Analyzing the Impact of GDPR on Storage Systems

Aashaka Shah, Vinay Banakar, **Supreeth Shastri**
Melissa Wasserman and Vijay Chidambaram

The University of Texas at Austin
Computer Science

The University of Texas at Austin
School of Law

Hewlett Packard
Enterprise

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

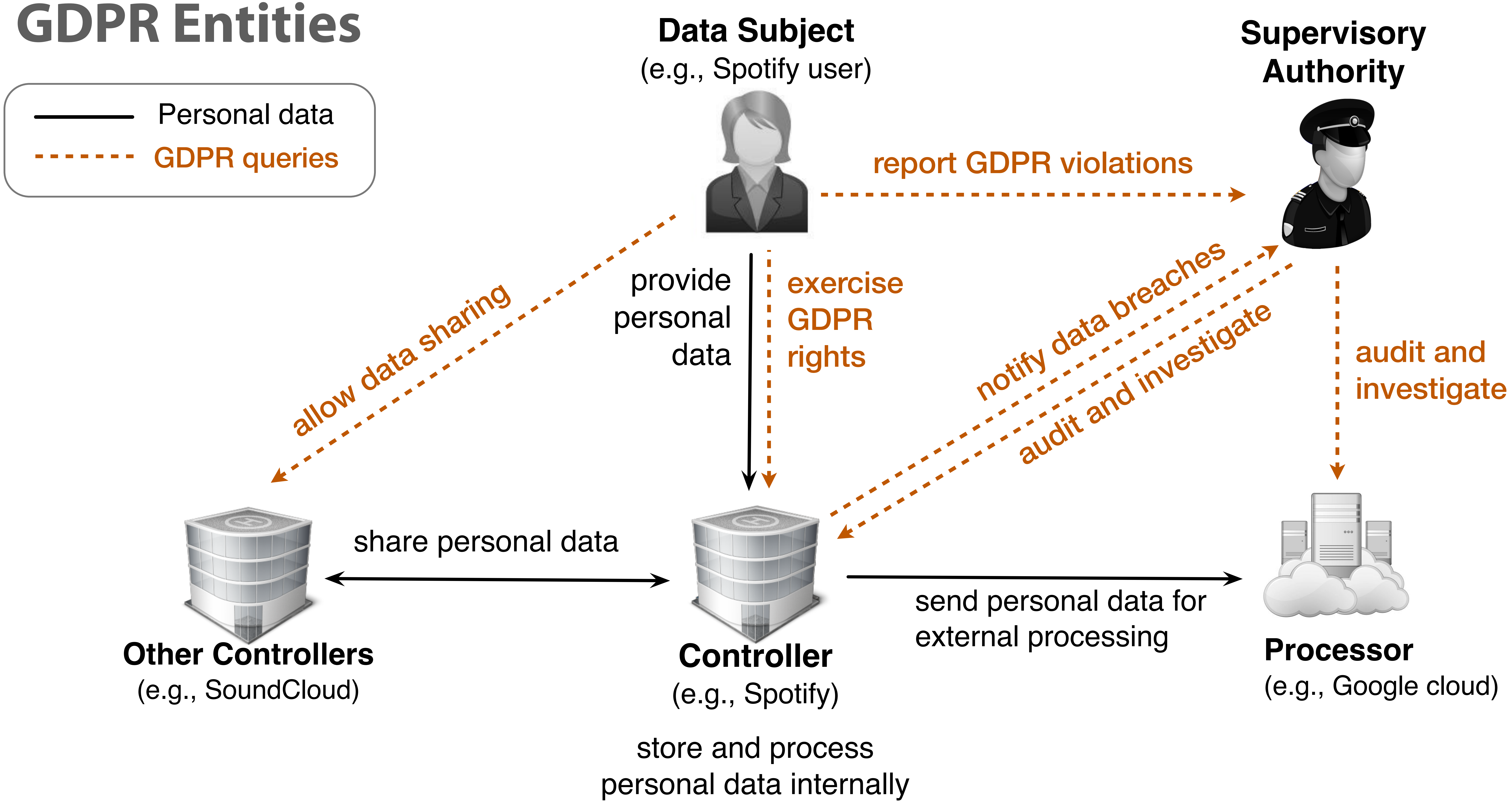
Covers entire lifecycle

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



Personal data

GDPR queries

Data Subject
(e.g., Spotify user)

