

***Event Stream  
Processing  
with **Kafka*****

Tim Lossen · Wooga

**Kafka?**



kafka download



### Search

About 4,380,000 results (0.08 seconds)

#### Everything

[The Trial by Franz Kafka - Project Gutenberg](#)

[www.gutenberg.org/ebooks/7849](http://www.gutenberg.org/ebooks/7849)

1 Apr 2005 - **Download** the free eBook: The Trial by Franz **Kafka**. ... Author, **Kafka**, Franz, 1883-1924. Translator, Wyllie, David. Title, The Trial. Language ...

#### Images

#### Maps

#### Videos

[Metamorphosis by Franz Kafka - Project Gutenberg](#)

[www.gutenberg.org/ebooks/5200](http://www.gutenberg.org/ebooks/5200)

17 Aug 2005 - **Download** the free eBook: Metamorphosis by Franz **Kafka**. ... Author, **Kafka**, Franz, 1883-1924. Translator, Wyllie, David. Title, Metamorphosis ...

#### News

#### Shopping

#### Applications

#### More

[Metamorphosis by Franz Kafka on Audio Download - Free Audio](#)

[www.learmoutloud.com/Free-Audio-Video/Literature/Short.../17038](http://www.learmoutloud.com/Free-Audio-Video/Literature/Short.../17038)

Metamorphosis, first published in 1915, is the story of Gregor Samsa, a young traveling salesman who lives with his family and financially supports his parents ...

Show search tools

[Metamorphosis by Franz Kafka - download here](#)

[www.e-booksdirectory.com Literature & Fiction](http://www.e-booksdirectory.com/Literature%20%26%20Fiction)

Metamorphosis by Franz **Kafka** - free book at E-Books Directory - **download** here.

[Free Kafka Downloads](#)

[www.kafka-music.co.uk/free\\_downloads.html](http://www.kafka-music.co.uk/free_downloads.html)

Free download of Mysterious Skin track 'Run with the Buffalo'. Click below. mysterious skin. More free **Kafka downloads** coming soon! radio.



**1883 – 1924**





kafka download



Search

About 4,380,000 results (0.08 seconds)

Everything

[Apache Kafka](#)

[incubator.apache.org/kafka/downloads.html](http://incubator.apache.org/kafka/downloads.html)

You can verify your **download** by following these procedures and using these KEYS. ...  
Apache **Kafka** is an effort undergoing incubation at The Apache Software ...

Images

Maps

Videos

[The Trial by Franz Kafka - Project Gutenberg](#)

[www.gutenberg.org/ebooks/7849](http://www.gutenberg.org/ebooks/7849)

1 Apr 2005 - **Download** the free eBook: The Trial by Franz **Kafka**. ... Author, **Kafka**, Franz, 1883-1924. Translator, Wyllie, David. Title, The Trial. Language ...

News

Shopping

Applications

[Metamorphosis by Franz Kafka - Project Gutenberg](#)

[www.gutenberg.org/ebooks/5200](http://www.gutenberg.org/ebooks/5200)

17 Aug 2005 - **Download** the free eBook: Metamorphosis by Franz **Kafka**. ... Author, **Kafka**, Franz, 1883-1924. Translator, Wyllie, David. Title, Metamorphosis ...

More

Show search tools

[Metamorphosis by Franz Kafka on Audio Download - Free Audio](#)

[www.learmoutloud.com/Free-Audio-Video/Literature/Short.../17038](http://www.learmoutloud.com/Free-Audio-Video/Literature/Short.../17038)

Metamorphosis, first published in 1915, is the story of Gregor Samsa, a young traveling salesman who lives with his family and financially supports his parents ...

[Metamorphosis by Franz Kafka - download here](#)

[www.e-booksdirectory.com](http://www.e-booksdirectory.com) > Literature & Fiction

Metamorphosis by Franz **Kafka** - free book at E-Books Directory - **download** here.

# Apache Kafka

*A high-throughput distributed messaging system.*

[download](#)

[api docs](#)

[quickstart](#)

[design](#)

[configuration](#)

[performance](#)

[operations](#)

[faq](#)

[wiki](#)

[bugs](#)

[mailing lists](#)

[powered by](#)

[developers](#)

◦ [code](#)

◦ [projects](#)

◦ [contributing](#)

◦ [coding guide](#)

◦ [unit tests](#)

## Apache Kafka is a distributed publish-subscribe messaging system

Apache Kafka is a distributed publish-subscribe messaging system. It is designed to support the following

- Persistent messaging with O(1) disk structures that provide constant time performance even with many TB of stored messages.
- High-throughput: even with very modest hardware Kafka can support hundreds of thousands of messages per second.
- Explicit support for partitioning messages over Kafka servers and distributing consumption over a cluster of consumer machines while maintaining per-partition ordering semantics.
- Support for parallel data load into Hadoop.

Kafka provides a publish-subscribe solution that can handle all activity stream data and processing on a consumer-scale web site. This kind of activity (page views, searches, and other user actions) are a key ingredient in many of the social feature on the modern web. This data is typically handled by "logging" and ad hoc log aggregation solutions due to the throughput requirements. This kind of ad hoc solution is a viable solution to providing logging data to an offline analysis system like Hadoop, but is very limiting for building real-time processing. Kafka aims to unify offline and online processing by providing a mechanism for parallel load into Hadoop as well as the ability to partition real-time consumption over a cluster of machines.

The use for activity stream processing makes Kafka comparable to [Facebook's Scribe](#) or [Apache Flume](#) (incubating), though the architecture and primitives are very different for these systems and make Kafka more comparable to a traditional messaging system. See our [design](#) page for more details.

*Apache Kafka is an effort undergoing incubation at The Apache Software Foundation (ASF), sponsored by the Incubator PMC. Incubation is required of all newly accepted projects until a further review indicates that the infrastructure, communications, and decision making process have stabilized in a manner consistent*

