eBook

Cybersecurity in Financial Services

Protecting financial institutions with advanced analytics and AI

databricks

Contents

INTRODUCTION

The State of the Industry

Cloud, cost and complexity of customer data and cybersecurity are top of mind for every financial services security leader today. As financial services institutions (FSIs) continue to accelerate their digital transformation, cybercriminals, fraudsters and state-sponsored actors continue with more sophisticated threats. The impact of these attacks ranges from the exposure of highly sensitive data to the disruption of services and the exploitation of backdoors for future attacks — all resulting in both financial and non-financial costs. Responding quickly to potential threats requires security tools capable of analyzing billions of threat signals in real-time.

Recently, it seems like every week reveals a new data breach or ransomware assault, and the cost is skyrocketing: more than $4 million per incident, up 10 percent from 2020, and about $401 million for a substantial breach at a large corporation.

Cybersecurity is no longer just a back-office cost and now poses critical business risks, such as:

- Operational disruption
- Material customer loss
- Increase in insurance premiums
- Lawsuits or fines
- Systemic destabilization
- Credit downgrade
- Reputational damage

Source: Navigating Cyber 2022, FS-ISAC, Annual Cyber Threat Review and Predictions

A New Commitment to Cybersecurity

It comes as no surprise that in recent years FSIs have seen an amplified commitment to cybersecurity. As business leaders look to new solutions, large portions of IT budgets are now devoted to leveraging data and AI to thwart cyberattacks.

Furthermore, regulators are taking notice of the increased risk of cybersecurity threats. Growing geopolitical tensions have also prompted federal agencies such as the Cybersecurity and Infrastructure Security Agency and the Federal Bureau of Investigation to warn that "tough sanctions imposed on Russia could prompt a spate of cyberattacks against critical infrastructure such as banks." Additionally, the Securities and Exchange Commission released its 2022 Exam Priorities, which include information security, and specifically "how firms are safeguarding their customers' records and assets from cyber threats, including oversight of third-party providers, identification of red flags related to identity theft, response to incidents, including to ransomware attacks and management of operational risk in light of 'a dispersed workforce.'"

However, as is often the case, implementing new cybersecurity strategies and processes is easier said than done.

In this eBook, we'll take a closer look at the challenges associated with replacing the infrastructure of a legacy data analytics system, and how financial institutions are solving them with Databricks.

Cybersecurity needs a transformation

… breaches, cost and complexity are growing

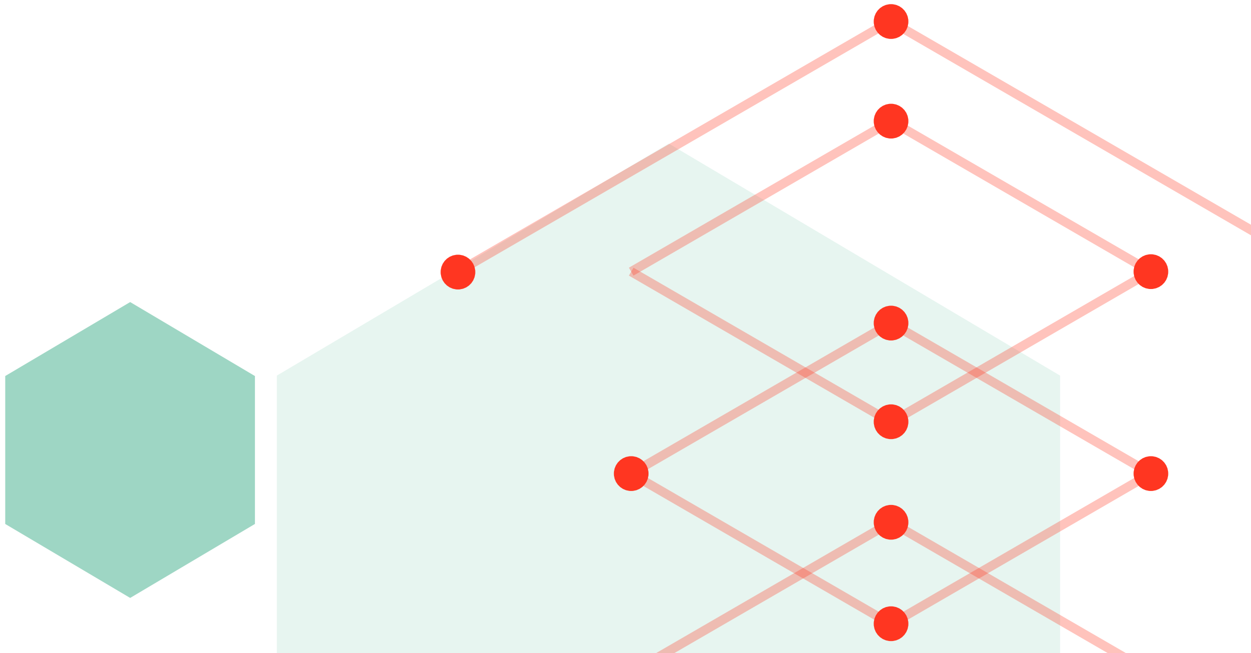
100%
of organizations surveyed have had breaches.
The average breach costs $4M

67%
of organizations were **breached at least three times**. A mega breach costs $401M.

