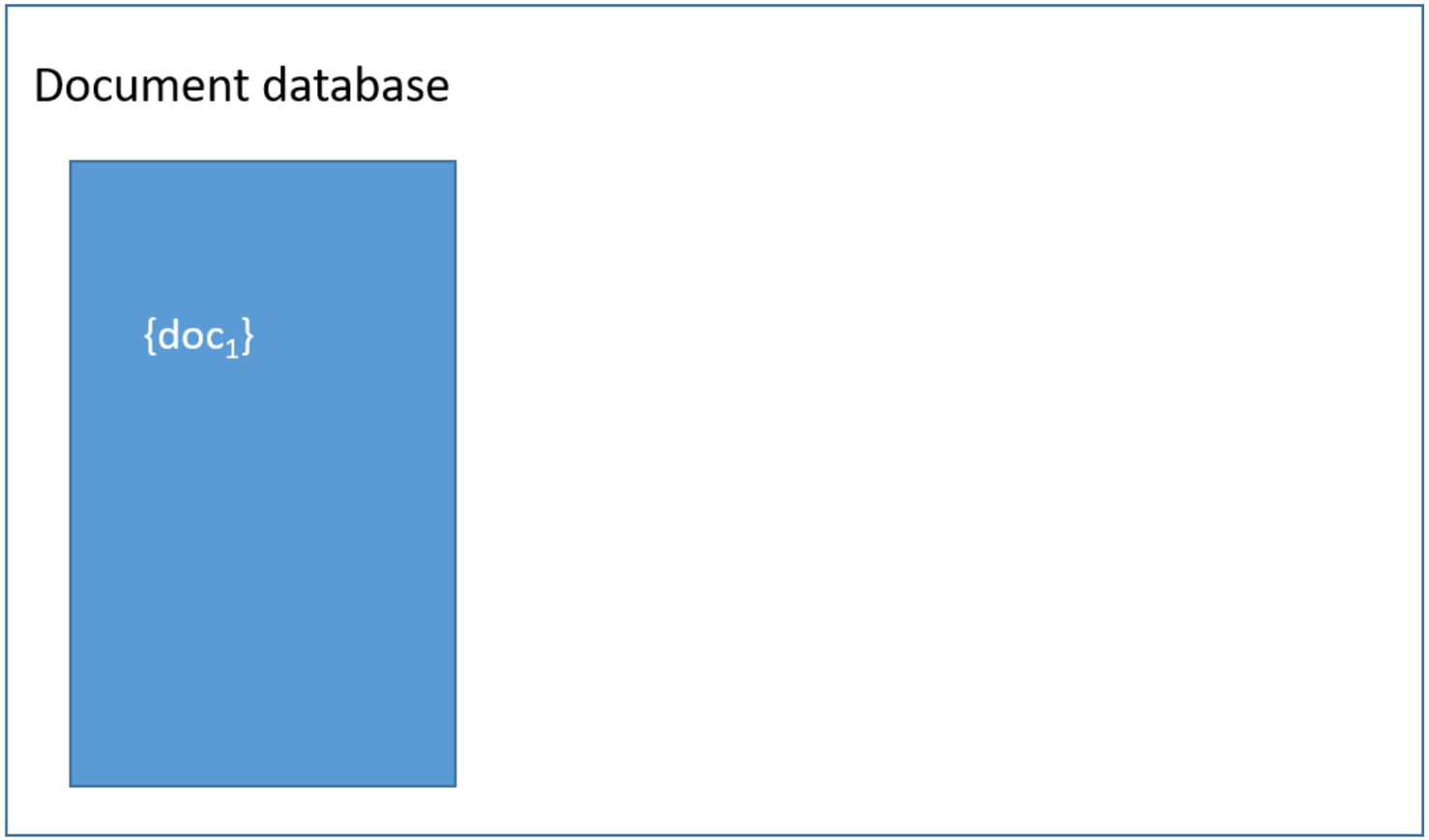
What is a document database?

