



From data lakes to rivers of insight:  
Our vision for the Oil & Gas  
industry and AWS partnership



For Oil & Gas companies, it's a time of great challenges, and even greater opportunities. We understand that uncertain times create pressures on every area of your business: Upstream, unexpected maintenance costs can have a big impact, while in Downstream, disrupters lurk. Meanwhile, geopolitical risks continue to give rise to unpredictable oil pricing.

AWS gives your company the agility to stay competitive in the face of ever changing prices, policy, technology and customer preferences. There are incredible opportunities to do more—especially around unleashing the business value in your data lakes.

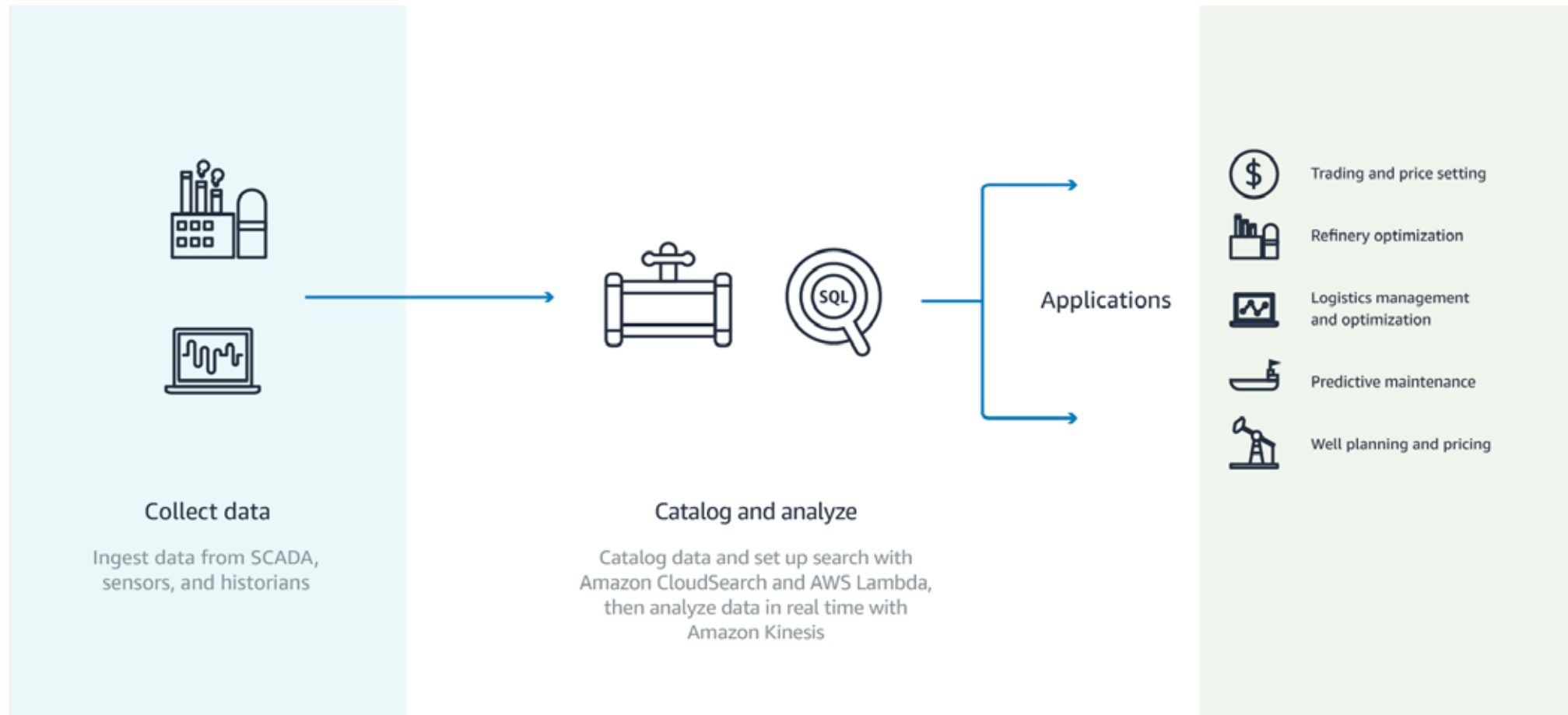
This is the time to explore how AWS can help your Oil & Gas company turn the vast information stored in your data lakes into fast-flowing rivers of insight for every part of the business.

