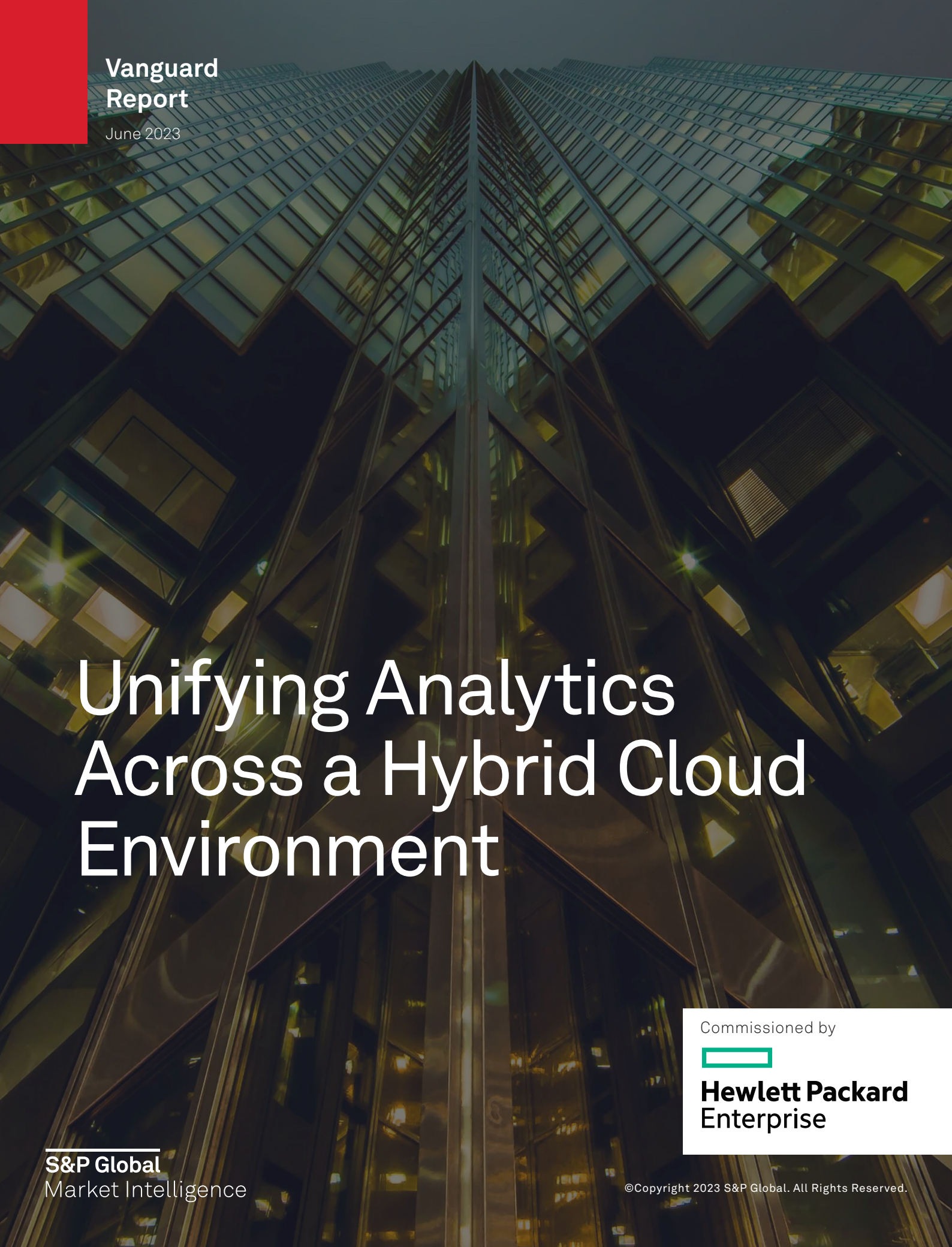


Vanguard Report

June 2023

Unifying Analytics Across a Hybrid Cloud Environment

Commissioned by

Hewlett Packard Enterprise

S&P Global Market Intelligence

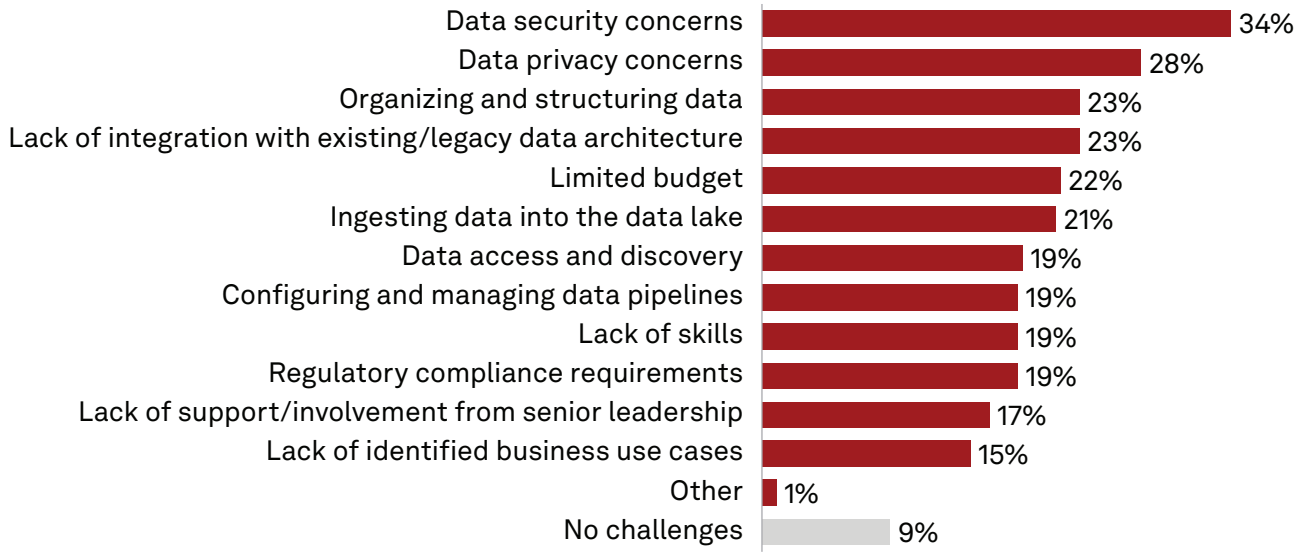
Introduction

Today's analytics platforms come with a good deal of promise, but the hard truth is that enterprises continue to face challenges and struggle to achieve the full potential of these platforms. There are a host of reasons why this might be the case. Implementing a standardized analytics toolset across a common platform is one challenge. Another is that different personas, including data engineers, data scientists and business analysts, all need access to the data.

Rising cloud costs, vendor lock-in, unpredictable performance, security vulnerabilities, limited services and a lack of support are a few more of the challenges faced by organizations that have struggled with implementing and managing their analytics environments. Working in a distributed, hybrid cloud environment where data and analytic processing tools reside in different locations can also be difficult.

Data from 451 Research's Voice of the Enterprise: Data Platforms, Data & Analytics 2022 reveals that 68% of respondent organizations' data platforms or services operate in a hybrid cloud environment, and another 18% have plans to move to a hybrid cloud environment. That means that roughly 86% of organizations in our study are either already deploying hybrid cloud or have plans to deploy it in the near future. Recent 451 Research data also identifies some of the top challenges organizations face when managing data, AI/machine learning (ML) and analytics in a distributed environment.

