.local
Stockholm

Implementing Right to Erasure with CSFLE

Pierre Petersson
Advisory Solutions Architect MongoDB

Tom McCarthy
Senior Solutions Architect MongoDB

€4.4b

€4.4b

Cumulative sum of GDPR fines as of October 2023

Source: https://www.enforcementtracker.com/

Agenda

Intro on right to erasure

The role of encryption

Simplifying it with MongoDB

Next generation of in use encryption

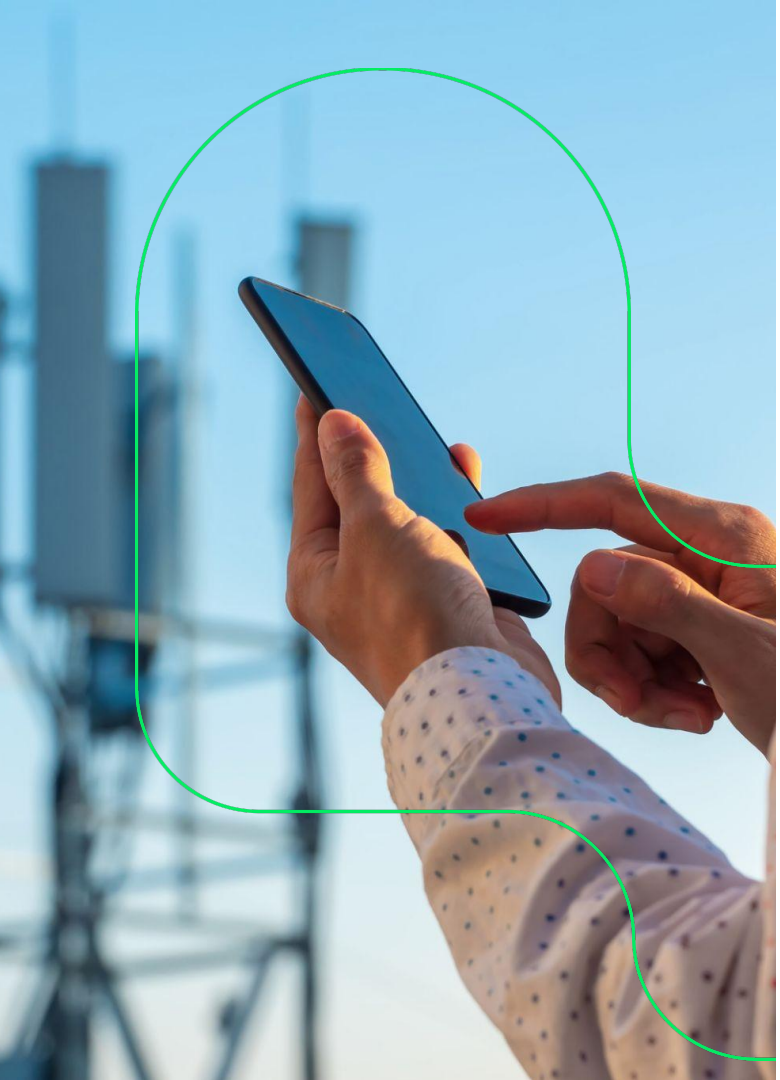
Tom McCarthy

Sr Solutions Architect

- Cloud expert
- Python ~~zealot~~ fan
- Developer experience advocate
- Definitely not a lawyer

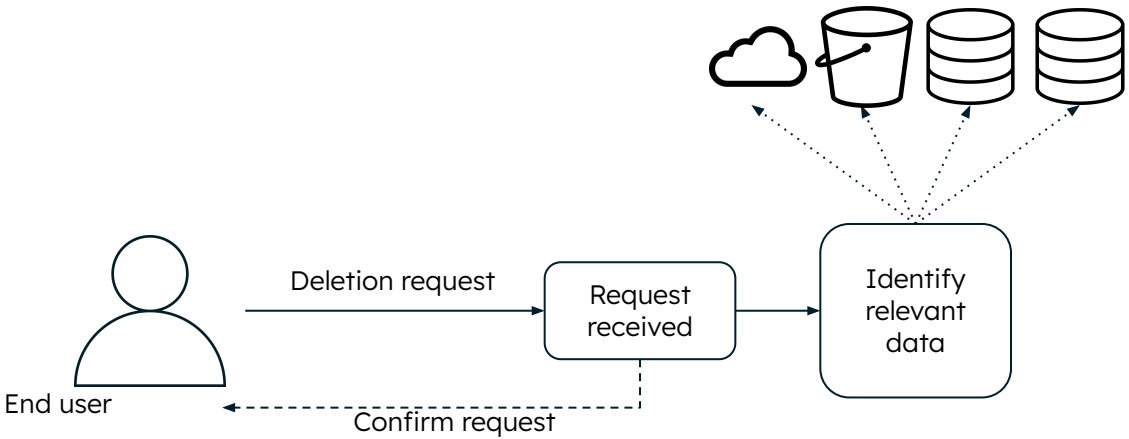
Right to Erasure

(the right to be forgotten)

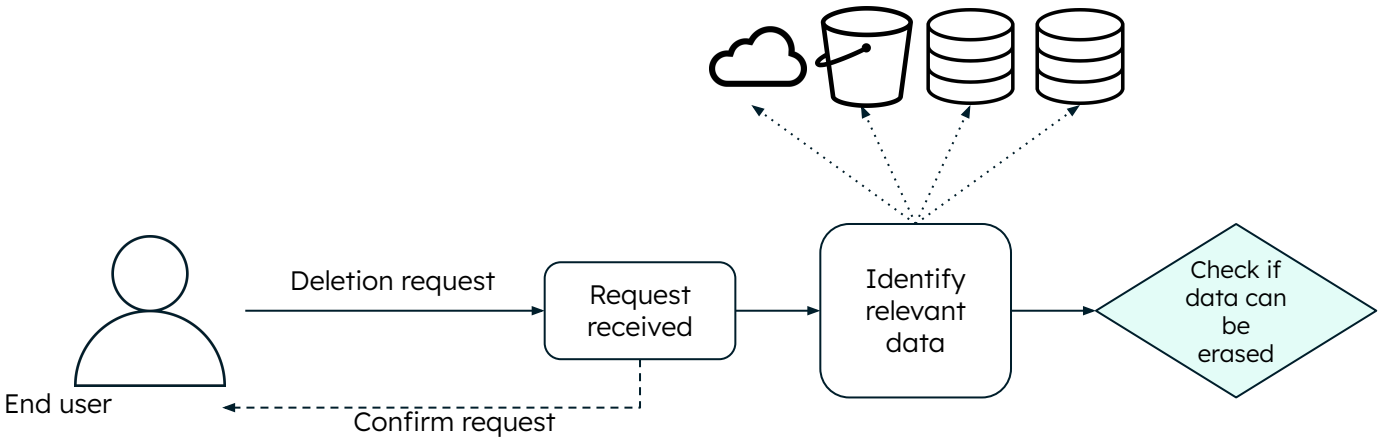


As an end customer or individual, I can request that a company delete all personal data they have about me.

