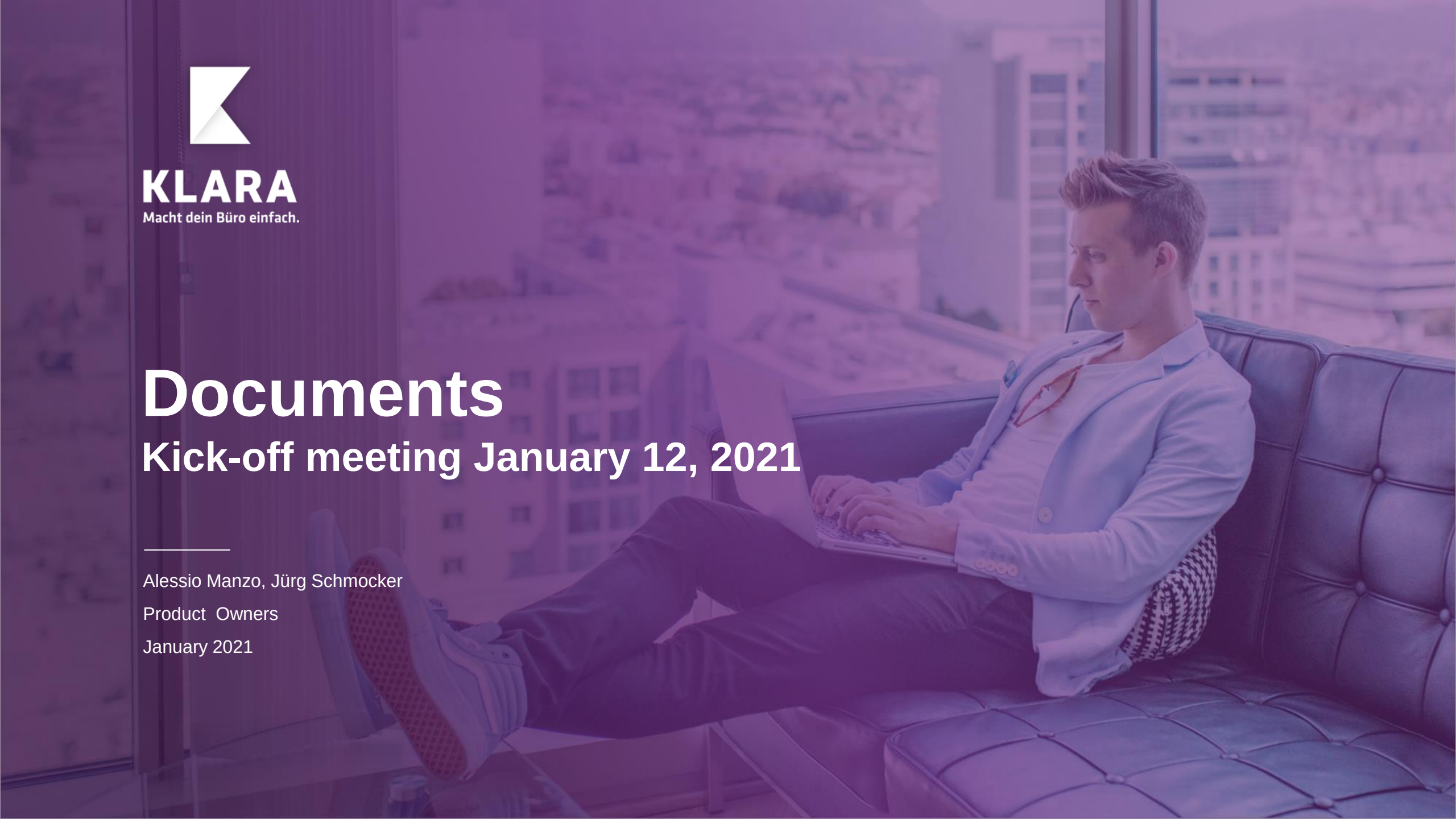


KLARA
Macht dein Büro einfach.

Documents

Kick-off meeting January 12, 2021

Alessio Manzo, Jürg Schmocker

Product Owners

January 2021

Documents

Agenda

• Business aspects	Alessio Manzo
• Architecture	Laurent Salamin
• POC	Tobias Hofer
• Next steps	Jürg Schmocker

Documents – Business aspects

New KLARA Project / Module

- Vision: develop a new document management system (DMS), supporting:
  - high volumes and high performance
  - high security and privacy
  - Can and will be used by different applications like KLARA, My Live, Digital letterbox etc.

- A document is a file (e.g. text, image, voice, etc.) including its metadata.

Documents – Business aspects

High volumes

Up to 10 million (10’000’000) of tenants

- 80 % of individuals
- 20 % of companies
  - 80 % of companies with same volumes than individuals
  - 20 % of companies with larger number of users and number of documents
- Each tenant will be
  - in general logged once a day
  - will stay connected for an average of 5 minutes

Up to 2 billion (2'000'000’000) documents

- documentsize is around 1 mb

Documents – Business aspects

High volumes / Performance

- Each tenant will receive between 5 and 10 files per day
- Concurrent tenants logged in:
  - On a 12 hours basis, we can estimate that approximatively 1 % of the tenants are logged in simultaneously (100’000 Tenants)
  - Be aware of some hot hours during which peaks will be encountered.

- Response time: for 95% of the volume < 1 second, no matter the function which is executed

Documents – Business aspects

Security

- Tenant specific encryption
- Authentication
- Authorization
- Ensure document integrity
- Log activities, Auditing

Documents – Business aspects

Requirements

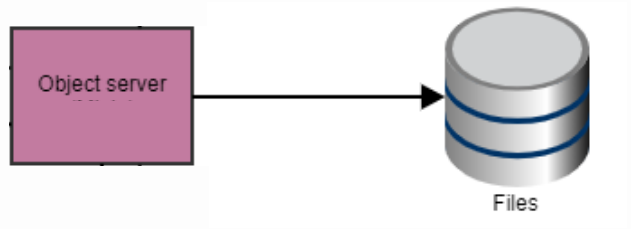
- Store documents
- Manage documents (CRUD)
- Retrieve documents
- Search documents / content
- Share documents (within / cross tenants)
- Enrich files with metadata (manually / automatically)
- Make documents and metadata accessible from everywhere and with any kind of devices