***high-throughput  
persistent  
publish-subscribe  
messaging***

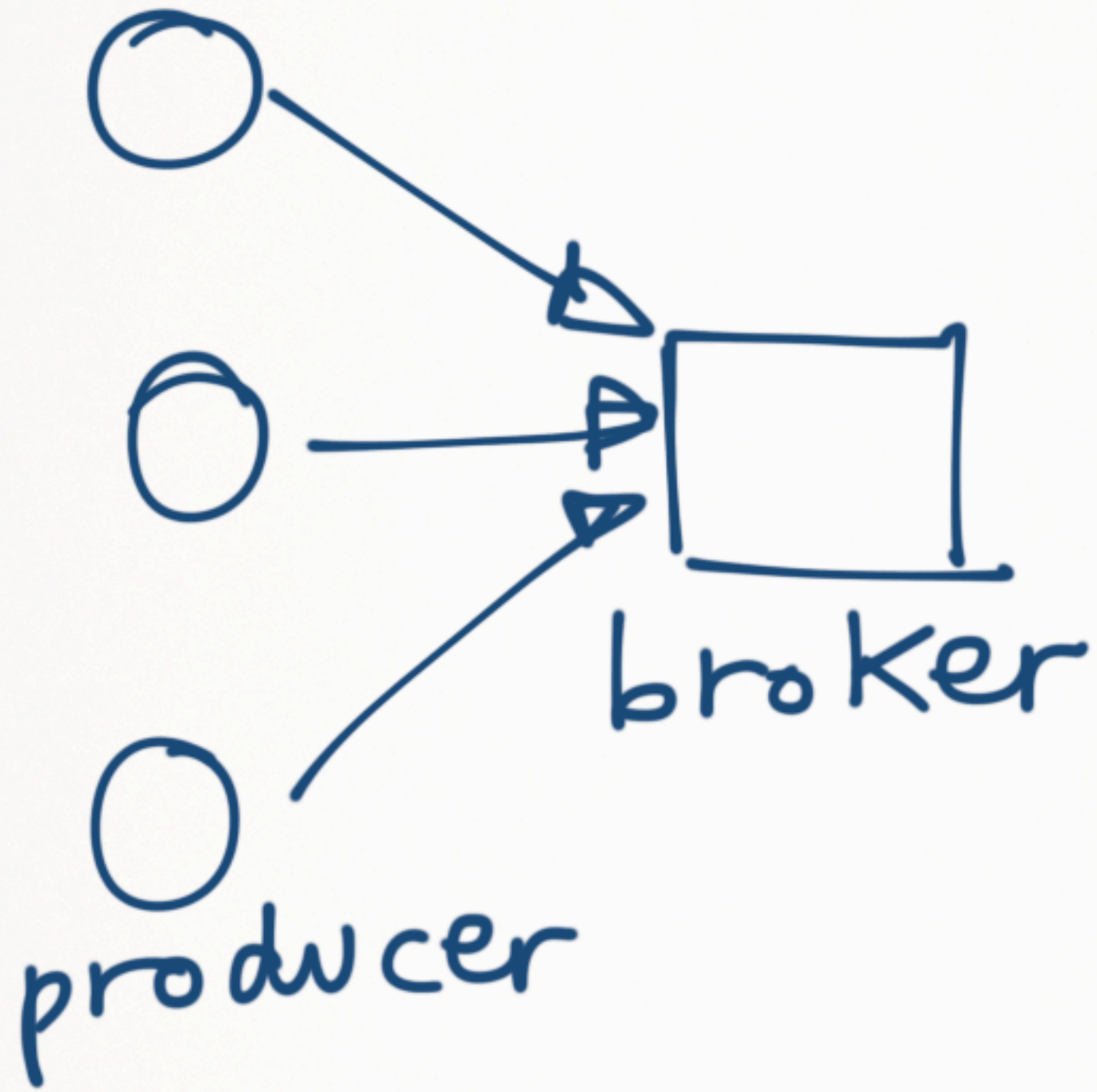


***high-throughput  
persistent  
publish-subscribe  
messaging***

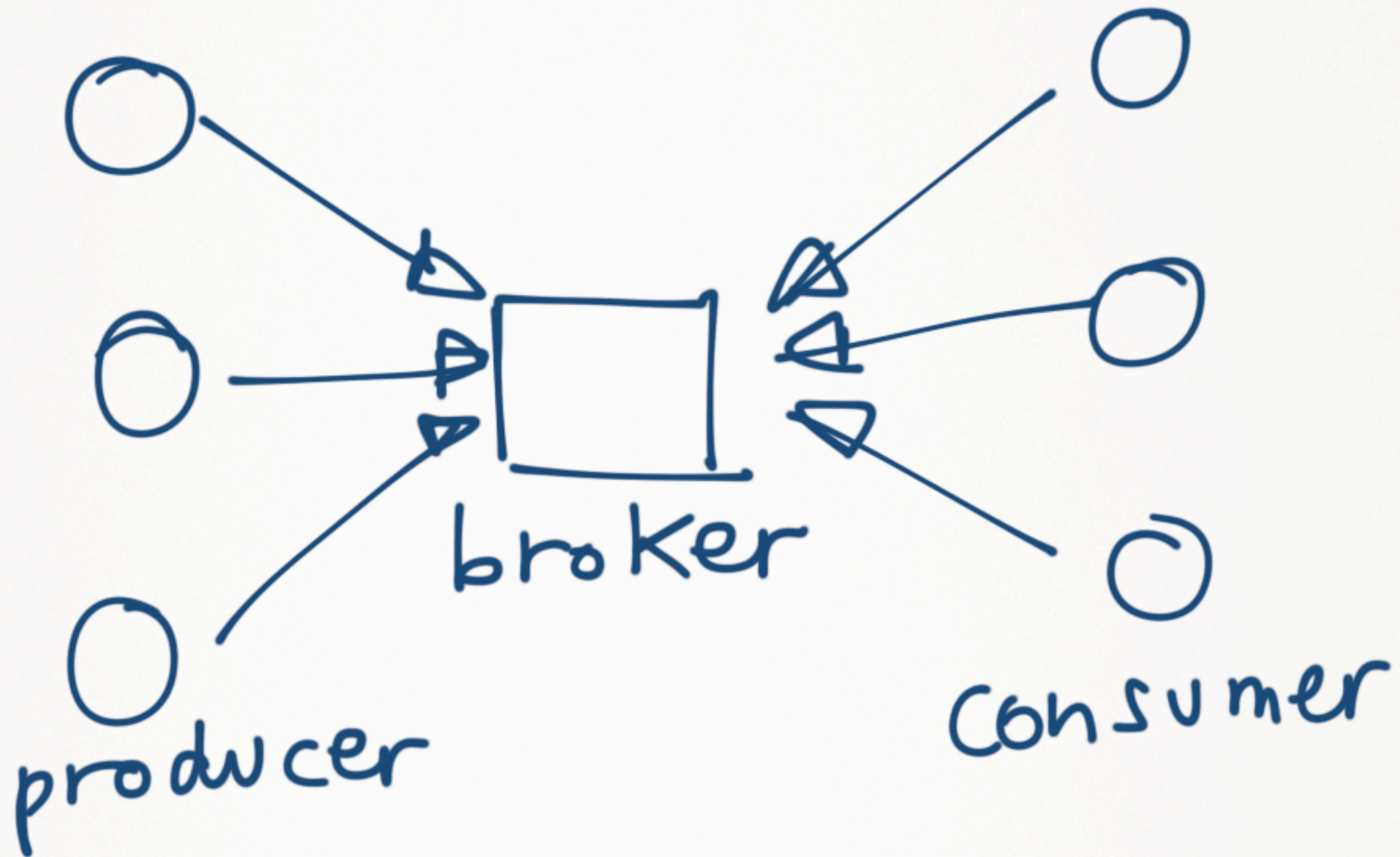




broker



