Databricks, Inc.

System and Organization Controls (SOC) 3

Report on Databricks, Inc.'s Assertion Related to Its Data Intelligence Platform Services System on Amazon Web Services, Microsoft Azure, and Google Cloud Platform Relevant to Security, Availability, and Confidentiality

Throughout the Period
November 1, 2023 to October 31, 2024

I. Independent Service Auditor’s Report on a SOC 3 Examination

Tel: 415-397-7900
Fax: 415-397-2161
www.bdo.com

One Bush Street, Suite 1800
San Francisco, CA 94104

Independent Service Auditor’s Report on a SOC 3 Examination

To the Management of
Databricks, Inc.
San Francisco, California

Scope

We have examined Databricks, Inc.’s (Databricks or service organization) accompanying assertion titled *Assertion of Databricks, Inc. Management* (assertion) that the controls within Databricks’ Data Intelligence Platform Services System on Amazon Web Services (AWS), Microsoft Azure (Azure), and Google Cloud Platform (GCP), (the System) were effective throughout the period November 1, 2023 to October 31, 2024 to provide reasonable assurance that Databricks’ service commitments and system requirements were achieved based on the trust services criteria relevant to security, availability, and confidentiality (applicable trust services criteria) *se*t forth in TSP Section 100, *2017 Trust Services Criteria for Security, Availability, Processing Integrity, Confidentiality, and Privacy*, in AICPA *Trust Services Criteria*.

Databricks uses subservice organizations to perform certain activities. A list of these subservice organizations and the activities performed is provided in Section III. The assertion indicates that complementary subservice organization controls that are suitably designed and operating effectively are necessary, along with controls at Databricks, to achieve Databricks’ service commitments and system requirements based on the applicable trust services criteria. Databricks’ description of the boundaries of the System in Attachment A presents the types of complementary subservice organization controls assumed in the design of Databricks’ controls but does not disclose the actual controls at the subservice organizations. Our examination did not include the services provided by the subservice organizations, and we have not evaluated the suitability of the design or operating effectiveness of such complementary subservice organization controls.

Service Organization’s Responsibilities

Databricks is responsible for its service commitments and system requirements and for designing, implementing, and operating effective controls within the System to provide reasonable assurance that Databricks’ service commitments and system requirements were achieved. Databricks has also provided the accompanying assertion about the effectiveness of controls within the System. When preparing its assertion, Databricks is responsible for selecting, and identifying in its assertion, the applicable trust services criteria and for having a reasonable basis for its assertion by performing an assessment of the effectiveness of controls within the System.

Service Auditor’s Responsibilities

Our responsibility is to express an opinion, based on our examination, on whether management’s assertion that controls within the System were effective throughout the period November 1, 2023 to October 31, 2024 to provide reasonable assurance that the service organization’s service commitments and system requirements were achieved based on the applicable trust services criteria. Our examination was conducted in accordance with attestation standards established by the AICPA. Those standards require that we plan and perform our examination to obtain reasonable assurance about whether management’s assertion is fairly stated, in all material respects. We believe that the evidence we obtained is sufficient and appropriate to provide a reasonable basis for our opinion.

Our examination included:

- Obtaining an understanding of the System and the service organization's service commitments and system requirements.
- Assessing the risks that controls were not effective to achieve the service organization's service commitments and system requirements based on the applicable trust services criteria.
- Performing procedures to obtain evidence about whether controls within the System were effective to achieve Databricks' service commitments and system requirements based on the applicable trust services criteria.

Our examination also included performing such other procedures as we considered necessary in the circumstances.

We are required to be independent and to meet our other ethical responsibilities in accordance with relevant ethical requirements relating to the engagement.

Inherent Limitations

There are inherent limitations in the effectiveness of any system of internal control, including the possibility of human error and the circumvention of controls.

Because of their nature, controls may not always operate effectively to provide reasonable assurance that the service organization's service commitments and system requirements are achieved based on the applicable trust services criteria. Also, the projection to the future of any conclusions about the operating effectiveness of controls is subject to the risks that controls may become inadequate because of changes in conditions or that the degree of compliance with the policies or procedures may deteriorate.

Opinion

In our opinion, management's assertion that the controls within Databricks' system were effective throughout the period November 1, 2023 to October 31, 2024 to provide reasonable assurance that Databricks' service commitments and system requirements were achieved based on the applicable trust services criteria is fairly stated, in all material respects.

Restricted Use

This report is intended solely for the information and use of Databricks, user entities of Databricks' system during some or all of the period November 1, 2023 to October 31, 2024, business partners of Databricks subject to risks arising from interactions with the System, practitioners providing services to such user entities and business partners, prospective user entities and business partners, and regulators who have sufficient knowledge and understanding of the following:

- The nature of the service provided by the service organization.

- How the service organization’s system interacts with user entities, business partners, subservice organizations, and other parties.
- Internal control and its limitations.
- Complementary subservice organization controls and how those controls interact with the controls at the service organization to achieve the service organization’s service commitments and system requirements.
- User entity responsibilities and how they may affect the user entity’s ability to effectively use the service organization’s services.
- The applicable trust services criteria.
- The risks that may threaten the achievement of the service organization’s service commitments and system requirements, and how controls address those risks.