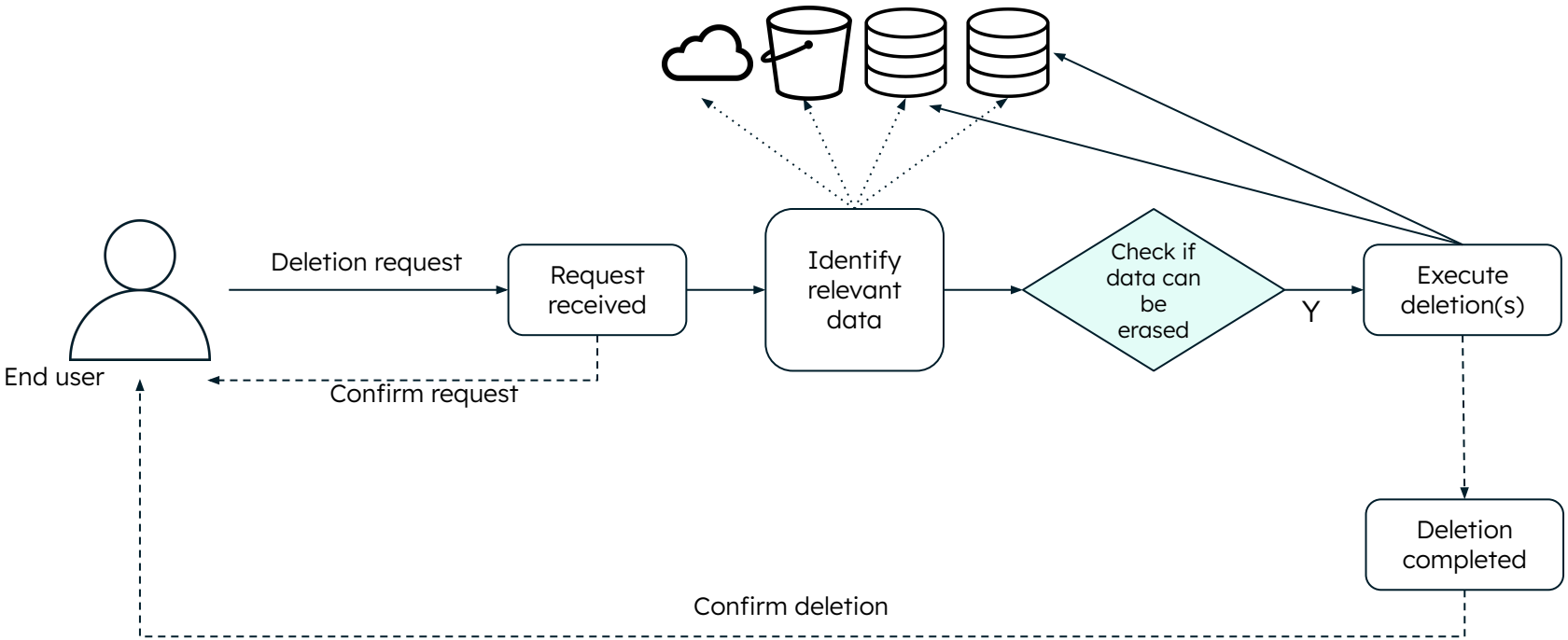
What does it mean for your business?



What does it mean for your business?



What does it mean for your business?



The role of encryption

What is Crypto Shredding

ssn: 901-01-001


```
{
    "customerID": "C12345",
    "ssn": "901-01-0001",
    "firstName": "John",
    "lastName": "Doe",
    "email": "johndoe@email.com",
    "address": {
        "city": "Springfield",
        "state": "IL"
    },
    "dateJoined": "2023-10-19"
}
```



```
{
    "customerID": "C12345",
    "ssn": "xxxxxxxxxx",
    "firstName": "xxxxxxxxxx",
    "lastName": "xxxxxxxxxx",
    "email": "xxxxxxxxxx",
    "address": {
        "city": "xxxxxxxxxx",
        "state": "xxxxxxxxxx"
    },
    "dateJoined": "2023-10-19"
}
```



```
{
    "customerID": "C12345",
    "ssn": "901-01-0001",
    "firstName": "John",
    "lastName": "Doe",
    "email": "johndoe@email.com",
    "address": {
        "city": "Springfield",
        "state": "IL"
    },
    "dateJoined": "2023-10-19"
}
```


User registration

Secure Key Management

Data Encryption

Customer Encryption Key

Reading data

ssn: 801-01-001


```
{
    "customerID": "D52345",
    "ssn": "801-01-0001",
    "firstName": "Adam",
    "lastName": "Doe",
    "email": "adamdoe@email.com",
    "address": {
        "city": "New York",
        "state": "NYC"
    },
    "dateJoined": "2023-10-20"
}
```



```
{
    "customerID": "D52345",
    "ssn": "xxxxxxxxxx",
    "firstName": "xxxxxxxxxx",
    "lastName": "xxxxxxxxxx",
    "email": "xxxxxxxxxx",
    "address": {
        "city": "xxxxxxxxxx",
        "state": "xxxxxxxxxx"
    },
    "dateJoined": "2023-10-20"
}
```



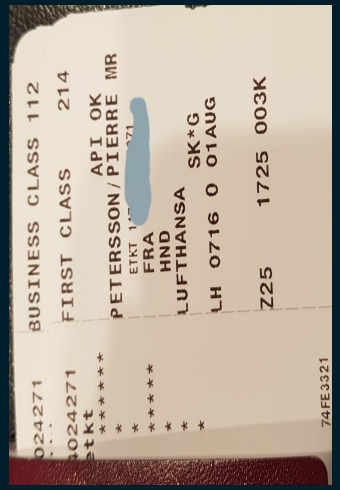
```
{
    "customerID": "D52345",
    "ssn": "801-01-0001",
    "firstName": "Adam",
    "lastName": "Doe",
    "email": "adamdoe@email.com",
    "address": {
        "city": "New York",
        "state": "NYC"
    },
    "dateJoined": "2023-10-20"
}
```

What is Crypto Shredding

ssn: 901-01-001

User registration | Secure Key Management | Data Encryption | Shredding the key | Proof of unreadability

ssn: 801-01-001



Father, Husband, Coder, Air mileage collector

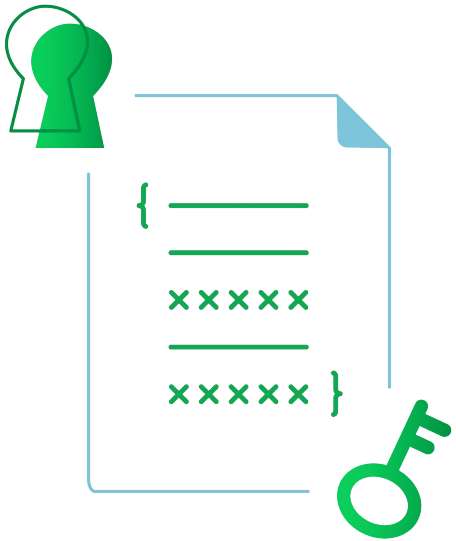
Pierre Petersson

Advisory solutions Architect MongoDB

20+ Years
Public Speaker
ex-Amazonian

SCAN ME

Simplifying it with MongoDB

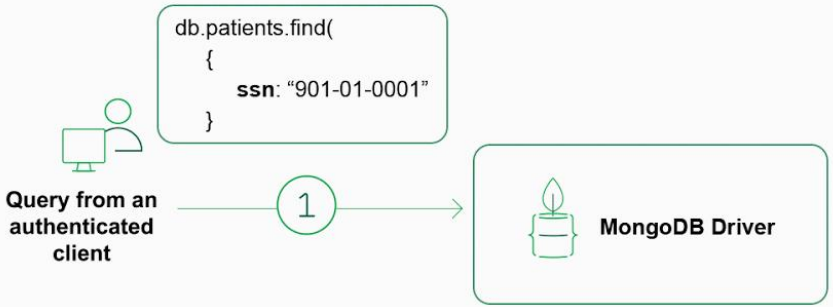


MongoDB Client Side FLE gives you a **safe and simple approach**

Select an encryption key, configure the fields to be encrypted in the MongoDB driver..and **GO**

