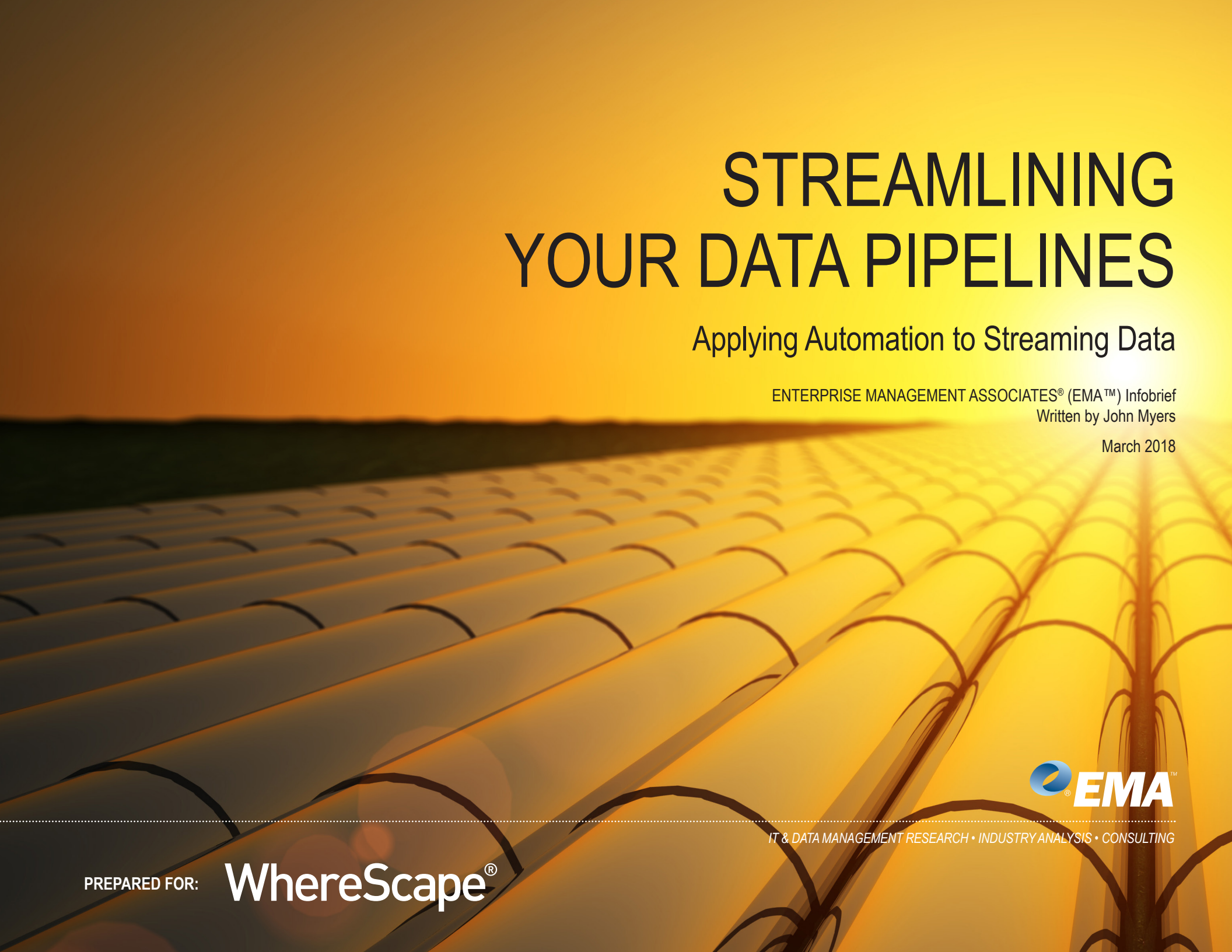


STREAMLINING YOUR DATA PIPELINES

Applying Automation to Streaming Data

ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) Infobrief
Written by John Myers

March 2018

EMA™

IT & DATA MANAGEMENT RESEARCH • INDUSTRY ANALYSIS • CONSULTING

PREPARED FOR: WhereScape®

TABLE OF CONTENTS

THE PROMISE AND CHALLENGE OF STREAMING DATA

Data-driven organizations push the envelope as they seek competitive advantage. They look for new, innovative ways to utilize the information they have to lower costs and improve operational efficiency and customer experience. This information comes from existing enterprise applications, such as enterprise resource planning, customer relationship management, and supply chain management, as well as new data sources such as social media platforms, like Facebook or Twitter, and big data sources, such as Hadoop and NoSQL.