This report is not intended to be, and should not be, used by anyone other than these specified parties.

BDO USA, P.C.

December 20, 2024

II. Assertion of Databricks, Inc. Management

databricks

Databricks Inc.
160 Spear Street, 13th Floor
San Francisco, CA 94105
1-866-330-0121

Section II

Assertion of Databricks, Inc. Management

We are responsible for designing, implementing, operating, and maintaining effective controls within Databricks, Inc.'s (Databricks or the service organization) Data Intelligence Platform Services System on Amazon Web Services (AWS), Microsoft Azure (Azure), and Google Cloud Platform (GCP), (the System) throughout the period November 1, 2023 to October 31, 2024, to provide reasonable assurance that Databricks' service commitments and system requirements relevant to security, availability, and confidentiality (applicable trust services criteria) were achieved. Our description of the boundaries of the System is presented in Attachment A, *Databricks, Inc.'s Description of the Boundaries of Its Data Intelligence Platform Services System on Amazon Web Services, Microsoft Azure, and Google Cloud Platform* and identified the aspects of the System covered by our assertion.

Databricks uses subservice organizations to perform certain activities. A list of these subservice organizations and the activities performed is provided in Section III. The description of the boundaries of the System in Attachment A indicates that complementary subservice organization controls that are suitably designed and operating effectively are necessary, along with controls at Databricks, to achieve Databricks' service commitments and system requirements based on the applicable trust services criteria. The description of the boundaries of the System presents the types of complementary subservice organization controls assumed in the design of Databricks' controls. The description of the boundaries of the System does not extend to the actual controls at the subservice organizations.

We have performed an evaluation of the effectiveness of the controls within the System throughout the period November 1, 2023 to October 31, 2024 to provide reasonable assurance that Databricks' service commitments and system requirements were achieved based on the trust services criteria relevant to the applicable trust services criteria set forth in TSP Section 100, *2017 Trust Services Criteria for Security, Availability, Processing Integrity, Confidentiality, and Privacy*, in AICPA *Trust Services Criteria*. Databricks' objectives for the System in applying the applicable trust services criteria are embodied in its service commitments and system requirements relevant to the applicable trust services criteria. The principal service commitments and system requirements related to the applicable trust services criteria are presented in Attachment B.

There are inherent limitations in any system of internal control, including the possibility of human error and the circumvention of controls. Because of these inherent limitations, a service organization may achieve reasonable, but not absolute, assurance that its service commitments and system requirements were achieved.

We assert that the controls within the System were effective throughout the period November 1, 2023 to October 31, 2024 to provide reasonable assurance that Databricks' service commitments and system requirements were achieved based on the applicable trust services criteria.

Databricks, Inc.

December 20, 2024

Attachment A — Databricks, Inc.’s Description of the Boundaries of Its Data Intelligence Platform Services System on Amazon Web Services, Microsoft Azure, and Google Cloud Platform

Attachment A — Databricks, Inc.’s Description of the Boundaries of Its Data Intelligence Platform Services System on Amazon Web Services, Microsoft Azure, and Google Cloud Platform

Scope and Description of the Boundaries of the System

This is a System and Organization Controls (SOC) 3 report and includes a description of the boundaries of Databricks, Inc.’s (Databricks, service organization, or Company) Data Intelligence Platform Services System on Amazon Web Services, Microsoft Azure, and Google Cloud Platform (the System) and the controls in place to meet the criteria for security, availability, and confidentiality (applicable trust services criteria) set forth in TSP Section 100, *2017 Trust Services Criteria for Security, Availability, Processing Integrity, Confidentiality, and Privacy*, in AICPA *Trust Services Criteria*, throughout the period November 1, 2023 to October 31, 2024 (the period) which may be relevant to the users of the System. It does not encompass all aspects of the services provided or procedures followed for other activities performed by Databricks.

Databricks uses subservice organizations to perform certain services. A list of these subservice organizations and the services performed is provided in the following table. The description does not disclose the actual controls at the subservice organizations.

Subservice Organization	Services Performed
Amazon Web Services, Inc. (Amazon)	Provides managed services over infrastructure, logging, and key management services.
Microsoft Corporation (Microsoft)	Provides managed services over infrastructure, logging, monitoring, vulnerability scan, and key management services.
Google LLC (Google)	Provides managed services over infrastructure, logging, and key management services.
Okta, Inc. (Okta)	Single sign-on (SSO) and authentication services.

Company Background

Databricks, Inc.’s founders started the Spark research project at UC Berkeley, which later became Apache Spark™. Databricks was founded in 2013 by the original creators of popular open-source projects, including Apache Spark, Delta Lake, MLflow, and Koalas. As the world’s first and only Lakehouse Platform (now renamed as Data Intelligence Platform) system in the cloud, Databricks is built on a lakehouse that combines the best elements of data lakes and data warehouses to offer an open, cost effective, and unified platform for data, analytic, and AI.