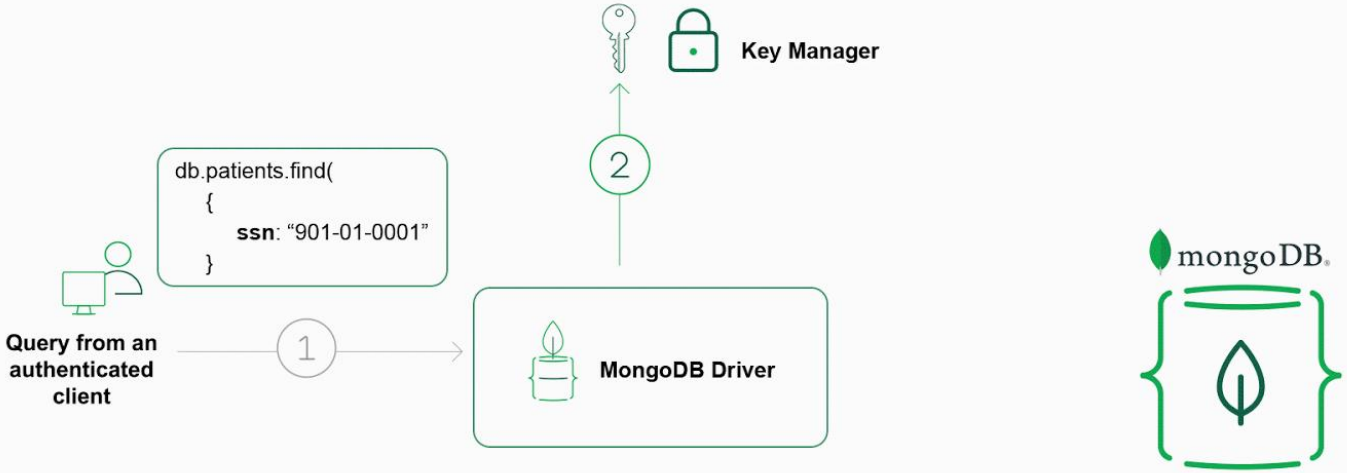
- Sensitive data **never leaves the application** without first being encrypted. Data **remains encrypted** server-side
- No need to **modify applications**
- Encrypted data is **still queryable**
- **Minimal performance impact** to the database and the application
- **Reduce Cognitive load,** intuitive and easy for developers to configure and setup

Query a document with CSFLE enabled



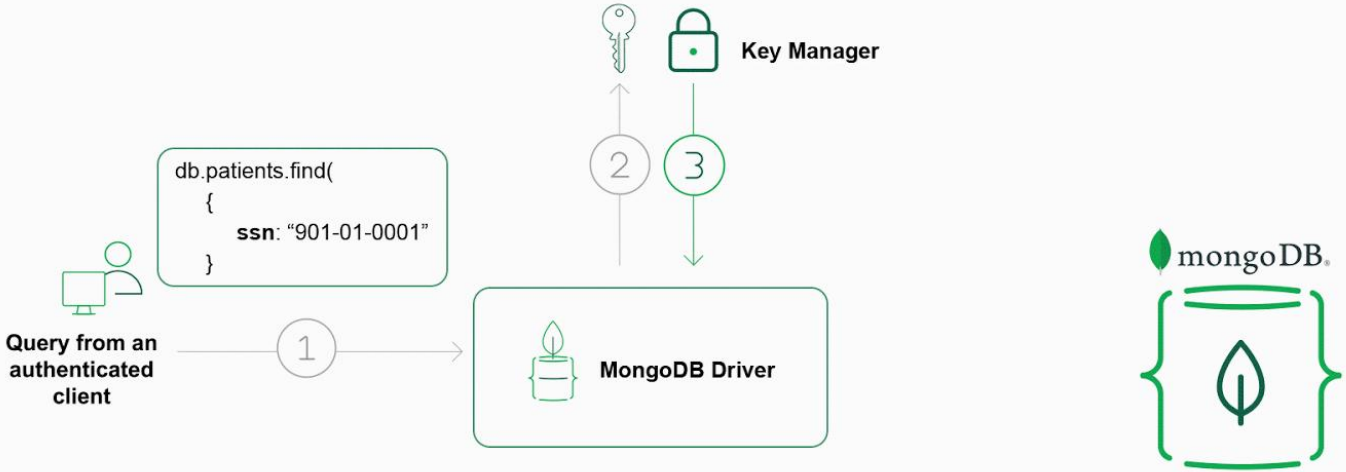
Upon receiving a query, the MongoDB driver checks to see if any encrypted fields are involved

Query a document with CSFLE enabled



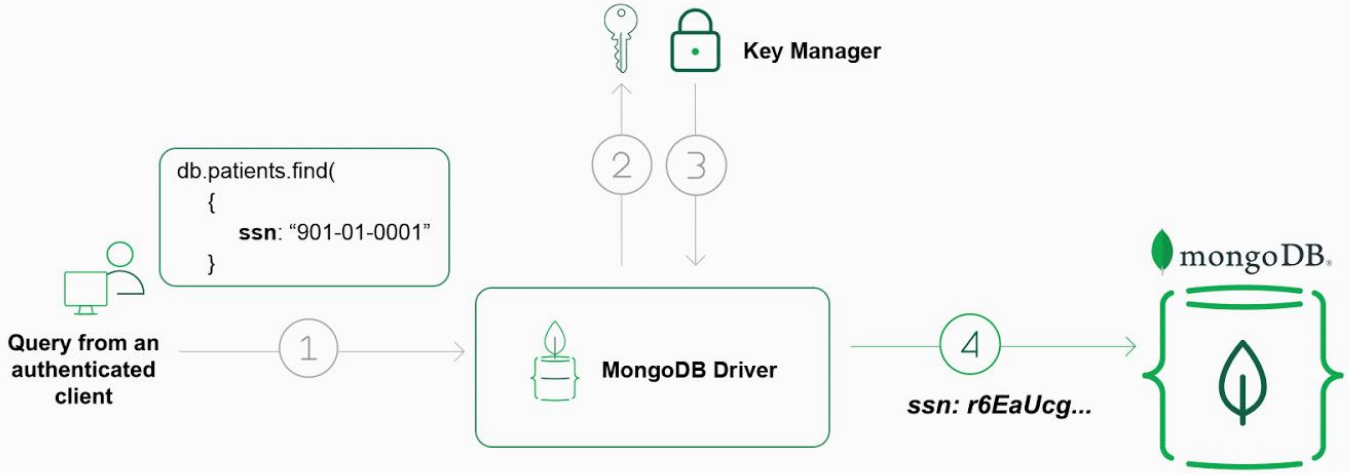
The driver requests the envelope encryption key from the KMS key provider

Query a document with CSFLE enabled



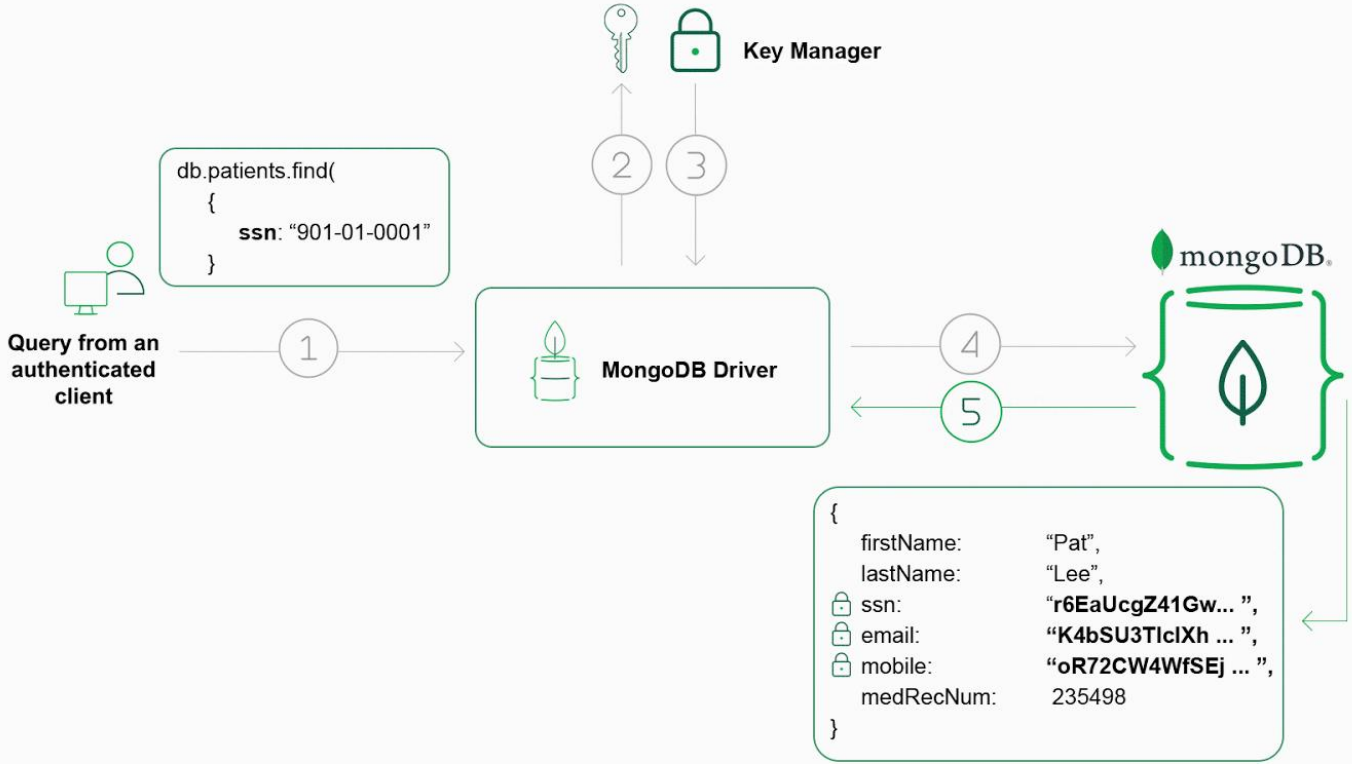
The driver decrypts the data encryptions keys using the envelope key, which then encrypts the sensitive fields

