Publication date:
June 2020
Authors:
Bradley Shimmin, Chief Analyst
Cassandra Mooshian, Senior Analyst

Meeting the New Intelligent Data Economy Head-on

Adopting Cloud-native, Highly Scalable, and AI-driven Data Storage Practices On-premise

In partnership with:

Cloudera and IBM

Brought to you by Informa Tech

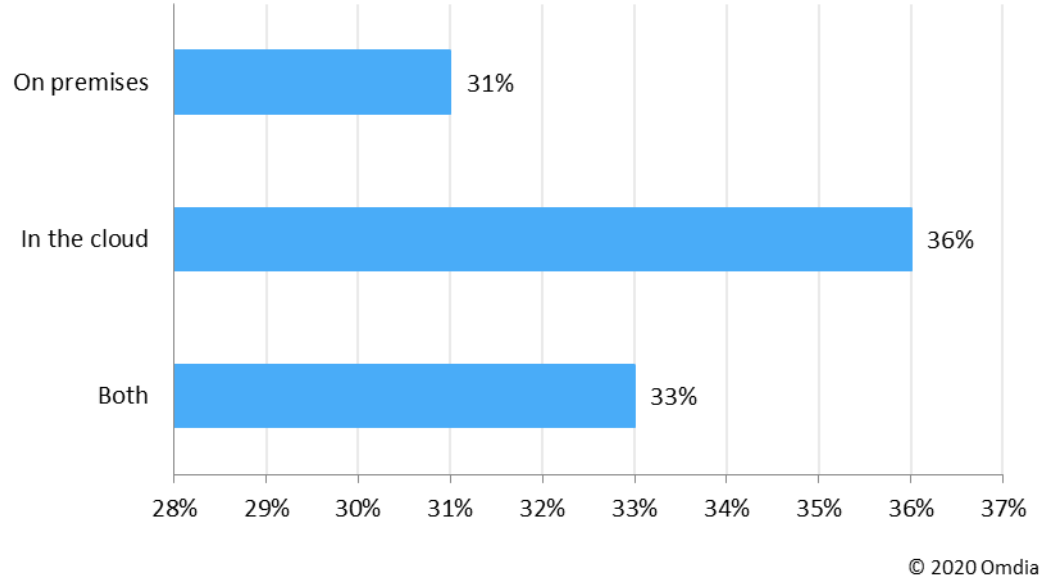
Overview

The data and AI economy is changing as the market evolves and enterprises become more eager for AI, an eagerness that is growing in response to the global COVID-19 pandemic. The current slowdown of IT investment due to the pandemic will continue past 2020, with large projects being significantly delayed. Funding will shift to projects that save money because they are more operationally efficient and resilient. Regardless, innovation will continue at a rapid pace, with buyers seeking new, data-driven business opportunities with clearly defined returns on investment.

In many ways, the pandemic has accelerated the digital transformation, with many enterprises forced to go digital in ways that before had been desired but tabled for several years. Many of those companies have been adapting to an entirely remote workforce and customer base. These solutions to new business problems are generating even more data than before 2020, when investment by enterprises in data architectures such as data lakes and warehouses was already growing exponentially.

This underscores the importance of establishing a well thought out and lasting strategy for data and AI, particularly for businesses seeking to use data as a strategic asset. One way an enterprise ensures a strong, enterprise-wide data architecture is to establish an AI center of excellence (CoE). With a CoE, an important part of the discussion is where to invest and how to prioritize investment. Fortunately, for many budget-strapped organizations, having a strong AI strategy does not necessarily require an IT overhaul. Instead, it may just entail optimizing what is already in place to facilitate new cloud-based data and AI solutions. After all, hybrid implementations have become a reality for most businesses.

Figure 1: Running AI on-premises or in the cloud



Source: Omdia. Results are from an April 2020 survey by InformationWeek and ITPro Today of 292 IT professionals using or planning to use AI technologies.

As part of a recent survey conducted by InformationWeek and ITPro Today, users were asked where they run their AI workloads (see Figure 1). Responses were almost evenly split between on premises, in the cloud, and both, which indicates that no single, all-in deployment paradigm dominates. Investment in on-premises solutions is still strong. In many ways, they serve as an anchor for the cloud.

