

Moduł 6

Bazy danych dla Big Data

Część 1 - DB zorientowane na dokumenty

iBigWorld:
Innovations for Big Data in a Real World

ULSIT team

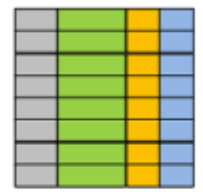
Disclaimer: Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the National Agency (NA). Neither the European Union nor NA can be held responsible for them.



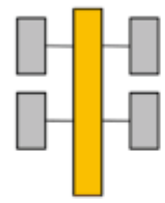
Typy baz danych NoSQL

SQL Database

Relational



Analytical (OLAP)



NoSQL Database

Column-Family



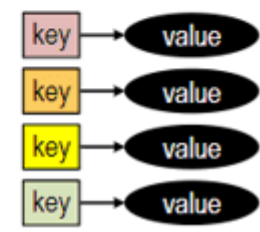
Graph



Document



Key-Value



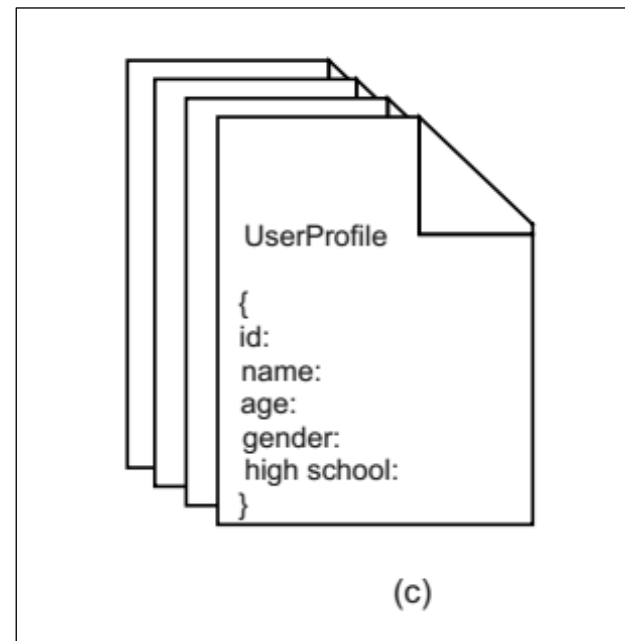
Rodzaje bazy danych NoSQL: Dokument

Zorientowany na dokument:

Model danych zorientowany na dokument jest podobny do struktury klucz-wartość i przechowuje dane w postaci klucza i wartości jako odniesienie do dokumentu.

Baza danych zorientowana na dokumenty określana jako sklep z dokumentami.

Bazy danych dokumentów obsługują bardziej złożone zapytania i relacje hierarchiczne. Ten model danych zazwyczaj implementuje format JSON i oferuje bardzo elastyczny schemat



**Przykład
modeli danych zorientowanych
na dokumenty:**

Rodzaje bazy danych NoSQL: Dokument

- Główne cechy
- Podobny do DB klucz-wartość, ale wartości przechowują XML, JSON lub BSON
- Oferuje przechowywanie złożonych danych jak drzewa, kolekcje, słowniki
- Nie obsługuje relacji i połączeń
- Brak języka zapytań
- Przykłady
- MongoDB, CouchDB, OrientDB

Do czego służą bazy danych dokumentów?

- Projektowanie
- Zapytanie
- Przechowywanie danych w formacie dokumentu (dokument JSON, XML, YAML lub formaty binarne, takie jak BSON i PDF)



Document data model

Popularne bazy danych o wartościach kluczowych

include secondary database models

53 systems in ranking, December 2021

Rank			DBMS	Database Model	Score		
Dec 2021	Nov 2021	Dec 2020			Dec 2021	Nov 2021	Dec 2020
1.	1.	1.	MongoDB +	Document, Multi-model i	484.67	-2.67	+26.95
2.	2.	2.	Amazon DynamoDB +	Multi-model i	77.63	+0.64	+8.51
3.	3.	3.	Microsoft Azure Cosmos DB +	Multi-model i	39.71	-1.11	+6.17
4.	4.	4.	Couchbase +	Document, Multi-model i	28.45	-1.42	-3.37
5.	5.	5.	Firebase Realtime Database	Document	19.96	-0.10	+3.01
6.	6.	6.	CouchDB	Document, Multi-model i	16.53	-0.27	-0.32
7.	↑ 9.	↑ 8.	Realm +	Document	9.80	+0.60	+0.65
8.	8.	↓ 7.	MarkLogic +	Multi-model i	8.94	-0.40	-2.00
9.	↓ 7.	9.	Google Cloud Firestore	Document	8.64	-0.92	+0.24
10.	↑ 12.	↑ 20.	Virtuoso +	Multi-model i	5.07	+0.25	+2.48
11.	↓ 10.	↓ 10.	ArangoDB +	Multi-model i	4.75	-0.35	-0.76
12.	↓ 11.	↓ 11.	Google Cloud Datastore	Document	4.44	-0.54	-1.07
13.	13.	↓ 12.	OrientDB	Multi-model i	4.40	-0.24	-0.89
14.	14.	↓ 13.	Oracle NoSQL	Multi-model i	4.23	-0.18	-0.19
15.	15.	15.	IBM Cloudant	Document	3.84	-0.42	-0.33
16.	16.	16.	RavenDB +	Document, Multi-model i	3.24	-0.19	-0.88
17.	↑ 18.	↓ 14.	RethinkDB	Document, Multi-model i	3.22	+0.03	-1.08
18.	↓ 17.	↓ 17.	PouchDB	Document	3.09	-0.25	-0.44
19.	19.	↓ 18.	Apache Drill	Multi-model i	2.74	-0.22	-0.40