Databricks is on a mission to simplify and democratize data and AI, helping data teams solve the world’s toughest problems. As the leader in Unified Data Analytics, Databricks helps organizations make all their data ready for analytics, empower data science and data-driven decisions across the organization, and rapidly adopt machine learning to outpace the competition. It is simple, open, and supports multi-cloud. By providing data teams with the ability to process massive amounts of data in the cloud and power AI with that data, Databricks helps organizations innovate faster and tackle challenges like treating chronic disease through faster drug discovery, improving energy efficiency, and protecting financial markets. More than 9,000 organizations worldwide rely on

Databricks to enable massive-scale data engineering, collaborative data science, full-lifecycle machine learning, and business analytics.

Databricks is venture-backed and headquartered in San Francisco, California with offices around the world and has over 1,200+ global partners, including Microsoft, Amazon, Google, Tableau, Informatica, Qlik, Dataiku, Fivetran, Accenture, Capgemini, and Booz Allen Hamilton.

Services Provided

Databricks provides a Data Intelligence Platform Service system for massive scale data engineering and collaborative data science in AWS, including AWS GovCloud Classic, as well as in Azure and GCP. The Databricks solution offers multi-tenant services that consist of a Control Plane, Compute Plane, and Customer Data Storage:

- *Control Plane* — The portion of the Data Intelligence Platform Service system that is always hosted in the Databricks-managed AWS, Azure, or GCP environment. It contains the core Databricks services such as web application, cluster management, access control to resources, and managing running or scheduled jobs. This report covers the Control Plane.
- *Compute Plane* — This is where the extensive data processing occurs (server clusters of the compute layer). The type of data processed is chosen by the customer. As of the report date, there are two types of Compute Planes:
    - *Serverless Compute Plane* — This resides in the customer's Databricks-managed and owned AWS, Azure, or GCP account rather than the customer-owned AWS, Azure, or GCP account. AWS, Azure, and GCP Serverless Compute Plane is in scope and covered by this report.
    - *Classic Compute Plane* — This resides in the customer-owned AWS, Azure, or GCP account and is not covered by this report.
- *Customer Data Storage* — This is where customer datasets are stored at rest when they use Databricks services. As of the report date, for the AWS Serverless Compute model environment, customers may select from the below storage options:
    - *Databricks-Managed Storage* — AWS Simple Storage Service (S3) bucket resides in the Databricks AWS account and is covered by this report.
    - *Customer-Managed Storage* — AWS S3 bucket resides in the customer-managed AWS account and is not covered by this report.

For other environments, customers store the data in their managed cloud account (i.e., AWS S3 bucket, Azure Blob Storage, Google Cloud Storage [GCS] bucket) and are excluded from the scope of the description of this report.

Databricks services enable customers to:

- Easily and quickly access data at scale.
- Process both structured data and unstructured data.
- Ingest from nontraditional data stores.
- Reduce the batch processing time.

- Deploy production-quality machine learning and streaming applications.
- Set up, tune, and scale Apache Spark clusters for the team.
- Keep clusters resilient and up to date with the latest versions.
- Schedule, run, and debug applications in production.
- Leverage more data science to aid in decision-making.
- Explore and visualize data interactively.
- Connect to business intelligence tools and build real-time dashboards.

System Incidents

A system incident is an incident that leads to the loss of, or disruption to, operations, services, or functions and results in Databricks' failure to achieve its service commitments or system requirements. Such an occurrence may arise from a security event, security incident, failure to comply with applicable laws and regulations, error, or other means. In determining whether a system incident occurred resulting in Databricks' failure to achieve one or more of its service commitments or system requirements, considerations may include, but are not limited to, the following:

- Whether the occurrence resulted from one or more controls that were not suitably designed or operating effectively.
- Whether public disclosure of the occurrence was required (or is likely to be required) by cybersecurity laws or regulations.
- Whether the occurrence had a material effect on the service organization's financial position or results of operations and required disclosure in a financial statement filing.
- Whether the occurrence resulted in sanctions by any legal or regulatory agency.
- Whether the occurrence resulted in the service organization's withdrawal from material markets or cancellation of material contracts.

Incidents and events relevant to Databricks' service commitments and system requirements based on the applicable trust services criteria are important in monitoring, identifying, and evaluating if a system incident has occurred; however, incidents and events relevant to Databricks' service commitments and system requirements based on the applicable trust services criteria do not always rise to the level of a system incident. The evaluation of an incident or event relevant to Databricks' service commitments and system requirements based on the applicable trust services criteria will make that determination.

Databricks did not identify any system incidents that occurred during the period November 1, 2023 to October 31, 2024 resulting in Databricks' failure to achieve one or more of its service commitments or system requirements based on these considerations.

Components of the System Used to Provide the Services

Infrastructure

Databricks is headquartered in San Francisco. Databricks does not operate its own data centers; instead, it has contracted with subservice organizations AWS, Azure, and GCP in support of its Data Intelligence Platform services to provide supporting infrastructure, logging, and key management services. Refer to section above titled *Scope and Boundaries of the System* for additional information on the service provided by AWS, Azure, and GCP and the associated complementary subservice organization controls.