Documents

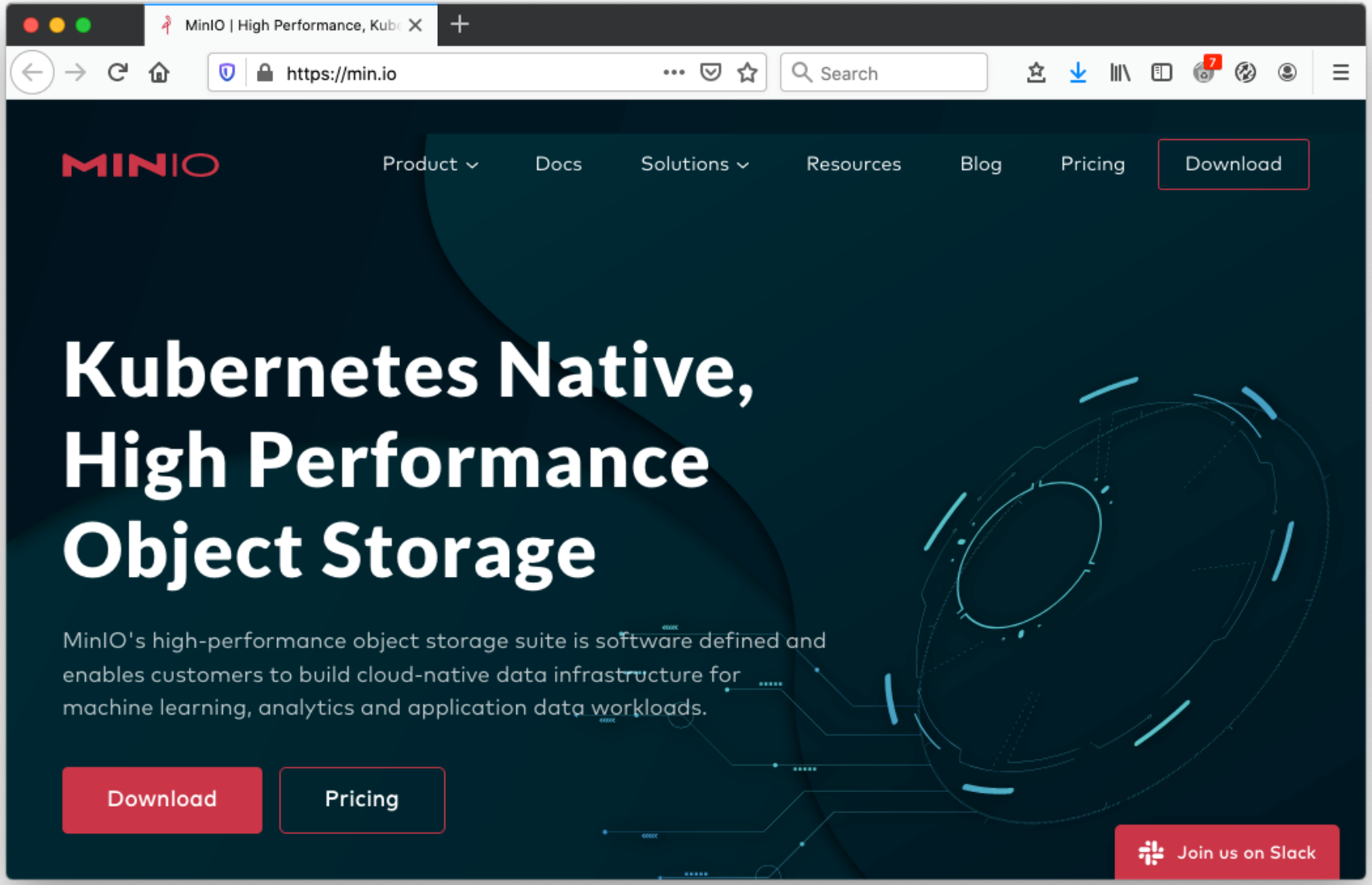
Agenda

• Business aspects	Alessio Manzo
• Architecture	Laurent Salamin
• POC	Tobias Hofer
• Next steps	Jürg Schmocker