85%
will increase their cyber budget next FY. Cybersecurity industry will grow to $366B by '28

Cost, Complexity, Cloud

- Hundreds of tools with expanding footprints
- Data locked in vendor proprietary tools
- Humans compensating for analytical and integration deficiencies



The Biggest Challenge With Security Analytics

For many FSIs, on-premises security incident and event management (SIEM) technologies have been the go-to solution for threat detection, analysis and investigations. However, these legacy technologies were built for a world where big data was measured in gigabytes, not today's terabytes or petabytes. This means that not only are legacy SIEMs unable to scale to today's data volumes, but they are also unable to serve the modern, distributed enterprise.

By now, the advantages of moving to the cloud are no secret to anyone. For FSIs, scalability, simplicity, efficiency and cost are absolutely essential components of success. Many within FinServ are looking to cloud computing to make this possible, adding detection and response in the cloud to the security team's responsibility.

Because legacy SIEMs predate the emergence of cloud, artificial intelligence and machine learning (AI/ML) in the mainstream, they're unable to address the complex data and AI-driven analytics needed for threat detection, threat hunting, in-stream threat intelligence enrichment, analytical automation and analyst collaboration.

In other words, legacy SIEMs are no longer suitable for the modern enterprise or the current threat landscape.

Counting the Financial Cost of Legacy SIEMs

The financial cost of the continued use of legacy SIEMs continues to rise because most SIEM providers charge their customers based on the volume of data ingested. While some legacy technologies are available in the cloud, they're either not designed to be cloud-native applications or confined to a single cloud service provider. As a result, security teams have to employ multiple tools for detection, investigation and response — or pay exorbitant egress charges for data transiting from one cloud provider to another. This causes operational slowdowns, errors driven by complexity, and inconsistent implementation of security policies.

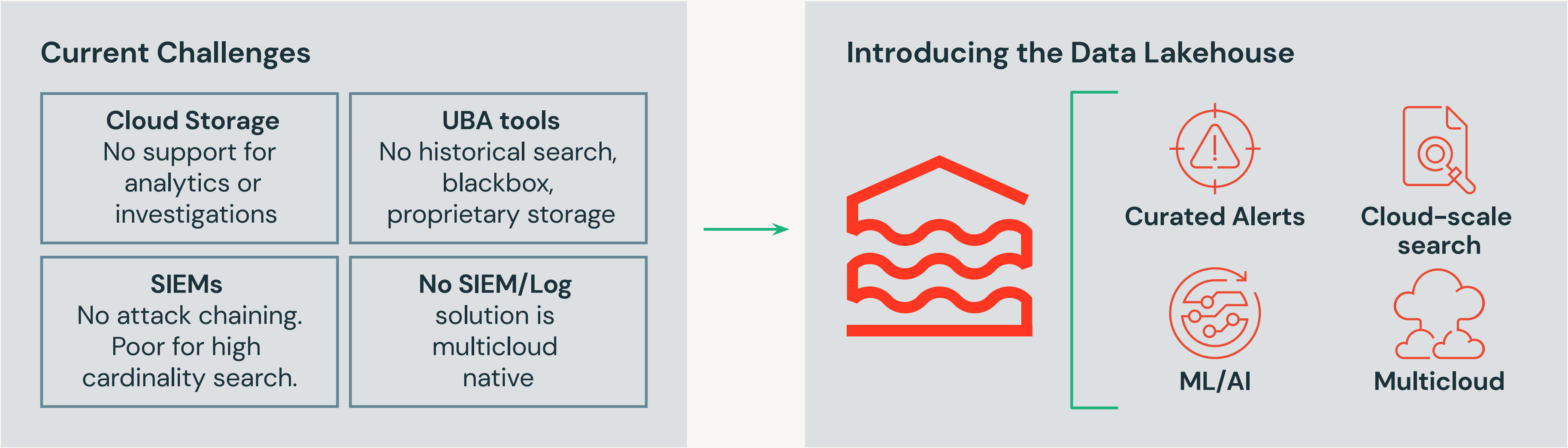
A lack of support for multiple clouds also means an increase in maintenance overhead. Security staff members are often stressed because analysts have to learn different tools for different cloud platforms. For some, it also creates an implicit cloud vendor lock-in, meaning that security teams are unable to support missions because their tools are not portable across multiple cloud providers.

Collectively, these drawbacks to legacy SIEMs result in a much weaker security posture for FSIs.