Primary infrastructure used to provide the Databricks-hosted Data Intelligence Platform Services system on AWS, Azure, and GCP includes, but is not limited to, the following:

Primary Infrastructure		
Hardware	**Type**	**Purpose**
Amazon Virtual Private Cloud (VPC)	Virtual Network	To provide network-level isolation between major areas of the infrastructure.
Amazon Elastic Compute Cloud (EC2)	Virtual Machines	Virtualized servers for hosting the web application and its supporting services.
Amazon Relational Database Service (RDS)	Database	To store customer metadata (such as notebook data, login credentials, and job definitions).
Amazon Elastic Block Store (EBS)	Block Storage	To store the operating systems, applications (as well as their configs), and any temporary Spark data (such as shuffle data).
Amazon Simple Storage Service (S3)	Object Storage	To provide general data storage that is primarily used for storing control plane artifacts for purposes of bootstrapping services.
AWS Key Management Service (KMS)	Cryptography	To provide cryptographic key management such as generation, distribution, and encryption for many components in the infrastructure.
AWS Identity and Access Management (IAM)	Identify and Access Management	To securely manage access to AWS services and resources.
Azure Resource Manager	Management Layer	To manage resources (e.g., create, update, and delete) inside Azure deployment such as virtual machine, storage account, web app, database, and virtual network.
Azure Virtual Network (VNET)	Virtual Network	To provide network-level isolation between major areas of the infrastructure.
Azure Virtual Machines (VMs)	Virtual Machines	Virtualized servers for hosting the web application and its supporting services.
Azure DB Database for MySQL	Database	To store customer metadata (such as notebook data, login credentials, and job definitions).

Primary Infrastructure		
Hardware	**Type**	**Purpose**
Azure Blob Storage	Object Storage	To provide a generic blob store that is primarily used for storing control plane artifacts for purposes of bootstrapping services.
Azure Key Vault (AKV)	Cryptography	To provide cryptographic key management such as generation, distribution, and encryption for many components in the infrastructure.
GCP Virtual Private Cloud (VPC)	Virtual Network	To provide network-level isolation between major areas of the infrastructure.
GCP Compute Engine	Virtual Machines	Virtualized servers for hosting the web application and its supporting services.
GCP Database	Database	To store customer metadata (such as notebook data, login credentials, and job definitions).
GCP Persistent Disk	Block Storage	To store the operating systems, applications (as well as their configs), and any temporary Spark data (such as shuffle data).
Google Cloud Storage (GCS)	Object Storage	To provide general data storage that is primarily used for storing control plane artifacts for purposes of bootstrapping services.
Google Cloud Key Management	Cryptography	To provide cryptographic key management such as generation, distribution, and encryption for many components in the infrastructure.
GCP Identity and Access Management (IAM)	Identify and Access Management	To securely manage access to GCP services and resources.

Software

Operating System and Database

Databricks uses Linux operating systems and SQL databases in AWS, Azure, and GCP environments.

Supporting Systems and Applications

Primary software tools and applications used to support Databricks' Data Intelligence Platform Services system on AWS, Azure, and GCP and operations are outlined below:

Primary Software and Application	
Software/Application	**Purpose**
People Operations	
Workday	Workday is a human capital management software. Workday is the source of truth for managing Databricks' employee and contractor accounts, payroll, and benefits. Workday is integrated with Okta.

Primary Software and Application	
Software/Application	**Purpose**
Change Management	
GitHub	GitHub is a centralized source code control system. It is implemented internally for the management of code repositories.
Jenkins/Deployment Service	Jenkins/Deployment Service is a tool that helps to automate the software development process with continuous integration and facilitating technical aspects of continuous delivery.
Spinnaker	Spinnaker is a continuous delivery platform that orchestrates the continuous delivery pipelines that include many stages throughout the release cycle.
Atlassian Jira	Jira is a ticketing system to document and track changes, including application changes, infrastructure changes, operational changes, etc.
FreshService	FreshService is a ticketing system used by the Corporate Engineering team and the People Operations team to provide and deliver employee support.
Customer Support	
Salesforce	Salesforce is the ticketing system used for customer support.
Identity and Access Management	
Okta	Okta is an SSO solution that manages and secures user authentication into modern applications. It provides a universal cloud-based platform to manage and secure identities.
YubiKey	YubiKey is a multifactor authentication (MFA) tool that must be used by Databricks personnel.
GlobalProtect VPN	GlobalProtect VPN is the virtual private network (VPN) solution provided by Palo Alto Networks. It is used to control remote access to the production systems via dedicated and on-demand encrypted tunnels.
Zscaler	Zscaler is a VPN solution that is used to control remote access to the production systems via dedicated and on-demand encrypted tunnels.
Genie	Genie is an in-house application that Databricks uses for managing and logging temporary access to production infrastructure.
Teleport	Teleport is an application that Databricks uses for managing and logging temporary access to production infrastructure.
Security and Performance Logging and Monitoring	
PagerDuty	PagerDuty is a software-as-a-service (SaaS) incident response platform that integrates with other applications and handles alert notifications and triaging.
M3	M3 is an open-source metrics engine that is Prometheus compatible. It provides monitoring solutions with an alerting toolkit to monitor network, machines, and applications.

