Embracing a Data-centric Culture Anchored by a Cloud Data Lake

By Mike Leone, ESG Senior Analyst
July 2019

This ESG White Paper was commissioned by Amazon Web Services and is distributed under license from ESG.
# Contents

**Introduction** .......................................................................................................................................................................... 3  
**The Right Way to Embrace a Data Lake** ................................................................................................................................ 4  
  **Current Data Lake Challenges** ........................................................................................................................................... 5  
  **On to the Cloud** ................................................................................................................................................................. 5  
**Essential Elements of a Cloud-based Data Lake** ................................................................................................................ 6  
**Comprehensive Database and Analytics Services with AWS** ................................................................................................. 7  
  **Addressing Cloud-based Data Lake Requirements with AWS** ........................................................................................... 9  
    **Data Movement with AWS** ........................................................................................................................................... 9  
    **Infrastructure and Management with AWS** .................................................................................................................. 9  
    **Analytics with AWS** ..................................................................................................................................................... 10  
  **Visualization with AWS** ............................................................................................................................................... 10  
  **Machine Learning with AWS** ....................................................................................................................................... 11  
**The Bigger Truth** ................................................................................................................................................................. 11
Introduction

The value of data has never been higher. Organizations are doubling down on their data analytics investments and initiatives in the hopes of gaining more valuable insights to shape and grow their businesses. In fact, more than half of organizations will increase their IT spend over the next 12-18 months on business intelligence, analytics, and/or big data.¹ Whether looking to improve business processes, reduce time to market for products and services, or deliver a game-changing customer experience that sets the business apart from the competition, the speed at which actionable and valuable insight can be attained (and, just as importantly, applied to the business) makes the difference between market laggard and market leader.

While any organization would love to immediately leverage their advanced analytics, numerous, ongoing challenges can create roadblocks that prevent organizations from reaching both short- and long-term goals. Constantly growing and changing data sets, operational silos spread across globally distributed environments, and the desire to incorporate different types of data into analytics projects are just the start of the complexity. Many businesses are stuck on traditional infrastructures—on-premises databases and data warehouse appliances with legacy hardware and software limitations—which simply can’t keep up with modern requirements.

ESG recently asked senior level IT personnel with knowledge of/responsibility for their organization’s analytics initiatives and goals about the factors that have had an adverse impact on their organization’s data analytics strategy and investments over the last 24 months. Cost has the largest adverse impact to an organization’s data analytics strategies and investments, as cited by 37% of respondents (see Figure 1). More than one-third of respondents cited too many disparate data sources, and 35% cited lack of skills necessary to properly manage data sets and derive value from them. Limited collaboration across the business, including various personnel involved in data strategy such as IT, data analysts, and line of business owners, was cited by 33% of organizations. And when accounting for the distributed nature of data, meeting security, governance, and compliance requirements was cited by 32% of organizations.²

**Figure 1. Top Five Areas that Adversely Impact Data Analytics Strategies and Investments**

<table>
<thead>
<tr>
<th>Which of the following has had an adverse impact on your organization’s data analytics strategies and investments over the last 24 months? (Percent of respondents, N=310, multiple responses accepted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
</tr>
<tr>
<td>Too many disparate data sources</td>
</tr>
<tr>
<td>Lack of skills necessary to properly manage data sets and derive value from them</td>
</tr>
<tr>
<td>Limited collaboration between IT, analysts, and/or line of business</td>
</tr>
<tr>
<td>Meeting security, governance, and compliance requirements</td>
</tr>
</tbody>
</table>

*Source: Enterprise Strategy Group*

² Source: ESG Research Survey, *2019 Data Analytics Trends*, June 2019. All ESG research references and charts in this white paper have been taken from this set of master survey results unless otherwise noted.
So how can organizations gain instant accessibility to data, infrastructure flexibility to handle changing and diverse data sets, on-demand scalability to handle real-time performance requirements to support high concurrent user counts, cost-effectiveness to keep budgets in check, and ensure reduced risk through a single security model that can be applied to all data? By turning to a cloud-based data lake.