Certainly, many companies have been unable to move to the cloud due to regulatory or security concerns. Others have not done it to better control costs or preserve autonomy. Regardless, much of the momentum behind AI is due to the cloud, and cloud adoption continues at an unprecedented pace, but the public cloud is not the platform. The platform consists of cloud-native technologies built increasingly on Red Hat OpenShift and Kubernetes. That is the "real" platform driving cloud spend, whether hosted on-premises or on a vendor's cloud. That layer of abstraction has enabled companies to invest in the cloud and premises without a great deal of compromise.

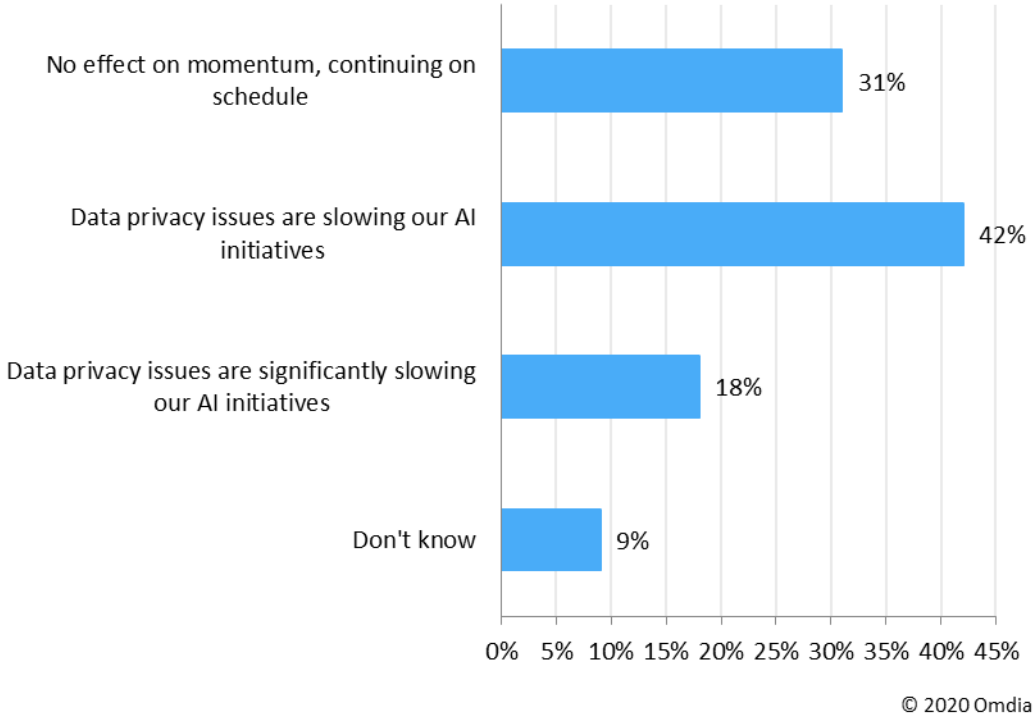
The question of survival lies in resilience and agility

A big area of investment is in efforts to make IT and business operations more resilient. In the currently environment, IT professionals must strive to balance innovation, optimization, and resiliency amid extreme upheaval. More importantly, companies must figure out how to survive a pandemic or weather a trade war and shifting regulations. Those efforts require them to consider many facets of operations, but arguably, none is more important than agility. Agility engenders resiliency.

Applying this notion to AI, enterprise data architectures must combine resiliency and agility. IT practitioners must understand the data flowing in and out of the business and

between departments while ensuring its security and privacy. Omdia recently surveyed over 350 enterprise IT practitioners and asked whether data privacy concerns were slowing innovation (see Figure 2).

Figure 2: The impact of data privacy issues on AI



Source: Omdia. Online survey of 365 enterprises conducted in January and February 2020. Surveyed companies ranged in size from small startups >100 employees to global companies with 10,000+ employees.

The results underscore that no matter how innovation-minded an organization is, the data privacy concerns of IT decision makers significantly, and negatively, affect the uptake of advanced analytics and AI workloads.