Journey of SecOps: Destination Lakehouse

How did security analytics get to this point? In the early days, there was a need to aggregate alerts from antiviruses and intrusion detection systems. SIEMs were born, built on data warehouses, relational databases or NoSQL database management systems. But as incident investigation needs evolved, those data warehouses weren't able to handle the volume and variety of data, which led to the development of data lakes. Data lakes were cost-effective and scalable but didn't have strong data governance and data hygiene, earning them the moniker of "data swamps." Simply integrating the two tech stacks is really complicated because of varying governance models, data silos and inconsistent use case support. Fast-forward to today, security teams now need AI/ML at scale in a multicloud world.

Why choose one or the other? The lakehouse architecture has emerged in recent years to help address these concerns with a single unified architecture for all your threat data, analytics and AI in the cloud. The governance and transactional capabilities of the data warehouse, the scale and flexibility of a data lake, AI/ML from the ground up and multicloud native deployments in one platform — this is a modern architecture called the lakehouse (data lake and data warehouse).



Current Challenges

Cloud Storage
No support for analytics or investigations

UBA tools
No historical search, blackbox, proprietary storage

SIEMs
No attack chaining. Poor for high cardinality search.

No SIEM/Log
solution is multicloud native

Introducing the Data Lakehouse

Curated Alerts

Cloud-scale search

ML/AI

Multicloud

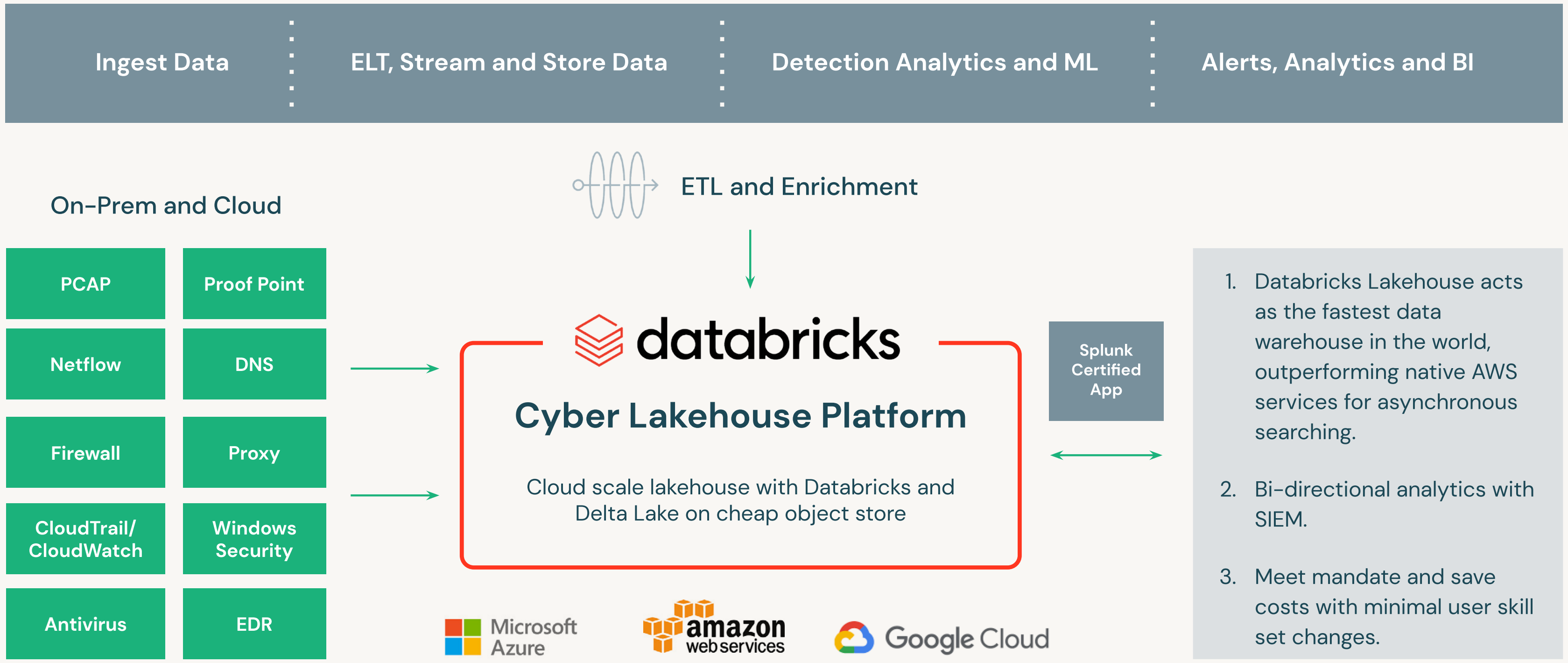
Rethinking Cybersecurity in Financial Services With Databricks Lakehouse

Databricks introduced the first data lakehouse platform to the industry, and today over 7,000 customers use it worldwide. With Databricks Lakehouse, FSIs that are ready to modernize their data infrastructure and analytics capabilities for better protection against cyber threats now have one cost-effective solution that addresses the needs of all their teams.

