E-book:

Building a Modern Data Estate on AWS

The complete guide to becoming a data-driven organization

ONICA by rackspace technology | aws

Introduction

Today, organizations are struggling with the performance and capability disparities that exist between legacy technologies and a modernized cloud. Companies know they must modernize. But they often choose to make incremental changes that can take months or years to implement. By contrast, businesses that want to become data-driven enterprises must evolve quickly. Rapid modernization in the cloud is the most direct path to reaping the benefits of the analytics and machine learning technologies that can position an organization as a market leader. Companies that unlock and activate their vast stores of data are likely to see significant top-line benefits, including higher customer acquisition and retention rates and increased profits.

If you want to completely understand your business environment, including your customers and value chain, you need three things: a modern data architecture, a deep understanding of your supply chain data, and the ability to continuously enhance and optimize your data. These three items form the foundation for intelligent cloud native applications that can utilize all of your corporate data, whenever and wherever it's needed.

However, the creation of a modern data strategy isn't easy. You'll need access to new skills, an understanding of new methodologies, and a willingness to curate, manage and integrate agile processes into your data and security solutions. You'll need to rethink how you handle data across your organization. Data is an important business asset, and thinking of data as such requires managing the data lifecycle and domains across the organization to produce better and more trusted data.

Despite the urgent call for modernization, few companies find immediate success when trying to move their data analytics projects into production. Instead, they often create additional complexity that can hamper attainment of business goals.

Common challenges to modernization include:

- **Ever-increasing data volumes** generated by the Internet of Things (IoT), customer and value chain interactions, and the need to store and process large amounts of unstructured data
- **Emerging technologies**, including serverless data platforms, data lakes, data mesh, data fabric architectures, DataOps processes and tools, blockchain and advances in AI/ML technologies
- **Data discovery challenges** stemming from unknown data sources, poor data quality, data silos and compliance restrictions
- **Costs**, including infrastructure costs, a lack of utility models, limited and expansive talent and large investments with no guaranteed return
- **Skills gaps** related to data analytics and a lack of specialized training and experience
- **Siloed data on outdated infrastructure** that cannot support shareability across teams or ensure data integrity.
- **A lack of governance** amid expanding data footprints that increase risk and complexity and require federated governance
- **Operating models** with poorly mapped ownership of roles and responsibilities for data domains that make it difficult to establish a Center of Excellence for data

What you'll learn

In this e-book, we examine the barriers and opportunities facing modern businesses and provide insight on how to overcome these challenges by deploying and managing a modern, cloud-based data platform on Amazon Web Services (AWS). We also explore what it means to be a data-led organization, identify the core design principles of a modern data estate, and outline what it takes to be successful in today's data-driven world.

Part 1: What does it mean to be a data-led organization?

