Deep Dive: How to Secure and Share Sensitive Information

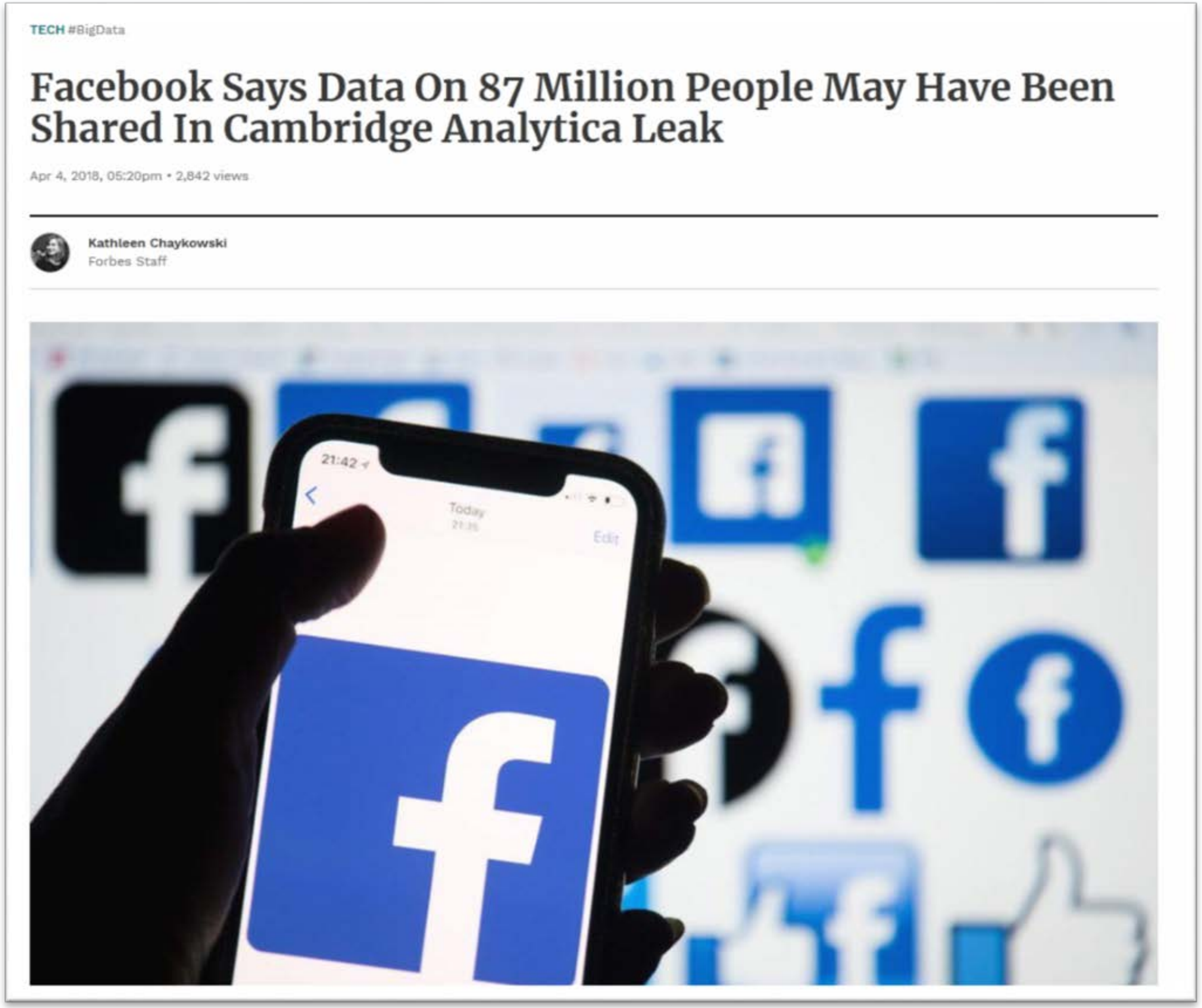
Caio Milani

Director, Product Management, MarkLogic
@cvbmilani

Charles Greer

Senior Staff Engineer, MarkLogic
@grechaw

Sharing Sensitive Information Gone Wrong



TECH #BigData

Facebook Says Data On 87 Million People May Have Been Shared In Cambridge Analytica Leak

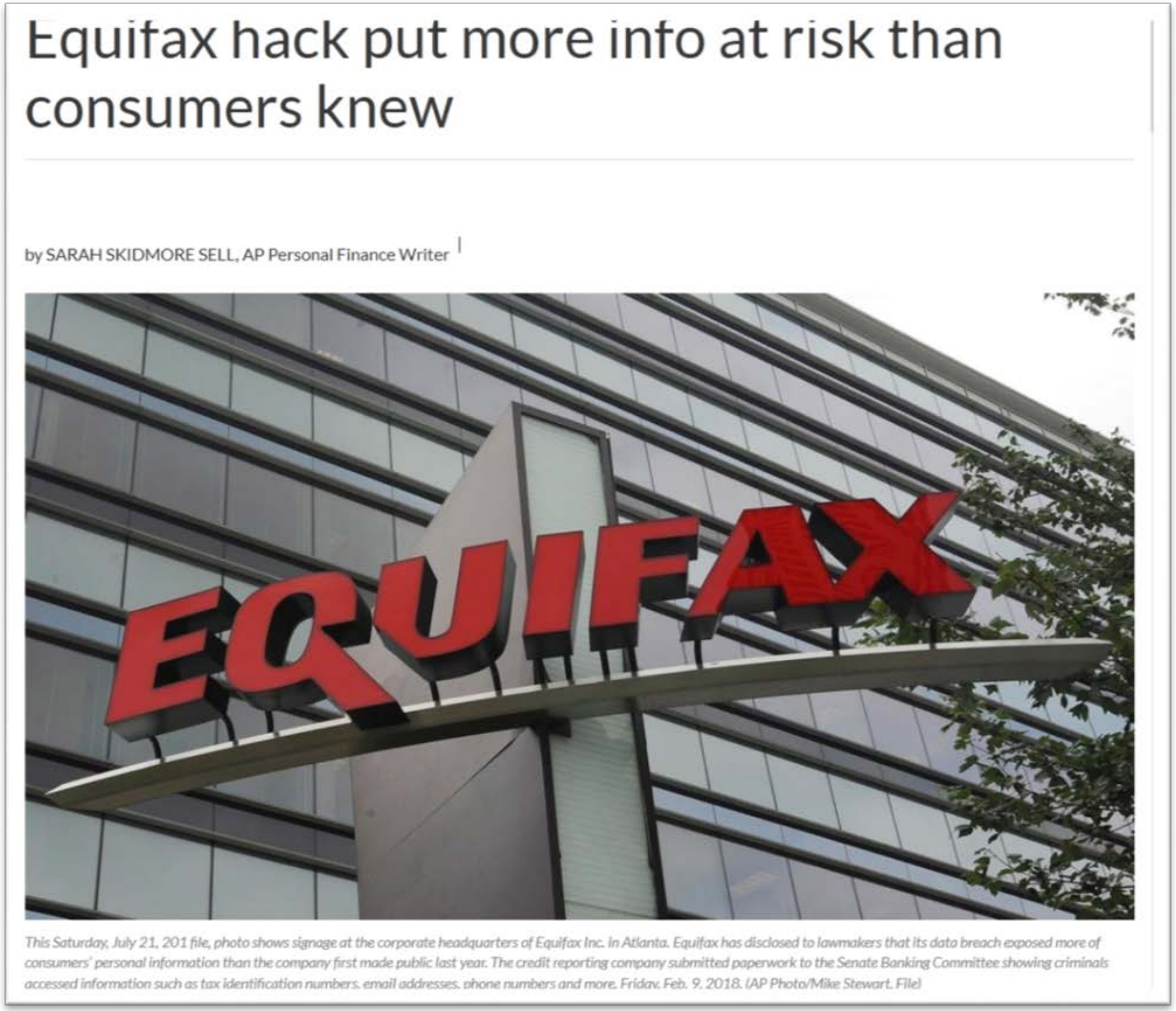
Apr 4, 2018, 05:20pm • 2,842 views

Kathleen Chaykowski
Forbes Staff

"…no systems were infiltrated, and no passwords or sensitive pieces of information were stolen or hacked"
-- Facebook

*"Among its privacy updates on Wednesday, Facebook said it will no longer allow people to **use phone numbers or email addresses** in its search tool to find other users on the social network."*
@Forbes

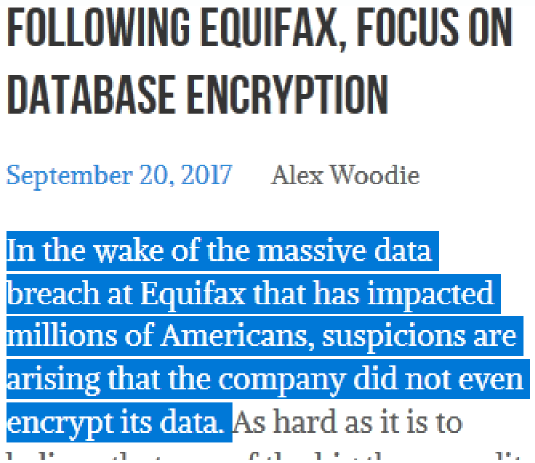
Securing Sensitive Information Failure



Equifax hack put more info at risk than consumers knew

by SARAH SKIDMORE SELL, AP Personal Finance Writer

This Saturday, July 21, 201 file, photo shows signage at the corporate headquarters of Equifax Inc. in Atlanta. Equifax has disclosed to lawmakers that its data breach exposed more of consumers' personal information than the company first made public last year. The credit reporting company submitted paperwork to the Senate Banking Committee showing criminals accessed information such as tax identification numbers, email addresses, phone numbers and more. Friday, Feb. 9, 2018. (AP Photo/Mike Stewart, File)

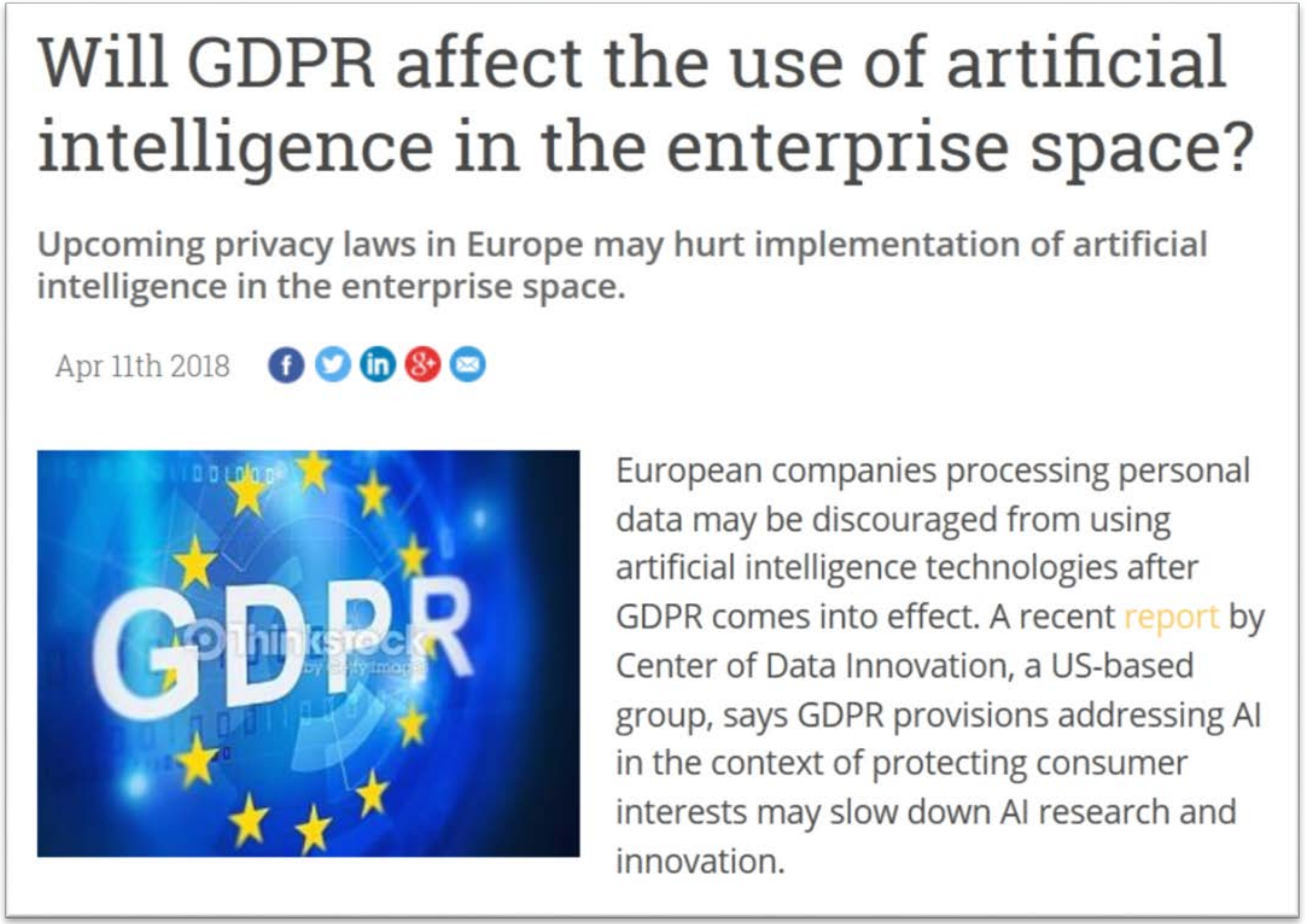


FOLLOWING EQUIFAX, FOCUS ON DATABASE ENCRYPTION

September 20, 2017 Alex Woodie

In the wake of the massive data breach at Equifax that has impacted millions of Americans, suspicions are arising that the company did not even encrypt its data. As hard as it is to

Sensitive Information Regulation



Will GDPR affect the use of artificial intelligence in the enterprise space?

Upcoming privacy laws in Europe may hurt implementation of artificial intelligence in the enterprise space.

Apr 11th 2018

European companies processing personal data may be discouraged from using artificial intelligence technologies after GDPR comes into effect. A recent report by Center of Data Innovation, a US-based group, says GDPR provisions addressing AI in the context of protecting consumer interests may slow down AI research and innovation.

Article 32 outlines methods to use in pursuit of data processing security, including anonymizing and encrypting personal data to ensure a level of security "appropriate to the risk."

COMPLEXITY OF DATA MANAGEMENT

Why Companies Fail

- Data Silos
- Multiple places to enforce policies
- Constant change in data models
- Multiple tools to understand

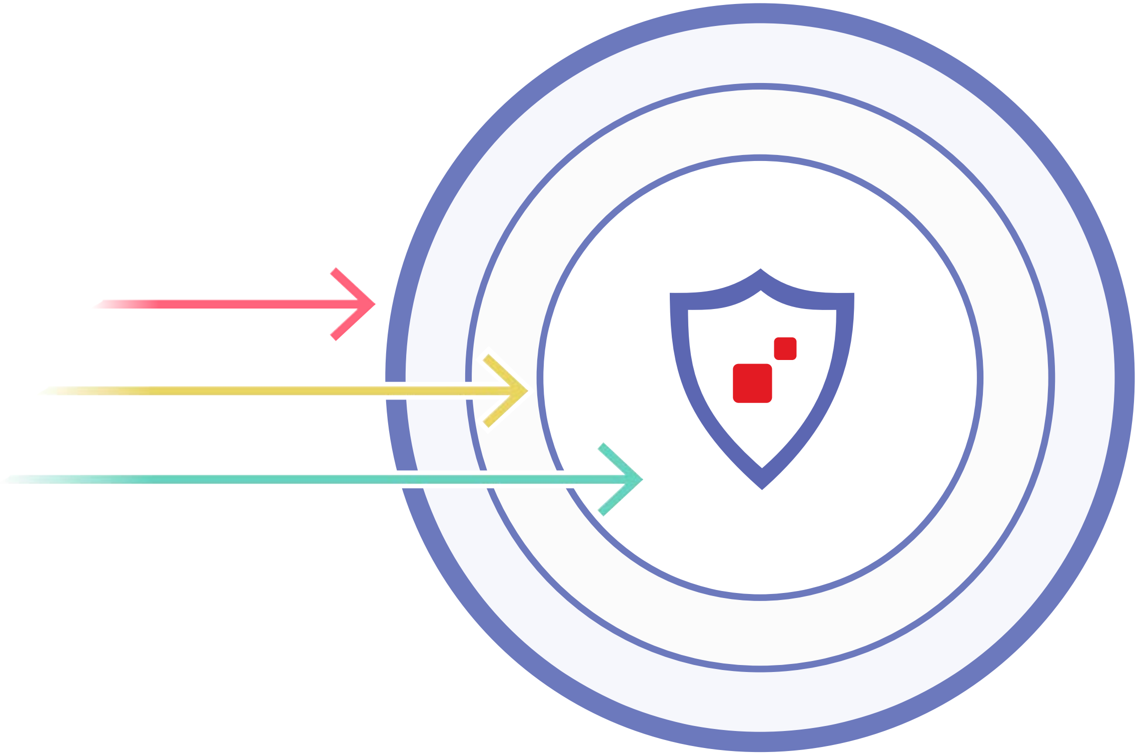
Compliance Officer

CEO/CFO/CIO

Developer

DBA

MarkLogic

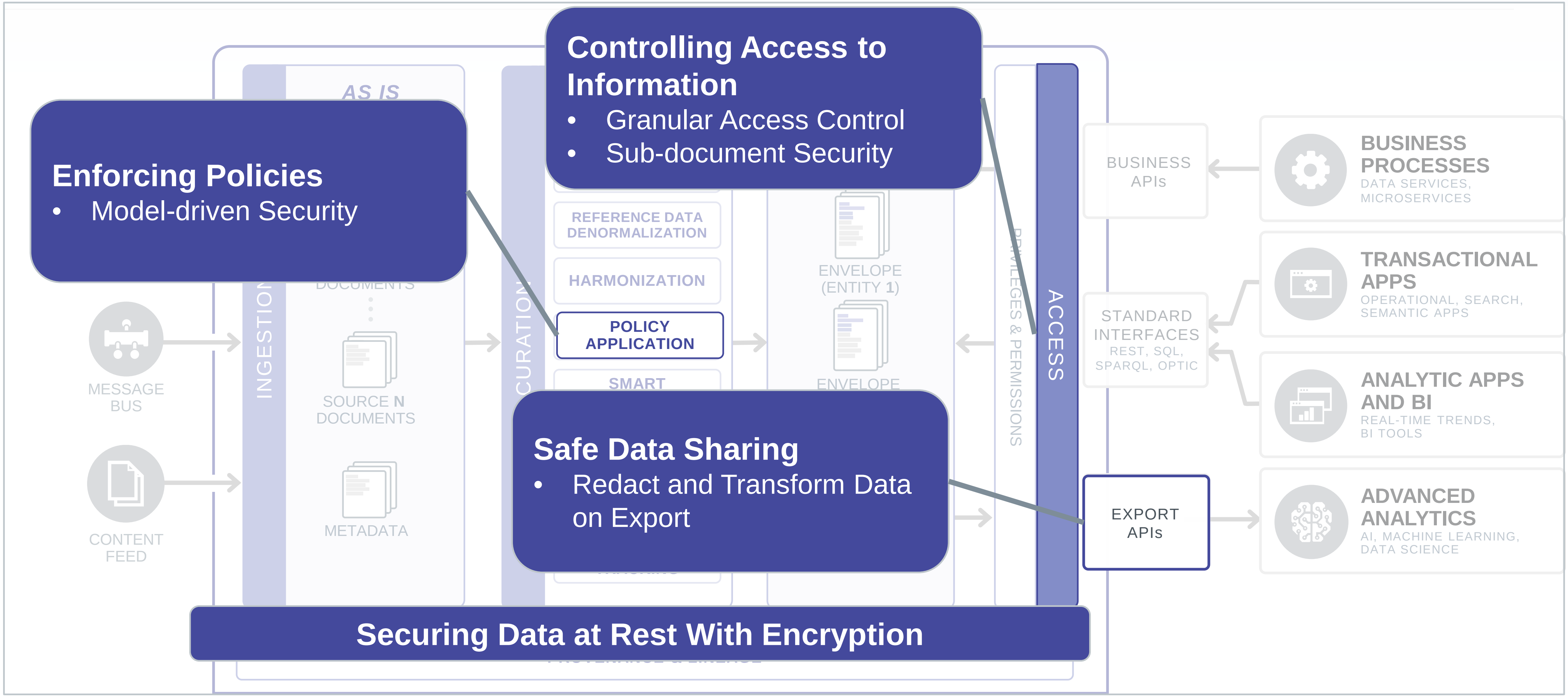


MARKLOGIC APPROACH

Secure and Share Sensitive Information

- Integrating data from silos
- Easily adapting to data model changes
- Enforcing policies in a single place
- Enabling safe data sharing

Secure and Share Sensitive Information



Enforcing Policies

- Model-driven Security

Controlling Access to Information

- Granular Access Control
- Sub-document Security

Safe Data Sharing

- Redact and Transform Data on Export

Securing Data at Rest With Encryption

Controlling Access to Information

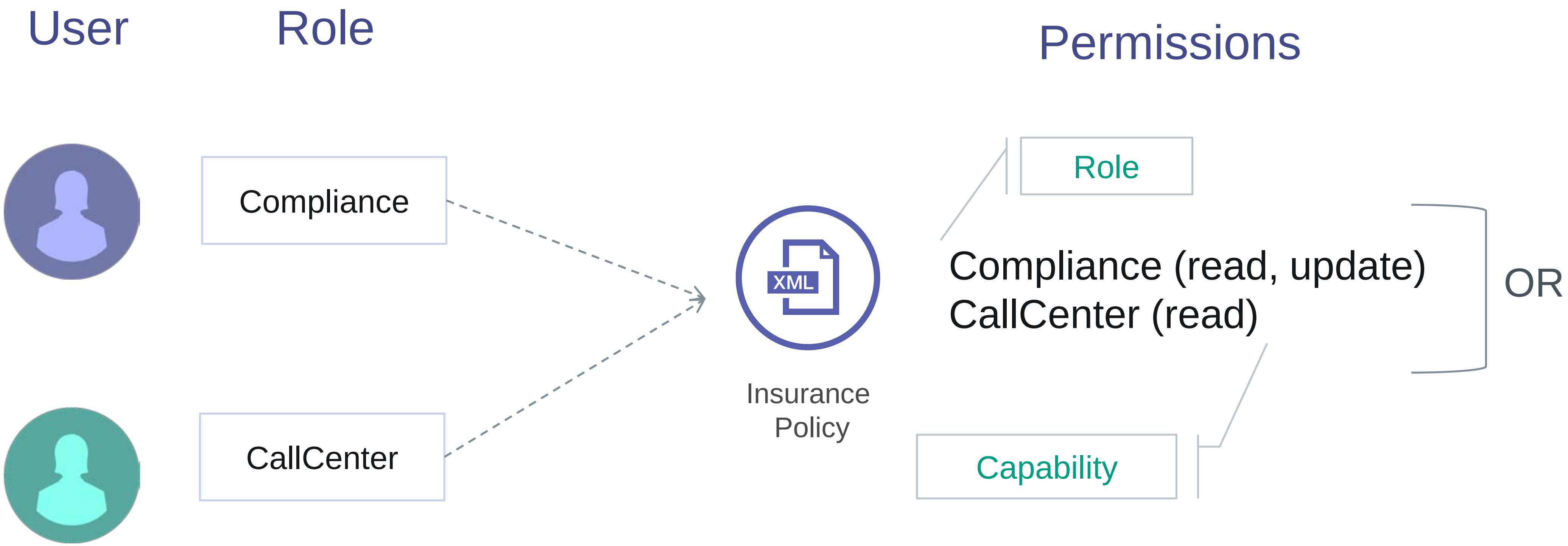
MarkLogic

HOW TO

Control Access to Documents

- Who is the user?
- What should the user see or do?

Role Based Access Control



User

Role

Permissions

Compliance

CallCenter

Insurance Policy

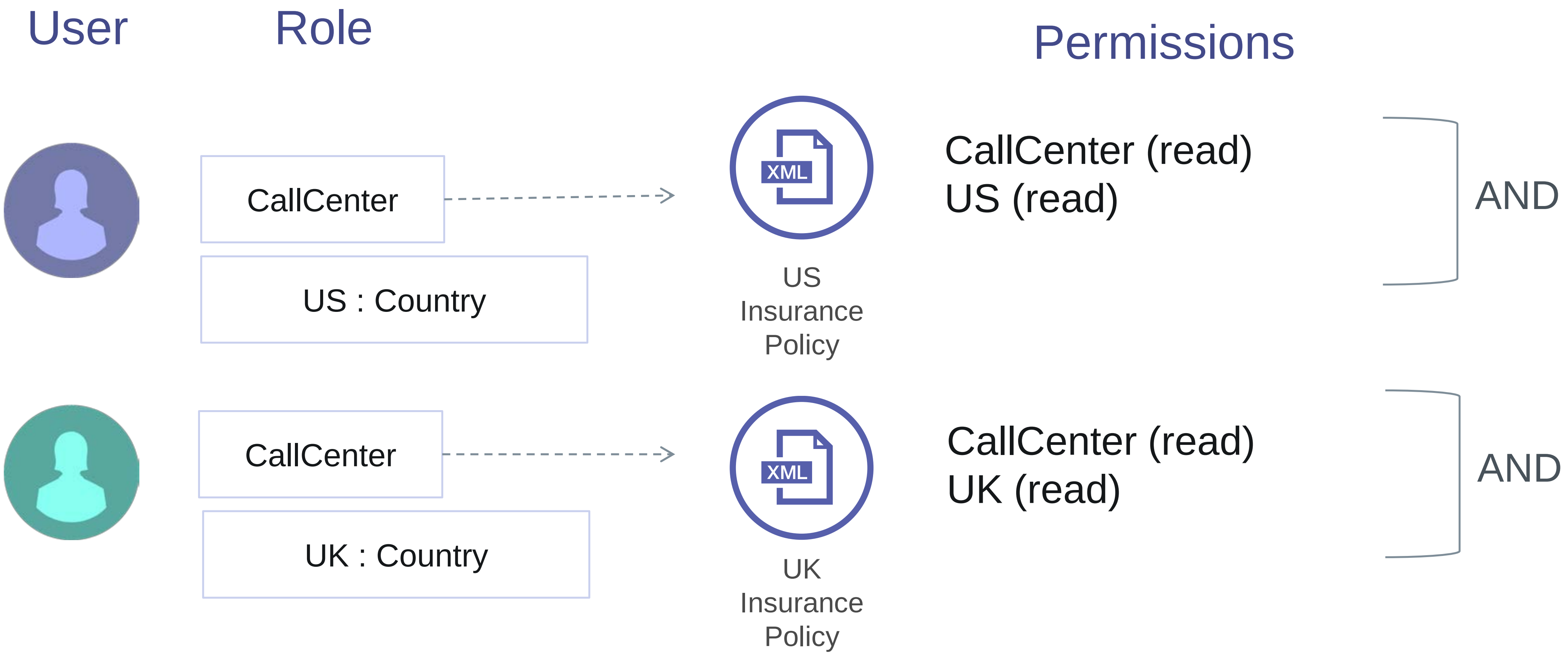
Role

Compliance (read, update)
CallCenter (read)

OR

Capability

RBAC – Compartment Security*



User

Role

Permissions

CallCenter

US : Country

US Insurance Policy

CallCenter (read)
US (read)

AND

CallCenter

UK : Country

UK Insurance Policy

CallCenter (read)
UK (read)

AND

* Part of the Advanced Security option

HOW TO

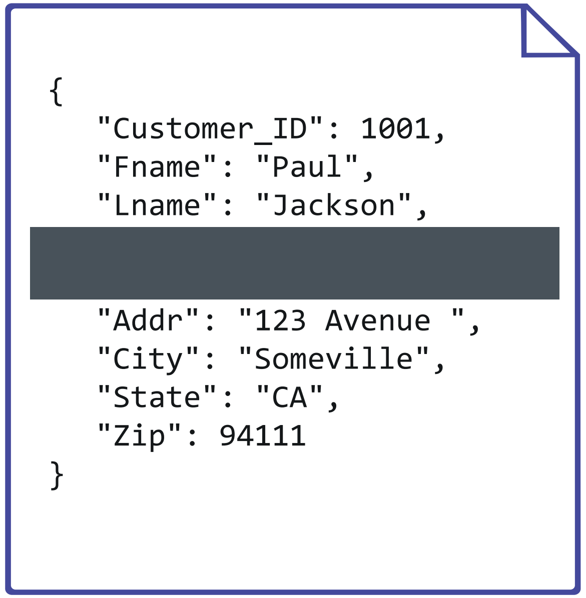
Control Access to Information Inside Documents

- What sensitive information has to be protected?
- How to enable authorized search only?

Role-Based, Element-Level Security

Granular Control On Information Visibility Based On User Roles

- Provide access control at the level of XML elements or JSON properties within documents
- Hide information from a user based on the user’s roles
- Out-of-the-box, in-database solution
- Real-time control enforced at the data layer for: search, queries, and updates



```
{
   "Customer_ID": 1001,
   "Fname": "Paul",
   "Lname": "Jackson",
[redacted]
   "Addr": "123 Avenue ",
   "City": "Someville",
   "State": "CA",
   "Zip": 94111
}
```

Ready for Evolving Document Models

- Based on Protected Paths that use XPath expressions to find information to conceal
- Each Protected Path is associated with roles and permissions
    - sec:protect-path*("//ssn", ("hr_role", "read"))
    - sec:protect-path("/root/reg[fn:matches(@access, ‘USA’)] ,(“USA_role", "read"))
    - sec:protect-path("/root/person[@cls=ts]", ("ts_role", "update"))

Only a user from HR can see the SSN

Only a Top Secret person can update documents classified as Top Secret

* Function signature simplified for illustration

Flexible Control

- Allows OR logic by creating protected path sets. Three protected path expressions*:
  - //agent[fn:contains(@releasableTo, "USA")] ,("Role_USA", "read"), "SetReleasableTo"
  - //agent[fn:contains(@releasableTo, "GBR")] ,("Role_GBR", "read"), "SetReleasableTo"
  - //agent[fn:contains(@releasableTo, "AUS")] ,("Role_AUS", "read") , "SetReleasableTo”

Set

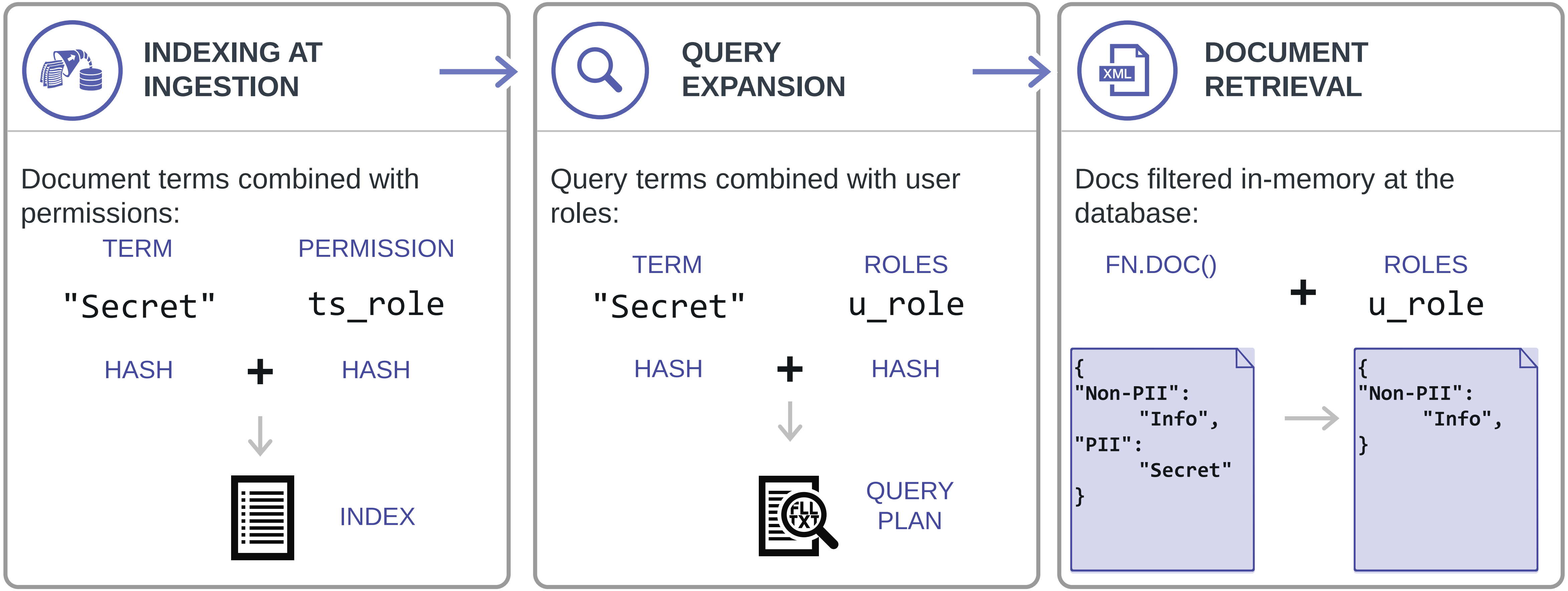
```
<agent releasableTo="USA GBR AUS">
    <name>Paul</name>
    <address>999 Broadway St</address>
    <phone>323-344-1555<phone>
    <country>US</country>
</agent>
```

Any of these roles can read
Logical OR

* Function signature simplified for illustration

Leak-Proof Security at Database Core

Information is protected at the database and never reaches the application



INDEXING AT INGESTION

Document terms combined with permissions:

TERM

"Secret"

HASH

PERMISSION

ts_role

HASH

+

INDEX

QUERY EXPANSION

Query terms combined with user roles:

TERM

"Secret"

HASH

ROLES

u_role

HASH

+

QUERY PLAN

DOCUMENT RETRIEVAL

Docs filtered in-memory at the database:

FN.DOC()

+

ROLES

u_role

```
{
"Non-PII":
      "Info",
"PII":
      "Secret"
}
```

```
{
"Non-PII":
      "Info",
}
```

Safe Data Sharing

MarkLogic®

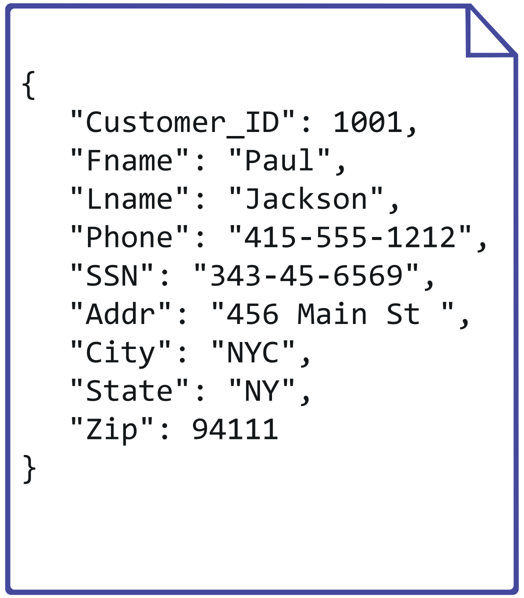
HOW TO

Share the Right Information

- Can I get a data dump with PII removed?
- How to give data to data scientists?
- How to get realistic data on QA/UAT?

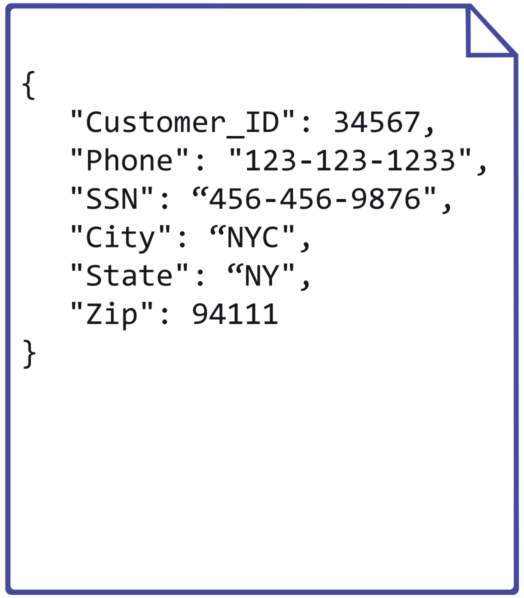
Rule-Based Redaction

Share Data While Preserving Privacy



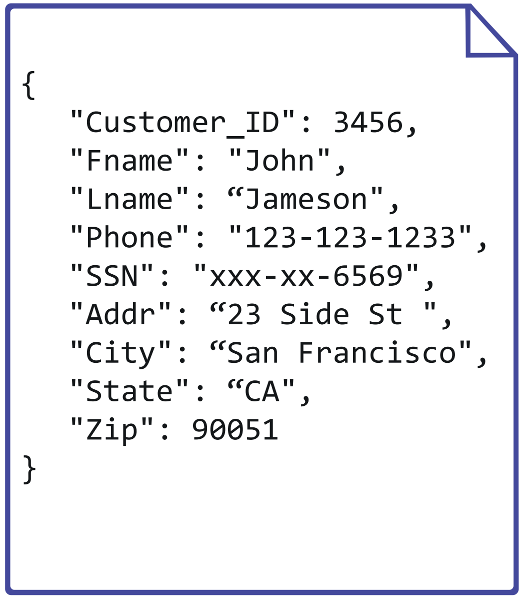
```
{
   "Customer_ID": 1001,
   "Fname": "Paul",
   "Lname": "Jackson",
   "Phone": "415-555-1212",
   "SSN": "343-45-6569",
   "Addr": "456 Main St ",
   "City": "NYC",
   "State": "NY",
   "Zip": 94111
}
```

Original document



```
{
   "Customer_ID": 34567,
   "Phone": "123-123-1233",
   "SSN": “456-456-9876",
   "City": “NYC",
   "State": “NY",
   "Zip": 94111
}
```

BI export



```
{
   "Customer_ID": 3456,
   "Fname": "John",
   "Lname": “Jameson",
   "Phone": "123-123-1233",
   "SSN": "xxx-xx-6569",
   "Addr": “23 Side St ",
   "City": “San Francisco",
   "State": “CA",
   "Zip": 90051
}
```

QA/Dev Export

- Mask or conceal sensitive information
- Use predefined functions
- Out-of-the-box, in-database solution

Rules

- Each rule uses XPath expressions to find information to conceal or mask
- Each rule defines what to do with the information by specifying a function
- Rules are documents in one collection, e.g. "DEVELOPERS_EXPORT

```
ruleClientInfo = {
  "rule": {
    "description": "Random #..",
    "path": "/policy/client/id",
    "method": {
      "function": "mask-random"  },
    "options": {
      "length": 10  }}};
```

```
 xdmp.documentInsert(
 "ruleClientInfo.json",ruleClientInfo,  {
     "collections": [
       DEVELOPERS_EXPORT ]});
```

Functions

- Ships with the following out-of-the-box functions:
  - Conceal
  - Masking: Random, Deterministic, Dictionary
    - Highly secure design to prevent linkage attacks
  - Patterns: SSN, US Phone, email, IPv4, Regex, Dates, Numbers
- Users can write custom functions

```
"method": {
       "function": "mask-deterministic"
     },
     "options": {
       "length": 10
        "salt": “a23sdas#4er"
        "extended-salt": "collection"
     }
```

```
 "method": {
     "function": "redact-us-ssn",
   },
   "options": {
     "level": "partial",
     "character": "X"
   }
```

Model Driven Security

User Story

- Given a customer record, a compliance officer must be able to verify a customer’s PII-Personally Identifiable Information.
- A clerk, on the other hand, must NOT see the customer’s PII.

```
{

  fullName : “Ellie Holland”,

  worksFor : “Supermemo Ltd”,

  email :eholland@supermemoltd.biz,

  ssn:   : “164-32-6412”

}
```

User Story

- Given a customer record, a compliance officer must be able to verify a customer’s PII.
- A clerk, on the other hand, must NOT see the customer’s PII.
- Provide a Java function to application developers that implements `getCustomerHistory(name)` which hides SSN if an unprivileged user requests the data.
- Provide a function `getCustomerHistoryBySSN(ssn)` that returns nothing if searched by a clerk.

```
{

  fullName : “Ellie Holland”,
  worksFor : “Supermemo Ltd”,
  email :eholland@supermemoltd.biz,
  ~~ssn:   : “164-32-6412”~~

}
```

Model Driven

- Rather than configure security for my requirements….
- I declare the requirement in a model.
- DHF tooling provides the rest!

```
{…

  Customer : {
      properties : { … },

      required : [ “fullName”,…],

      “pii” : [ “ssn” ]

  },...

}
```

Terms in the Demo

- PII: Personally Identifiable Information
- DHF: MarkLogic Data Hub Framework
- ES: Entity Services
  - Models/Types/Properties
- ELS: Element Level Security
  - Protected Paths/Query Rolesets

The Task List – Already Done

- Created test dataset.
- Created Model for Calls, Employees, and Customers.
- Database function `getCustomerHistory(name)` implemented/verified.
- Java function provided to application developers.
- DHF Flows created and run (using test data).
- Roles defined:
  - “clerk” can read documents.
  - “compliance-officer” can read secured elements.

MarkLogic

TASK LIST – DEMO TODO

- Indicate what properties are PII in your entity services model.
- Generate security configuration.
- Deploy security configuration
- Verify

- Edit a model file
  ./gradlew mlLoadModules
- ./gradlew hubGeneratePii
- ./gradlew mlDeploySecurity
- ./gradlew getCustomerHistory

Under the Hood: hubGeneratePii

- All properties in an Entity Services model refer to known locations in documents
- A new role is shipped with MarkLogic 9.0-5: `pii-reader`
- `hubGeneratePII()` looks at all the entity services models in your project and:
  - Creates a protected path configuration to each PII property, securing with pii-reader
  - Creates a query roleset for accessing those properties with the pii-reader role
  - Saves these configurations in your DHF project.
- Use `mlDeploySecurity` task to send these configurations to your server.

<demo/>

Securing Data at Rest With Encryption

HOW TO

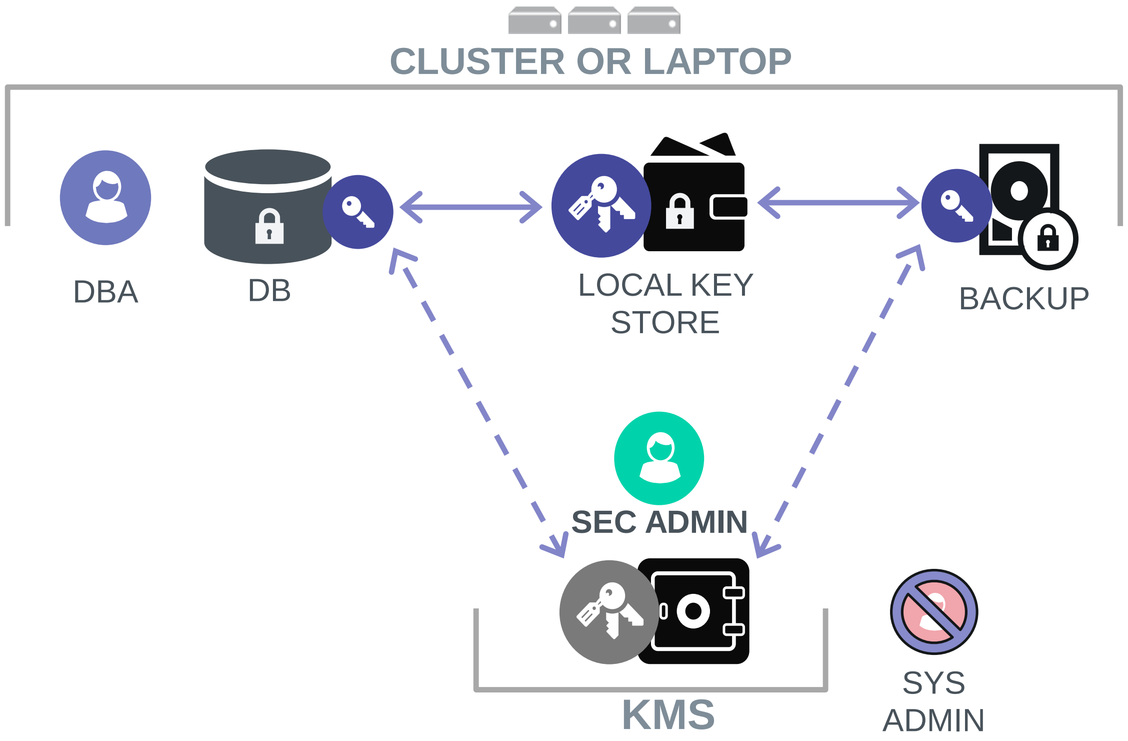
Prevent Direct Access to Files

- Is my data secure on disk?
- Can the cloud sys admin see the data?
- Can someone modify the data on disk?
- Can you erase traces of wrongdoing?

Advanced Encryption

Transparent Encryption of Data, Configuration and Logs

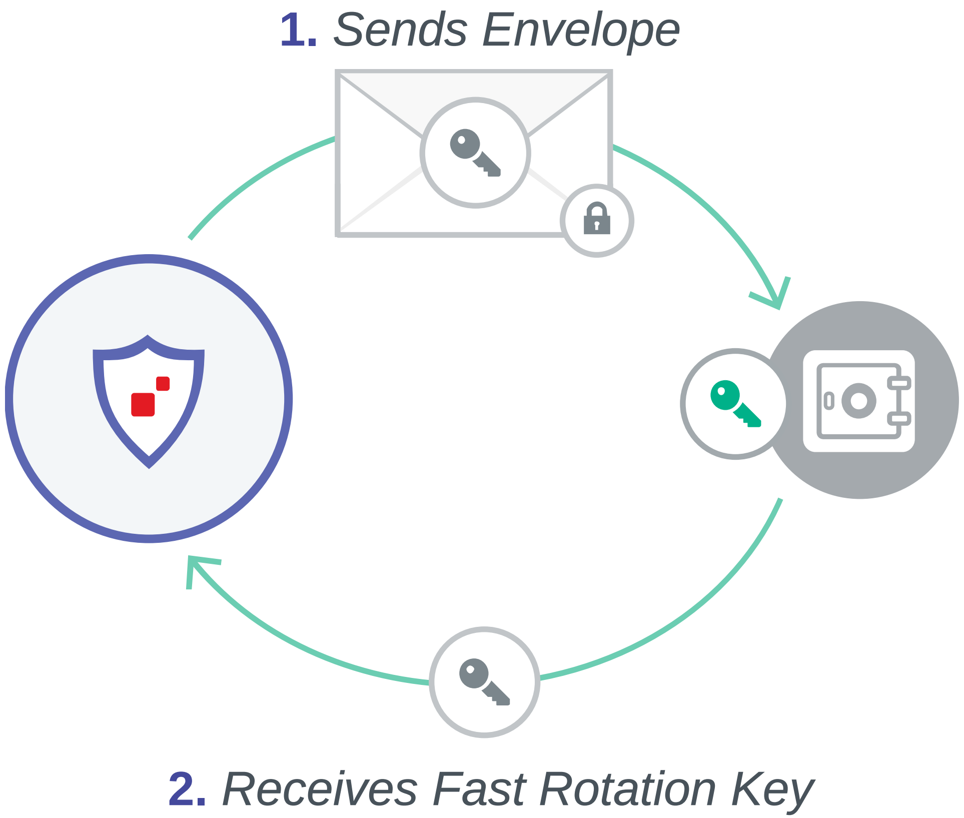
- Transparent: No code modification
- Local or external KMS
- Prevent tampering of information on disk, including logs
- High performance encryption



Secure by Design

Reliable and easy to manage

- To access low level keys and read files, MarkLogic sends the envelope to the KMS, that then sends back the unencrypted key
- For external KMS, MarkLogic has no access to envelope keys
- No access to KMS – No access to files, no ingestion, and no compromises



Concluding Thoughts

- Security in the database
  - Comprehensive granular access to and sharing of information
  - Auditable and fail-secure
  - Personal data is protected consistently for all users and applications
  - Simple, effective high performance encryption
- Governed by default
  - Metadata for datasets: test means developers can’t see PII
  - Certification for applications: only when PII is verified can the app go to production
- Model-driven governance
  - An alternative to configuring security

MarkLogic

Key Resources

- Documentation: docs.marklogic.com
- Whitepaper:
  - Enabling Regulatory Compliance
  - EU General Data Protection Regulation: The Path to Compliance
- Videos: Your GRC Strategy: When Is Enough Tooling Enough?
- Blogs:
  - A security model for data integration
  - How to prepare your security model for Brexit
  - Protect Against Linkage Attacks

MarkLogic

Questions?