Primary Software and Application	
Software/Application	**Purpose**
AWS CloudTrail	AWS CloudTrail is an AWS service that enables governance, compliance, operational auditing, and risk auditing of Databricks AWS accounts.
AWS GuardDuty	AWS GuardDuty is a threat detection service used to monitor malicious activity and unauthorized behavior to protect Databricks AWS accounts.
Google Cloud Audit	Google Cloud Audit is a GCP service that enables governance, compliance, operational auditing, and risk auditing of Databricks GCP accounts. It provides Admin Activity, Data Access, System Event, and Policy Denied audit logs for each cloud project, folder, and organization.
GCP Event Threat Detection	GCP Event Threat Detection is a threat detection service used to monitor malicious activity and unauthorized behavior to protect Databricks GCP accounts.
GCP Security Command Center	GCP Security Command Center is a security and risk management platform for Google Cloud. Google Cloud Audit and GCP Event Threat Detection services are integrated with the GCP Security Command Center, and the findings are available to view there.
Secfood	Secfood is a Databricks application instance used by the Security Team as the security information and event management (SIEM) tool.
CrowdStrike	CrowdStrike provides next-generation antivirus software for Databricks end-user endpoints that use a combination of artificial intelligence, behavioral detection, machine learning algorithms, and exploit mitigation.
Intune	Intune provides management software that enforces policies on Databricks end-user Microsoft Windows endpoints (i.e., host-based firewall, laptop disk encryption, etc.).
Jamf	Jamf provides management software that enforces policies on Databricks end-user MacOS endpoints (i.e., host-based firewall, laptop disk encryption, etc.).
Vulnerability Management	
Qualys	Qualys Cloud Platform is a third-party product used to perform infrastructure and web application vulnerability assessments.
HackerOne	HackerOne is a third-party platform used to facilitate penetration testing bug bounty programs to identify vulnerabilities on an ongoing basis.
Tenable	Tenable Security Center provides infrastructure and web application vulnerability assessments for the AWS GovCloud environment.
Configuration Management	
AWS CloudFormation	AWS CloudFormation is an infrastructure-as-code service to deploy cloud resources in AWS.

Primary Software and Application	
Software/Application	**Purpose**
Terraform	Terraform is an open-source, infrastructure-as-code, software tool to deploy cloud resources in Azure and GCP.

People

The Databricks staff provides support for the above services in each of the following functional areas:

Functional Areas	Responsibilities
Executive Management	• Develops and executes on the Company's strategy and oversees operations to achieve business objectives. • The executive team includes: Chief Executive Officer (CEO), Global Field Operations, Chief Financial Officer (CFO), Chief Security Officer (CSO), Senior Vice President of Engineering, Chief Operations Officer (COO), Chief Information Officer (CIO), Chief People Officer (CPO), Senior Vice President of Products, Senior Vice President & General Counsel.
Information Security	• Led by the CSO to meet security and compliance requirements. • Security policies and procedures documentation and review. • Overall security of the platform, maintaining compliance with ISO 27001, ISO 27018, ISO 27017, ISO 27701, HITRUST, SOC 1, SOC 2, HIPAA, PCI-DSS, and various industry standards. • Security architecture, engineering, and operations. • Security monitoring and incident response. • Vulnerability scanning and penetration testing. • Organizational security awareness and training. • Security risk management. • Vendor security compliance.

Functional Areas	Responsibilities
Engineering	• Managing, monitoring, and supporting the cloud infrastructure required to run the product. • Responsible for day-to-day maintenance of AWS, Azure, and GCP cloud systems security, availability, and confidentiality, as applicable. • Managing, monitoring, and supporting the cloud services required to support the product (i.e., Genie, Teleport, Secret Management, etc.). • Managing base images. • Building services to ensure Databricks' infrastructure runs on a secure and performant base operating system. • Overseeing the development of the product and release management. • Developing new features and fixing bugs and vulnerabilities. • Monitoring and handling availability and capacity alerts.
Field Engineering	• Performing pre- and post-sales engineering. • Working with potential customers to try out Databricks' products and solutions.
Product Management	• Working closely with Engineering teams to ensure product improvements are tracked, implemented, and released in a timely manner. • Working closely with customers to get feedback and gather requirements for product improvements.
Customer Success	• Single point of contact for handling and directing customer issues. • Day-to-day customer support.
Sales Operations	• Customer sales onboarding, renewal, and account management. • Working with field engineering to provide training and professional services to customers.
Legal	• Managing contractual agreements with customers and third parties. • Addressing Databricks' legal concerns and handling legal-related external matters.
People Operations	• Handling talent acquisition and employee benefits. • Reviewing and updating the Employee Handbook and Databricks Code of Conduct. • Handling employee termination and internal investigations. • Managing performance enablement reviews and focal programs.
IT	• Managing IT systems, applications, and office networks. • Managing employee onboarding and offboarding such as laptop setup, new hire account creation, laptop collection and disposal, applications assignment in Okta, and account removal from Okta.

Functional Areas	Responsibilities
Facilities	• Managing Databricks office locations and office physical security controls (for example, door access control, security camera, space management, building maintenance).
Professional Services	• Delivering customer projects from ideation to production. • Providing industry-leading expertise on Spark, machine learning, cloud, streaming, data science, and engineering. • Empowering and enabling customers with the technical skills needed to deploy, maintain, and enhance their platforms.