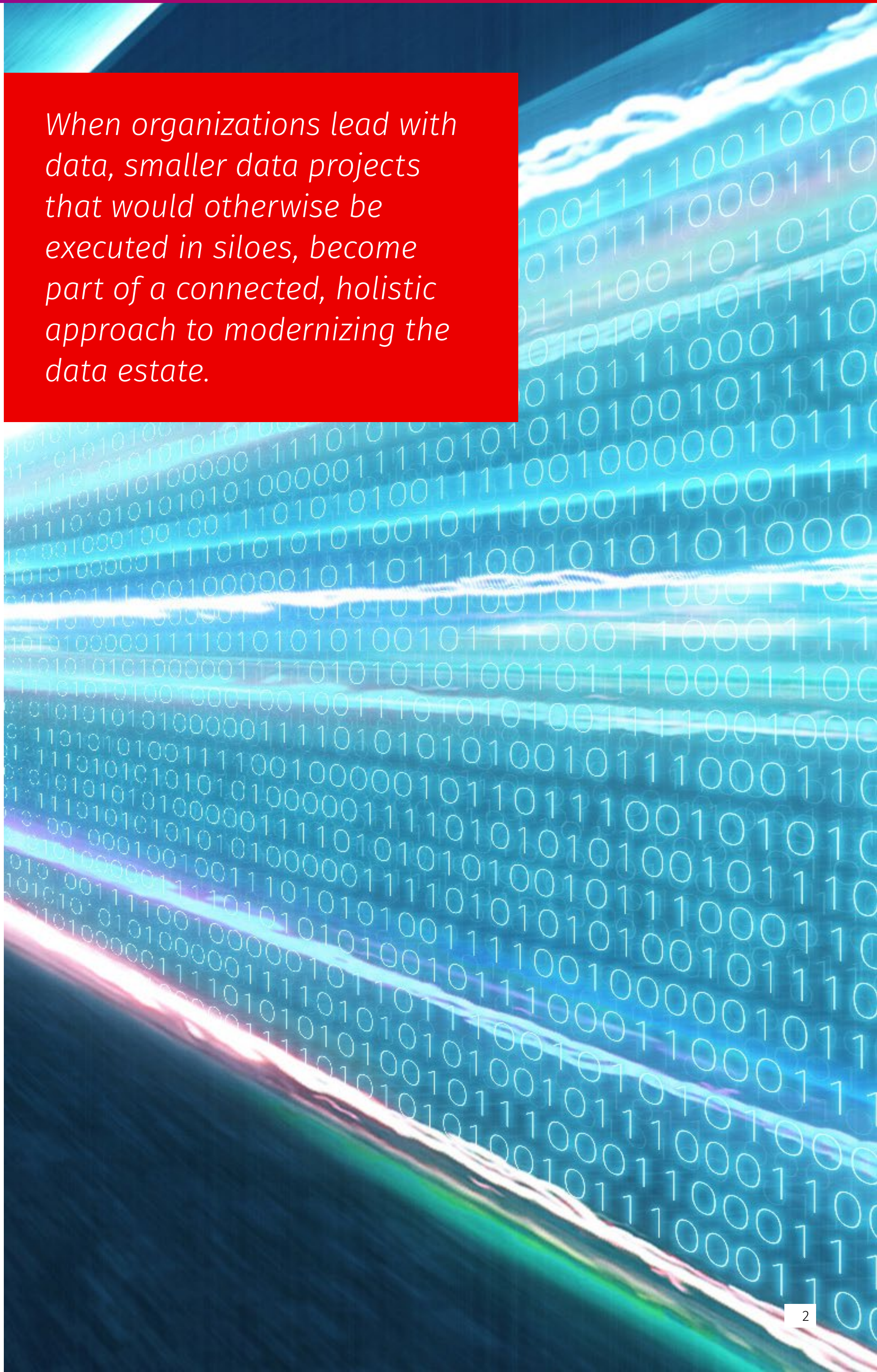
A data-led modernization initiative differs from an application-led approach in several ways. In an application-led approach, most organizations begin their data modernization journey with a cloud readiness assessment, which is a standard methodology for helping make key decisions regarding cloud migration. The assessment is typically followed by a cloud migration assessment, during which the application landscape is inventoried to identify application dependencies that will later help shape a cloud migration and modernization plan.

Typically, these plans are focused on lift-and-shift cloud migrations or application modernization. In both scenarios, data is often a secondary consideration, and data improvement tasks are usually piecemealed into different application and workload silos. This leave organizations with no clear view of how to ensure trust or connect the data across the entire application portfolio.