Query a document with CSFLE enabled



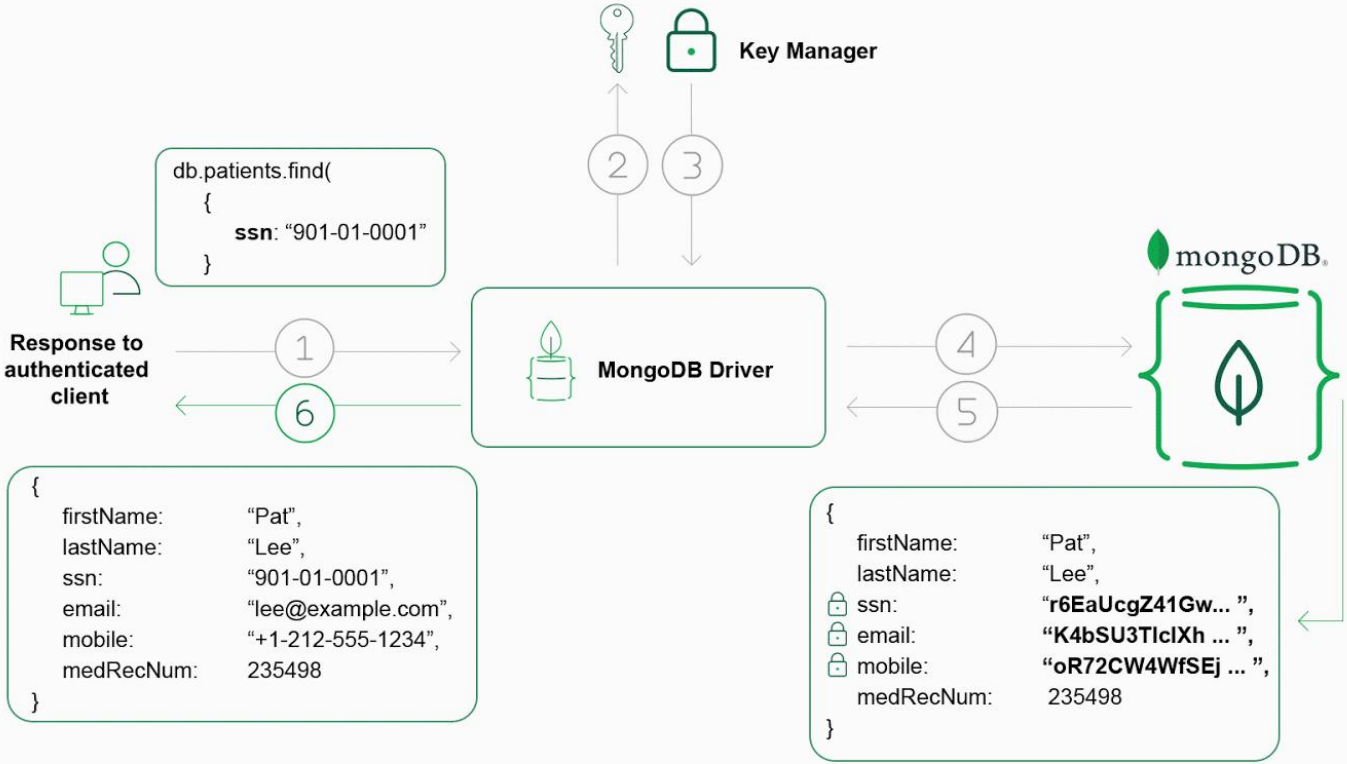
The driver submits the query to the MongoDB server with the encrypted fields rendered as ciphertext

Query a document with CSFLE enabled

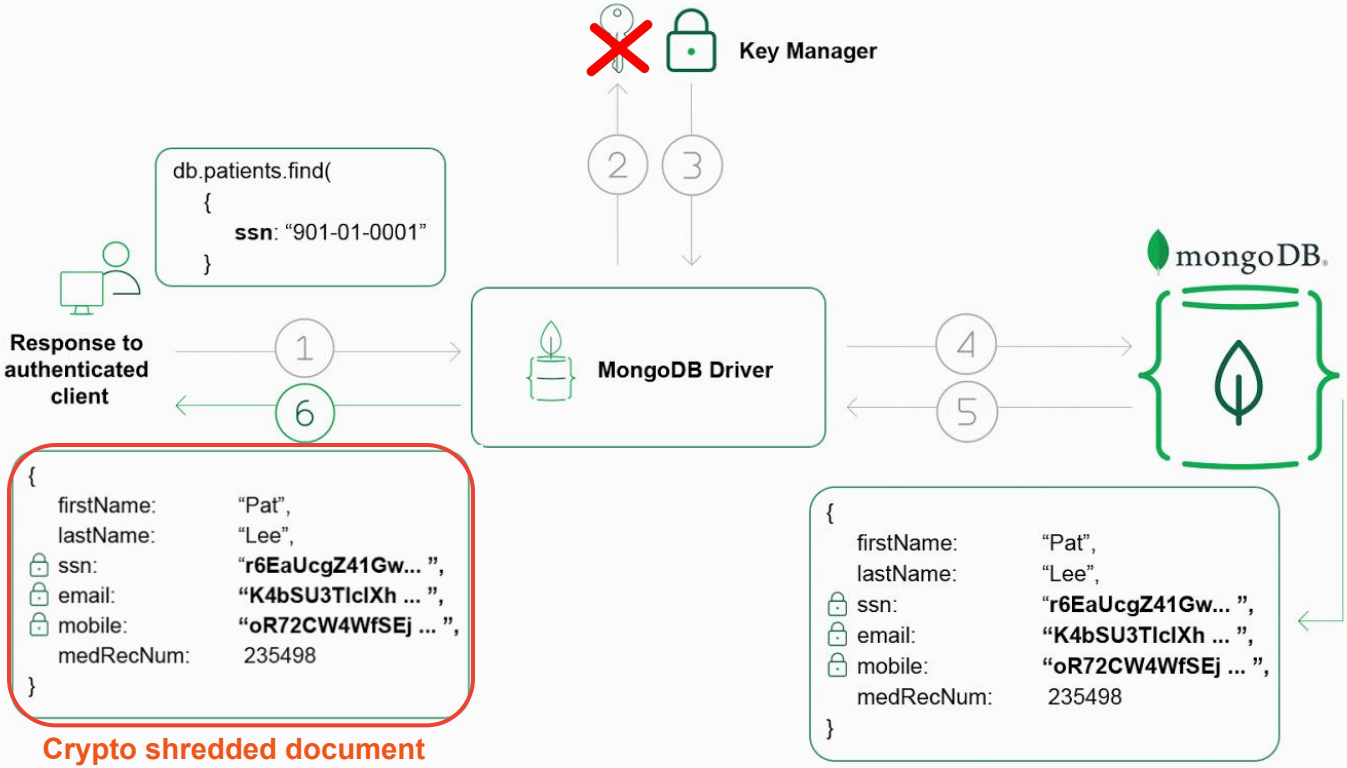


MongoDB returns the encrypted results of the query to the driver

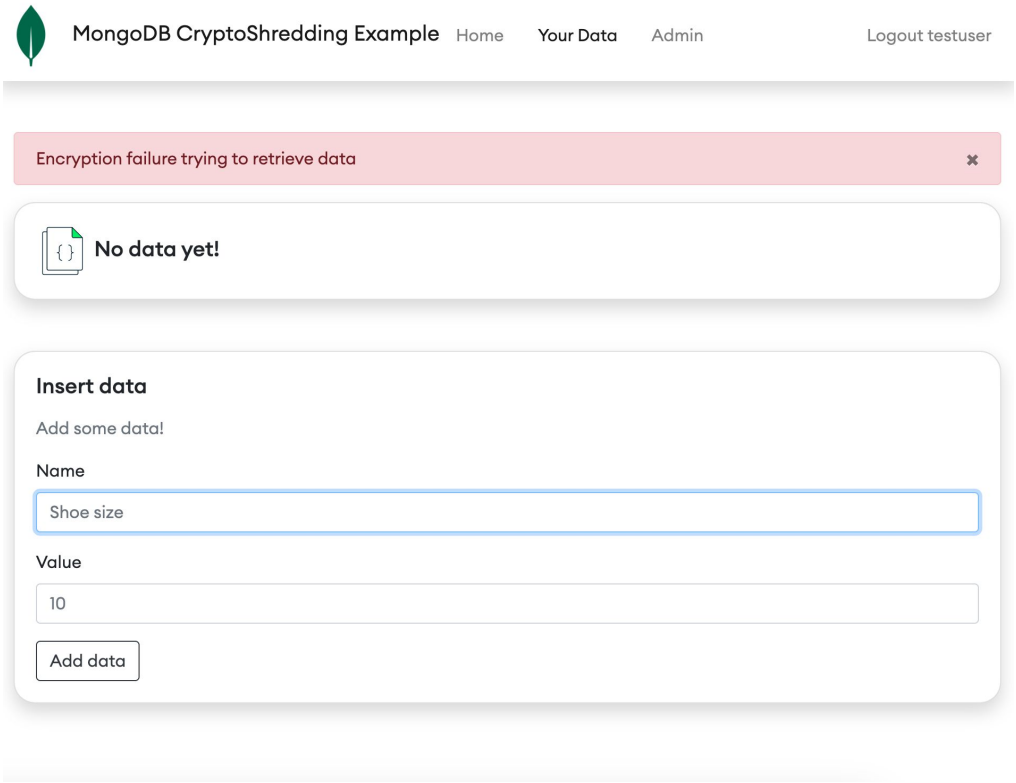
Query a document with CSFLE enabled

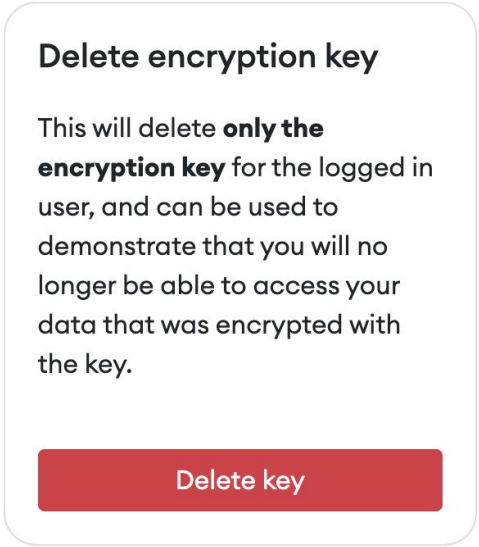


CSFLE Right for erasure by removing encryption key



Let's See it In Action!





https://github.com/to-mc/flask-mongodb-cryptoshredding-example

Next generation of in use encryption: Queryable encryption

Use Cases

Industry: Financial Services
Bank application needs to find transactions using a **range of dates** or **dollar amounts** for fraud detection

Industry: Human Resources
HR system allows searching for employees by the **last 4 digits** of their social security number

Industry: Health Care
Customer support agents needs to find patient records by searching for the **first few characters** of their name

CSFLE vs Queryable Encryption Trade-offs

	Inserts	Find (equality)	Find (range, prefix, suffix, substring)	Storage Overhead	Frequency Leakage
FLE	Fast	Fast	No	Minimal	Possibly
Queryable Encryption	Slower	Fast	Yes	Yes	None

https://www.mongodb.com/collateral/queryable-encryption-technical-paper

Client-Side FLE / In-use encryption

Design Goal

Provide some of the strongest levels of data privacy and security for regulated workloads

- Simplify compliance with modern privacy regs: “Right to Erasure”
- End to end protection of your most sensitive data
- Increase your confidence in moving to managed services in the cloud
- Enable querying on encrypted data, in a secure, fast and scalable way

Additional Resources

Resources CSFLE
Client-Side Field Level Encryption The Next Generation of Privacy & Security MDBW22 Video

Whitepaper CSFLE

CSFLE Multi Cloud Environments MDBW22 Video

Queryable Encryption Resources
Docs (very much a WIP): https://www.mongodb.com/docs/upcoming/core/queryable-encryption/
Blog post: https://www.mongodb.com/blog/post/mongodb-releases-queryable-encryption-preview
Product page & FAQ: https://mongodb.com/products/queryable-encryption