Processes and Procedures

Databricks has put into place a set of policies and procedures designed to provide requirements and guidance for management and employees regarding security, availability, and confidentiality. Databricks reviews those policies annually and makes them available to all Databricks personnel. Teams are expected to implement procedures that define how services are delivered. Exceptions to Databricks' policies are documented, tracked, and reviewed. Databricks personnel that fail to comply with Databricks' security policies are subject to a disciplinary process.

Databricks' security policies cover, but are not limited to, the following domains:

- Anti-Malware
- Application Development
- Asset Management
- Backup
- Business Continuity and Disaster Recovery
- Change Management
- Cryptography
- Data Classification and Handling
- Data Collection, Storage, and Use
- Data Retention and Destruction
- Incident Response Management
- Logging and Monitoring
- Network Security
- Personnel Management
- Remote Access and Teleworking
- Removable Media
- Risk Assessment
- Security Governance and Compliance
- Third Party Management
- Training and Awareness
- Vulnerability Management

Data

Databricks handles a variety of sensitive data, which, depending on contract, may include personal identifiable information (PII) and protected health information (PHI). Therefore, Databricks has implemented security controls throughout the data lifecycle.

Collection

The customer provides credentials to connect to the data sources. At runtime, the clusters access these data sources using temporary authorizations that were generated from customer-provided credentials. Clusters access customer data during processing. Communications between the customer and Databricks services always use Hypertext Transfer Protocol Secure (HTTPS).

Handling and Storage

Databricks SQL or instructions within notebooks are stored in databases that reside in the Databricks-managed Control Plane. Databases in the Databricks-managed Control Plane are encrypted at rest with 256-bit keys. Encryption and retention of the data stored in customer-controlled data sources (e.g., AWS S3 bucket, Azure Blob Storage, GCS bucket, etc.) are the customer's responsibility.

Transmission

Data in transit between the Control Plane (Databricks-managed AWS/Azure/GCP account) and the Compute Plane is encrypted using the Transport Layer Security (TLS) 1.2 or 1.3 cipher suite. Notebook source data submitted from users of the Databricks web application to the Control Plane is encrypted in transit with TLS 1.2 or 1.3. Customers are responsible for using encrypted connections to their data sources. Note that for Azure Databricks, communication between the Control Plane and the customer Compute Plane traverses the Azure network backbone, not across the public internet.

Modification

Data is protected from modification through encryption and access controls.

Release/Disclosure

Databricks will not release or disclose customer data to a third party unless required by applicable law.

Processing

With the Databricks Serverless Compute platform, data is processed by clusters residing in Databricks' cloud account. With the Databricks Classic Compute platform, data is processed by clusters residing in the customer's AWS, Azure, or GCP account.

Use

Databricks will use customers' data only to perform activities contractually agreed to. Datasets are stored and managed by the customer unless Databricks-managed storage service is used. Databricks SQL or instructions within Databricks Notebooks are retained for a time specified according to the customer's contract. After that time, such data is deleted.

Complementary Subservice Organization Controls

In some instances, a service organization's controls cannot provide reasonable assurance that its service commitments and system requirements were achieved without the subservice organizations performing certain activities in a defined manner. Such activities are referred to as complementary subservice organization controls (CSOCs). The following CSOCs are those controls that Databricks' management assumed, in the design of the System, would be implemented by a subservice organization and are necessary, in combination with controls at Databricks, to provide reasonable assurance that the service organization’s service commitments and system requirements are achieved.

Number	CSOC	Applicable Trust Services Criteria and HIPAA Security and Breach Notification Requirements
Amazon Web Services, Inc., Microsoft Corporation, Google LLC, and Okta, Inc.		
1.	Physical access to facilities housing the production servers and other system components is restricted to authorized personnel only.	CC6.4, §164.310(a)(1), §164.310(a)(2)(ii), §164.310(a)(2)(iii), §164.310(a)(2)(iv)
2.	Physical access to data centers is approved by an authorized individual.	
3.	Physical access is revoked in a timely manner of the employee or vendor record being deactivated.	
4.	Physical access to data centers is reviewed on a quarterly basis by appropriate personnel.	
5.	Logs and maintenance records for physical access to data centers are retained for a defined period.	
6.	Physical access points to server locations are recorded by closed-circuit television cameras. Images are retained for a defined period, unless limited by legal or contractual obligations.	
7.	Physical access points to server locations are managed by electronic access control devices.	
8.	The entity implements logical access policies, security software, infrastructure, and architectures over protected information assets to protect them from unauthorized access and security events.	CC6.1, CC6.6, §164.312(a)(1), §164.312(a)(2)(i) §164.312(a)(2)(iii), §164.312(a)(2)(iv), §164.312(b), §164.312(c)(2), §164.312(d), §164.312(e)(1), §164.312(e)(2)(ii)
9.	Prior to issuing system credentials and granting system access, the entity registers and authorizes new internal and external users whose access is administered by the entity. For those users whose access is administered by the entity, user system credentials are removed when user access is no longer authorized.	CC6.2, §164.308(a)(3)(ii)(C), §164.308(a)(4)(ii)(B) §164.308(a)(5)(ii)(D), §164.312(a)(1), §164.312(a)(2)(i), §164.312(a)(2)(iii)