By **DB-Engines Ranking**

<https://db-engines.com/en/ranking/key-value+store>



MongoDB

- MongoDB to system do przetwarzania baz danych dokumentów.
- MongoDB jest rodzajem nierelacyjnej bazy danych (NoSQL) i jest open-source. Jest rozwijany jak wysoce dostępne, skalowalne i odporne na błędy rozwiązanie zorientowane na dokumenty. Zamiast przechowywania informacji w tabelach, jak w tradycyjnych relacyjnych bazach danych

•

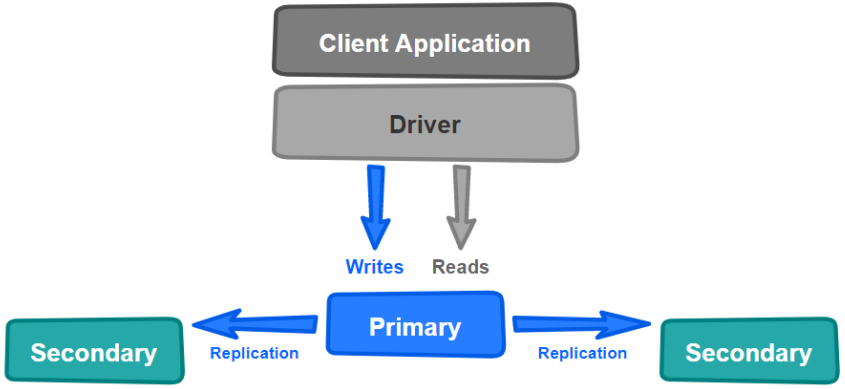


Mongo DB

- Wolne i otwarte źródło bazy danych zorientowanej na dokumenty
- Dokumenty mogą zawierać wiele różnych par klucz-wartość lub pary klucz-array, a nawet dokumenty zagnieżdżone
- Oferuje wyszukiwanie, indeksowanie i agregację
- Wbudowana wysoka dostępność, skalowanie poziome i dystrybucja geograficzna
- Biblioteki dla wielu języków programowania (C, Go, PHP, Ruby, itp.)
- Obsługiwane na systemach Windows, Linux i macOS
- Rozwiązanie hostowane jest również dostępne przez MongoDB Atlas

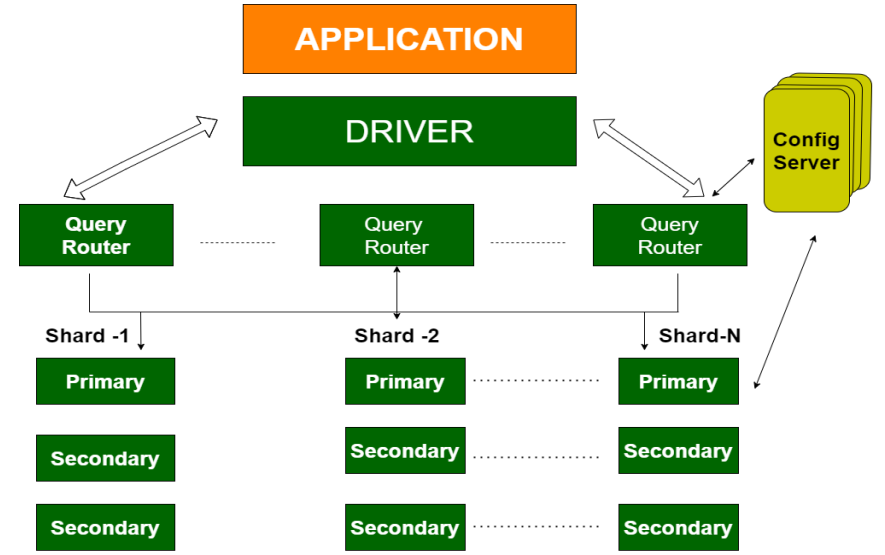


Diagram architektury MongoDB



- Single

Sharing



https://www.researchgate.net/publication/330970143_A_comparative_study_of_NoSQL_databases_MongoDB_and_HBase/figures?lo=1
<https://geekflare.com/mongodb-sharding/>
https://www.researchgate.net/publication/330970143_A_comparative_study_of_NoSQL_databases_MongoDB_and_HBase/figures?lo=1