The Right Way to Embrace a Data Lake

Many organizations stream, collect, store, and analyze both structured and unstructured data in their own siloes, whether physically or operationally. These organizations have existing technologies that serve a specific business unit, analytic function, or use case. Between relational and non-relational databases, enterprise data warehouses (EDWs), and business intelligence (BI) applications, organizations are finding legacy architectures imposing a long list of limitations, which include a constant struggle to keep the business secure by mitigating threats across an ever-widening attack surface, as well as unrelenting efforts to enforce access policies across the organization. Then add scale and performance challenges, integration gaps between tools in the data pipeline (if a data pipeline even exists yet), accessibility restrictions, an inability to support modern application requirements, and high costs. With many business processes relying on data found in different silos, it’s not surprising to find inefficiencies across discrete systems. To integrate, process, and analyze all this disparate data requires a great deal of effort to extract, transform, and load (ETL) the data, leading to delays in accessing data in a timely manner or worse, simply avoiding data integration altogether, which can limit insights or yield incorrect insights.

Data lakes were introduced to address these limitations. ESG research shows that while just 13% of organizations currently rely on a data lake, an additional 47% are planning to implement or explore usage of the technology in the next 12 months. By serving as a centralized repository or collection of potentially relevant data that unites disparate data silos into a virtual single entity (regardless of structure, size, type, or dependency), data lakes can help organizations better organize, discover, analyze, understand, and gain value from their data. In the same research, ESG asked respondents what their organization’s objectives are for utilizing a data lake. Figure 2 highlights the top five objectives and demonstrates that leveraging a cloud data lake can clearly enable organizations to successfully meet each objective.

Figure 2. Top Five Objectives for Utilizing a Data Lake

<table>
<thead>
<tr>
<th>Objective</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improve scalability</td>
<td>39%</td>
</tr>
<tr>
<td>Merge structured and unstructured data</td>
<td>32%</td>
</tr>
<tr>
<td>Improve application development times</td>
<td>28%</td>
</tr>
<tr>
<td>Improve data sharing and collaboration</td>
<td>27%</td>
</tr>
<tr>
<td>Analyze data in place</td>
<td>24%</td>
</tr>
</tbody>
</table>

Source: Enterprise Strategy Group
The cloud offers limitless scalability to address the demands of a dynamic organization. Whether organizations require additional storage and compute capacity or need to incorporate different types of storage and compute, the right cloud vendor will present the resources to best meet analytics requirements. This includes the ability to bring together structured and unstructured data into a single repository that can be shared across an organization, from IT through data scientists to developers. With a single repository of all data and numerous tools that enable in-place analysis, organizations can recognize shorter time to value, improved insights, and faster development cycles.

**Current Data Lake Challenges**

While organizations’ first instance of collecting data from multiple sources, curating it, and centrally storing it for analysis was done in an enterprise data warehouse (EDW), limitations around scalability, performance, type of supported data, and cost have forced organizations to look for other ways to analyze diverse data sets. The initial architecture they turned to included node-based, scale-out architectures deployed on generic hardware with distributed storage and processing resources (i.e., Hadoop). This served as an effective means to initially satisfy multiple business units that desired a more comprehensive view of data, while utilizing a more cost-effective, scalable infrastructure. Organizations could migrate certain data sets and workloads from their enterprise data warehouse to their data lake.

This continues to be a common practice and is supported by ESG research, which shows the top two ways enterprise data warehouse and data lake technologies interact are first, to gain key optimizations for certain workloads by offloading from an EDW to a data lake, and second, to enable organizations to better incorporate next-generation technologies like AI and ML. But as pressure continues to be placed on different areas of the business to gain real-time, actionable insight, organizations are experiencing unforeseen challenges—for starters, the inability to satisfy real-time requirements and mitigate management complexity. Then add to the list the less-than-efficient speed at which data can be ingested, an inability to accurately discover and catalog data effectively, and immature security and governance tools.

The cost and complexity of buying, building, maintaining, and supporting a rapidly changing and expanding hardware and software environment consisting of numerous open source projects is simply more trouble than it’s worth. With more than 70% of organizations implementing an on-premises data lake today, ESG data suggests a move to cloud could be helpful.

**On to the Cloud**

According to ESG research, cloud-based analytics is the most often cited area within data analytics where organizations are making the most significant investment over the next 12-18 months.

For data lakes specifically, the cloud offers organizations numerous advantages compared with implementing an on-premises data lake. And to avoid having to make tradeoffs, organizations should prioritize comprehensiveness—the ability to deliver on all the promises the cloud can offer a data lake implementation, including both current and future data-driven goals. Those promises include:

- **Unified data** – Enable organizations to reduce the number of globally distributed data and management silos into a single, globally accessible data store.