The Databricks Lakehouse Platform combines the best elements of data lakes and data warehouses, delivering the low-cost, flexible object stores offered by data lakes and the data management and performance typically found in data warehouses. This unified platform simplifies existing architecture by eliminating the data silos that traditionally separate analytics, data science and ML. It's built on open source, open data and open standards to maximize flexibility, and its inherent collaborative capabilities accelerate the ability to work across teams and innovate faster. Moreover, because it's multicloud, it works the same way no matter which cloud provider is used.

Databricks Cyber "Multi-Tier" Architecture

A lakehouse platform for scalable, real-time threat analytics

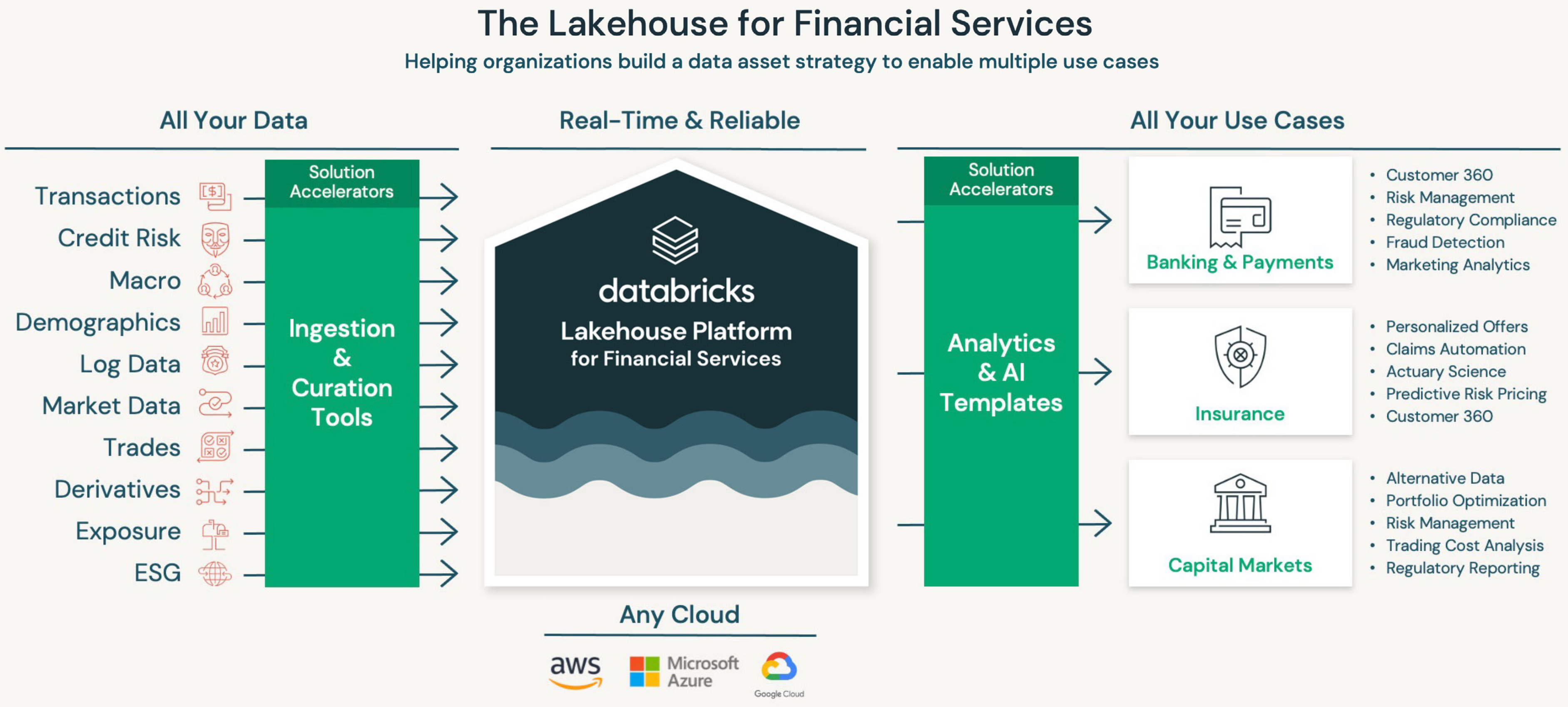


1. Databricks Lakehouse acts as the fastest data warehouse in the world, outperforming native AWS services for asynchronous searching.
2. Bi-directional analytics with SIEM.
3. Meet mandate and save costs with minimal user skill set changes.

Lakehouse in Financial Services

By unifying data with analytics and AI, Lakehouse allows FSIs to easily access all their data for downstream advanced analytics capabilities to support complex security use cases. Lakehouse facilitates collaboration between threat intelligence teams and cyber operations, enables security operations teams to detect advanced threats, and reduces human resource burnout through analytical automation and collaboration. Importantly, Lakehouse also accelerates investigations from days to minutes.

