

Data integration via Apache kafka, mongoDB and Grafana for online analysis



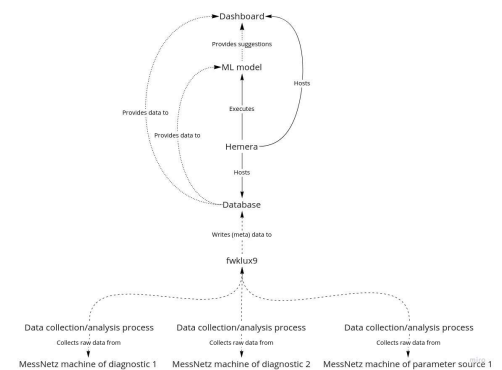
Friedrich Bethke, Nico Hoffman, Arie Irman, Susanne Schoebel, Patrick Ufer, Michael Bussmann

Prototype for electron experimental data

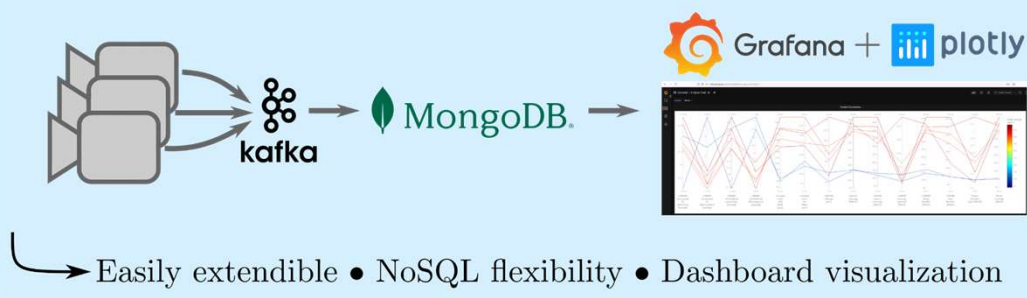
- **Problem statement:** Data is collected into a variety of machines and python Jupyter notebooks that are used for on- and offline analysis
- **Goal:** We seek to create a more autonomous data pipeline with simpler access and higher availability to a flexible database, complete with a data interaction interface, primarily through utilization of existing resources
- **How to Get There:**
 - Kafka: message passing tool
 - MongoDB: flexible database
 - Grafana + Plotly: real-time flexible visualization



Existing architecture



ELBE / HZDR

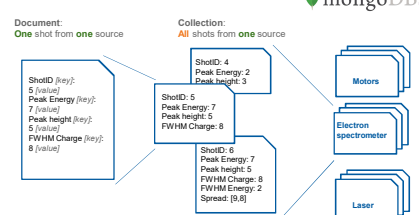


Apache kafka

- Send messages between different PC's
- One PC hosts the kafka message broker server which accepts messages
- Instances that send messages are called kafka producers
- Instances interested in the messages of the server are called kafka consumers



mongoDB

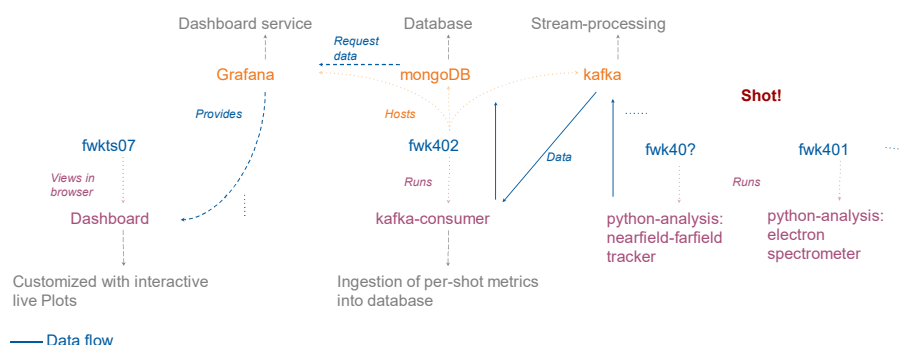


Grafana

- Used to create **dashboards** for data visualization and system monitoring
- Active plugin community
- Multiple different data sources possible
- Existing **plotly** plugin for interactive python-like plots (written in JavaScript)
- Can be hosted locally
- Dashboards can be accessed from multiple clients at once
- Provides large toolbox for configuration



Current prototype



Next Steps

- Add more diagnostics
- Add input parameters (gas pressure, ...)
- Connection to hemera: easier access to data

Feature requests:

- Add image visualization (data already collected)
- Set analysis tool
- Manual labeling input (new service needed)

