Whitepaper

DORA, Operational Resilience & Beyond

Version 1.0

Author
Abhi Arikapudi
Sr. Director, Security Engineering, Databricks

Table of Contents

This page is intentionally left blank.

Introduction

The Digital Operational Resilience Act (DORA), Regulation (EU) 2022/2554, is a landmark European Union regulation designed to enhance the cybersecurity and operational resilience of financial institutions and their critical third-party providers. Set to come into effect on January 17, 2025, DORA introduces a comprehensive framework for managing Information and Communication Technology (ICT) risks across the financial sector. As financial institutions increasingly rely on data-driven technologies and artificial intelligence (AI) platforms, DORA's requirements will significantly impact how these entities manage ICT risks, facilitate business continuity, and maintain compliance.

Overview of DORA

DORA was introduced as part of the EU's broader Digital Finance Package to address the growing reliance on digital technologies in the financial sector. The regulation aims to ensure that financial institutions can withstand, respond to, and recover from ICT-related disruptions such as cyberattacks or system failures. The scope of DORA includes banks, insurance companies, investment firms, crypto-asset service providers, and critical third-party ICT providers such as cloud services, data centers, and AI platforms.

Key Objectives of DORA

1. ICT Risk Management: Financial entities must implement robust ICT risk management frameworks that cover risk identification, mitigation strategies, and continuous monitoring.
2. Incident Reporting: Entities are required to report significant ICT-related incidents to relevant authorities within stipulated timeframes.
3. Operational Resilience Testing: Regular testing of ICT systems, including penetration testing and disaster recovery exercises, is mandated.

4. Third-Party Risk Management: Financial institutions must ensure that their critical third-party ICT providers comply with DORA's resilience standards.
5. Information Sharing: Encourages collaboration between financial entities on cyber threat intelligence to improve collective resilience.

When do organizations need to be compliant with DORA?

- Financial entities must comply by January 17, 2025.
- Critical ICT third-party service providers will need to comply with the Oversight Framework within one month after their designation by European Supervisory Authorities (ESAs).

How is Databricks approaching DORA?

Databricks has implemented a comprehensive approach to facilitate compliance with DORA’s requirements. This includes:

- Security by Design: Security is embedded into every layer of the Databricks platform. We follow best practices in secure software development, vulnerability management, and penetration testing.
- Enhanced Security and Compliance: Customers can leverage enhanced security options such as serverless compute workloads protected by multiple layers of isolation.
- Compliance Certifications: Databricks holds multiple certifications relevant to regulated industries, including ISO 27001, SOC 2 Type II, and others. These certifications demonstrate our commitment to meeting stringent regulatory requirements.
- Operational Resilience: Our platform includes features such as disaster recovery (DR) plans and business continuity strategies that align with DORA’s focus on operational resilience.

We will continue to update our internal processes and contractual documentation for alignment with DORA's evolving requirements.

Is Databricks a Critical ICT Provider?

Critical ICT third-party service providers are designated by the European Supervisory Authorities (ESAs) after the January 17, 2025 effective date. In the event Databricks is designated as a critical ICT third-party service provider under DORA, our processes will be updated for compliance with the Oversight Framework by the ESA.

Databricks' Key Focus Areas for DORA

Digital Resilience

Databricks provides a robust Disaster Recovery (DR) framework for business continuity in the event of natural or human-caused disruptions. This framework is designed to recover critical infrastructure and minimize downtime for data analytics and AI workloads. Below are the essentials components of Databricks' DR capabilities that customers can leverage, based on their specific needs:

Replication and Infrastructure Setup

Databricks emphasizes the importance of replicating critical assets such as notebooks, job configurations, and cluster definitions across regions. Key strategies include:

- **Terraform Integration**: Infrastructure as code (IaC) using Terraform enables that environments can be recreated in a secondary region.
- **Data Replication**: Data stored in cloud storage (e.g., AWS S3, Azure ADLS, GCP) should be replicated to a disaster recovery region. For AWS users, S3 Multi-Region Access Points can be used to span data across multiple regions.
- **CI/CD Pipelines**: Continuous Integration/Continuous Deployment (CI/CD) tools should push updates to both primary and secondary regions simultaneously, ensuring consistency.

