

# Contents

<b>1</b>	<b>TODO: choose a title</b>	<b>1</b>
1.1	Plan . . . . .	1
1.2	The Elephant Graveyard . . . . .	2

## 1 TODO: choose a title

TODO: tl;dr: ... (3 sentences max)

TODO: introduction (20 lines max)

### 1.1 Plan

TODO: Remove the detailed plan

- Start with the end
  - show a pulse
  - explain what is simple / hard
- The situation before
  - pb with volume (MongoDB / PHP, etc...)
  - securities issues
  - pb with abilities
  - angular complexity
  - refactoring issues
  - deployment issues
- The choices
  - why clojure?
  - why Haskell?
  - why not full Haskell?
  - why reagent?
  - why Kafka?
  - why Mesos / Marathon?
  - why Druid?
  - why still MongoDB?
- The firsts weeks
  - first impressions
  - what was harder?

- what was easier?
- Once used to clojure
  - how does it feels?
  - was it a mistake?
  - Do we have any doubts?
- One year later (maintenance and impressions)

## 1.2 The Elephant Graveyard

Imagine you could get all tweets in realtime.

Imagine you need to count them. Imagine you need to filter them by keywords. Imagine you need to answer complex questions about them in realtime. For example, how many tweets from women, containing the word `clojure` expressing a positive sentiment during the last hour. Imagine the same question about the last year.

How would you do it?

First you'll need to absorb tweets in realtime. The twitter streaming API is here for that. But you are limited to 1% of all twitter volume. If you want not to be limited, you need either deal directly with twitter or use GNIP.

Next, you'll need to keep only tweet of interest. By example, you'll need to filter by keyword.

Just after that, you need to add informations for each received tweet. You need to enrich them by information it doesn't already possess. For example, the gender of the author of a tweet must be guessed. The same for the sentiment expressed by the tweet.

In the end you'll need to display all these informations in real-time. By real-time we mean with a very low latency.

Under the minute is generally acceptable. But under the hood, we generally have less than half a second of latency.

Most of the lantency is due to twitter (about 2s) or gnip (about 15s).