Understanding and Benchmarking the Impact of **GDPR** on **Database Systems**


Supreeth Shastri


Vinay Banakar


Melissa Wasserman


Arun Kumar


Vijay Chidambaram

TEXAS
The University of Texas at Austin

Hewlett Packard Enterprise

UC San Diego

vmware

General Data Protection Regulation (GDPR)

Privacy and protection of personal data is a fundamental right of natural persons

General Data Protection Regulation (GDPR)

Privacy and protection of personal data is a fundamental right of natural persons

99 Legal Articles

Regulate the collection, processing, protection, transfer and deletion of personal data

General Data Protection Regulation (GDPR)

Privacy and protection of personal data is a fundamental right of natural persons

99 Legal Articles

Regulate the collection, processing, protection, transfer and deletion of personal data

Grants Rights to People

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

General Data Protection Regulation (GDPR)

Privacy and protection of personal data is a fundamental right of natural persons

99 Legal Articles

Regulate the collection, processing, protection, transfer and deletion of personal data

Grants Rights to People

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

Assigns Responsibilities to Companies

Those who collect and process personal data are solely responsible for its privacy and protection

General Data Protection Regulation (GDPR)

Privacy and protection of personal data is a fundamental right of natural persons

99 Legal Articles

Regulate the collection, processing, protection, transfer and deletion of personal data

Grants Rights to People

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

Assigns Responsibilities to Companies

Those who collect and process personal data are solely responsible for its privacy and protection

Hefty Penalty

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

Complying with GDPR has been a *challenge*

Complying with GDPR has been a *challenge*

Google

€50M

French Data Protection Authority, Jan 2019

Marriott HOTELS & RESORTS

$123M

UK Data Protection Agency, Jun 2019

BRITISH AIRWAYS

£183M

UK Data Protection Agency, Jun 2019

Complying with GDPR has been a *challenge*

Google

€50M

French Data Protection Authority, Jan 2019

Marriott HOTELS & RESORTS

$123M

UK Data Protection Agency, Jun 2019

BRITISH AIRWAYS

£183M

UK Data Protection Agency, Jun 2019

Public *Complaints*

144,376

EU-wide (Year 1)

Personal Data

any information relating to an identified or identifiable natural person

GDPR §4(1)

Personal Data

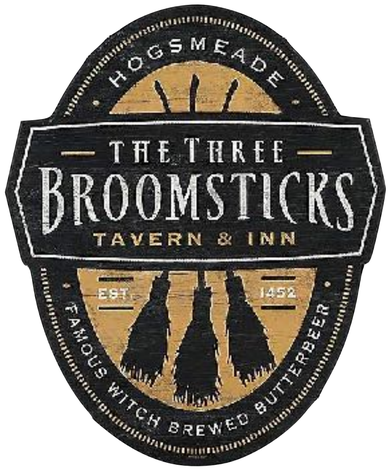
any information relating to an identified or identifiable natural person

GDPR §4(1)

Personal Data

any information relating to an identified or identifiable natural person

GDPR §4(1)



Personal Data

any information relating to an identified or identifiable natural person

GDPR §4(1)

Prof. Albus Dumbledore

- *Has a phoenix as pet*
- *Drinks coffee at 8am*
- *Published a paper at VLDB 2020*

Personal Data

any information relating to an identified or identifiable natural person

GDPR §4(1)

Prof. Albus Dumbledore

- *Has a phoenix as pet*
- *Drinks coffee at 8am*
- *Published a paper at VLDB 2020*

I have
eight rights!

Right to know, access, rectify, erase, object, port, restrict processing, and withdraw from automated processing

Personal Data

any information relating to an identified or identifiable natural person

GDPR §4(1)

Prof. Albus Dumbledore

- *Has a phoenix as pet*
- *Drinks coffee at 8am*
- *Published a paper at VLDB 2020*

I have
eight rights!

Right to know, access, rectify, erase, object, port, restrict processing, and withdraw from automated processing

I have
responsibilities

To obtain consent, track data usage, keep it secure, notify breaches etc.

Personal Data

any information relating to an identified or identifiable natural person

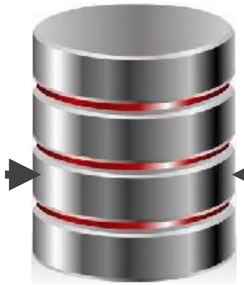
GDPR §4(1)

Prof. Albus Dumbledore

- *Has a phoenix as pet*
- *Drinks coffee at 8am*
- *Published a paper at VLDB 2020*

I have
eight rights!

Right to know, access, rectify, erase, object, port, restrict processing, and withdraw from automated processing



I have
responsibilities

To obtain consent, track data usage, keep it secure, notify breaches etc.

*How to build a **GDPR-compliant** database system for storing personal-data?*

Analyze

Translate GDPR articles into system-level capabilities and characteristics

Build

Implement GDPR requirements in Redis and PostgreSQL

Measure

Benchmark compliant systems against GDPR workloads

Store Data with a Timeline for Deletion

Store Data with a Timeline for Deletion

§ 5(1)(E): Storage Limitation

"[…] kept for no longer than is necessary for the purposes for which the personal data are processed […]"

Store Data with a Timeline for Deletion

§ 5(1)(E): Storage Limitation

"[…] kept for no longer than is necessary for the purposes for which the personal data are processed […]"

§17: Right To Be Forgotten

(1) The data subject shall have the right to obtain from the controller the erasure of personal data without undue delay […]

Store Data with a Timeline for Deletion

§ 5(1)(E): Storage Limitation

"[...] kept for no longer than is necessary for the purposes for which the personal data are processed [...]"

§17: Right To Be Forgotten

(1) The data subject shall have the right to obtain from the controller the erasure of personal data without undue delay [...]

GDPR-compliant datastore should:

Associate a **time-to-live** attribute with all data

Have support for **timely deletion** of data

Store Data with a Timeline for Deletion

§ 5(1)(E): Storage Limitation

"[...] kept for no longer than is necessary for the purposes for which the personal data are processed [...]"

§17: Right To Be Forgotten

(1) The data subject shall have the right to obtain from the controller the erasure of personal data without undue delay [...]

GDPR-compliant datastore should:

Associate a **time-to-live** attribute with all data

Have support for **timely deletion** of data

Keep Record of Data Processing Activity

Store Data with a Timeline for Deletion

§ 5(1)(E): Storage Limitation

"[...] kept for no longer than is necessary for the purposes for which the personal data are processed [...]"

§17: Right To Be Forgotten

(1) The data subject shall have the right to obtain from the controller the erasure of personal data without undue delay [...]

GDPR-compliant datastore should:

Associate a **time-to-live** attribute with all data

Have support for **timely deletion** of data

Keep Record of Data Processing Activity

§ 30: Records of Processing Activities

(1) Each controller [...] shall maintain a record of processing activities under its responsibility.

Store Data with a Timeline for Deletion

§ 5(1)(E): Storage Limitation

"[...] kept for no longer than is necessary for the purposes for which the personal data are processed [...]"

