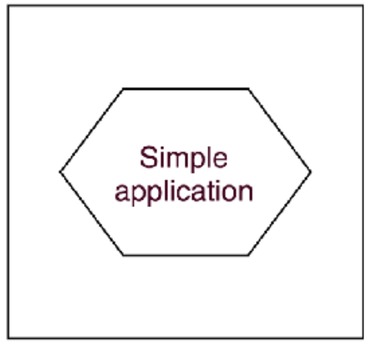
Centralize your logs with Kafka and Elasticsearch

Maciej Ciołek

CTO @ codeheroes_

Simple application



We have a single node application

It`s a web application

It outputs access logs
and application logs

Simple application

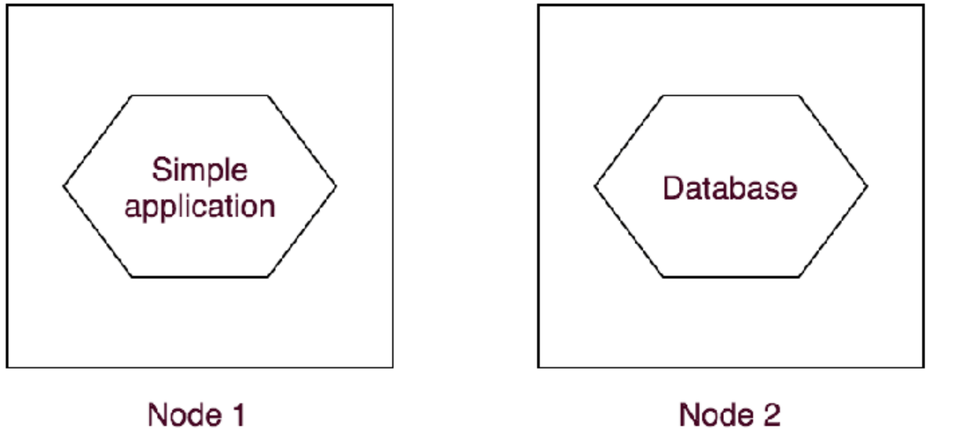
We can just log in via SSH and use:

cat / tail -fn

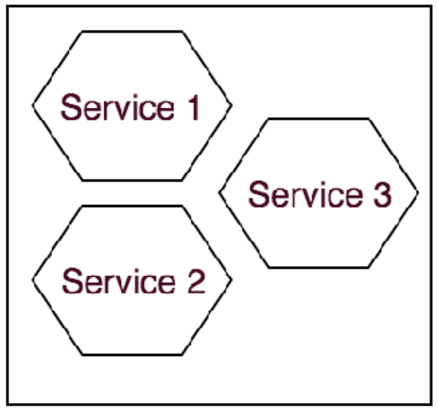
journalctl

or any other local tool

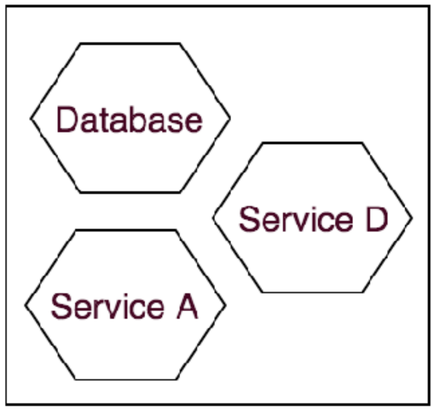
Advanced application



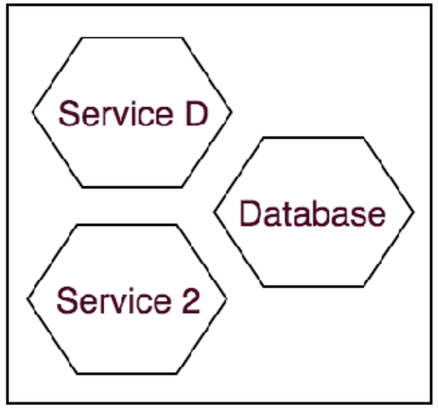
More advanced application



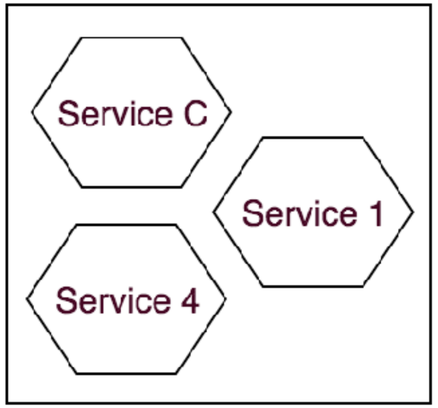
Node 1



Node 2



Node 3



Node 4

New software architectures

Over last few years we can notice a huge changes in software development approach

Our applications don`t look like those in the past

Our logging layer must adopt to new approach

Whats new?

