

# Manual

---

## TLDR

[Setup](#)

[Run](#)

[Usage](#)

## Detailed Manual

[initialization.sh](#)

### Docker Related

[Dockerfile](#)

[docker-compose.yml](#)

### Backend Related

[backend/backend.py](#)

[backend/generate\\_beread.py](#)

[backend/generate\\_popular\\_rank.py](#)

[backend/requirements.txt](#)

[backend/start.sh](#)

[backend/templates](#)

### MongoDB Related

[bulk\\_load\\_data.py](#)

[post\\_bulk\\_load\\_data.py](#)

[mongo\\_dump.sh](#)

[backup\\_ddbs1.sh](#)

[mongo\\_drop.sh](#)

[clear\\_ddbs1.sh](#)

[clear\\_ddbs1\\_bak.sh](#)

[mongo\\_restore.sh](#)

[restore\\_in\\_ddbs\\_1.sh](#)

[restore\\_in\\_bak\\_ddbs1.sh](#)

[fuse\\_ddbs1\\_from\\_bak.sh](#)

### FastDFS Related

[bulk\\_load\\_file.sh](#)

[update\\_file\\_path.py](#)

[configs/storage.conf](#)

[configs/storage0.conf](#)

[configs/storage1.conf](#)

### Nginx Related

[configs/nginx.conf](#)

### Others

[utils.py](#)

# TLDR

## Setup

First clone the repository

```
git clone https://github.com/Danielznn16/DDBS-Final.git
cd DDBS-Final
```

Extract the data generation to `db-generation` directory, then run `python3 genTable_mongoDB10G.py`, the resulting file should look similar to

```
./
├── Dockerfile
├── backend
│   ├── backend.py
│   ├── generate_beread.py
│   ├── generate_popular_rank.py
│   ├── requirements.txt
│   ├── start.sh
│   └── templates
├── backup_ddbs1.sh
├── bulk_load_data.py
├── bulk_load_file.sh
├── clear_ddbs1.sh
├── clear_ddbs1_bak.sh
├── configs
│   ├── nginx.conf
│   ├── storage.conf
│   ├── storage0.conf
│   └── storagel.conf
├── db-generation
│   ├── article.dat
│   ├── articles
│   ├── bbc_news_texts
│   └── genTable_mongoDB100G.py
├── ...
├── docker-compose.yml
├── fuse_ddbs1_from_bak.sh
├── initialization.sh
├── mongo_drop.sh
├── mongo_dump.sh
├── mongo_restore.sh
├── post_bulk_load_data.py
├── presentation.pptx
├── restore_in_bak_ddbs1.sh
├── restore_in_ddbs_1.sh
└── startup_log.log
```

```
|— update_file_path.py
|— utils.py
```

Make sure to setup docker and docker-compose. For details see [Docker Manual](#). Personally I recommend to use Docker-Engine on servers.

Then with python3 install, run

```
pip3 install tqdm
```

## Run

To start the system, run

```
chmod +x ./initialization.sh
initialization.sh
```

If everything is set up correctly, and assuming you have already pulled related images before(If you haven't pulling these images and building docker images should be done automatically), you should get something similar to this

```
Line 5: Tue Jan  9 16:21:18 CST 2024 - Command took 0 seconds
rm: db-generation/articles/mapping_results.txt: No such file or directory
Line 10: Tue Jan  9 16:21:19 CST 2024 - Command took 1 seconds
...
...
Loading for ddbb_mongo_1_bak
2024-01-09T08:45:15.923+0000  connected to: mongoddb://localhost/
2024-01-09T08:45:16.292+0000  30479 document(s) imported successfully. 0 document(s)
failed to import.
Loading for ddbb_mongo_2_bak
2024-01-09T08:45:16.408+0000  connected to: mongoddb://localhost/
2024-01-09T08:45:16.929+0000  30479 document(s) imported successfully. 0 document(s)
failed to import.
Line 43: Tue Jan  9 16:45:16 CST 2024 - Command took 2 seconds
```

