PADDLEGURU

THE ROAD TO SUMMINGBIRD

Stream Processing at (Every) Scale

Sam Ritchie :: @sritchie :: Data Day Texas 2014


Σ
@summingbird

https://  /  /summingbird

AGENDA

AGENDA

- Logging and Monitoring in the Small

AGENDA

- Logging and Monitoring in the Small
- Scaling toward Summingbird - Tooling Overview

AGENDA

- Logging and Monitoring in the Small
- Scaling toward Summingbird - Tooling Overview
- What breaks at full scale?

AGENDA

- Logging and Monitoring in the Small
- Scaling toward Summingbird - Tooling Overview
- What breaks at full scale?
- Summingbird's Constraints, how they can help

AGENDA

- Logging and Monitoring in the Small
- Scaling toward Summingbird - Tooling Overview
- What breaks at full scale?
- Summingbird's Constraints, how they can help
- Lessons Learned

WHAT TO MONITOR?

WHAT TO MONITOR?

- Application "Events"

WHAT TO MONITOR?

- Application "Events"
- 

on certain events or patterns

WHAT TO MONITOR?

- Application "Events"
- 

on certain events or patterns
- Extract metrics from the event stream

WHAT TO MONITOR?

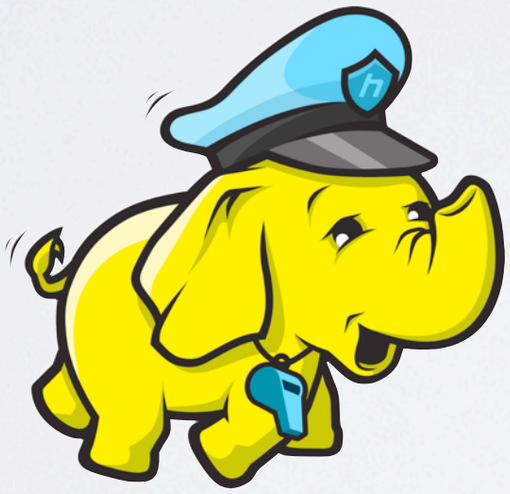
- Application "Events"
- 

on certain events or patterns
- Extract metrics from the event stream
- Dashboards?


I AINT GOING TO KFC
MOTHERFUCKER

PREPPING FOR SCALE

PREPPING FOR SCALE



PREPPING FOR SCALE

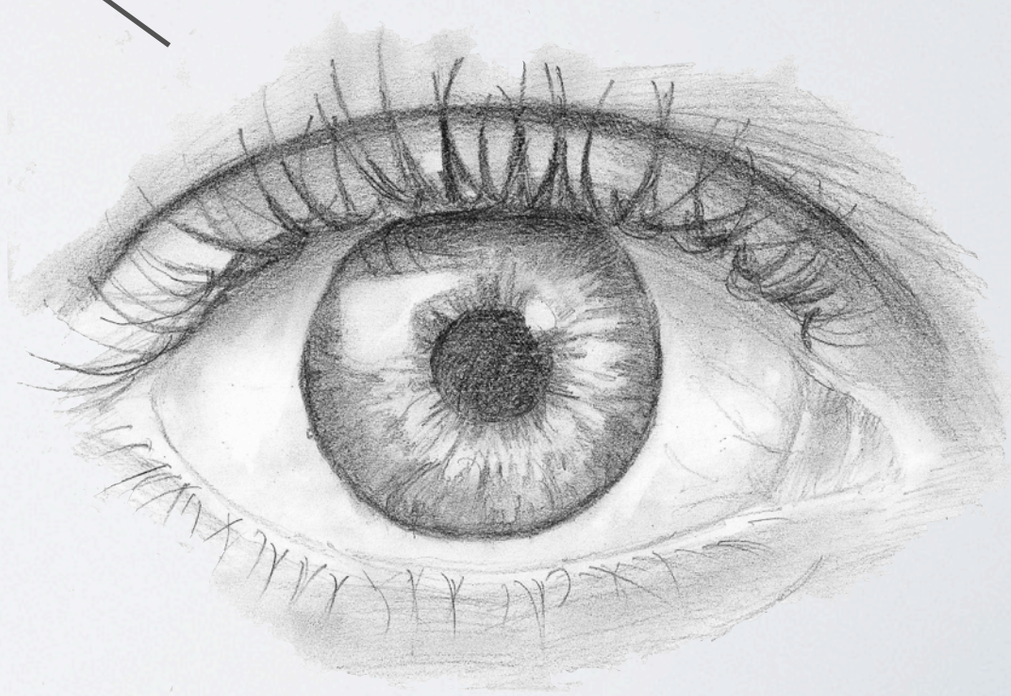
LOG STATEMENTS

```
(defn create-user! [username]
  (log/info "User Created: " username)
  (db/create {:type :user
              :name username
              :timestamp
              (System/currentTimeMillis)}))
```

```
INFO [paddleguru.views.race-page] - AJAX:  {:regatta-title ReturntothePi
er]: at=info method=GET path=/races/ReturntothePier/get-total-cost?eventn
.150.56" dyno=web.2 connect=1ms service=227ms status=200 bytes=8
er]: at=info method=GET path=/races/ReturntothePier/get-total-cost?eventn
50.56" dyno=web.2 connect=5ms service=238ms status=200 bytes=8
er]: at=info method=GET path=/races/ReturntothePier/photos/logo host=padd
s status=200 bytes=41964
er]: at=info method=GET path=/races/2014HanohanoHuki0ceanChallenge/photos
ms service=312ms status=200 bytes=133732
er]: at=info method=GET path=/races/ValentinesRace/photos/logo host=paddl
 status=200 bytes=46053
er]: at=info method=GET path=/races/LazyDogSUPFestandRace/photos/logo hos
e=350ms status=200 bytes=457674
er]: at=info method=GET path=/races/SaintPatricksScramble/photos/logo hos
e=250ms status=200 bytes=46053
er]: at=info method=GET path=/races/YoYoRace/photos/logo host=paddleguru.
s=200 bytes=46053
er]: at=info method=GET path=/athletes/nhickmet host=paddleguru.com reque
s=3452
```