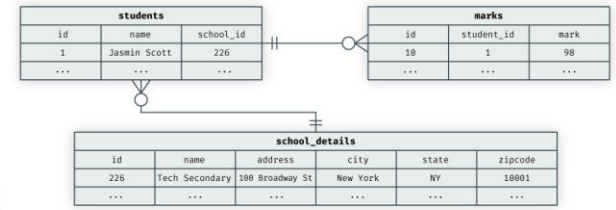
Obraz jest wart tysiąca słów

mongoDB.

```
{
  "_id": 1,
  "student_name": "Jasmin Scott",
  "school": {
    "school_id": 226,
    "name": "Tech Secondary",
    "address": "100 Broadway St",
    "city": "New York",
    "state": "NY",
    "zipcode": "10001"
  },
  "marks": [98, 93, 95, 88, 100],
}
```

```
mongo
> db.students.find({"student_name": "Jasmin Scott"})
```

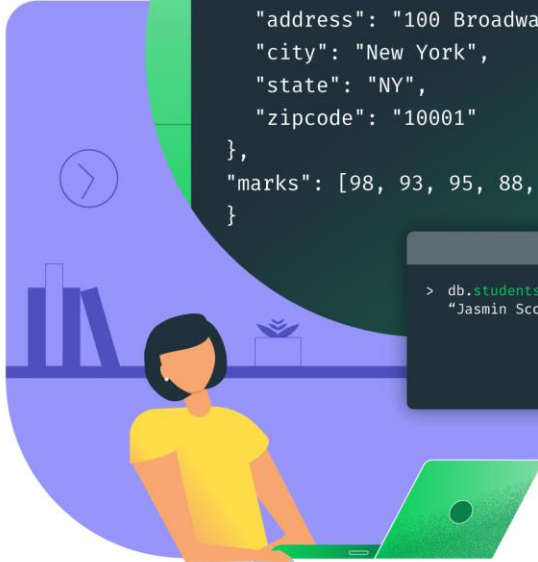
SQL



Results

name	mark	school_name	city
Jasmin Scott	98	Tech Secondary	New York

```
sql
SELECT s.name, m.mark, d.name as "school_name",
d.city
FROM students s
INNER JOIN marks m ON s.id = m.student_id
INNER JOIN school_details d ON s.school_id = d.id
WHERE s.name = "Jasmin Scott";
```



<https://docs.mongodb.com/>

Przypadki użycia dla MongoDB

Przypadki użycia baz danych zorientowanych na dokumenty obejmują profile użytkowników w sieciach społecznościowych, analizę stron internetowych oraz złożone aplikacje danych transakcyjnych.

- **Facebook**
- **LinkedIn**
- **Book Database**
- **McAfee**
- **SAP**
- **Yandex**