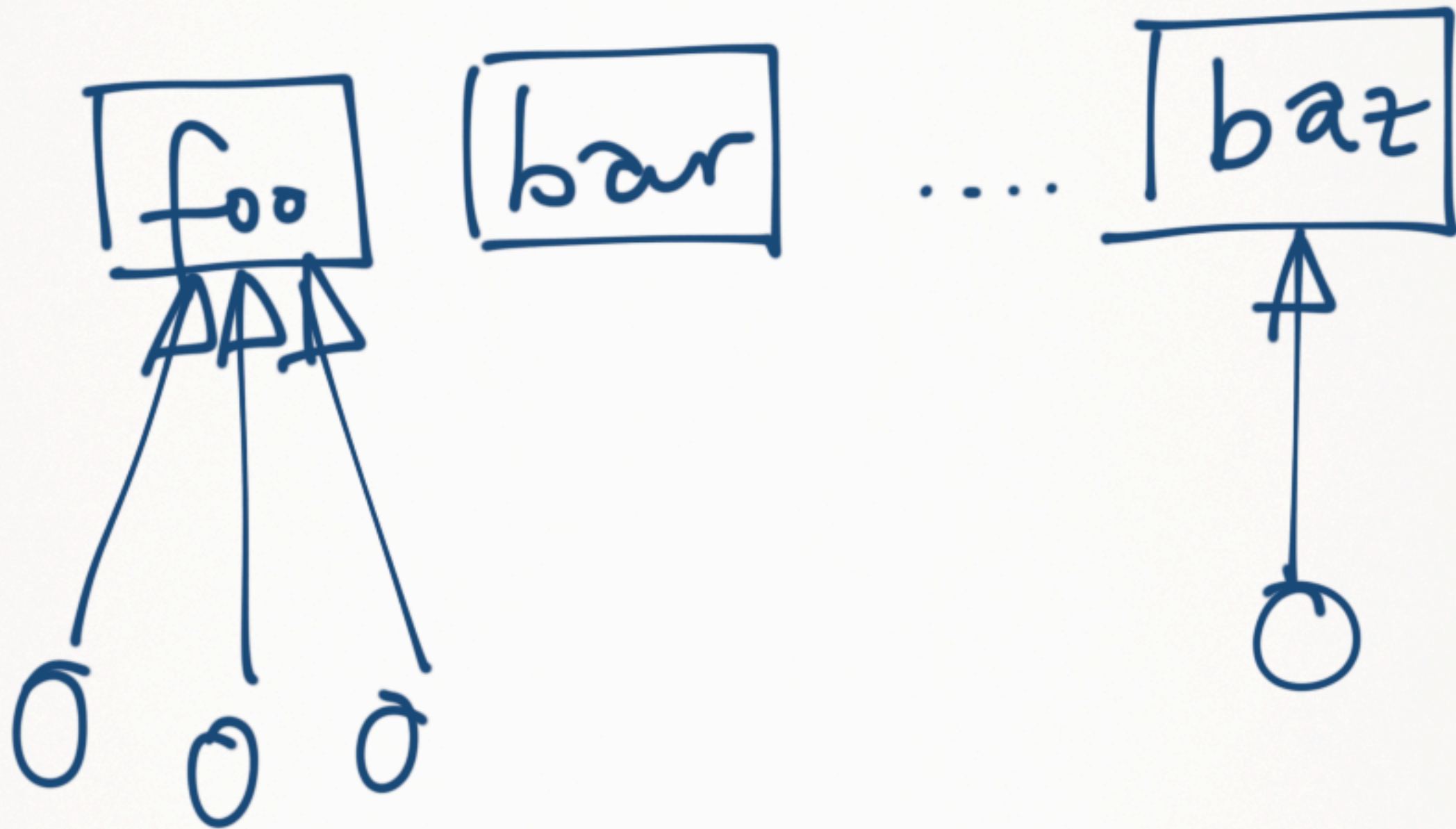




***high-throughput  
persistent  
publish-subscribe  
messaging***



foo bar ... baz



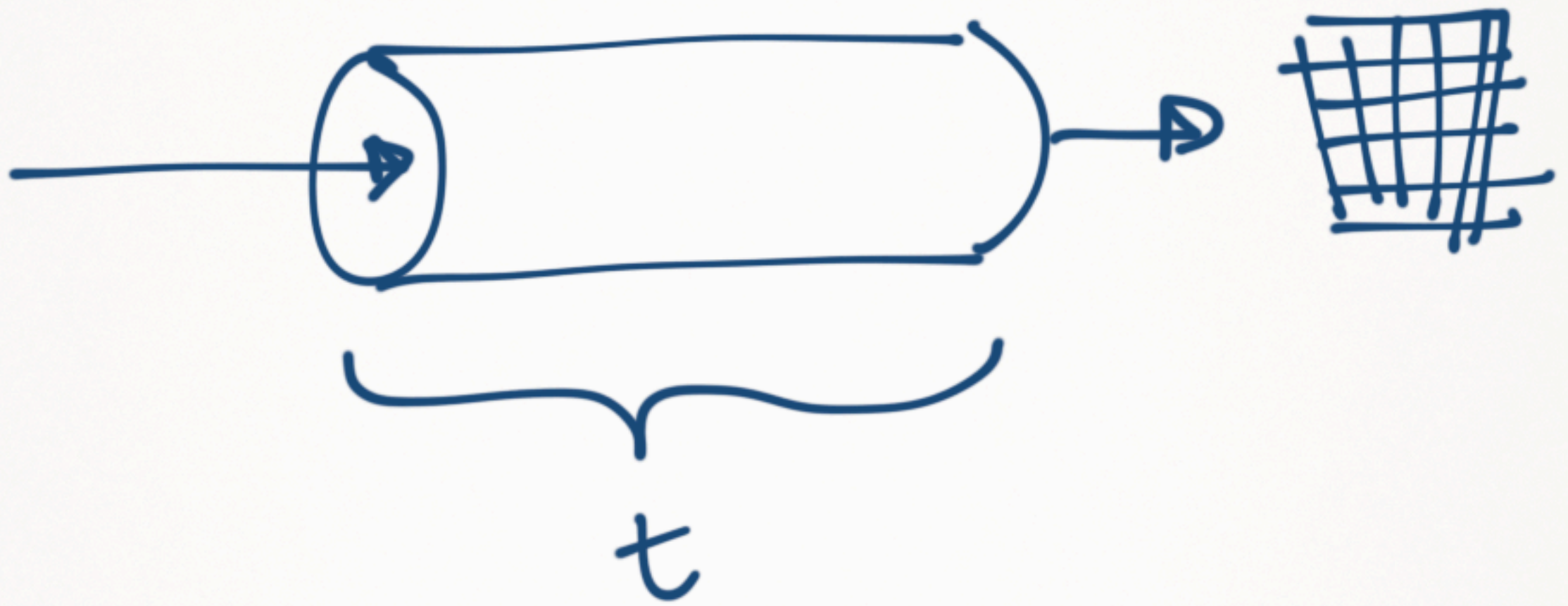
***high-throughput***  
***persistent***  
***publish-subscribe***  
***messaging***





