Twitter Thread by Philip Vollet

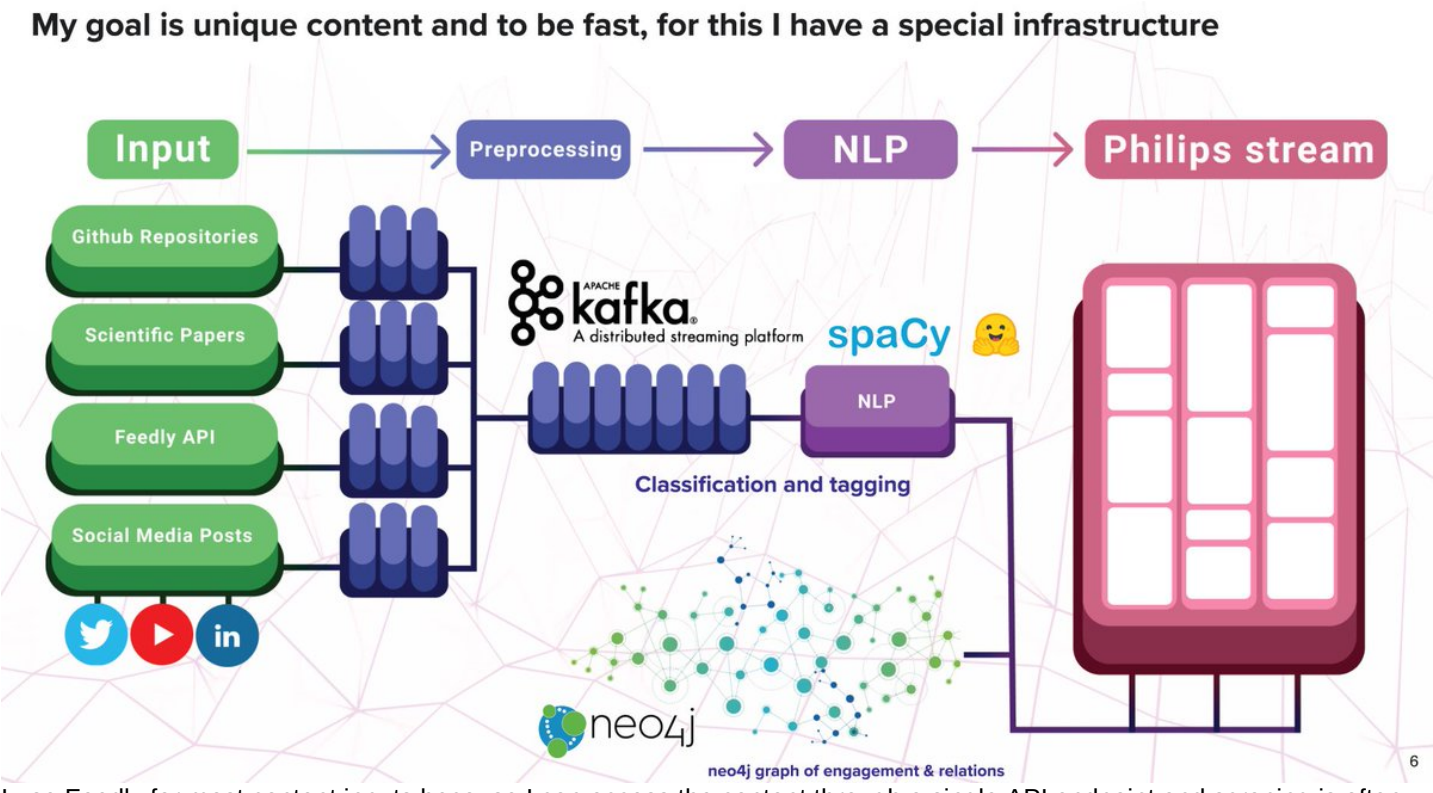
Philip Vollet
@philipvollet

Insights from an open source influencer

I'm often asked how I get my content, over the years I've built an unusual technology stack for it

https://t.co/FdIAeZc8GF

Some insights:



I use Feedly for most content inputs because I can access the content through a single API endpoint and scraping is often pure pain.