Your App

Heroku Logs



CENTRALIZED LOGGING

loggly

papertrail

```
  fwd="23.241.201.220" dyno=web.2 connect=7ms service=1549ms status=302 bytes=0
Jan 10 18:38:55 web-production app/web.2: INFO [paddleguru.util.mandrill] - Sending email to:  mikedavisconst@hotmail.com
Jan 10 18:38:55 web-production heroku/router: at=info method=GET path=/races/ReturntothePier host=paddleguru.com request_id=1bb7
  dyno=web.2 connect=0ms service=727ms status=200 bytes=6786
Jan 10 18:38:56 web-production heroku/router: at=info method=GET path=/races/ReturntothePier/racers? host=paddleguru.com request_
  fwd="23.241.201.220" dyno=web.1 connect=2ms service=429ms status=200 bytes=2904
Jan 10 18:39:07 web-production heroku/router: at=info method=GET path=/races/ReturntothePier/get-coordinates host=paddleguru.com
  fwd="23.241.201.220" dyno=web.2 connect=1ms service=172ms status=200 bytes=23
Jan 10 18:39:13 web-production heroku/router: at=info method=GET path=/ host=www.paddleguru.com request_id=3cc016423b8ad2dda18a31
  connect=3ms service=5ms status=301 bytes=0
Jan 10 18:39:25 web-production heroku/router: at=info method=GET path=/ host=paddleguru.com request_id=5a76df599cbc9b392fe70a6521
  connect=8ms service=110ms status=200 bytes=8891
Jan 10 18:39:43 web-production heroku/router: at=info method=GET path=/organizers host=paddleguru.com request_id=df20e7ad2648065
  connect=1ms service=165ms status=200 bytes=5332
Jan 10 18:39:46 web-production heroku/router: at=info method=GET path=/races/FloridaCup2014 host=paddleguru.com request_id=45fe58
  dyno=web.1 connect=5ms service=683ms status=200 bytes=4771
Jan 10 18:39:47 web-production heroku/router: at=info method=GET path=/races/FloridaCup2014/sponsors/1 host=paddleguru.com reques
  fwd="72.184.147.95" dyno=web.2 connect=2ms service=346ms status=200 bytes=129955
Jan 10 18:39:47 web-production heroku/router: at=info method=GET path=/races/FloridaCup2014/photos/logo host=paddleguru.com reque
  fwd="72.184.147.95" dyno=web.2 connect=0ms service=330ms status=200 bytes=44662
Jan 10 18:39:47 web-production heroku/router: at=info method=GET path=/races/FloridaCup2014/sponsors/2 host=paddleguru.com reques
  fwd="72.184.147.95" dyno=web.1 connect=2ms service=330ms status=200 bytes=114432
Jan 10 18:39:47 web-production heroku/router: at=info method=GET path=/races host=paddleguru.com request_id=af68dada5cec4f6e1d61
  connect=2ms service=442ms status=200 bytes=10515
Jan 10 18:39:47 web-production heroku/router: at=info method=GET path=/races/FloridaCup2014/sponsors/3 host=paddleguru.com reques
  fwd="72.184.147.95" dyno=web.1 connect=2ms service=227ms status=200 bytes=68862
Jan 10 18:39:48 web-production heroku/router: at=info method=GET path=/races/BobHannaClassic/photos/logo host=paddleguru.com requ
  fwd="74.176.174.36" dyno=web.2 connect=2ms service=259ms status=200 bytes=52976
Jan 10 18:39:48 web-production heroku/router: at=info method=GET path=/races/ValentinesRace/photos/logo host=paddleguru.com reque
  fwd="74.176.174.36" dyno=web.1 connect=2ms service=265ms status=200 bytes=46053
Jan 10 18:39:48 web-production heroku/router: at=info method=GET path=/html/tables.html host=paddleguru.com request_id=c3c0645a98
  dyno=web.2 connect=1ms service=6ms status=200 bytes=158
Jan 10 18:39:48 web-production heroku/router: at=info method=GET path=/html/timing.html host=paddleguru.com request_id=0e63253ea1
  dyno=web.1 connect=2ms service=18ms status=200 bytes=1761
Jan 10 18:39:48 web-production heroku/router: at=info method=GET path=/races/2014HanohanoHuki0ceanChallenge/photos/logo host=pado
  request_id=efdc4d04b1b9ef545a6dfcb252314876 fwd="74.176.174.36" dyno=web.2 connect=0ms service=317ms status=200 bytes=133732
```

Your App

S3

WHAT DO YOU GET?

WHAT DO YOU GET?

- Ability to REACT to system events

WHAT DO YOU GET?

- Ability to REACT to system events
- Long-term storage via S3

WHAT DO YOU GET?

- Ability to REACT to system events
- Long-term storage via S3
- Searchable Logs

WHAT'S MISSING?

WHAT'S MISSING?

- How many users per day?

WHAT'S MISSING?

- How many users per day?
- How many times did this exception show up vs that?

WHAT'S MISSING?

- How many users per day?
- How many times did this exception show up vs that?
- Was this the first time I've seen that error?

WHAT'S MISSING?

- How many users per day?
- How many times did this exception show up vs that?
- Was this the first time I've seen that error?
- Pattern Analysis requires Aggregations

STRUCTURED LOGGING

IMPOSE STRUCTURE

```
(log/info "User Created: " username)
```

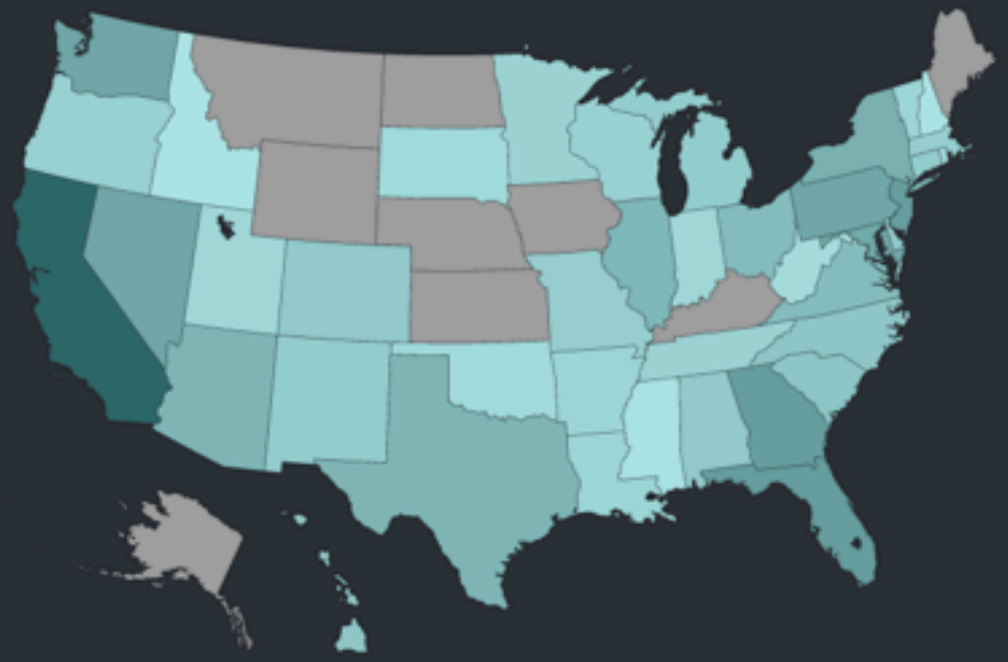
```
(log/info {:event "user_creation"
           :name "sritchie"
           :timestamp (now)
           :request-id request-id})
```