Along with a more modern architecture, the Lakehouse Platform includes Delta Lake, which unifies all security data in a transactional data lake to feed advanced analytics. The analytics and collaboration are done in notebooks, and security teams can use multiple languages — SQL, Python, R and Scala — in the same notebook. This makes it easy for security practitioners to explore data and develop advanced analytics and reporting using their favorite methods. Additionally, a separation of compute from storage means performance at scale without impacting overall storage costs.



The Lakehouse for Financial Services

Helping organizations build a data asset strategy to enable multiple use cases

- Customer 360
- Risk Management
- Regulatory Compliance
- Fraud Detection
- Marketing Analytics

- Personalized Offers
- Claims Automation
- Actuary Science
- Predictive Risk Pricing
- Customer 360

- Alternative Data
- Portfolio Optimization
- Risk Management
- Trading Cost Analysis
- Regulatory Reporting

CASE STUDY

When It Comes to Security, Data Is the Best Defense*

Protecting HSBC's 40 million customers begins with collecting and processing data from billions of signals to make previously impossible threat detection possible

The old way of thinking about security — stronger locks, higher walls — is outdated and ineffective. "When defending an organization, too often we just focus heavily on tools, technology, and reactive scenarios," said T.J. Campana, managing director of global defense and chief technology officer at HSBC, the multinational bank. "But the security business is a data business. And the data always has a story to tell us."

The quality of security, he added, is proportional to the information that can be distilled from petabytes of data that endlessly flows through company networks. That means "empowering people to get the right insights, in the right way to quickly prevent, detect, and respond to threats, wherever and whenever they occur," said George Webster, executive director of global cybersecurity science and analytics at HSBC.

If a big organization is made up of tens of millions of parts that must click together seamlessly, security keeps those seals tight. Data gathering, analytical tools, and human intellect work together as one. This involves fusing the data science and security operation departments, creating an enhanced relationship that results in better defenses, insight into the security posture of the organization, and the ability to respond at the pace of the adversary.

But working across years of data at petabyte scale is not an easy task, especially when a long time is measured in minutes and the adversary is constantly working against you. To put this in perspective, the security teams at HSBC intake 10 times the amount of data contained in all of the books in the U.S. Library of Congress every day, and must process months, if not years, of data at a time. That is where innovative design, smart people, and leveraging the right technology come into play. "We have to break the paradigm of the tool being the end goal of defense and instead view the tools as an enabler of our people," said Webster. "It is always about the people," added Campana.

HSBC turned away from the common security paradigm by leveraging the big data processing techniques from Azure Databricks. In many ways, their open source Delta Lake is the key enabler, with Spark being the engine. Delta Lake allows these teams to structure, optimize, and unlock data at scale, while Spark allows multiple complex programs to seamlessly crunch through the data. This enables HSBC's security teams to constantly evolve their defenses, create new capabilities at pace, and perform investigations that were previously impossible. When a new threat emerges, the bank doesn't have the luxury to wait for the security market to identify, respond, and mitigate. Instead, the bank turns to its people and creates what is needed at breathtaking speed.

CASE STUDY: CONTINUED

It's an essential function for HSBC, which needs to continually think about how to keep more than 40 million customers in 64 countries and territories safe. Taken together, it's an all-brains-on-deck moment with data and people guiding the ship. It's also a tall task for a company as massive and multifaceted as HSBC. Headquartered in the UK, it is one of the largest global banks (total assets: a whopping $2.968 trillion), with operations across Africa, Europe, Asia, and the Americas. It's also the largest bank in Hong Kong and even prints some of the local currency, which bears the HSBC name.

The bank's cybersecurity approach involves fusing the data science and security operation departments, creating an enhanced relationship that results in more efficient threat discovery, rapid development of operational use cases and AI models. This enables the continuous creation of capabilities that stop adversaries before they even start. "We have to get out of the mindset that security is a walled garden," said Webster. "We must create truly collaborative environments for our people to enable the business to operate," said Campana.

Staffing this symbiotic power center will be someone Campana optimistically calls "the analyst of the future," a description that's both mindset and skillset: threat hunter and data scientist.

In addition, when another organization is hit by cybercrime, HSBC analyzes it to understand how it may have responded and then improves its defenses accordingly. That's in contrast to the industry norm; a Ponemon survey revealed that 47 percent of organizations have not assessed the readiness of their incident response teams. That means the first time they test their plans will be at the worst possible time — in the middle of a cyber attack.

The proactive approach is a far cry from the old reactive conveyor belt model of security when alert tickets were received from tooling and processed in a slow and linear way. Today, cross-disciplinary security teams don't just react; they continually search for the signals in the noise — tiny aberrations that indicate something's not right – and send up red flags in real-time. "We're scanning hundreds of billions of signals per day. I cannot wait. We need situational awareness right now," said Campana.

