**Milestone Report – 2**

**Capstone Project: News Article Classification**

* **Dayakar & Krupasindhu (16th Oct 2021)**

1. **Project Architecture Design Completed** (Attached .PPT document for reference)
2. **ML Approach Documentation completed** (Attached .PPT document for reference)
3. **Developed Python script for Data preparation**
   1. Install the below packages

Install json and requests, mongodb and mongosrv libraries in case you want to scrap the data using beautiful soup install bs4

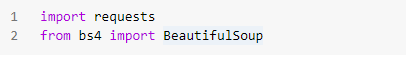
!pip3 install requests -- required for data extraction

!pip3 install bs4 – web scrapping package

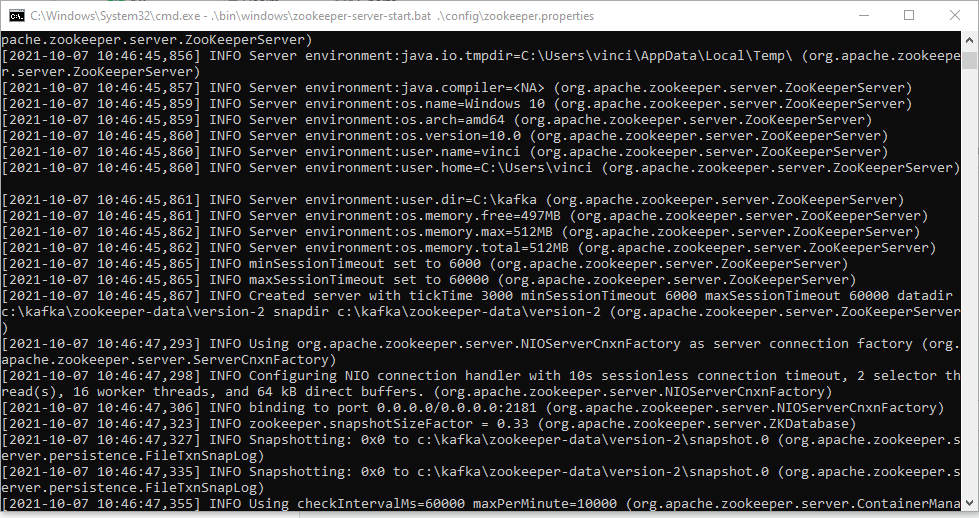
!pip3 install json - used for json conversion

!pip3 install pymongo pymongo[srv] convert the json string to json object - this important step

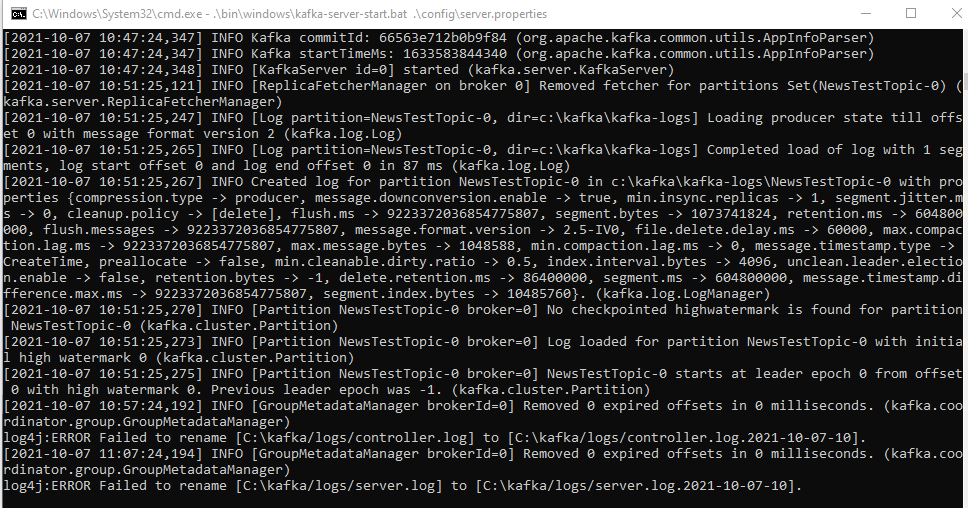
* 1. Import these libraries



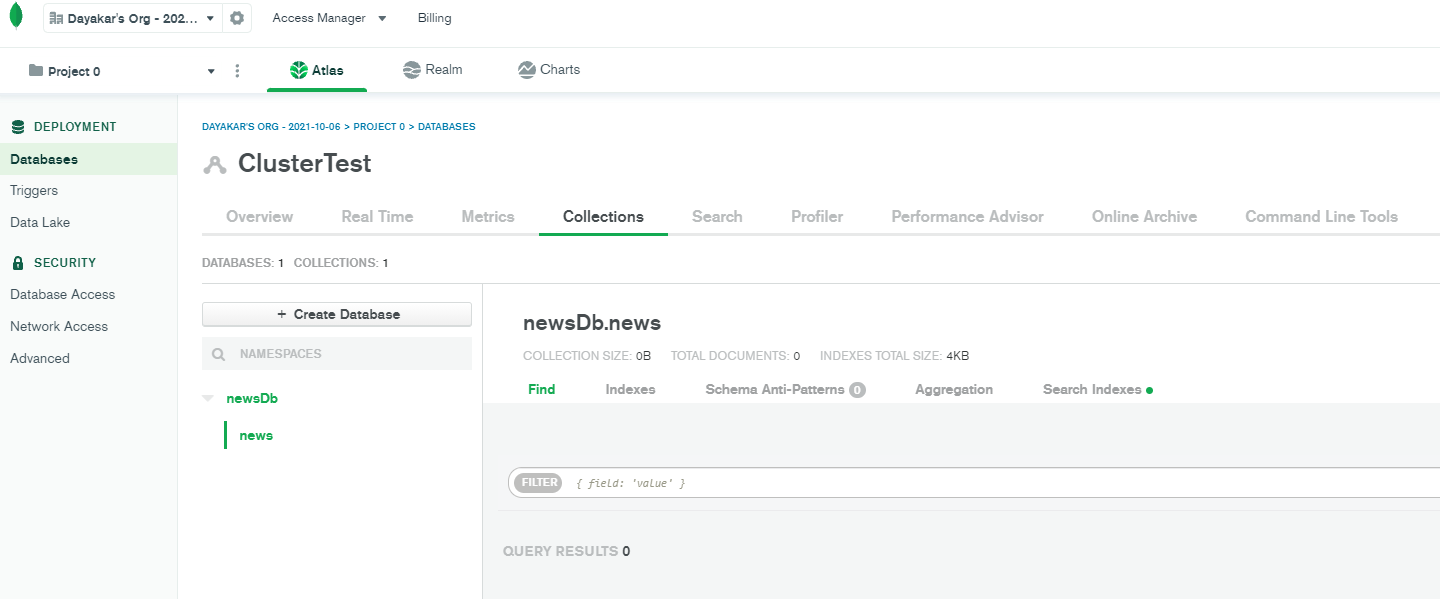
1. **Kafka Set up Completed**
   1. Running the zookeeper