@feedly

Feedly saves me a lot of time and manual work.

https://t.co/CeHju43rwJ

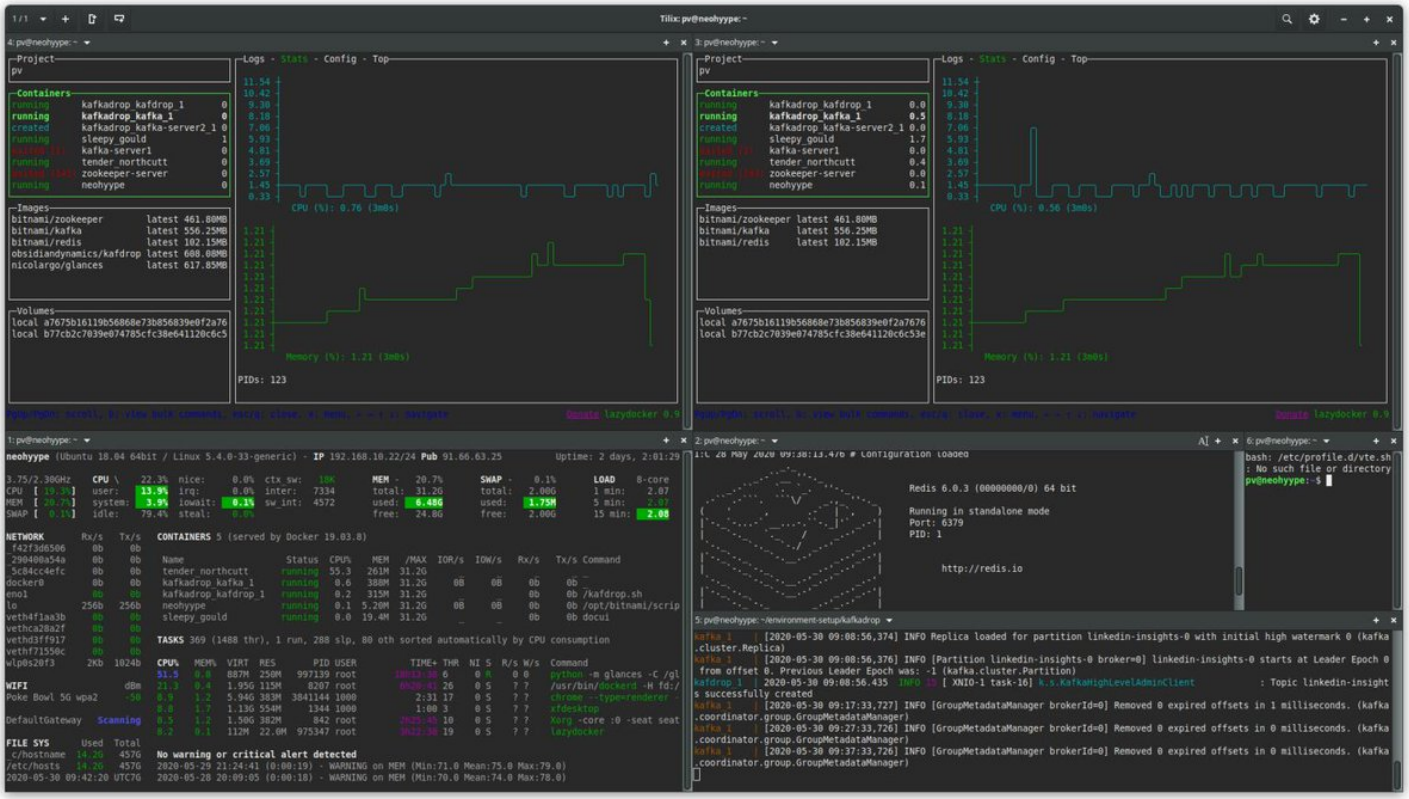
To pull and enrich my GitHub content I use ghapi from @fastdotai which provides a 100% always-updated coverage of the entire GitHub API