MongoDB - wiersz poleceń w akcji

```

C:\> mongosh mongodb+srv://cluster0.mubsb.mongodb.net/myFirstDatabase
Using MongoDB:      4.4.10
Using Mongosh:     1.1.6

For mongosh info see: https://docs.mongodb.com/mongosh-shell/

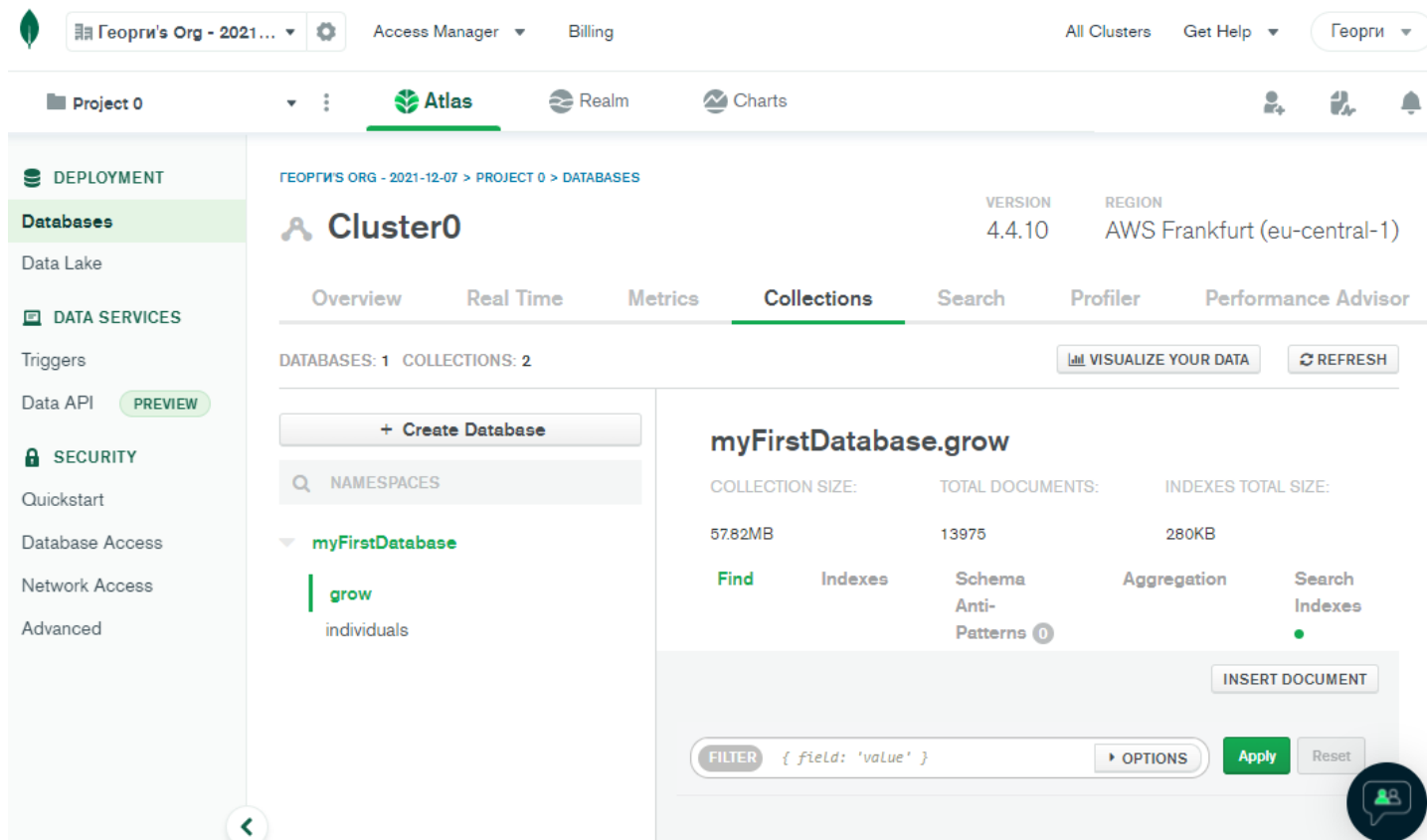
Atlas atlas-irmhob-shard-0 [primary] myFirstDatabase> db.createCollection('customers');
{ ok: 1 }
Atlas atlas-irmhob-shard-0 [primary] myFirstDatabase>

Atlas atlas-irmhob-shard-0 [primary] myFirstDatabase> show collections;
customers
grow
individuals
Atlas atlas-irmhob-shard-0 [primary] myFirstDatabase> db.customers.insert({customer_id: "i-001", name: "Gerogi Dimitrov",
balance: 1000});
DeprecationWarning: Collection.insert() is deprecated. Use insertOne, insertMany, or bulkWrite.
{
  acknowledged: true,
  insertedIds: { '0': ObjectId("61b06d5e37f970d54b29fb27") }
}
Atlas atlas-irmhob-shard-0 [primary] myFirstDatabase> db.customers.find();
[
  {
    _id: ObjectId("61b06d5e37f970d54b29fb27"),
    customer_id: 'i-001',
    name: 'Gerogi Dimitrov',
    balance: 1000
  }
]
Atlas atlas-irmhob-shard-0 [primary] myFirstDatabase>

```



MongoDB - Kompas w działaniu



The screenshot displays the MongoDB Atlas web interface. At the top, there's a navigation bar with 'Geopriv's Org - 2021...' and 'Access Manager'. The main header shows 'Project 0' and 'Atlas'.

The left sidebar contains navigation options: DEPLOYMENT, Databases, Data Lake, DATA SERVICES, Triggers, Data API (PREVIEW), SECURITY, Quickstart, Database Access, Network Access, and Advanced.

The main content area shows the 'Cluster0' overview. It includes a breadcrumb trail: 'GEOPRIV'S ORG - 2021-12-07 > PROJECT 0 > DATABASES'. The cluster details show 'VERSION 4.4.10' and 'REGION AWS Frankfurt (eu-central-1)'. Below this are tabs for Overview, Real Time, Metrics, Collections (selected), Search, Profiler, and Performance Advisor.

The 'Collections' tab shows 'DATABASES: 1' and 'COLLECTIONS: 2'. A '+ Create Database' button is visible. Under 'NAMESPACES', there's a tree view showing 'myFirstDatabase' with a sub-entry 'grow' containing 'individuals'.

The right panel displays details for 'myFirstDatabase.grow':

