## Githublink : <https://github.com/JangasMonsterPlace>

<http://34.125.55.184:5000/>

## 

## Delivery

A Framework for creating an event driven architecture with a special focus on NLP for getting insights from Review data.

1. Creating a Source connector Framework and sending data via etl into company internal architecture. Supported sources:
   1. Twitter (get tweets by hashtag)
   2. CSV (load CSV files from GCP Storage)
2. Getting Insights from internal (CSV) and external (Twitter) data by using N-Grams for finding patterns with length 2 and 3 and by using LDA Topic modelling (hard coded count: 4, shame on us!). It’s possible to create this analysis by time intervals (USP)
3. Coordinating Microservices and Sync data in a scalable Event Driven Microservice architecture. Scalable means:
   1. each component is extendable without affecting other components
   2. each compontent can be deleted / changed without affecting other components
   3. it’s easy to add new components without affecting other components
   4. kafka & Kafka connect and each component can be parallelized for scaling with higher throughput (well …. potentioally. we would need to use Consumer Groups for making this possible. A single line code to change)

## Challenges we wanted to tackle

* Get insights from text data (it could have been solved much better!)
* creating a framework for working at best practices to improve working in complex architectures
* most innovative solution I - not by our massive techstack! but by the idea to allow to create an NLP based on specific data filters (e.g. just take data from 12.01.2021 to 25.01.2021 for our internal CSV Data and do analysis on prem for this filter)
* most innovative solution II - cooperate between AI and users: Specially unsupervised learning is always doing just something without explaining the why. We wanted to provide an interface which allows to understand and writing down the meaning of the created clusters

## Team

Talha Abdur Rahman, India, GMT+3:50, Ops and Dev (Python)

Fabio Batti, Germany, MEZ, Dev (Python) and Frontend

Julia, Germany, MEZ, Frontend

Janis, Wuppertal, MEZ, Architect and Lead

## Techstack

* Python Flask & Bootstrap for Web Application
* Python NLTK & Gensim for Bigrams and LDA Topic Modelling
* Python Tweepy for fetching data from Twitter
* Kafka Connect (JDBC Source & Sink, Elasticsearch Sink) for Synchronising data between databases with **no Code**!
* Kafka as an Event handler
* Elasticsearch & Kibana for storing and querying for text patterns
* Logstash for sending Server Logs to Elasticsearch
* GitHub Actions for creating CD Pipelines and allow fast delivery (but no tests - lol)
* GCP as a cloud Provider including Postgres Database, Virtual Machine for our Services, hosting Confluent Cloud (Kafka) and Elasticsearch

## Log

### Friday, 10th Dec, 2021

* Kickoff Meeting and initial architecture planning
* Talha wrote a message and wanted to join the team
* Ticket Planning with Trello (around 20 Tickets)
* Project Setup at Git

### Saturday, 11th Dec, 2021

* Fabio joined the party in the morning
* Tooling Setup: Postgres, Elasticsearch, Kafka, Ubuntu Servers, GCP Storage
* Create Twitter Consumer for fetching additional language data
* Load CSV from Flask Webapp to GCP Storage, monitor the GCP Storage and load files into system
* NLP planning & Writing to database
* Julia joined the party in the evening

### Sunday, 12th Dec, 2021

* Julia left the party
* Frontend day and creating APIs
* improve logging and use logstash to stream logs into elasticsearch & display them via Kibana
* Setup Kafka & Kafka Connect for connecting Microservices (lol - we would have been dead if that was not running …)
* Implement Filter Options into NLP (another critical thing, it’s our USP - we want to aggregate Feedback in timeframes! This was just created by the very last breath)
* Create Presentations