- **Pervasive security** – Enforce centrally controlled policies (if not, a single security policy for all) to manage and control all data and user access, while incorporating all the aggregated knowledge and best practices of thousands of organizations, learning from each customer’s requirements. While organizations generally understand which policies they need to enforce, implementation can present challenges.

- **Performance and scalability** – Utilize on-demand, decoupled resources that provide predictable, scale-out performance with a wide selection of configurations for memory, processors, and storage. The elasticity delivered by the cloud enables organizations to break free of common limitations experienced in an on-premises data lake.

---

implementation, including idle resources, data locality to nodes/applications, and limited accessibility across analytic use cases. As an example, if a workload or use case requires the use of 100 GPUs for an hour or simply a different resource profile, the time, cost, and risk associated with making configuration adjustments is far too high on-premises.

- **Agility** – Reduce the burdens of resource procurement and management that face a typical enterprise and let data dictate how it should be analyzed, through flexible, one-click integrations of different services, tools, processing engines, frameworks, and policies based on specific goals and use cases.

- **Reliability and availability** – Leverage built-in redundancy throughout the entire technology stack, using appropriate and efficient processes to avoid any interruption of service, even spanning geographic zones.

- **Self-service enablement** – Expose as much relevant data as possible to as many people as possible, unlocking a greater potential for valuable insights, all within comprehensive security and privacy guardrails.

- **Cost** – Benefit from massive economies of scale through offers of resources, such as auto-scaling and spot instances, and simplified management that are more than what most businesses could afford on their own, including a pay-as-you-go pricing model to minimize upfront investments. Additionally, organizations use less expensive storage and storage tiering for the data, rather than more costly storage within various analytics tools/platforms.

- **Future proof** – Anchor a forward-looking, data-centric strategy by using a platform that can foster growth and ensure integration of next-generation technologies, like AI and ML, that can easily access large pools of training data.

### Essential Elements of a Cloud-based Data Lake

Organizations want the promise of an all-encompassing data lake and while the cloud may help, all solutions are not created equal. ESG research shows that a top consideration when it comes to a data lake technology solution is the vendor ecosystem, including the comprehensiveness of the infrastructure and analytics portfolio. And it’s not just the current portfolio, but the vendor’s roadmap and proven track record of meeting roadmap timelines. What advancements, integrations, and enabling technologies are going to be added to that portfolio in the next 12-18 months? For a data lake to meet the current and equally important future demands of the data-centric and data-driven businesses, organizations must account for four architectural pillars:

1. **Data movement** accounts for the migration and streaming of data. This pillar should have answers for all questions associated with loading and synchronizing data: Where is the data coming from and going to? At what rate is the data coming and going? How frequently will that data be accessed and used for analytics purposes? Is the chosen vendor capable of handling all of the organization’s requirements?

2. **Data lake infrastructure and management** addresses where the data gets stored, as well as how the data is prepared for analytics, including security, data catalogs, and ETL functionality. This includes the ability to deliver a highly available and therefore reliable infrastructure with granular security controls to ensure reduced risk. Additionally, ensuring the right level of price/performance through intelligent data management provides organizations with peace of mind. Vendors should be answering questions such as: How accessible is the data? Who has access to what data sets? How can high data quality be ensured? To what extent does infrastructure availability impact the ability to derive insights? What type of security controls are available?
3. **Analytics** is all about leveraging the right tools to effectively analyze data. A data warehouse engine can be leveraged to analyze structured data. Big data processing with Apache Spark, Hive, and HBase satisfy the workloads that must incorporate unstructured data. On-demand resource allocation with serverless analytics enables organizations to focus on the data and less on scaling the infrastructure. In-place interactive querying with SQL enables organizations to query data with the most widely known query language available. Operational analytics for application monitoring ensures applications are running at peak performance regardless of load. And log and clickstream analytics ensures end-users are satisfied based on current demand by making streaming data available to modern applications that require real-time insight. Critical questions in this category include: Is the right tool or service available and integrated? Can multiple use cases and supporting applications meet strict SLAs associated with concurrency, performance, and scalability?