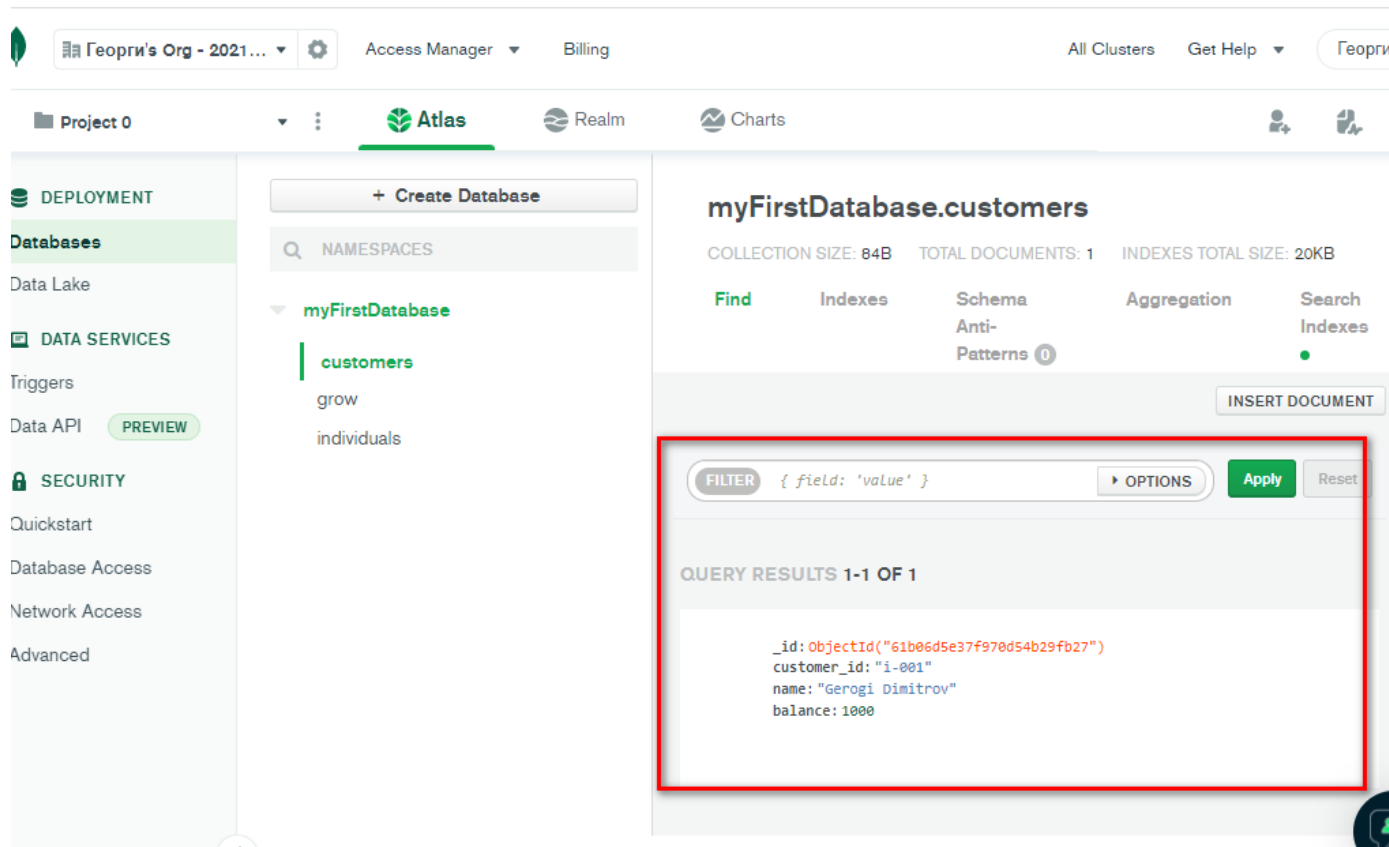
COLLECTION SIZE:	TOTAL DOCUMENTS:	INDEXES TOTAL SIZE:
57.82MB	13975	280KB

 Below the table are tabs for 'Find', 'Indexes', 'Schema Anti-Patterns 0', 'Aggregation', and 'Search Indexes'. An 'INSERT DOCUMENT' button is at the bottom right. A filter bar at the bottom contains the text '{ field: 'value' }' and 'Apply' and 'Reset' buttons.

<https://www.mongodb.com/>



MongoDB - Kompas w działaniu



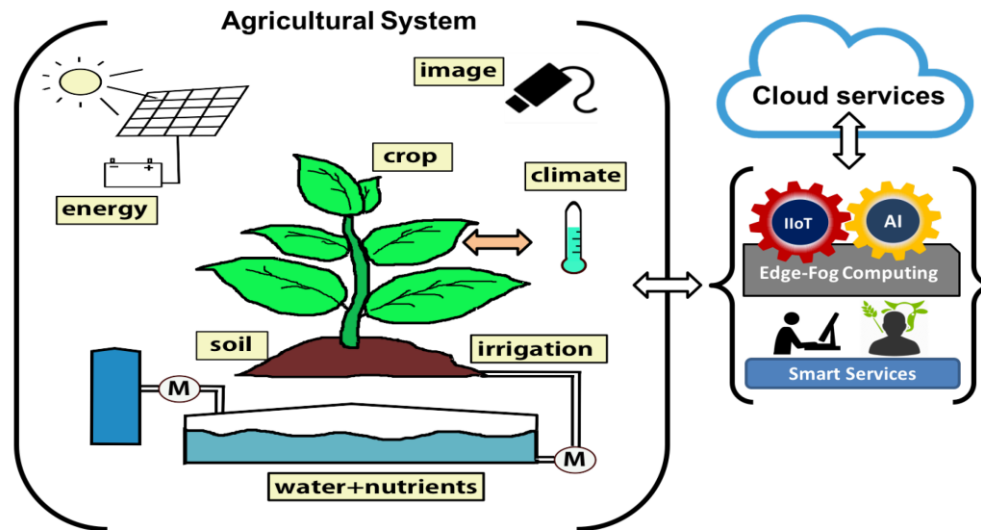
The screenshot displays the MongoDB Atlas web interface. The top navigation bar includes 'Geopriv's Org - 2021...', 'Access Manager', 'Billing', 'All Clusters', 'Get Help', and 'Geopriv'. The main interface shows 'Project 0' with 'Atlas' selected. The left sidebar contains navigation options: DEPLOYMENT, Databases, Data Lake, DATA SERVICES, SECURITY, and Quickstart. The central panel shows a '+ Create Database' button and a search for 'NAMESPACES'. Under 'myFirstDatabase', the 'customers' namespace is selected. The right panel displays 'myFirstDatabase.customers' with collection statistics: COLLECTION SIZE: 84B, TOTAL DOCUMENTS: 1, INDEXES TOTAL SIZE: 20KB. A 'Find' button is active, and a filter is applied: { field: 'value' }. The query results show one document:

```
{
  "_id": ObjectId("61b06d5e37f970d54b29fb27"),
  "customer_id": "i-001",
  "name": "Gerogi Dimitrov",
  "balance": 1000
}
```