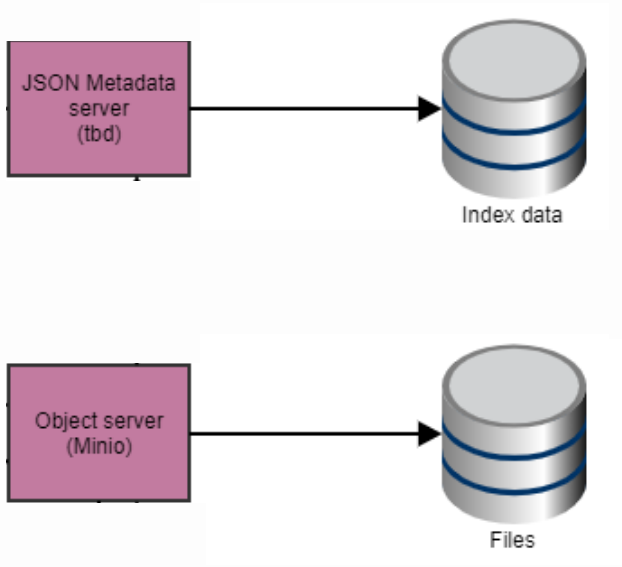
Global architecture overview



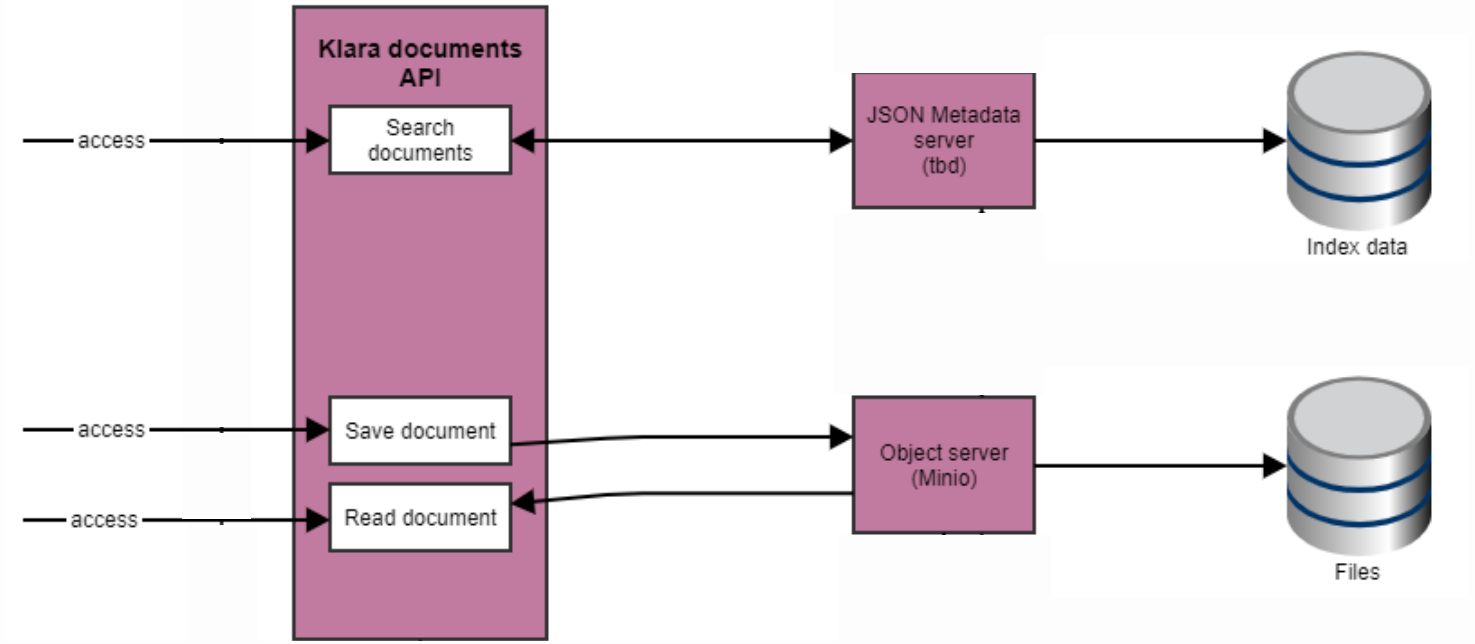
Global architecture overview



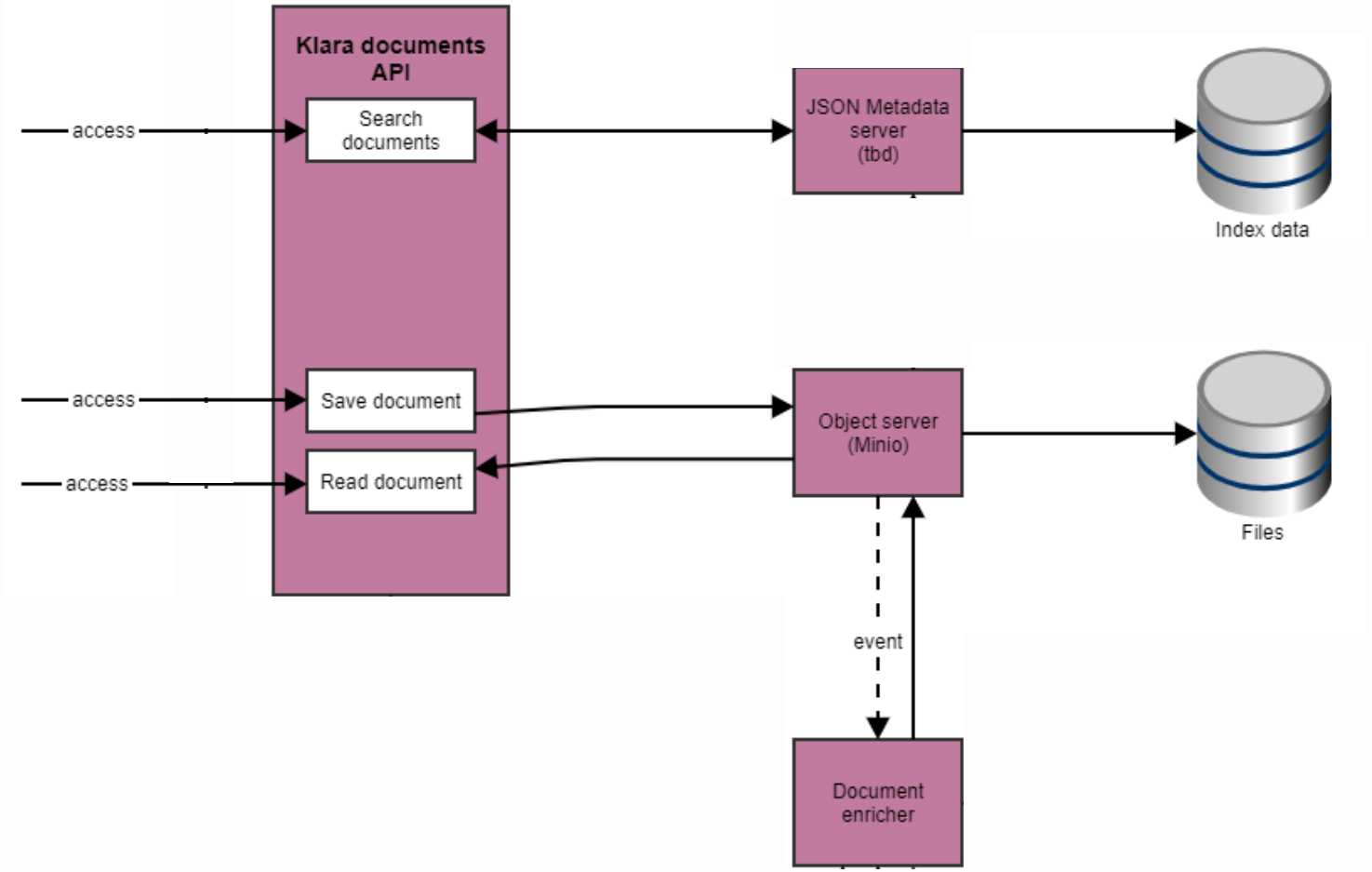
Global architecture overview



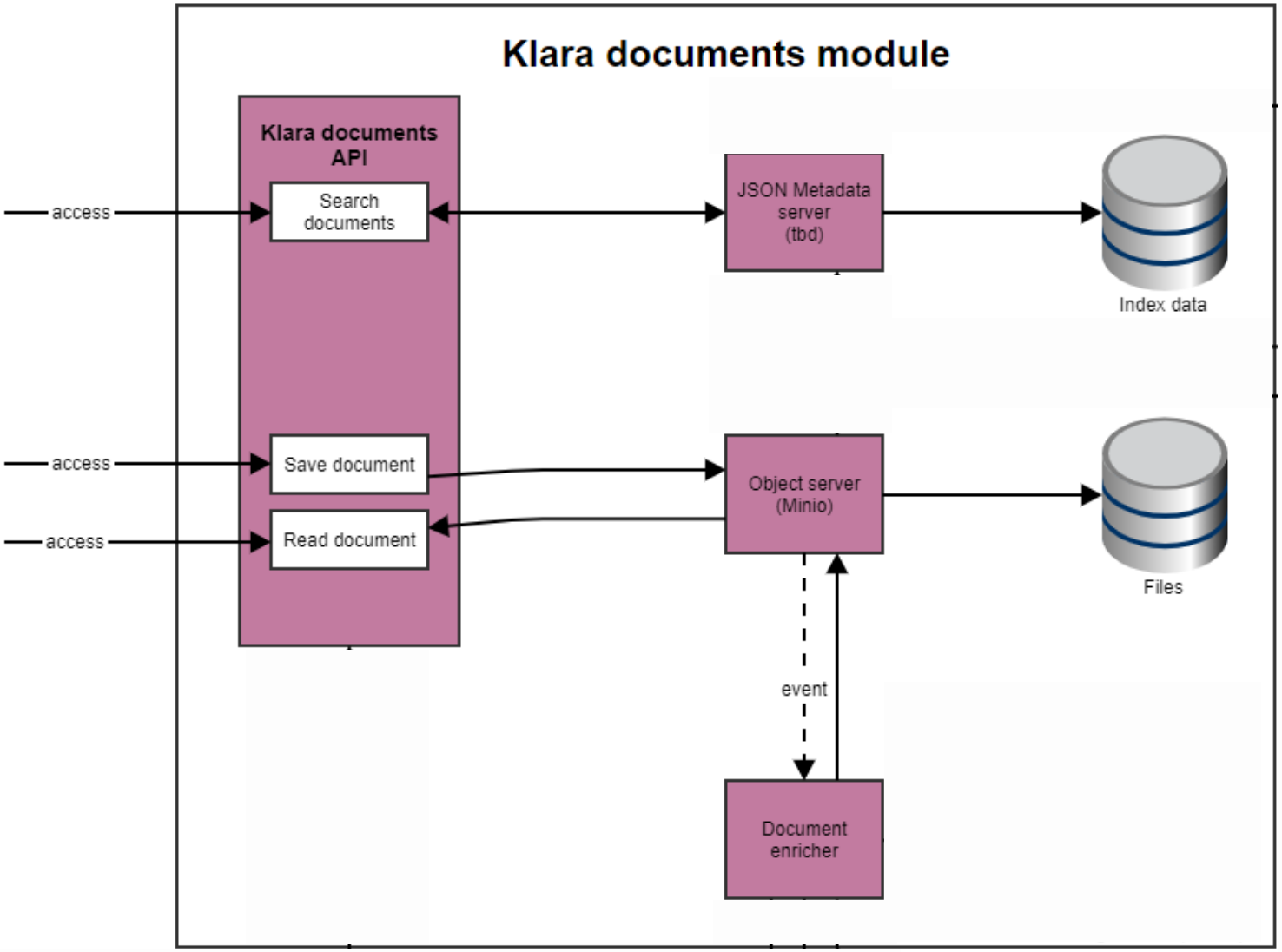
Global architecture overview



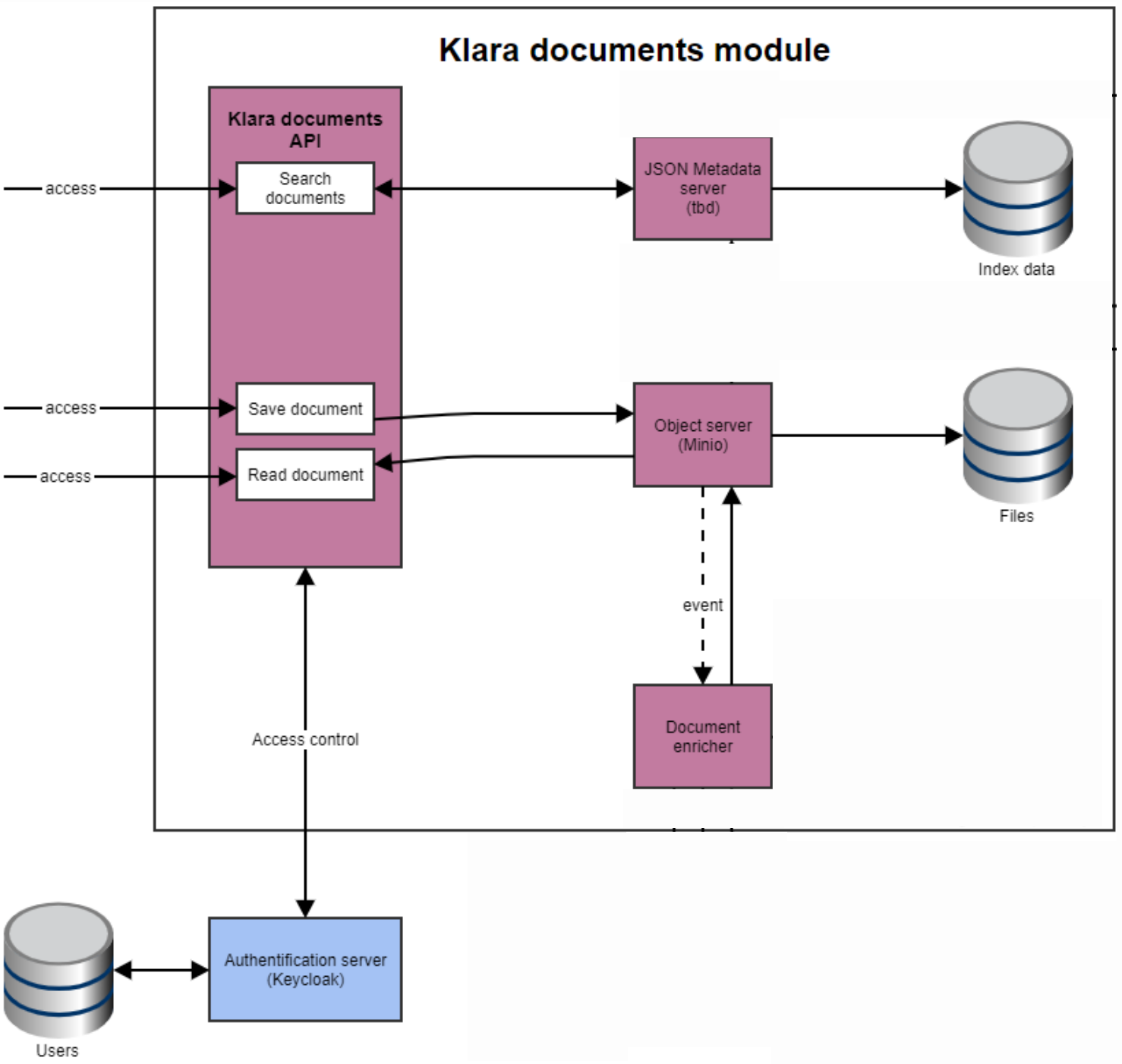
Global architecture overview



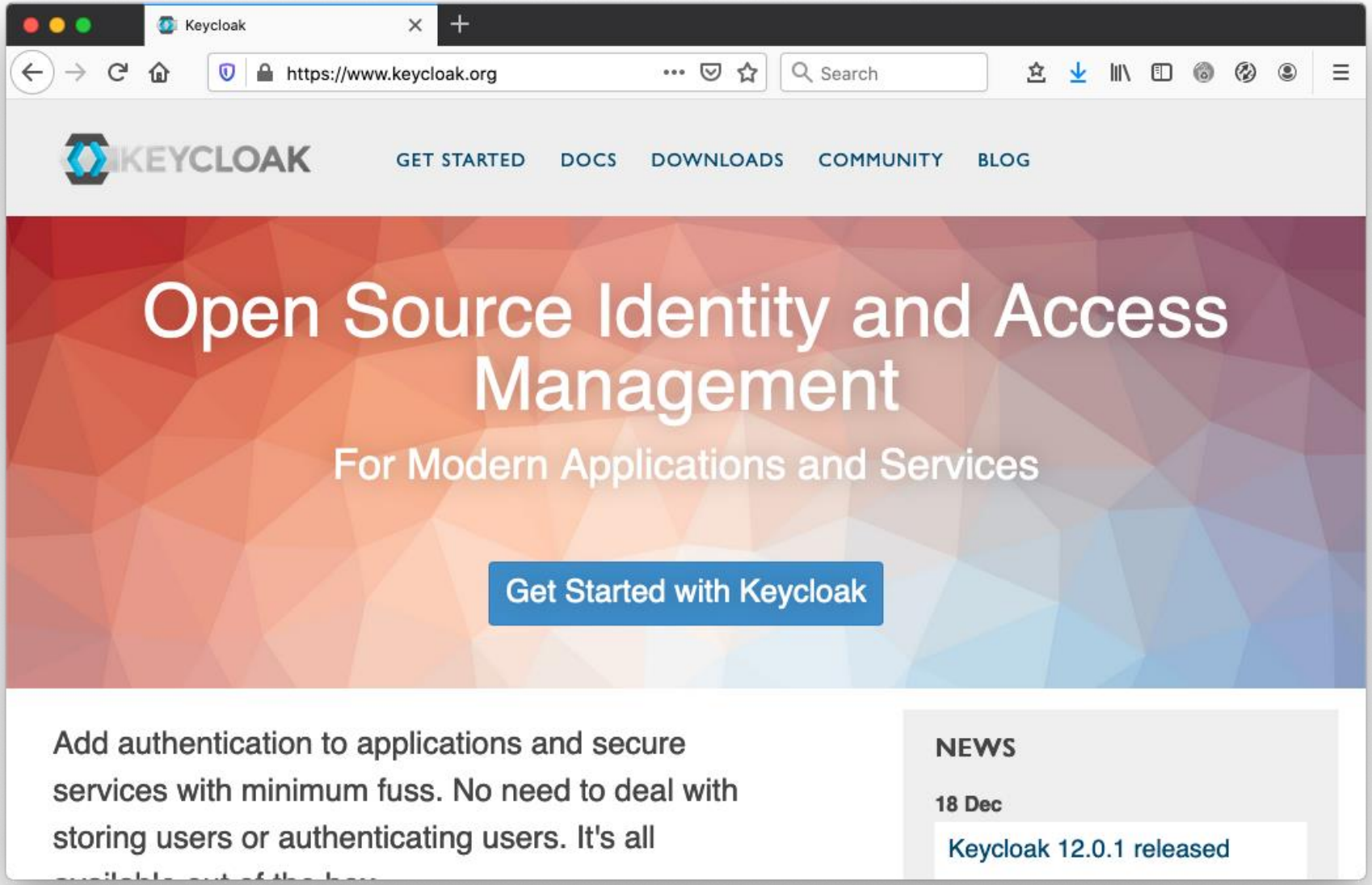
Global architecture overview



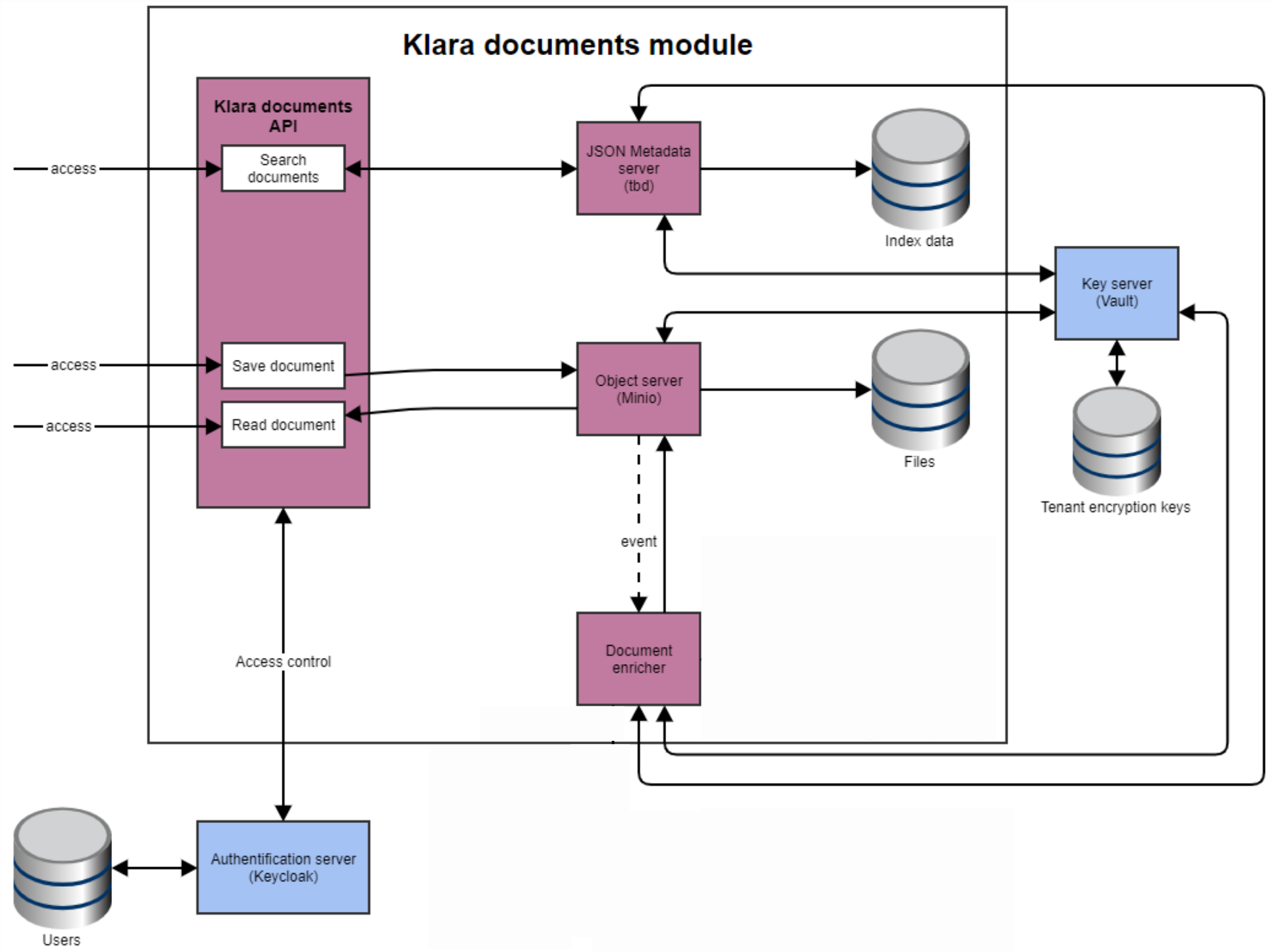
Global architecture overview



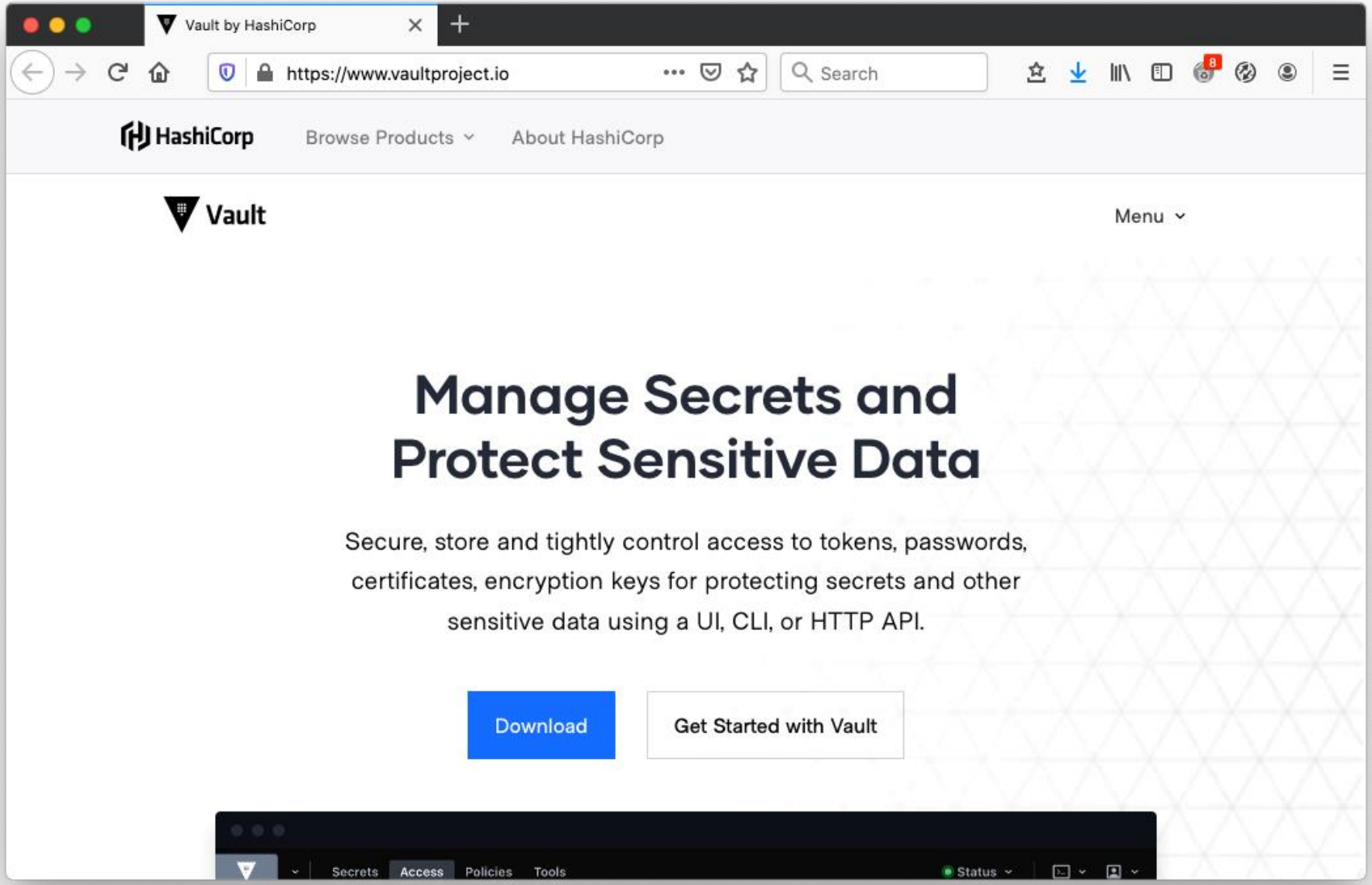
Global architecture overview



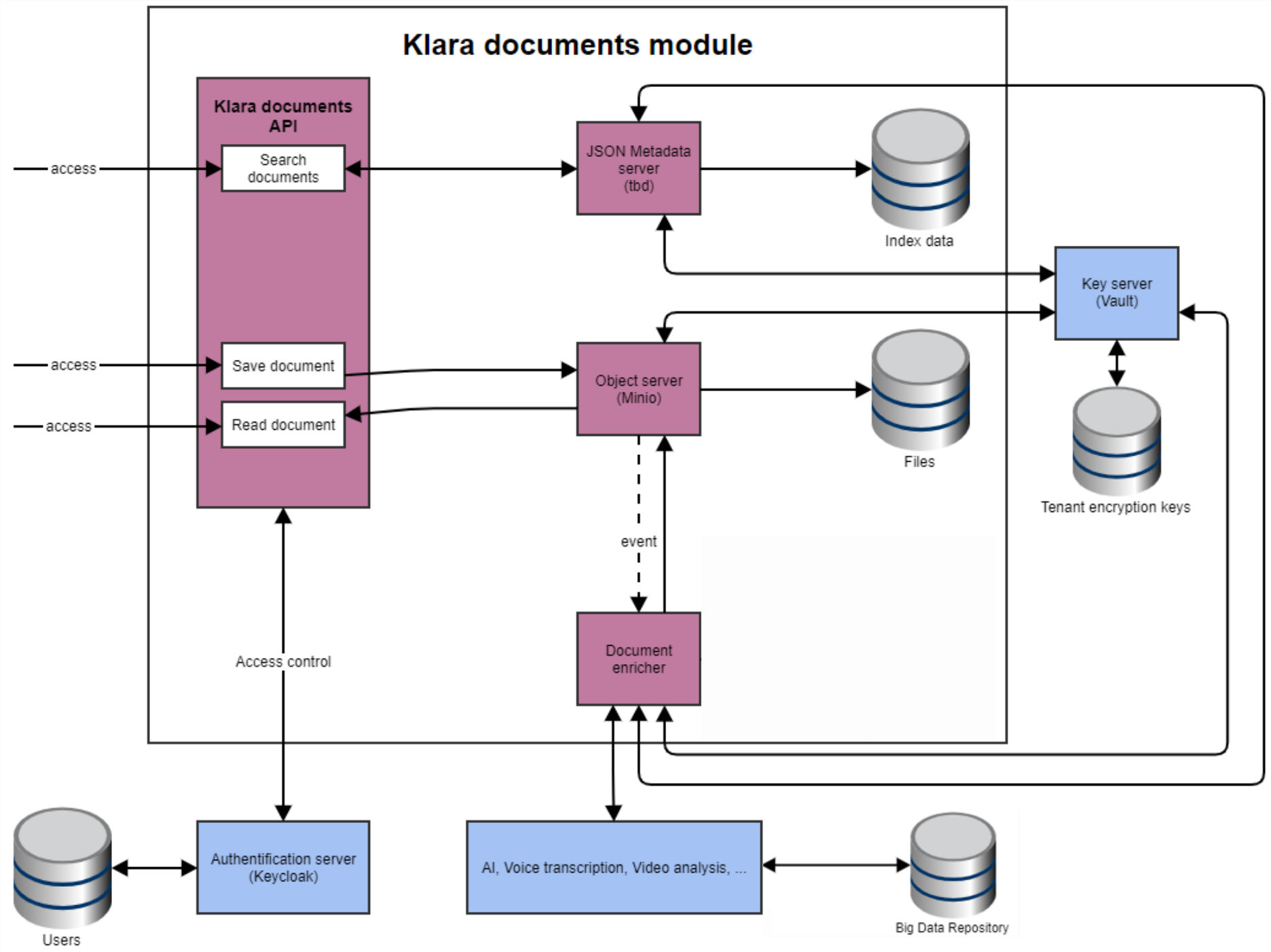
Global architecture overview



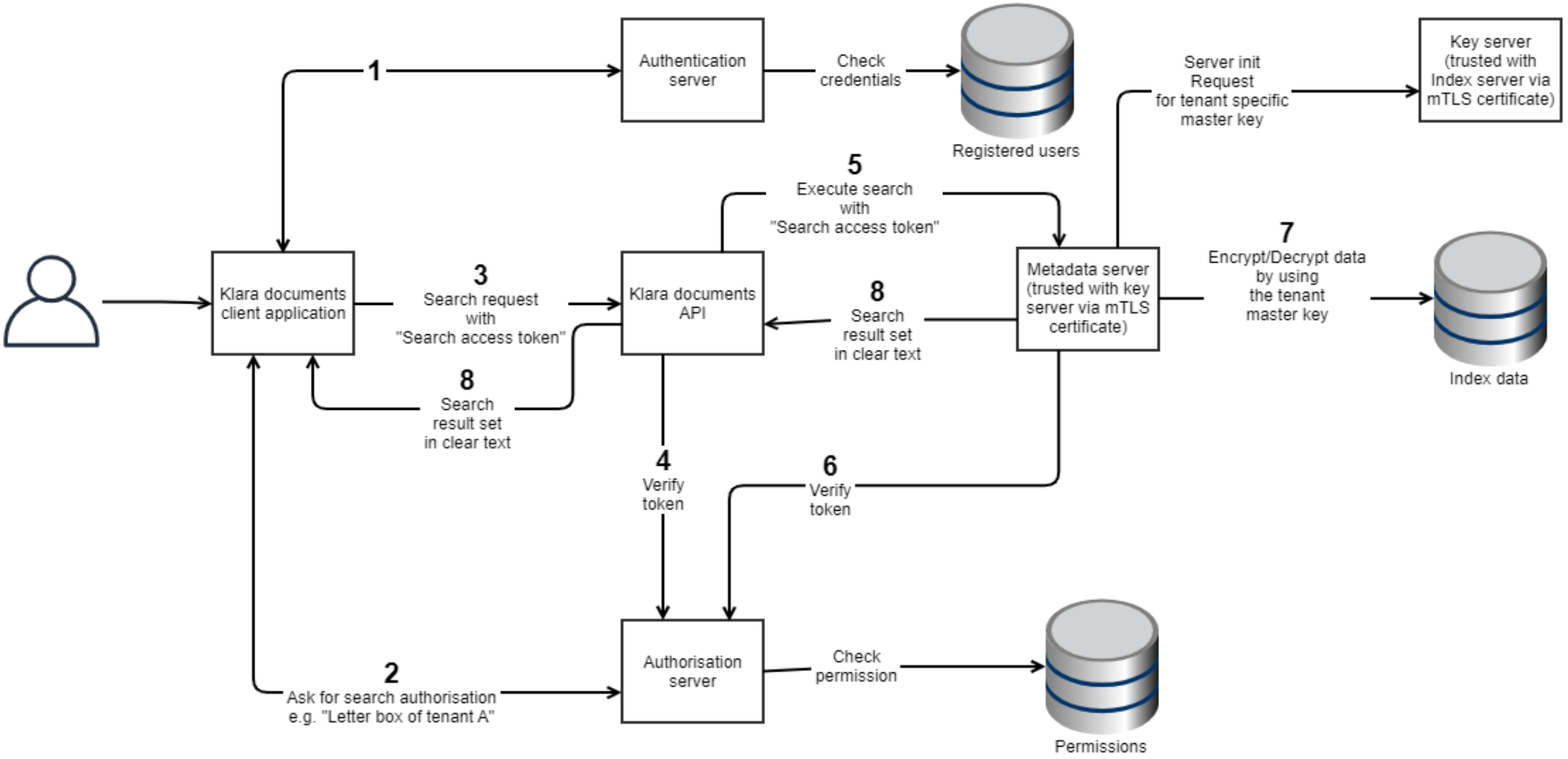
Global architecture overview



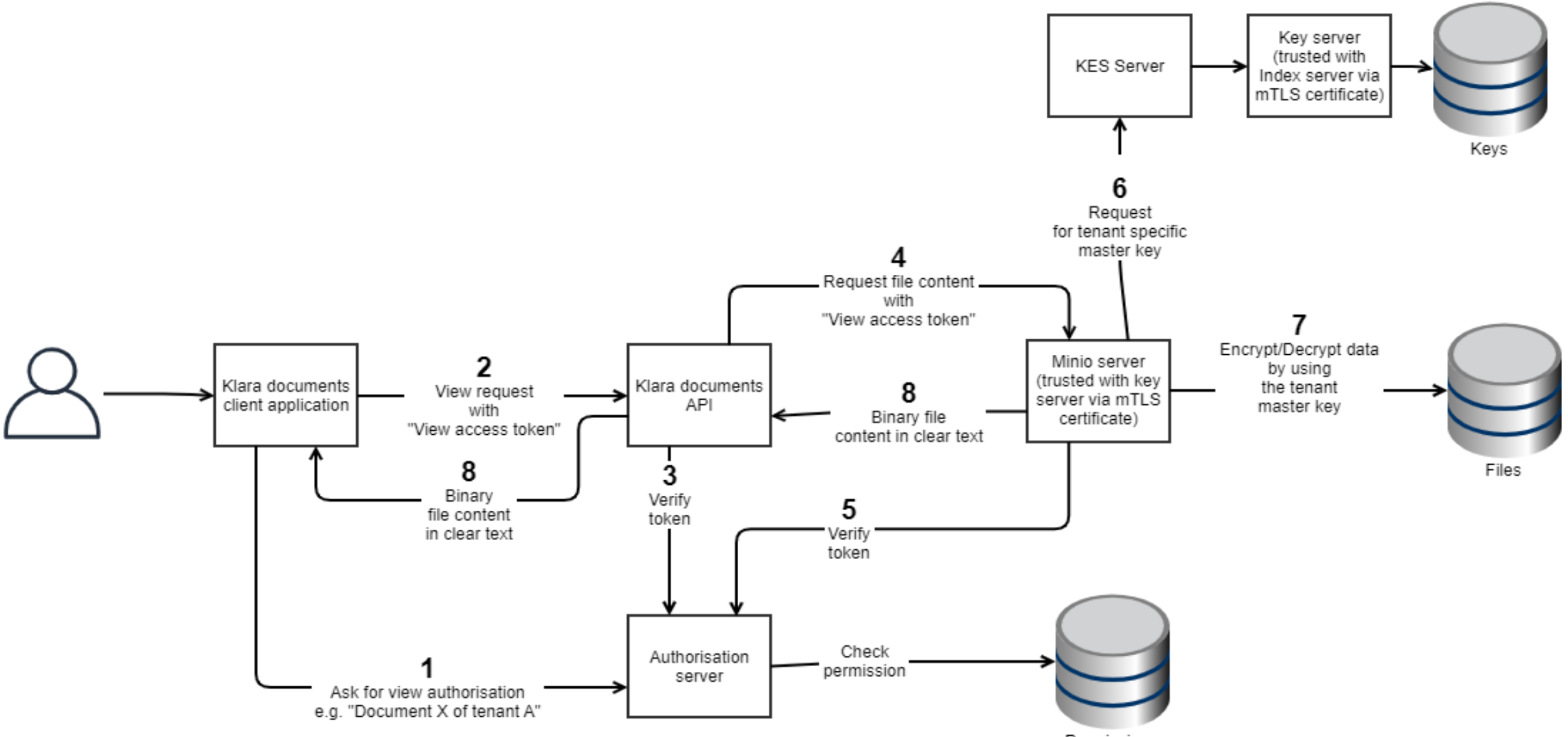
Global architecture overview



Search documents – Server side encryption flow