https://t.co/c0HwjoShvg

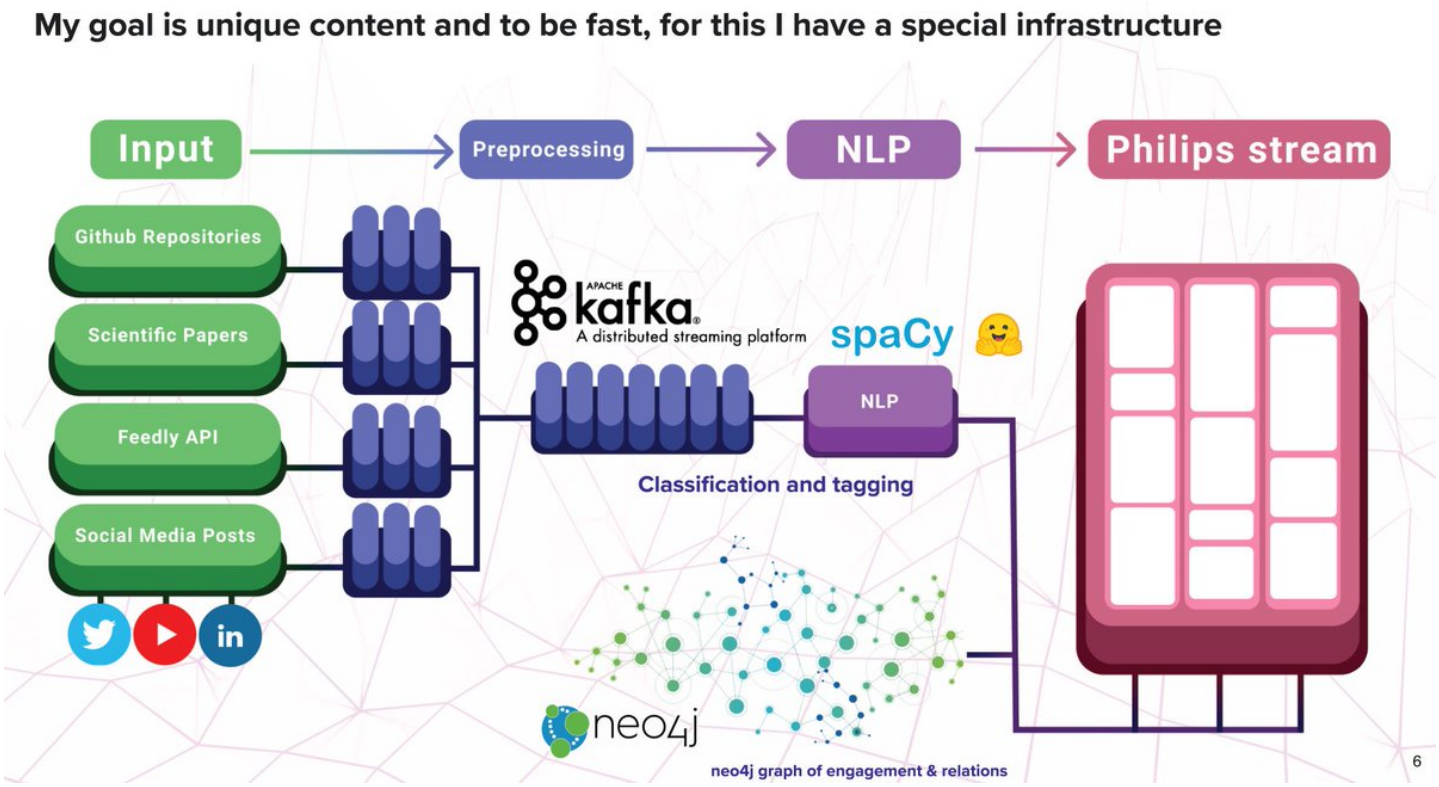
The actual magic happens on my server infrastructure.

This is where the various data streams converge and are enriched.

The whole thing is based on Kafka streams so that if a pipeline stops it can be started up again without any problems.



Then a wild mix of machine learning NLP pipelines between @huggingface and @spacy_io is used to classify, score and tag the content.