Kibana

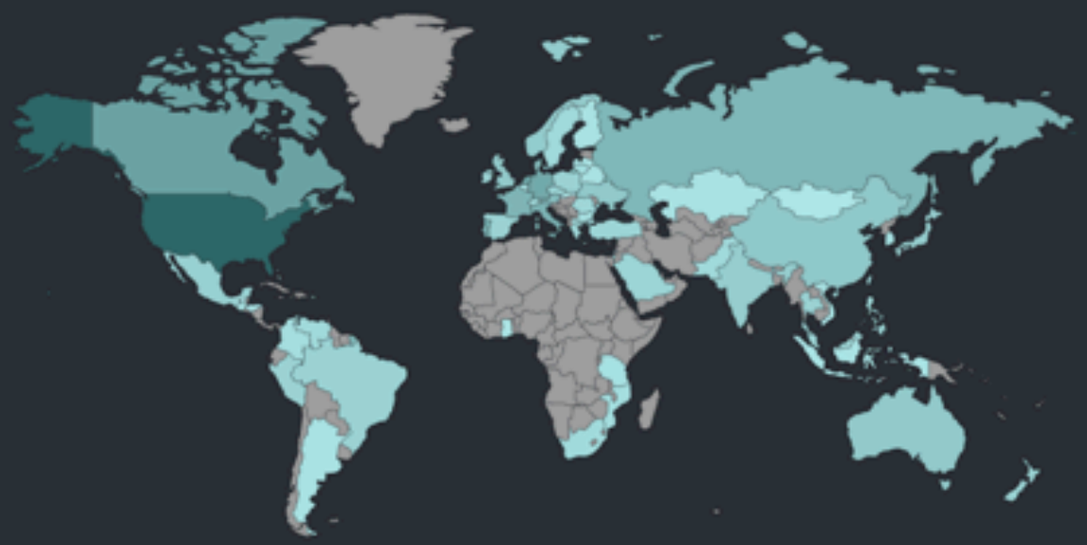
elasticsearch.