Supervisory Authority

report GDPR violations

allow data sharing

provide personal data

exercise GDPR rights

notify data breaches

audit and investigate

audit and investigate

share personal data

send personal data for external processing

Other Controllers
(e.g., SoundCloud)

Controller
(e.g., Spotify)

store and process personal data internally

Processor
(e.g., Google cloud)

GDPR in the Wild

Terminated — KLOUT, iab.europe

Adapted — USA TODAY, The New York Times

Advertised compliance — BigTech

Assumed compliance —

everyone else

<50%

estimated compliance

By the end of 2018 [Gartner 2018]

94,622

complaints from people

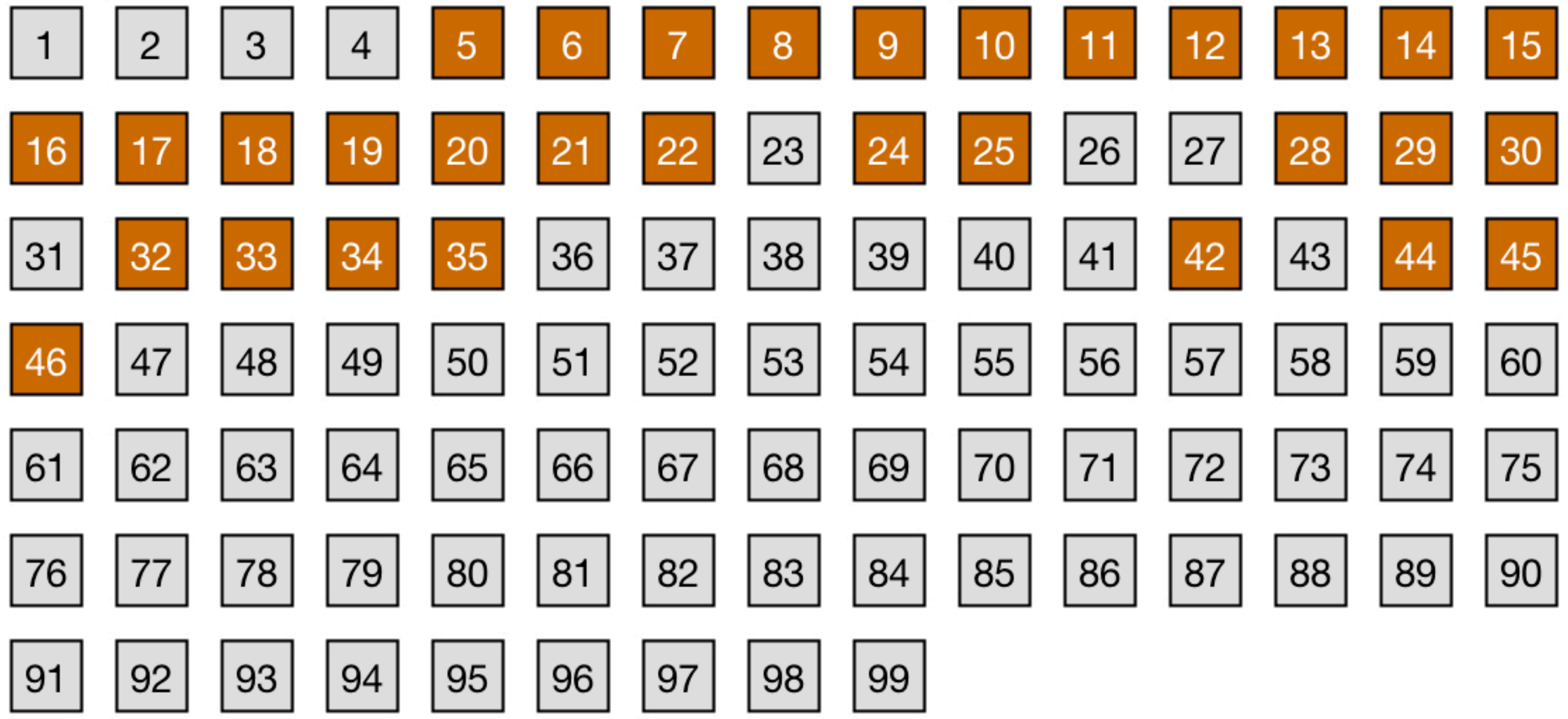
In the first 9 months of GDPR rollout

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



1	2	3	4	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
16	**17**	**18**	**19**	**20**	**21**	**22**	23	**24**	**25**	26	27	**28**	**29**	**30**
31	**32**	**33**	**34**	**35**	36	37	38	39	40	41	**42**	43	**44**	**45**
46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99						

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**

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

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

HYPOTHESIS

Despite needing to implement a **small set** of new features for **GDPR**-compliance, storage systems would experience **significant** performance impact.