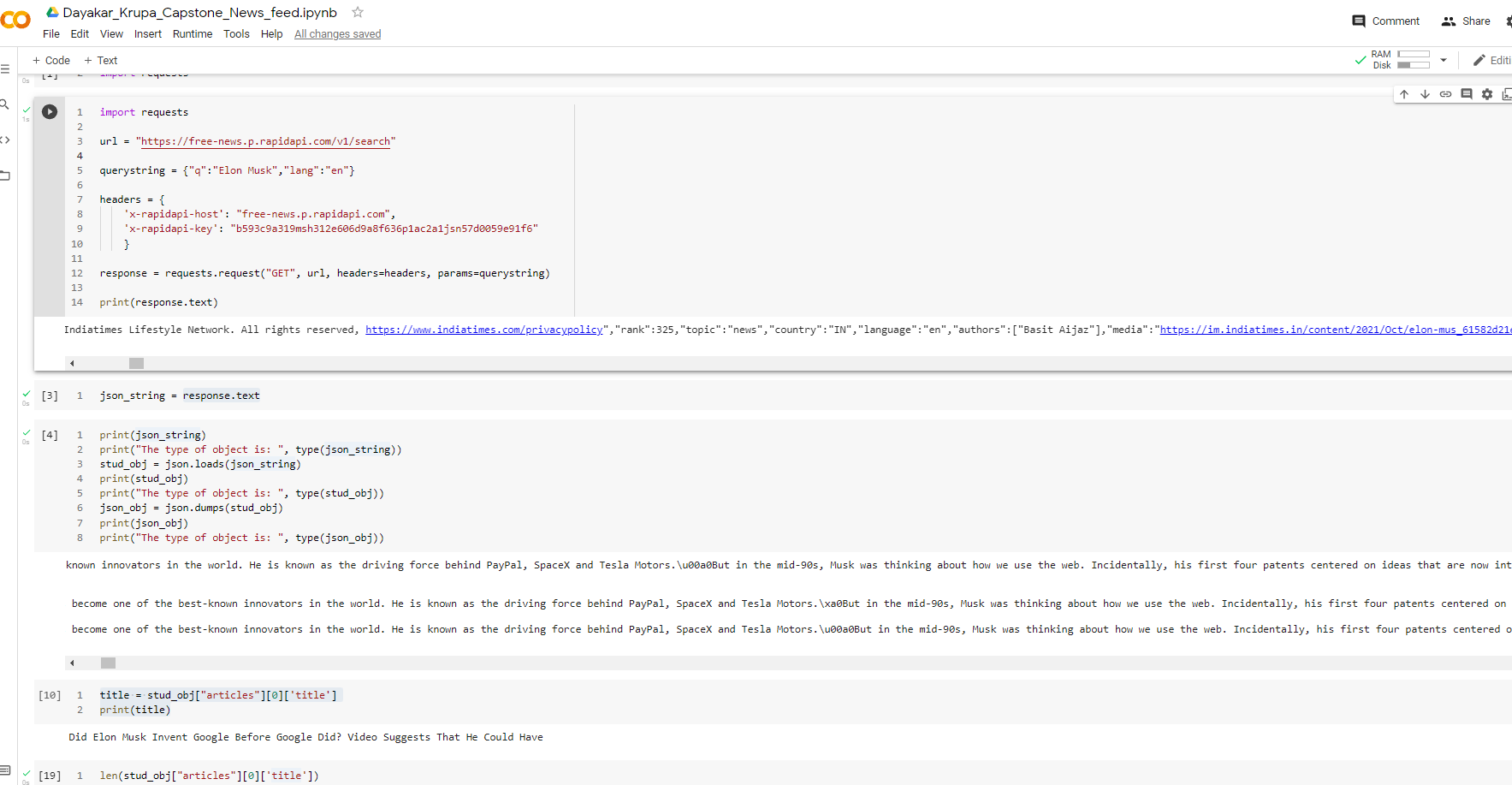
* 1. Running the Kafka-server



1. **MongoDB Set up Completed**



1. **Initial Data Ingestion Python code development is completed**



1. **GitHub Link / URL:**

<https://github.com/bksindhu/Capstone---News-articles-classifier>

<https://github.com/DayakarKodirekka/Capstone---News-articles-classifier>

1. **Challenges:**

Connecting to mongodb

Below is the error we are facing while connecting to mongodb