FILTERING

US HITS

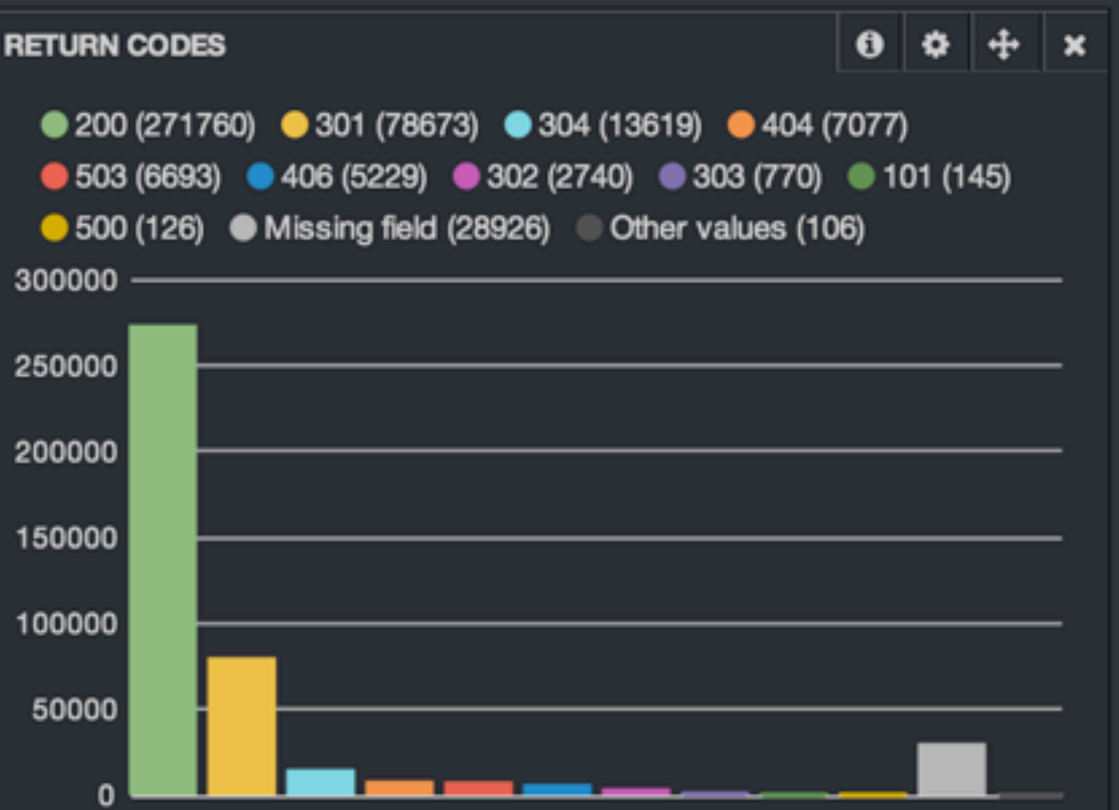


WORLD HITS

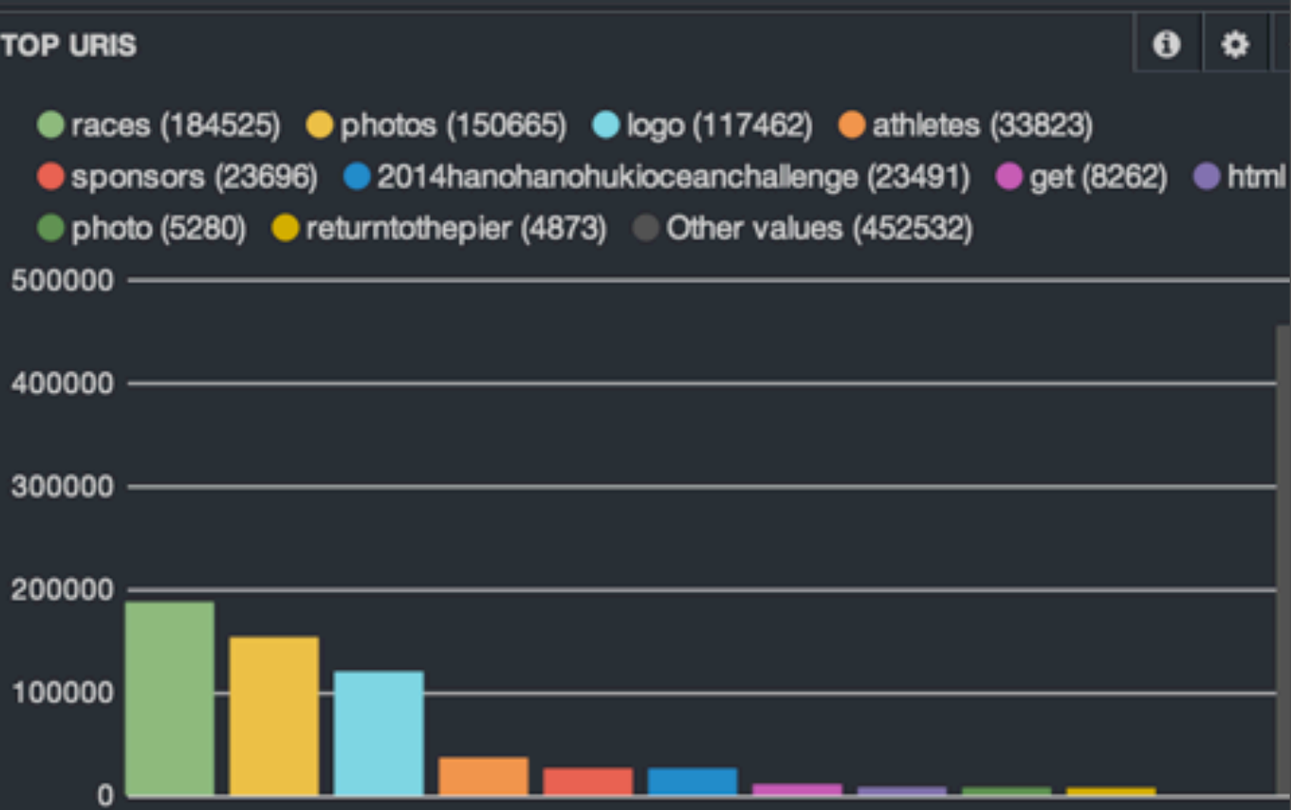


LATL

RETURN CODES



TOP URIS

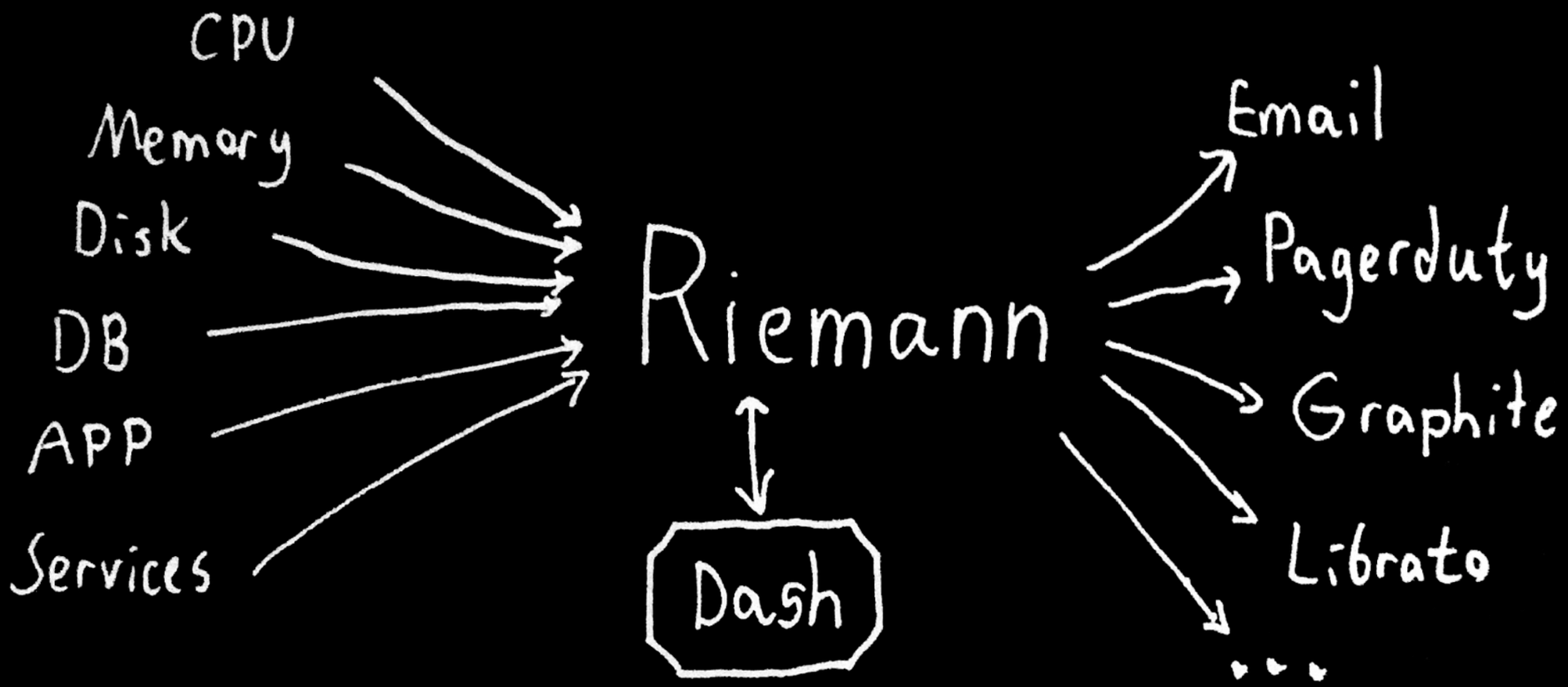


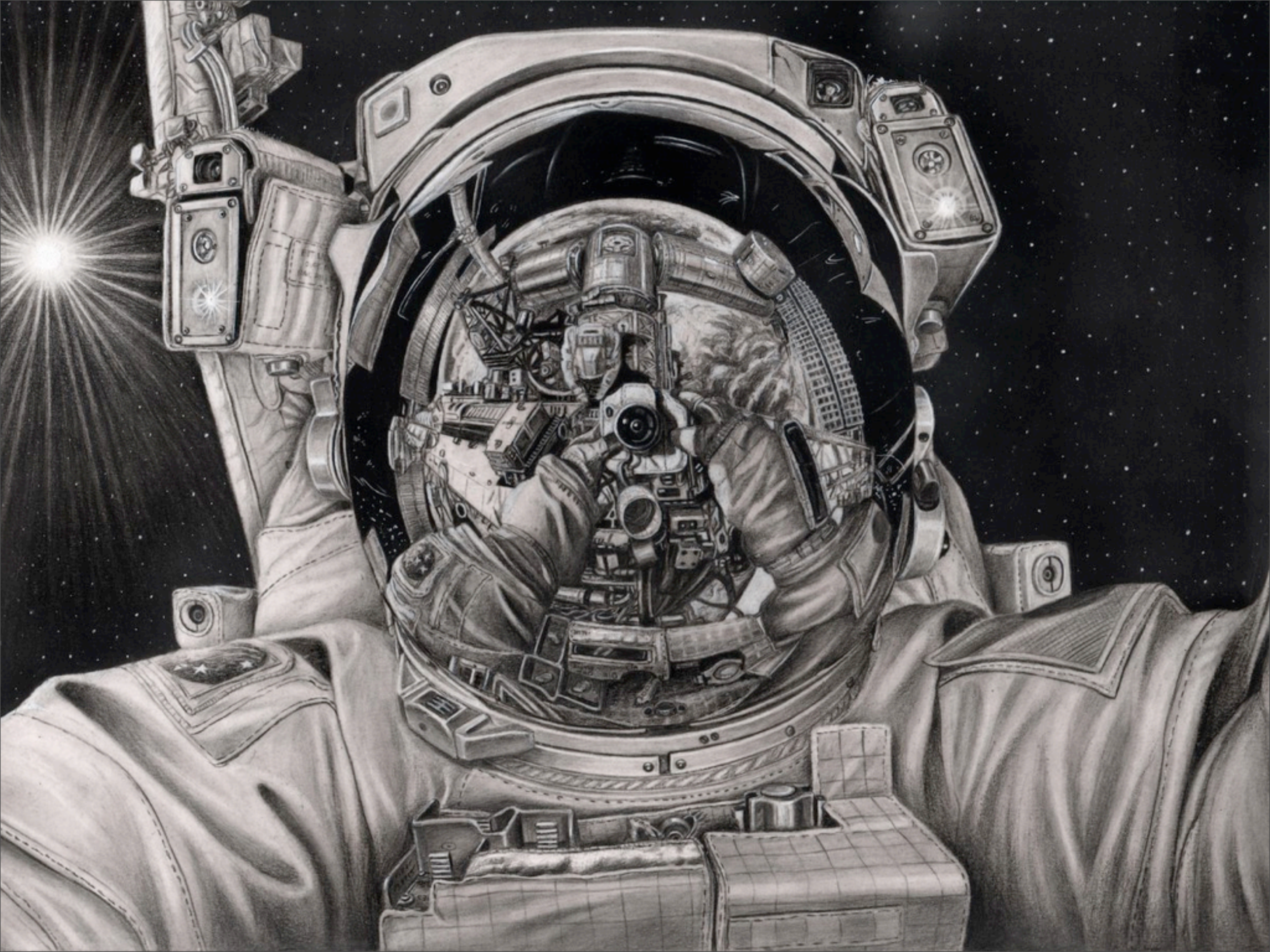
DOCUMENTS