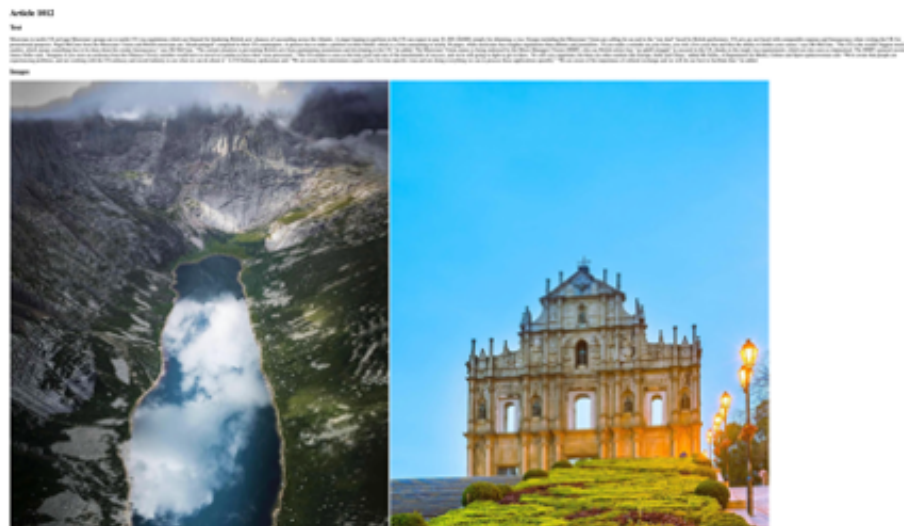
Then the system should be completely started.

## Usage

We created three major APIs

1. <http://localhost:9090/frontend/article/1012>  
Feel free to change the 1012 to other article ids

You should get something like this



## 2. [http://localhost:9090/frontend/popular\\_rank/daily/1](http://localhost:9090/frontend/popular_rank/daily/1)

This api is in the form of grainularity and popular\_rank id, make sure to check the mongodb for ids. Resulting output should look like

### Popular Articles - Daily Rank

Starting from: 2017-09-26

#### Top Articles

##### Article 1512

South Africa swept top awards South Africa's Schalk Burger was named player of the year as the Tri-Nations champions swept the top honours at the International Rugby Board's awards. The flustered capped a list which included Ireland's Gordon D'Arcy and Australian sensation Matt Giteau. John White claimed the coaching award while his side held off Great Britain winners France to take the team award. England player Stuart Annett beat teammate Ben Gollings and Argentina's Lucio Lopez (kicking) to win the scrum award. Burger's second name just a week after he won the equivalent prize from his fellow international players and White, who also coached Burger at under-21 level, paid tribute to him. "Schalk's emergence as a major force has meant a lot to South African rugby, but has also influenced world rugby," said White. "It's because in South Africa to play what Jarryd Hooker was in South Africa created. It's amazing what he has achieved in such a short time so far in his international career." Annett, who won a captain's armband in his season's opening IRB Sevens tournament, the Dubai Sevens, which started on Thursday, was delighted with his award. "There are so many great players on the circuit at the moment that it is a great honour," said the Gloucester fly-half.

##### Article 1895

Queen secret singer for new tour The remaining members of rock band Queen are to go on tour next year with former Free and Bad Company singer Paul Rodgers taking Freddie Mercury's place. Graham Nash has said he expects to be on the road with Rodgers and drummer Roger Taylor from April. May said: "Suddenly the Queen (Queen) is doing again from the ashes and will rise phoenix even... we know." Queen have played with many different singers since Mercury's death in 1991 but have reportedly not issued May performed with Rodgers as a concert in celebration of the 30th anniversary of the band's last live performance in London in September. "We were both in a moment of the country that was going on in (Free) (All Right Now), that suddenly it seems suddenly obvious that there was something happening here." May wrote on his website. They teamed up again for a concert in March 2006 in London at the UK Music Hall of Fame, and were joined by Taylor. "The show went so incredibly well from our point of view, and we got so many rave reviews from our fans, we decided almost then and there that we would look at it as a new chapter." May wrote. Queen went to another one in 2006 with a version of the 1978 rock 'n' roll song by the band Free, and they have also played with Robbie Williams, WU. Shang and Bob Geldof. Queen's leader John Deacon has currently retired from the stage. Rodgers was singer with early 1970s medium Free, who had a global hit with All Right Now before joining Bad Company, a successful "supergroup" with members of King Crimson and Mott the Hoople. He has also been in The Firm with Led Zeppelin's Jimmy Page and The Law with The Small Faces and The Who drummer Kenny Jones.