This book outlines our vision for how AWS data solutions will bring business benefits to each stage of your company's operations. We would love to explore the topics discussed in more detail with you in person over the coming weeks.



# Our vision for Oil & Gas companies

The future of successful business lays in creating fast-flowing rivers of insight from the data lakes of information kept by modern organizations. In this document we set out what that could mean for each stage of your operations. But before we jump into exploring that vision, let's first set out exactly what a data lake is.



# What is a data lake?

Organizations are tasked with managing greater volumes of data, from more sources, and containing more types of data than ever before. In the face of massive, heterogeneous volumes of data, many organizations are finding that they need a storage and analytics solution that offers more speed and flexibility than legacy systems.

A data lake is a new, increasingly popular way to store and analyze data that addresses many of these challenges, by allowing an organization to store all data in one centralized repository. Since data can be stored in its original form, there is no need to convert it to a predefined schema before ingestion. This provides a way to store all your data—both structured and unstructured—with minimal lead time. And by eliminating the need to store data in any proprietary format, which may not exist in a few years' time, your data is future-proofed. With a data lake on AWS, you no longer need to know what questions you want to ask of your data before you store it, giving you a flexible platform for data analysis.

A data lake on AWS can also help break down silos between domains and operating units, revealing information you didn't know existed—the proverbial needle in a data haystack that can make a multimillion-dollar difference in decision making. Users have access to data reservoirs around the world, providing a truly global perspective on your data. And modelling that once took months can now be done in hours.

At the heart of an AWS-based data lake is Amazon Simple Storage Service (S3), which provides secure, cost-effective, durable, and scalable cloud storage. AWS also offers an extensive set of services to help you provide strong security for your data lake—even as you increase access to it, for both internal employees and external partners. Your AWS data lake includes granular access controls, virtual isolation, 256-bit encryption, logging and monitoring, and more.

There's a variety of ways to transfer data to your data lake, including through services like:

## **Amazon Kinesis**

Enables real-time data ingestion.

## **AWS Import/Export**

A service for sending a portable storage device with your data to AWS.

## **AWS Import/ Export Snowball**

A secure appliance AWS sends you for ingesting data in batches.

## **AWS Storage Gateway**

Connect on-premises software appliances with your AWS Cloud-based storage.

## **AWS Direct Connect**

A dedicated network connectivity between your data center and AWS.

“ With a data lake on AWS, you no longer need to know what questions you want to ask of your data before you store it, giving you a flexible platform for data analysis. ”

# Truly connected Upstream operations

In Upstream, the biggest benefit of a data lake is the de-risking of huge investment decisions, which speeds up time to first prospect or oil; and knowledge mining. Our vision for your company is to create a truly connected Upstream operation—one that achieves more stable, predictable profits over the long term. Here's how we see that coming to life.

## **Predictable profits: Reduce time taken to extract resources from the ground**

In a world where large scale oil and gas sites are harder to come by, the powerful data science tools of AWS can help quickly determine where to drill from existing structured and unstructured seismic data. This will help current assets work harder and provide the analytics platforms that allow you to discover new resources.

A data lake can positively impact Upstream operations by providing master data management—ensuring globally consistent data, such as unique well identifiers, consistent coordinate systems, and standardized geological data while eliminating duplicate data.

Products like Amazon Sagemaker and Amazon S3 can be quickly put to work to accelerate and optimize exploration, drilling, and production. Using AWS machine learning and big data tools to quickly and easily extract powerful insights from data automates time-consuming, error-prone manual tasks, and improves the speed and accuracy of decision making.