MongoDB - Real Use Case Project "IncYields"

“Badania i kontrola wilgotności gleby w celu optymalizacji kosztów nawadniania i zwiększenia plonów.



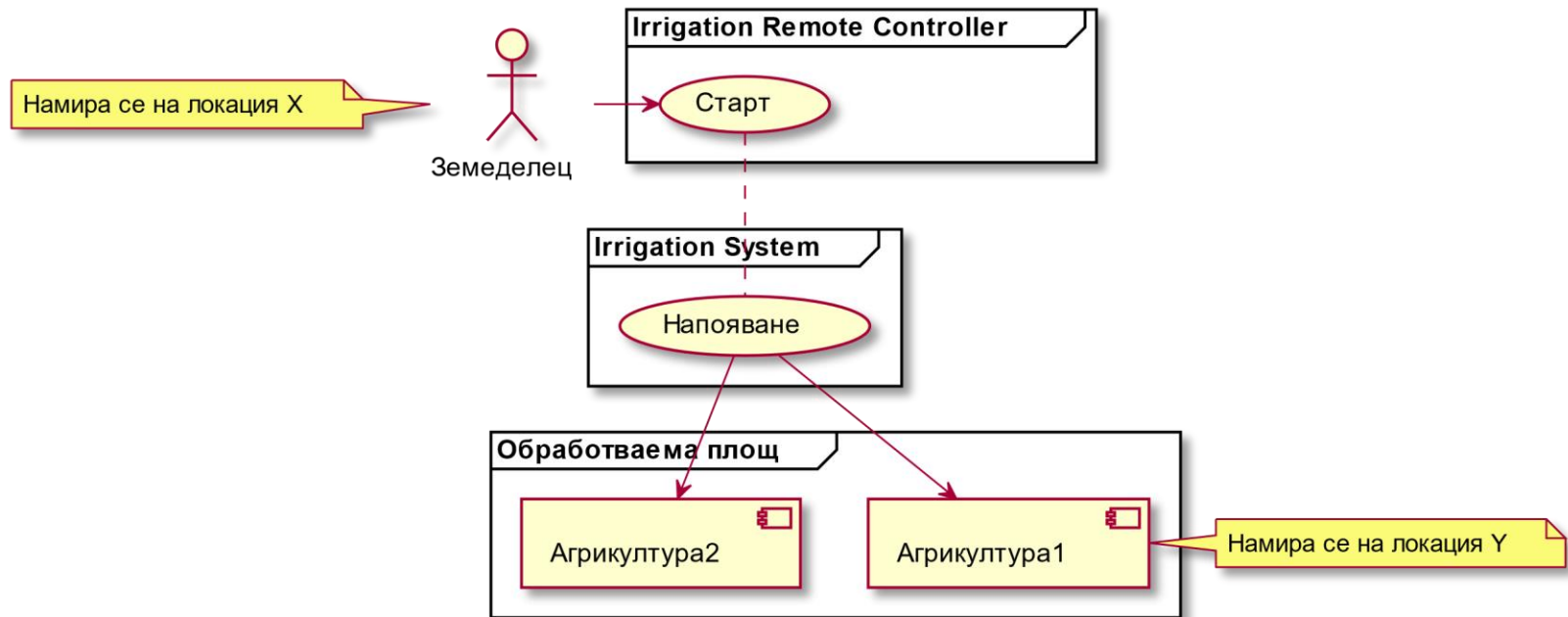
https://news.criticghana.com/single_post.php?nid=39

MongoDB - Real Use Case Project "IncYields"

Woda rozprzestrzenia się z różną prędkością i w różny sposób w różnych podłożach. Pomiar tych wartości pozwala nam na optymalizację zużycia wody do nawadniania tylko w niezbędnych granicach i we właściwym czasie. To z kolei prowadzi do tańszych i bardziej licznych produktów.