GDPR-Compliant **Redis**

benchmark with ***YCSB***

Redis' support for GDPR features

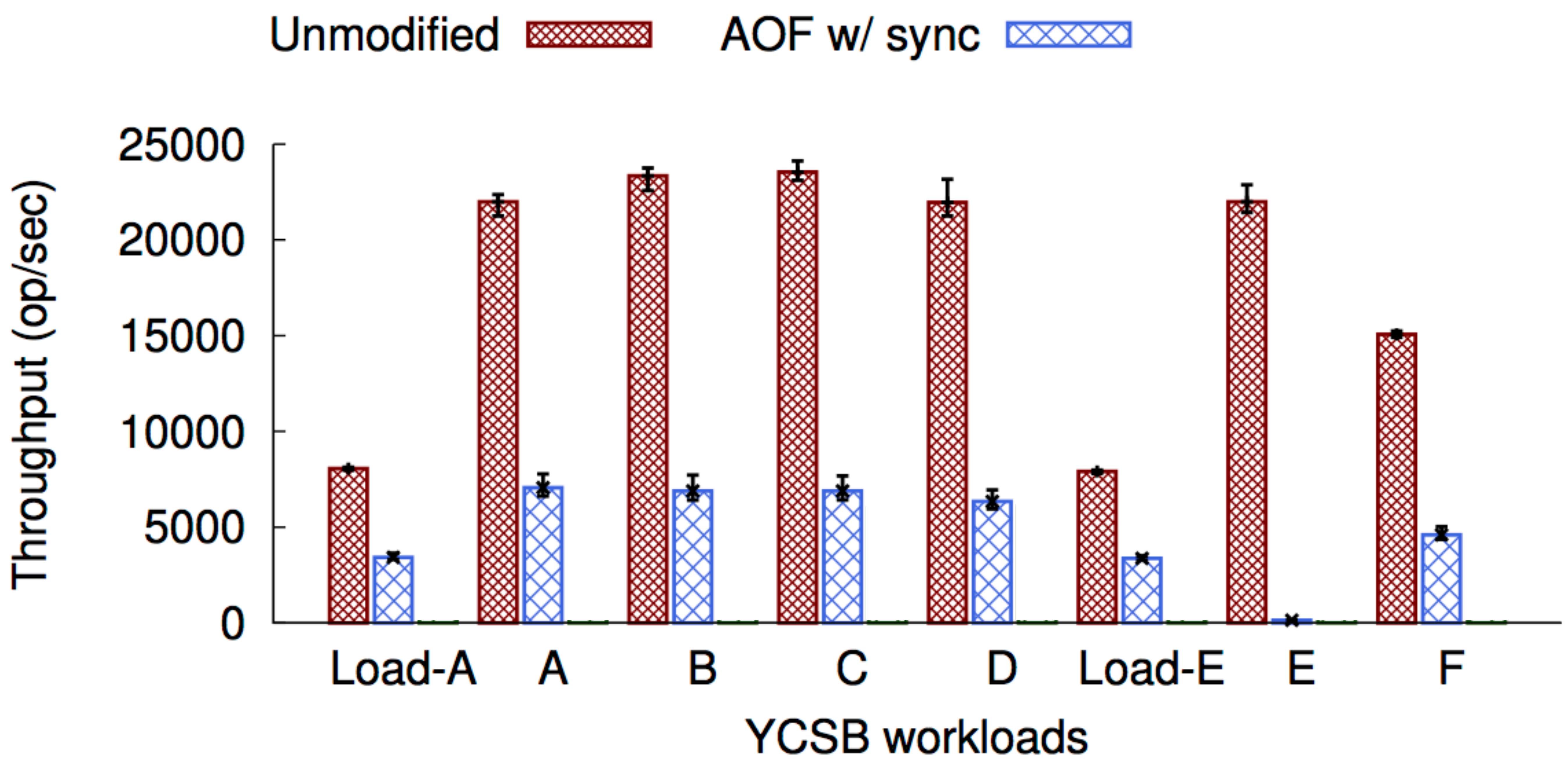
FULL	PARTIAL	NO
Monitoring & Logging	**Timely deletion**	**Encryption**
Manage data Location	Metadata indexing	Access control