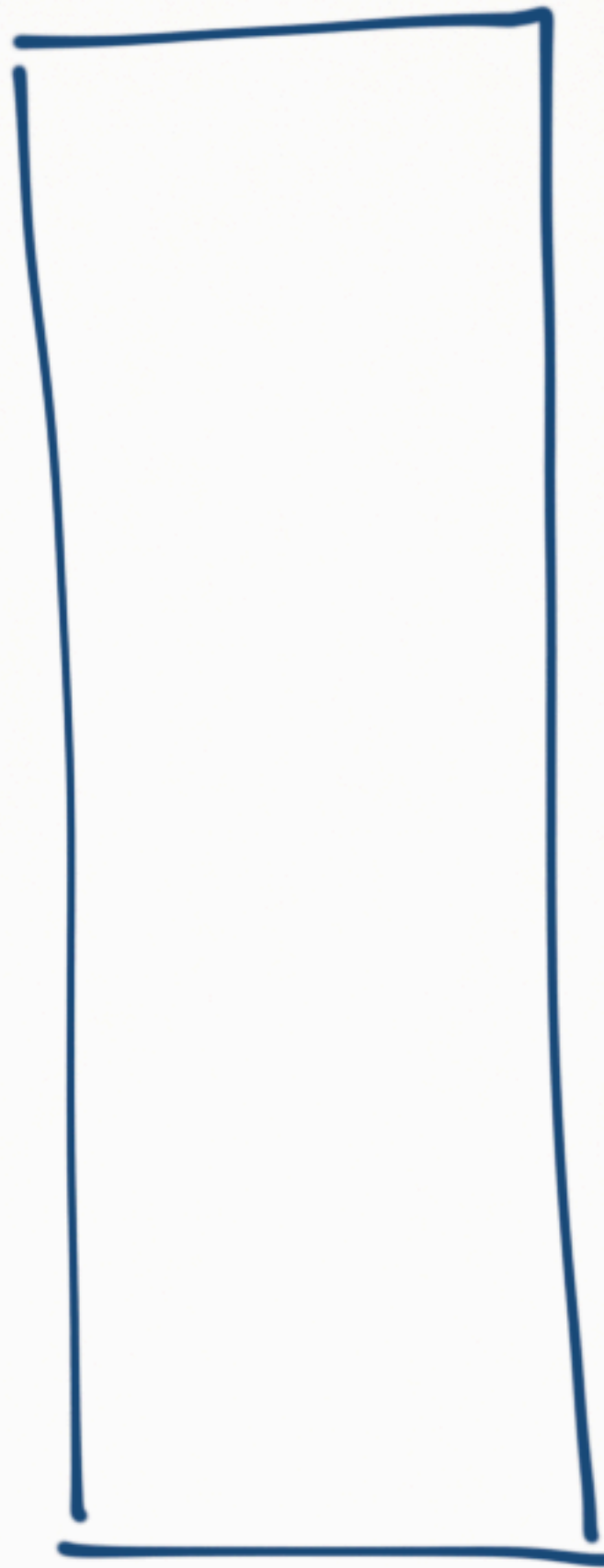
BAMBOO PAPER





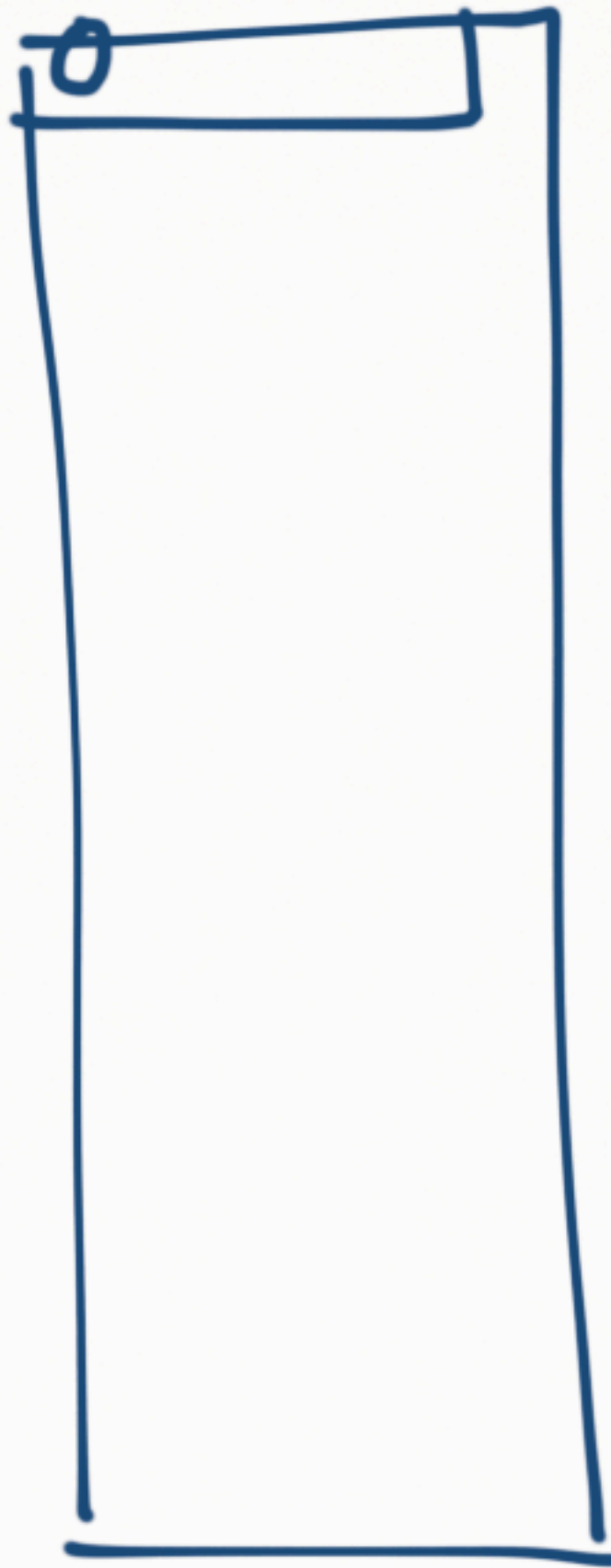
BAMBOO PAPER

***high-throughput***  
***persistent***  
***publish-subscribe***  
***messaging***



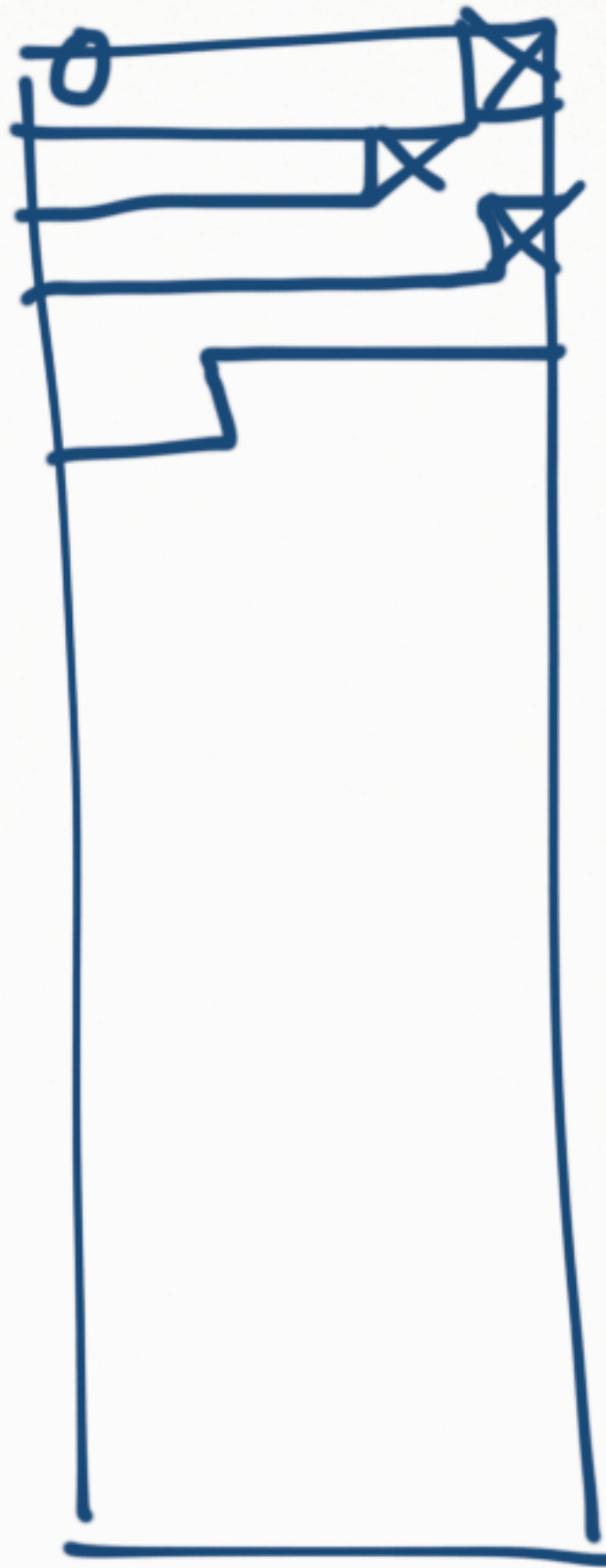
topic



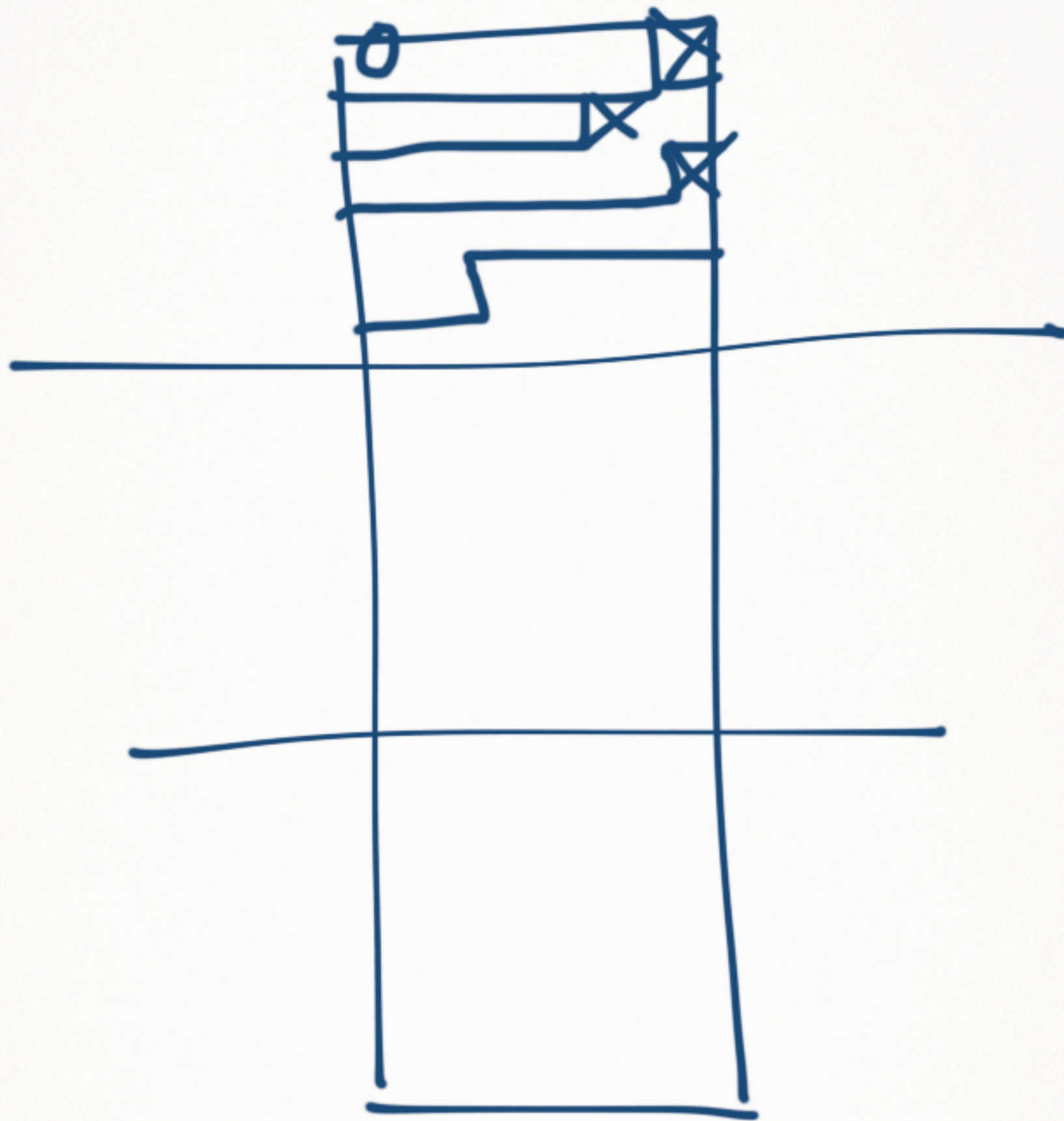


topic





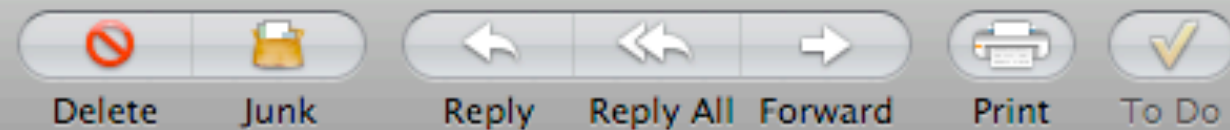
topic



topic

# Case Study: Tracking





**From:** Report Guy  
**Subject:** Daily Happy Hospital report: 2011-06-05  
**Date:** June 6, 2011 7:35:26 AM GMT+02:00  
**To:** HH-Report

