

Open in app ↗



Search

Write



Nifi and Docker to index CSV data straight from the web into ElasticSearch/Kibana without installing, downloading, or converting anything upfront...?



Thomas Jaensch

2 min read · Sep 12, 2019



6



Apache Nifi has been around for a few years now, but anyone trying to implement data pipelines using Nifi will sooner or later run into questions that the official documentation alone will not answer, and resources on the Internets, especially on stackoverflow, are still rather sparse.

To add another layer of complexity, how to run Nifi in Docker in combination with the ELK stack (ElasticSearch-Logstash-Kibana) without having to mess with a bunch of loner containers and endless complicated flags and arguments? Docker Compose to the rescue, but how to configure it?

Also, ever had to process a bunch of data from the web but didn't want to download some chunky large file and take it from there? Wouldn't it be nice to be able to process stuff in memory but not have to write some complicated boiler plate thing in an IDE first?

Well, a few weeks into actively using Nifi I sort of know my way around all of the above now, but man was it frustrating at times... So I thought documenting this kind of workflow in a concise little example project could probably help someone else, I myself was happy for sure for the things I found that guided me along the way.

So, what's the goal now? Basically getting a rather large CSV file >4MB from here:

VSRR Provisional Drug Overdose Death Counts

This data contains provisional counts for drug overdose deaths based on a current flow of mortality data in the...

catalog.data.gov

... straight into ElasticSearch/Kibana via Apache Nifi with everything needed running in Docker containers that can talk to each other.

If this sounds like something you may want to be trying out, here is how I did it:

tjaensch/nifi_docker_elasticsearch_demo

This Nifi workflow will pull down the VSRR Provisional Drug Overdose Death Counts CSV file with over 23,000 records...

[github.com](#)

Happy coding!

Elasticsearch

Docker

Data Science

Software Development

Apache Nifi



Written by Thomas Jaensch

[Edit profile](#)

21 Followers

WFH, staring at code