Data-driven organizations are also focusing on the adoption and utilization of streaming event data sources coming from real-time applications associated with orders, fulfillment and payment, Internet of Things (IoT) device and sensor event data, and mobile app and online portal navigation information. These streaming event sources promise to accelerate the collection, analysis, and action taken to build competitive advantage. Organizations that can spot and act on trends in their operations, customer tastes, and overall market conditions in real time have a clear advantage over their competitors, who are taking days, if not weeks or months, to find those same opportunities.

Streaming Complexity Prevails

However, the data management environments that support streaming event data collection, transformation, orchestration, and storage are complex. They require specialized and cutting-edge skillsets that are scarce, and in high demand, to construct, configure, and maintain. These environments require the ability to scale data infrastructure teams to meet the challenges of variation in mobile app and IoT device configuration. They are far accelerated in terms of rate and pace of change than traditional, batch-oriented data management environments.

For technology teams and departments already weighed down with the workloads associated with traditional environments, making the leap from traditional data management practices to those associated with streaming data can present a significant challenge.

IN THE KNOW

WHO: Executives, architects, and business stakeholders from data-driven organizations.

WHEN: As streaming event data sources present opportunities for real-time data orchestration and analysis.

WHAT: How to apply automation practices and strategies to the implementation of streaming data pipelines.

Applying Automation to Streaming

To manage the new challenges of constructing and maintaining streaming data pipelines, automation strategies and practices provide organizations with the ability to capitalize on the opportunities of streaming data without taking on a challenge destined for failure. Key considerations for implementing automation in association with streaming include:

1. Growth in the adoption of streaming data strategies with IoT, enterprise applications, and mobile and online data
2. Complexity of streaming data management and engineering across multiple points
3. Value of automation for data infrastructure engineering initiatives to increase scale and maintainability
4. Strategic impact of automation on streaming environments
5. Value of business outcomes for automation practices

THE GROWTH OF STREAMING

The adoption of streaming strategies is not just for innovators. Streaming use cases and applications are now part of IT strategy for a large majority of organizations. This can take the shape of Internet of Things (IoT) sensor and device data, or it can be the ingestion of real-time event data from enterprise applications, such as ordering, fulfillment, and payment platforms. Streaming use cases can come from geolocation information from mobile apps, or clickstream information from online portals and webpages.

In two separate EMA end-user research studies,[1] respondents indicated that their organizations are adopting the strategies of streaming data at levels over 70 percent. In one study, nearly 9 of 10 respondents indicated they are adopting IoT strategies. In another, over 7 of 10 indicated that their organization is adopting streaming data strategies.

These adoption rates show that organizations are not just nibbling at the edges of streaming data strategies, but integrating streaming into their operations with sensors and devices, monitoring and managing their processes with real-time event data, and maximizing customer experience with mobile and online presence information.

STREAMING DATA MYTH #1

MISCONCEPTION: Streaming event data is only for companies on the bleeding edge of technology.

TRUTH: Organizations across the spectrum of company size, industry, and technology maturity are looking at streaming practices and strategies.

85% OF RESPONDENTS INDICATED THAT IOT WAS PART OF THEIR ORGANIZATIONAL STRATEGY

72% OF RESPONDENTS INDICATED THAT STREAMING DATA WAS PART OF THEIR ORGANIZATIONAL STRATEGY

[1] "The Rise of the Internet of Things" and "Charting the Expanding Horizons of Big Data"

THE COMPLEXITY OF STREAMING

Streaming data involves a level of complexity that is not inherent with traditional data management techniques. Over and above the nature of the data associated with streaming data applications is a constant flow of information 24 hours a day, 365 days a year. There are increased areas of data preparation time and turnaround, and changing requirements to support various workloads.

