Impact of General Data Protection Regulation GDPR on systems & storage[1]

Contents:

What it is; to whom it applies

In May 2016, the European Parliament published the General Data Protection Regulation EU 2016/679, which is compulsory in each member state starting 25 May 2018. It replaces the data protection directive 95/46/EC from 1995, and does not require any enabling legislation from national governments. The intention is a stronger data protection for individuals within the European Union, and a unified regulation of the export of personal data outside the EU.

GDPR could be described as anything between a *logical evolution* and a *privacy earthquake*. To help reduce speculation, this paper highlights how IT systems capabilities and IT systems software are affected by GDPR when used for processing personal data. GDPR defines precise measures around storing and processing that data:

- The *pseudonymisation* and encryption of personal data.
- Measures to support the *"right to erasure"* and irrecoverable deletion of personal data.
- Measures to restore the availability and access to data in the event of a data breach, with an obligation to notify authorities of the breach.
- An obligation to notify affected individuals of any data breach, alleviated only if the exposed data was demonstrably anonymized and/or adequately encrypted to prevent misuse.
- Measures to ensure resilience of systems and services processing data.
- Frequent testing of the effectiveness of the security measures.

Affected systems & software capabilities

Two aspects directly touch storage systems and data management software: (1) Encryption of *processed personal data*, which calls for media-level encryption of data at rest, besides application-level encryption; (2) Controlled data placement and tracking of physical copies in a central inventory, or *copy data management*.
IBM offers infrastructure/software solutions and assistance for your GDPR assessment activities regarding these aspects.

Media-level data encryption

An often overlooked potential data protection breach involves replacements of failed storage components by a service technician: *"Failed"* does not imply *"unreadable"* – in most cases data can still be retrieved even if the component was part of a larger data distribution scheme, like a RAID set. Various technologies may help mitigate this vulnerability:

Encryption for Flash & legacy storage media

At the lowest media layer, protection is usually offered by native encryption, like the always-encrypted Flash modules of the IBM **FlashSystems** family; idem for disk drives.
At the SAN layer, IBM offers inline encryption for other vendor's legacy systems with its IBM **Spectrum Virtualize** software, deployed as-is or inside a storage virtualization appliance (SAN Volume Controller, FlashSystem V series, Storwize). Note also that data reduction measures like real-time compression or de-duplication fail on encrypted data, so IBM software applies encryption *after* data reduction.

Encryption and certified deletion for file spaces

At file level, both IBM **Spectrum Scale** (formerly GPFS) and IBM **Spectrum Protect** (formerly TSM) encrypt files after applying compression and de-duplication techniques, thus preserving compressibility. Per-file encryption is a simple way to provide irrevocable deletion on commodity hardware, while avoiding cumbersome physical erasure as a preventive measure against fraudulent low-level access. Per-file encryption also means that IBM **Spectrum Scale** can execute deletions of such files in a *single* metadata operation (key re-wrapping), reducing the performance impact from mass deletions, e.g. after a migration.

Protection for geographically dispersed data

In geographically dispersed environments where global key management is complex, advanced protection is available in IBM **Cloud Object Storage** (formerly Cleversafe): A mathematical encoding known as *all-or-nothing transform* dictates the retrieval of numerous dispersed data chunks to successfully decode a data sequence. Incomplete fraudulent retrievals remain illegible and protected from misuse. This method avoids having to manage global encryption keys[*].
Note: IBM Cloud Object Storage also supports AES-256.

[*] www.usenix.org/legacy/event/fast11/tech/full_papers/Resch.pdf

IBM storage helps mitigate the risk of data breach from uncontrolled media access.

Media encryption, file deletion by re-keying, and the *all-or-nothing transform* help mitigate risks associated with fraudulent hardware level access. Without demonstrably appropriate data protection, any accidental or unlawful personal data breach results in a mandatory notification to each and every affected individual (article 32.2).

Mixing encryption and de-duplication

Multiple layers of encryption are not generally problematic, hence the best practice is to encrypt or anonymize personal data at the earliest possible stage, following "privacy-by-design" guidelines. Already encrypted data will be transparently re-encrypted by most self-encrypting media without further performance loss.
However, the same is not true for data reduction like de-duplication or compression: IT shops heavily relying on these techniques may see their provisioning pipeline disrupted by unanticipated growth of incompressible pre-encrypted content. WAN data compression will also be affected, multiplying the necessary bandwidth. We suggest using data pseudonymisation rather than full application encryption if compression and de-duplication are strategic. See enclosed link.