Achieving Scalable Data Intelligence with BigID & MongoDB

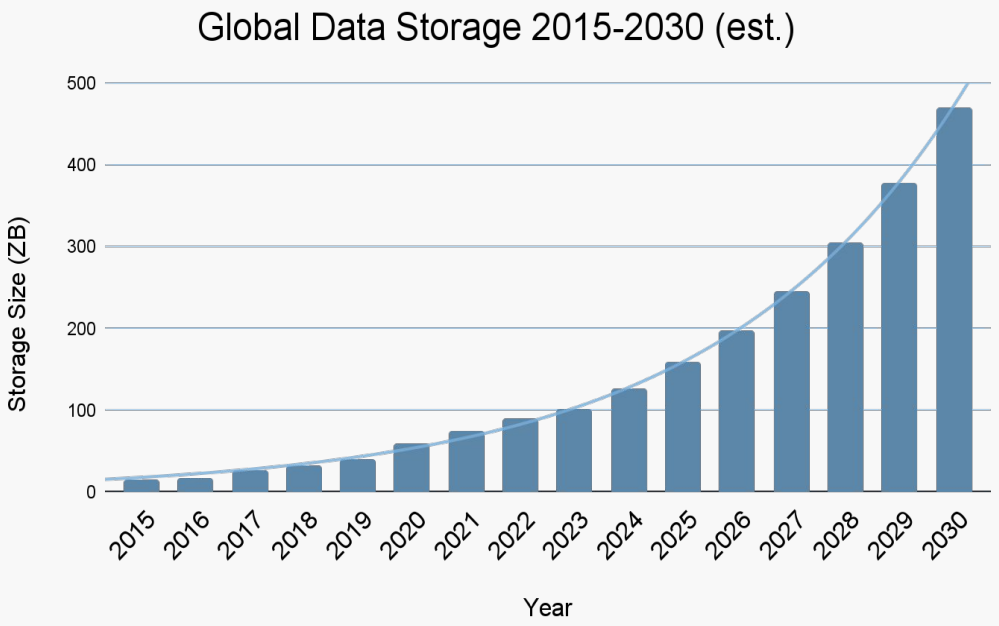
Andreas Holmström

Sr. Solutions Engineer - BigID

MongoDB World Tour
.local STHLM

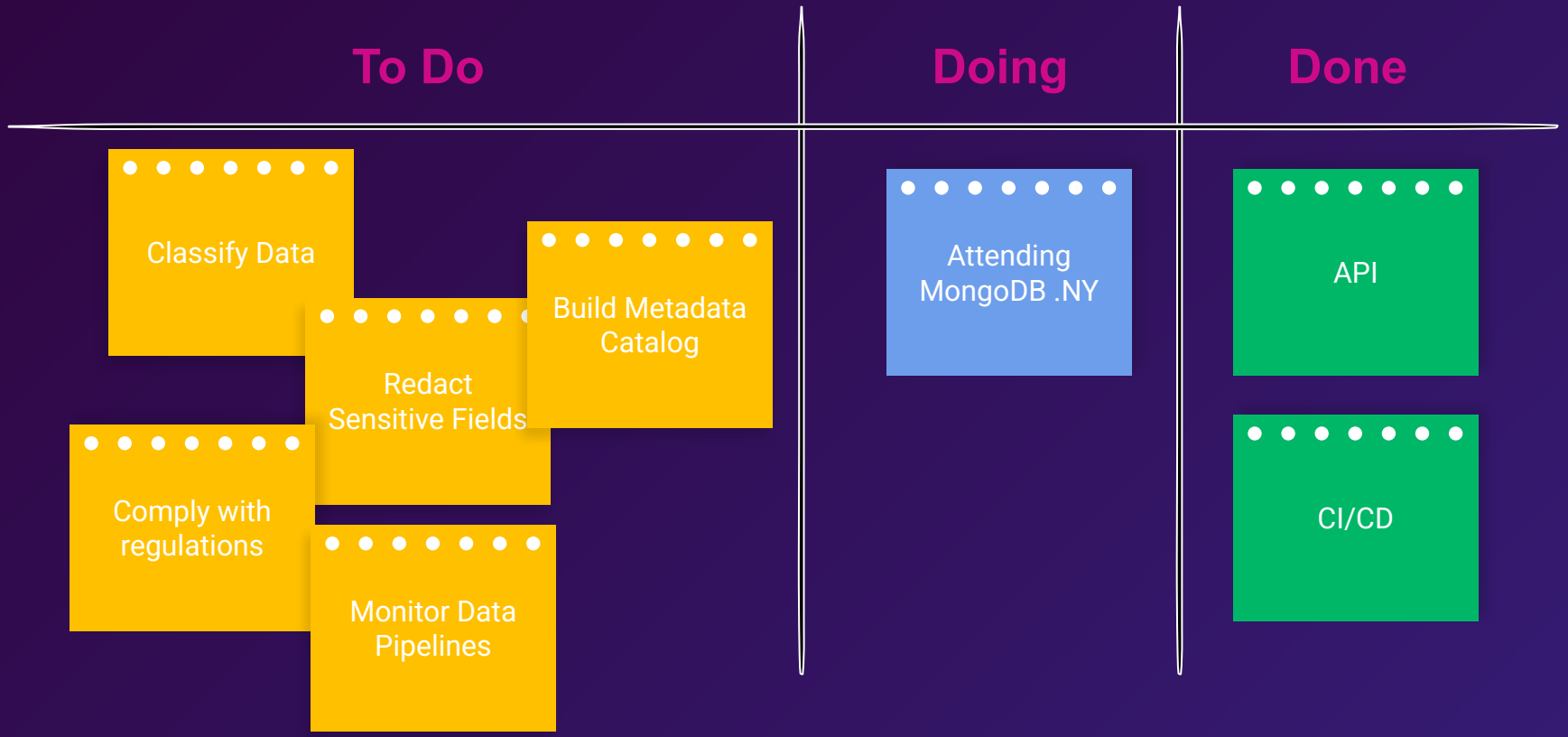
“Data is a precious thing and will last longer than the systems themselves”

- Tim Berners Lee



BigID

The Problem



BigID

We'll discuss

How to reduce risk and gain visibility to your data using an API, dev first approach

- The Problem - You have lots of data. You need to control it.
- The Solution - Data Discovery & Classification
- The Method - Shift Left ←

BigID

BigID and MongoDB

- **Data Intelligence Platform for Privacy, Security & Governance**
- **Using MongoDB since 2016:**
  - BigID's internal database (containerized & via Atlas)
  - Connectors to scan, classify and action data in MongoDB
- **Strong Partnership**
- **Come meet us at our stand!**

BigID & MongoDB

How MongoDB powers BigID's Scalable Data Intelligence Platform

Flexible Data Model

- Component of BigID Data Catalog
- Avoid "Migration Script"
- Use of MongoDB as both "Documents Store" and "Full DB" with aggregations, joins and transactions

Availability

- High availability through the automated distribution of data across multiple servers
- Data remains accessible even in the event of hardware failures or other disruptions within a cluster

Scalability

- Vertical scaling based on CPU, Memory, and Disk utilization
- Horizontal scaling enabled using sharding and replica sets
- All managed automatically by Atlas

Support for KMS

- Encrypt data at rest using a cloud providers Key Management Service
- Support for Bring-your-own-key with customers for enhanced management capability

BigID

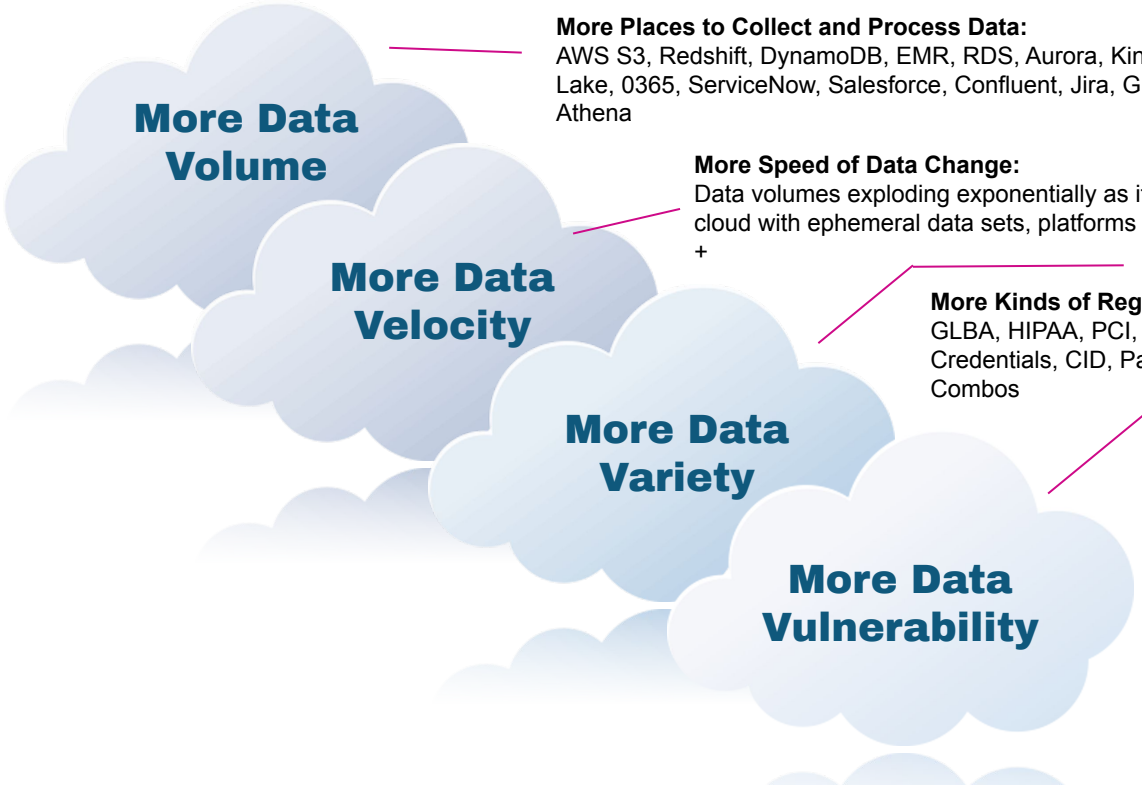
Know Your Data

Data Discovery & Classification

With great ~~power~~ *data* comes great responsibility

Know your data

This was Hard in Data Center, Harder in the Cloud



More Places to Collect and Process Data:
AWS S3, Redshift, DynamoDB, EMR, RDS, Aurora, Kinesis, Databricks, Snowflake, Azure Data Lake, 0365, ServiceNow, Salesforce, Confluent, Jira, GitHub, GitLab, Cohesity, Confluence, Athena

More Speed of Data Change:
Data volumes exploding exponentially as it becomes easier to add, copy, move in cloud with ephemeral data sets, platforms like snowflake and databricks, email, files +

More Kinds of Regulated, Sensitive, Critical Data To Track
GLBA, HIPAA, PCI, SOX, GDPR, CCPA, PII, PI, PCI, Credentials, CID, Patents, Research, ML Models, Data Combos

How do you measure & remediate risk:
open access, excessive privileges, duplicate data, sovereign data, misconfigurations, ROT data, credential leaks, embedded code