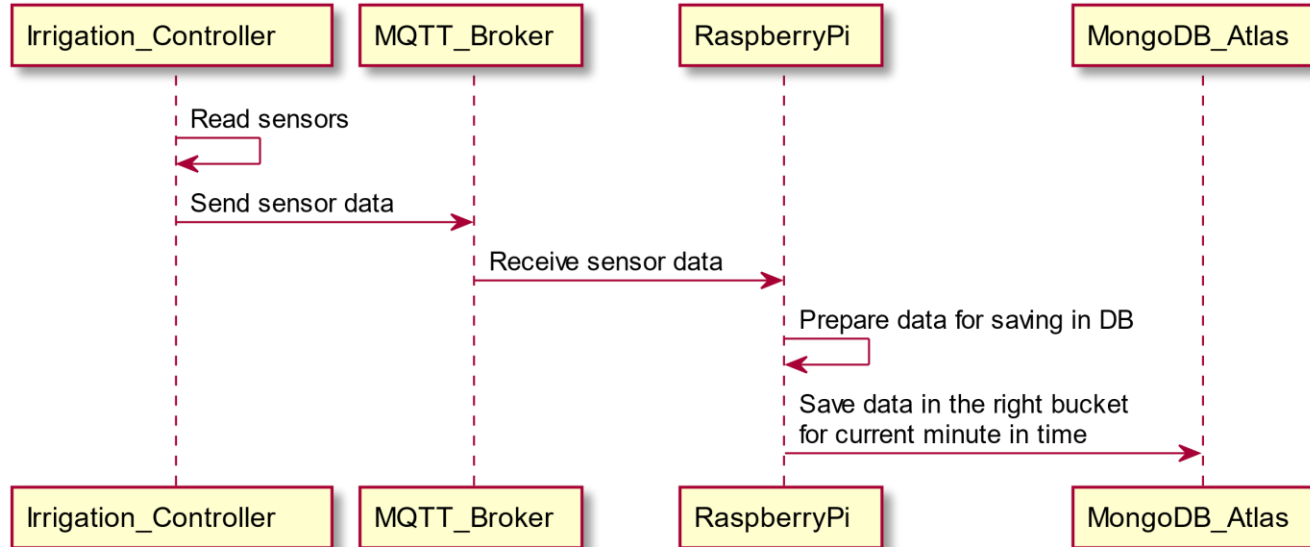
MongoDB - rzeczywisty przypadek użycia - GROW

Wyślij proces danych rzeczywistych -> do zbioru MongoDB "Grow"



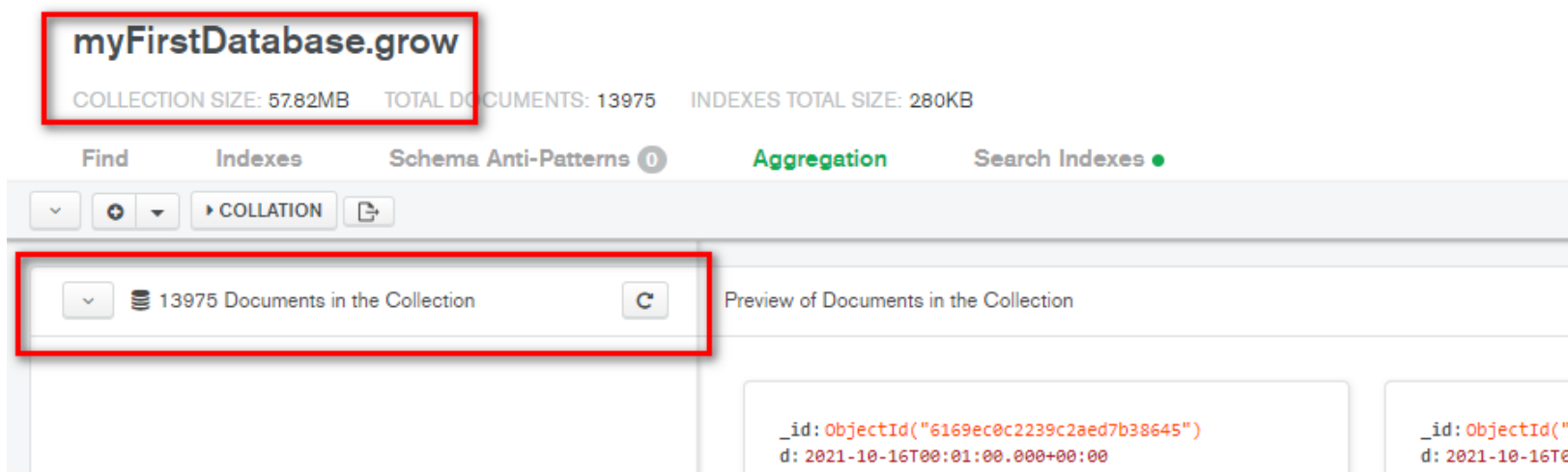
MongoDB - rzeczywisty przypadek użycia - GROW