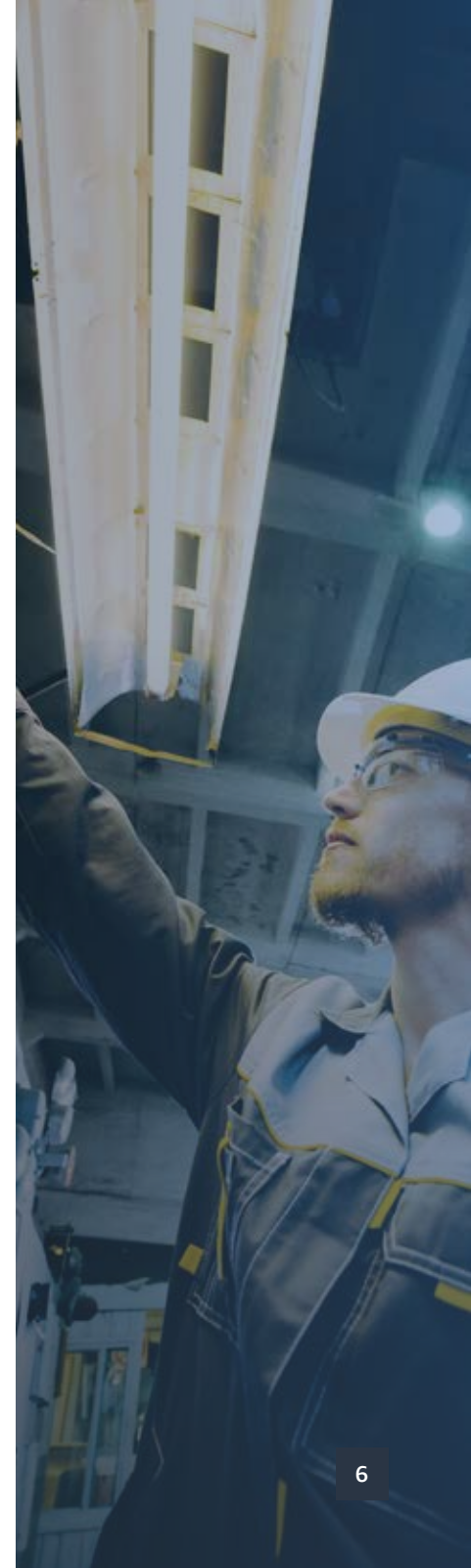
This data can be quickly delivered to Upstream operations staff and business units across the organization for immediate use and impact: facilitating better strategic decision making. Planning and evaluation. Increasing predictability in Upstream output at the engineering level. Creating more stable profits for the whole of your company.

## **Stability of production: IoT ecosystem to ensure near 100% production uptime**

Upstream extraction operations are extremely costly and difficult to set up. Extended periods of downtime and unexpected maintenance issues can have a serious impact on your profit margins.

Our vision is to mitigate the impact of extended maintenance periods on your production flows with a leading approach to IoT infrastructure. With the AWS suite of IoT infrastructure products, you can improve operational efficiencies by connecting sensors at scale, such as predictive maintenance of oilfield pumps.

By connecting AWS IoT infrastructure with near limitless high-performance computing resources and Amazon Quicksight, our data visualization tools unlock the business value in Upstream data by presenting it a way that the whole business can use.



# Rivers of insight at every point of Downstream operations

The barriers to entry in Downstream are lower than in Upstream; your company likely is more susceptible to disruption here than in other areas of business. To protect against new market entrants and meet ever-changing customer demands, it is vital that your company has the flexibility and Insights needed to stay ahead of the game. Here's our vision for how AWS can deliver just that for you:

## **Efficient service: Leveraging Amazon insights to create a seamless supply chain**

Amazon has built a world-class global supply chain for its retail businesses on top of AWS infrastructure—in fact, Amazon S3 technology is the cornerstone of global supply chains in enterprises across the world.

For our existing Oil and Gas customers, it ensures that all data generated in the Upstream process is properly captured, categorized and stored. Amazon S3 provides unparalleled levels of stability and security that continue to meet the demands of the world's most globalized organizations.

A data lake can also combine operating and equipment data (from IoT sensors) with product data (oil production, polypropylene or diesel output, etc.) with financial and economic data, along with other relevant data like environmental or weather conditions.

The analytics capabilities of your AWS data lake can then find hidden relationships between data and their effects, leading to greater operational efficiency: cost reductions, avoidance of downtime, more accurate forecasting, and more.

The AWS-based IoT infrastructure creates rivers of data that must be processed in near-real time—AWS Greengrass meets this task and runs these IoT applications seamlessly across the cloud. For your company, this means that shipping of unrefined oil and gas resources can be monitored and tracked in real time. Total visibility across this complex network means refineries are ready to receive unprocessed goods as soon as they arrive. This reduces production lag and maintains the stable flow of raw oil and gas products into refineries and on to the end consumers.