NOSQL CONCEPTS

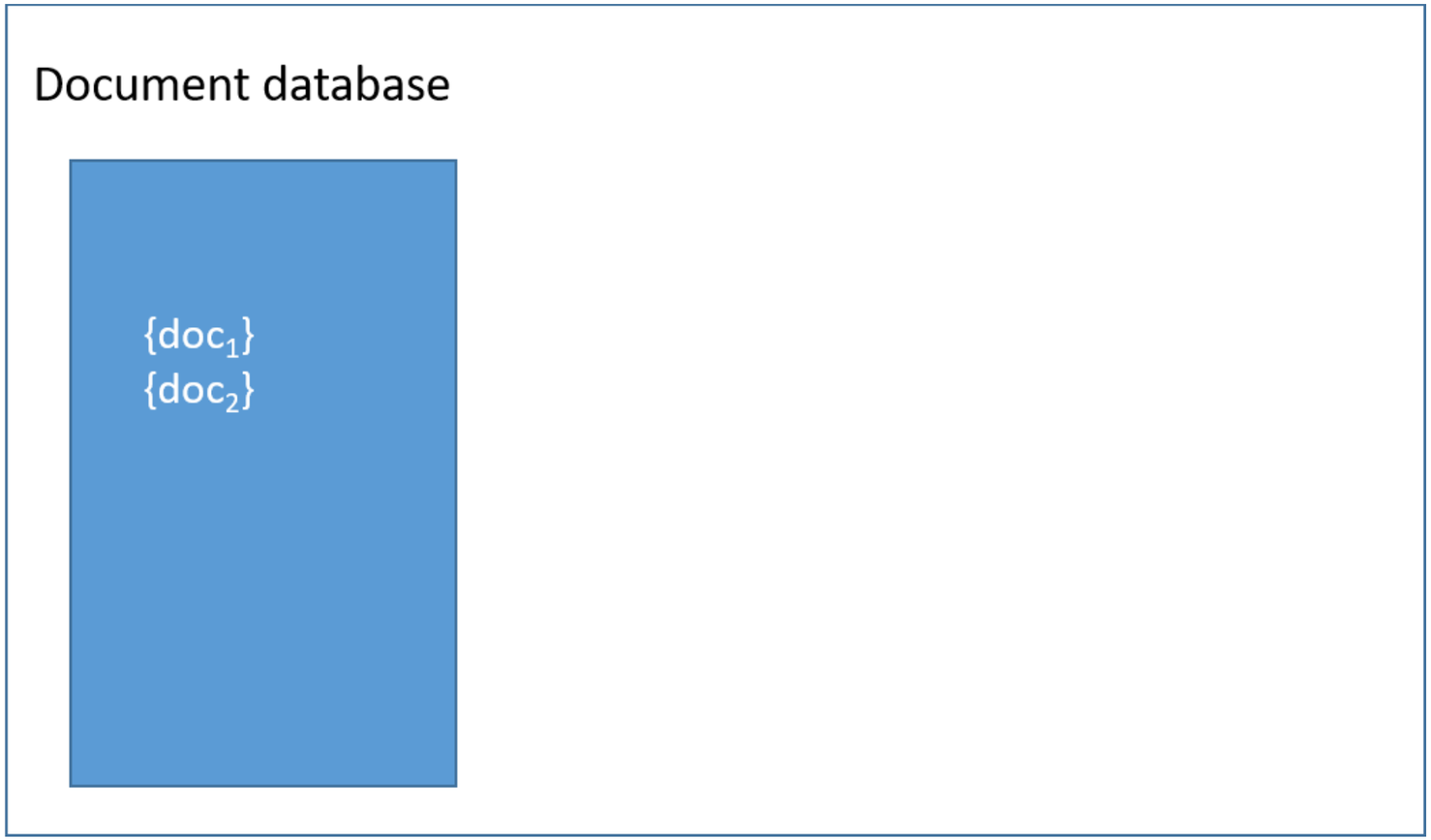
Miriam Antona
Software engineer

Document database - overview

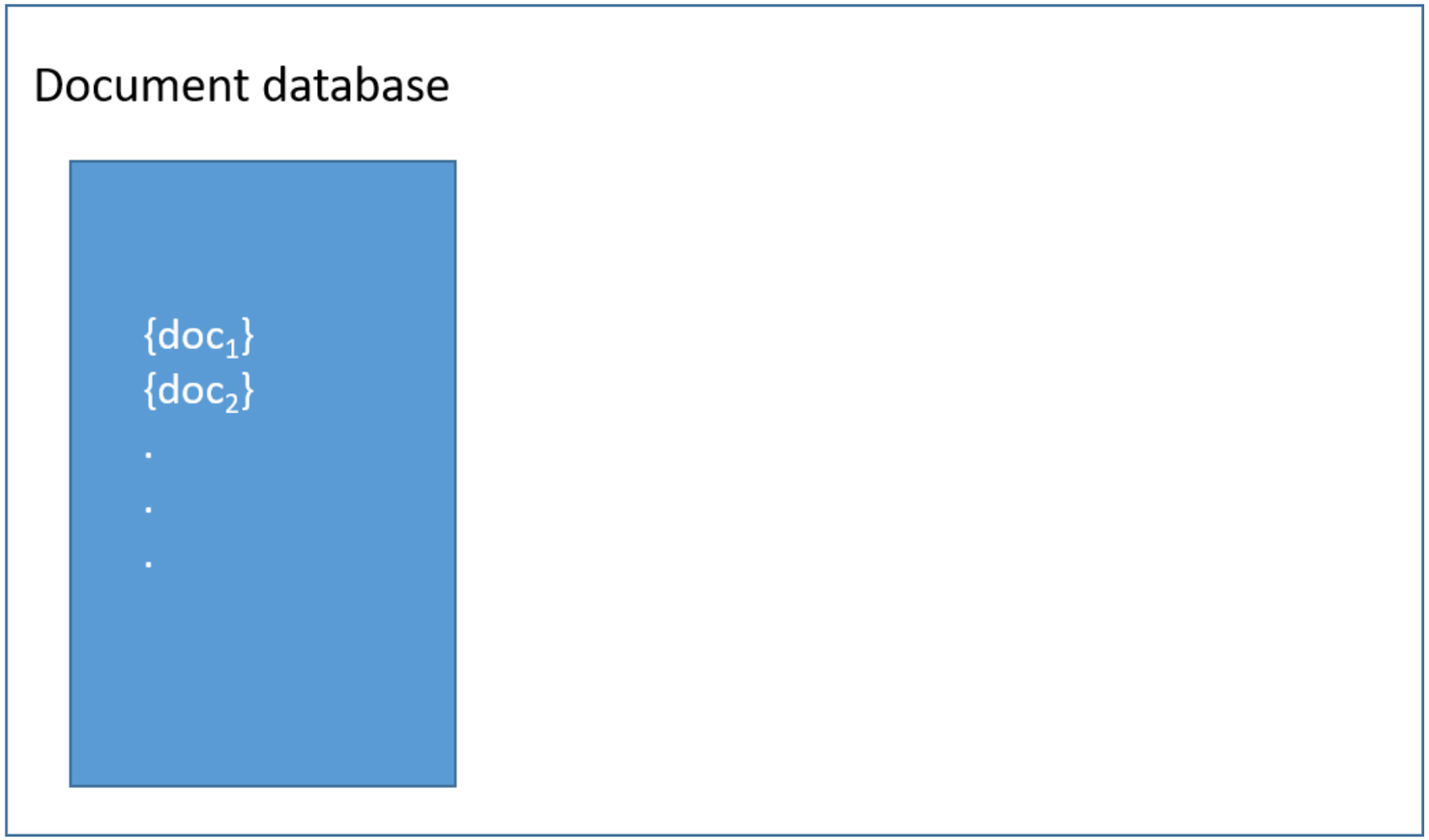
Document database - overview



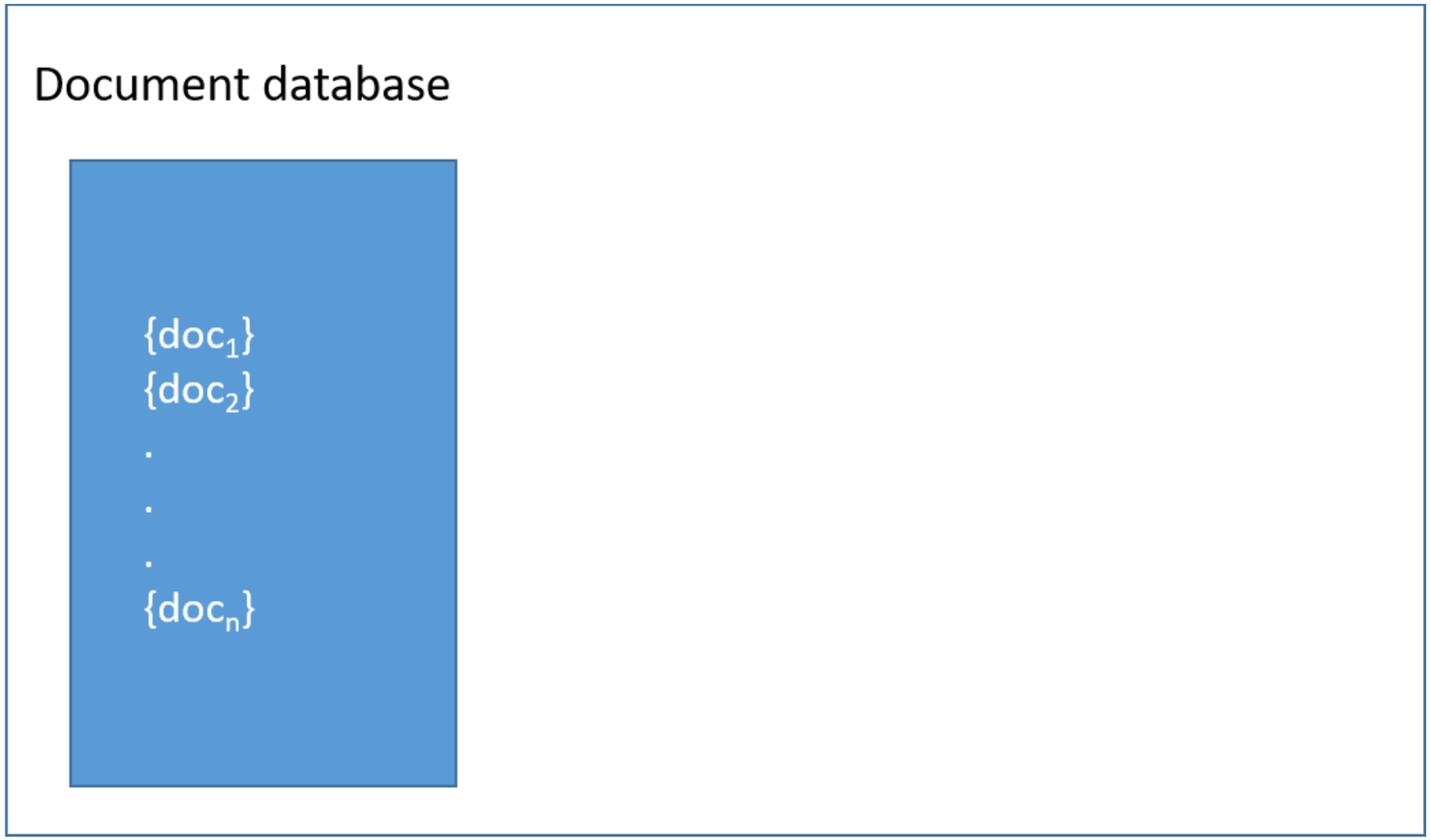
Document database - overview



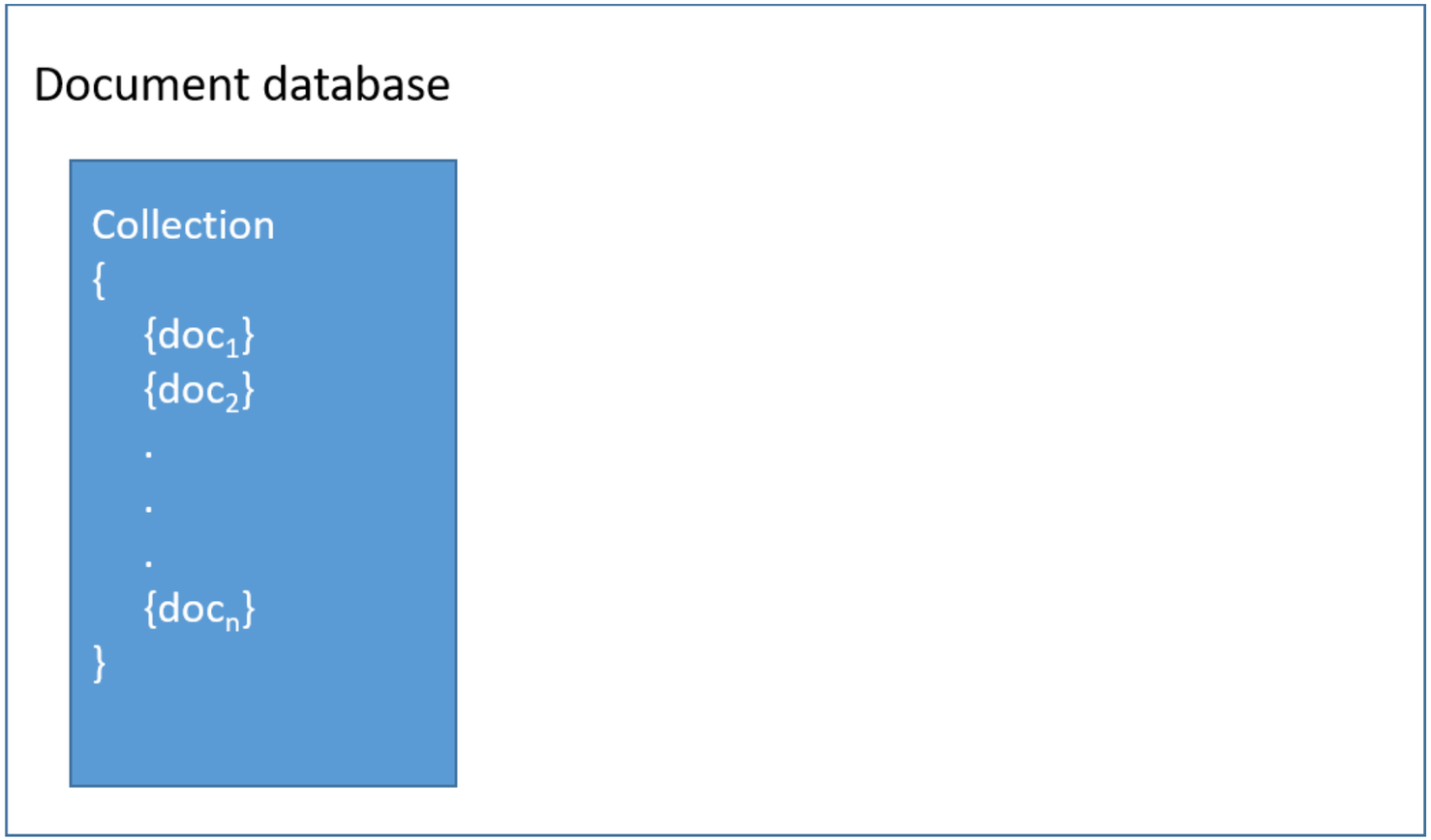
Document database - overview



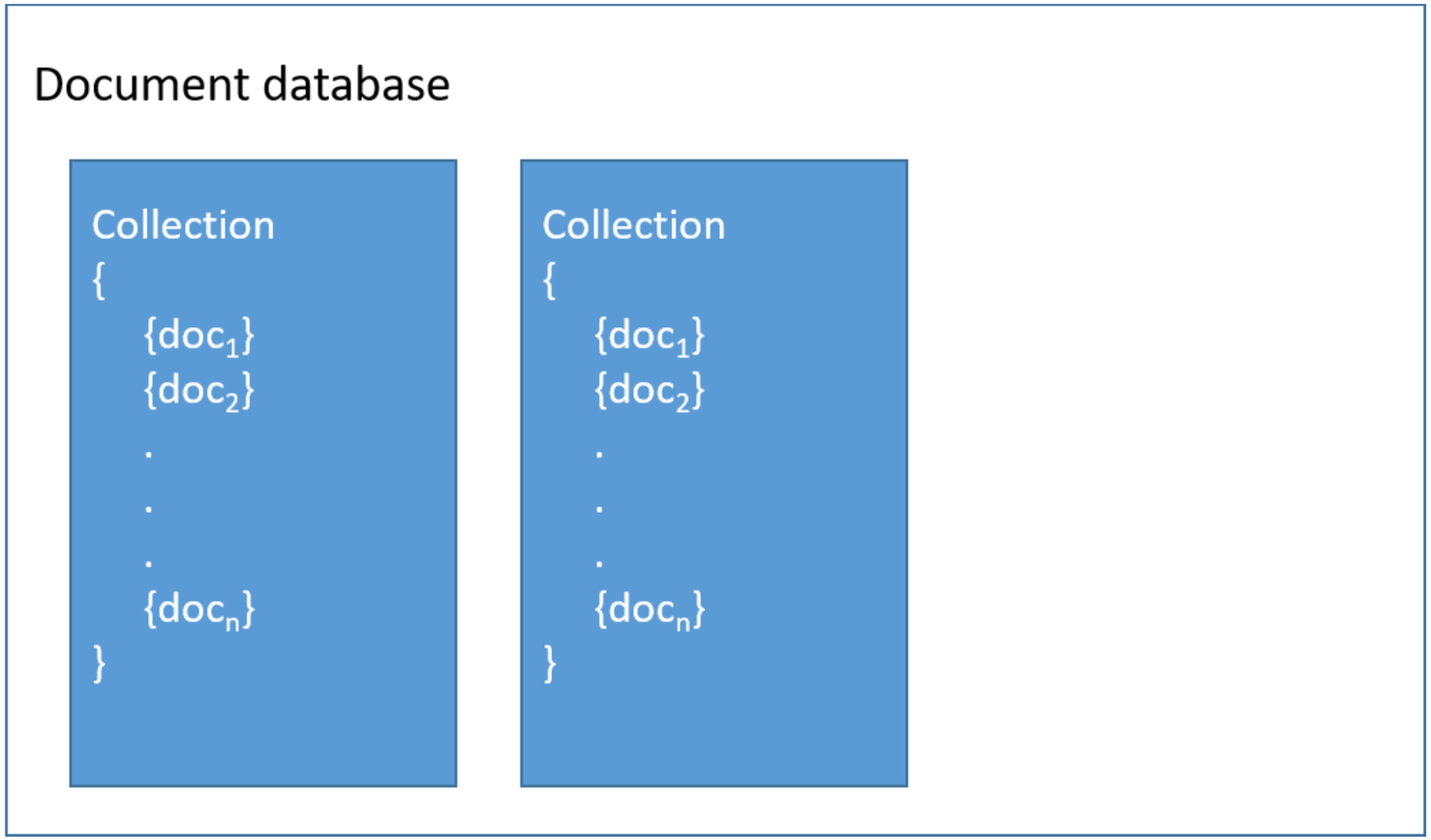
Document database - overview



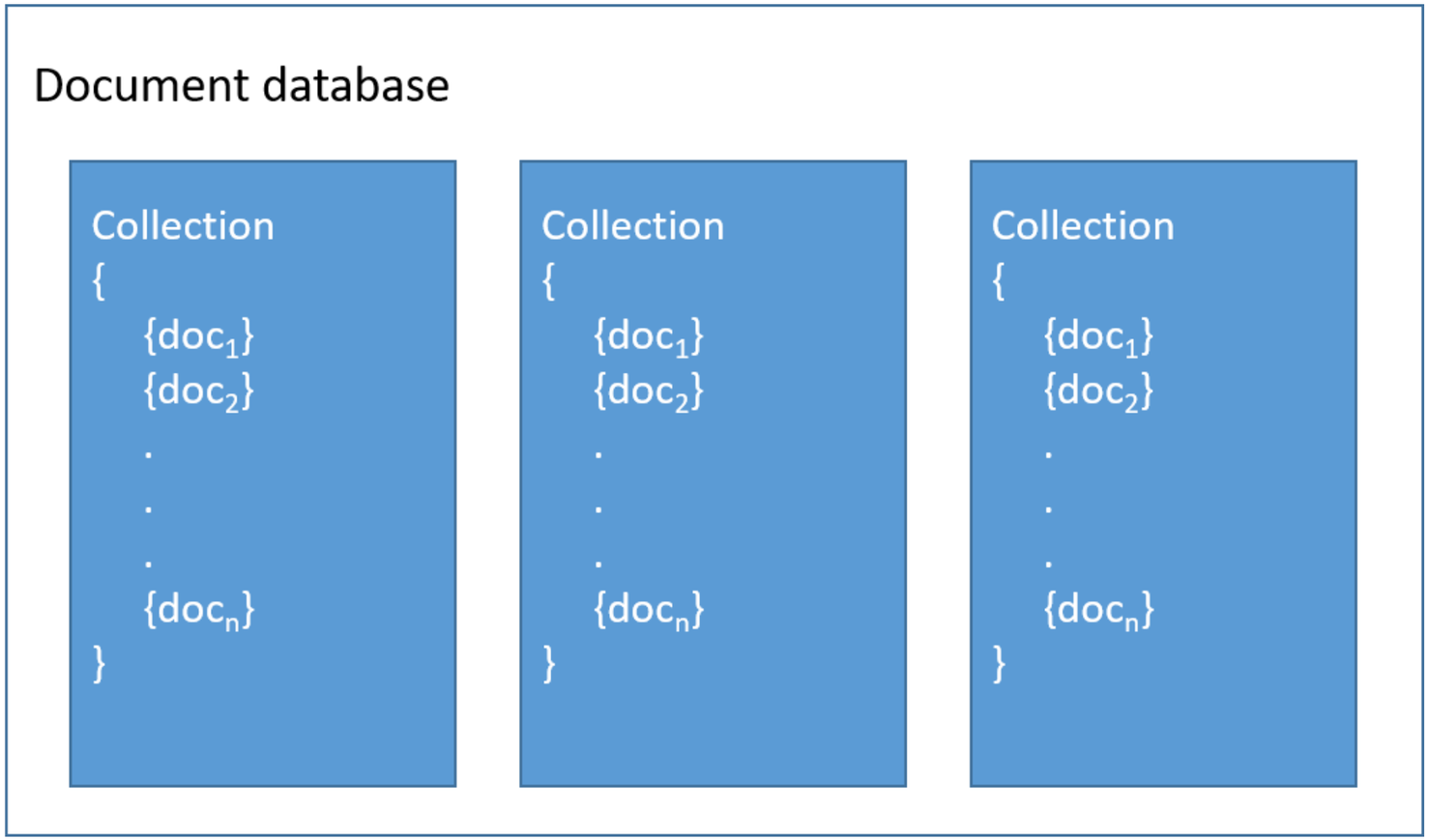
Document database - overview



Document database - overview



Document database - overview



- Documents -> rows
- Collections -> tables

Documents