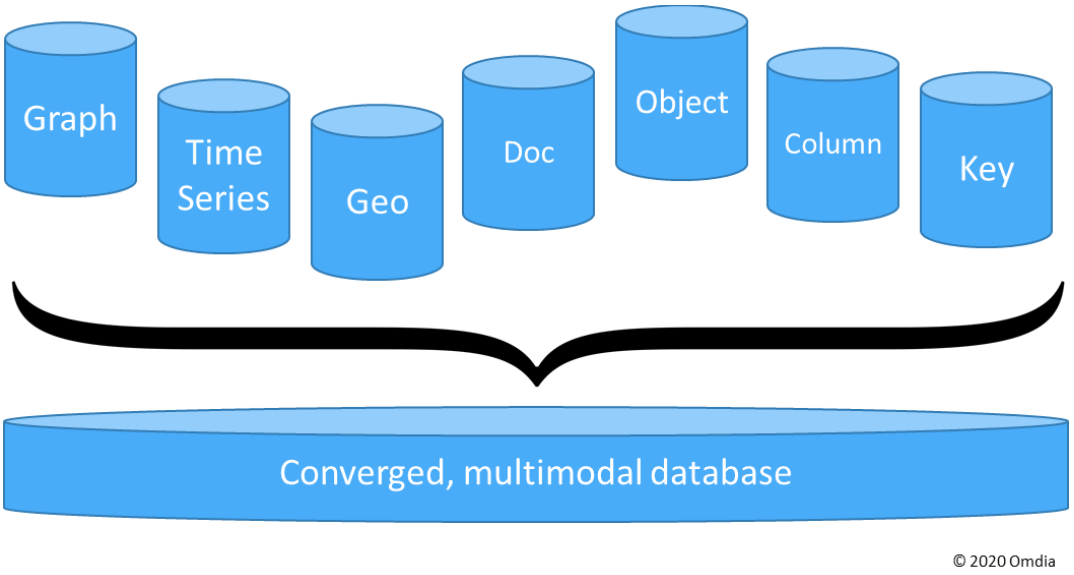
Building an agile data architecture

Despite the very real concerns regarding data privacy, Omdia ICT Enterprise Insights 2019/20 found that 29.8% of enterprises plan to invest in data platform architectures, eclipsing all other areas of investment, including data quality, privacy, and analysis.

There are nearly twenty different types of databases to choose from, such as graph, object, column, key, etc. Each delivers a specific and highly valuable functionality. The abundance of choices has led to a noticeable shift of the database market toward a more flexible data architecture, one that can handle transactions and analytics and that can behave like a time series, graph, key-value, or document data store (See figure 3). There are not many multi-modal or converged databases on the market, and there are many limitations, caveats, and compromises when adopting one. Regardless, they can have a

significant impact on organizations looking to undo data fragmentation and break down data silos for organizations able to work within those confines.

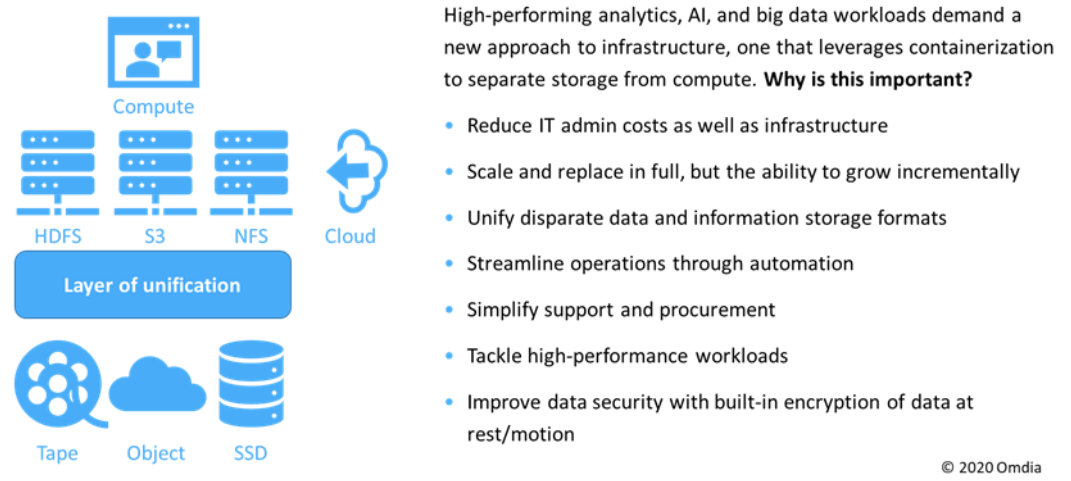
Figure 3: Specialized vs. converged databases



Source: Omdia

One of the most impactful inventions of the last thirty years was a specific technology, but rather the idea of virtualization. The fundamental theorem of software engineering states that "we can solve any problem by introducing an extra level of indirection." Indirection means abstraction. For many high performing analytics, AI, and big data workloads, abstraction is important. Leveraging abstraction to separate storage from compute can solve several important problems, including unifying data from disparate origins and formats, reducing infrastructure costs, and streamlining operations (see Figure 4).