Containers, containers everywhere - Docker

Resource abstraction - Mesos / DCOS / Kubernetes

Different frameworks = different logging approach

Sophisticated business logic = complex logs

What do we need?

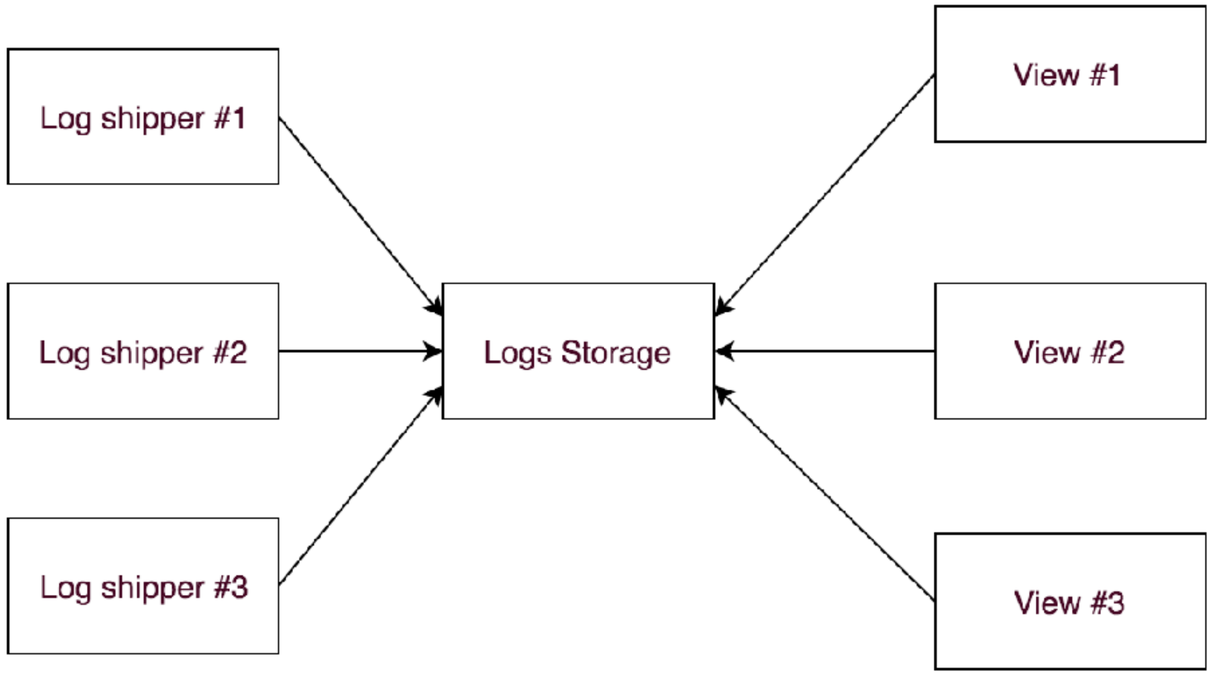
centralized source of logs

scaleable solution

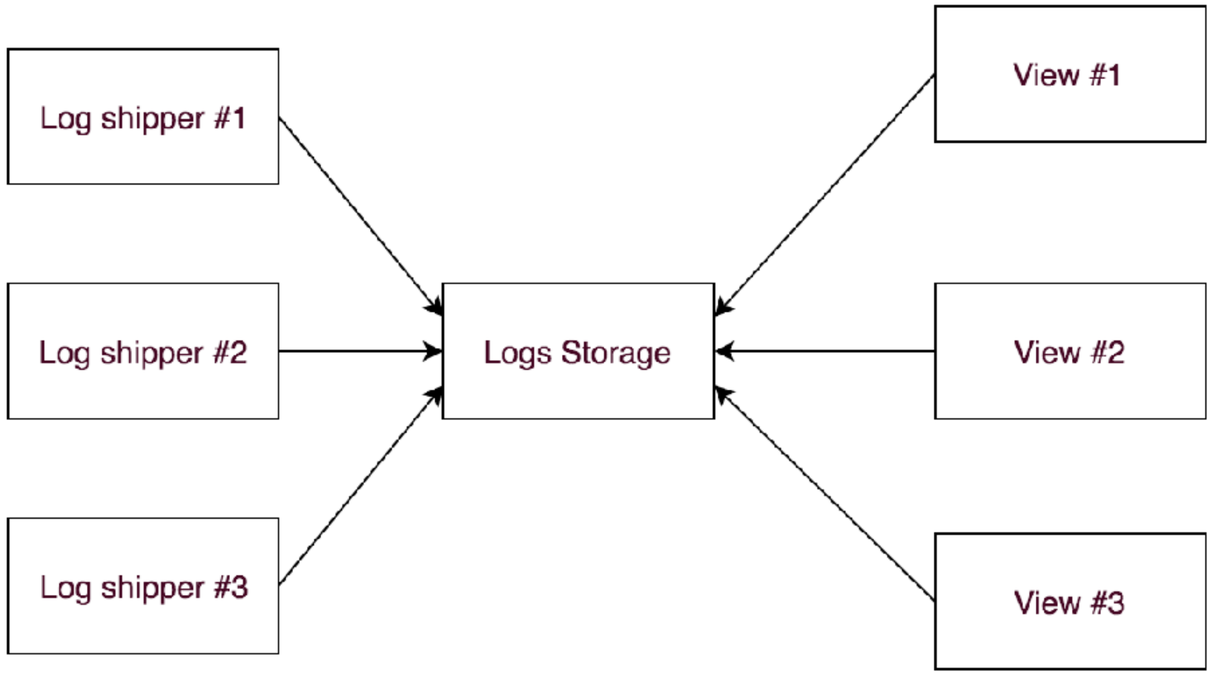
streaming API

ability to handle different views
(historical analysis, real monitoring, alerting)

#1 Logging infrastructure design

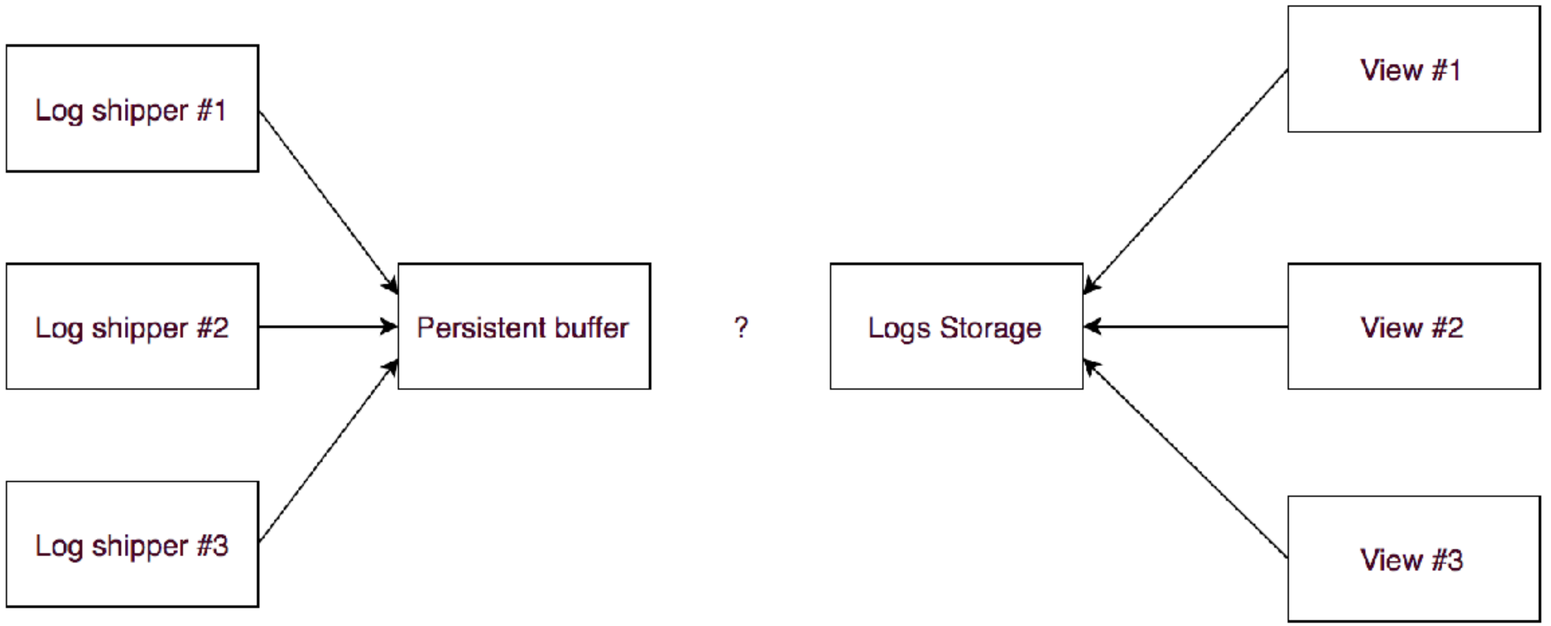


#2 Logging infrastructure design

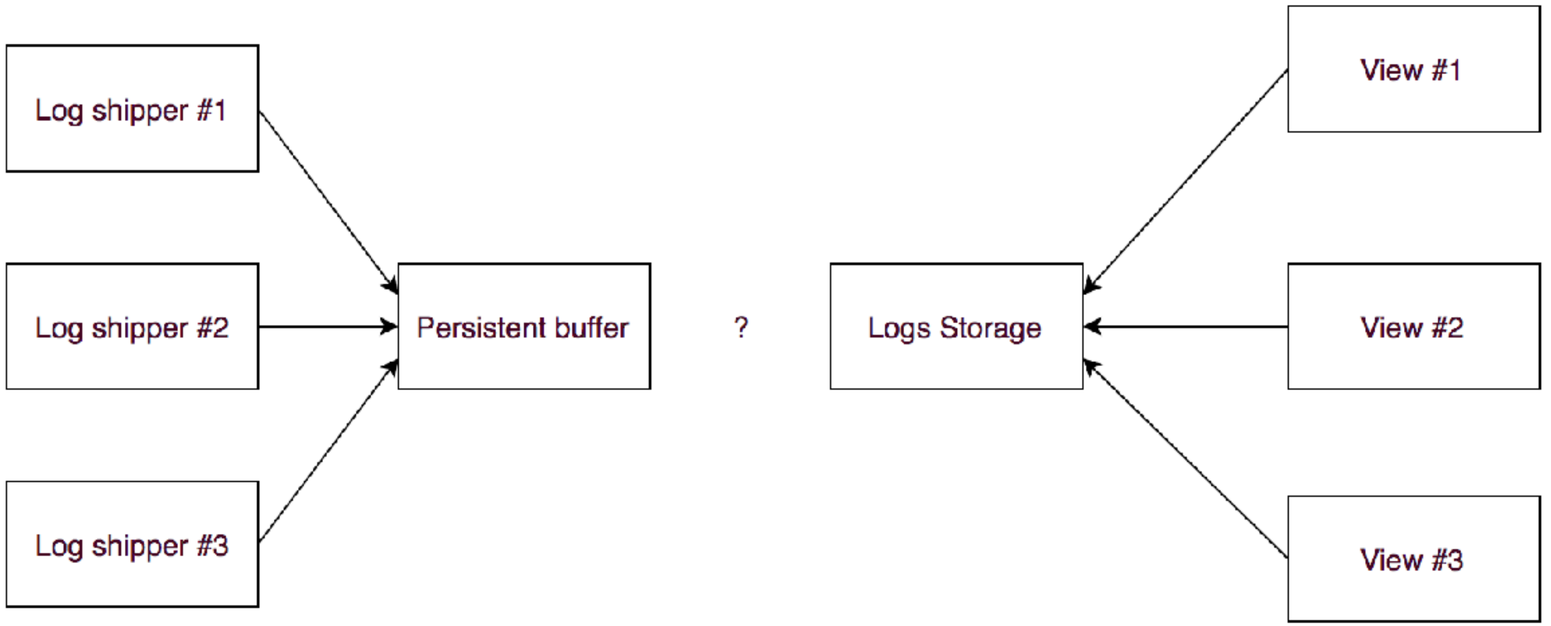


Log shipper must buffer data on its side

#3 Logging infrastructure design

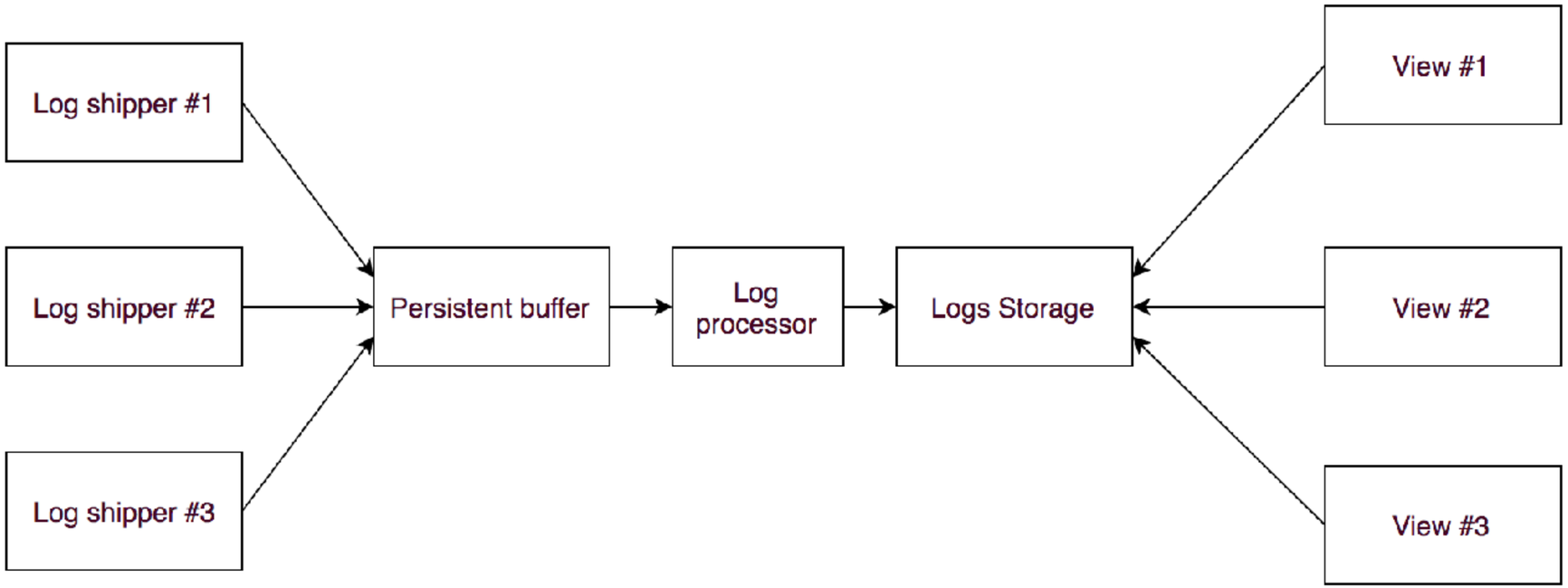


#3 Logging infrastructure design

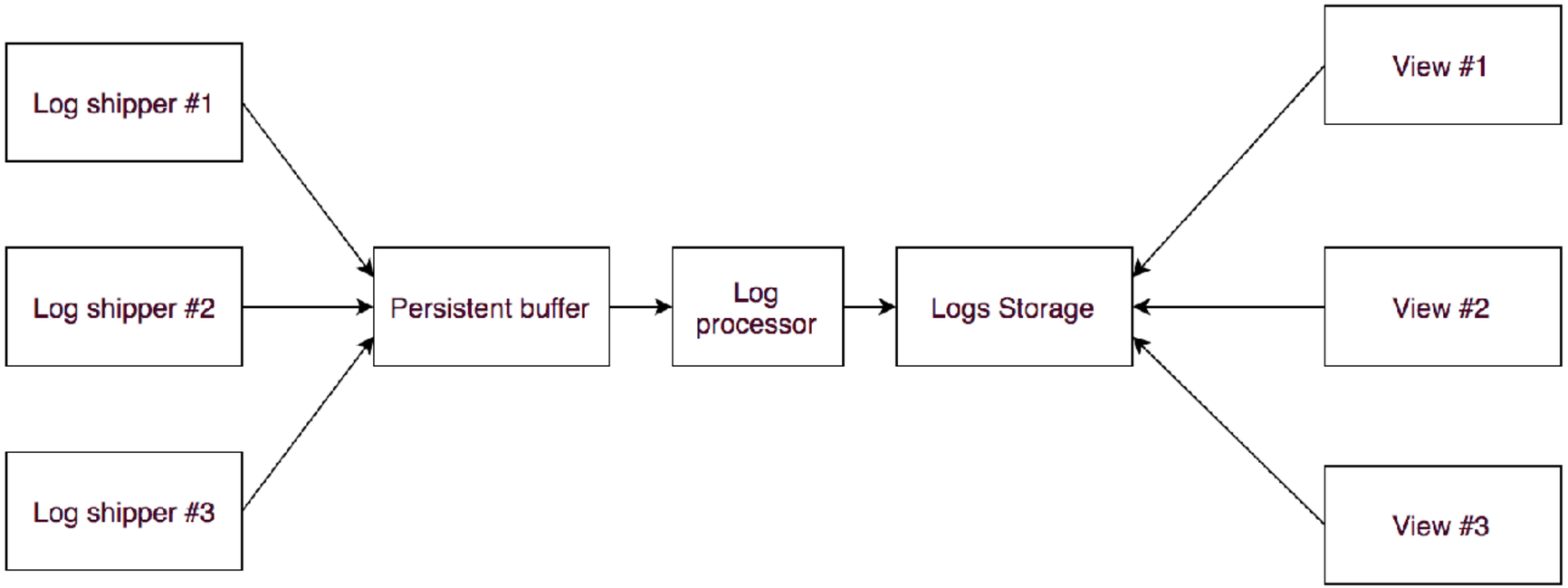


We need to connect buffer to storage

#4 Logging infrastructure design

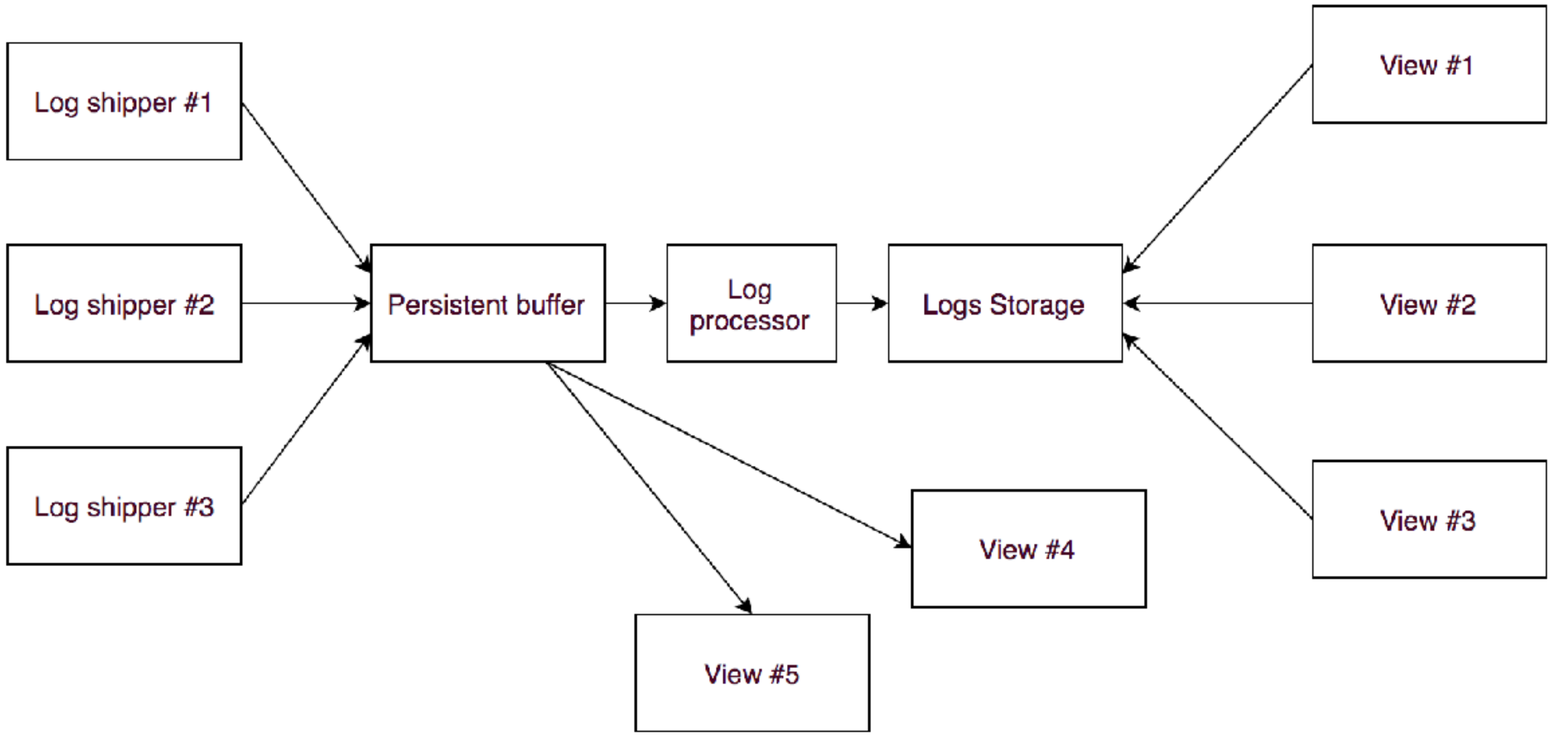


#4 Logging infrastructure design

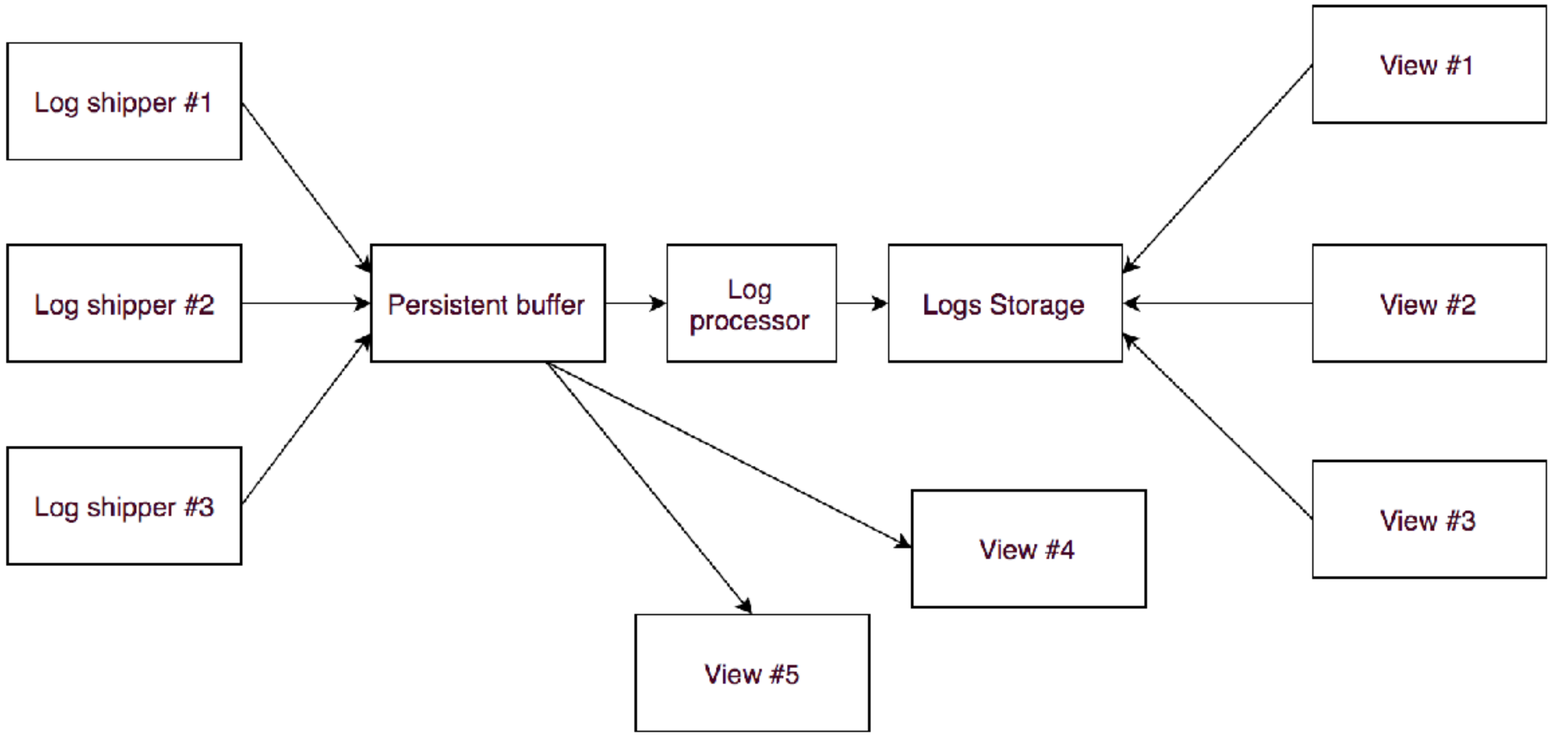


Maybe we can add more views?

#5 Logging infrastructure design



#5 Logging infrastructure design



Let`s match existing tools to our boxes

Kafka

distributed persistent queue

horizontally scaleable

stores record in fault-tolerant way

battle tested by many companies

streaming API

Elasticsearch

search and analytics engine

realtime search

distributed and partition tolerant

REST API

Logstash

data processing pipeline

reads, processes and sends to target

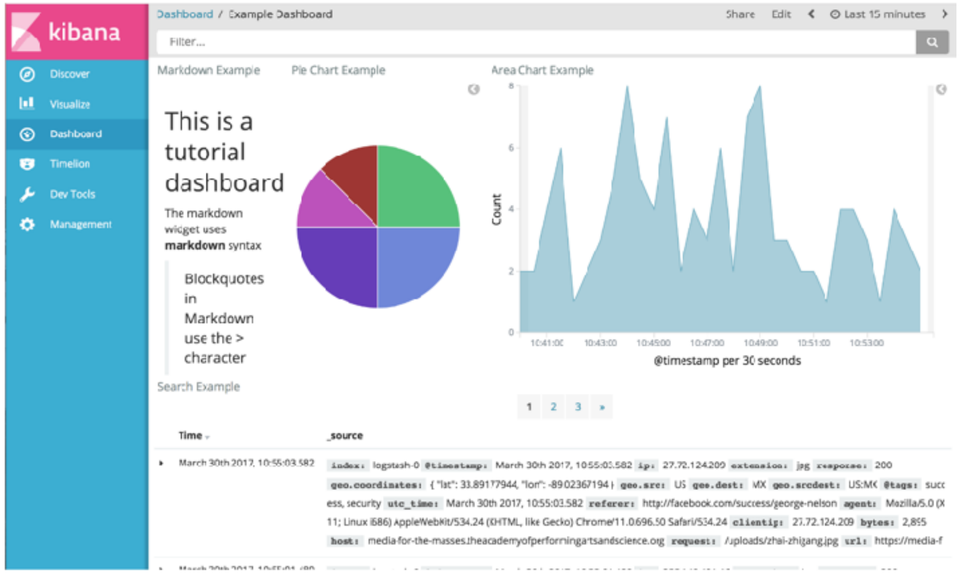
simple configuration

Kibana

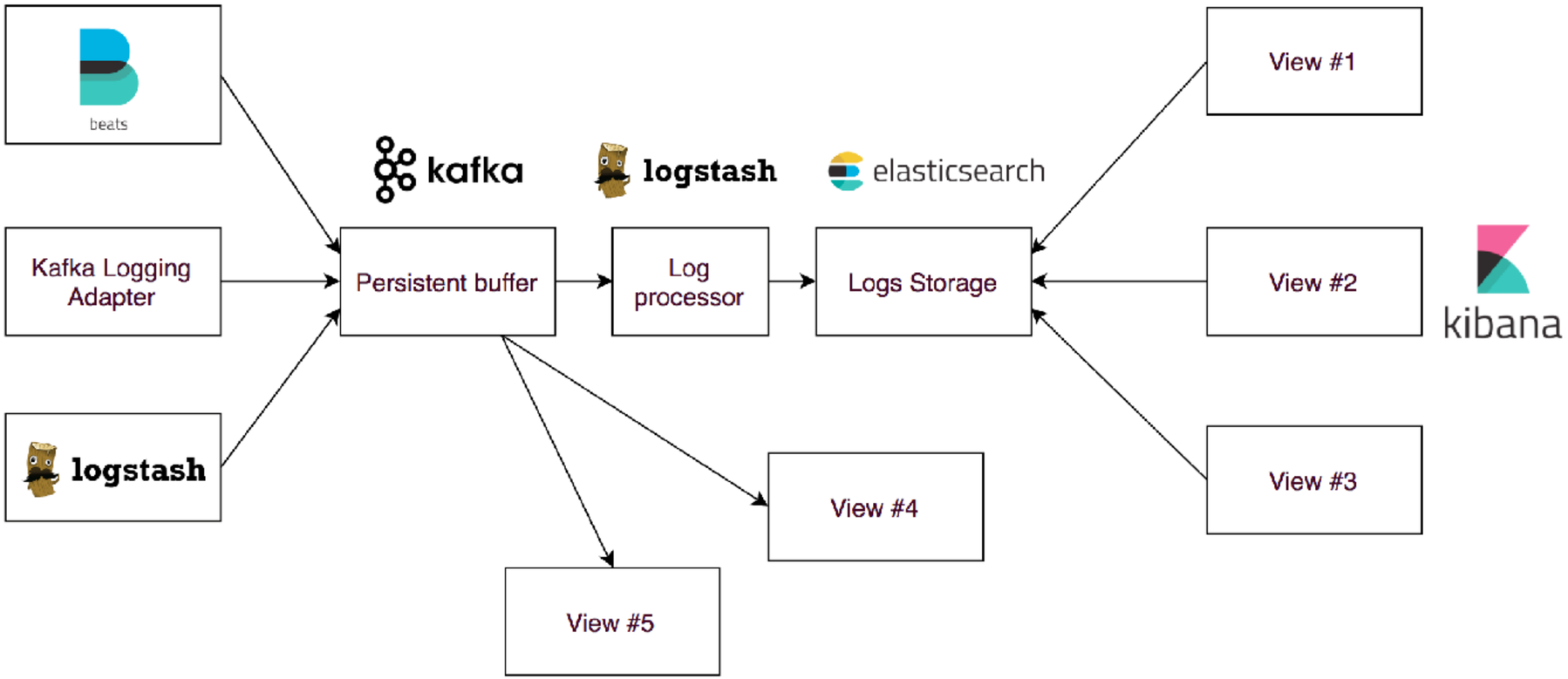
tool for visualising data

build on the top of Elasticsearch

easily save queries, build dashboards



#6 Logging infrastructure design



DEMO

Thank you!

We are using this approach in couple of products
and we are satisfied with it.

Code available here:
https://github.com/codeheroesdev/logging-infrastructure-example

Maciej Ciołek

maciej@codeheroes.io

linkedin.com/in/maciejciolek

@MaciejCiolek