Fields

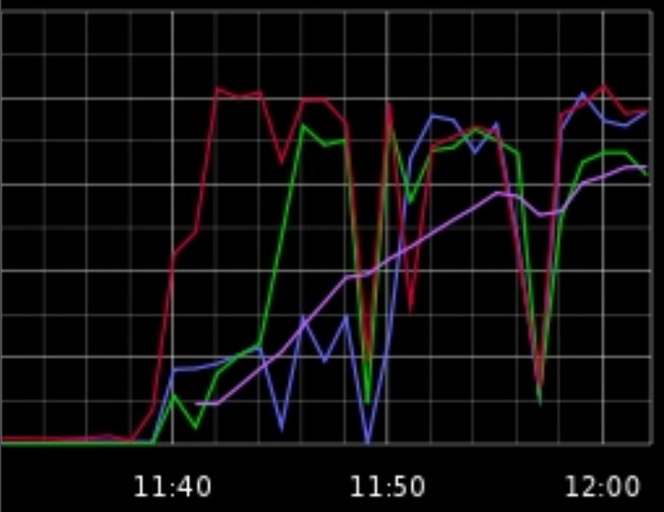
0 to **100** of 500 available for paging

EVENT PROCESSING

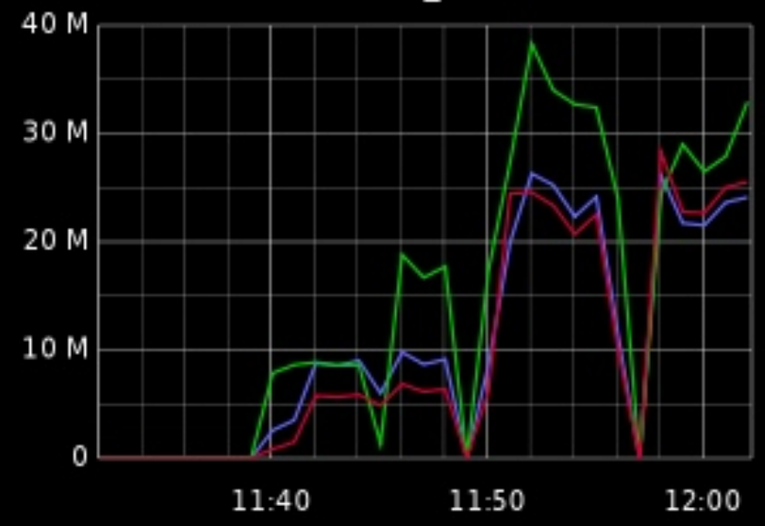

CPU
Memory
Disk
DB
APP
Services
Riemann
Dash
Email
Pagerduty
Graphite
Librato
...