- Set of key-value pairs
- **Keys:** strings
- **Values:** numbers, strings, booleans, arrays or objects
- **Schemaless:** no need to specify the structure
- **Formats:** JSON, BSON, YAML, or XML

Documents - JSON format

```
{
  "user_id": 512,
  "name": "Carol",
  "last_name": "Harper",
  "email": "carolharper@datazy.com",
  "address": {
    "street": "123 Sesame Street",
    "city": "New York City",
    "state": "New York",
    "country": "USA"
  },
  "hobbies": [
    "hiking",
    "painting"
  ]
}
```

Documents - queries

```
{
  "user_id": 512,
  "name": "Carol",
  "last_name": "Harper",
  "email": "carolharper@datazy.com",
  "address": {
    "street": "123 Sesame Street",
    "city": "New York City",
    "state": "New York",
    "country": "USA"
  },
  "hobbies": [
    "hiking",
    "painting"
  ]
}
```

- All the users who live in New York and like hiking
- All the users older than 40
- User's data by user_id
- ...

Documents - polymorphic model

```
{
  "user_id": 512,
  "name": "Carol",
  "last_name": "Harper",
  "email": "carolharper@datazy.com",
  "address": {
    "street": "123 Sesame Street",
    "city": "New York City",
    "state": "New York",
    "country": "USA"
  },
  "hobbies": [
    "hiking",