Monitoring and Detection

Early detection of failures is critical for minimizing downtime:

- **Service Status Notifications**: Databricks provides a status page for monitoring core services. Additionally, third-party tools like *DataDog* and *Dynatrace* can monitor infrastructure health.
- **Health Checks**: Both shallow and deep health checks monitor system performance. Shallow checks assess basic functionality (e.g., cluster uptime), while deep checks analyze system metrics like CPU usage and Spark performance.

Failover Process

In the event of a failure, Databricks supports two failover scenarios:

- **Temporary DR Site**: The DR site is used temporarily until the primary site is restored.
- **Permanent DR Site**: The DR site is promoted to the new primary site, eliminating the need for failback.
  Failover involves shutting down services at the primary site to prevent data corruption and activating clusters, jobs, and schedules at the DR site.

Incident Management

Databricks has developed a comprehensive Incident Management program designed to respond efficiently to security threats, incidents, and investigations, enabling the protection of customer data, employee information, and enterprise systems. The program is built on a proactive approach to incident detection, response, and resolution, leveraging Databricks' own platform for real-time analytics and forensics.

Proactive Monitoring & Detection

Databricks employs advanced monitoring tools and techniques to detect potential security incidents before they escalate. The platform processes over 10 Terabytes of data daily using structured streaming pipelines and machine learning models to identify unusual activity or security threats. This proactive approach facilitates that incidents are detected early, allowing for quicker response times.

Automated Response & Forensics

Databricks leverages its own platform for automating parts of the incident response process. By using Databricks as a Security Information and Event Management (SIEM) system, the team can automate log analysis, alert triage, and forensic investigations. This automation enhances the speed and accuracy of responses while reducing manual effort.

Incident Resolution & Postmortems

Once an incident is detected, the IR team immediately triages the event by correlating logs from multiple sources to investigate the root cause. The team operates on a 24/7 schedule to provide continuous coverage. After resolving incidents, Databricks conducts thorough postmortems to analyze the event and derive actionable insights that can improve future responses.

Continuous Improvement & Training

Databricks emphasizes continuous improvement in its Incident Management processes by regularly updating its playbooks based on evolving threats and lessons learned from past incidents. The IR team also collaborates with internal stakeholders to refine detection mechanisms and introduce new security automation capabilities.

Subcontracting

Databricks uses subcontractors, such as cloud service providers, to deliver certain services more effectively. Databricks is updating its customer and subcontractor contracts for requirements for subcontracting ICT services supporting critical or important functions as provided in the Regulatory Technical Standard under DORA'..

Internal access

Databricks applies strict policies and controls to internal employee access to our production systems, customer environments and customer data.

Cloud Console Access

We require multifactor authentication to access core infrastructure consoles such as the cloud service provider consoles (AWS, GCP and Azure). Databricks has policies and procedures to avoid the use of explicit credentials, such as passwords or API keys, wherever possible. For example, only appointed security team members can process exception requests for new AWS IAM principals or policies.

Separation of Duties

Our internal security standards call for the separation of duties wherever possible. For example, we centralize our cloud identity provider's authentication and authorization process to separate authorizing access from granting access.

Least Privilege

We prioritize least privilege access, both in internal systems and for our access to production systems. Least privilege is explicitly built into our internal policies and reflected in our procedures. For example, most customers can control whether Databricks employees have access to their workspace, and we programmatically apply numerous checks before access can be granted and automatically revoke access after a limited time.

Security and Risk Management

Databricks provides a comprehensive suite of security and risk management features designed to protect data, enable compliance, and maintain operational resilience across its cloud-based data analytics platform.

Below is a detailed summary of the key security features and risk management practices offered by Databricks.

Data Encryption and Customer-Managed Keys

Databricks enables that all data is encrypted both at rest and in transit, using industry-standard encryption protocols such as AES-256 and TLS 1.2+. To provide customers with greater control over their data, Databricks offers Customer-Managed Keys (CMK). This feature allows customers to manage their own encryption keys through integration with cloud-native key management services (e.g., AWS KMS, Azure Key Vault, Google Cloud KMS). By managing their own keys, customers can enforce stricter access controls and meet specific regulatory requirements for data protection.

Network Security and Private Connectivity

Databricks provides robust network security features to safeguard communication between customer environments and Databricks workspaces. One of the key offerings is Private Link, which enables secure, private connections between customer networks and Databricks without exposing traffic to the public internet. This reduces the attack surface and enhances security for sensitive workloads.

In addition, Databricks offers Network-level controls capabilities including firewall rules, IP whitelisting, and virtual network (VNet) peering to restrict access to authorized users only.

Enhanced Security and Compliance

Databricks offers an Enhanced Security and Compliance package tailored for organizations with stringent regulatory requirements. This package includes advanced security features such as:

- **Audit Logging**: Tracks all user activities for compliance reporting.
- **Compliance Certifications**: Supports regulatory requirements like GDPR, HIPAA, SOC 2 Type II.

Unified Security for Data and AI Governance

Databricks provides a unified approach to security governance across both data analytics and AI workloads. With integrated data governance tools like Unity Catalog, organizations can enforce consistent policies across all their datasets while maintaining visibility into how data is accessed and used. This is crucial for meeting complex regulatory requirements like financial services. Please see here for additional details regarding Unity Catalog.

Security Best Practices & Deployment Tools

- **Security Best Practices Whitepaper**: To help customers deploy secure environments quickly and efficiently, Databricks has developed a set of best practices (AWS, Azure and GCP), based on years of experience working with enterprise clients. These best practices are encapsulated in a comprehensive whitepaper that outlines guidelines for securing Databricks deployments.

- **Terraform Templates**: Additionally, Databricks offers pre-configured templates through its Security Reference Architecture (SRA) using Terraform, enabling organizations to automate the deployment of secure workspaces across AWS, Azure, or Google Cloud.

Monitoring & Risk Management Tools

Databricks offers a powerful tool called the Security Analysis Tool (SAT) that continuously monitors workspaces against security best practices. SAT uses API calls to programmatically verify that deployments adhere to recommended configurations and reports any deviations by severity level. This proactive monitoring helps organizations identify potential risks early and take corrective actions before they escalate.

Please see here for additional information regarding Databricks SAT.

Shared Responsibility Model

Databricks' Shared Responsibility Model for Security and Resilience divides configuration/enforcement responsibilities between Databricks and its customers.

Databricks' Responsibilities

Databricks is responsible for securing the infrastructure and Platform Services that it provides. This includes:

- **Platform Security:** Databricks enables the platform itself to be secure by implementing industry-standard security frameworks such as NIST 800-53 and ISO 27001. The company manages the security of the control plane, which governs the Platform Services, and facilitates that the infrastructure is hardened with proper security measures like firewalls, encryption, and access controls.
- **Data Encryption:** Databricks handles encryption for data in transit and at rest, ensuring that customer data is protected while being processed on its platform.
- **Compliance and Certifications:** Databricks undergoes regular third-party audits and maintains certifications such as SOC 1 Type II, SOC 2 Type II, ISO 27001/17/18, HIPAA, and PCI DSS, ensuring that its security practices meet stringent regulatory requirements.

- **Monitoring & Incident Response:** Databricks monitors its systems for unauthorized access and employs intrusion detection systems. It also has a Security Incident Response Team (SIRT) to manage security breaches.
- **Business Continuity & Disaster Recovery:** Databricks enables platform resilience through business continuity and disaster recovery plans.

Customer Responsibilities

Customers are responsible for securing their own data and configurations within the Databricks environment. This includes:

- **Data Management & Encryption:** Customers must configure encryption for their own data if they are using customer-managed services (e.g., AWS S3). While Databricks provides tools to encrypt data in transit and at rest, customers must ensure these configurations align with their internal policies.
- **Access Control & Identity Management:** Customers are responsible for managing user access to their Databricks workspaces. This includes configuring role-based access control (RBAC) and integrating identity management systems like Azure Active Directory (AAD) to ensure proper permissions are applied.
- **Configuration & Compliance:** Customers need to evaluate and correctly configure the security features available on the platform to meet their risk profiles. This includes setting up secure network configurations (e.g., VPC peering, IP access lists) and ensuring compliance with internal policies or applicable regulatory requirements.
- **Backups & Data Resilience:** While Databricks provides tools for synchronizing notebooks with external repositories like GitHub, customers are responsible for backing up their own data stored within their cloud environments.