## User Summary (FB)

|   |         |          |
|---|---------|----------|
| DAU (login last 24h)                                | 426.315 | (+3.305) |
| New users (last 24h)                                | 58.417  | (-119)   |
| 1-Day Retention                                     | 25,3%   | (+0,9%)  |
| DRU/WAU   | 32,02%  | (+0,12%) |
| Unique user visits coming from Requests as % of DAU | 17,7%   | (+1,8%)  |
| Unique user visits coming from Feeds as % of DAU    | 2,2%    | (+0,0%)  |

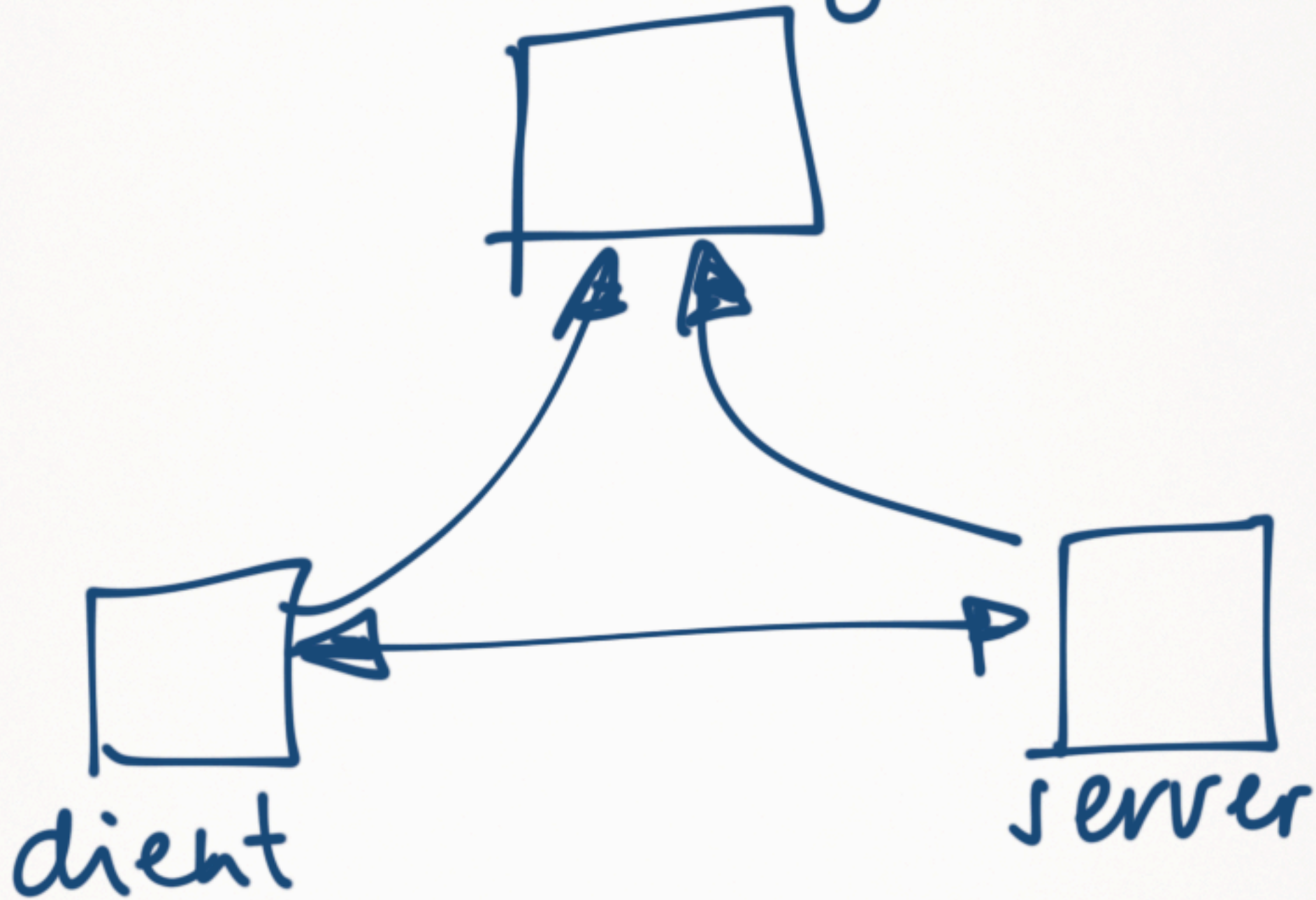
## User Summary (FB)

|                           |           |           |
|---------------------------|-----------|-----------|
| DAU (login last 24h)      | 426.315   | (+3.305)  |
| % of DAU from A-Countries | 45,4%     | (+2,7%)   |
| DRU                       | 367.898   | (+3.424)  |
| WAU                       | 1.149.013 | (+6.548)  |
| MAU (login last 30d)      | 2.572.241 | (+34.306) |
| Sticky Factor (DAU/MAU)   | 16,6%     | (-0,1%)   |