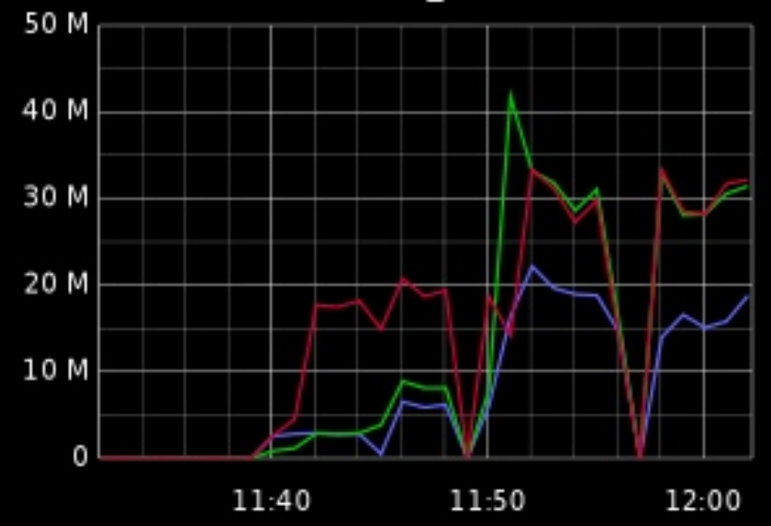
CPU

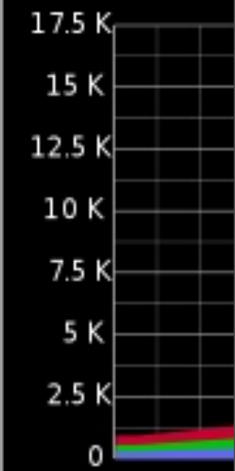


Network_RX

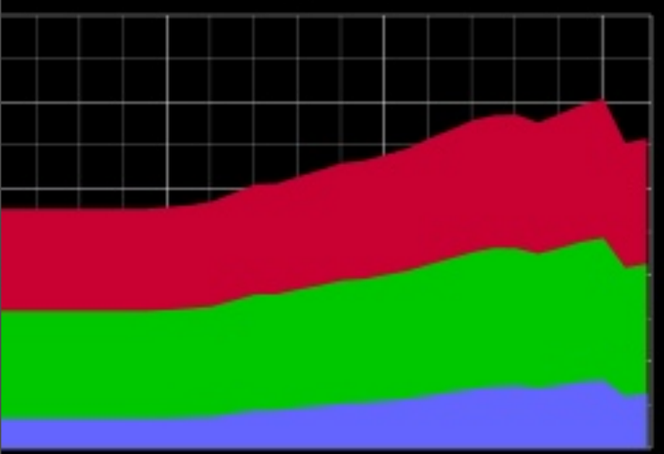


Network_TX

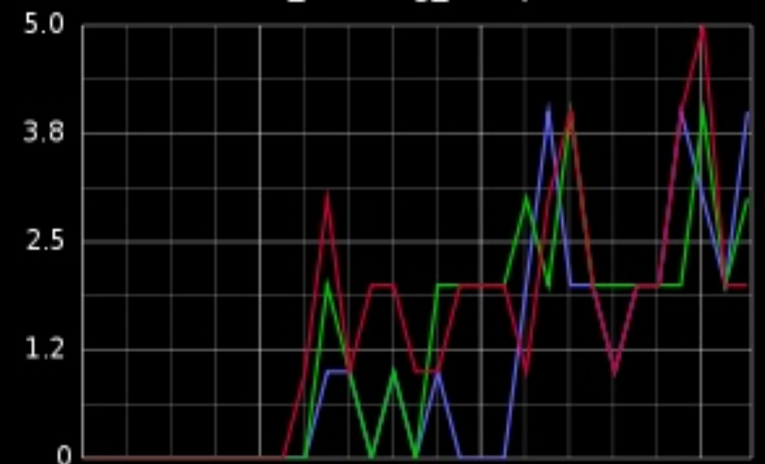




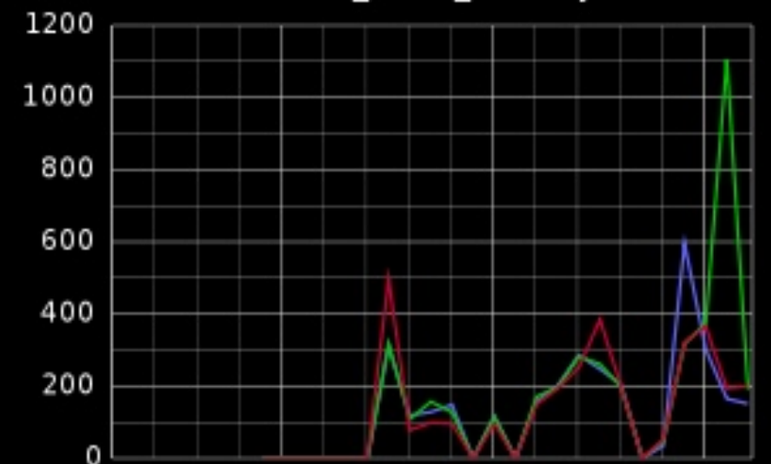
Cassandra_Total_Disk

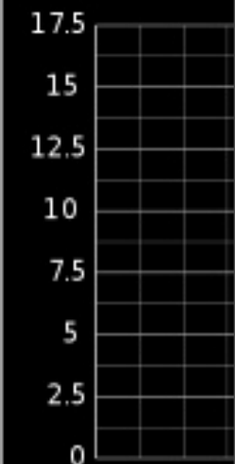


Cassandra_Pending_Compactions

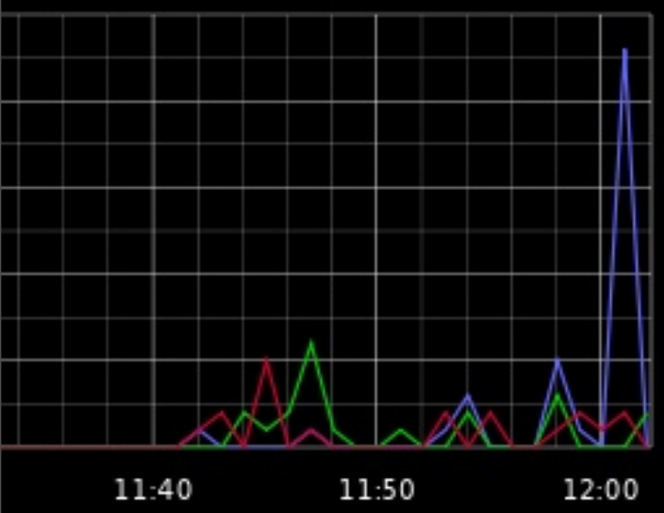


Cassandra_Read_Latency

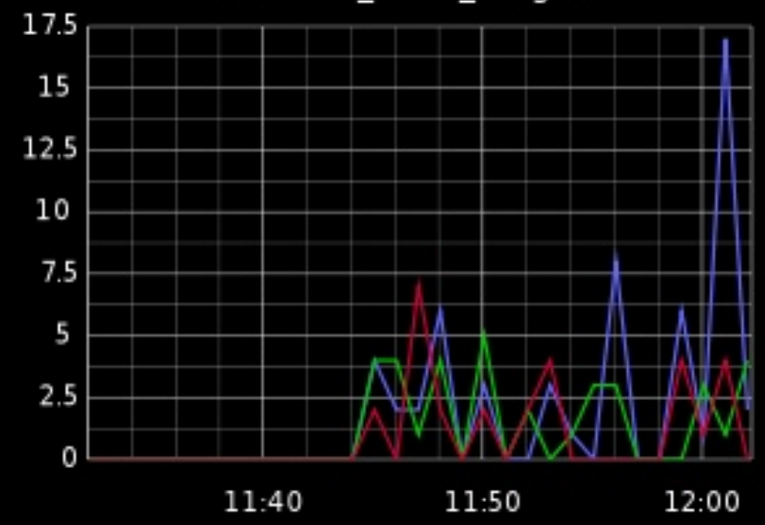




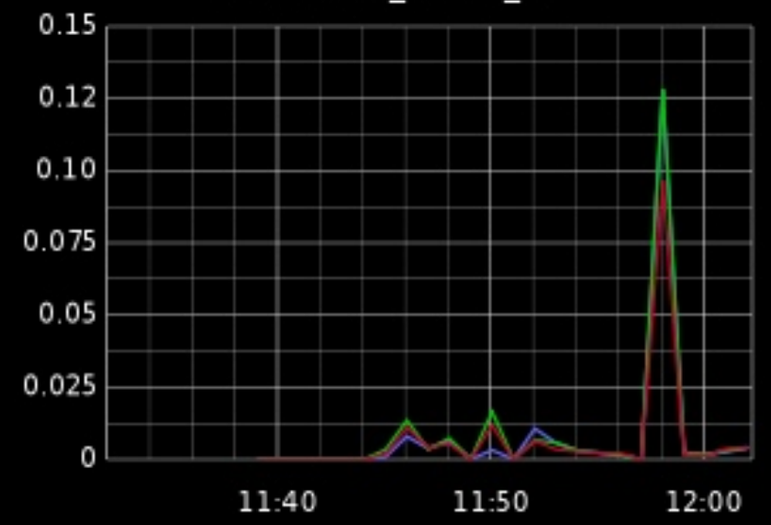
Cassandra_Mutation_Stages

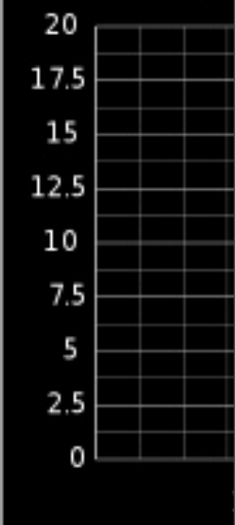


Cassandra_Read_Stages



Cassandra_Bloom_FP

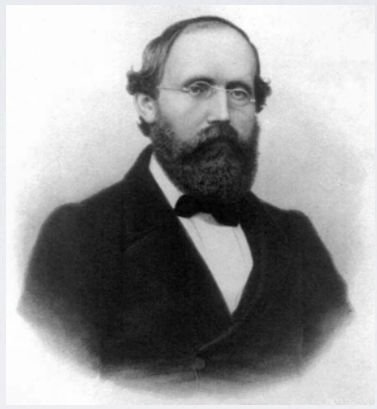




Your App

S3

Github



Mandrill

elasticsearch.

Kibana

EVENT PROCESSORS

EVENT PROCESSORS

- FluentD (http://fluentd.org/)

EVENT PROCESSORS

- FluentD (http://fluentd.org/)
- Riemann (http://riemann.io/)

EVENT PROCESSORS

- FluentD (http://fluentd.org/)
- Riemann (http://riemann.io/)
- Splunk (http://www.splunk.com/)

EVENT PROCESSORS

- FluentD (http://fluentd.org/)
- Riemann (http://riemann.io/)
- Splunk (http://www.splunk.com/)
- Simmer (https://github.com/avibryant/simmer)

EVENT PROCESSORS

- FluentD (http://fluentd.org/)
- Riemann (http://riemann.io/)
- Splunk (http://www.splunk.com/)
- Simmer (https://github.com/avibryant/simmer)
- StatsD + CollectD (https://github.com/etsy/statsd/)

EVENT PROCESSORS

- FluentD (http://fluentd.org/)
- Riemann (http://riemann.io/)
- Splunk (http://www.splunk.com/)
