cloudera®

WHITE PAPER

Simplify Your Response to General Data Protection Regulation (GDPR)

While Gaining Insights for your Business with Cloudera Enterprise

Version: 104

cloudera

Table of Contents

Cloudera does not provide legal advice and all users of Cloudera products should consult their own legal counsel for compliance.

cloudera

Overview

After dialogue and debate for much of this decade, the European Parliament approved the General Data Protection Regulation (the "**GDPR**") in April 2016. Given this is a Regulation (rather than a Directive), this legislation will apply automatically in every EU country beginning on **May 25th, 2018**. Many of the requirements are similar to those set out in Directive 95/46/EC (the "**EU Directive**"); however there are several expanded requirements as well as dramatically increased consequences for non-compliance.

For the majority of organisations, GDPR compliance is an enterprise-wide business problem — a massive cross-departmental effort that impacts oversight, technology, processes and people. This document highlights Cloudera's capabilities which may be useful as part of the technology portion of an organization's overall preparations for GDPR.

Broad Application

Compliance is required by every organization that processes data within the EU. GDPR also reaches outside of the EU, where personal data about EU residents is processed in connection with the "offering of goods and services" and monitoring behaviour. This includes the tracking of individuals online to create profiles.

Twenty million reasons to become compliant

GDPR violations can lead to fines of up to €20 million or 4% of global annual turnover for the preceding financial year, whichever is the greater.

Will GDPR be strictly enforced?
Yes, if pre-GDPR EU privacy cases are any indication. For example, in November 2016, the Hague administrative court in the Netherlands penalised WhatsApp for non-compliance with a requirement to appoint a representative in the EU. The court upheld an administrative penalty of €10,000 each day up to a maximum of €1,000,000 until WhatsApp complied with the regulation. Under GDPR, regulators will have even greater sanctioning powers.

"...customers' perceptions of privacy strongly affect the quality of customer interactions."
(the "Data Privacy Metrics That Matter To The Business" Forrester report)

To many, compliance leads to the avoidance of risk - avoiding big fines and damaged reputation. Beyond that, there is also the opportunity to build an increased reputation and level of trust with your customers - both those inside and outside the EU - many of whom value companies that in turn value their privacy.

Big Data and GDPR – A Challenge and an Opportunity

GDPR applies to data and IT systems across the enterprise and reaches across multiple layers where changes to comply with the regulation are likely needed: from customer-facing website workflows to backend data storage and archiving - and everything in between. This document focuses primarily on challenges related to data storage.

As companies consolidate more and more data sets from legacy relational databases and data warehouses into big data environments, they benefit from having a single place where data security and governance policies can be managed and enforced consistently. That means that certain new security controls implemented for GDPR might only need to be implemented once, rather than needing to be implemented separately, likely with differing methods, on different data systems. For example, access controls, encryption and audit controls configured consistently with the GDPR principles described below.

Yet big data environments can also bring their own challenges: they typically contain large amounts of rapidly updating data from a wide range of sources, stored in a multitude of formats, while at the same time, a variety of tools is used to process it. The incredible flexibility these environments offer enable everyone from call center analysts to data scientists to get the answers they need. However, these same characteristics can make these environments a challenge to govern. For example, when individuals exercise their privacy rights, organisations need to ensure they can not only locate all relevant personal data about those individuals, but also extract or delete that data upon request.

Cloudera does not provide legal advice and all users of Cloudera products should consult their own legal counsel for compliance.

cloudera

GDPR Principles

GDPR is a complex and detailed regulation, and this document only deals with a portion of it. Article 5 lists principles that organizations are obligated to follow, with regard to processing of personal data:

- Lawfulness, fairness and transparency
- Purpose limitation
- Data minimisation
- Accuracy
- Storage limitation
- Integrity and confidentiality
- Accountability

In addition to the general principles above, GDPR also includes the information security concepts of:

- Data protection by design and by default
- Preventing unlawful data transfers outside the EU
- Data breach management and notification

How can Cloudera help?

Achieving GDPR compliance requires a combination of changes to technology and process. Cloudera can help organisations on their journey to compliance with both. A typical GDPR readiness program starts with an enterprise-wide privacy assessment resulting in a list of requirements. Once that has occurred, Cloudera's software technology and services can assist customers in meeting specific requirements within their Cloudera big data environment.

Apache Kudu for Rapid Update & Erasure

One design pattern consistent with GDPR principles involves the use of Apache Kudu for personal data. Kudu is storage for fast analytics on fast data—providing a combination of fast inserts, deletes and updates alongside efficient columnar scans to enable multiple real-time analytic workloads across a single storage layer.

Technology

From a technology perspective, Cloudera's Enterprise Data Hub can help organizations address the challenges, principles and concepts highlighted in the sections above. It provides a single unified platform for organisations to store data, cost-effectively and reliably, for as long as needed, and lets users access that data in a variety of ways while being governed by security and compliance policies. Specifically, Cloudera's capabilities simplify and accelerate efforts toward complying with the various key GDPR requirements where data is stored in the Cloudera platform. Some of the relevant capabilities are described below.