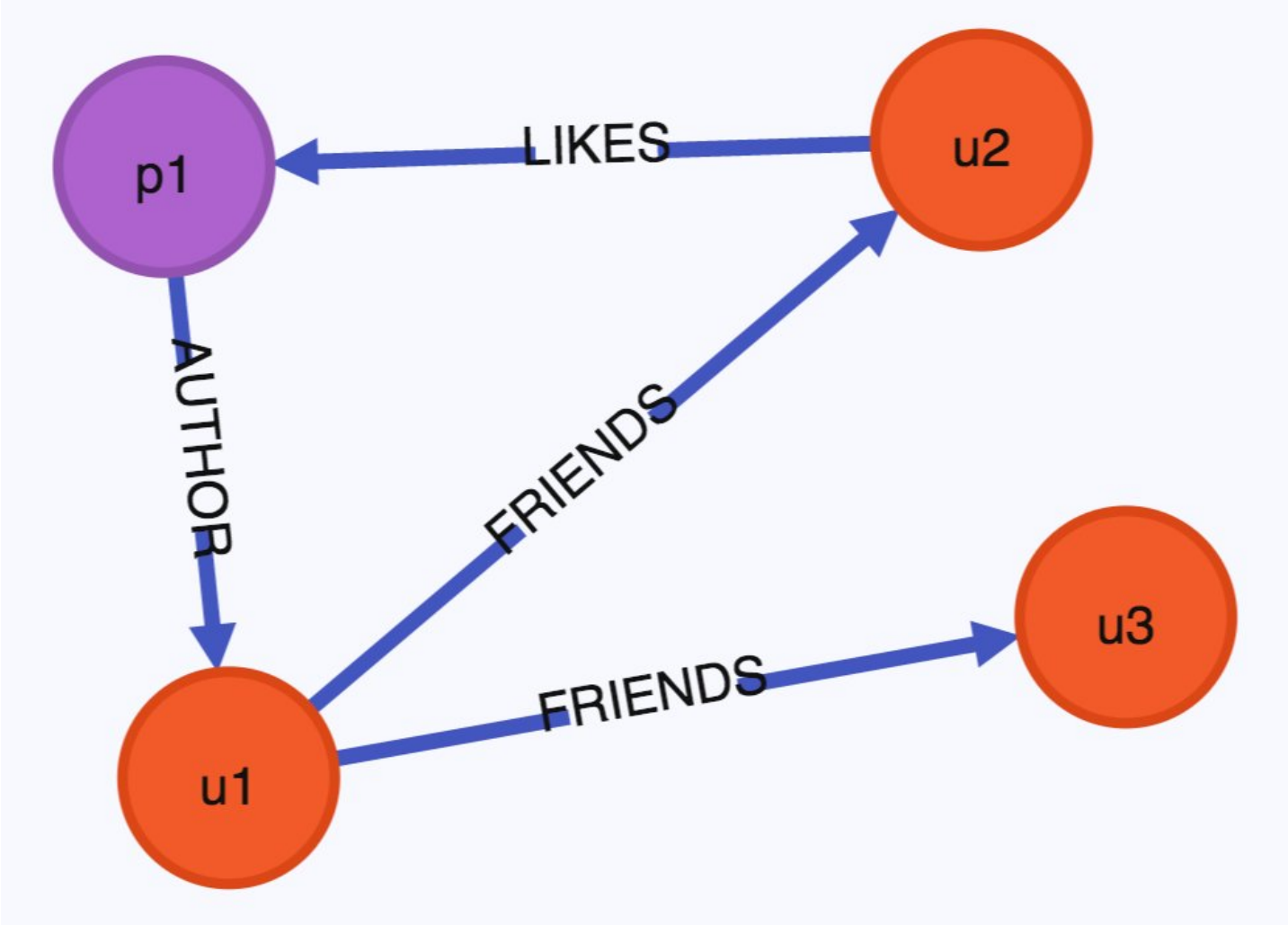


For analysis I have a @neo4j instance to analyze my network relations and the engagement.

This is also used to find potential new sources.

Why?

Because it's about engagement, influence and adding value



Spread the open source love!

If you know an amazing project drop me a message
@philipvollet

https://t.co/FdIAeZc8GF