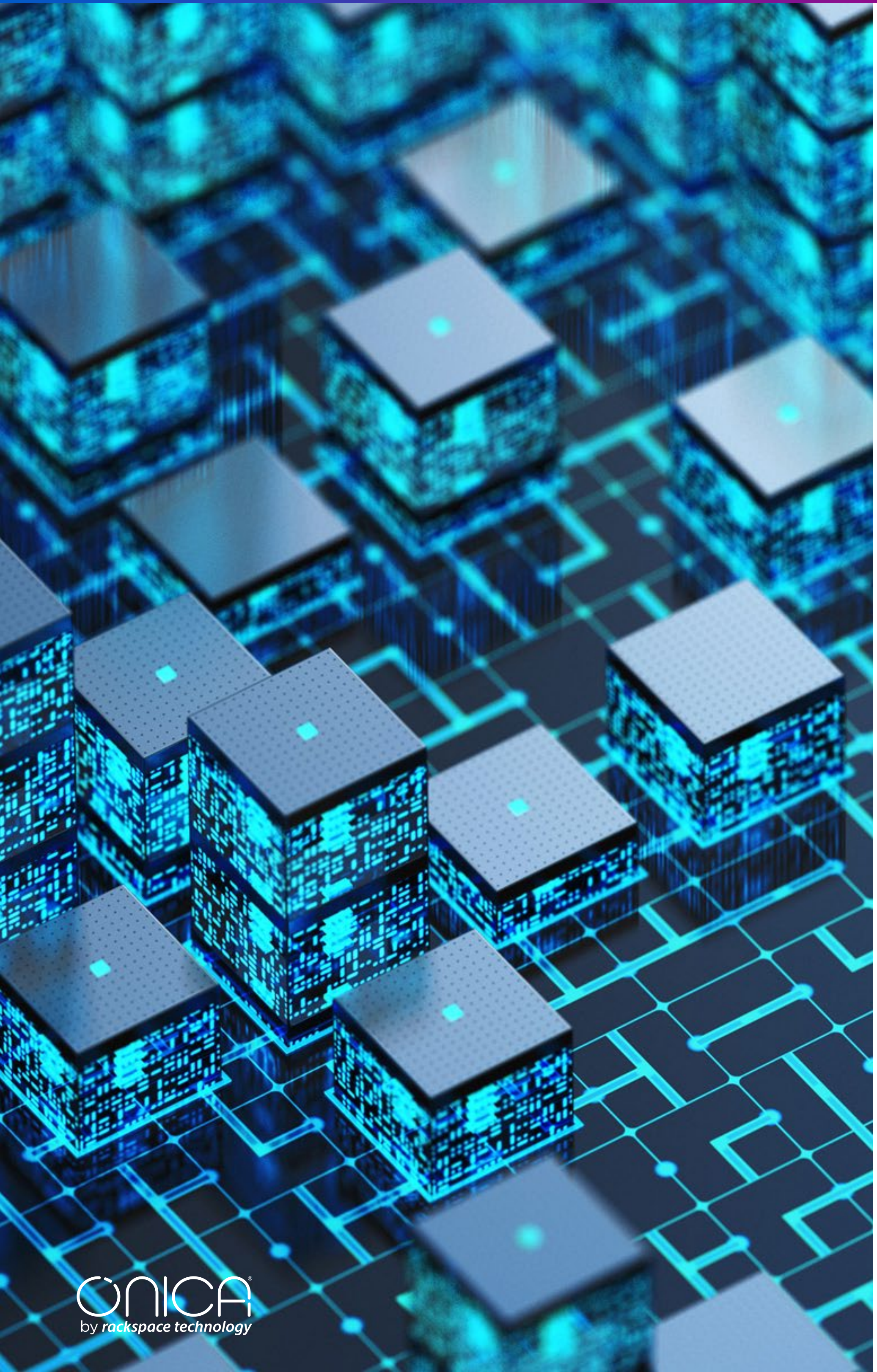
By contrast, a data-led organization approaches modernization by first realizing the value of its data. This opens the door to broader discussions about data estate modernization.

For some, the data journey occurs independently from the application journey, which can accelerate the pace of data estate modernization. For others, the data and application journey will run in parallel — often converging to build intelligent tools and processes that unleash the possibilities of digital innovation.

When organizations lead with data, smaller data projects that would otherwise be executed in siloes, become part of a connected, holistic approach to modernizing the data estate. You can start with relatively small business challenges, such as financial reporting, data quality or security incidents. Or you can start with a business need such as point-of-sale data collection, IoT project integrations, improved efficiencies, reduced costs, or business insights and visualization. Once value and ROI are proven for the first business challenge, you can then move on to large-scale data estate transformations.



