

Data Foundations System

# MinIO Data Search (Backend)

## Progress Evaluation Report

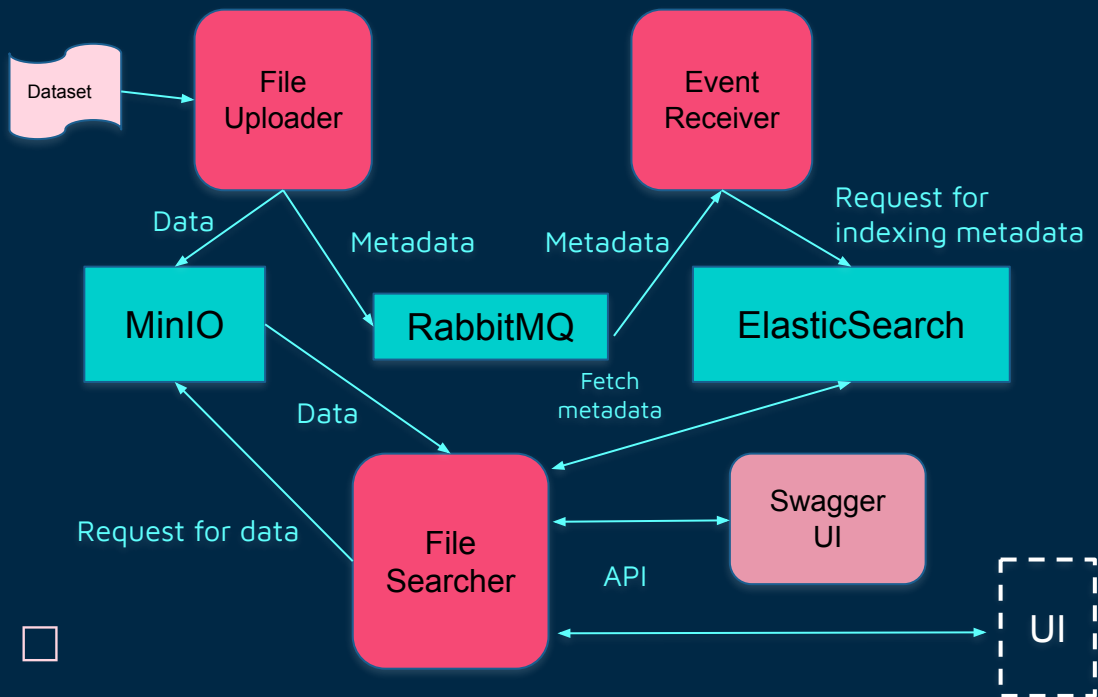
### Team Celeron:

Hari Krishnan UM - 2021202022

Manu Gupta - 2021202025

Shaon Dasgupta - 2021201068

# Work Done Till Now



- ✓ MinIO and Elasticsearch setup.
- ✓ RabbitMQ setup.
- ✓ File Uploader as a CLI tool.
- ✓ Event Receiver.
- ✓ [File Searcher] API to search over dataset names and general metadata.
- ✓ [File Searcher] API to search inside datasets.
- ✓ [File Searcher] Pagination and generic filters.
- ✓ Swagger UI.

# File Uploader

- Consists of a .env file with all the necessary details regarding minio, rabbit mq, and ElasticSearch, etc.
- To upload a dataset, copy the dataset folder to the FileUploader directory and run the following command:
- `python3 file_uploader.py --dir_path <path_to_data_set_folder> --bucket <name_of_the_bucket_on_minio_where_you_want_to_upload>`

```
MINIO_ENDPOINT= ...  
MINIO_ACCESS_KEY= ...  
MINIO_SECRET_KEY= ...  
RM_QUEUE= <queue name>  
RM_HOST= ...  
RM_PORT= ...  
RM_USERNAME= ...  
RM_PASSWORD= ...
```

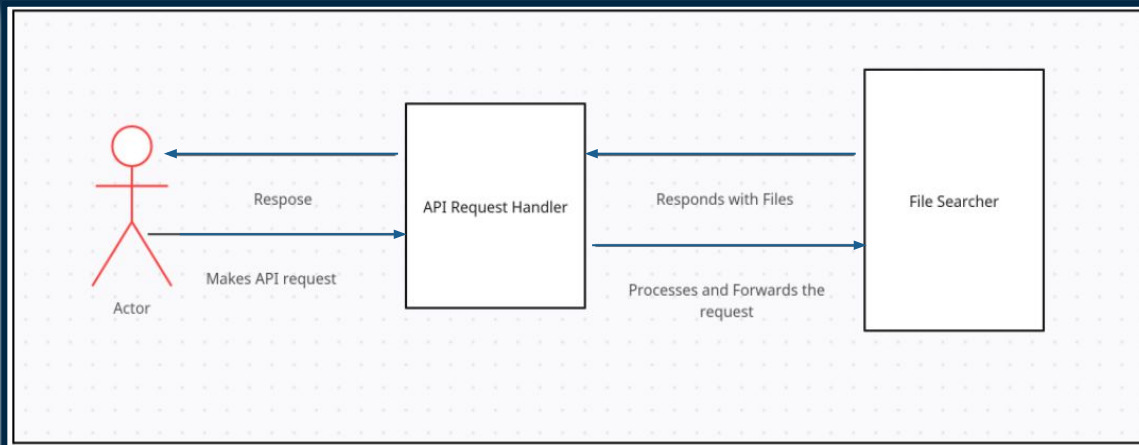
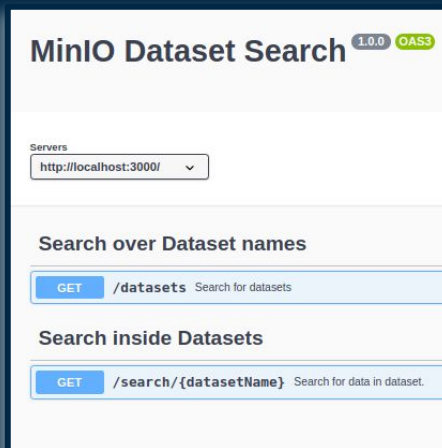
# Event Receiver

- Similar to File Uploader, update the .env file with all the necessary details regarding RabbitMQ, and ElasticSearch, etc.
- Event Receiver will always stay alive and keep listening, RabbitMQ triggers it whenever it has something new and unread.
- Event Receiver then uploads the metadata to ElasticSearch which can then later be used for fast querying on this data.

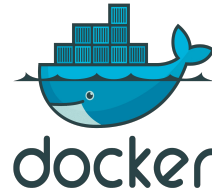
```
RM_QUEUE= <queue name>
RM_HOST=...
RM_PORT=...
RM_USERNAME=...
RM_PASSWORD=...
ELASTICSEARCH_HOST= <host-name>:<elastic-search-port (9200 by default)>
```

# File Searcher

- When an API is hit, the file searcher builds a search query and send it to elasticsearch, the results received from the elasticsearch are displayed as results.



# Technologies Used



# Work Left To Be Done

- Collecting dataset and metadata information from the stakeholders and adjusting our architecture to fit their requirements.
- Implement the Download API for the data.
- Deployment
- Some more additional features if time permits (like making the cli interactive, autocomplete on the search bar, showing top searches, etc.).



# Contribution

Hari Krishnan UM - File Uploader

Manu Gupta - Event Receiver

Shaon Dasgupta - File Searcher