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**

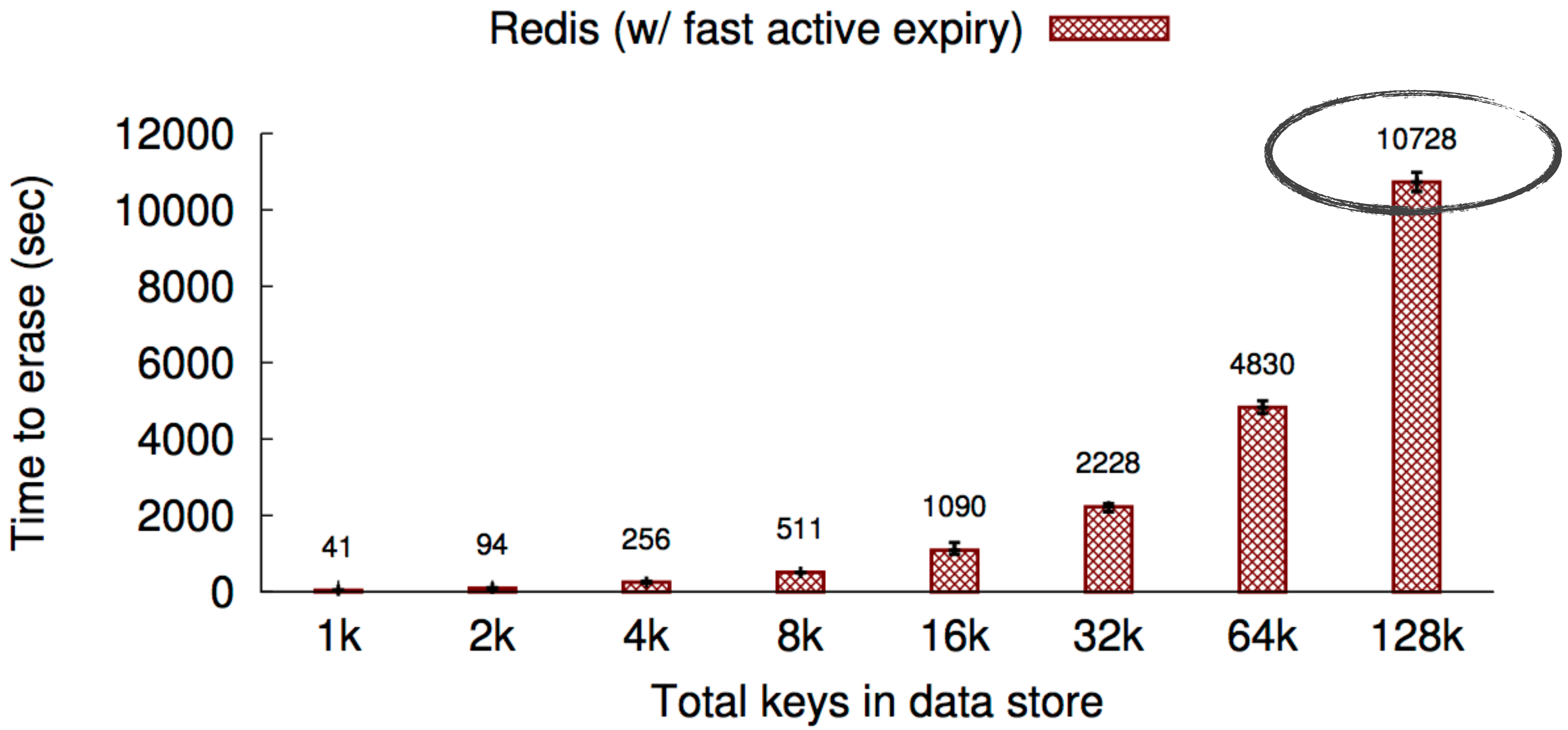
GDPR-Compliant Redis: **Timely Deletion**