View document – SSE S3 encryption flow



Metadata and Document concepts

Metadata

Two separate data categories :

1. Properties :
    1. ID
    2. Creation date
    3. Last change date
    4. ...
2. Business information :
    1. Type
    2. Description
    3. Author
    4. Tags
    5. Structured data
        1. Invoice details
        2. Accounting information

Metadata and Document concepts

Metadata

Document metadata can be :

1. Provided together with the document files
2. Entered at the keyboard by the user classifying the document
3. Extracted from the document files
4. Enriched by external data and AI processing

Metadata and Document concepts

Document

- Represents the end-user view of an information
- Can be represented by any kind of information content :
  - pdf text
  - JSON text
  - XML text
  - Video
  - Sound
  - Non human readable binary content
  - Etc.

Metadata and Document concepts

Document

- Is not equal to a single file
- Within Klara documents, it will consist of
  - One JSON structured **metadata**
  - A **collection** of **files** representing information in different views

Metadata and Document concepts

Document

Each file is identifed uniquely by the following :

- Document ID
- File type
- Version number

File types :

- source document
- canonical document (REFERENCE)
- Thumbnail in different sizes (e.g. thumbnail-S, thumbnail-M, thumbnail-L)
- displayed document (in PDF format)
- other document's representations like XML or JSON structured data of the document

Metadata and Document concepts

Document

- File properties :
  - ID
  - Creation date
  - Last change date
  - Size
  - Media type e.g. :
    - application/pdf
    - application/xml
    - application/json

Documents

Agenda

• Business aspects	Alessio Manzo
• Architecture	Laurent Salamin
• POC	Tobias Hofer
• Next steps	Jürg Schmocker

PoC – Proof of Concept

Goals

- Proof new **access management** standard
- Proof that Authentication and Authorization Server will not become a **bottleneck**
- Proof ability of **third-party components** to secure access to their data
- Proof that tenant specific server-side **encryption** works

PoC – Proof of Concept

New Access Management Standard

- OAuth (Open Authorization)
  → Killed the ‘password anti-pattern’
- OpenID Connect
  → Identity API (on top of OAuth2)
- UMA (User Managed Access)
  → Access Management API (on top of OAuth2)

PoC – Proof of Concept

Starting situation

- Start from scratch (but reuse existing code where applicable)
- Use most recent version of our technology stack (e.g. application server, libraries)
- Use standards