Figure 4: Unified data infrastructures



Source: Omdia

Cloud-native AI development practices are flexible, with code typically written once and deployed anywhere. This is critical to enterprise IT agility and resiliency. The portability inherent in containerization not only allows practitioners to move their applications closer to the data, rather than the inverse, but it also provides an additional layer of resiliency for enterprise IT and enables truly hybrid environments. Additionally, utilizing cloud-native development and portability within a unified data architecture addresses the growing need to extend data access controls across cloud and on-premises environments.

Predicting the future of data architecture and management

Countering AI bias

Even with a cloud-native approach to data, analytics, and AI architectures, companies will find themselves thrust into new challenges that start small but could undermine agility and resiliency. For example, they must work against human and machine bias and provide transparency and accountability across all production AI workloads while securing machine learning (ML) workloads, thus protecting against theft and hostile modifications. Solving this will require coordination between IT and the business and a heavy investment in AI that provides more than the basic business outcomes.

Companies will need to become proficient at using AI to combat bias that may be hiding data, the algorithm, or even the business hypothesis. IT practitioners will need to make their ML models transparent, explainable, and understandable such that the company can explain how and why that model made a given decision. Fortunately, companies can use ML to do this, exploring what is happening within a deep learning (DL) black box by looking at the features and target parameters. There is also AI-based intelligent automation that can help relieve the task burdens of IT staffs and automate IT (and

business) processes to make organizations more resilient and agile. More tools are using AI to find the right data and ensure that requested data is properly used and secured.

Tearing down data silos

Setting up and supporting a unified and intelligent data architecture can overcome many of the challenges that practitioners face with on-premises data, analytics, and AI environments. Breaking down data silos is critical to a strong data architecture and allows enterprises to get the most out of their data. Databases must handle a multitude of different workloads, whether or not the data is structured, unstructured, graph, time series, etc. This underscores the need for multimodal databases, particularly if they are infused with AI. Having AI within a database optimizes and automates many database administrative tasks and makes it easier to connect to various data sources across various environments (vendor hosted, third-party hosted, on-premises).

Operationalizing data, analytics, and AI

Centralizing master data management is no longer the best way to understand and control data assets company-wide, especially since data assets important to business operations can often be outside the company. An emerging solution for this problem is lightweight data catalogs, which can see not just across departments, but also beyond the corporate firewall, leaving the data management work to the data platform. Data cataloging creates a silo-free data fabric that provides a catalog of consistent data services across different data warehouses and different private and public clouds. Additionally, it helps practitioners to better anticipate potential issues, such as compliance with regional data privacy laws.

DataOps is gaining traction, which is the application of DevOps ideals to the problems of data management. It converts data from an amorphous mix of data warehouses and lakes into a self-service collection of actionable data assets. This is highlighted in tools such as self-service data hubs and programmatic, end-to-end data pipelines. It means treating data as a tactical resource that takes shape only as needed and in an agile manner. This is also seen in serverless computing and functional programming, where the applications built do not need to know about supporting or surrounding infrastructures.

All roads lead to DataOps

Enterprise data practitioners should re-think integration patterns using DataOps, or agile, end-to-end DevOps-style data pipelines. It means creating serverless, transactional data flows and patterns that can be built, deployed, even shared just as they would with any cloud-native, containerized resource. Further, those pipelines could then be managed as resources within a data catalog. These changes help reduce the amount of energy required to build and maintain what were brittle, script-based integration flows. More importantly, they help an enterprise to more rapidly and safely accommodate new data sources and data use cases, making it more agile.

Table 1: Buyer recommendations

Theme	Recommendation
Defeating data silos	Re-think brittle integration patterns using agile, end-to-end data pipelines.
Putting AI to work for IT	AI can relieve IT staffs of task burdens, elevating the work they are able to do.
Opening and securing data as a commodity	Pursue data security and privacy because they will be differentiators.
Optimizing, not abandoning on-premises	Recent and ongoing investments in on-premises solutions are not obsolete despite cloud advancements, but they can be optimizing by bringing both together.