## Staying competitive: Flexibility to innovate at the customer level

Your enterprise doesn't just need to get raw oil and gas resources to refineries; it must also package, market, and sell the refined oil and gas products to the end consumer. Unfortunately, the barriers to entry in this space are far lower than in your Upstream business. In recent years, new entrants to this market have disrupted the established practices with new and improved customer service models.

With AWS, your company will have all the tools it needs to stay nimble and innovate quickly to meet evolving customer service demands. Because our service offering is built to be flexible, we can quickly meet the demands of customer-facing digital product prototypes; provide deep insights into their successes and failures, and scale according to business decisions without incurring any long-term costs.

## Single view of the customer: Boosting loyalty and improving customer experience

Our vision is to help you reinvent your retail environment to match this demand and serve the changing needs of end consumers. The AWS infrastructure lets you keep customer data in a secure environment. Helping your teams to understand your customers and serve them better.

AWS provides you with a wide range of reliable, scalable, and secure data storage and warehousing services that allow you to collect, store, and optimize this data for future fast analytics queries. Dozens of our customers in the retail space are already using AWS solutions to reinvent the instore customer experience by uncovering as yet unidentified trends in customer behaviour and flexing new prototypes to match this. We can't wait to get you started on this journey.





# The AWS vision for maintaining stability in uncertain times

A fragmenting geopolitical landscape and a swing towards economic nationalism has introduced more risk into the system—meaning that the only certainty in business is uncertainty itself, and that puts your profits at risk.

To monitor these fluid situations and maintain accurate modelling for the future, your analysts will have to process more data and run more scenarios to keep making the best decisions. This means that they will need ever-bigger data lakes and ever higher computational power.

By working with AWS, Oil & Gas companies are able to rethink their operating models and implementing cost-saving measures that increase performance. The results include accelerated go-to-market speeds, strengthened security, enriched customer experiences, and better data-driven decision making. Our vision for your business is to provide access to the largest ecosystem of technology and consulting partners to deploy industry-specific business solutions that elevate every aspect of your organization. Let's explore what that looks like.

## **Security**

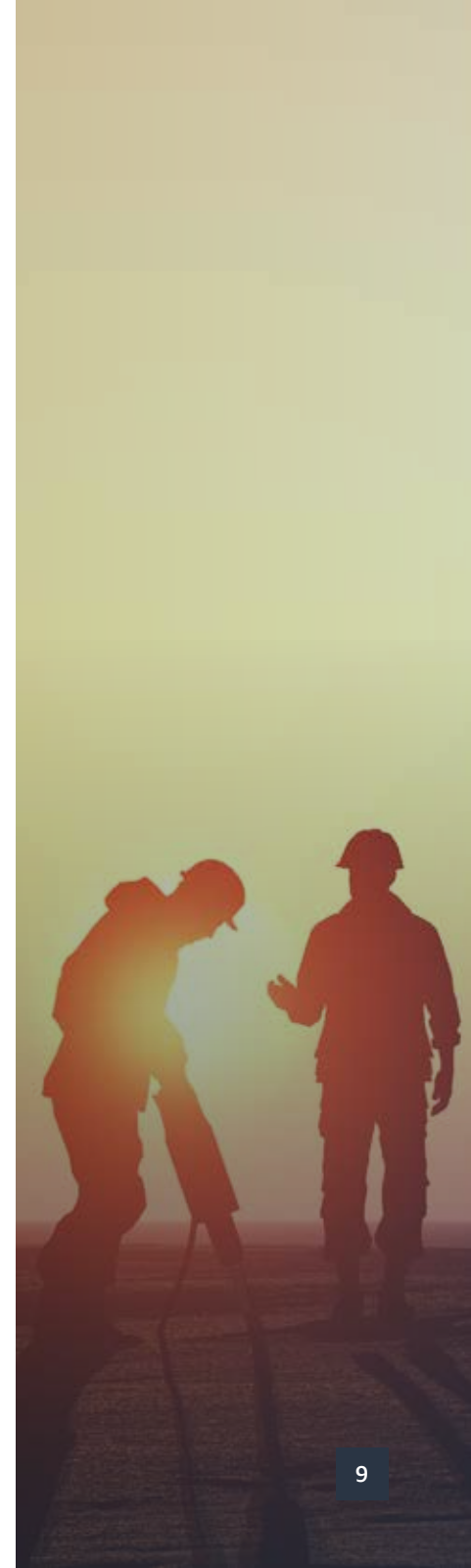
At AWS, security is job zero. Our platform is designed for the most data-sensitive organizations, and our partners offer hundreds of tools and features to help meet objectives around visibility, auditability, controllability, and agility.