Services

For both existing and new customers, Cloudera also offers a variety of packaged service engagements. Our two week GDPR architecture review engagement helps organisations map the requirements they have identified for meeting GDPR to standardised design patterns using the platform. The engagement may also include implementing a single data set in the new architecture.

Project prerequisites include:

- An active Cloudera subscription
- A representative development cluster set up to Cloudera technical pre-requisites, including security
- Datasets that will be subject to GDPR
- Data Protection Impact Assessments (DPIA) for datasets under consideration
- Identified use cases for / access patterns against those datasets

Cloudera also offers a one-week security assessment, as well as a two or three-week comprehensive security deployment.

Cloudera does not provide legal advice and all users of Cloudera products should consult their own legal counsel for compliance.

cloudera

Addressing GDPR Principles with Cloudera Capabilities

The table below lists the principles from Article 5 of GDPR, together with relevant Cloudera capabilities that may help address each principle.

Principle	Description	Cloudera Enterprise capabilities that may be useful as part of an organization's plan to address GDPR
Lawfulness, fairness and transparency	Personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject	**Cloudera Navigator** can be used to apply metadata classification on ingest so that data sets are tagged automatically and consistently and the organisation doesn't lose track of where that data is located. Tags can indicate the presence of personal data and what type. A data protection officer could then query Navigator to locate personal tagged datasets. In addition to customer-provided scripts, certified Cloudera partners provide data discovery tools which can automatically apply classification tags.
Purpose limitation	Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes	In addition to the above, **Cloudera Navigator** also tracks the details of how data has been used. Column-level data lineage to track personal data as it is combined with other datasets. Other environments typically only track lineage at the file or table level. **Cloudera Data Science Workbench** (**CDSW**) enables enables data scientists to perform work with the flexibility of a workstation environment while data remains on governed servers.
Data minimisation	Personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed	The same capabilities listed in the 2 rows above also apply here. For example, **Cloudera Navigator** tags could indicate the purpose(s) of a particular data field.
Accuracy	Personal data shall be accurate and, where necessary, kept up to date	Data classification and tagging with **Cloudera Navigator** may be used to identify personal data, to assist with locating all of a user's personal data when an update is needed. When personal data is stored in **Kudu**, updates can be made to individual records without the need to wait for a re-write of the entire file.
Storage limitation	Personal data shall be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed	**Apache Kudu** is a big data storage engine with immediate record-level erasure. File-based storage systems including HDFS and cloud object stores don't support individual record erasure, but may still be part of a GDPR compliant data environment by periodically re-writing files to remove the identifiable personal data that must be erased. Cloudera professional services can assist with designing data architectures that limit personal data to a relatively small number of lookup tables, so that data erasure may be streamlined. **Cloudera Backup and Disaster Recovery** (**BDR**) can be used to copy only carefully selected data objects to a backup site, either in an on-premise data center or in a regional public cloud.
Integrity and confidentiality	Personal data shall be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures	Comprehensive encryption at rest and in motion, including separation of duties so that even administrators can't access user data. Fine-grained authorization ensures users who should have permission to access datasets are the only ones with the ability to do so.
Accountability	The controller/processor shall be responsible for, and be able to demonstrate compliance with the GDPR	Comprehensive, inescapable audit trail, ensures that the organizations can always perform a detailed analysis of exactly what data was accessed by users.

Cloudera does not provide legal advice and all users of Cloudera products should consult their own legal counsel for compliance.

cloudera

In Summary

Achieving GDPR compliance is not as simple as deploying a single policing application or flicking the 'GDPR mode' switch on existing systems. Fundamentally, it comes down to understanding the data landscape and being able to pinpoint and operate on data with surgical precision. Cloudera helps organisations on their journey to GDPR compliance with both technology and services. Contact your Cloudera representative today to find out more.

Disclaimer

This document is intended to help organizations understand how Cloudera software and services can be used to help comply with certain aspects of EU General Data Protection Regulation (GDPR) requirements. Applicability of any of these capabilities depends on each organization's own requirements specific to their business. Every organization should determine its own needs with regard to GDPR and then evaluate solutions for suitability to those needs. The information contained in this document is not intended to be and should not be construed to be legal advice. Customers and prospective customers must not rely on the information herein and they should obtain legal advice from their own legal counsel or other professional legal services provider.

Cloudera does not provide legal advice and all users of Cloudera products should consult their own legal counsel for compliance.

cloudera®

Notes

cloudera

About Cloudera

Cloudera delivers the modern platform for machine learning and advanced analytics built on the latest open source technologies. The world's leading organizations trust Cloudera to help solve their most challenging business problems by efficiently capturing, storing, processing and analyzing vast amounts of data. Learn more at cloudera.com.

cloudera.com

1-888-789-1488 or 1-650-362-0488
Cloudera, Inc. 1001 Page Mill Road, Palo Alto, CA 94304, USA

© 2017 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice.