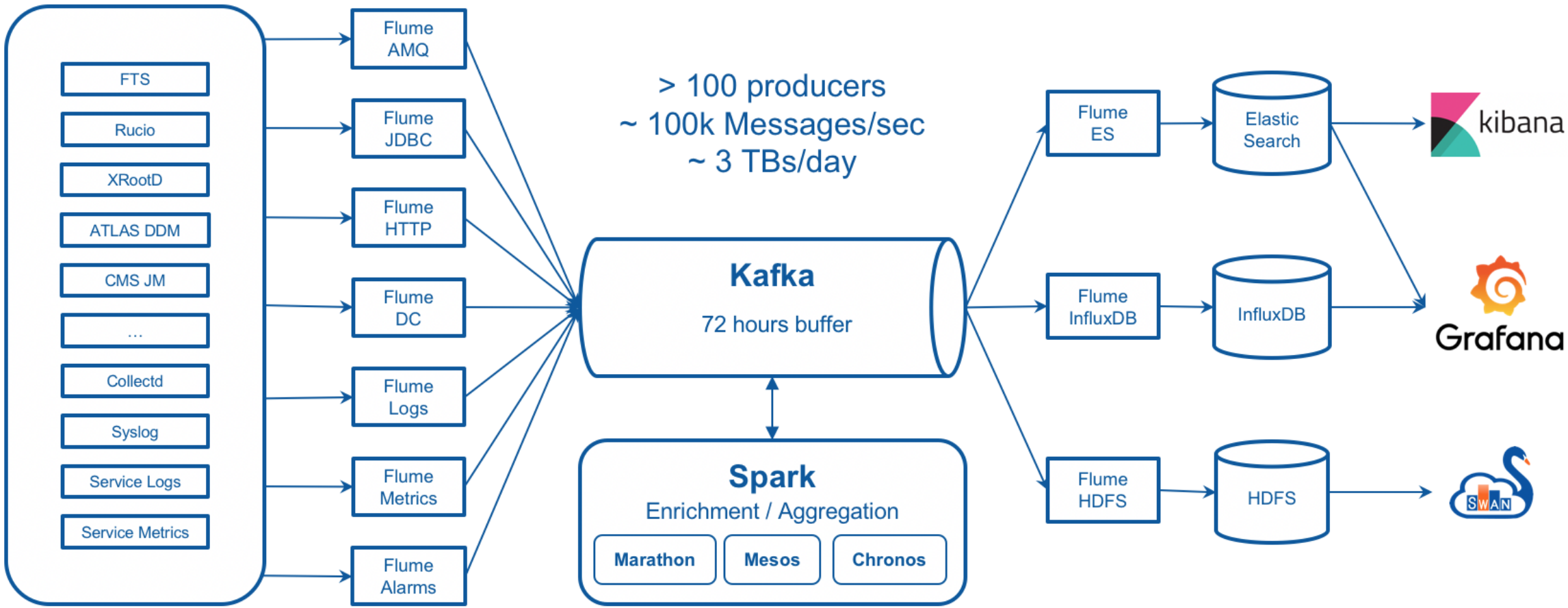
Automatic Alert's Triage

Summer Student Project 2019

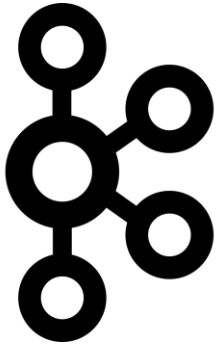
CMS Monitoring Team

Sources > Transport > (Processing) > Storage > Access



Multiple sources and notification channels

- Each subsystem produce its own data, but we have a common messaging service: Kafka



- E-mail
- Tickets (Jira/Snow)
- Messaging Applications

Different expected response time.

Complex patterns

E.g.:

- Timeout message has been repeated 5 times in a 1 hour window for a given node.
- Memory increse alert but there is not a request increase alert.

Alerts Classification

- Notification Channel as target variable
- Features to be determined
- Explore multiples alternatives (e.g. Decision trees, Random Forest, content based recommendation techniques)

Results visualization and presentation

- Storytelling with data guidelines.
- Paper with results

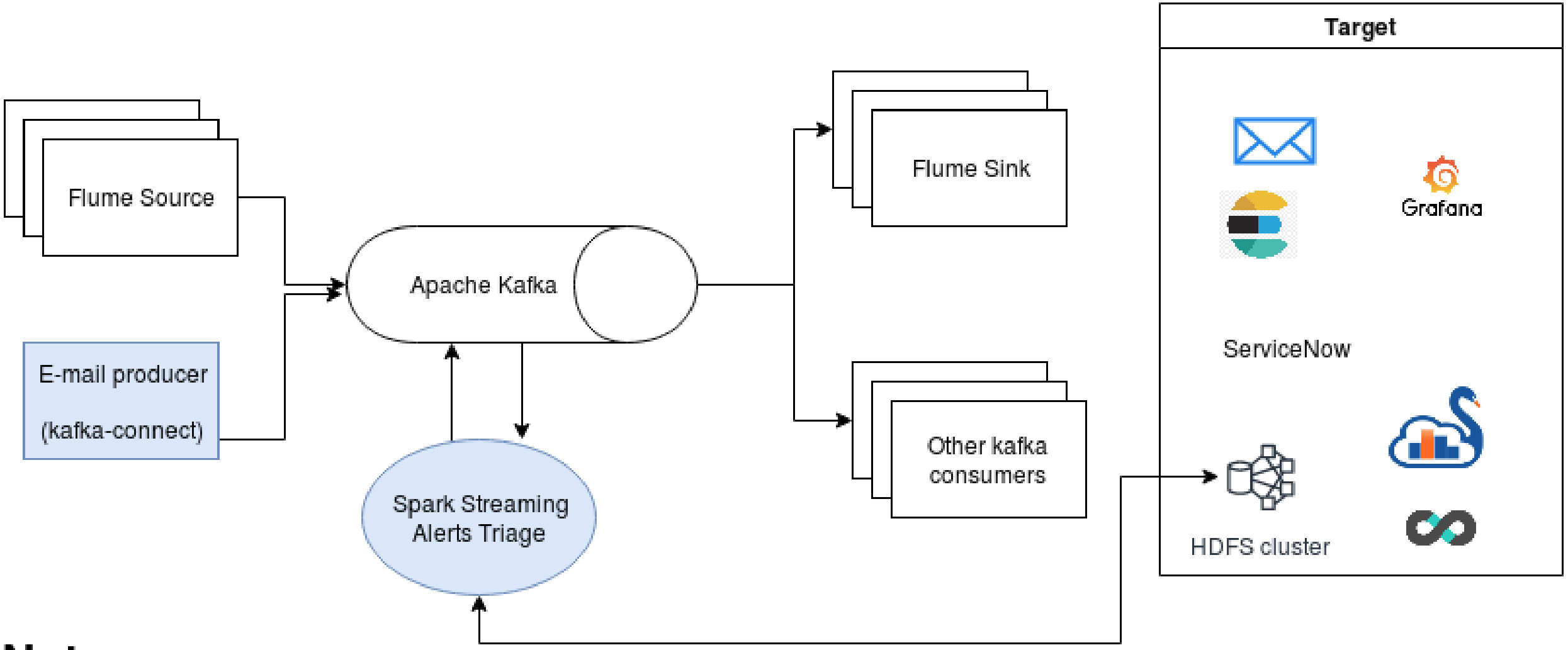
Restrictions

Python and Bash only.

The project must be finished during the summer stay.

Tools

- Apache Spark
- Scikit-Learn
- Spark streaming
- Apache Flink



Notes

Consuming from Kafka, we can use this architecture in a project-agnostic way. Initially, we will work with currently produced alerts, but the architecture will help to produce new alerts from other sources in the future. We are looking forward to new applications.

The project will be released with an Open Source license.