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)*



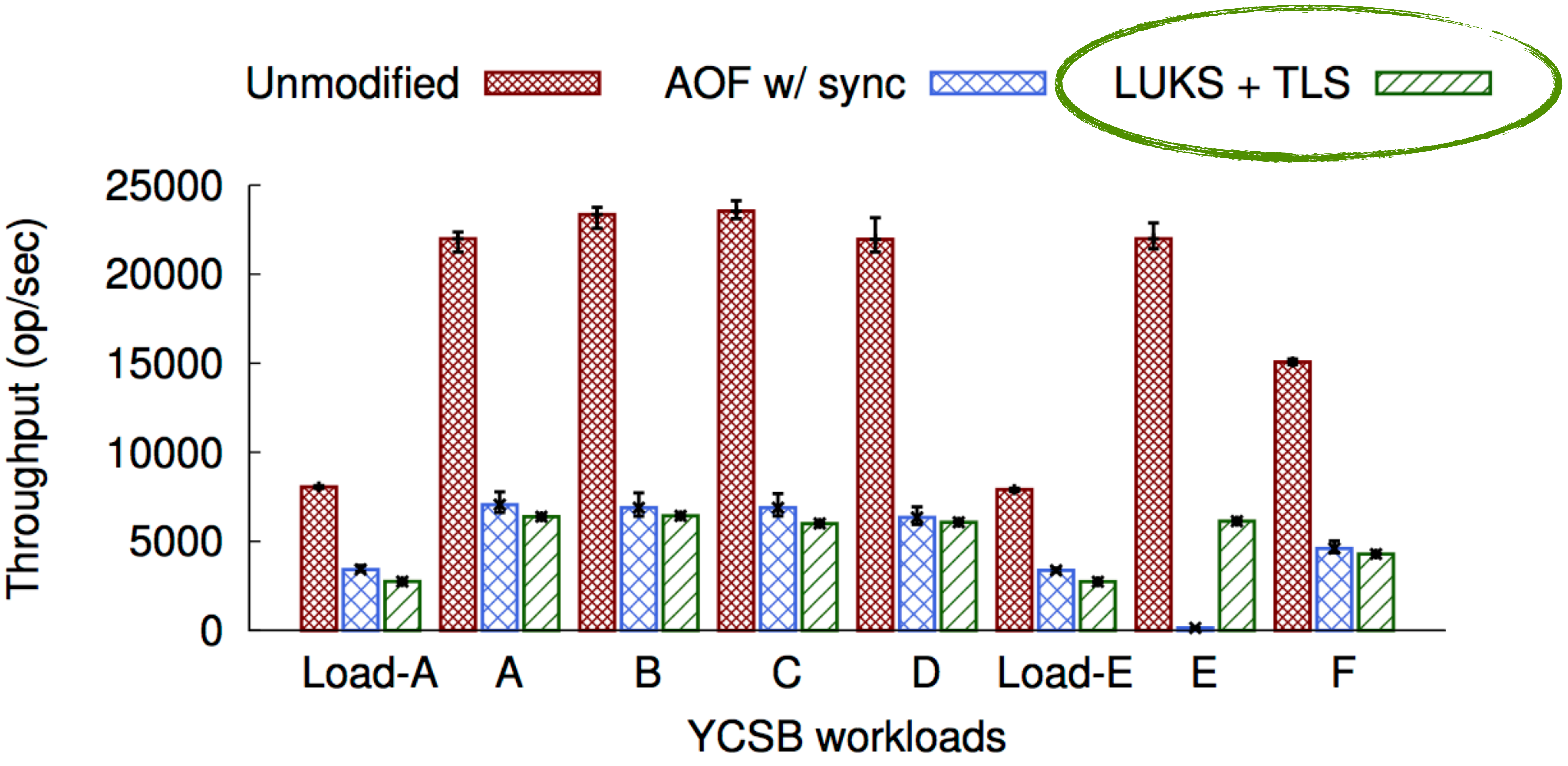
System internals should be carefully analyzed to determine the ***degree of compliance***

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/