

NYT Articles and Comments Analysis with MongoDB and Docker

Apr 17, 2022

Analyzing New York Times Articles and Comments (2020) with MongoDB on a Docker container.

Instructions

You can find a more detailed description of this repository in this [Blog Post](#), which also includes a [Video Tutorial on YouTube](#).

Below you'll find the commands used in this example step by step, as well as the MongoDB queries in this [file](#).

Run the Docker Container

Commands to setup the docker container and copy the data into it.

- Link to the official [MongoDB Docker Container](#)
- Link to the [data of NYT](#)

```
# Run the docker container
docker run --name some-mongo -p 27018:27017 -d mongo:latest # make su

# Check if the container is running
docker ps

# Copy data from host to docker container
docker cp nyt-articles-2020.csv some-mongo:/nyt.csv

# Open bash on docker container
docker exec -it some-mongo bash
```

Create a database and import

Commands to create a database and import the data into a collection.

```
# Run mongo shell
mongo

# Show existing databases
show dbs

# Create and use new database
use nyt

# Exit mongo shell
quit

# Import the previously copied csv into the new database
mongoimport --db nyt --collection nyt_articles --type csv --headerlin

# Import with specific datatypes
mongoimport --db nyt --collection nyt_articles_2 --type csv --columns

# Return to mongo shell
mongo

# Check the databases again to see if nyt is listed now
show dbs

# Use the nyt database
use nyt
```

Explore and analyze the data

Now that we have stored the data in a database we can execute queries on it. For this part please refer to the separate [query document](#).

Using MongoDB Atlas and Charts for visualization

In my [additional blog post](#) I explain how to use MongoDB Atlas to create a cluster, import the data there and visualize it with MongoDB Charts.

Data Source

- [New York Times Articles & Comments \(2020\)](#), Benjamin Dornel, 04/2022

License



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).

Domenic's Blog

Domenic's Blog

 [goseind](#)

On my personal blog you can find out more about me and read some of my blog posts about topics I'm interested in.