PoC – Proof of Concept

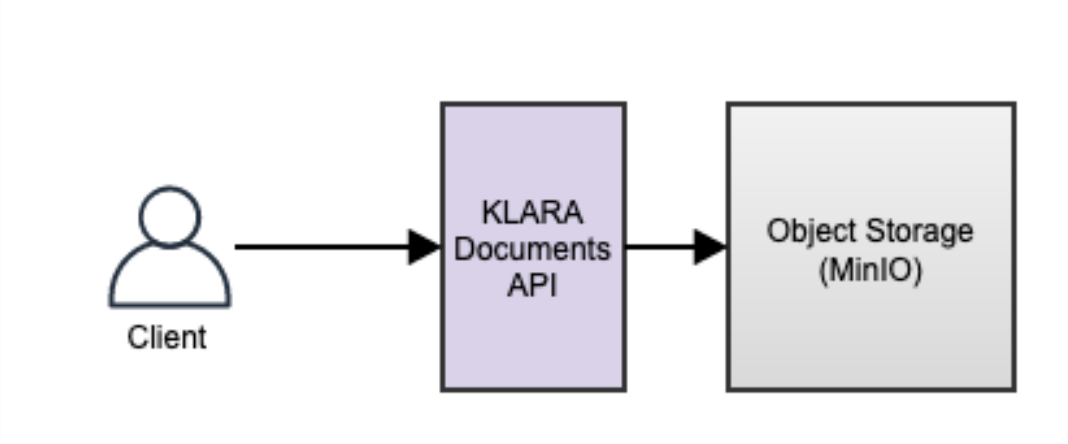
Step by step approach

- Start with a basic setup
- Add complexity incrementally
- Test system performance continuously

PoC – Proof of Concept

Story 1 – Basic setup

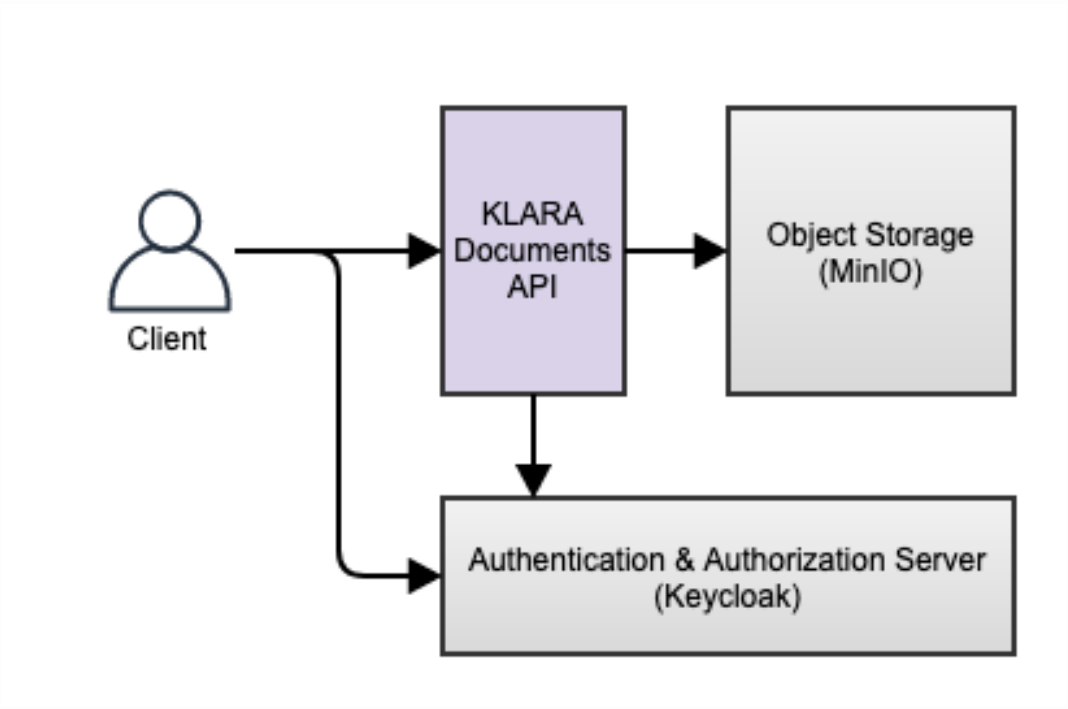
- Small API to store, list and retrieve files



PoC – Proof of Concept

Story 2 – Authentication

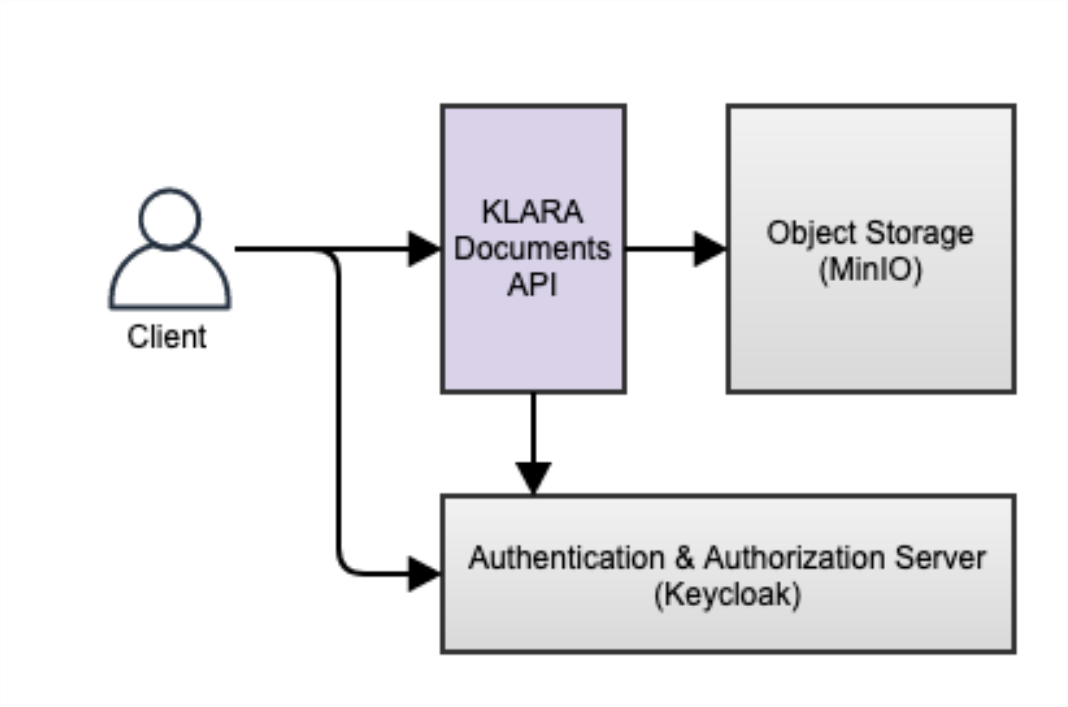
- Grant access to identified users only



PoC – Proof of Concept

Story 3 – Authorization

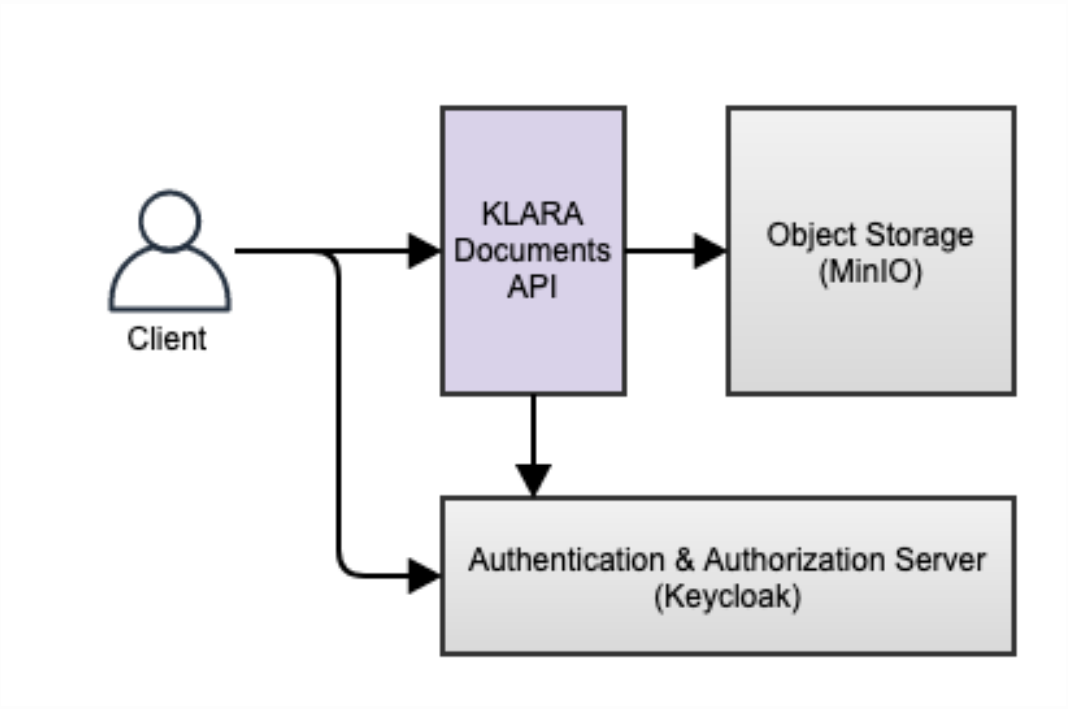
- A user is only allowed to access resources of its tenant



