

GDPR Compliance with Databricks

About the General Data Protection Regulation (GDPR)

GDPR is a regulation that requires businesses around the world to protect the personal data and privacy of EU citizens. Starting on May 25, 2018 GDPR puts in place certain restrictions on the collection, use, and retention of personal data.

Important Things to Know



PERSONAL DATA

Defined as information relating to an identified or identifiable natural person. This includes data such as name, email, and phone number in addition to data that may be less obvious like IP address, GPS location, phone ID and more.



APPLICABILITY

GDPR applies to any company processing personal data from the EU regardless of where that company is headquartered or conducting business.



NON-COMPLIANCE

Penalties for non-compliance can be significant, with fines ranging up to the greater of 20 million euros or 4% of “global annual turnover” (i.e., global annual revenues).



DATA SUBJECT RIGHTS

GDPR grants EU residents a series of rights such as the right to know what personal data is collected, how it is used and to have that data changed or permanently removed.

Addressing GDPR Compliance

Key Areas of Focus

1 PERSONAL DATA PROCESSING AND STORAGE

Ability to demonstrate compliant collection, use and retention of personal data

2 DATA SUBJECT REQUESTS

Responding to requests for how personal data is used and removing it if requested within prescribed timelines

3 DATA PROCESSING AGREEMENTS AND SUBPROCESSORS

Enacting proper agreements with your subprocessors (like Databricks) that process personal data on your behalf

The Databricks Approach

Databricks is actively taking steps to provide our customers with the capabilities and support they need to achieve GDPR compliance by May 25, 2018 and is on track to be compliant by the May deadline. Data security and privacy has always been at the heart of what we do and complying with GDPR is a natural extension of this focus.

Processing and Storing Personal Data

DATA SECURITY

As part of our ongoing commitment to providing and continually improving data security, Databricks is 3rd party audited on an annual basis and is currently certified under SOC 2 Type II and ISO27001 and in the process of attesting to ISO27018, the internationally recognized standardized approach for protecting personal data in the cloud.

DATA STORAGE

Unlike other big data service providers, Databricks does not require customers to migrate their data into our platform. Databricks processes and stores your data within your cloud storage account. The small amount of data that is stored in our notebook environment (which is visible to the user) can be permanently erased to comply with GDPR.

DATA PROCESSING BEST PRACTICES

- Data Minimization - our platform is architected to store and process data in your environment to minimize our access to personal data.
- Pseudonymization - We employ best practice pseudonymization techniques (e.g., removing user email address and replacing it with a unique ID in our logs) to protect your personal data in the unlikely event of a breach and to prevent internal misuse of personal data.

Responding to Data Subject Requests

SELF HELP TOOLS

On or before May 25, Databricks will make available controls to enable you to permanently erase data in the Databricks notebook environment. Data will be erasable at a broad notebook or individual cell level.

Data Processing Agreements and Subprocessors

DATA PROCESSING AGREEMENT (DPA)

Databricks offers a DPA with the EC Standard Contractual Clauses (aka the Model Clauses) updated to explicitly address GDPR for customers who need it. The DPA includes language to cover the lawful transfer of Personal Data from the EEA to regions outside the EEA, including the US and India.

SUBPROCESSOR AGREEMENTS

Like other SaaS providers, Databricks utilizes certain subprocessors to provide the Databricks Services. Our subprocessors who may process Personal Data for our customers are held to the same standards defined in our DPA.

DEDICATED RESPONSE TEAMS

We have a dedicated team to help you respond to data subject requests within prescribed timelines. Additionally, in the unlikely event of a breach, our team of security professionals are ready to mitigate a potential breach and make any necessary notifications to our customers.

Contact us to learn more about the Databricks Unified Analytics Platform or start a POC today at databricks.com/contact-databricks.  databricks