Number	CSOC	Applicable Trust Services Criteria and HIPAA Security and Breach Notification Requirements
10.	Implement procedures for the authorization and/or supervision of workforce members who work with electronic protected health information or in locations where it might be accessed.	§164.308(a)(3)(ii)(A)
11.	Industry standard encryption algorithms are used to manage production infrastructure.	CC6.1, CC6.7, §164.312(a)(2)(iv), §164.312(e)(2)(ii)
12.	External and internal security vulnerability assessment and penetration testing is performed on an annual basis.	CC7.2, §164.308(a)(1)(ii)(A), §164.308(a)(1)(ii)(B)
13.	Procedures for evaluating security events and managing security incidents are implemented. Security incidents are communicated as appropriate.	CC7.3, CC7.4, CC7.5, §164.308(a)(1)(ii)(D), §164.308(a)(6), §164.308(a)(6)(ii)
14.	Changes (including emergency/nonroutine) to production infrastructure and applications are recorded, authorized, tested, and approved prior to migration.	CC8.1, §164.312(e)(2)(i)
15.	Environmental protections of systems at the data centers are designed, developed, implemented, operated, maintained, and monitored to meet availability commitments and requirements.	A1.2, A1.3, §164.308(a)(7), §164.308(a)(7)(ii)(A), §164.308(a)(7)(ii)(B), §164.308(a)(7)(ii)(C), §164.308(a)(7)(ii)(D)
16.	Data centers are protected by fire detection and suppression systems.	
17.	Data centers are air-conditioned to maintain appropriate atmospheric conditions. Personnel and systems monitor and control air temperature and humidity at appropriate levels.	
18.	Data centers' electrical power systems are designed to be fully redundant and maintainable without impact to operations, and Uninterruptible Power Supply units provide backup power in the event of an electrical failure for critical and essential loads in the facility.	
19.	Data centers have generators to provide backup power in case of electrical failure.	
20.	Preventative maintenance is performed on environmental protections on at least an annual basis.	
21.	Data centers' contingency plan (e.g., disaster recovery plan, business continuity plan, etc.) is in place. The contingency plan is reviewed and tested on at least an annual basis.	

Number	CSOC	Applicable Trust Services Criteria and HIPAA Security and Breach Notification Requirements
22.	Access to encryption keys and recovery key materials is appropriately restricted.	CC6.1, CC6.7, §164.312(a)(1), §164.308(a)(3)(ii)(A), §164.312(a)(2)(iv), §164.312(e)(2)(ii)
Microsoft Corporation and Google LLC		
23.	Databases are encrypted, securely backed up, and stored in separate data center facilities.	CC6.1, CC6.6, CC7.5, A1.2, C1.1, §164.308(a)(7)(ii)(A) §164.310(a)(2)(i) §164.310(d)(2)(iv) §164.312(a)(2)(iv), §164.312(c)(1), §164.312(e)(2)(ii)
Microsoft Corporation		
24.	The environment is continuously monitored for security-related events.	CC7.1, CC7.2, CC7.3, §164.308(a)(1)(ii)(D), §164.308(a)(5)(ii)(C), §164.312(b), §164.312(c)(2)

Databricks management, along with the subservice organizations, defines the scope and responsibility of the controls necessary to meet all the relevant trust services criteria through written contracts, such as SLAs. In addition, Databricks performs monitoring of the subservice organization controls, including the following procedures:

- Holding periodic discussions with vendors and subservice organizations.
- Reviewing attestation reports over services provided by critical cloud service vendors and subservice organizations annually.
- Monitoring external communications, such as customer complaints relevant to the services by the subservice organizations.

User Entity Responsibilities

User entities must perform specific activities in order to benefit from Databricks' services. These activities may affect the user entity's ability to effectively use Databricks' services but do not affect the ability of Databricks to achieve its service commitments and system requirements. These activities may be specified in agreements between user entities and Databricks, user manuals, and/or other communications. These activities are referred to as user entity responsibilities (UERs).

UERs are listed in the following table. They are the responsibility of the user entities of the System and are expected to be in operation at user entities to complement Databricks' controls. The list of

UERs does not represent a comprehensive set of all the controls that should be employed by user entities. Other controls may be required at user entities.