PoC – Proof of Concept

Story 4 – Advanced Authorization

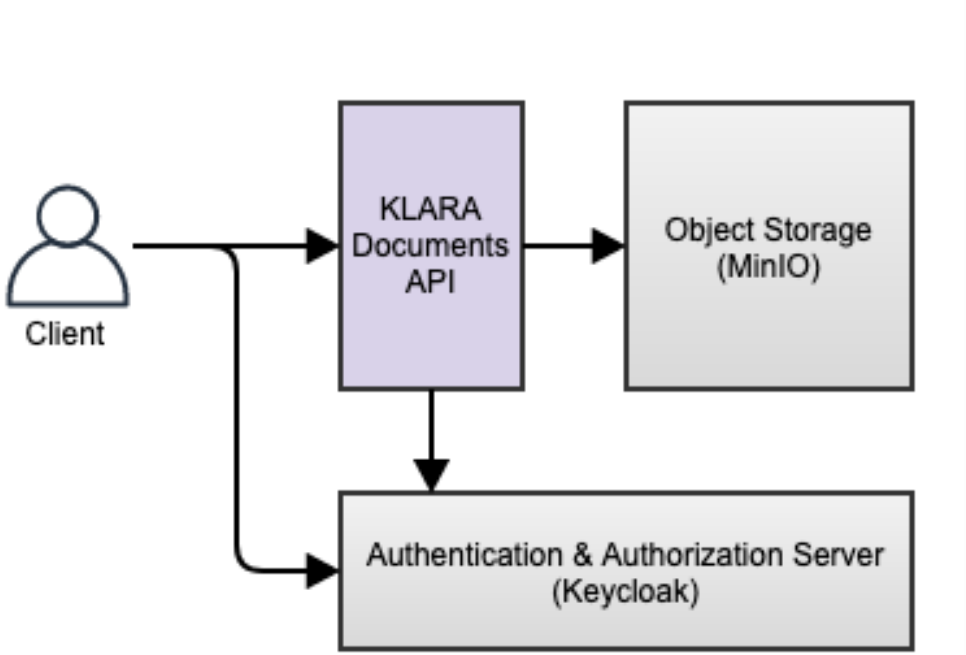
- Restrict access to resources for users of same tenant
- Allow cross-tenant sharing of resources



PoC – Proof of Concept

Story 5 + 6 – Stress Tests

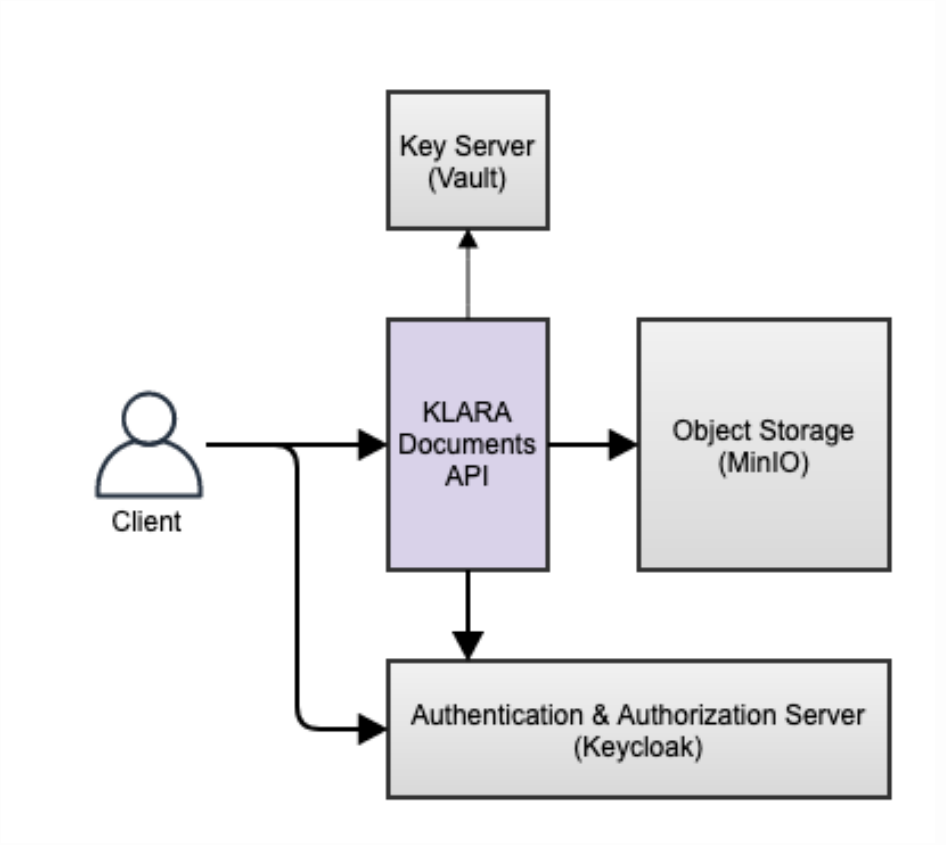
- Authentication Stress Test
  - Can we scale out for 10 million users and 100’000 users in parallel
- Authorization Stress Test
  - Can we scale out for 20 million resources



PoC – Proof of Concept

Next steps – Client-side Encryption

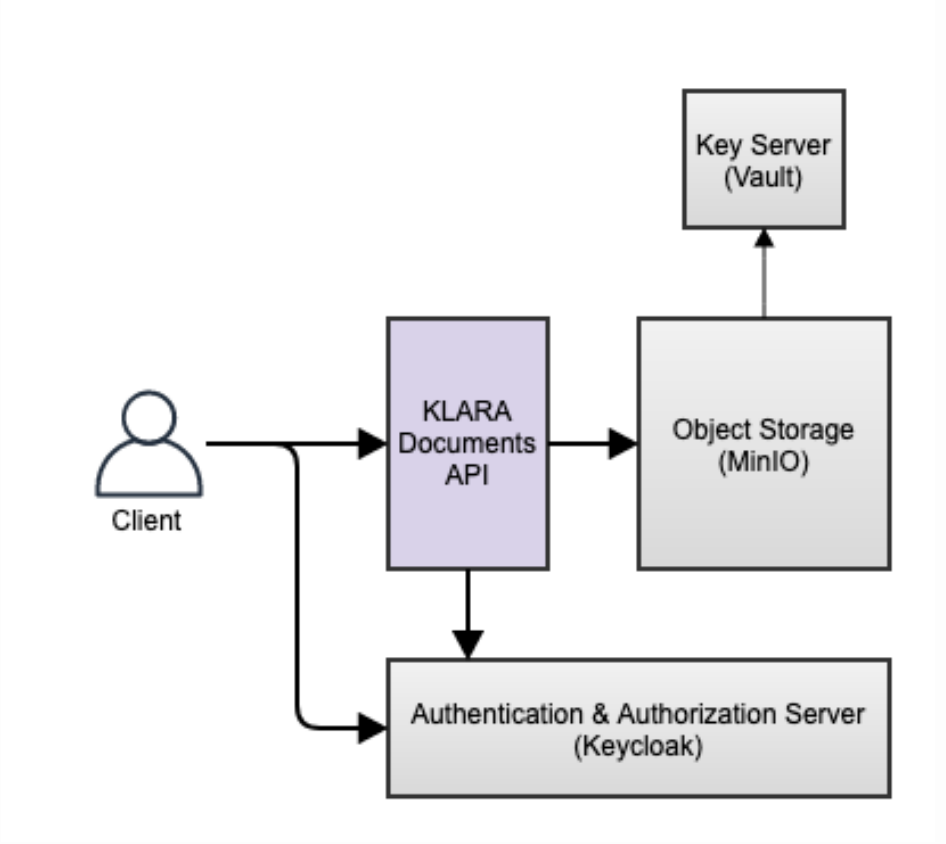
- Secure access to tenant specific encryption and decryption services



PoC – Proof of Concept

Next steps – Server-side Encryption

- Transparent, tenant specific encryption



Documents

Agenda

• Business aspects	Alessio Manzo
• Architecture	Laurent Salamin
• POC	Tobias Hofer
• Next steps	Jürg Schmocker

Documents – Next steps

Kepler team

- Start with POC of architecture in sprint beginning 19.1.2021
- Alessio will take over PO role for Kepler team from this sprint on
- Focus for Kepler: Documents project and Vaudoise Widget
- For technical POC questions: Documents chat in Skype

KLARA

Macht dein Büro einfach

WWW.KLARA.CH