##### Article 6461

US peer-to-peer pirates convicted The first convictions for piracy ever peer-to-peer networks have been handed down in the US. New York's William Thurnidge and Timon Michael Chiorine have pleaded guilty to charges that they infringed copyright by illegally sharing music, movies and software. The two men faced charges following raids in August on suspected pirates by the FBI. The pair face jail terms of up to five years and a \$250,000 (A\$390,000) fine. In a statement the US Department of Justice said the two men operated the central hub in a piracy ecosystem organised across the Direct Connect peer-to-peer network. The piracy group called itself the Underground Network and membership of it demanded that users share between one and 100 gigabytes of files. Direct Connect allows users to set themselves up as central servers that act as coordinating spots for users. Users would swap files, such as films and music, by exchanging data over the network. During its investigation FBI agents reportedly downloaded 54 movies, 40 software programs, 13 games and 178 "sound recordings" from the file hosts that made up the larger piracy groups. The raids were reported under the umbrella of Operation Digital Desecration, which was aimed at fighting "criminal copyright theft on peer-to-peer networks". In total, six raids were carried out in August. Prior to the raids, the names of suspected copyright thieves and users in a peer-to-peer network. The Department of Justice said that but its own pleaded guilty to its own criminal copyright infringement. They also pleaded guilty to aiding for criminal copyright infringement. The two men are due to be sentenced on 26 April.

##### Article 1522

Rodriguez poised for Helen Southampton are set to unveil Harry Redknapp as their new manager at a news conference at 1500 GMT on Wednesday. The former Portsmouth boss replaces Steve Wigley, who has been relieved of first team duties after just one win in 14 league games in charge. Redknapp, 57, left his Premier League position on 24 November and vowed: "I will not go down the road of no chance." Pompey coach Kevin Bond is poised to join Redknapp, who will be second third boss of the season. Redknapp's first game in charge will be at home to Millwall although ex-Sunderland Portsmouth chairman Miles Mander said he was "disappointed" by the news and claimed Redknapp had been in talks with Southampton for "some time". "It was all agreed that negotiations over this have been going on for some time," Mander said. Ex-Portsmouth's official website. "I am surprised and a little shocked that the chairman of Southampton has not ended his silence and kept us informed." According to Mander, Redknapp would be the second and last South coast coach who he left Portsmouth. "I used to have 15000 fans and

### 3. <http://localhost:9090/frontend/user/1012>

Feel free to change user id from `1012` into other ids



## Detailed Manual

In this section we give a detailed manual of which file does what

## initialization.sh

Overall startup script, contains all steps of startup in one script. it also logs the time spend for each operation

```
#!/bin/bash

# Bring down any currently running containers
SECONDS=0
echo "Line $LINENO: $(date) - Command took $SECONDS seconds"; SECONDS=0

docker-compose down

rm db-generation/articles/mapping_results.txt

echo "Line $LINENO: $(date) - Command took $SECONDS seconds"; SECONDS=0

# Create directories
mkdir -p ./ddbbs_1_data
mkdir -p ./ddbbs_2_data

# Start the Docker Compose services in detached mode
docker-compose up -d
echo "Line $LINENO: $(date) - Command took $SECONDS seconds"; SECONDS=0

# Run your Python script
python3 bulk_load_data.py
echo "Line $LINENO: $(date) - Command took $SECONDS seconds"; SECONDS=0
```

```

sleep 5;
echo "Line $LINENO: $(date) - Command took $SECONDS seconds"; SECONDS=0
python3 post_bulk_load_data.py
echo "Line $LINENO: $(date) - Command took $SECONDS seconds"; SECONDS=0
docker exec -it python-app bash -c "cd /usr/src/app/ && python3 ./generate_beread.py"
echo "Line $LINENO: $(date) - Command took $SECONDS seconds"; SECONDS=0
docker exec -it python-app bash -c "cd /usr/src/app/ && python3
./generate_popular_rank.py"
echo "Line $LINENO: $(date) - Command took $SECONDS seconds"; SECONDS=0

docker cp bulk_load_file.sh storage0:/etc/fdfs_buffer/
echo "Line $LINENO: $(date) - Command took $SECONDS seconds"; SECONDS=0

echo "Uploading Files"
docker exec -it storage0 bash -c "cd /etc/fdfs_buffer/ && bash ./bulk_load_file.sh"
echo "Line $LINENO: $(date) - Command took $SECONDS seconds"; SECONDS=0

mv db-generation/articles/mapping_results.txt backend/mapping_results.txt
python3 ./update_file_path.py
echo "Line $LINENO: $(date) - Command took $SECONDS seconds"; SECONDS=0