- Simmer (https://github.com/avibryant/simmer)
- StatsD + CollectD (https://github.com/etsy/statsd/)
- Esper (http://esper.codehaus.org/)


I AINT GOING TO KFC
MOTHERFUCKER

What Breaks at Scale?

SERIALIZATION

SERIALIZATION

- Thrift (http://thrift.apache.org/)

SERIALIZATION

- Thrift (http://thrift.apache.org/)
- Protocol Buffers (https://code.google.com/p/protobuf/)

SERIALIZATION

- Thrift (http://thrift.apache.org/)
- Protocol Buffers (https://code.google.com/p/protobuf/)
- Avro (http://avro.apache.org/)

SERIALIZATION

- Thrift (http://thrift.apache.org/)
- Protocol Buffers (https://code.google.com/p/protobuf/)
- Avro (http://avro.apache.org/)
- ~~Kryo (https://github.com/EsotericSoftware/kryo)~~

LOG COLLECTION

LOG COLLECTION

- Kafka (https://kafka.apache.org/)

LOG COLLECTION

- Kafka (https://kafka.apache.org/)
- LogStash (http://logstash.net/)

LOG COLLECTION

- Kafka (https://kafka.apache.org/)
- LogStash (http://logstash.net/)
- Flume (http://flume.apache.org/)

LOG COLLECTION

- Kafka (https://kafka.apache.org/)
- LogStash (http://logstash.net/)
- Flume (http://flume.apache.org/)
- Kinesis (http://aws.amazon.com/kinesis/)

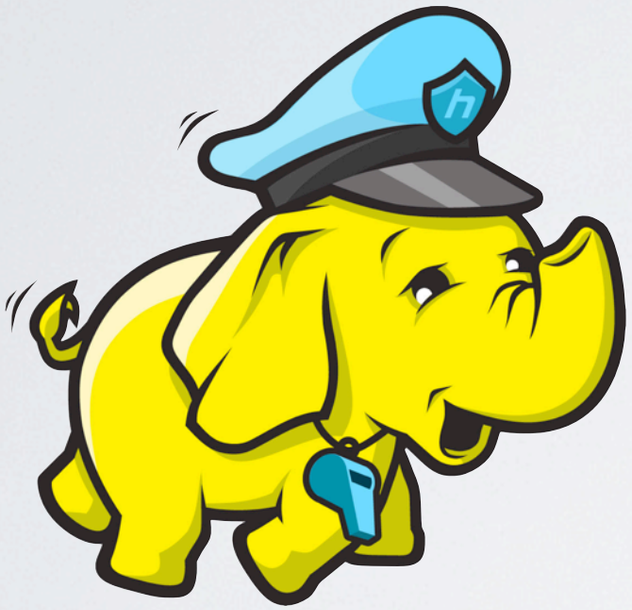
LOG COLLECTION

- Kafka (https://kafka.apache.org/)
- LogStash (http://logstash.net/)
- Flume (http://flume.apache.org/)
- Kinesis (http://aws.amazon.com/kinesis/)
- Scribe (https://github.com/facebook/scribe)

EVENT PROCESSING

Spark

Lightning-Fast Cluster Computing



akka


Σ
@summingbird

What is Summingbird?