Figure 1. Challenges generating insight in a distributed data processing (data lake) environment



Q. What are the most significant challenges your organization faces in generating insight from its data lake environment? Please select all that apply.
Base: Respondents whose organization currently has a data lake in use, in a proof-of-concept or pilot stage, or plan to use within the next 12 months (n=290).
Source: 451 Research's Voice of the Enterprise: Data & Analytics, Data Platforms 2022.

And while organizations grapple with implementing analytics across a hybrid cloud environment, with a variety of personas needing to access that data, there is strong organizational pressure to use AI/ML and analytics to leverage data for strategic business decisions. As a result, organizations find themselves in the perfect storm: managing an AI/ML and analytics platform in a hybrid cloud environment while simultaneously trying to drive strategic decisions.

Despite the challenges and complexities of working in a distributed, hybrid cloud environment, a unified analytics platform provides a path forward for enterprises. A unified analytics platform consists of a standardized set of enterprise-ready tools that work with a broad set of data types and formats across a variety of locations for a variety of data users.

The Take

While some cloud-based analytics environments require customized deployments or do-it-yourself approaches to assemble open-source tools, a unified analytics platform takes a different approach. It not only provides a simplified platform for AI/ML and analytics at scale, but it also delivers a SaaS-like experience that runs in a hybrid cloud environment. It also addresses the challenges of working with and supporting open-source software for data science and analytics workloads.

One specific challenge enterprises face is the lack of standardized tools, especially open-source tools favored by data scientists and data engineers. Many offerings that are marketed as unified analytics platforms consist of a disparate collection of open-source tools, but a true unified analytics platform incorporates best-of-breed open-source tools that have been integrated to ensure compatibility and interoperability between them. This eliminates the step of integrating open-source software, which can be burdensome for organizations.

Supporting and maintaining those open-source tools is another challenge. While open-source software can be less expensive, help avoid vendor lock-in and offer ongoing innovation, the downside is that the enterprise must provide support for each open-source software package. This can be costly to maintain because many open-source software tools are not enterprise-ready or lack the reliability of proprietary software. In a unified analytics platform, support and maintenance for these open-source software tools are baked into the platform.

Even though hybrid cloud environments are becoming the dominant deployment pattern for AI/ML and analytics workloads, there is no guarantee that open-source software will function correctly in that environment. Enterprises using a do-it-yourself analytics platform would need to “lift and shift” as well as re-architect open-source code to run it in the cloud. However, to drive environment-agnosticism, they would need to use cloud-native technologies. Organizations adopting a unified analytics platform benefit from open-source software that integrates seamlessly, comes with simplified management tooling and is expected to run smoothly across a hybrid cloud environment.

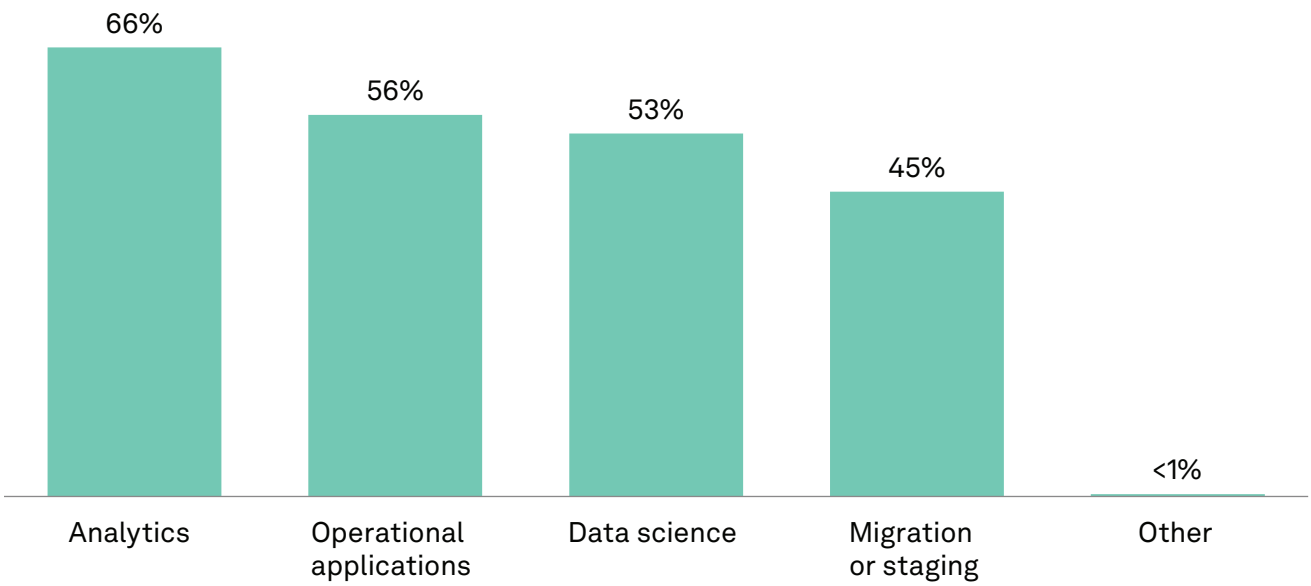
Key attributes of a unified analytics environment

While a unified analytics platform is not defined by its technologies, it ***is*** defined by how the technologies are integrated and function together, as well as by the types of users and use cases that it enables. For instance, user personas may include business-oriented users, such as executives and business operations leaders, as well as business analysts; operations-oriented users may consist of individuals concerned with the management, maintenance and access of the environment, such as data engineers, data analysts and data scientists.

Mixed workloads

The wide spectrum of users provides a platform for running multiple analytics workloads (see Figure 2). Additionally, 66% of respondents to the Data Platforms survey say their data analytics environments support mixed workloads: 28% are running analytics, data science and operational workloads, 21% are running both analytics and data science workloads, 12% are running both analytics and operational workloads, and 5% are running data science and operational workloads.

Figure 2. Use cases in a distributed data processing (data lake) environment



Q. What application use cases run (or will run) on your primary data lake environment? Please select all that apply.
Base: Respondents whose organization currently has a data lake in use, in a proof-of-concept or pilot stage, or plan to use within the next 12 months (n=294).
Source: 451 Research's Voice of the Enterprise: Data & Analytics, Data Platforms 2022.

Global data access

Underlying the unified analytics platform is typically a data fabric that provides global access to data regardless of where it resides. Without data access, data scientists, data engineers and other analysts would be severely hindered in their efforts.

An enabling driver for data science and analytic workloads is agility, a core tenet of a unified analytics platform. Organizations need to be able to adapt to market changes quickly, especially when game-changing events happen without warning. Unified analytics platforms offer value through their ability to access different data types from a variety of data sources, and store that data in a hybrid cloud environment to ensure the consistent delivery of high-integrity data to data "consumers" and other data-dependent individuals within the enterprise. This unfettered access to the data — enabled by a data fabric — expands the use cases and analytical workloads that can be carried out on the platform.

Supports the full spectrum of AI and analytics workloads

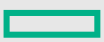
Because a unified analytics platform provides a robust suite of data science and analytics tools, it means data scientists, data engineers, data analysts and other data users can address the full life cycle of analytics from building data pipelines to developing and deploying AI, machine learning models and analytics workloads. Hybrid cloud infrastructure and technologies also serve as a foundational element of a unified analytics platform that enables organizations to carry out data science and analytics workloads in any environment.

Open and extensible

Because a unified analytics platform is based on an open architecture that integrates open-source software, it also provides a flexible environment so organizations can add open-source tools, custom applications or third-party ISV tools for specific users and use cases as needed. For instance, data scientists may use tools such as Python, Java, R, C/C++ and SQL, while business users, who are generally non-technical, may feel more comfortable working in visually oriented analytical environments. The open architecture approach, which is core to a unified analytics platform, not only reduces unpredictable costs, but it also helps organizations avoid vendor lock-in based on proprietary software.

Conclusion

Many organizations are frustrated with the ongoing challenges of successfully deploying data science and analytics platforms, which can be highly complex to deploy, require significant resources to maintain and be costly to manage. A unified analytics platform can help address many of these challenges by increasing collaboration and productivity, removing the dependence on proprietary solutions and lowering costs, which enables enterprises to focus on developing and deploying AI and analytics workloads.



Hewlett Packard
Enterprise

HPE Ezmeral Unified Analytics Software helps customers develop and deploy data and analytic workloads faster by providing fully managed, secure, enterprise-supported versions of the most popular open-source analytics tools spanning the complete lifecycle of AI and analytics workloads in a consistent, consumption-based SaaS experience that provides predictable, transparent economics and precise controls to optimize costs. Hybrid by design, it can be run at business edges, on-premises, or in public cloud environments, allowing you to choose where to develop and deploy analytics workloads that best fit your business requirements.

About the author

James Curtis

Senior Research Analyst, Data, AI & Analytics

James Curtis is a Senior Research Analyst for the Data, AI & Analytics Channel at S&P Global Market Intelligence. He has had experience covering the BI reporting and analytics sector and currently covers Hadoop, NoSQL and related analytic and operational database technologies.

James has over 20 years' experience in the IT and technology industry, serving in a number of senior roles in marketing and communications, touching a broad range of technologies. At iQor, he served as a VP for an upstart analytics group, overseeing marketing for custom, advanced analytic solutions. He also worked at Netezza and later at IBM, where he was a senior product marketing manager with responsibility for Hadoop and big data products. In addition, James has worked at Hewlett-Packard managing global programs and as a case editor at Harvard Business School.

James holds a bachelor's degree in English from Utah State University, a master's degree in writing from Northeastern University in Boston, and an MBA from Texas A&M University.

About S&P Global Market Intelligence

S&P Global Market Intelligence's Technology, Media and Telecommunications (TMT) Research provides essential insight into the pace and extent of digital transformation across the global TMT landscape. Through the 451 Research and Kagan products, TMT Research offers differentiated insight and data on adoption, innovation and disruption across the telecom, media and technology markets, backed by a global team of industry experts, and delivered via a range of syndicated research, consulting and go-to-market services, and live events.

CONTACTS

Americas: +1 800 447 2273
Japan: +81 3 6262 1887
Asia Pacific: +60 4 291 3600
Europe, Middle East, Africa: +44 (0) 134 432 8300

www.spglobal.com/marketintelligence
www.spglobal.com/en/enterprise/about/contact-us.html