When organizations lead with data, smaller data projects that would otherwise be executed in siloes, become part of a connected, holistic approach to modernizing the data estate.



Part 2: Design principles of a modern data estate

A modern data estate capitalizes on the power of the cloud to ingest, process, store, serve and visualize structured and unstructured data from multiple sources to execute a variety of tasks. The core components of a modern data estate should include:

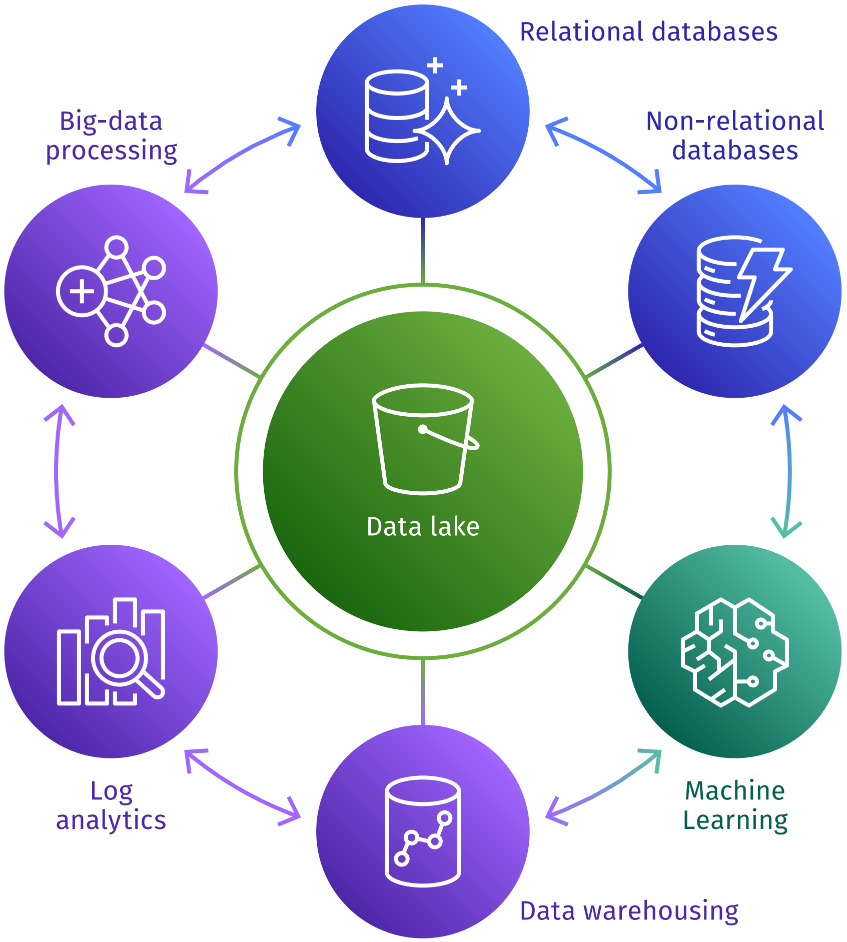
- **Data-aware applications:** Building data-driven applications that power business innovations, augmented with predictions and interactions that leverage AI and machine learning
- **Connected streams of data:** Enabling new business models that drive new revenue streams and leverage federated or event-driven architectures for faster collaboration
- **Adaptive data architectures:** Enabling business agility to meet rapidly shifting trends using reusable patterns to accelerate data product development
- **Governed data and policies:** Leveraging trusted datasets that are self-describing and available on an internal data marketplace that is compliant, secure and audited according to a zero-trust framework
- **Scalable data platform:** Democratizing data through sustainable architectures and platforms, which helps improve the total cost of ownership and investment ROI
- **Resilient operations:** Adding automation and observability to drive operational efficiency, so you can create self-healing systems, efficiently manage the data lifecycle, and support growth and expansion of the data estate