Please see additional details here: AWS, Azure, or GCP.

Penetration Testing & Bug Bounty

Databricks performs penetration testing through a combination of its in-house offensive security team, qualified third-party penetration testers and a year-round public bug bounty program. We use a mixture of fuzzing, secure code review and dynamic application testing to evaluate the integrity of our platform and the security of our application. We conduct penetration tests on major releases, new services and security-sensitive features. The offensive security team works with our incident response team and security champions within engineering to resolve findings and infuse learnings throughout the company.

We typically perform 8-10 external third-party penetration tests and 15-20 internal penetration tests per year, and all material findings must be addressed before a test can be marked as passed. As part of our commitment to transparency, we publicly share our platform-wide third-party test report in our due diligence package.

Bug Bounty

Our public bug bounty program, facilitated by *HackerOne*, allows a global collective of cybersecurity researchers and penetration testers to test Databricks for security vulnerabilities. Some of the key decisions we've made to make the program successful include:

- Encouraging an engaged community of hackers to be active on our program by providing transparency to our HackerOne program statistics such as response rate and payouts
- Promptly responding to bug bounty submissions, with an average time-to-bounty under a week
- Performing variant analysis on every valid submission to identify alternative ways that an exploit may be used, and verifying 100% of fixes
- Adding bonuses that drive attention to the most important areas of the product

Customer Penetration and Vulnerability Testing

We want our customers to have confidence in the workloads they run on Databricks. If your team would like to run a vulnerability scan or penetration test against Databricks, we encourage you to:

- Run vulnerability scans on data plane systems located inside of your cloud service provider account.
- Run tests against your code, provided that those tests are entirely contained within the data plane (or other systems) located in your cloud service provider account and are evaluating your controls.
- Join the Databricks Bug Bounty program to access a dedicated deployment of Databricks to perform penetration tests. Any penetration test against our multi-tenant control plane requires participation in the program.

Customers can also conduct their own penetration tests under agreed terms.

Due Diligence and Audit

Customers can perform their own due diligence on Databricks using our self-service security review package. This package includes documentation on our compliance certifications and security controls. Please see Databricks Security & Trust Center for more details.

In addition, Databricks customers have the right to request an audit of Databricks' controls as per agreed-upon terms, if they are unable to meet Databricks' compliance requirements through our independent audit reports.

Databricks also offers several other security features that allow customers to perform their own independent due diligence, including:

- **Auditing**: Admins can monitor user activity to detect security issues, such as unusual login times or simultaneous remote logins.

- **Compliance Security Profile (CSP)**: This baseline for the data plane helps customers meet compliance requirements for HIPAA, PCI-DSS, and FedRAMP Moderate workloads.
- **Access Control**: Security privileges can be mapped to each customer's operating model.
- **Audit Logs**: Account admins can enable the audit log system table to access audit logs of user activity

Regulator Cooperation

Databricks will fully cooperate with financial services regulators in connection with supervisory activities over our financial services customers and their use of the Databricks platform as detailed in our FSA. We provide transparency into our security practices through detailed documentation available on our Trust Center.

Access, Recovery, and Return of Data

Customers retain full ownership of the Customer Content processed by Databricks on their behalf while providing the Platform Services as provided in the MCSA:

- They can access their data at any time via self-service tools provided within the platform.
- Data recovery processes are aligned with our disaster recovery plans to minimize downtime during incidents.

Conclusion

Databricks is committed to supporting financial services organizations in meeting their obligations under the Digital Operational Resilience Act (DORA). As we approach the January 2025 deadline for compliance, we will continue updating our documentation and working closely with customers.