4. **Visualization** enables organizations to see and leverage data through interactive and actionable dashboards infused with predictive insights. In terms of visualizations, organizations should ask: How quickly is data available to my visualization tools? Is self-service analytics attainable through intuitive interfaces that can scale on demand?

5. **Machine learning** can be incorporated seamlessly through a foundational technology in a modern data lake. Organizations can be on the best path to leverage next-generation technologies like AI and ML, while empowering self-service throughout a business that prioritizes data-centricity. This includes the ability to incorporate preferred ML frameworks, notebooks, and development environments, and simplified ML services that accelerate adoption. IT leaders should ask: How seamlessly can advanced technologies like AI and ML be integrated into existing analytics workflows? Is the right set of tools and services available to support my use case? What level of expertise is required to utilize available tools so my organizations can benefit from AI and ML?

**Figure 3. Essential Elements of a Data Lake**

Comprehensive Database and Analytics Services with AWS

As a leading cloud service provider, AWS offers a comprehensive portfolio of database and analytics services that enables organizations to embrace data-centricity. Through the availability of more than 50 data services, customers can simplify
the process of gaining insight from their data—starting with a cloud data lake and ending with a future-proof platform that enables the integration of next-generation technology (such as AI) into modern, distributed applications.

Based on ESG research, organizations looking to adopt new technologies to support data analytics initiatives cited security as their top capability/attribute to look for in a solution. Reliability, performance, and cost followed security as the most important capabilities to look for in a platform to support data-driven goals. AWS addresses these requirements by offering key services and approaches across the entire data analytics stack.

AWS offers one of the most secure cloud platforms available today, with Amazon S3 data lake storage solution providing almost unlimited scalability. With AWS CloudTrail, customers gain visibility into every API access and data event, while leveraging machine learning (ML) to alert customers when anomalous data activity is detected. Amazon Macie security service discovers, classifies, and protects sensitive data stored in AWS (e.g., personally identifiable information, personal health information (PHI), intellectual properties (IP), and financial data), alerting customers when critical data is accessed or moved. Amazon offers services that not only encrypt data in flight when replicating data across regions, but also reports on encryption status to satisfy compliance standards. The solution also offers fine-grain access control, at both a file and object level. Further, organizations are able to launch their AWS services in virtual networks using Amazon VPC and deliver granular access and permission control using AWS Identity and Access Management and Kerberos for authentication.

With agility and choice being a key value proposition from cloud providers, it’s no surprise they serve as core pillars to the AWS approach. By enabling customers to store and analyze data the way they want using open formats and open standards and providing a constantly growing list of partner solutions with prebuilt integrations, organizations gain flexibility in ensuring their current and future needs are met regardless of their stage of data and analytics maturity.

As it relates to data lakes built on AWS, organizations can benefit from numerous intelligent services that automatically help ease the cost burden. With Amazon Glacier, infrequently accessed data can be moved to the most cost-effective AWS storage for long-term backup or archive, and with Amazon S3 management, data access patterns can be analyzed to determine the type of movement, whether on-demand or based on an applied lifecycle policy. Regardless of where the data is stored, it can be queried, and through services like Amazon S3 Select and Glacier Select, subsets of data can be retrieved to improve query performance. And with a pay-as-you-go or reserved instance option, organizations gain flexibility in how they pay for consumed processing resources.

Two great examples of AWS services that incorporate intelligence and automation can be found in Amazon Managed Streaming for Apache Kafka (MSK) and Kinesis Video Streams. Amazon MSK automatically provisions highly available Apache Kafka clusters with settings and configuration based on Apache Kafka’s deployment best practices, continuously monitors cluster health, and automatically replaces unhealthy nodes with no downtime to the application. Kinesis Video Streams can automatically provision and elastically scale all the infrastructure needed to ingest streaming video data.
Addressing Cloud-based Data Lake Requirements with AWS

ESG research shows that 49% of organizations would consider a single vendor to satisfy their data lake and data pipeline requirements—but the top reason preventing organizations from choosing a single vendor is that they do not believe a single vendor could satisfy all their requirements. Although it may appear that few, if any, vendors could satisfy every organizational requirement, organizations should perform due diligence to find the vendor that can satisfy most of their requirements and then supplement the chosen vendor with third-party software partners. A comprehensive review can help an organization eliminate the time and expense of learning, maintaining, and paying for two different environments. In essence, AWS can offer the tools and solutions to satisfy each stage of an organization’s analytics journey, from movement, infrastructure, and management, to analytics, visualization, and machine learning.

