About Databricks

Databricks is a hosted end-to-end data platform powered by Apache® Spark™. Databricks makes it easy to turn data into value, from ingest to production, without the hassle of managing complex infrastructure, systems and tools.

Databricks is currently hosted on Amazon Web Services (AWS), including Amazon Elastic Compute Cloud (Amazon EC2) and Amazon Simple Storage Service (Amazon S3), and runs within each customer’s AWS account. AWS offers a highly secure and reliable platform used by thousands of businesses and institutions. Being on AWS allows Databricks and our customers to take full advantage of the cutting-edge security expertise and features offered by one of the top cloud infrastructure providers today.

We understand that the big data challenge extends beyond just data processing and analytics. An integral part of our vision is to help our customers protect their most valuable asset - the data and derived insights. Our technology has been designed to provide best-in-class security capabilities in the follow 3 key categories:

- **INFRASTRUCTURE SECURITY**
- **DATA SECURITY**
- **SECURITY AUDITS AND CONTROLS**
Infrastructure Security

ISOLATION WITH VIRTUAL PRIVATE CLOUD (VPC)
Databricks deploys Spark nodes in an Amazon Virtual Private Cloud (VPC) running in the customer’s own AWS account, giving the customer full control over their data and instances. VPCs enable customers to isolate the network that is used by their Databricks deployment from the rest of the networks they use on AWS, control access to it via Security Groups, and revoke permissions at any time. Data does not leave the customer’s storage infrastructure.

CONTROLLING ACCESS WITH IDENTITY AND ACCESS MANAGEMENT (IAM)
AWS Identity and Access Management (IAM) provides the ability to manage how users utilize and interact with different AWS services such as S3 or EC2 instances. Databricks customers can continue to leverage their existing AWS IAM credentials to limit access to their datasets on S3. For example, customers have the ability to limit what datasets a given Spark cluster has access to. They can also bring their IAM credentials to Databricks Notebooks and limit what actions a given Notebook can perform on a hosted S3 dataset.

STRONG PHYSICAL PROTECTION
AWS’ data centers are physically protected by electronic surveillance, multi-factor access control systems, and trained security guards staffed 24x7. Physical access is authorized strictly on a least privileged basis. Likewise AWS software service has been architected to provide the strictest security. The network traffic within AWS travels only over private network segments by default, fully isolated from the public Internet and not routable externally.
Data Security

Databricks takes a holistic approach to solving the enterprise security challenge by building all the facets of security — encryption, identity management, role-based access control, data governance, and compliance standards — natively into the data platform with the Databricks Enterprise Security (DBES) Framework.

**BUILT-IN ACCESS CONTROL**

Databricks is a multi-user platform built for large teams to work securely with shared resources. Our customers can limit access to confidential information, credentials, or execution of code on an individual basis through a simple, centralized interface.

**DATA ENCRYPTION AT REST AND IN FLIGHT**

Databricks encrypts all derived metadata (notebooks, logs and other artifacts) while they are at rest on the Databricks platform. Data is also secured during transit with strong, industry-standard encryption. Databricks utilizes TLS to encrypt the connection between the customer and Databricks’ user interface as well as the communication between Spark nodes and the storage layer.

**CONNECTS TO YOUR EXISTING MANAGED STORAGE**

When our customers deploy Databricks, all raw data stay in their storage platform of choice, such as S3, Cassandra and others. This architecture gives Databricks customers the flexibility to protect and manage their data with their preferred security infrastructure.

Download the DBES white paper to learn more.

[go.databricks.com/protecting-enterprise-data-on-apache-spark](go.databricks.com/protecting-enterprise-data-on-apache-spark)
Security Audits and Controls

AUDITED SECURITY
The Databricks platform is assured through rigorous third party auditing, to ensure the highest security standards. Databricks leverages iSEC Partners’ recognized world-class security expertise with periodic source-code auditing, proactively identifying and mitigating potential security issues that may adversely affect the integrity of the Databricks platform.

VERIFIED COMPLIANCE
Databricks has successfully completed SOC 2 Type 1 certification and can offer a HIPAA-compliant solution. The provider of our cloud infrastructure, Amazon Web Services, has achieved ISO 27001 certification. AWS also undergoes SOC 1 audits annually and have been successfully evaluated at the Moderate level for Federal government systems as well as DIACAP Level 2 for DoD systems.

Evaluate Databricks with a trial account today.
databricks.com/try-databricks