## **Breadth of data abilities**

AWS offers more data services than any other platform, so you can collect, store, process, analyze, and visualize big data in the cloud. Deploy virtually any application regardless of volume, velocity, and variety of data.

## **Speed of innovation**

AWS enables companies to rapidly and reliably build, deliver, and iterate innovative products and services. AWS can help simplify provisioning and managing infrastructure, deploying code, and automating software updates.



# Building a data lake on AWS

We hope that you agree that an ongoing partnership between your Oil & Gas company and AWS is hugely exciting—and AWS can make this vision a reality much faster than you might think.

At its core, a data lake solution on AWS leverages Amazon Simple Storage Service (Amazon S3) for secure, cost-effective, durable, and scalable storage. You can quickly and easily collect data into Amazon S3, from a wide variety of sources by using services like AWS Snowball or Amazon Kinesis Firehose delivery streams. Amazon S3 also offers an extensive set of features to help you provide strong security for your data lake, including access controls & policies, data transfer over SSL, encryption at rest, logging and monitoring, and more.

For the management of the data, you can leverage services such as Amazon DynamoDB and Amazon Elasticsearch to catalog and index the data in Amazon S3. Using AWS Lambda functions that are directly triggered by Amazon S3 in response to events such as new data being uploaded, you easily can keep your catalog up-to-date. With Amazon API Gateway, you can create an API that acts as a “front door” for applications to access data quickly and securely by authorizing access via AWS Identity and Access Management (IAM) and Amazon Cognito.

For analyzing and accessing the data stored in Amazon S3, AWS provides fast access to flexible and low-cost services, like Amazon EMR, Amazon Redshift, and Amazon Machine Learning, so you can rapidly scale any analytical solution. Example solutions include data warehousing, clickstream analytics, fraud detection, recommendation engines, event-driven ETL, and IoT processing. By leveraging AWS, you can easily provision exactly the resources and scale you need to power any Big Data applications, meet demand, and improve innovation., meet demand, and improve innovation.



AWS is perfectly placed to help create data lakes that provide fast flowing rivers of insight to line of business units across your enterprise. Benefits to hosting your data lake on AWS include:

### **Cost-effective data storage**

Amazon S3 provides cost-effective and durable storage, allowing you to store nearly unlimited amounts of data of any type, from any source. Because storing data in Amazon S3 doesn't require upfront transformations, you have the flexibility to apply schemas for data analysis on demand. This enables you to more easily answer new questions as they come up and improve time-to-value.

### **Easy data collection and ingestion**

There's a variety of ways to ingest data into your data lake, including services such as Amazon Kinesis, which enables you to ingest data in real-time; AWS Snowball, a secure appliance AWS sends you for ingesting data in batches; AWS Storage Gateway, which enables you to connect on-premises software appliances with your AWS Cloud-based storage; or through AWS Direct Connect, which gives you dedicated network connectivity between your data center and AWS.

### **Security and compliance**

When hosting your data lake on AWS, you gain access to a highly secure cloud infrastructure and a deep suite of security offerings designed to keep your data secure. As an AWS customer, you will benefit from a data center and network architecture built to meet the requirements of the most security-sensitive organizations. AWS also actively manages dozens of compliance programs in its infrastructure, helping organizations to easily meet compliance standards such as PCI DSS, HIPAA, and FedRAMP.

### **Most complete platform for big data**

AWS gives you fast access to flexible and low-cost IT resources, so you can rapidly scale virtually any big data application including data warehousing, clickstream analytics, fraud detection, recommendation engines, event-driven ETL, serverless computing, and Internet of Things (IoT) processing. With AWS, you don't need to make large, upfront investments in time and money to build and maintain infrastructure. Instead, you can provision exactly the right type and size of resources you need to power big data analytics applications.



# Featured data lake partners



## 47Lining

47Lining is an AWS Advanced Consulting Partner with Big Data Competency designation.