That increased speed is critical for threat assessment. Information theft may be the most expensive and fastest-rising consequence of cybercrime, but data is not the only target. Core systems are being hacked in a dangerous trend to disrupt and destroy. Regulators are also increasingly asking banks for controls in place to detect and preempt financial crimes. That's where big data tooling like Delta Lake and Spark shine, and where it will continually be called on to address the security needs of new initiatives.

"Digital security is about organically adjusting to risks," said Webster. "It's a journey of continual discovery with one central goal: to protect customers. They want things easy and they want them quick. It's our job to make sure that it's secure."

*This story previously appeared in WIRED Brand Lab for Databricks.

Advantages of a Lakehouse

A cost-efficient upgrade

Databricks customers only pay for the data they analyze, not for what they collect. This means that security teams can collect any amount of data without worrying about ingest-based pricing, and only pay for the data that's actually used for analysis — for example, an incident investigation or a data call for an audit. This pricing model enables security teams to collect data that was previously out of reach, such as netflow data, endpoint detection and response data, and application and services data.

Further, Databricks is a fully managed service, meaning that security teams don't have to pre-commit to hardware capital expenditures. With no hardware to manage and no big data implementations to maintain, security teams can significantly reduce their management and maintenance costs.

Multicloud

Databricks is cloud-native on AWS, Microsoft Azure and Google Cloud. This creates freedom for the security teams to use whatever cloud provider they like. Additionally, teams can acquire and maintain operational consistency across all providers when they have multiple cloud footprints. This enables consistent policy implementation, reduced complexity for staff and increased efficiency.

Additionally, Databricks enables faster detection, investigation and response across the enterprise because analytics can be reused across the major cloud providers through a unified platform that centralizes data for easy sharing and fosters collaboration across teams.

Enterprise security and 360° risk management

The Lakehouse Platform is easy to set up, manage, scale and, most importantly, secure. This is because Lakehouse easily integrates with existing security and management tools, enabling users to extend their policies for peace of mind and greater control.

With multicloud management, security admins and data teams get a consistent experience across all major cloud providers. This saves valuable time and the resources required to upskill talent on proprietary services for data, analytics and AI.

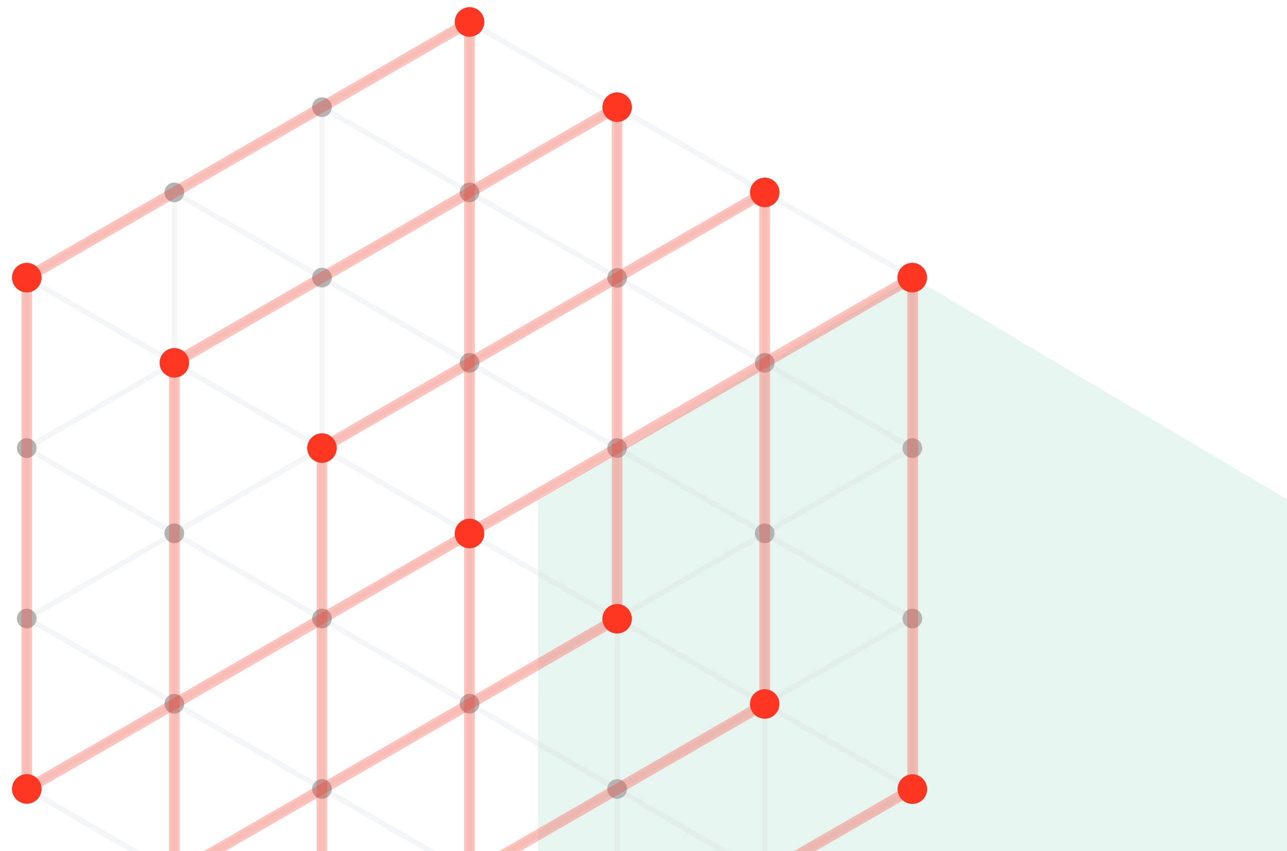
Security, risk and compliance leaders are also able to give team members a range of security permissions that come with thorough audit trails. This allows teams to quickly spin up and wind down collaborative workspaces for any project and to manage use cases from end to end — from enabling user access and controlling spend to auditing usage and analyzing activity across every workspace to enforce user and data governance.