Also, a modern data estate should be architected with the six core pillars from the AWS Well-Architected Framework, including:

1. Operational excellence
2. Security
3. Reliability
4. Performance efficiency
5. Cost optimization
6. Sustainability

Also, a modern data estate brings advanced capabilities to these critical functions:

- **Data integration:** Fragmented data in structured and unstructured formats creates challenges for users. Data should be cleaned, reformatted and filtered before it's used in analytics. Creating a single uniform platform will improve data management and efficiency, and data quality and usability.
- **Common data platforms (e.g., lake houses):** Creating a central repository for commonly used data is an essential component of a modern data estate. Using a lake house pattern to store data can help consolidate data sources, keep historical data clean, improve data quality and enhance accuracy.
- **Business intelligence:** Making smarter, data-driven decisions means centralizing around a tool that can scale with you as you create more reports and dashboards.
- **Streaming analytics:** Real-time data can help you gain a strategic advantage in business planning, operations, monitoring and security. Getting started with the right framework supports both streaming and batch workloads.
- **Search and analytics:** Mining for operational intelligence is a key process in IT departments. However, data varies in format and frequency, so it needs to be formatted so it can be used by most analytic tools. Amazon OpenSearch helps manage and monitor operational data. It's easy to deploy, is scalable, integrates with other AWS services and allows companies to pay for only what they use.



Source: aws.amazon.com/blogs/big-data/build-a-lake-house-architecture-on-aws

Data analytics and AWS

AWS is a popular choice for building modern data estates thanks to its comprehensive suite of cloud services that offer flexibility, scalability and security. AWS's wide range of data storage, processing, governance and analytics services are designed to work seamlessly. This allows you to more easily integrate your data storage, analytics and machine learning workflows.

AWS's suite of services includes Amazon Athena, Amazon EMR, Amazon Redshift, Amazon Kinesis, Amazon OpenSearch Service, Amazon QuickSight, AWS Glue, Amazon S3, Amazon DataZone and Amazon SageMaker. Here's a quick look at each:

Amazon Athena: An interactive query service that enables you to analyze data in Amazon S3 using standard SQL.

Amazon EMR: A managed Hadoop framework that allows you to process large amounts of data using open-source tools, like Apache Hadoop, Apache Spark and Presto.

Amazon Redshift: A data warehouse service that allows you to store and analyze large amounts of data using SQL.

Amazon Kinesis: A service that enables you to collect, process and analyze real-time streaming data, such as video, audio and IoT device data in a managed streaming framework.

Amazon OpenSearch: An open source distributed search and analytics suite that helps you perform interactive log analytics, real-time application monitoring, observability and website search.

Amazon QuickSight: A business intelligence service that allows you to create and share interactive dashboards and reports.

AWS Glue: A fully managed ETL service that makes it easy to move data between data stores.

Amazon S3: A simple storage service that allows you to store and retrieve data from anywhere on the web.

Amazon DataZone: A governance service for discovering and sharing data at scale across organizational boundaries with access controls.

Amazon SageMaker: A fully managed machine learning service that enables developers and data scientists to build, train and deploy machine learning models at scale.

Scalability and security

One of the primary reasons why companies choose AWS for their modern data estate is its ability to scale. AWS provides an elastic infrastructure that can easily scale up or down to accommodate your changing needs. This allows you to easily adjust your infrastructure to meet fluctuating demands without incurring up-front capital expenditures. Also, AWS's pay-as-you-go pricing model means you only pay for what you use, which helps reduce your overall cloud spend.

Another key advantage of using AWS for your modern data estate is security. AWS has a range of security and compliance measures in place to protect data, including encryption, access controls and network isolation. AWS also adheres to a range of industry-standard compliance certifications, including HIPAA and PCI DSS. This helps to protect your data from unauthorized access and breaches.

Together, these advantages make AWS a comprehensive and reliable choice for supporting a modern data estate.

