# Trabalho de final de módulo – Mineração de Texto com ELK - INFNET

# Professor: Felipe Fink Grael

# Aluno: Marcelo Lopes da Silva

# Dezembro 2020

## Instalação dos pré requisitos

Para a presente solução serão utilizados o Elasticsearch e o Kibana executados via docker. O acesso ao elasticsearch será feito utilizando jupyter notebook. Outras dependências podem ser instaladas utilizando o gerenciador de pacotes pipenv. Os exercícios foram realizados em máquina com sistema operacional Ubuntu versão 20

1 – Instalar o docker/docker compose

<https://www.digitalocean.com/community/tutorials/how-to-install-and-use-docker-on-ubuntu-20-04-pt>

2 – Instalar o pipenv

Em um terminal digite o comando pip install pipenv

3 – Instalar as dependências utilizadas no exercício

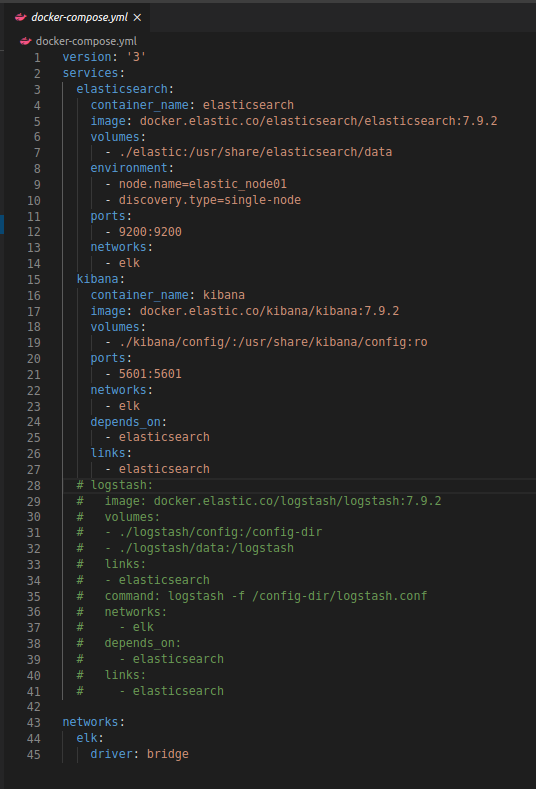
No diretório onde serão executados os notebook digite o comando pipenv install. Note que os arquivo Pipfile e Pipfile.loc tem que estar no mesmo diretório.

A execução dos notebooks tem que ser feita no ambiente virtual criado pelo pipenv em que as dependências foram instaladas. Para isso no terminal digite o comando pipenv shell.

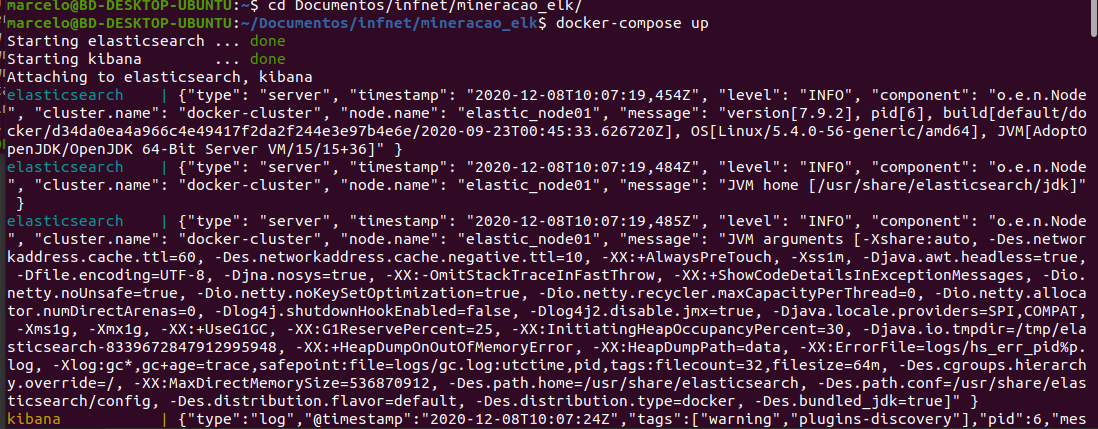
## Execução do Elastic e Kibana

Como informado anteriormente a execução é feita utilizando docker. Um arquivo docker compose é fornecido junto a solução.

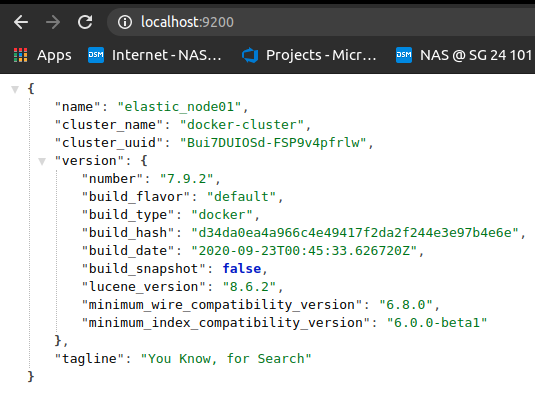
docker-compose.yml



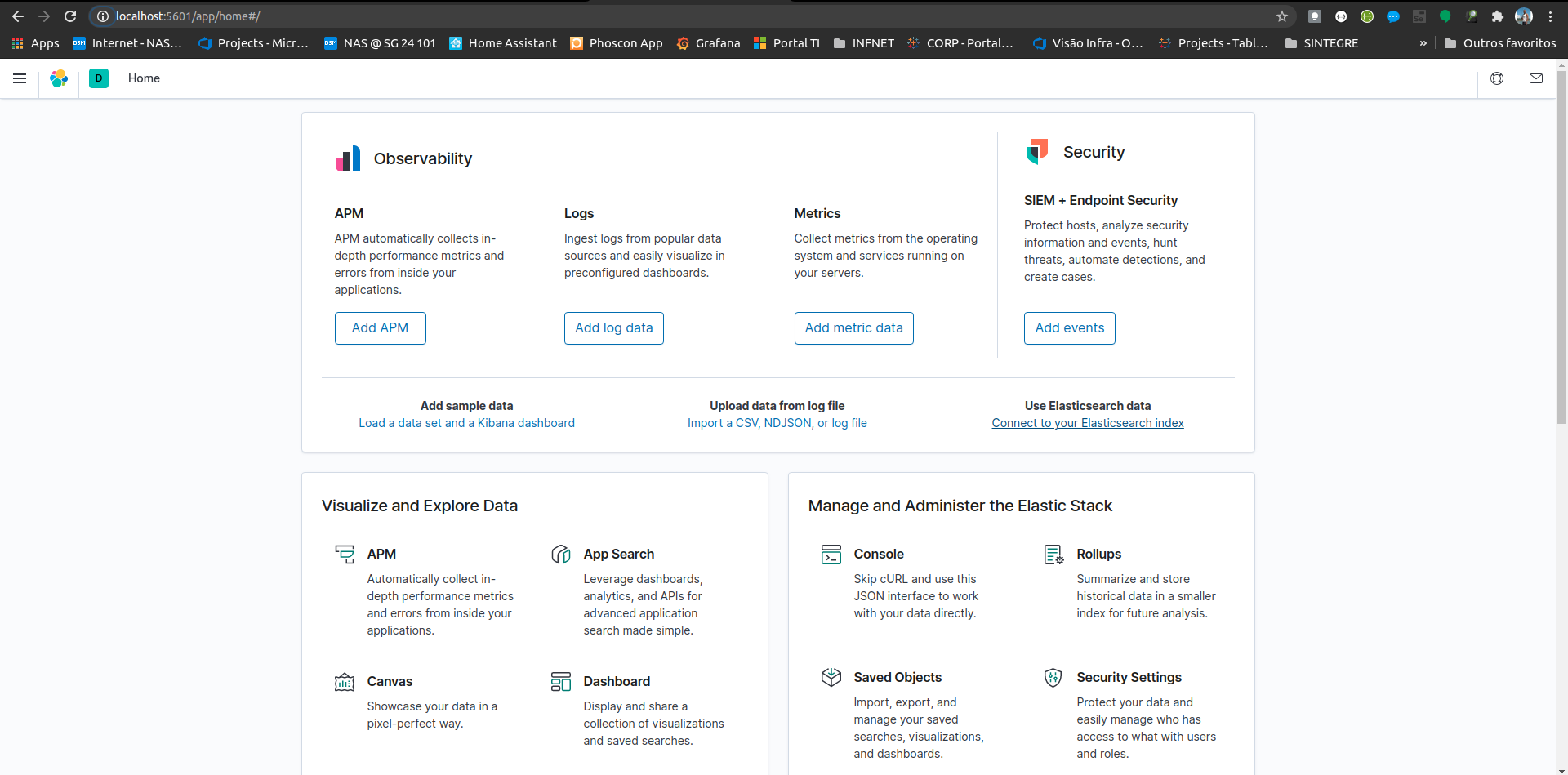
execute o comando docker-compose up para iniciar os serviços



Após algum tempo verifique se o elasticsearch está respondendo, abrindo um browser e apontando para localhost:9200



Depois verifique se o kibana está respondendo, apontando o browser para localhost:5601



## Notebooks

Os notebooks de carga de informações e consultas podem ser observados nos arquivos: 1\_carga\_informacoes\_dou\_es.ipynb e 2\_consultas\_informacoes\_dou\_es.ipynb

## Dashboard

Dashboard criado no kibana com visualizações dos top 10 registradores de informações, distribuição de documentos por tipo, frequência de termos e nuvem de palavras com os 70 termos mais relevantes.