```

## Docker Related

### Dockerfile

A build file describing how to build the docker image for the python container

### docker-compose.yml

Stores the configuration of all containers started by docker-compose.

```

version: "3"

networks:
  ddbb_network:
    driver: bridge

services:
  tracker:
    image: delron/fastdfs
    container_name: tracker
    networks:
      - ddbb_network
    ports:
      - "22122:22122"
    command: "tracker"

storage0:

```

```
image: delron/fastdfs
container_name: storage0
environment:
  - TRACKER_SERVER=tracker:22122
volumes:
  - ${PWD}/db-generation/articles:/etc/fdfs_buffer/
  # - ${PWD}/dfs_1_data:/etc/fdfs_buffer/
  - ${PWD}/configs/storage0.conf:/etc/fdfs/storage.conf
  - ${PWD}/configs/storage.conf:/usr/local/nginx/conf/nginx.conf
depends_on:
  - tracker
...
```

## Backend Related

### backend/backend.py

backend implementation for backend service.

### backend/generate\_beread.py

Used to generate the beread table, implemented with raw mongo requests over pymongo.

### backend/generate\_popular\_rank.py

Used to generate the popular rank table, implemented with mongo's aggregate api.

### backend/requirements.txt

The requirement packages needed for the python container, will be installed during build

### backend/start.sh

Used to start the backend service with delay so docker bridge network's routing can be added to the service

```
#!/bin/bash
sleep 10 # Waits 10 seconds
python backend.py
```

### backend/templates

Stores the html templates used to render the response webpage for backend service

## MongoDB Related

## **bulk\_load\_data.py**

Used to process the input files into files to be imported with `post_bulk_load_data.py`

## **post\_bulk\_load\_data.py**

Used to upload the processed data files into mongo.

## **mongo\_dump.sh**

Used to dump all data for a mongo deployment.

## **backup\_ddbs1.sh**

Used to dump the content in ddbs1's first replica.

## **mongo\_drop.sh**

Used to drop all collections for a mongo deployment.

## **clear\_ddbs1.sh**

Simulated dropping mongo deployment by clearing all data stored in ddbs1's first replica.

## **clear\_ddbs1\_bak.sh**

Simulated dropping mongo deployment by clearing all data stored in ddbs1's second replica.

## **mongo\_restore.sh**

Used to restore all data for a mongo deployment from dumped data.

## **restore\_in\_ddbs\_1.sh**

Used to restore the first replica of ddbs1 from files.

## **restore\_in\_bak\_ddbs1.sh**

Used to restore the second replica of ddbs1 from files.

## **fuse\_ddbs1\_from\_bak.sh**

Transfer the data stored in ddbs1's second replica to its first replica.

## **FastDFS Related**

### **bulk\_load\_file.sh**

Used to load files into FastDFS.



## **update\_file\_path.py**

Used to upload the path mapping yielded by `bulk_load_file.sh` into mongo deployments.

## **configs/storage.conf**

Used to overwrite nginx configuration within storage nodes.

## **configs/storage0.conf**

Used to overwrite the Storage configuration in the first storage node.

## **configs/storage1.conf**

Used to overwrite the Storage configuration in the second storage node.

## **Nginx Related**

### **configs/nginx.conf**

Used to overwrite nginx configuration in Nginx container.

## **Others**

### **utils.py**

Stores frequently used utils maintained by Daniel and only necessary utils for this project are kept.