Number	UER
1.	Customers are responsible for setting up separate development and production Data Intelligence Platform workspaces to isolate the production work from development, where applicable.
2.	Customers are responsible for implementing an SSO solution and MFA, where applicable, for controlling user access to the Data Intelligence Platform.
3.	If available in their subscription tier, customers are responsible for enabling Data Intelligence Platform audit log delivery features and implementing appropriate monitoring and incident response processes.
4.	Customers are responsible for reviewing Data Intelligence Platform audit logs of actions performed by the customer support personnel, if applicable.
5.	Customers are responsible for making any changes to data stored within the Data Intelligence Platform, where applicable.
6.	Customers are responsible for (1) backing up data that is stored in customer-controlled storage locations (e.g., AWS S3 bucket/Azure Blob Storage/GCS bucket), and (2) backing up notebooks with tools provided by Databricks within the product. For the data that is backed up by customers, customers are responsible for retaining the data for a specified retention period to protect from erasure or destruction.
7.	Customers are responsible for developing their disaster recovery and business continuity plans that address the inability to access or use Databricks services.
8.	Customers are responsible to determine, based on their needs, including on the relative sensitivity of the data they are processing, which product tier (and features) are appropriate for their intended uses.
9.	Customers are responsible for managing their organization’s instance(s) of the Data Intelligence Platform through any customized security solutions or automated processes using setup features, application development tools, and API integration tools.
10.	Customers are responsible for ensuring that authorized users are appointed as organizational administrators, and the authorized users should follow a defined logical access management process to manage and protect the admin account user credentials to administer the Data Intelligence Platform, where applicable. Customers are responsible for appropriately managing user access to the Data Intelligence Platform and customer-managed AWS, Azure, or GCP account(s).
11.	Customers are responsible for data classification and the implementation of encryption features available where deemed necessary by customer-defined requirements. Additionally, customers are responsible for their choices of libraries to prevent the inadvertent transmission of sensitive data (including PHI) over the wire in an unencrypted fashion.
12.	Customers are responsible for securing and protecting account credentials and security tokens for REST API access.

Number	UER
13.	Customers are responsible for securing and protecting the encryption keys that are stored in the customer-managed environment, where applicable.
14.	Customers are responsible for encrypting their sensitive data at rest (e.g., S3 bucket encryption, Azure Blob Storage encryption, GCS bucket encryption) that is used and processed by the Databricks services. Customers are responsible for proper use of encryption tools, including, but not limited to, up-to-date web browsers and REST API client tools, when accessing Data Intelligence Platform services (e.g., use of tools that properly implement TLS 1.2 encryption). Where applicable, customers are responsible for encrypting data in transit to Databricks personnel or Microsoft personnel (for Azure Databricks customers) over the network for interactions outside of the Data Intelligence Platform services (e.g., using TLS 1.2 encryption).
15.	Where applicable, customers are responsible for authorizing temporary access to customer support personnel from Databricks or Microsoft Azure (e.g., for technical support and troubleshooting).
16.	Customers are responsible for notifying Databricks of any unauthorized use of any password or account, or any other known or suspected breach of security, related to the use of the Data Intelligence Platform.
17.	Customers are responsible for communicating issues and incidents related to security, availability, and confidentiality to the support personnel (Databricks and Microsoft Azure, as appropriate) through identified channels.

Attachment B — Principal Service Commitments and System Requirements

Attachment B — Principal Service Commitments and System Requirements

In order to meet the security, availability, and confidentiality commitments of cloud services offered on the Data Intelligence Platform, Databricks has designed and implemented several technologies, processes, and procedures across its environments. Databricks documents and communicates its service commitments to its customers in the form of customer agreements. Service commitments include, but are not limited to:

Security

- The Data Intelligence Platform is designed to permit system users to access the information they need based on their role, while restricting them from accessing information not required for their role.
- Encryption technologies are implemented to help protect customer data.
- Access and activity logging with appropriate audit controls are in place to support incident management. Security incident response planning and notification requirements procedures exist to monitor, react, notify, and investigate incidents.
- Vulnerability scans are performed on a regular basis, with external third-party penetration tests executed at least annually for each cloud. Remediation follows industry acceptable vulnerability management processes, defined exception processes, and defined service-level agreements (SLAs).
- Industry-recognized application development standards (e.g., OWASP Top 10) are followed.

Availability

- Support is provided according to contract and is available 24 hours a day, seven days a week, with committed response times based on severity of the issue.
- Availability monitoring and capacity monitoring are in place to identify and alert on spikes in activity above predefined critical availability and capacity thresholds. An investigation is conducted and remediation is performed as needed, based on the nature and severity of the incident.
- Business continuity and disaster recovery mechanisms are in place to minimize data loss and ensure return to operations in the case of a disaster using industry-standard mechanisms. Business Continuity and Disaster Recovery Plans are tested at least annually.
- For databases that reside in the Databricks-managed AWS, Azure, and GCP accounts, backups are performed daily and retained.

Confidentiality

- Record retention policies are in place to retain information for periods defined by Databricks policies.
- Databricks identifies and maintains confidential information according to its objectives and applicable regulatory standards.

- Databricks deletes customer accounts per customer request in accordance with applicable laws and requirements outlined within customer contracts.

Databricks establishes operational requirements that support the achievement of security commitments, relevant laws and regulations, and other system requirements. Such requirements are communicated in Databricks' system policies and procedures, system design documentation, and contracts with customers. Information security policies define an organization-wide approach to how systems and data are protected. These include policies around how the service is designed and developed, how the system is operated, how the internal business systems and networks are managed, and how employees are hired and trained. In addition to these policies, standard operating procedures have been documented on how to carry out specific manual and automated processes required in the operation and development of the Data Intelligence Platform. The HIPAA-specific product offering of Databricks is subject to the security and privacy requirements of the Health Insurance Portability and Accountability Act, as amended, including relevant regulations, as well as state privacy security laws and regulations in the jurisdictions in which Databricks operates.

For more details on the security features of the Databricks product, please see the Security Features section of databricks.com/trust.