Part 3: Beyond the technology: ongoing data management

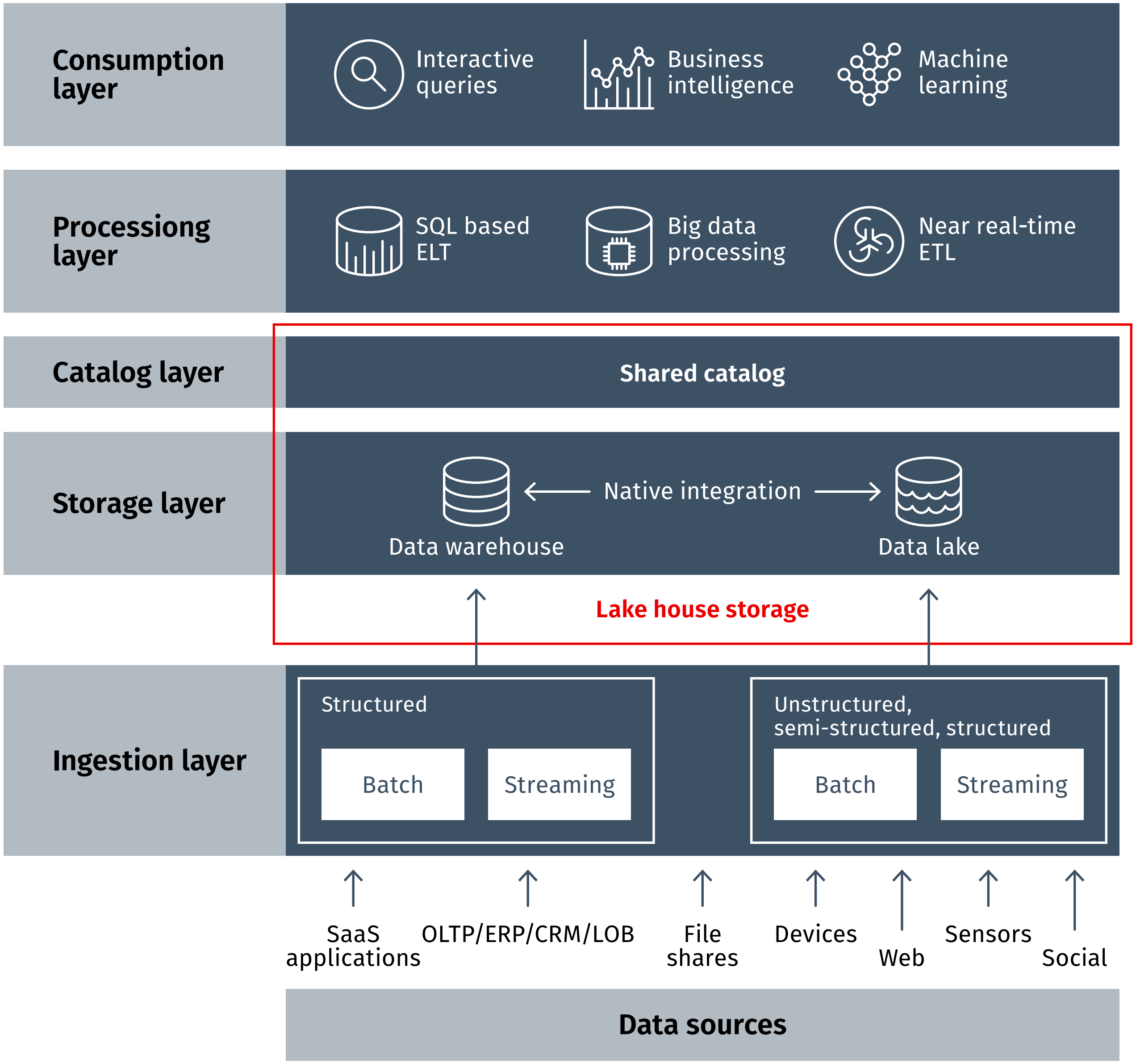
A successful data modernization strategy requires more than just technology to deliver business outcomes. While the technology is critical to success, so are your change management capabilities. You have to closely consider how best to manage the experience of your end users when you roll out new applications. After all, end-user acceptance is critical to the successful use of any new application.