Modules



MongoDB - Rekordy w kolekcji Grow

Liczba przechowywanych dokumentów

A screenshot of the MongoDB Compass web interface. The database name 'myFirstDatabase.grow' is highlighted with a red box. Below it, the collection statistics are shown: 'COLLECTION SIZE: 57.82MB', 'TOTAL DOCUMENTS: 13975', and 'INDEXES TOTAL SIZE: 280KB'. The 'Aggregation' tab is selected. A dropdown menu is open, showing '13975 Documents in the Collection' and a refresh icon, also highlighted with a red box. Below the dropdown, a preview of documents is visible, showing fields like '_id' and 'd' with their respective values.

myFirstDatabase.grow

COLLECTION SIZE: 57.82MB TOTAL DOCUMENTS: 13975 INDEXES TOTAL SIZE: 280KB

Find Indexes Schema Anti-Patterns 0 Aggregation Search Indexes ●

COLLATION

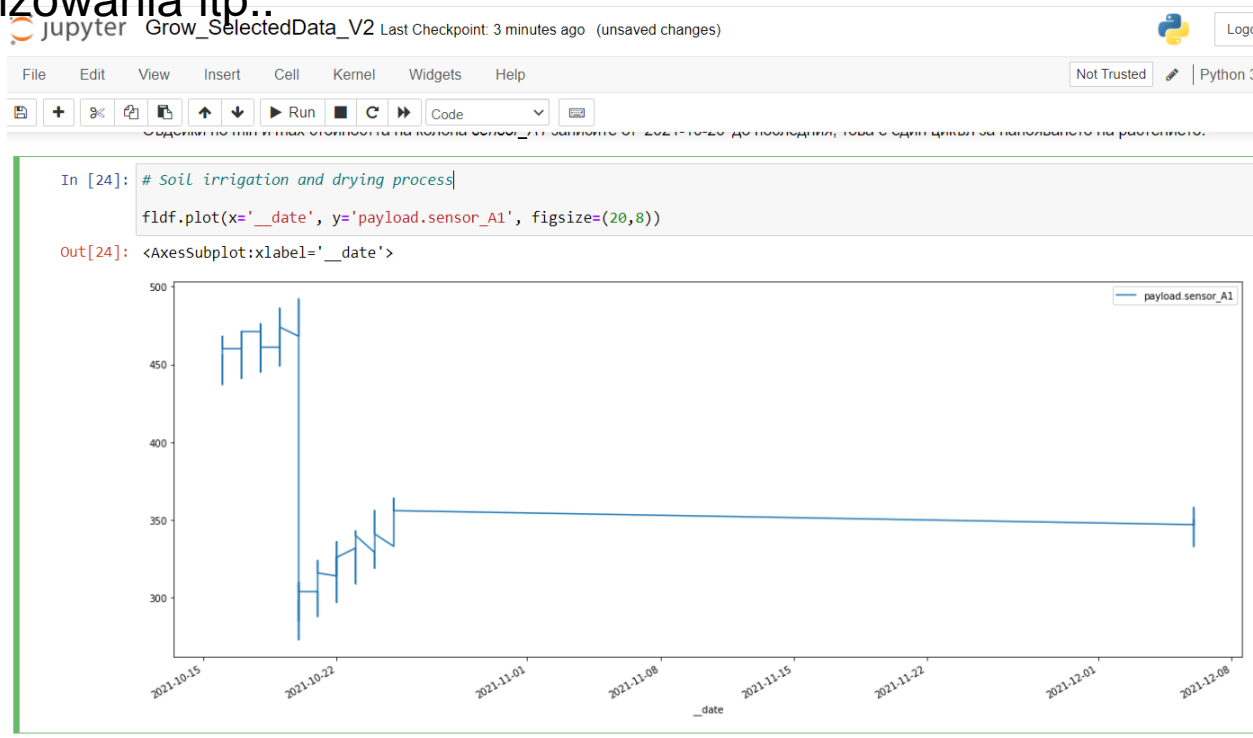
13975 Documents in the Collection Preview of Documents in the Collection

```
{ "_id": ObjectId("6169ec0c2239c2aed7b38645"), "d": "2021-10-16T00:01:00.000+00:00" }
```

```
{ "_id": ObjectId("6169ec0c2239c2aed7b38645"), "d": "2021-10-16T00:01:00.000+00:00" }
```

MongoDB - wykorzystanie BIG danych

Wykorzystanie danych BIG w zewnętrznej aplikacji do pokazywania, analizowania itp..



Python Jupyter Notebook



Referencje

- Amani, Nidamanuri; Rajesh, Yelchuri (2021): Comparative Study of Open-Source NOSQL Document-Based Databases. In Tomonobu Senjyu, Parikshit N. Mahalle, Thinagaran Perumal, Amit Joshi (Eds.): Information and Communication Technology for Intelligent Systems. Singapore, 2021. Singapore: Springer Singapore, pp. 297–303.