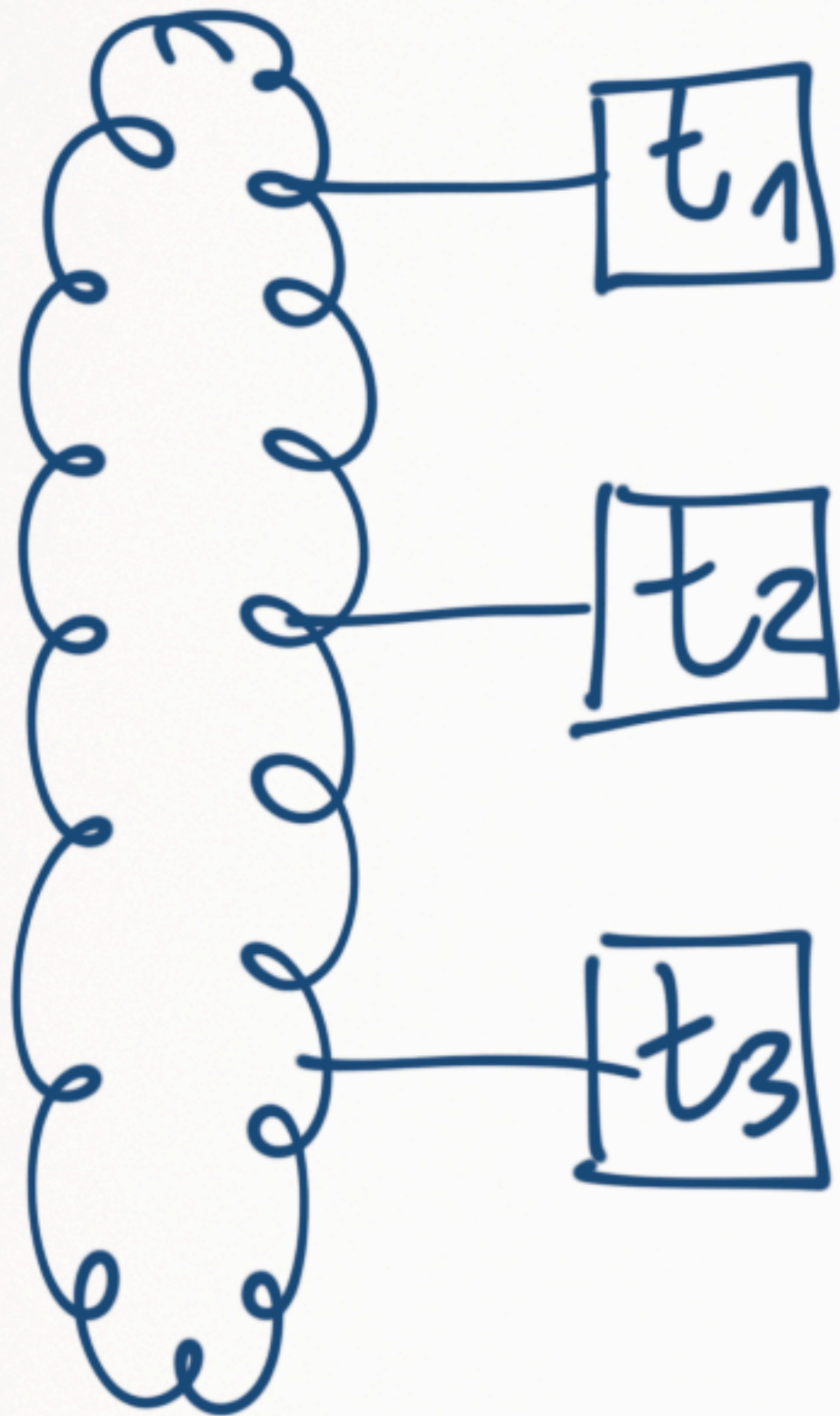




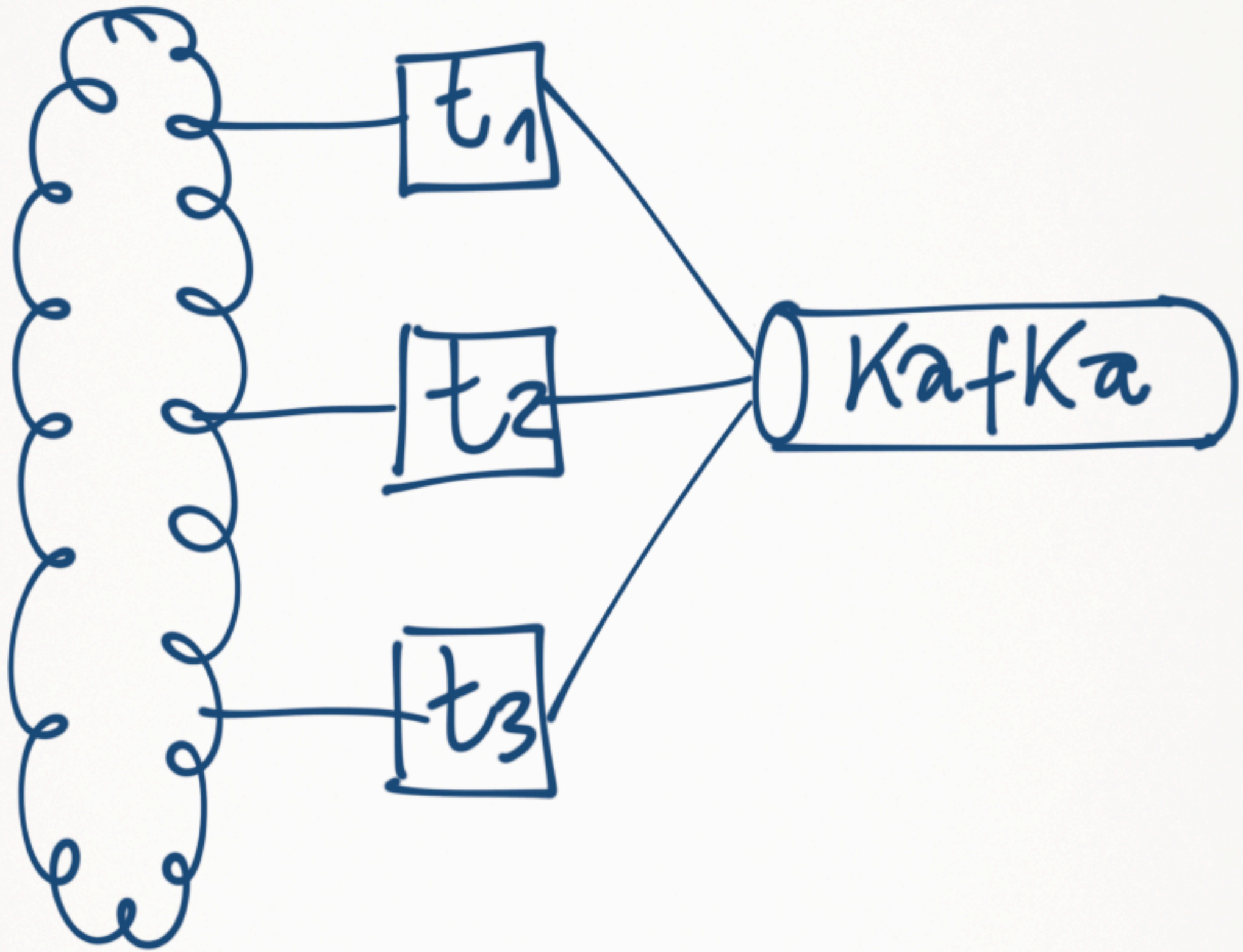
tracking



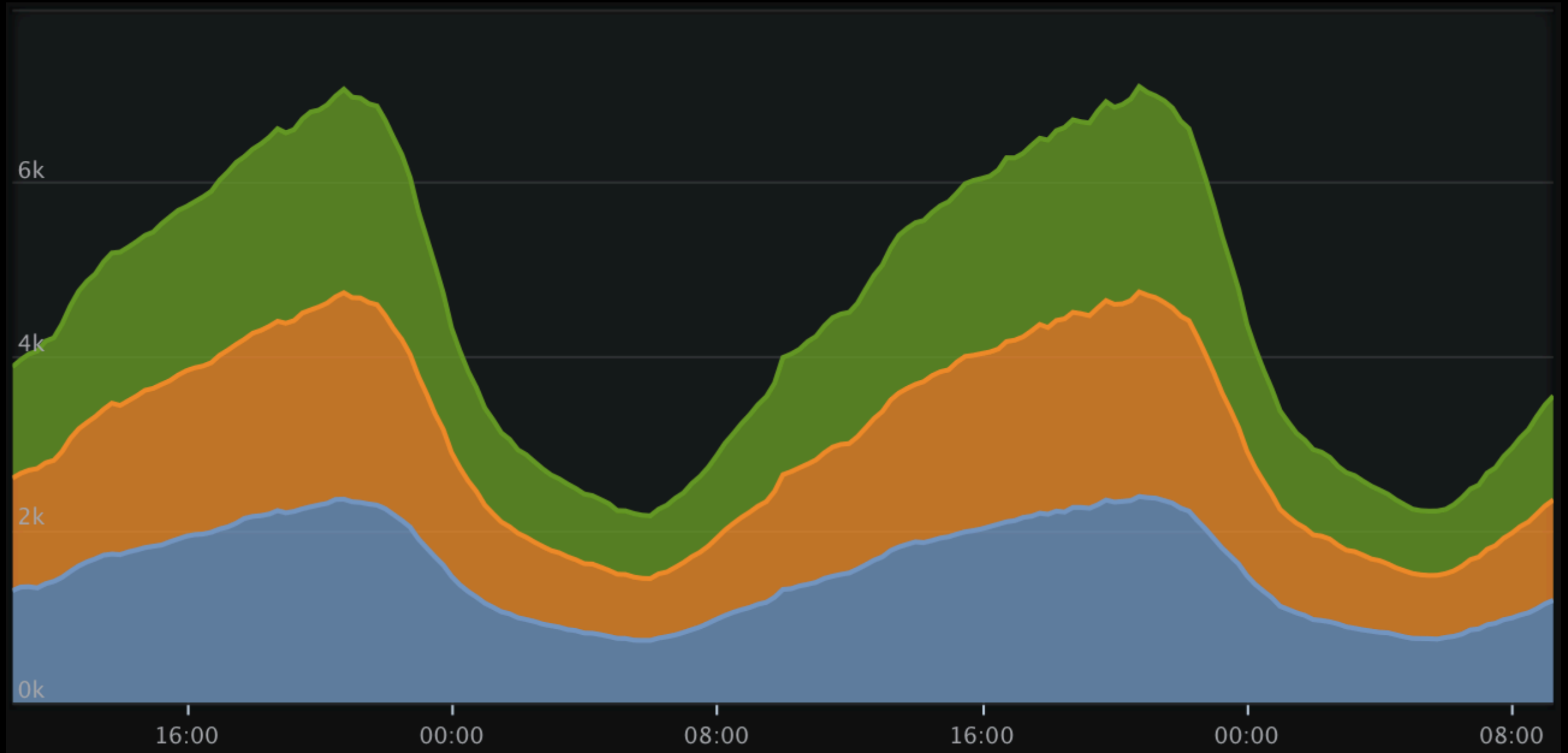




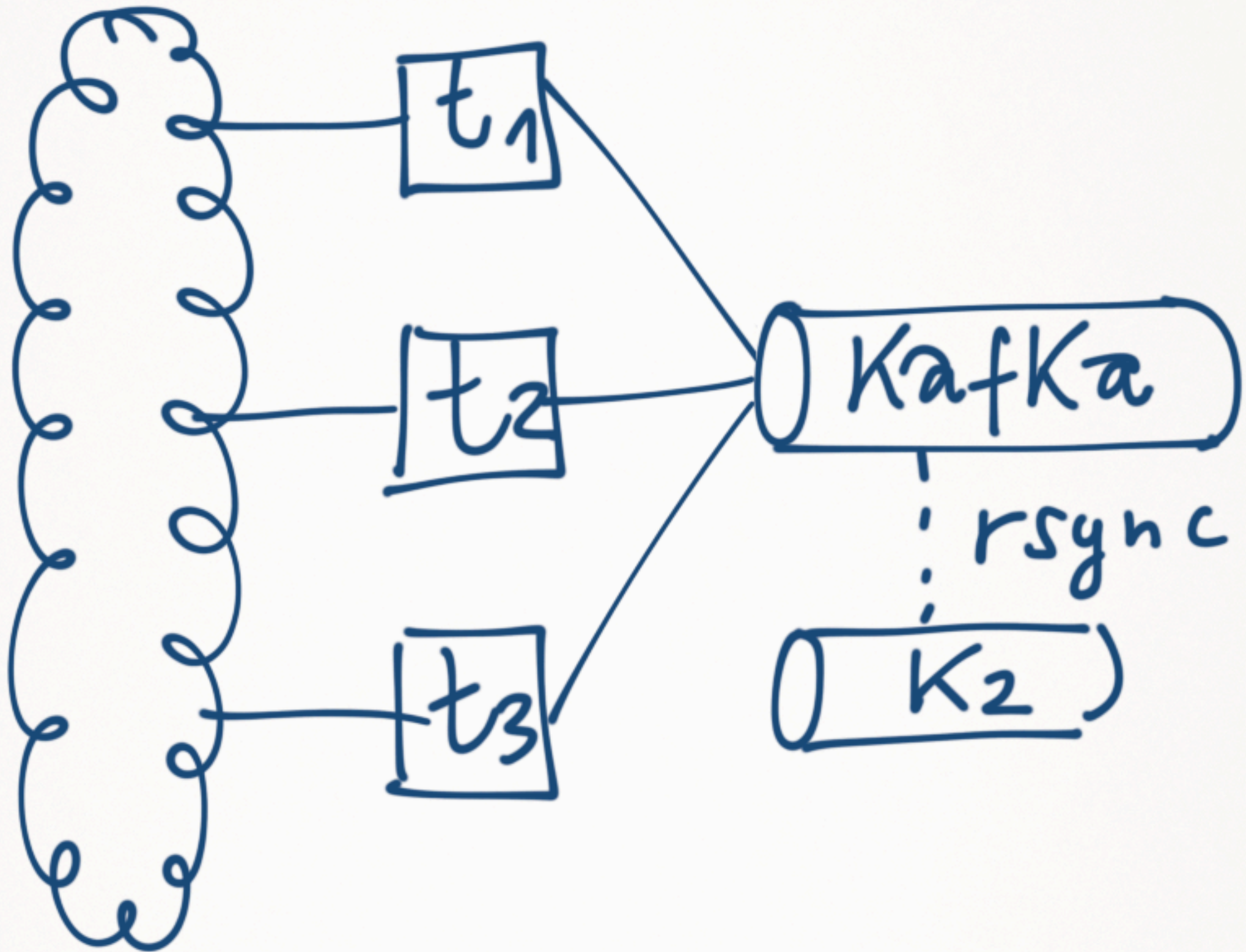




# *requests per second*











mysql



mysql



S3





mysql



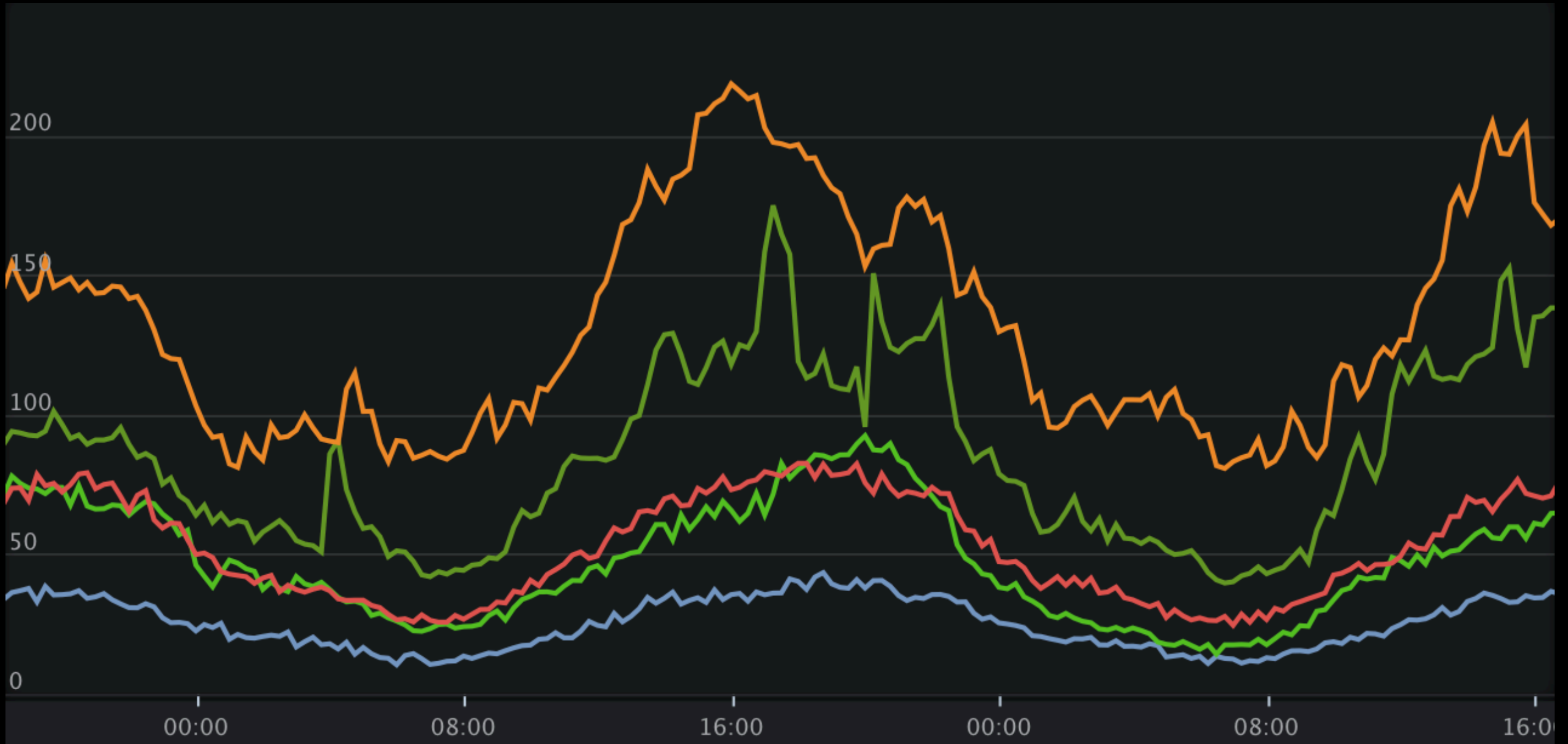
S3



librato



# *installs per minute*





mysql



S3



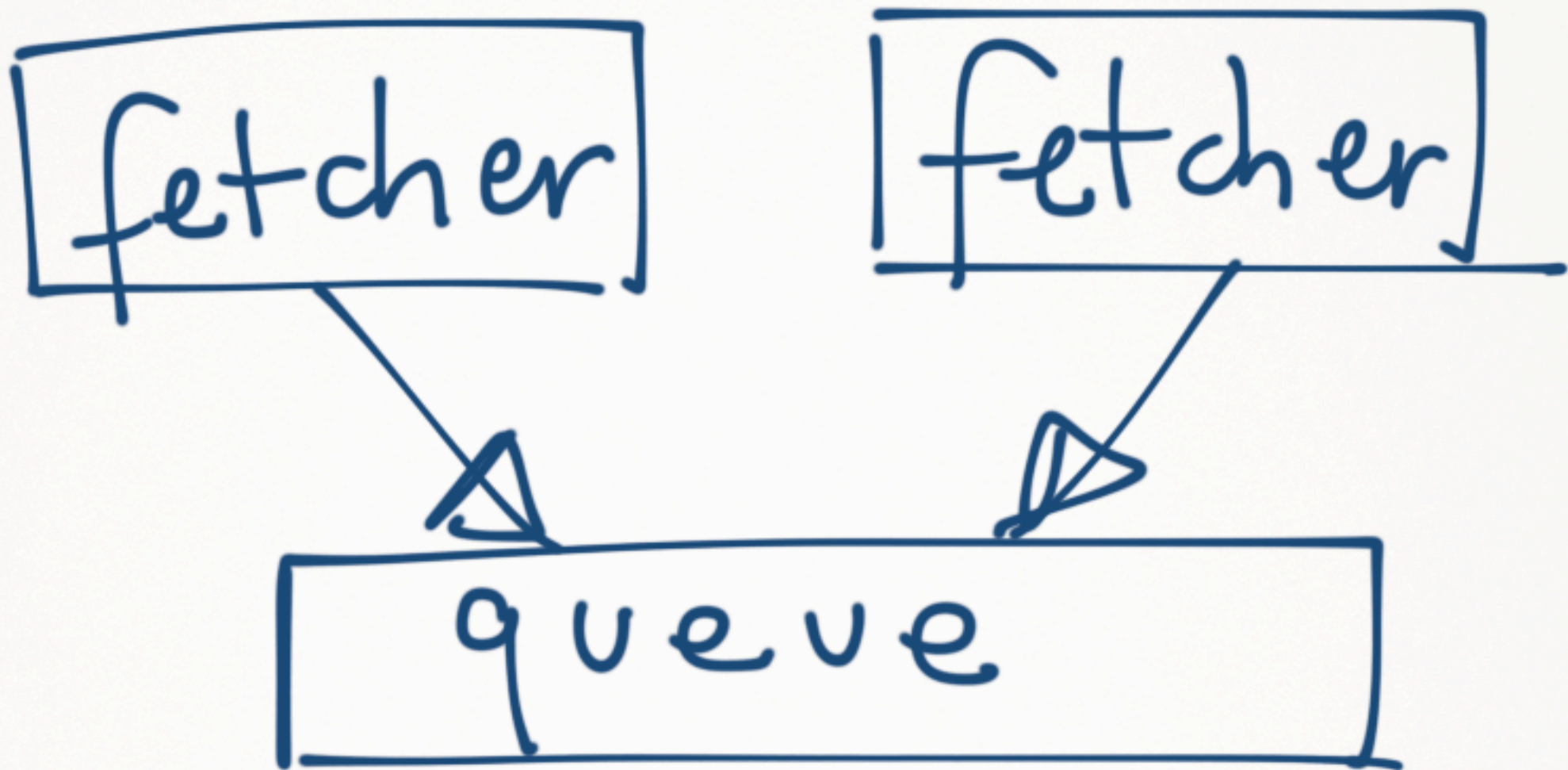
librato

.....

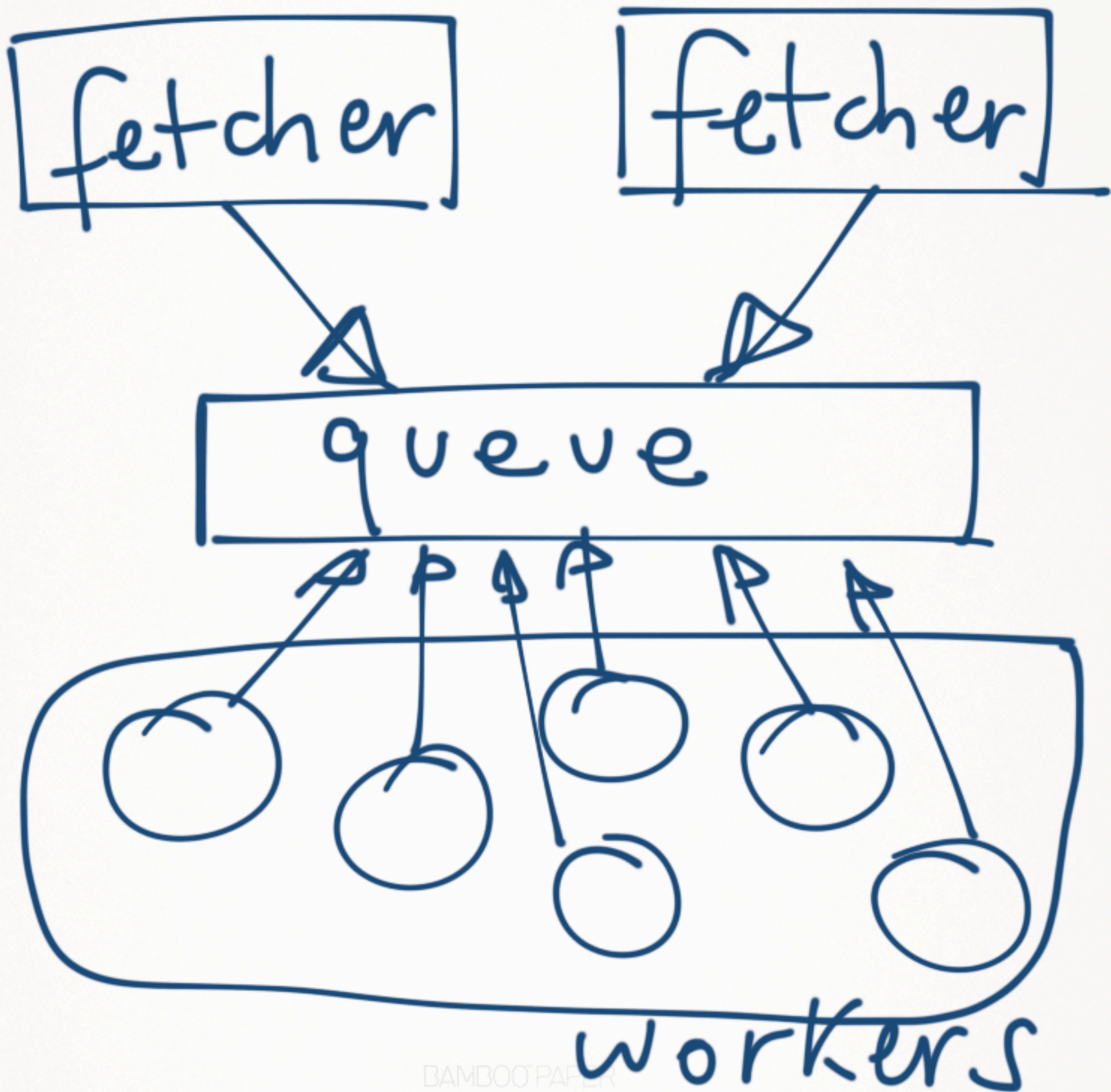
**Kafkaesque**



queue









```
1 require 'kafkaesque'
2
3 Kafkaesque::Consumer.new(
4   :handler => StatsHandler,
5   :kafka => {
6     :host => '192.168.0.22',
7     :topics => StatsHandler::GAMES
8   }
9 ).start
```

```
1 require 'redis'
2
3 class StatsHandler
4   GAMES = %w[mw hh dd ml bi]
5   TIMEOUT = 4 * 60 * 60 #4 hours
6
7   def initialize(config)
8     @redis = Redis.new(:host => config[:kafka][:host], :db => 0)
9   end
10
11  def handle(event)
12    @redis.pipelined do
13      key = event.time.strftime("stats:#{event.game}:%Y_%m%d_%H%M")
14      @redis.hincrby(key, event.call, 1)
15      @redis.expire(key, TIMEOUT)
16    end
17  end
18 end
```

***github.com/wooga/kafka-rb***

***github.com/wooga/kafka-erlang***

***github.com/wooga/kafkaesque***



***Questions?***





# ★ WANTED ★

~~DEAD OR ALIVE~~



**BACKEND DEVELOPER**

★ ★ ★ **REWARD** ★ ★ ★  
**WWW.WOOGA.COM/JOBS**



# ★ WANTED ★

~~DEAD OR ALIVE~~



**BACKEND DEVELOPER**

**REWARD** ★ ★ ★

★ ★ ★ **WWW.WOOGA.COM/JOBS** ★ ★ ★