In addition, organizations often fail to drive the process changes required to execute a smooth transition to a new operating model. These failures are often exacerbated by internal resources that haven't been developed yet to provide the specialized expertise required to deploy a wide range of new and emerging technologies.

Your data modernization goals should include:

- ✓ **A robust data platform:** Build a durable platform that can manage the flow of data, create stability and help a data-driven environment thrive.
- ✓ **Outcome-focused use cases:** Create business use cases and applications that focus on customer outcomes and include well-articulated metrics of success.
- ✓ **Developing expertise:** Remember to account for access to the expertise you'll need to design, deploy, manage and optimize the data platform.

This diagram illustrates a modern lake house architecture on AWS.



Source: aws.amazon.com/blogs/big-data/build-a-lake-house-architecture-on-aws

- ✓ **Detailed processes:** Implement processes that ensure data from new sources is properly ingested and managed
- ✓ **Cloud native development:** Integrate automation and orchestration into application design on a cloud-based platform.
- ✓ **The reduction of manual tasks:** Use machine learning to help automate tasks like data entry, workflow builds and raw data conversions.
- ✓ **Active data governance:** Manage the lifecycle and expectations of data to improve its usability and quality for end users, such as eliminating silos and enabling manageable cross-team sharing.

Service providers bring the expertise

As businesses move toward a modern data platform, the need for professional services intensifies. The right service provider should be able to function as an extension of your team, filling skills gaps where needed, and providing expertise across the entire data journey — design, build, manage and optimize.

Working with the right partner can help ensure that your data modernization projects are stickier, quicker-to-market and deliver faster time-to-value, compared to enterprise lines of business applications that may take years to modernize and move to the cloud.

We know that a modern data estate alone does not guarantee long-term success and profitability. As businesses and their data grow, monitoring and managing large volumes of data ingestion across several data pipelines throughout an entire data analytics platform becomes labor-intensive. This often depletes development teams time, which could be better spent on core application and engineering tasks.

Common barriers to change

To accelerate innovation and remain competitive, companies must adopt a collaborative data management practice focused on improving communication, integration and the automation of data flows between managers and consumers. Of course, achieving predictable delivery and change management doesn't come without challenges.

Among the common barriers to change management are:

- Complex data landscapes and processes with highly diverse data sources and tools
- A lack of agility, making it hard to respond to rapidly changing business requirements
- The need to manage a growing number of reports and dashboards over the entire lifecycle
- An inability to understand business needs or quickly respond to changing requirements
- The challenge of transforming machine learning models from the experimental stage to production
- A lack of alignment among data scientists, product owners and data engineers

Deliver value faster

Onica by Rackspace Technology™ has DataOps figured out. We have designed and deployed a platform called **Guzzle** for DataOps that enables analytics engineers to build data pipelines for their data warehouses and data lakes. It allows the creation, deployment and monitoring of data pipelines, which consist of ingestion, processing, reconciliation, data quality and house-keeping activities. Guzzle provides extraction, transformation, loading, validation and reconciliation, while leveraging a domain specific language (DSL) to simplify configurations.

While your organization stays focused on high-level configuration settings and business logic for data pipelines, Guzzle can handle the lower-level implementation details for you. Built on the foundation of Apache Spark, Guzzle leverages Spark connectors to extract and load data at massive scale on most common datastores, both relational data warehouses and data lakes.

What makes Guzzle so powerful?