    "painting"
  ]
}
```

```
{
  "user_id": 513,
  "name": "Benjamin",
  "last_name": "Lieberman",
  "email": "benjaminlieberman@datazy.com",
  "date_of_birth": "07/04/1984",
  "hobbies": [
    "reading"
  ]
}
```

Collections

- Sets of documents
- Store the **same type of entities**
- **Organize** documents and collections by thinking about the **queries**

Popular document databases

DynamoDB

Couchbase

Firebase
Realtime Database

Let's practice!

NOSQL CONCEPTS

Advantages and limitations of document databases

NOSQL CONCEPTS

Miriam Antona
Software engineer

datacamp

Advantages - flexibility

- **Don't need** to predefine the **schema**
- **Documents can vary** over time
  - Avoids schema migrations
- Embedded documents **avoid joins**
  - Better times
- One of the **first reasons** to choose document databases

Advantages - intuitive for developers

- **Natural** way to work
- JSON is **human-readable**
- **Documents map objects** in code
  - Less coding
  - Simpler and faster development
  - Start coding and storing objects as documents are created
- **Easier** for new developers

Advantages - horizontal scalability

- Sharding

Limitations - more responsibility

- Care about data in the **application code**
  - e.g. check required email
- Care about **redundant data**
  - e.g. modify duplicated name

Let's practice!

NOSQL CONCEPTS

When to use document databases

NOSQL CONCEPTS

Miriam Antona
Software engineer

datacamp

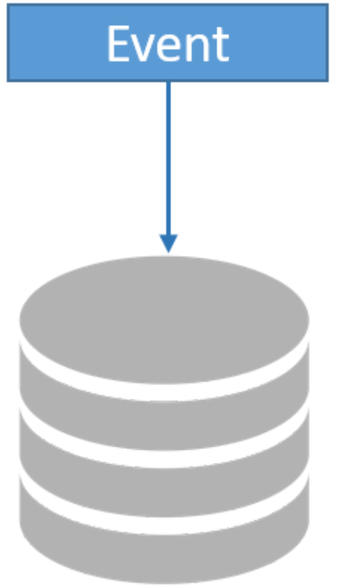
Suitable cases

- **Catalogs**
  - E-commerce web sites/applications store product information
  - Different attributes between the products
  - Embed related information

```
{
  "product_id": 879,
  "name": "Fashion shirt",
  "category": {
    "category_id": 15,
    "name": "Tops & t-shirts",
    "type": "Shirt"
  }
}
```

Suitable cases

- **Event logging**
  - Types of events:
    - User logging
    - Product purchase
    - Errors
    - ...
  - Sharding by:
    - Time
    - Type of event
    - ...




```
{
  "type": "info",
  "message": "user_logged",
  "user_id": 551,
  ...
}
```

Suitable cases

- **User profiles**

```
{
  "user_id": 512,
  "name": "Carol",
  "last_name": "Harper",
  "email": "carolharper@datazy.com",
  "address": {
    "street": "123 Sesame Street",
    "city": "New York City",
    "state": "New York"
  },
  ...
}
```

- Information may vary
- Document flexibility

Suitable cases

- **Content management systems**
  - Blogs, video platforms, etc.
  - Users' content
    - Comments
    - Images
    - Videos
    - ...


```
{
  "id": 458,
  "url": "myblog/datazy.com",
  "title": "How to write a blog entry at Datazy",
  "tags": [
    "Datazy",
    "Blog"
  ],
  "last10comments": [
    { "name": "Eliza", "comment": "Great!" },
    { "name": "Eric", "comment": "Thank you!"}
  ]...
}
```

Suitable cases

- **Real-time analytics**
  - Page views, unique visitors...
  - Easy to store the information

```
{
  "_id": "1000241",
  "hour": "Sat Jun 12 2021 16:40:00 GMT+0200 (EST)",
  "site": "datazy",
  "uniques": 5,
  "pageviews": 15,
  ...
}
```

Unsuitable cases

- Very structured data
- Always have consistent data

Let's practice!

NOSQL CONCEPTS

MongoDB case study

NOSQL CONCEPTS

Miriam Antona
Software engineer

datacamp

MongoDB - overview

- Popular document database
- BSON (Binary JSON) format

MongoDB - features

- MongoDB Query Language (**MQL**)

```
db.users.find({ "address.zipcode" : "10245" })
```

- Native drivers for **programming languages**: C#, Java, Python, Scala, etc.
- **Indexes** on any field
- **ACID** transactions (Atomicity, Consistency, Isolation, and Durability)
- **Joins** in queries

MongoDB - features

- **Scale horizontally**
  - native sharding
  - add/move shards
- **Replication**
  - 50 copies of our data

MongoDB - products

- **MongoDB Compass:**
  - Free GUI
  - Explore schema, create queries visually...
- **MongoDB Atlas:**
  - Cloud service
  - AWS, Azure, Google Cloud
- **MongoDB Enterprise Advanced:**
  - Run MongoDB in our infrastructure

MongoDB - products

- **MongoDB Atlas Lake:**
  - Query and analyze data
  - AWS S3 and MongoDB Atlas
  - MQL
- **MongoDB Charts:**
  - Visualizations of the data
- **Realm Mobile Database:**
  - Store data locally on iOS or Android

MongoDB - popular uses

- **Single view applications:** financial services, government, high tech, retail...
- **Gaming:** player profiles, leaderboards...
- **Catalogs:** financial services, government, high tech, retail...
- **Real-time analytics**
- **Content management**
- **Internet of Things**

MongoDB - customers

TOYOTA

ThermoFisher
SCIENTIFIC

Shutterfly case study - overview

- Online photography service
  - Share personalized photo albums
  - Products with printed photographs
- Millions of customers
- More than six billion images

Shutterfly case study - problem and solution

- Massive data growth
- Performance limits of Oracle
- Long time to build applications
- Applications didn't perform quickly enough
- Oracle became too expensive
- Chose MongoDB

Shutterfly case study - results

- Performance improvement
  - Inserts 400 ms. -> 2 ms.
- Horizontal scaling
- Flexible schema -> quickly development
  - Tags, comments, etc. are not difficult to implement
- New query patterns
- Cost reduction

Shutterfly case study - results

- Performance improvement
  - Inserts 400 ms. -> 2 ms.
- Horizontal scalling
- Flexible schema -> quickly development
  - Tags, comments, etc. are not difficult to implement
- New query patterns
- Cost reduction

[1] https://www.mongodb.com/who-uses-mongodb

Let's practice!

NOSQL CONCEPTS