Big Data solutions developed by 47Lining are built from underlying AWS building blocks like Amazon Redshift, Amazon Kinesis, Amazon S3, Amazon DynamoDB, Amazon Machine Learning and Amazon Elastic MapReduce (EMR). What does all this mean? Put simply thanks to 47Lining, you can expect fast flowing rivers of insight from your data lakes quicker.



## C3IoT

C3IoT is an Advanced Technology Partner in the AWS Partner Network with IoT and AI competencies. With over nine years' experience, C3IoT has the skills and expertise to create a data lake that meets the unique needs of any business. With the C3 Enterprise data lake, you can auto-scale storage and compute resources, minimizing your costs while maximizing the return on your data investment.



## Optika Solutions

Optika Solutions, an Australian software and product solutions development company, is another partner that brings real business benefits. Optika empowers businesses to easily work with big data, collaboratively, in real time, and effortlessly find insights in their data.



# Case study: Royal Dutch Shell

Like any large corporation with international exposure, Shell needs to protect itself against cybersecurity threats. The oil and gas industry in particular has seen many cyberattacks in recent years, and as a result Shell took a critical look at its security information and event management (SIEM) solution and considered ways to improve it.

Shell also wanted to incorporate cyber threat hunting—the ability to analyze data to proactively identify any vulnerabilities. The company's SIEM solution had also reached the physical limits of what it could do. They were already pushing more data through it than the architecture could handle. They needed a SIEM environment that could easily scale.

## Moving to AWS Cloud

Shell expanded its SIEM solution by adopting Splunk Enterprise and Splunk Enterprise Security, a platform that can rapidly search and analyze historical machine and log data from its various systems. It chose to host the platform on AWS because AWS offered the scalability and flexibility needed to accommodate Shell's huge global footprint.

The company now has a data lake of multiple petabytes to use for historical analysis, which means its cyber defense team can engage in identifying trends and proactive cyber threat hunting. It can also prevent incidents from occurring by identifying vulnerabilities through data analysis and closing them upfront.

## Benefits of AWS

- Scalable infrastructure for expanding cybersecurity solution
- Ingests four terabytes of data each day into Splunk Enterprise
- Stores a two-petabyte data lake for trending analysis and cyberthreat hunting
- Utilizes AWS storage options for hot and cold data, gaining optimal performance and cost
- Increases detection and remediation of potential security breaches by 100%

[Read the full case study](#)

“ Our cyber defense team is now finding more than twice as many events that could have resulted in security incidents and breaches. We are really happy with the flexibility, scalability and functionality of our Splunk SIEM solution on AWS, compared to our old on-premises solution. ”

**Oskar Brink, Cyber Defence Manager, Royal Dutch Shell**

# Case study: Woodside

Woodside owns and manages gas plants around the world. In order to improve maintenance efficiency, the company wanted to use its plants' sensors to predict maintenance needs and help avoid issues that, if unattended, can shut down the plant for weeks or months.

## Moving to AWS Cloud

Using IoT based sensors that run on AWS, Woodside was able to predict up to one week prior to when issues would occur—and could act to keep the plant running. At one plant, the implementation of the monitoring technology took only six weeks. Woodside scaled up to activate 200,000 sensors in the AWS Cloud.

## Benefits of AWS

- Scaled with flick of a switch
- Runs 6,000 analytics models and three million daily calculations
- Monitors equipment and enables preventative maintenance before major issues occur
- Looking to expand to and connect sensors on onshore plants, offshore plants, and vessels

“ When I started, we had a culture of technology evolving over years; now it's months and down into weeks. It has had such a positive impact on what traditionally was a heavy industry, to one that's pushing the boundaries of IoT and making a difference. ”

**Shaun Gregory, Senior Vice President and CTO, Woodside**

# Why AWS?

With AWS, your company can have durable, available and high-performance services can scale as much as you need, with a combination of a low storage cost and no minimum commitments. Together with our partner network, AWS makes finding insight a matter of exploration, not excavation.

**How can the Oil & Gas industry achieve the agility to accelerate innovation and stay competitive?**

Learn more at:

[aws.amazon.com/oil-and-gas](https://aws.amazon.com/oil-and-gas)

[aws.amazon.com/oil-and-gas/data-insights](https://aws.amazon.com/oil-and-gas/data-insights)

**Ready to get started?**

Get in touch with us.

[aws.amazon.com/contact-us](https://aws.amazon.com/contact-us)