In terms of data preparation or data engineering, streaming data requires organizations to spend more time with their data. Since streaming data sources tend to have less rigorous requirements for data schemas and will often use the flexible JSON data format to support that flexibility, organizations need to spend more time assessing and accounting for adjustments to those formats for proper data management. Included in this process is also the fact that there are many more sources and points of change within a streaming environment than a traditional environment. The number and types of sensors in IoT implementations, the versions and operating systems associated with mobile apps, and the variants of app developer preferences and customer demands are all factors to consider.

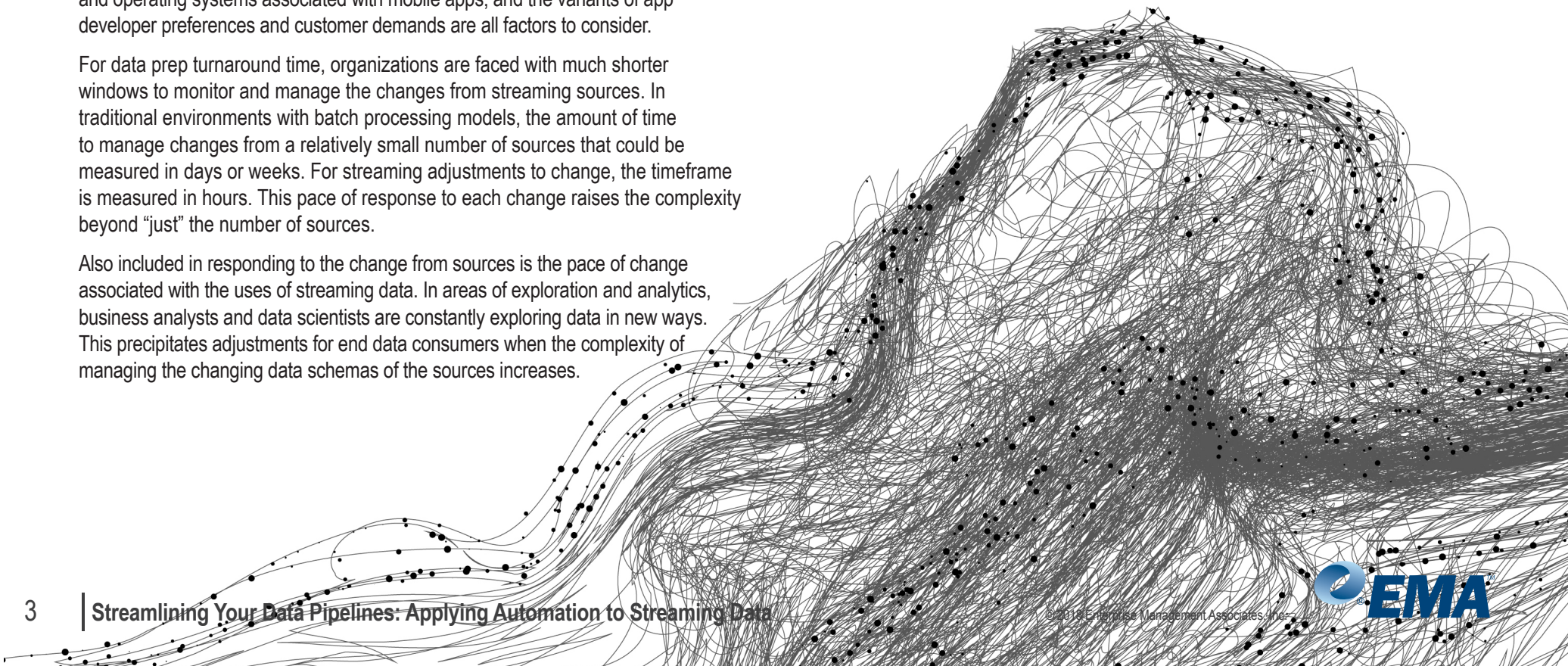
For data prep turnaround time, organizations are faced with much shorter windows to monitor and manage the changes from streaming sources. In traditional environments with batch processing models, the amount of time to manage changes from a relatively small number of sources that could be measured in days or weeks. For streaming adjustments to change, the timeframe is measured in hours. This pace of response to each change raises the complexity beyond "just" the number of sources.

Also included in responding to the change from sources is the pace of change associated with the uses of streaming data. In areas of exploration and analytics, business analysts and data scientists are constantly exploring data in new ways. This precipitates adjustments for end data consumers when the complexity of managing the changing data schemas of the sources increases.

BIG DATA CLOUD MYTH #2

MISCONCEPTION: Data management practices for streaming event data are the same as traditional batch-oriented practices.

TRUTH: The real-time nature of streaming event data and the practices and techniques to support it are significantly different than those associated with traditional environments, such as the data warehouse and the data lake.



THE VALUE OF AUTOMATION

Automation for analytical environments has proven to be a valuable avenue for organizations to manage change and increase the success of their data infrastructure initiatives. Organizations with strong automation strategies and practices have shown that they can support more change events, improve the timely and cost-effective implementation, and have a greater capability to maintain existing environments.

For organizations who experience large amounts of change within their data management and engineering environments, automation strategies and practices enable organizations to handle more change than those who are not adopting the same principles. In EMA end-user research, organizations with strong automation strategy maturity are 2.5x more likely to be able to support changes to their analytical environments at least weekly, and often more frequently, than those who do not adopt automation.

Business success is a fundamental outcome of any IT initiative. Data management is no different. With its core components in terms of standardization and repeatability, automation enhances the success of companies to achieve data infrastructure implementations. Respondents with strong automation adoption in EMA end-user research indicated that they achieved nearly 2x the success of organizations that were not adopting automation as part of their implementation strategy.

The organizations with strong automation strategies and adoption not only handle more change and provide more successful outcomes, but they are able to maintain those efforts in less time. EMA research shows that organizations with strong automation adoption reduce the amount of time spent on the ongoing maintenance of analytical environments much more than those without automation practices.

STREAMING DATA MYTH #3

MISCONCEPTION: Current IT teams and infrastructure using manual-based techniques can support streaming data pipelines.

TRUTH: Streaming data pipelines represent a significant change in terms of complexity. Manual-based techniques and processes cannot keep up with the scale of streaming environments.

2x ORGANIZATIONS ACHIEVED NEARLY TWICE THE SUCCESS OF ORGANIZATIONS USING AUTOMATION STRATEGIES THAN THOSE THAT WERE NOT ADOPTING AUTOMATION

THE IMPACT OF AUTOMATION ON STREAMING ENVIRONMENTS

Automation is enabling organizations to establish technical environments faster and more efficiently than when they used traditional, manual approaches. As a result, IT teams can focus on the strategic aspects of streaming data provisioning and management.

When streaming event sources are first introduced into an organization, data infrastructure teams can use automation to establish the data pipelines from the sources directly or from a streaming platform, such as Kafka. This allows efficient data pipelines to scale as more streaming event sources are brought online. Effective data pipelines will be particularly important in IoT device and sensor data scenarios, and those associated with mobile applications. The amount of variability in the initial data formats will require flexibility in the data pipeline configuration. Later, as the number of devices proliferates across the organization and multiple vendor configurations are implemented—allowing for variation in mobile apps by operating system, platform, customer, and scale—automation will provide the ability to match that scale by letting data infrastructure focus on classes or types of streaming event sources as opposed to individual ones.

As companies establish these data pipelines, automation allows data engineers to quickly build exploratory or discovery environments for the evaluation and validation of data. These temporary prototype environments are often referred to as sandboxes. As business stakeholders profile and explore the streaming event data, both at rest and as it flows from streaming data sources, engineering is refining and validating the data's operational and analytical applications. Automation enables IT to quickly ramp sandboxes up and down as teams in operations, marketing, customer care, and finance grow to better understand and validate streaming data sources.

Once the data requirements for a particular streaming application are established, automation allows for the implementation of the application to occur quickly and seamlessly. Automation can either "move" the data pipeline from a sandbox to a production destination, or create a parallel pipeline to another stream platform or a data management platform for storage and historical queries.

In each of these instances, automation takes on many of the tactical burdens for the data infrastructure teams. This allows individual team members to handle a wider scope in terms of technical implementation and the overall organization to have greater bandwidth regarding the technical landscape to be managed and maintained.

THE BUSINESS VALUE OF AUTOMATED DATA PIPELINES

The ability of data infrastructure teams to deliver consistent quality access to streaming data sources means business stakeholders can focus on the business impact of data-focused applications to create competitive advantage.