§17: Right To Be Forgotten

(1) The data subject shall have the right to obtain from the controller the erasure of personal data without undue delay [...]

GDPR-compliant datastore should:

Associate a **time-to-live** attribute with all data

Have support for **timely deletion** of data

Keep Record of Data Processing Activity

§ 30: Records of Processing Activities

(1) Each controller [...] shall maintain a record of processing activities under its responsibility.

§ 33: Notification of A Data Breach

(1) the controller shall without undue delay and not later than 72 hours after having become aware of it, notify [...] (3) The notification shall at least describe the nature of the personal breach."

Store Data with a Timeline for Deletion

§ 5(1)(E): Storage Limitation

"[...] kept for no longer than is necessary for the purposes for which the personal data are processed [...]"

§17: Right To Be Forgotten

(1) The data subject shall have the right to obtain from the controller the erasure of personal data without undue delay [...]

GDPR-compliant datastore should:

Associate a **time-to-live** attribute with all data

Have support for **timely deletion** of data

Keep Record of Data Processing Activity

§ 30: Records of Processing Activities

(1) Each controller [...] shall maintain a record of processing activities under its responsibility.

§ 33: Notification of A Data Breach

(1) the controller shall without undue delay and not later than 72 hours after having become aware of it, notify [...] (3) The notification shall at least describe the nature of the personal breach."

GDPR-compliant datastore should:

Associate an **audit trail** with all data

Implement support for **monitoring/logging** of all data accesses

Translating GDPR Articles into Systems-Level Attributes and Actions

Translating GDPR Articles into Systems-Level Attributes and Actions

We analyzed all the 99 articles of GDPR, both individually and collectively…

Translating GDPR Articles into Systems-Level Attributes and Actions

We analyzed all the 99 articles of GDPR, both individually and collectively…

GDPR Metadata

Associate seven behavioral attributes with personal data

Translating GDPR Articles into Systems-Level Attributes and Actions

We analyzed all the 99 articles of GDPR, both individually and collectively…

GDPR Metadata

Associate seven behavioral attributes with personal data

1	***Purpose***
2	***TTL***
3	***Audit trail***
4	***Objections***
5	***Origin of data***
6	***Externally shared?***
7	***Use in automated decision-making?***

Translating GDPR Articles into Systems-Level Attributes and Actions

We analyzed all the 99 articles of GDPR, both individually and collectively…

GDPR Metadata

Associate seven behavioral attributes with personal data

1	***Purpose***
2	***TTL***
3	***Audit trail***
4	***Objections***
5	***Origin of data***
6	***Externally shared?***
7	***Use in automated decision-making?***

GDPR Capabilities

*Implement **five features** in the database system*

Translating GDPR Articles into Systems-Level Attributes and Actions

We analyzed all the 99 articles of GDPR, both individually and collectively…

GDPR Metadata

Associate seven behavioral attributes with personal data

1	*Purpose*
2	*TTL*
3	*Audit trail*
4	*Objections*
5	*Origin of data*
6	*Externally shared?*
7	*Use in automated decision-making?*

GDPR Capabilities

*Implement **five features** in the database system*

Encryption

Monitoring

Access control

Timely deletion

Metadata-based querying

Translating GDPR Articles into Systems-Level Attributes and Actions

We analyzed all the 99 articles of GDPR, both individually and collectively…

GDPR Metadata

Associate seven behavioral attributes with personal data

1	*Purpose*
2	*TTL*
3	*A*
	rigin of data
6	*Externally shared?*
7	*Use in automated decision-making?*

Storage overhead

GDPR Capabilities

*Implement **five features** in the database system*

Encryption

Monitoring

Access control

Timely deletion

Metadata-based querying

Translating GDPR Articles into Systems-Level Attributes and Actions

We analyzed all the 99 articles of GDPR, both individually and collectively…

GDPR Metadata

Associate seven behavioral attributes with personal data

1	*Purpose*
2	*TTL*
3	*A*
	…rigin of data
6	*Externally shared?*
7	*Use in automated decision-making?*

Storage overhead

GDPR Capabilities

*Implement **five features** in the database system*

…toring

Access control

Performance overhead

Timely deletion

Metadata-based querying

GDPR-Compliant Storage Systems

GDPR-Compliant Storage Systems

Goal: Introduce GDPR-compliance into two representative storage systems and measure its impact

GDPR-Compliant Storage Systems

Goal: Introduce GDPR-compliance into two representative storage systems and measure its impact

GDPR-Compliant Storage Systems

Goal: Introduce GDPR-compliance into two representative storage systems and measure its impact

redis PostgreSQL

- Encryption
- TTL/Timely deletion
- Monitoring/Logging
- Metadata Indexing
- Access control
- GDPR queries

GDPR-Compliant Storage Systems

Goal: Introduce GDPR-compliance into two representative storage systems and measure its impact

	redis	PostgreSQL
Encryption	*3rd party lib*	*3rd party lib*
TTL/Timely deletion	*Code change*	*Scripting*
Monitoring/Logging	*Code change*	*Configure*
Metadata Indexing	*Scripting*	*Configure*
Access control	*Scripting*	*Configure*
GDPR queries	*Code change*	*Scripting*

GDPR-Compliant Storage Systems

Goal: Introduce GDPR-compliance into two representative storage systems and measure its impact

	redis	PostgreSQL
Encryption	*3rd party lib*	*3rd party lib*
TTL/Timely deletion	*Code change*	*Scripting*
Monitoring/Logging	*Code change*	*Configure*
Metadata Indexing	*Scripting*	*Configure*
Access control	*Scripting*	*Configure*
GDPR queries	*Code change*	*Scripting*

Performance overhead in
Yahoo! Cloud Serving Benchmark (YCSB)

GDPR-Compliant Storage Systems

Goal: Introduce GDPR-compliance into two representative storage systems and measure its impact

	redis	PostgreSQL
Encryption	*3rd party lib*	*3rd party lib*
TTL/Timely deletion	*Code change*	*Scripting*
Monitoring/Logging	*Code change*	*Configure*
Metadata Indexing	*Scripting*	*Configure*
Access control	*Scripting*	*Configure*
GDPR queries	*Code change*	*Scripting*

80% ↓

50% ↓

Performance overhead in
Yahoo! Cloud Serving Benchmark (YCSB)

*How to benchmark database systems against **GDPR workloads?***

How to benchmark database systems against GDPR workloads?

We build a new open-source benchmark called GDPRbench

How to benchmark database systems against GDPR workloads?

We build a new open-source benchmark called GDPRbench

GDPR Queries

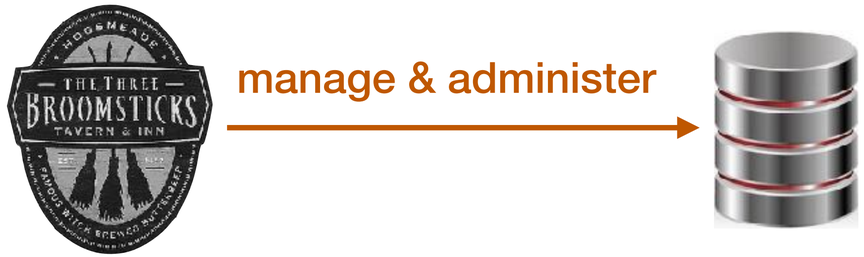
Control- and data-path operations performed on GDPR datastores

How to benchmark database systems against GDPR workloads?

We build a new open-source benchmark called GDPRbench

GDPR Queries

Control- and data-path operations performed on GDPR datastores

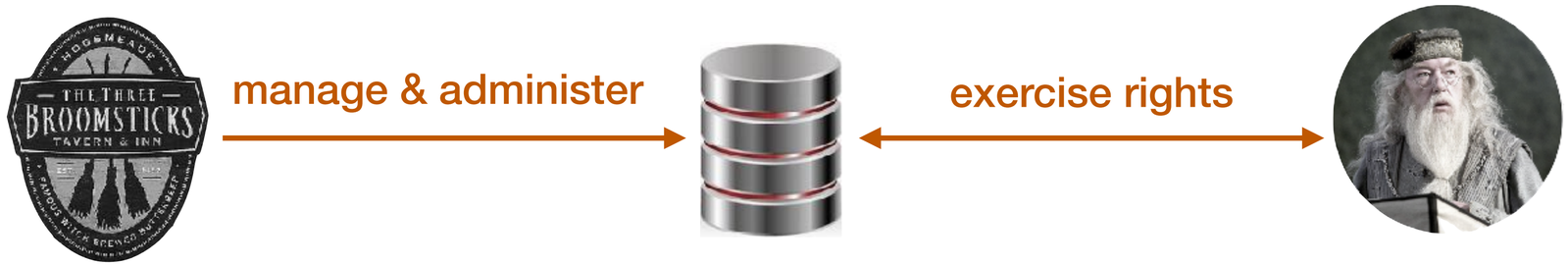


How to benchmark database systems against GDPR workloads?

We build a new open-source benchmark called GDPRbench

GDPR Queries

Control- and data-path operations performed on GDPR datastores

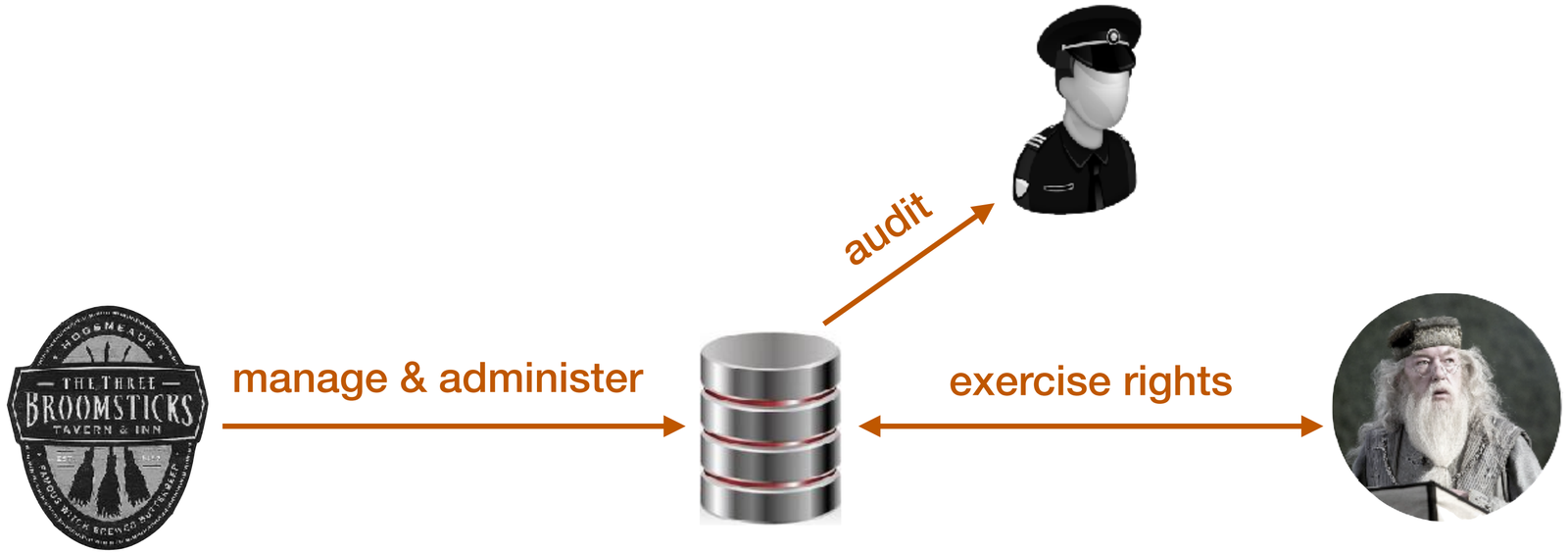


How to benchmark database systems against GDPR workloads?

We build a new open-source benchmark called GDPRbench

GDPR Queries

Control- and data-path operations performed on GDPR datastores

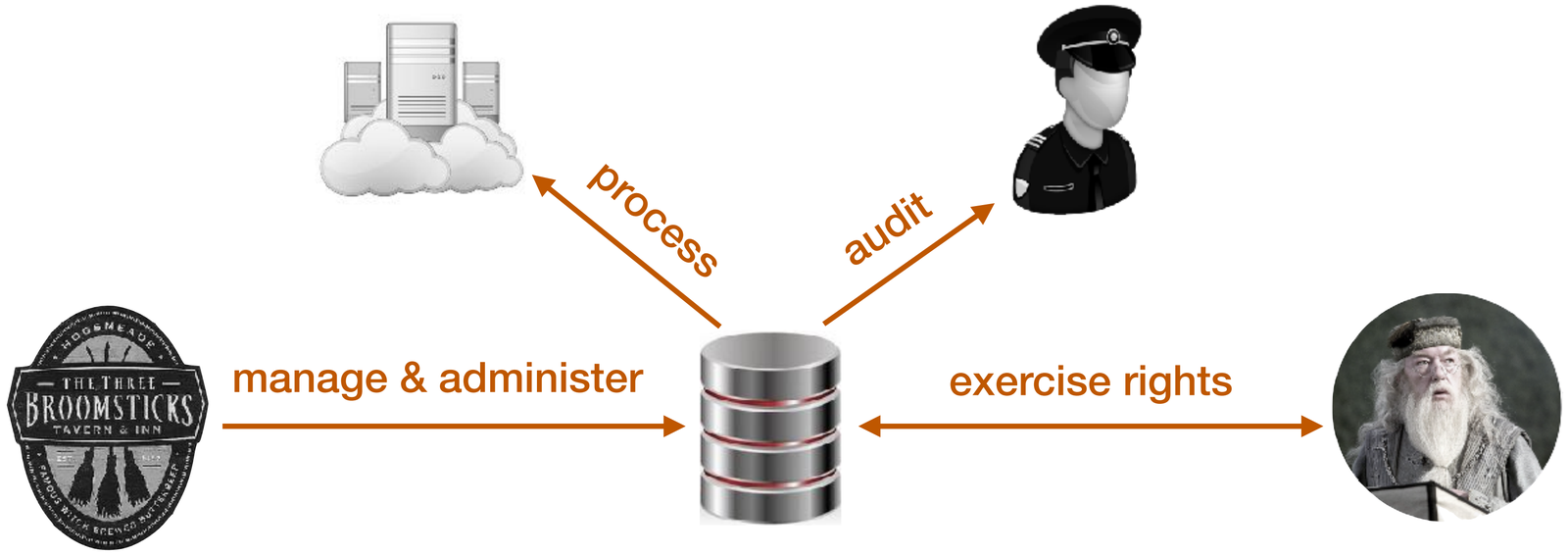


How to benchmark database systems against GDPR workloads?

We build a new open-source benchmark called GDPRbench

GDPR Queries

Control- and data-path operations performed on GDPR datastores

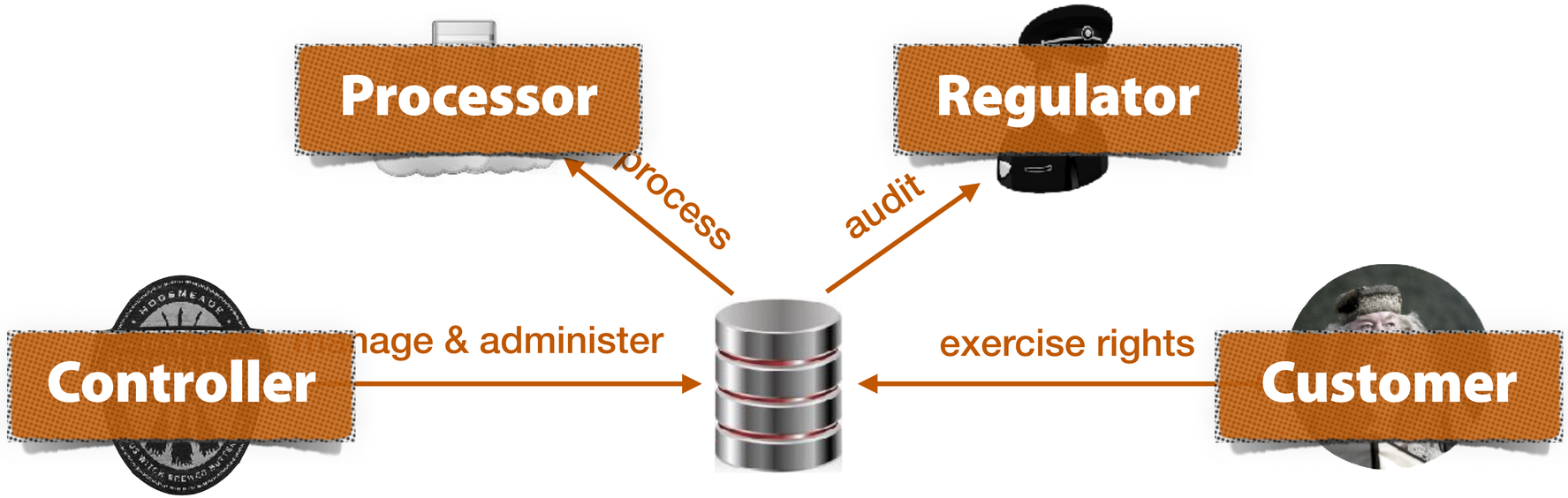


How to benchmark database systems against GDPR workloads?

We build a new open-source benchmark called GDPRbench

GDPR Queries

Control- and data-path operations performed on GDPR datastores

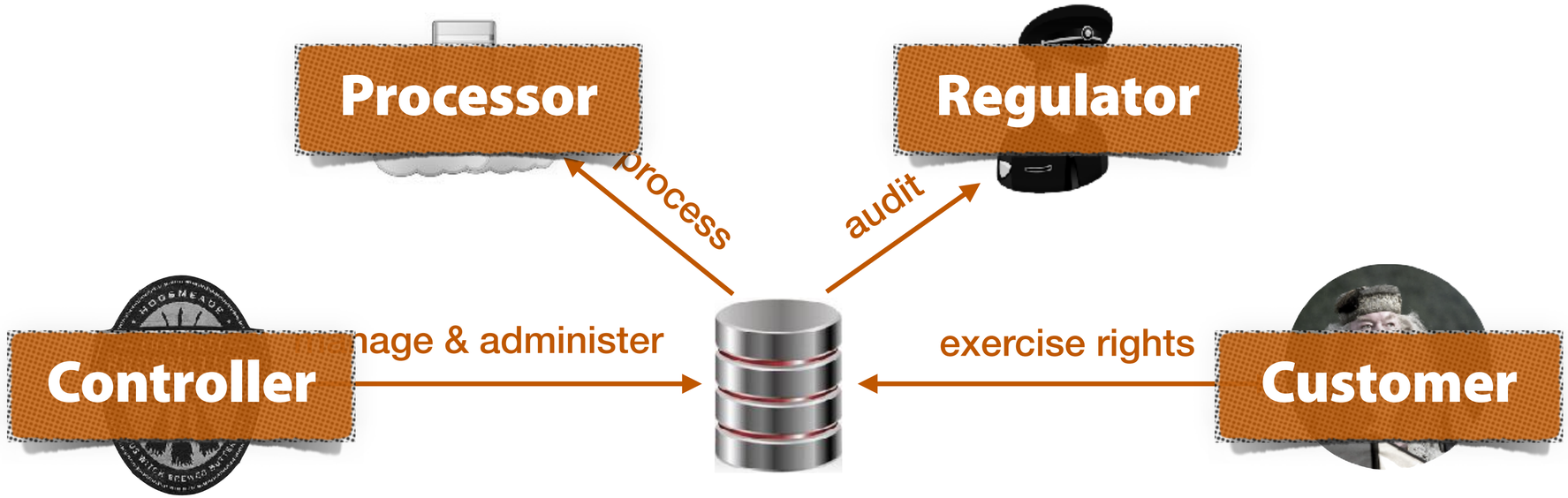


How to benchmark database systems against GDPR workloads?

We build a new open-source benchmark called GDPRbench

GDPR Queries

Control- and data-path operations performed on GDPR datastores



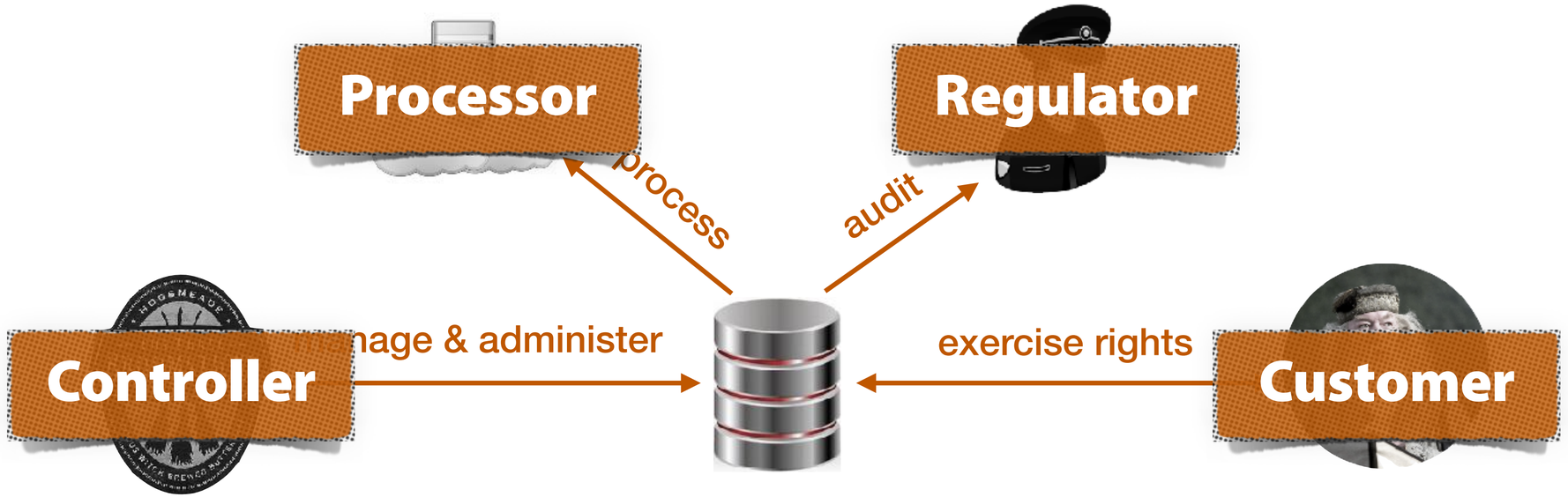
Benchmark Metrics

How to benchmark database systems against GDPR workloads?

We build a new open-source benchmark called GDPRbench

GDPR Queries

Control- and data-path operations performed on GDPR datastores



Benchmark Metrics

Correctness	Completion Time	Space Overhead
% responses that match the expected results	*Time to complete all the workloads*	*Ratio of total DB size to size of personal data*

*How do our compliant systems perform against **GDPRbench**?*

*How do our compliant systems perform against **GDPRbench**?*

*How do our compliant systems perform against **GDPRbench**?*

3.5X
space overhead

*How do our compliant systems perform against **GDPRbench**?*

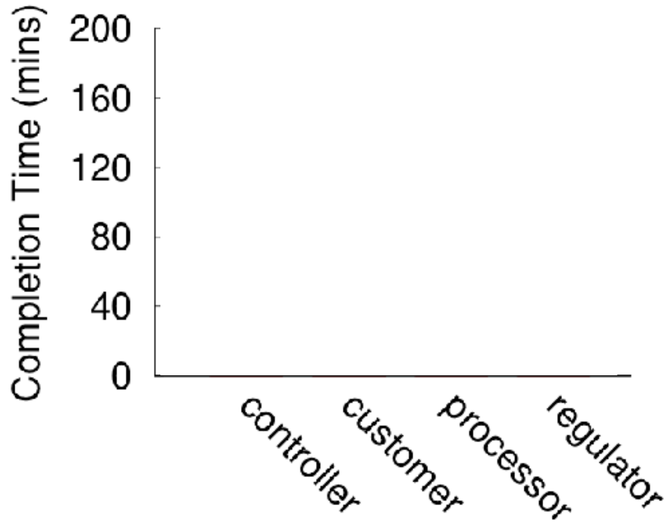
3.5X
space overhead

5.95X
space overhead w/
metadata indices

*How do our compliant systems perform against **GDPRbench**?*

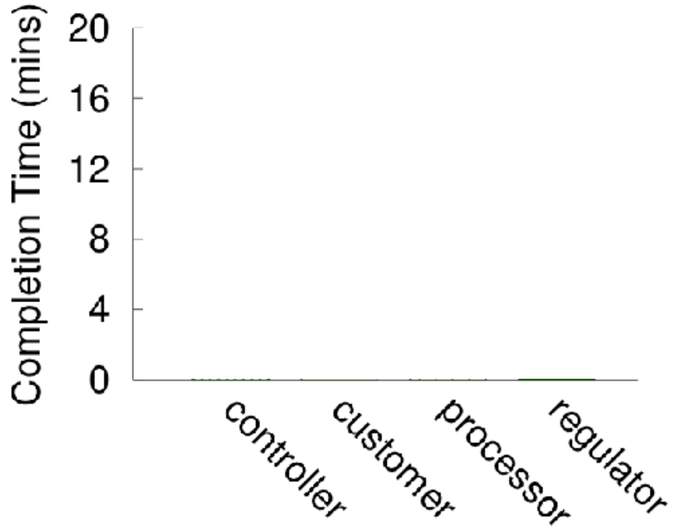
redis

3.5X
space overhead



PostgreSQL

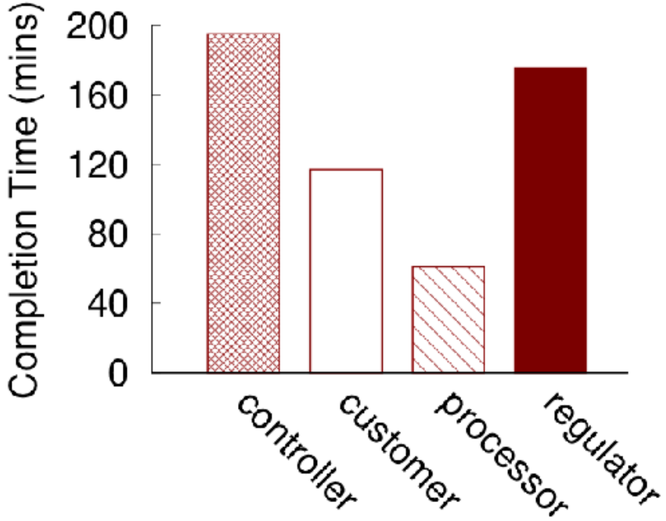
5.95X
space overhead w/
metadata indices



*How do our compliant systems perform against **GDPRbench**?*

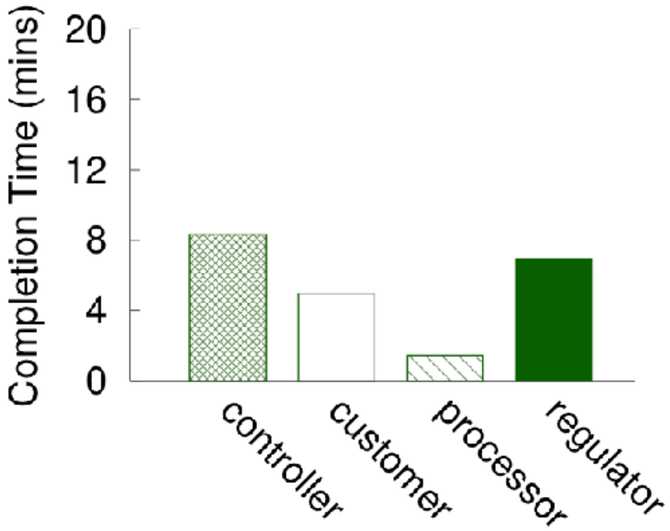
redis

3.5X
space overhead



PostgreSQL

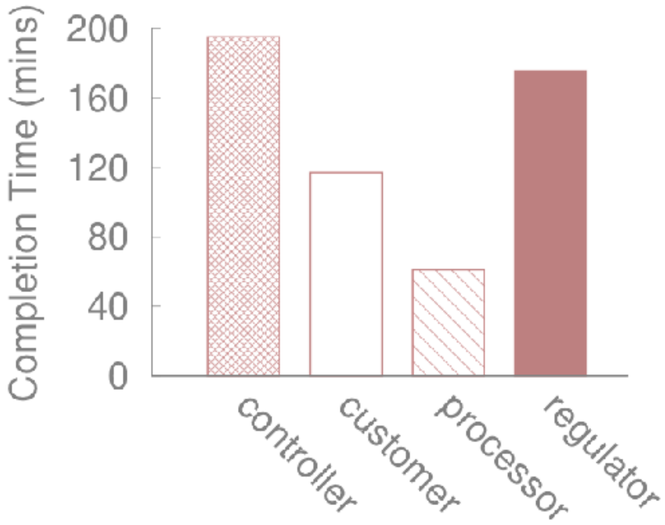
5.95X
space overhead w/
metadata indices



*How do our compliant systems perform against **GDPRbench**?*

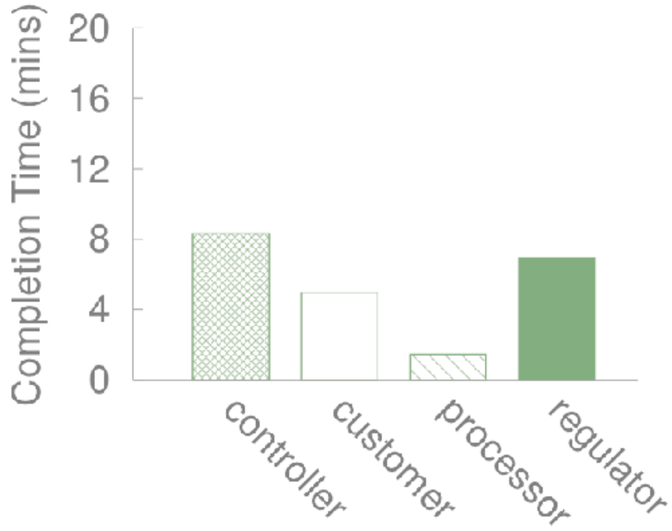
redis

3.5X
space overhead



PostgreSQL

5.95X
space overhead w/
metadata indices



*GDPR workloads **run faster** and **scale better** on SQL than NoSQL databases*

Real-World Implications

Real-World Implications

Compliance may result in high performance overheads

Production system should be carefully analyzed before enabling/claiming compliance

Real-World Implications

Compliance may result in high performance overheads

Production system should be carefully analyzed before enabling/claiming compliance

Compliance is easier in RDBMS than in NoSQL

Redis needed more involved changes and had much higher overhead

Real-World Implications

Compliance may result in high performance overheads

Production system should be carefully analyzed before enabling/claiming compliance

Compliance is easier in RDBMS than in NoSQL

Redis needed more involved changes and had much higher overhead

Compliance is a spectrum

Examine tradeoffs b/w strictness vs. efficiency

Need mechanisms for efficient auditing/timely deletion/indexing

We want to hear from you!

We want to hear from you!

Find out more at

https://www.GDPRbench.org/

Our Interpretation of GDPR

> “ In Law, nothing is certain but the expense ”

— *Samuel Butler*

Our Interpretation of GDPR

> “ In Law, nothing is certain but the expense ”
>
> *— Samuel Butler*

Prof. Melissa Wasserman
Law faculty, UT Austin

Our Interpretation of GDPR

> “ In Law, nothing is certain but the expense ”
>
> — *Samuel Butler*

Prof. Melissa Wasserman
Law faculty, UT Austin

Response Time

Real-time
Complete GDPR tasks synchronously

Eventual
Complete GDPR tasks asynchronously

Our Interpretation of GDPR

> “ In Law, nothing is certain but the expense ”
>
> — *Samuel Butler*

Prof. Melissa Wasserman
Law faculty, UT Austin

Response Time

Real-time
Complete GDPR tasks synchronously

Eventual
Complete GDPR tasks asynchronously

Granularity of Rights

Per data item
Support right for every piece of data

Per service/~~person~~
Support rights at the level of service

Our Interpretation of GDPR

> “ In Law, nothing is certain but the expense ”
>
> — *Samuel Butler*

Prof. Melissa Wasserman
Law faculty, UT Austin

Response Time

Real-time
Complete GDPR tasks synchronously

Eventual
Complete GDPR tasks asynchronously

Granularity of Rights

Per data item
Support right for every piece of data

Per service/~~person~~
Support rights at the level of service

Strict interpretation *that will benchmark worst-case performance overhead*

GDPR-Compliant Storage Systems

redis *Support for TTL and Timely Delete*

GDPR-Compliant Storage Systems

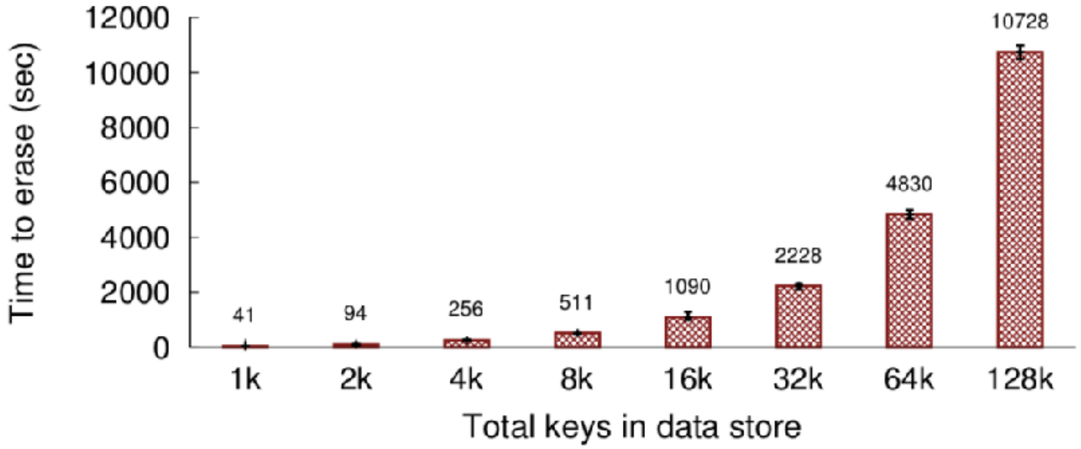
Support for TTL and Timely Delete

Redis has built-in support for TTLs but… it internally erases expired keys using a **lazy randomized algorithm**

GDPR-Compliant Storage Systems

Support for TTL and Timely Delete

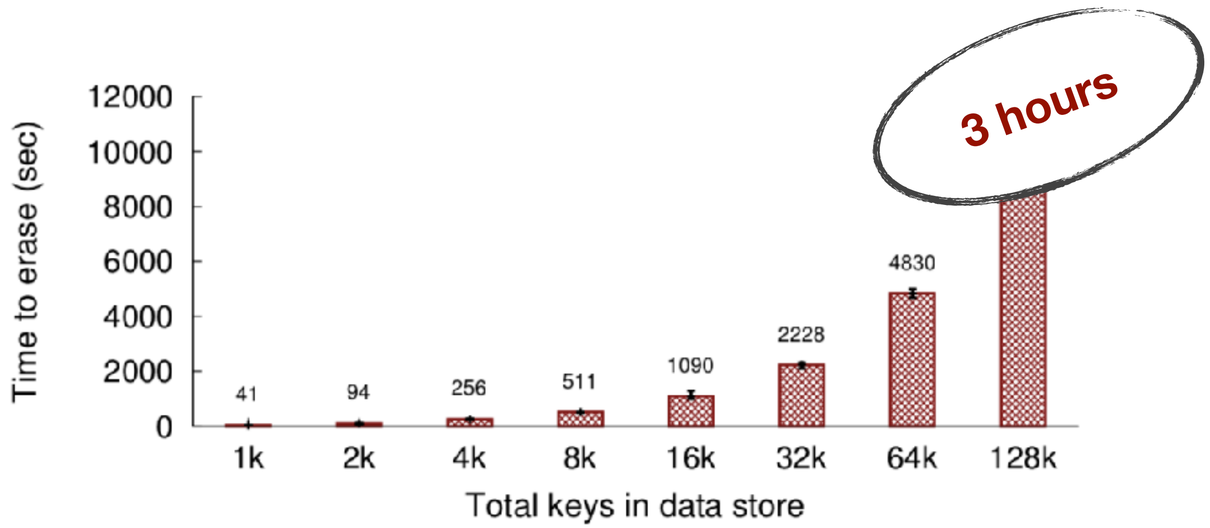
Redis has built-in support for TTLs but… it internally erases expired keys using a **lazy randomized algorithm**



GDPR-Compliant Storage Systems

Support for TTL and Timely Delete

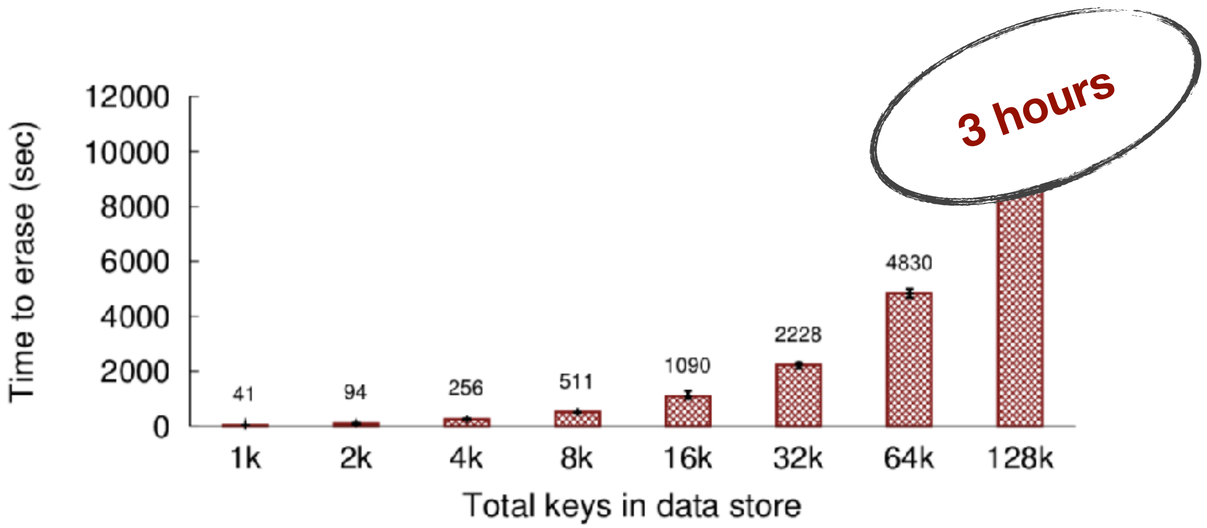
Redis has built-in support for TTLs but… it internally erases expired keys using a **lazy randomized algorithm**

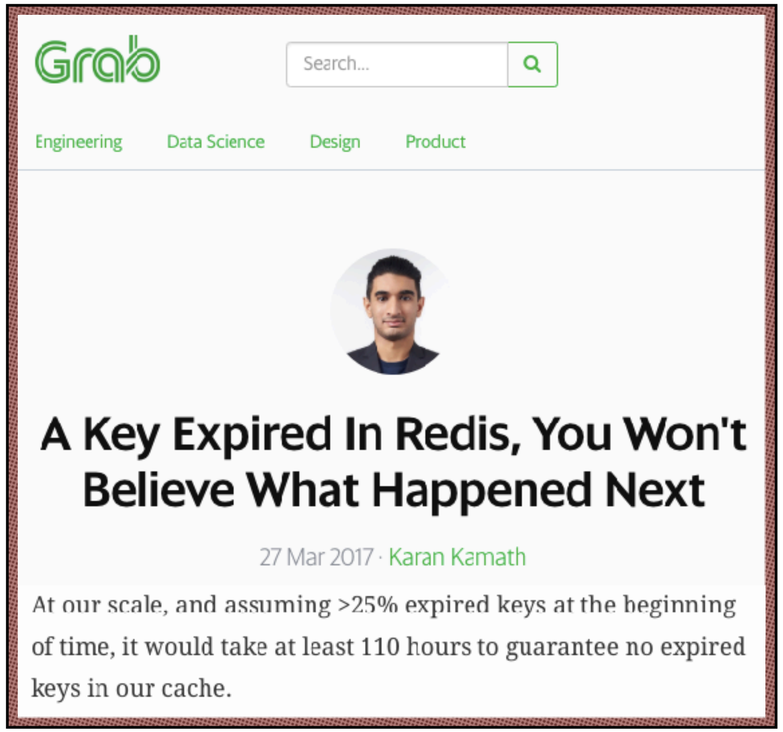


GDPR-Compliant Storage Systems

redis *Support for TTL and Timely Delete*

Redis has built-in support for TTLs but… it internally erases expired keys using a **lazy randomized algorithm**

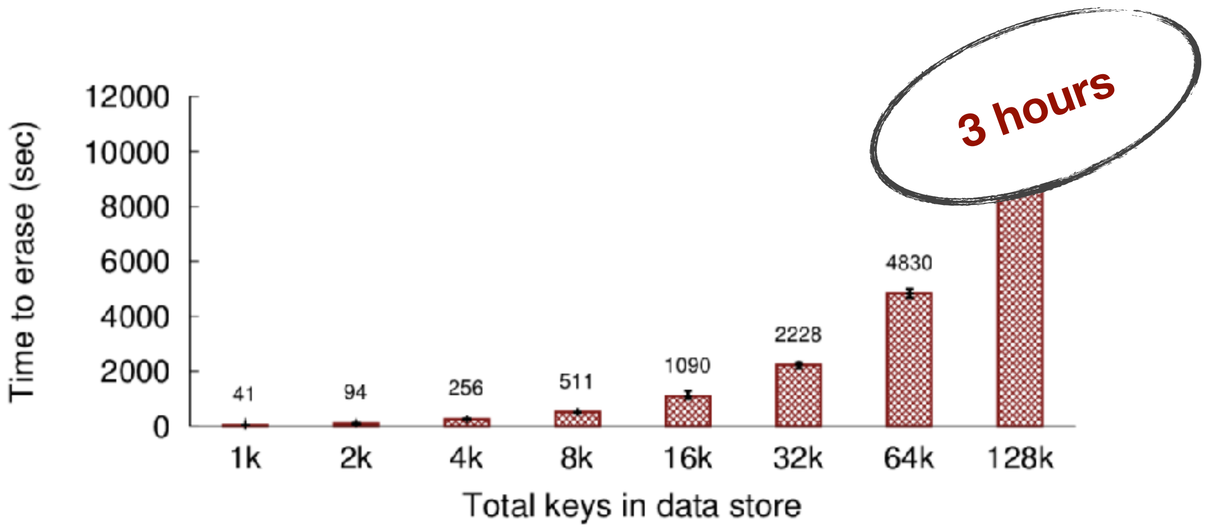




GDPR-Compliant Storage Systems

redis *Support for TTL and Timely Delete*

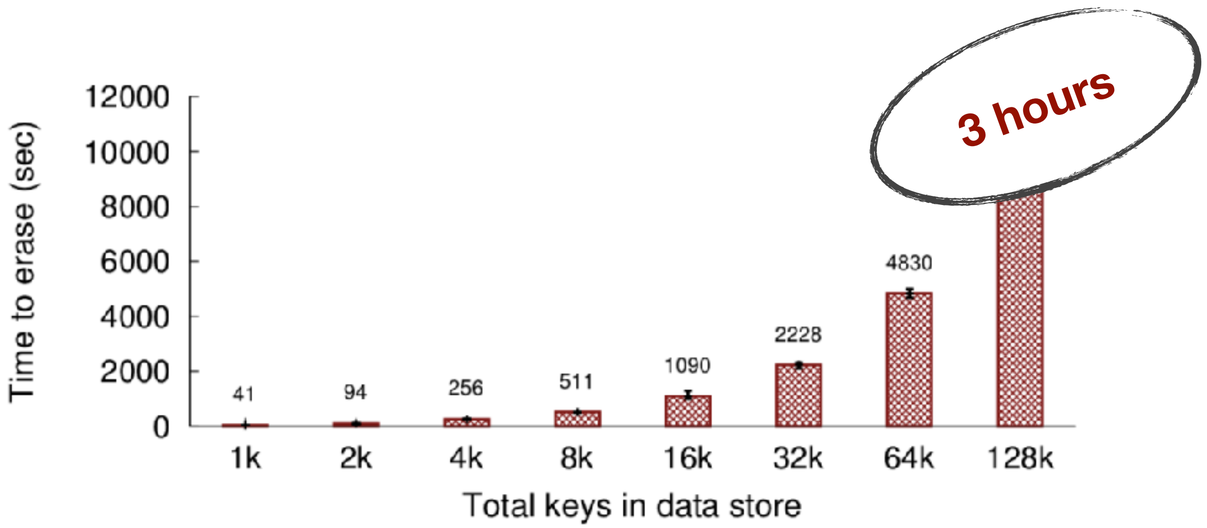
Redis has built-in support for TTLs but… it internally erases expired keys using a **lazy randomized algorithm**

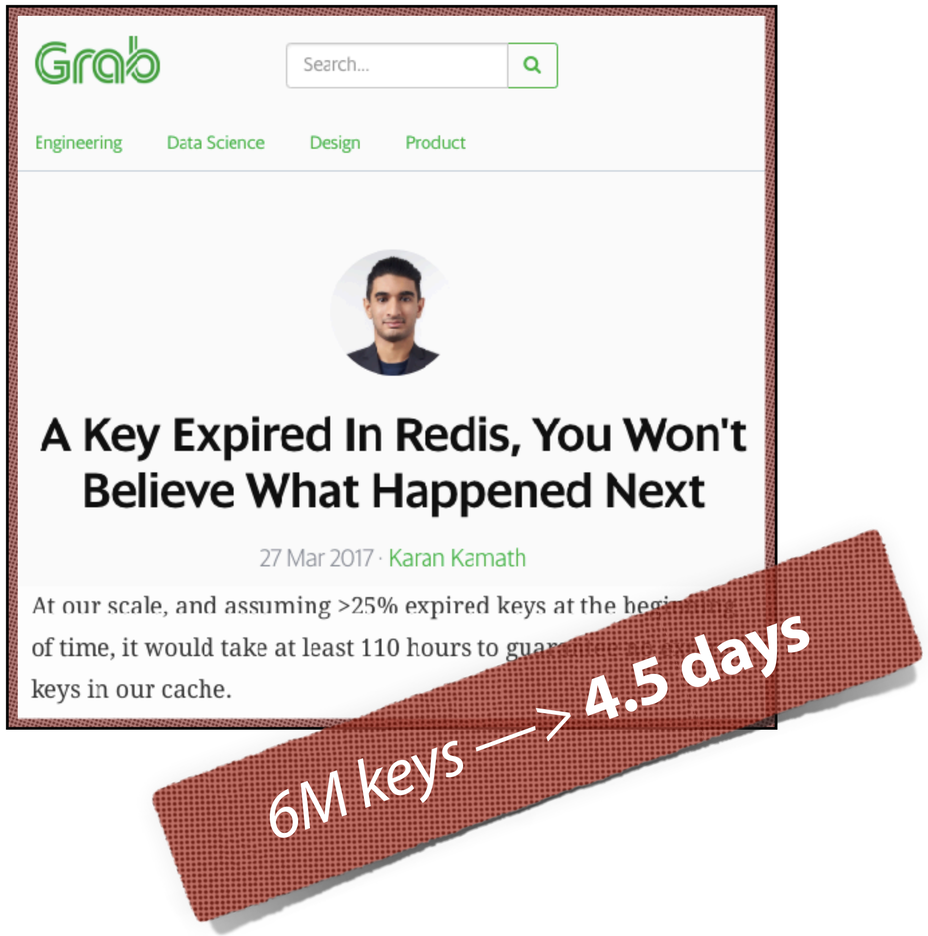


GDPR-Compliant Storage Systems

redis *Support for TTL and Timely Delete*

Redis has built-in support for TTLs but… it internally erases expired keys using a **lazy randomized algorithm**





Code change: we changed the expiry algorithm to be deterministic