Automatic data placement

Automatic data placement designates the policy-driven choice of storage devices or media depending on the data type. It is most common for hierarchical storage management optimizing cost versus performance (low latency data on fast Flash, write-rarely data on read-intensive dense Flash, read-never data on Tape).
But automatic data placement also enables users to:

- enforce encrypted media for certain critical data;
- manage data erasure independently of the current media.

Example:
The policy engine in IBM Spectrum Scale can be programmed to prevent sensitive application data from being placed on any legacy storage deemed insecure, including the data copies required for HA/DR. Beyond that, IBM Storage integrates with IBM Spectrum Copy Data Management, giving you general control over data copies generated on-the-fly for test & development purposes.

IBM storage assists you in protecting personal information by controlling its physical copies through the lifecycle.

Imperative Copy Data Management

Besides data breaches, GDPR also outlines potential consequences for failures to fulfil the *"right to erasure"* and the irrecoverable deletion of specific personal data sets.
It is therefore advisable to track storage-level copies from application data, and to avoid *stray* copies that may contain personal user information. Stray copies often appear during cloning for development testing (should be anonymized), during migrations (leaving behind the original), and during safety fallback snapshots before software upgrades.

IBM **Spectrum Copy Data Management** offers a three-fold functionality for cloning/snapshot administration by:

- allowing application operators to trigger select cloning workflows without any access to the storage layer;
- retaining an auditable history of copies that have been generated including their revocation or deletion date;
- prohibiting copy workflows that do not incorporate data masking for pseudonymisation, e.g. in Oracle or SAP database cloning for development testing.

Copy Data Management provides ways to track data copies (incl. Disaster Recovery copies) that you determine might require irrecoverable deletion for the *"right to erasure"*.
Copy Data Management combined with automatic data placement also regulates the storage selection for sensitive data copies by providing only appropriate workflows.

Tracking backups and archived data

Archives and backups are subject to policies defined by GDPR, like the *"right to erasure"*. We therefore suggest that long-living data that is cyclically copied to newer media or to remote locations is logically tracked.
IBM **Spectrum Protect** offers this kind of tracking, with an industry-leading scalability in terms of managed number of objects and capacity. Spectrum Protect can identify and delete all managed copies of a data set by means of meta-data operations. Spectrum Protect can encrypt data regard-less of the underlying media. These protective measures also help address fraudulent data access from outside or inside. Finally Spectrum Protect and its archive sibling **Spectrum Protect for Data Retention** can help avoid data loss or access loss by offering best-in-class resiliency, high availability and disaster recovery features.

Is there a "GDPR compliance" certificate?

No. Unlike for immutable archiving, there is no certification label for GDPR at the time of writing this article. Some product-related certificates like the KPMG compliance label accredit the quality of mandatory features like resistance to tampering for immutable archive solutions. Such thing is not defined for GDPR.
GDPR readiness is rather achieved by adopting policies and procedures around processing of personal data, as defined in the regulation. While storage & data management products can help achieving this, they do not constitute a compliance guarantee. Minimum standards are required by the regulation e.g. for encryption complexity. However, calling a product "GDPR compliant" solely based on that feature would be misleading.

Conclusion

GDPR is a process-oriented regulation, therefore do not rely solely on *"GDPR-compliant"* product stickers. Data management tools will better support you meeting your obligations while processing personal data. Think encryption, pseudonymisation, data breach prevention (fraudulent or accidental), policy-based data placement, tracking of copies and storage resilience. For pseudonymisation, review the IBM Identity Mixer demo (link enclosed). For data management, consider IBM Spectrum Scale & Spectrum Virtualize, IBM Cloud Object Storage, IBM Spectrum Protect, and IBM Spectrum Copy Data Management to support your GDPR activities.

For more information

How IBM Storage supports GDPR: **ibm.biz**/nordics-gdpr

IBM 360° assessments for GDPR readiness: **ibm.com**/gdpr

Pseudonymisation demo: **ibm.biz**/identity-mixer
(an innovative way to reduce the cost of privacy safekeeping)

The GDPR FAQ from the European Union:
europa.eu/rapid/press-release_MEMO-15-6385_en.htm

Latest update of this paper (IBM login required):
ibm.biz/gdpr-storage-paper

[1] The customer is responsible for ensuring compliance with legal requirements. It is the customer's sole responsibility to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer is in compliance with any law or regulation.

IBM Corporation Systems Group
May 2017