Lakehouse and SIEM: The Pattern for Cloud-Scale Security Operations

According to George Webster, head of cybersecurity sciences and analytics at HSBC, Lakehouse and SIEM is the pattern for security operations. What does it look like? It leverages the strengths of the two components: Lakehouse for multicloud native storage and analytics, SIEM for security operations workflows. For Databricks customers like HSBC, there are two general patterns for this integration that are both underpinned by what Webster calls the cybersecurity data lake with Lakehouse.

In the first pattern, Lakehouse stores all the data for the maximum retention period. A subset of the data is then sent to the SIEM and stored for a fraction of the time. This pattern has the advantage of allowing analysts to query near-term data using the SIEM while having the ability to do historical analysis and more sophisticated analytics in Databricks. It also lets them manage any licensing or storage costs for the SIEM deployment.

The second pattern is to send the highest-volume data sources to Databricks — for example, cloud-native logs, endpoint threat detection and response logs, DNS data and network events. Low-volume data sources such as alerts, e-mail logs and vulnerability scan data go to the SIEM. This pattern enables Tier 1 analysts to quickly handle high-priority alerts in the SIEM. Threat-hunt teams and investigators can leverage the advanced analytical capabilities of Databricks. This pattern has a cost-benefit of offloading processing, ingestion and storage from the SIEM.