Source: Omdia

Data needs to be treated not just as a resource but also as a commodity, as a means of driving business value internally and externally. For this reason, data practitioners and IT professionals should prioritize data security and privacy as a corporate center of excellence. If data is indeed a commodity, then it must be something they can demonstrate and validate externally, whether they are sharing data with the supply chain or participating in a public data exchange. Their data is the company's differentiator and livelihood.

Architecting and modernizing infrastructure in a way that enables software abstraction and supports more advanced workloads is core to successfully scaling AI and analytics workloads to the enterprise level. Enterprise AI workloads cannot be run without the supporting infrastructure and proper information architecture. This is where the topic of optimization versus innovation really comes into play. Optimizing on-premises infrastructure versus overhauling or abandoning it in favor of cloud-native options can be practical, cost-effective, and timely. Regardless, if the current infrastructure is not capable of supporting dynamic AI workloads, then modernizing and rearchitecting the infrastructure may be the best option.

Data and trust in data are important because they are the engine that facilitates innovation and the capacity to withstand unanticipated change. Data residing on-premises demands an underlying storage infrastructure that accommodates a rich array of data, analytics, cloud-native, and AI workloads securely and reliably at scale. Companies that embrace these ideals will be well positioned to weather current and future market disruptions by placing data front and center as a means of gaining both operational efficiency and technical resiliency. Doing so will also create a solid foundation for future data-driven business opportunities, particularly those fueled by AI.

Appendix

Methodology

This paper is based on ongoing Omdia research, analysis, and knowledge of the enterprise AI market. Additionally, select data from recent surveys completed by Omdia, ITPro Today, Interop, and InformationWeek were used to elevate the discussion further. Regarding Figure 1, InformationWeek, ITPro Today, and Interop surveyed 292 IT and business professionals in April 2020 regarding their current or planned AI deployments as part of the *2020 State of AI* report. Regarding Figure 2, Omdia surveyed 365 enterprises regarding the impact that data privacy concerns have on their adoption of AI in January and February 2020 as part of the *AI Market Maturity* study and report.

Authors

Bradley Shimmin
Chief Analyst, AI and Intelligent Automation
askananalyst@omdia.com

Cassandra Mooshian
Senior Analyst, AI and Intelligent Automation
askananalyst@omdia.com

Get in touch

www.omdia.com
askananalyst@omdia.com

Omdia consulting

Omdia is a market-leading data, research, and consulting business focused on helping digital service providers, technology companies, and enterprise decision-makers thrive in the connected digital economy. Through our global base of analysts, we offer expert analysis and strategic insight across the IT, telecoms, and media industries.

We create business advantage for our customers by providing actionable insight to support business planning, product development, and go-to-market initiatives.

Our unique combination of authoritative data, market analysis, and vertical industry expertise is designed to empower decision-making, helping our clients profit from new technologies and capitalize on evolving business models.

Omdia is part of Informa Tech, a B2B information services business serving the technology, media, and telecoms sector. The Informa group is listed on the London Stock Exchange.

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Omdia's consulting team may be able to help your company identify future trends and opportunities.

About Cloudera

Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises. Learn more at Cloudera.com.

About IBM

For more information about IBM visit, www.ibm.com. Statements regarding IBM's future direction and intent are subject to change or withdrawal without notice and represent goals and objectives only.

Copyright notice and disclaimer

The Omdia research, data and information referenced herein (the “Omdia Materials") are the copyrighted property of Informa Tech and its subsidiaries or affiliates (together “Informa Tech”) and represent data, research, opinions or viewpoints published by Informa Tech, and are not representations of fact.

The Omdia Materials reflect information and opinions from the original publication date and not from the date of this document. The information and opinions expressed in the Omdia Materials are subject to change without notice and Informa Tech does not have any duty or responsibility to update the Omdia Materials or this publication as a result.

Omdia Materials are delivered on an “as-is” and “as-available” basis. No representation or warranty, express or implied, is made as to the fairness, accuracy, completeness or correctness of the information, opinions and conclusions contained in Omdia Materials.

To the maximum extent permitted by law, Informa Tech and its affiliates, officers, directors, employees and agents, disclaim any liability (including, without limitation, any liability arising from fault or negligence) as to the accuracy or completeness or use of the Omdia Materials. Informa Tech will not, under any circumstance whatsoever, be liable for any trading, investment, commercial or other decisions based on or made in reliance of the Omdia Materials.