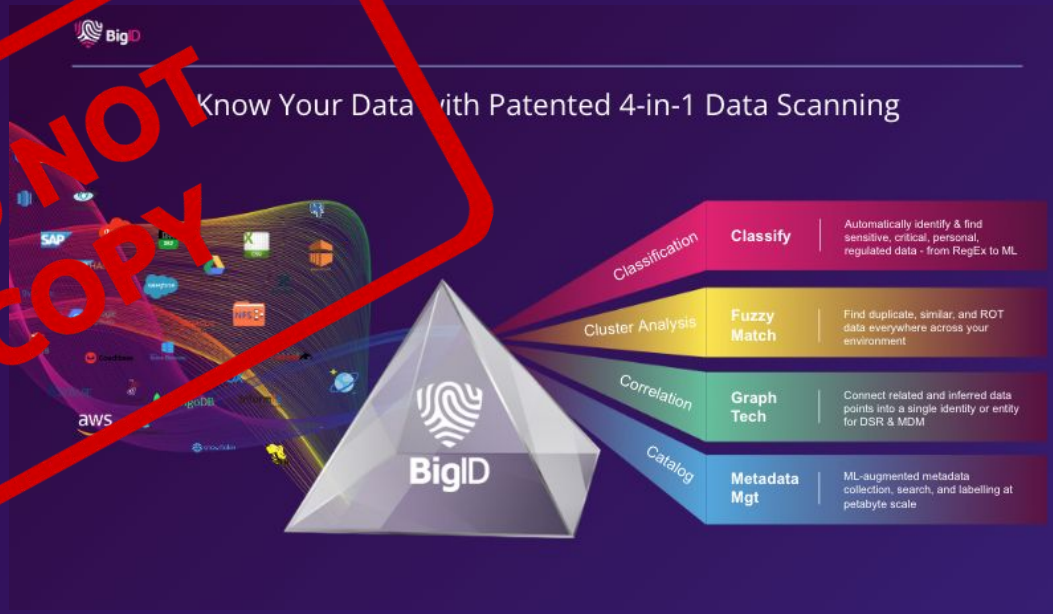
Data Discovery & Classification

Map & Connect

Enumerate & Scan

Process & Analyze

Report & Query



BigID

API & Dev First

Shift←Left

Shift Left
Breakdown

Definition

Practice of doing X earlier in the process or value chain

E.g: Automatic QA Testing; Code Vulnerability Scans

In Practice

- Integrate & automate as code and data are created
- Extending new & existing tools
- Applies to more fields: Privacy, Compliance, etc

The Benefits

- Addressing issues before they reach Production
- Reduce risks and costs
- Speed up dev efficiency

Leverage tools with Open API

Plug into Data Pipelines

CI/CD

Examine The Code

Extend Capabilities

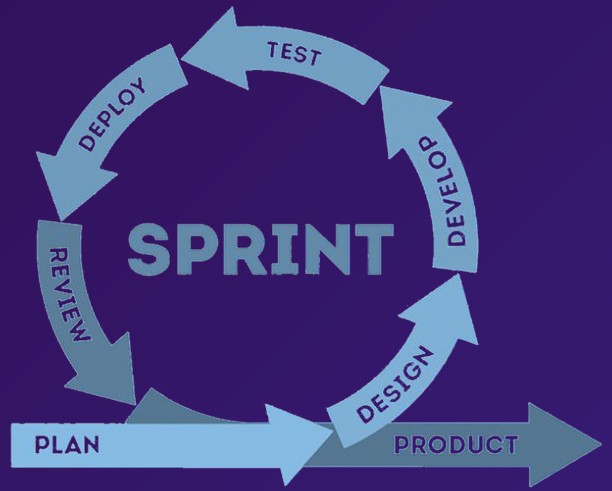
Part of your Code Lifecycle

Scanning Code Repositories (bitbucket, github, gitlab etc)

- Find sensitive data like passwords or certificates
- Detect customer data processed by apps / microservices
- Notify on PII without an approved purpose of processing
- Locate DB tables referenced in queries that use PII

CI/CD

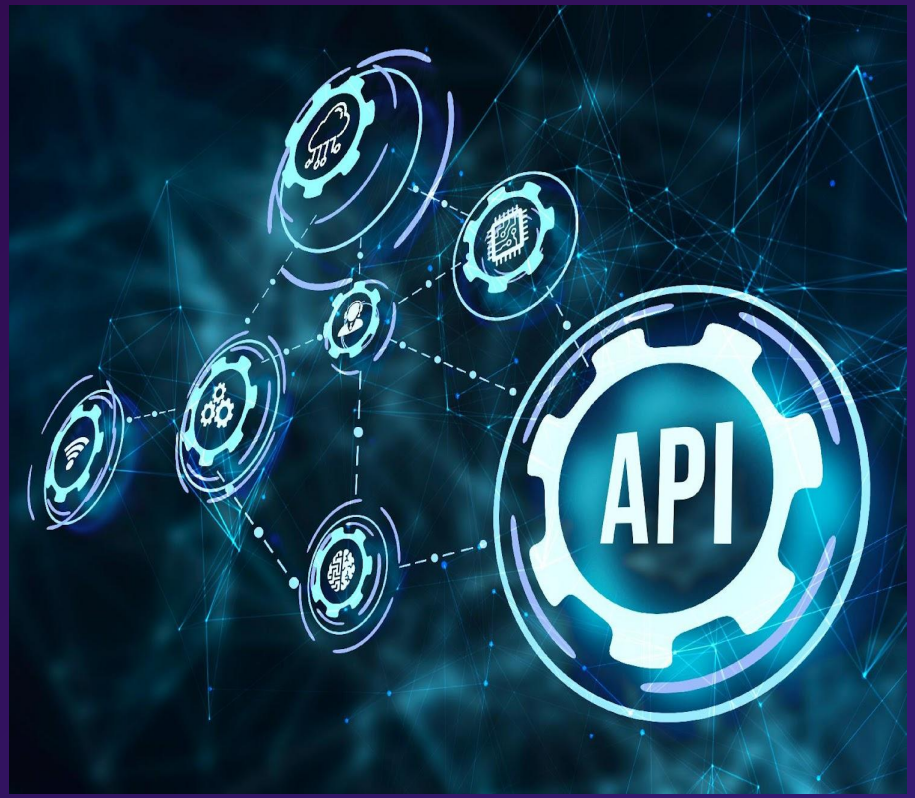
- Run as part of build (e.g. Jenkins)
- Monitor api gateway (e.g. nginx)



Open API

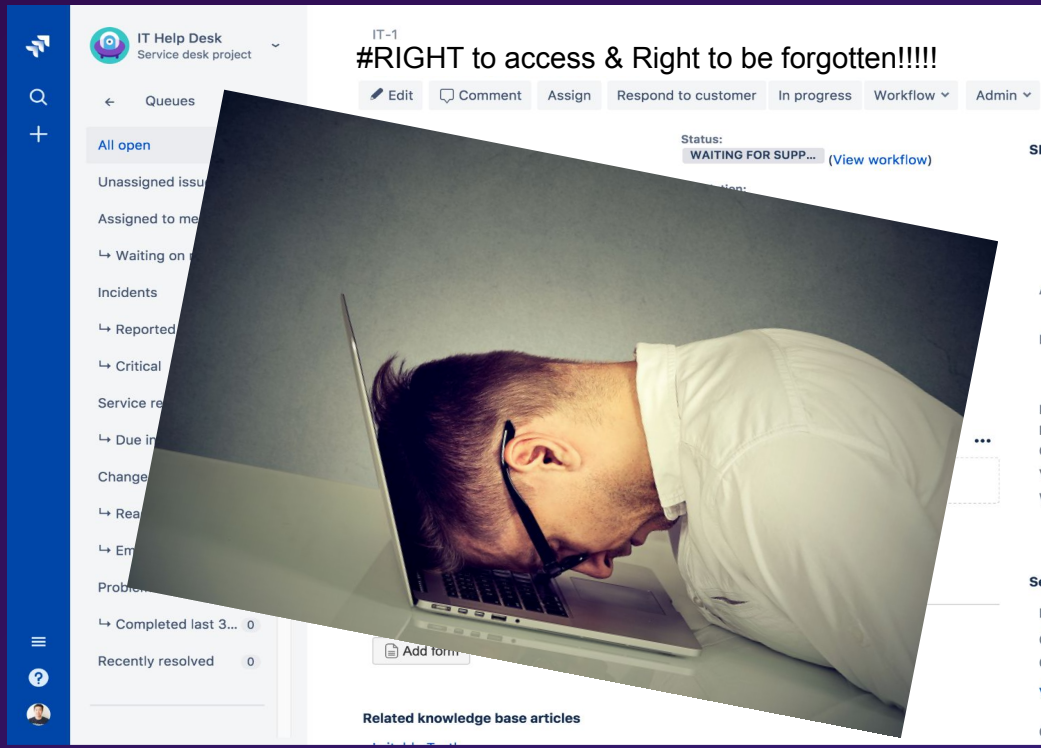
Developers can automate recurring tasks, integrate home-grown systems and extend capabilities better if they have tools with:

- Open, documented, comprehensive APIs
- Standard communication and authentication protocols
- Tech stack / language agnostic integrations



Privacy automation

- End to End user experience to manage, consent, access, deletion, rectification, completing requests in minutes or hours rather than many days or weeks!
- CSFLE deletion, much better as access is only required for one DB, which reduces the attack surface and speeds up the deletion process
- Allows auditor to track privacy requests!
- Limit internal tickets and manual workflows

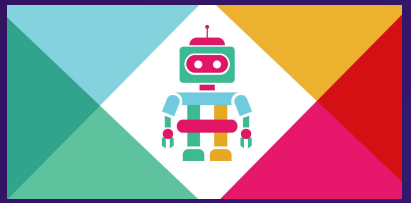


Extensibility

- A FRAMEWORK for developers
- Plugins (or apps) that allow
  - Adding new capabilities
  - Integrating multiple systems
- Industry Examples
  - Splunk add-ons
  - Kafka Connect
  - Slack apps & bots
  - Snowflake & DataBricks external Functions

splunkbase

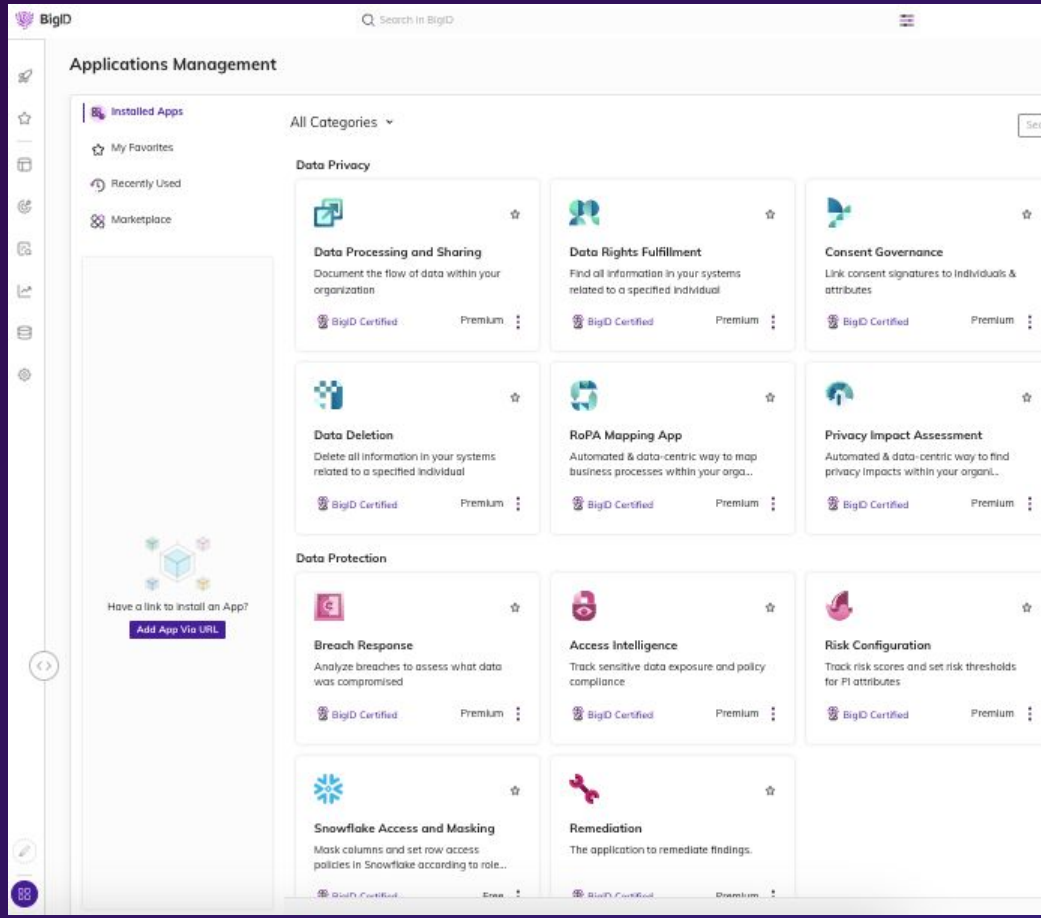
kafka



Application Framework SDK

A call for developers to:

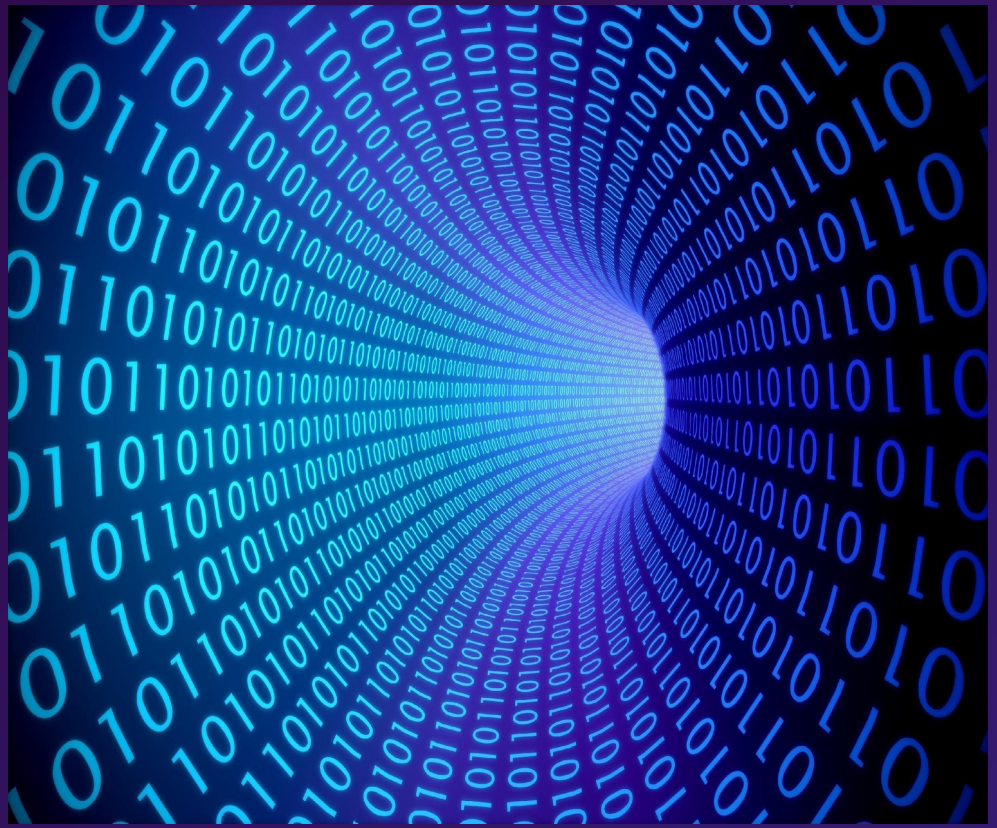
- Leverage the platform's REST APIs
  - Scan new data stores as they are created
  - Automate compliance
- Integrate it in pipelines
  - API GW + Data-in-Motion
  - Classification SDK
- Extend it
  - Connect to systems – connector SDK
  - Create new applications – APP FW SDK
- Find additional Apps and plugins on the BigID Marketplace



Data Pipeline

Plug into processes like:

- Data-in-motion: Data moves from one store to another
- Ingesting Data from Partners
- Data Sharing with 3rd Parties
- Processing, aggregation and analytics



BigID

Solutions and Examples

Please come speak to Us!