Analyzing the Impact of GDPR on Storage Systems

Aashaka Shah, Vinay Banakar, Supreeth Shastri
Melissa Wasserman and Vijay Chidambaram
General Data Protection Regulation (GDPR)

May 25, 2018
Adopted after 2 years of public debate.
All but 2 EU countries have legislated.

Fundamental right
Grants all European people a right to protection and privacy of personal data

Personal data
Any information relating to a natural person;
Broad in scope unlike FERPA, HIPAA

Covers entire lifecycle
Collection, processing, protection, transfer and deletion; Regulated via 99 articles

Hefty penalty
Max penalty of 4% of global revenue
or €20 million, whichever is greater
GDPR Entities

- **Controller** (e.g., Spotify)
  - store and process personal data internally
  - send personal data for external processing
- **Processor** (e.g., Google cloud)
  - notify data breaches
  - audit and investigate
- **Data Subject** (e.g., Spotify user)
  - exercise GDPR rights
  - provide personal data
- **Supervisory Authority**
  - report GDPR violations
  - audit and investigate
- **Other Controllers** (e.g., SoundCloud)
  - share personal data
  - allow data sharing

**Personal data**

**GDPR queries**
**GDPR in the Wild**

- **Terminated**
  - KLOUT
  - iab.europe

- **Adapted**
  - USA TODAY
  - The New York Times

- **Advertised compliance**
  - BigTech

- **Assumed compliance**
  - everyone else

*Less than 50% estimated compliance*

By the end of 2018 [Gartner 2018]

*94,622 complaints from people*

In the first 9 months of GDPR rollout
Investigate how GDPR-compliance impacts Storage Systems

- What effort is needed to make a modern storage system, GDPR-compliant?
- What is the resulting performance impact?
- Is it possible to achieve strict compliance in an efficient manner?
Analyzing GDPR: Two Key Observations

31 of the 99 GDPR articles directly pertain to data storage.

GDPR’s goal of data protection by design and by default conflicts with the traditional system design goals of performance, cost, and reliability.
Key GDPR Articles concerning Storage Systems

Rights of data subjects

[15] Right of access
[17] Right to be forgotten
[20] Right to data portability
[21] Right to object

Responsibilities of Data Controllers

[5] Purpose / Storage limitations
[13] Conditions for data collection
[25] Protection by design & by default
[30] Records of processing activities
[33] Notification of data breaches
# Translating GDPR Articles into Storage Features

<table>
<thead>
<tr>
<th>GDPR article</th>
<th>Key requirement</th>
<th>Storage feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 Right of access by users</td>
<td>Allow customers to access all their data</td>
<td>Metadata indexing</td>
</tr>
<tr>
<td>17 Right to be forgotten</td>
<td>Find and delete groups of data</td>
<td>Timely deletion</td>
</tr>
<tr>
<td>21 Right to object</td>
<td>Avoid using data for any objected reasons</td>
<td>Metadata indexing</td>
</tr>
<tr>
<td>25 Protection by design and by default</td>
<td>Safeguard and restrict access to data</td>
<td>Access control, Encryption</td>
</tr>
<tr>
<td>30 Records of processing activity</td>
<td>Store audit logs of all operations on data</td>
<td>Logging</td>
</tr>
<tr>
<td>33 Notify data breaches</td>
<td>Share insights and logs from affected systems</td>
<td>Monitoring</td>
</tr>
<tr>
<td>46 Transfers subject to safeguards</td>
<td>Control where the data resides</td>
<td>Managing location</td>
</tr>
</tbody>
</table>
Features of GDPR-Compliant Storage

Timely deletion
Associate TTL to all personal data; it can be static value or a policy criterion

Metadata indexing
Provide quick and efficient access to groups of data

Encryption
Encrypt data at rest, and while in transit

Manage data Location
Ability to find and control the location of personal data at all times

Access control
Limit access to permitted entities, for established purposes, and for predefined duration of time

Monitoring & Logging
Save the audit trail of all internal actions and external interactions
GDPR-Compliance is a Spectrum

**Response Time**

- **Real-time**
  - Complete GDPR tasks synchronously in real-time

- **Eventual**
  - Complete GDPR tasks asynchronously

**Capability**

- **Full**
  - Support all GDPR features natively

- **Partial**
  - Support for some GDPR features is lacking or coarse-grained
Despite needing to implement a small set of new features for GDPR-compliance, storage systems would experience significant performance impact.
# Redis’ support for GDPR features

<table>
<thead>
<tr>
<th>FULL</th>
<th>PARTIAL</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monitoring &amp; Logging</strong>&lt;br&gt;Manage data Location</td>
<td><strong>Timely deletion</strong>&lt;br&gt;Metadata indexing</td>
<td><strong>Encryption</strong>&lt;br&gt;Access control</td>
</tr>
</tbody>
</table>
GDPR-Compliant Redis: Monitoring & Logging

Three built-in options

- **MONITOR** debug command
- Configure `slowlog` option
- Piggyback on AoF

*modified AoF code to include read/scan operations*

Even fully supported features can cause significant **performance overheads**
GDPR-Compliant Redis: Timely Deletion

Three options to delete

- **DEL** and **UNLINK**
- **FLUSH** `{DB | ALL}`
- **EXPIRE** and **EXPIREAT**

Redis erases expired keys using a lazy randomized algorithm.

We changed it to a static scheme (== sub-second latency for up to 1M keys)

System internals should be carefully analyzed to determine the degree of compliance.

![Time to erase (sec) vs. Total keys in data store for Redis with fast active expiry](image)
GDPR-Compliant Redis: Encryption

No native support
- Encryption at rest w/ LUKS
- Encryption in transit w/ STunnel

Investigated key-level encryption using Themis (== similar performance overhead)

Retrofitting new features not aligned with the core design principles of the system will result in excessive performance overheads
Concluding Remarks

“In law, nothing is certain but the expense.” — Samuel Butler

GDPR-compliant Redis
Performance impact of GDPR on a modern storage system

Research challenges
Efficient Logging; Efficient Deletion; Efficient Metadata indexing

Beyond GDPR
California’s CCPA is going into effect 1/1/2020

We want to hear from you!

https://utsaslab.github.io/research/gdpr/