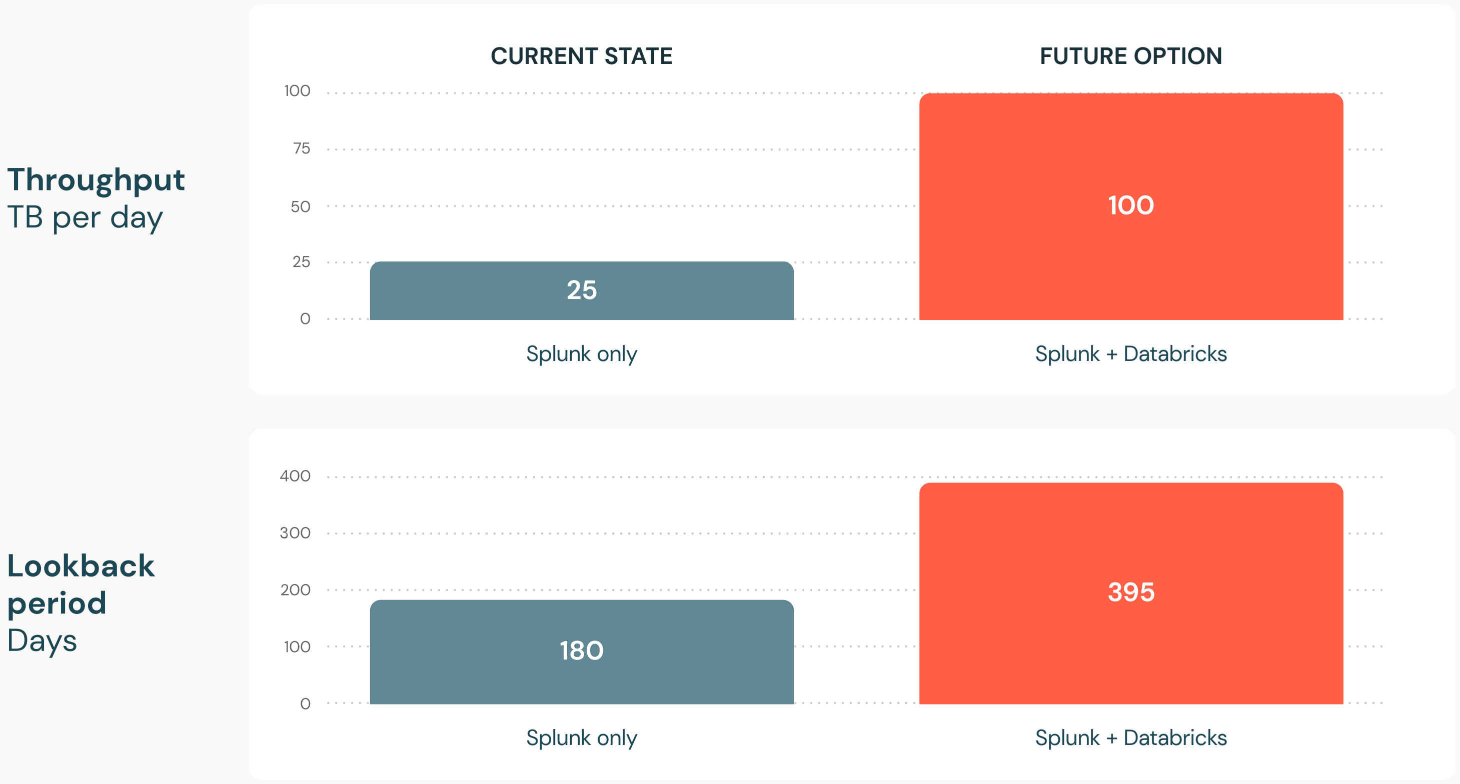


Databricks and Splunk: A Case Study in Cost-Savings

Databricks integrates with your preferred SIEM, like Splunk, and the Splunk-certified Databricks add-on can be used to meet SOC needs without changing the user interface. This example features a global financial institution's security operation, where the organization grew throughput from 25TB per day with only 180 days lookback, to 100TB per day with 395 days lookback using the Databricks SIEM augmentation. The total cost of ownership savings, including infrastructure and license costs, saved tens of millions (more than $80mn per year) in cloud costs.

FinServ Security Operations

Databricks + Splunk **Drastically** Lowered Costs



TCO savings with Splunk and Databricks vs. Splunk only solution: $81M

Common Use Cases

As FSIs focus on modernizing their data analytics and warehousing capabilities, the Databricks Lakehouse Platform brings a new level of empowerment to FSIs, allowing them to unlock the full potential of their data to deliver on their objectives and better serve their customers.

Common use cases include:

- **Threat hunting:** Empower security teams to proactively detect and discover advanced threats using months or years of data
- **Incident investigation:** Gain complete visibility across network, endpoint, cloud and application data to respond to incidents
- **Phishing threat detection:** Uncover social engineering attacks that are often used to steal user data, including log-in credentials and credit card numbers
- **Supply chain monitoring:** Leverage ML to identify suspicious behavior within your software supply chain
- **Ransomware detection:** Scope the impact and spread of ransomware attacks to inform complete mitigation and remediation
- **Credentials-abuse detection:** Identify and investigate anomalous credential usage across your infrastructure
- **Insider-threats detection:** Find and respond to malicious threats from people within an organization who have inside information about security practices, data and computer systems
- **Network traffic analysis:** Examine real-time network availability and activity to identify anomalies, vulnerabilities and malware
- **Analytics automation:** Automatically contextualize and enrich multiple streaming and batch analytics to accelerate analyst workflows and decision-making
- **Augmenting anti-money laundering practices (AML):** Using structured and unstructured data to maintain a list of politically exposed individuals, often referred to as PEP, to augment a bank's AML processes. This includes pulling data from an organization externally (keeping the PEP list up-to-date including out-of-country officials and diplomats) as well as internally (including critical personnel, network admins, etc.) who need extra scrutiny.

Getting Started With Databricks for Cybersecurity

Getting up and running on Databricks to address your cybersecurity needs is easy with our Solution Accelerators. Databricks Solution Accelerators are highly optimized, fully functional analytics solutions that provide customers with a fast start to solving their data problems.

- Cybersecurity analytics and AI at scale with Splunk and Databricks: Rapidly detect threats, investigate the impact and reduce risks with the Databricks add-on for Splunk
- Threat detection at scale with DNS analytics: Recognize cybercriminals using DNS, threat intelligence feeds and ML

Databricks Solution Accelerators are free. Join the hundreds of Databricks customers using Solution Accelerators to drive better outcomes in their businesses.

If you'd like to learn more about how we are helping financial services institutions securely leverage data and AI, please visit us at dbricks.co/fiserv or reach out to us at cybersecurity@databricks.com.

About Databricks

Databricks is the data and AI company. More than 7,000 organizations worldwide — including Comcast, Condé Nast, Acosta and over 40% of the Fortune 500 — rely on the Databricks Lakehouse Platform to unify their data, analytics and AI. Databricks is headquartered in San Francisco, with offices around the globe. Founded by the original creators of Apache Spark,™ Delta Lake and MLflow, Databricks is on a mission to help data teams solve the world's toughest problems. To learn more, follow Databricks on Twitter, LinkedIn and Facebook.

Get started with a free trial of Databricks and start building data applications today

START YOUR FREE TRIAL

To learn more, visit us at:

dbricks.com/fiserv

databricks

© Databricks 2022. All rights reserved. Apache, Apache Spark, Spark and the Spark logo are trademarks of the Apache Software Foundation. Privacy Policy | Terms of Use