For organizations with IoT initiatives, this means focusing on productivity improvement. For connected fleet scenarios, this can include how to best understand the geographic routes of shipping containers and delivery schedules for orders and packages. For manufacturing organizations, information from the assembly line on product quality and process efficiencies is validated against observations. In both cases, when the data is evaluated and validated, automation allows organizations to quickly adjust to a production environment to operationalize the benefits of the data.

For enterprises using mobile apps and online websites to extend their organizational reach with customers and partners, the automated flow of events and clickstream enables teams in customer service and partner management to better understand customer experience components. Streaming data provides organizations insight in real time, or near-real time, about their customers and their experiences.

For companies examining streaming data from their real-time applications for ordering, fulfillment, and payment, streaming events provide key understanding for operational efficiency within their existing practices and procedures. This allows organizations to "turbocharge" their improvement programs such as lean manufacturing or lean distribution and lower their operational costs to improve margins.

STREAMING DATA MYTH #4

MISCONCEPTION: You can retrieve past streaming event data at a later date if you need it.

TRUTH: Streaming data has an element of "freshness" or availability to it that traditional transactional event and historical analysis data does not. Organizations should decide whether or not to leverage streaming data as soon as possible, including how they will use it and/or if they should store it in a data lake for future use. Otherwise, they will risk data loss.

ABOUT OUR SPONSOR

WhereScape®

WhereScape helps IT organizations of all sizes leverage automation to design, develop, deploy, and operate data infrastructure faster. More than 700 customers worldwide rely on WhereScape automation to eliminate hand-coding and other repetitive, time-intensive aspects of data infrastructure projects to deliver data warehouses, vaults, lakes, and marts in days or weeks rather than in months or years. WhereScape has global operations in the USA, UK, Singapore, and New Zealand.

WhereScape® Automation with Streaming

IT organizations no longer have months or years to deliver the analytics and insights business leaders expect. Conventional approaches to data infrastructure development are too time-consuming and expensive to support today's frequent business requests, the increasing complexity and volume of data, and the scarcity of IT resources. It's time for a new approach—one that applies agile and DevOps principles to data infrastructure delivery. WhereScape automation with Streaming automates the design, development, deployment, and operation of data infrastructure on leading data platforms. Whether a data infrastructure is on-premises, in the cloud, or a hybrid of both, WhereScape automation can help companies get to production faster, with less cost and risk. With WhereScape, data architects can develop and iterate collaboratively with users to deliver projects faster and respond more quickly to changing business requirements.

With out-of-the-box support for popular data platforms, WhereScape automation eliminates hand-coding, streamlines deployment, and auto-generates documentation to boost developer productivity fivefold. WhereScape automation with Streaming can help you:

- Get to production fast by automating up to 95 percent of your coding effort.
- Deliver impactful projects to the business in days or weeks with less cost and risk.
- Collaborate with users to rapidly create working solutions that deliver value immediately.
- Tame hybrid data environment complexities and have the flexibility to easily leverage both batch-based and real-time streaming sources of data.
- Take advantage of the native strengths of cloud and on-premises data platforms with optimized code.
- Adapt to change quickly, with automatic creation and management of metadata and documentation.
- Automate virtually any data project, including data warehouses, data lakes, data vaults, data marts, or big data integration.

To learn more about WhereScape automation with Streaming, visit www.wherescape.com.

About Enterprise Management Associates, Inc.

Founded in 1996, Enterprise Management Associates (EMA) is a leading industry analyst firm that provides deep insight across the full spectrum of IT and data management technologies. EMA analysts leverage a unique combination of practical experience, insight into industry best practices, and in-depth knowledge of current and planned vendor solutions to help EMA's clients achieve their goals. Learn more about EMA research, analysis, and consulting services for enterprise line of business users, IT professionals, and IT vendors at www.enterprisemanagement.com or blogs.enterprisemanagement.com. You can also follow EMA on Twitter, Facebook, or LinkedIn.

Corporate Headquarters:
1995 North 57th Court, Suite 120
Boulder, CO 80301
Phone: +1 303.543.9500
Fax: +1 303.543.7687
www.enterprisemanagement.com

3691.032318