Data Movement with AWS

AWS helps customers move data into its cloud easily and reliably with a collection of comprehensive services depending on where the data currently resides. AWS Snowball, AWS Snowball Edge, and AWS Snowmobile enable customers to leverage physical appliances to move petabytes of data from on-premises environments to AWS. AWS Storage Gateway can be used to redirect data from on-premises applications to AWS. AWS Direct Connect enables faster data transfers with a dedicated network connection from the data center to AWS. Combining AWS S3 Transfer Acceleration and AWS’s distributed edge locations, customers can accelerate data transfers across longer distances. And Amazon Kinesis and AWS IoT Core can best enable organizations looking to capture or load streaming or IoT data.

For databases, the AWS Database Migration Service (DMS) helps organizations migrate databases to AWS quickly and securely while ensuring the source database remains fully operational during the migration, minimizing downtime to applications that rely on the database. The AWS DMS can not only migrate data to and from the most widely used commercial and open source databases in the market, but with the AWS Schema Conversion Tool, migrations are also more predictable due to automatic conversion of the source database schema to a format compatible with the target database. And with data integration being a component of data movement, AWS Glue serves as a fully managed extract, transform, and load (ETL) service that makes it easy for customers to prepare and load their data for analytics.

Infrastructure and Management with AWS

Leveraging the AWS infrastructure for a data lake enables customers to have a single, centralized data repository that collects and stores all data in its original format. Durability, availability, and resiliency is gained through replication across multiple available zones. And factoring in features like cross-region replication using company-specific keys stored by Amazon’s key management service, maintaining an effective security posture is simplified.

Further, Amazon S3 offers the ability to audit how and when individual objects are accessed and by whom. AI-based security services deliver automatic monitoring, alerting, and reporting, enabling features like real-time anomaly detection and mitigation, encryption status, and daily inventories. Object-level control enables custom and granular policies that can be applied to data, such as automatic tagging, tiering, and access restriction. With S3, organizations can gain insight into their data, such as request and access patterns, tiering recommendations for cost optimization, content-type classification, and custom reports with heat mapping to highlight frequency of object groups being accessed.

Georgia-Pacific built a centralized data lake on Amazon S3 with data streamed real-time by Amazon Kinesis from manufacturing equipment to ingest and analyze data at scale.

“We are using AWS storage and analytics to predict precisely how fast converting lines should run to avoid tearing. By reducing paper tears, we have increased profits by millions of dollars for one production line.”

Vice President of IT/Digital Transformation, Georgia-Pacific
The Amazon S3 reduced redundancy storage option allows organizations to store non-critical, reproducible data at lower levels of redundancy, providing 99.99% durability and 99.99% availability of objects over a given year. This durability level corresponds to an average annual expected loss of 0.01% of objects.

With tight integrations from traditional storage vendors, tens of thousands of consulting systems integrators, independent software vendor partners, and 3,500+ software listings in the AWS marketplace, S3 pre-integration is designed to help customers get up and running without delay.

**AWS Lake Formation**

Within the infrastructure and management pillar of AWS’s data-centric services stack, the recently announced AWS Lake Formation is a service that simplifies the process of creating a data lake on AWS. Organizations are able to define where data resides, as well as access and security policies that enable Lake Formation to collect, catalog, classify, clean, and secure data as it moves into the Amazon S3 data lake. This allows organizations to access a centralized catalog of data, including descriptions of specific data sets and the best way to leverage data across the wide range of analytics and machine learning services AWS provides.

**Analytics with AWS**

AWS offers numerous analytics services that enable customers to analyze data their way. With Amazon Redshift, organizations can utilize a cloud data warehouse to query petabytes of structured data or utilize Redshift Spectrum to run SQL queries directly on structured or unstructured data in S3. Amazon EMR enables customers to utilize Spark and Hadoop frameworks as a managed service, with support for 17 additional open source projects, including HBase and Presto, which are updated within 30 days of a new version release.

Amazon Kinesis enables real-time analytics by collecting, processing, and analyzing streaming data from IoT devices, application logs, and website clicks. With the Amazon Elasticsearch Service, IT is able to search, explore, filter, aggregate, and visualize IT operational data in real time. For interactive analysis, Amazon Athena provides a serverless architecture to analyze data directly in S3 and Glacier using standard SQL. And speaking of serverless, AWS is focused on delivering its services in a serverless manner, ensuring customers spend less time worrying about infrastructure and more time focusing on deriving insights.

**Visualization with AWS**

The fourth pillar of the AWS portfolio is visualization. By delivering services that enable all personnel within an organization to see relevant data and derive insight, AWS is empowering the level of self-service that organizations are demanding. With Amazon QuickSight, customers can cost-effectively build powerful, interactive dashboards that can be accessed anywhere, on any device, and scale to meet the growing number of supported users.

“Beeswax uses Amazon S3 and AWS Glue Data Catalog to build a highly reliable data lake that is fully managed by AWS. Our platform leverages the AWS Glue Data Catalog integration with Amazon EMR in Hive and Spark SQL applications to deliver reporting and optimization features to our customers.”

**CTO, Beeswax**

“20 percent of our queries now complete in less than one second. Best of all, we didn’t have to change anything to get this speed-up with Redshift, which supports our mission-critical workloads.”

**Executive Director of Technology, Edmunds**

“The QuickSight pay-per-session dashboard access is perfect as it allows secure, fast and cost-effective access to interactive data. As a cloud-based solution, QuickSight automatically scales to our needs.”

**Product Manager, Siemens**
Machine Learning with AWS

The final pillar of the AWS portfolio is about leveraging next-generation technology and advanced analytics for predictive insights. Regarding machine learning, AWS delivers three layers that enable different personas to leverage the power of predictive insights. Services like Amazon SageMaker and AWS Deep Learning AMI satisfy the expert machine learning practitioners in the frameworks and interface layers. The platform services layer extends the availability of machine learning to developers by using SageMaker to eliminate the complexity and guesswork across every step of the machine learning pipeline. Lastly, the application services layer provides prebuilt AI services that can easily be leveraged by developers possessing no ML skills. These services include Amazon Rekognition for image analysis, Amazon Transcribe for speech to text, Amazon Polly for text to speech, Amazon Lex for conversational apps, and Amazon Comprehend for unstructured text analysis.

The Bigger Truth

As organizations continue to embrace a data-centric mantra, data lakes can serve as a foundation to the modern enterprise. But data lake technology won’t be successful if it remains in the rarefied realm of data scientists, or even data engineers. With adequate security in place, data lakes should be made easily accessible to a wide range of users, from CIOs and line of business owners, to developers and business analysts. If implemented properly, a data lake will enable all personnel throughout a business to ask more questions, come together with new ideas, and walk away with useful and actionable insights. And to ensure a data lake can satisfy both the current and future demands of a business, it’s essential to architect it with surrounding technologies and services that enable cost-effective simplicity, agility, scalability, and security. With the right tools and best practices, an organization can use all its data, making it accessible to more users and fueling better business decisions.

When looking to properly architect a data lake solution, ensure the key elements are satisfied: data movement, infrastructure, management, analytics, visualization, and machine learning. Organizations must be able to bring data to the data lake on their terms. They must be able to reliably store, manage, and prepare the data for analytics. They must be able to access the data from a wide range of analytics solutions based on the size and type of data, the speed at which it needs to be analyzed, and the cost associated with analyzing it. And finally, they must be able to see and interact with the data through visualizations and dashboards—while being placed on a path to easily integrate next-generation technologies like machine learning for predictive insights.

With AWS, customers can check all the boxes with a comprehensive portfolio of data-centric services, many of which are fully managed by AWS. This enables organizations to focus less on maintaining the infrastructure and more on the data and gaining valuable and timely insights. With tight integration across the continuously growing AWS big data and data analytics services stack, organizations gain a future-proof data platform anchored by an endlessly scalable data lake.

AWS is helping businesses of all sizes build productive data lakes (marquee customers include Nasdaq, FINRA, Capital One, and GE Oil & Gas). And while every business might not be ready to transition from an on-premises approach to a cloud-based data lake, AWS is well positioned to help organizations explore new and better ways to store, process, and analyze their most valuable asset—data.

“With AWS and our move to the cloud, we can build a truly modern machine learning ecosystem, with all of our data connected and fully available.”

Managing Vice President of Machine Learning, Capital One

© 2019 by The Enterprise Strategy Group, Inc. All Rights Reserved.