application or tool. This makes it quicker and easier to provide complete, integrated, fit-for-purpose data to the business users who need to get those actionable insights for better decision making. By providing access to data from the original source, integrating and transforming it in real time, the Data Virtualization platform provides complete and current data to the users—without having to set up complex data replication and synchronization pipelines to try to copy data into a central repository or data lake. Not having to copy data means that Data Virtualization can provision new data sets to users in a fraction of the time required by traditional data integration techniques—indeed, users have reported that what previously took two weeks with ETL took only half a day with Data Virtualization.



All relevant data from all data sources—internal, Cloud, or external to the organization—in real time and fit-for-purpose, and provisioned in a fraction of the time of traditional techniques? That's what the business users want and have been demanding for quite some time. ■

DENODO
www.denodo.com

How HVR Enables Continuous Data Integration for the Modern Enterprise

Most organizations rely on information from multiple systems to optimize business decisions. Until recently, they built operational data stores (ODS) and data warehouses to consolidate data for this analysis and populated these data stores using batch extract, transform and load (ETL) jobs that ran nightly, monthly or quarterly.

But organizations today face a number of changes that ETL solutions are ill equipped to handle. Organizations are now looking to perform real-time analytics to support faster, more dynamic and agile business decisions. As organizations increasingly adopt applications that live in the cloud, data consolidation must encompass these solutions as well. At the same time, growing deployment of internet of things (IoT) sensors is causing data volumes to explode while demanding that data be analyzed closer to real-time.

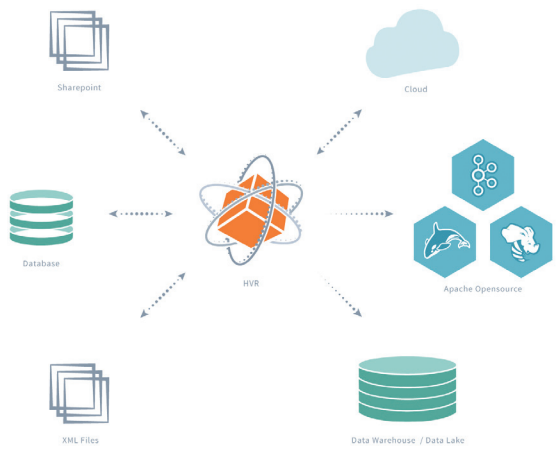
Traditional batch oriented ETL systems were never designed to address these use cases. Organizations today collect more data than ever, which ETL solutions must process during the batch window. Yet because these organizations provide services to customers across multiple geographic regions, there's no longer a lengthy natural batch window outside office hours when users are not impacted by batch loads. If changes are made during the extract, performance for end users can suffer and data available for analysis can become out of sync with operational data. The result is that decisions must be made based on the analysis of historical data alone.

Moreover, traditional ETL solutions were not designed specifically to integrate data from cloud-based systems or to take advantage of native functionality in streaming environments.

Organizations that wish to perform analytics in real-time need a continuous data integration strategy. Such a strategy must integrate data from a wide range of on-premises, cloud, and streaming data sources at the time each one is updated.

HOW HVR ENABLES CONTINUOUS DATA INTEGRATION

HVR was designed specifically to support organizations that require continuous data integration to enable real-time analytics with features that include log-based change data capture (CDC), fast data capture and transmission, heterogeneity, and scalability.



CDC—Our log-based CDC eliminates the need for a batch window. It captures changes as soon as the transaction is committed in the database's log, rather than making larger updates all at once. This reduces the risk that data in production databases and data warehouses will be out of sync and ensures that decisions can be made based on the analysis of the latest data.

Speed—the HVR solution delivers high-speed data capture and transmission locally and globally. Change data capture delivers excellent throughput by replicating only data changes to minimize impact on network traffic and performance. It has no effect on the performance of the source system. Advanced compression algorithms dramatically reduce traffic flowing over the network to minimize latency. HVR further improves performance through smart network utilization and use of native technology wherever possible.

Heterogeneity—HVR enables you to exchange data between heterogeneous systems in real-time. Our solution works with dozens of the most popular relational and analytical databases, Hadoop targets and file locations including Amazon S3. It also bridges the gap between on-premises and the cloud, as well as between different clouds. APIs are available to enable integration with additional systems.

Scalability—With HVR, your organization can easily add applications and platforms to your existing architecture. Our API enables you to bring on new platforms if we don't support a specific platform, and we continually add support for new platforms to the product. HVR also allows you to support ever increasing volumes of data. Since our CDC functionality processes incremental changes and compression functionality, we enable organizations to move data at tremendous speed. As your organization grows, you don't need to worry about data access slowing you down.

As organizations increasingly look to take advantage of IoT solutions as well as become more agile by making real-time business decisions, they need a solution that can continuously consolidate the requisite data. HVR picks up where traditional ETL systems leave off. It allows your organization to capture changes in real-time, with excellent performance, from your on-premises and cloud-based data sources, with the scalability to add more data sources and ever increasing volumes of data as your business grows. ■

HVR
www.hvr-software.com/