

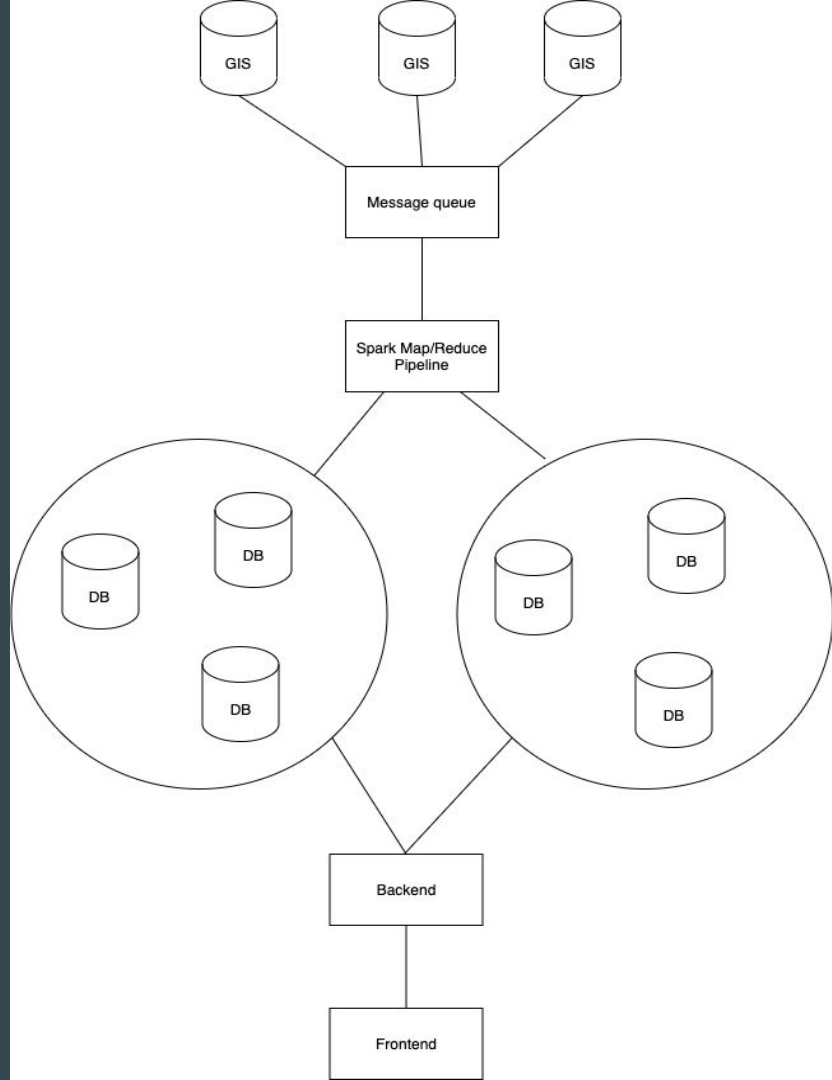
# Analysing GIS Data

...

Radu Catarambol, Shubham Jinde, Richard Westerhof

# General Architecture

- Every entity lives in a Docker container
- Every level is replicated via Kubernetes nodes



# Interacting with and storing the processed data

Storage:

- Output of the Map/Reduce process will be stored in Elasticsearch

Frontend:

- Kibana will act as our frontend for the processed data

Backend:

- Python backend for more complex interactions with the data

# Spark Map/Reduce pipeline

- Decided on using Python => we will use Spark via the **pyspark** package



# Msg queue

- We want a distributed message queue with persistent storage and fault tolerance
- It serves as an intermediary between the GIS databases and the Map/Reduce pipeline
- Kafka, Zookeeper



# GIS Database

- Elasticsearch + Cassandra => Elassandra?
- Data source idea?