Here are some of the benefits of using Guzzle for DataOps:

- Native to Apache Spark and big data
- Simple to deploy and use
- Encapsulates commonly occurring design patterns in data pipelines
- Supports of wide array of source and target technology
- Provides details about traceability and provenance of job runs
- Delivers deeper support for DevOps, including out of box integration with Git (and Git workflows), test-automation and auto-deployment

Pivot now to become a data-driven organization

To fully capture and optimize the opportunities in a data-driven estate, you must rethink the way you approach your cloud journey. You need to shift from an application-only approach to a data-led approach that includes applications and data platform management. This means conducting discovery and assessing the data estate early in the process, apart from any singular application, to ensure repeatability and scalability across current and future customer use cases that impact business outcomes.

Organizations that want to adopt a data-led approach should consider these key tenants:

- ✓ A data-led approach should be designed with data used as a product within your organization with various stakeholders, users and streams of data.
- ✓ If you want to completely understand your business environment, including your customers and your value chain, you need a modern data architecture and the ability to continuously enhance and optimize it. This architecture becomes the foundation for intelligent cloud native applications that incorporate all data needed when it's needed.
- ✓ This will require new skills, methodologies and a willingness to bring agility to data curation, management, integration and security. To maximize value, look for data service providers who can support you across every stage of your data journey

These tenets offer a roadmap for organizations that are looking to innovate with data. They become the linchpin for companies in a time where data has the potential to determine the winners and losers of our current fourth industrial age.

We're ready to help you modernize across your data estate

We think about data from end-to-end. We start with building the appropriate strategy based on a thorough assessment of your data needs. The Well-Architected Review for Data helps us evaluate key AWS technologies and gives you access to our data experts who can provide you with the guidance and help you need to execute a successful data modernization initiative.

Ready to tackle data modernization? Talk to an expert today.

Whether you're just getting started with data modernization, or you're ready to conduct a big data strategy assessment, Onica can analyze your needs and create a road map for deploying a modern data platform on AWS across your enterprise.

1-800-961-2888
www.rackspace.com/data/aws-data

ONICA
by rackspace technology

About Onica by Rackspace Technology

Onica by Rackspace Technology™ is the dedicated Amazon Web Services (AWS) business unit at Rackspace Technology serving North America. Onica® helps customers drive innovation, agility, cost savings and operational efficiency on AWS. Onica delivers professional services and engagement practices focused on bringing the most cutting-edge AWS capabilities to every customer through deep expertise in AWS strategy, cloud native development, containers, application modernization, AI and machine learning, and IoT. With in-depth advisory services built on 15 competencies and experience with 1,000+ customer launches, Onica is here to help you accelerate cloud native transformation on AWS.

Learn more at www.rackspace.com or call 1-800-961-2888.

THE INFORMATION CONTAINED IN THIS DOCUMENT IS A GENERAL INTRODUCTION TO RACKSPACE TECHNOLOGY SERVICES AND DOES NOT INCLUDE ANY LEGAL COMMITMENT ON THE PART OF RACKSPACE TECHNOLOGY.

You should not rely solely on this document to decide whether to purchase the service. Rackspace Technology detailed services descriptions and legal commitments are stated in its services agreements. Rackspace Technology services' features and benefits depend on system configuration and may require enabled hardware, software or additional service activation.

Except as set forth in Rackspace Technology general terms and conditions, cloud terms of service and/or other agreement you sign with Rackspace Technology, Rackspace Technology assumes no liability whatsoever, and disclaims any express or implied warranty, relating to its services including, but not limited to, the implied warranty of merchantability, fitness for a particular purpose, and noninfringement.

Although part of the document explains how Rackspace Technology services may work with third party products, the information contained in the document is not designed to work with all scenarios. any use or changes to third party products and/or configurations should be made at the discretion of your administrators and subject to the applicable terms and conditions of such third party. Rackspace Technology does not provide technical support for third party products, other than specified in your hosting services or other agreement you have with Rackspace Technology and Rackspace Technology accepts no responsibility for third-party products.

Rackspace Technology cannot guarantee the accuracy of any information presented after the date of publication.

Rackspace-Ebook-Modern-Data-Estate-on-AWS-TSK-8336 :: March 20, 2023