- Declarative Streaming Map/Reduce DSL
- Realtime platform that runs on Storm.
- Batch platform that runs on Hadoop.
- Batch / Realtime Hybrid platform


ONE DOES NOT SIMPLY
"USE STORM"

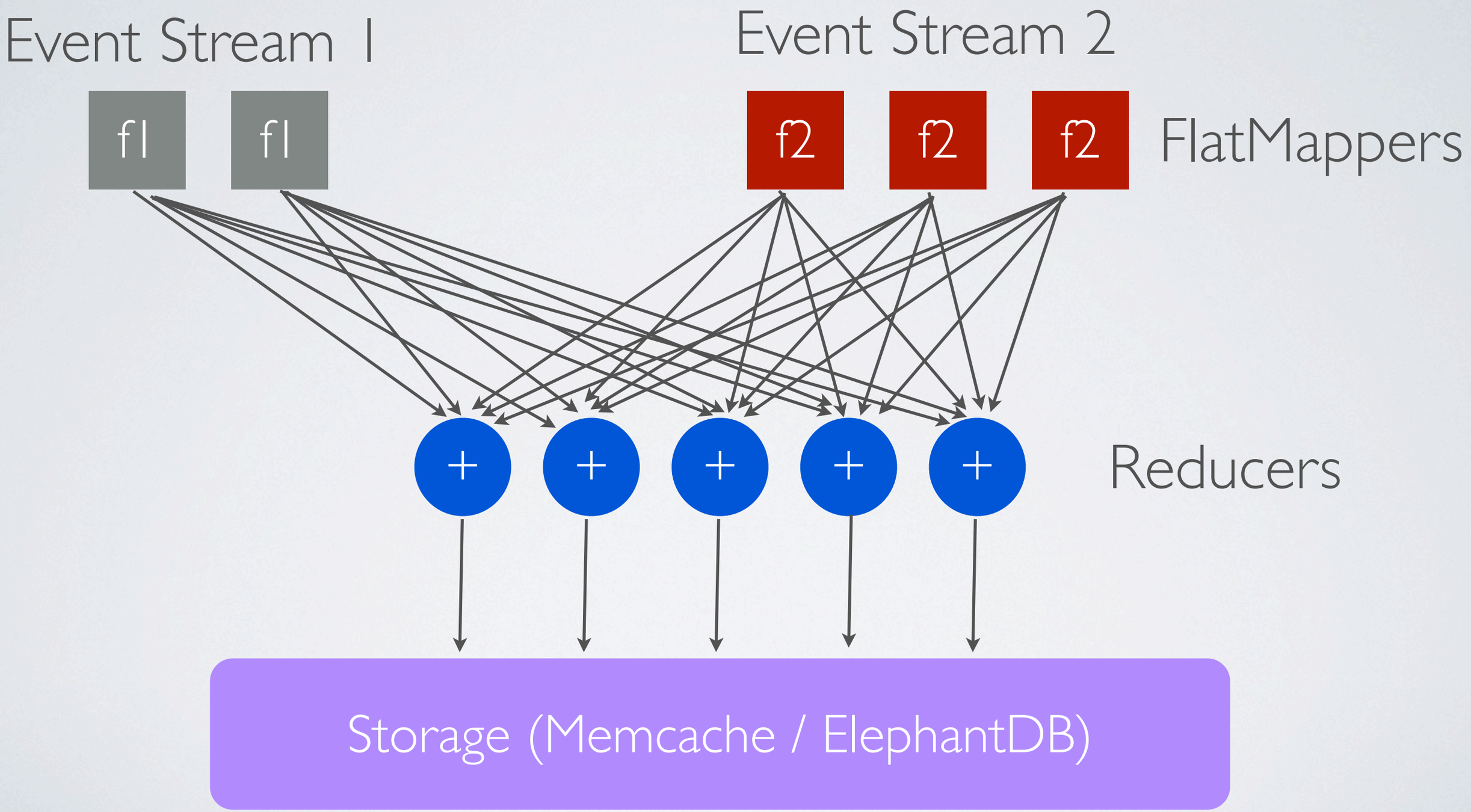
```
val impressionCounts =
  impressionHose.flatMap(extractCounts(_))

val engagementCounts =
  engagementHose.filter(_.isValid)
    .flatMap(engagementCounts(_))

val totalCounts =
  (impressionCounts ++ engagementCounts)
    .flatMap(fanoutByTime(_))
    .sumByKey(onlineStore)

val stormTopology =
  Storm.remote("stormName").plan(totalCounts)
val hadoopJob =
  Scalding("scaldingName").plan(totalCounts)
```

MAP/REDUCE


Event Stream 1
Event Stream 2
f1
f1
f2
f2
f2
FlatMappers
+
+
+
+
+
Reducers
Storage (Memcache / ElephantDB)

- Source[+T]
- Service[-K, +V]

- Store[-K, V]
- Sink[-T]

The Four Ss!

- `Source[+T]`
- `Service[-K, +V]`

- `Store[-K, V]`
- `Sink[-T]`

Store[-K, V]:

What values are allowed?

```
trait Monoid[T] {
  def zero: T
  def plus(l: T, r: T): T
}
```

Tons O'Monoids:

CMS,
HyperLogLog,
ExponentialMA,
BloomFilter,
Moments,
MinHash,
TopK

https://github.com/twitter/algebird/tree/develop/algebird-core/src/main/scala/com/twitter/alg

d Libra | RB | My Delicious | SavePublishing | Bookmark on Delicio | Add to Wish List

algebird / algebird-core / src / main / scala / com / twitter / **algebird** /

Merge pull request #136 from ccsevers/add_foldM ···

johnynek authored 3 days ago

..

mutable	a month ago	Adds priority queue aggregator [johnyne
AdjoinedUnitRing.scala	18 days ago	Cleans up intTimes [johnynek]
AffineFunction.scala	a month ago	test [sritchie]
Aggregator.scala	a month ago	test [sritchie]
Approximate.scala	a month ago	test [sritchie]
AveragedValue.scala	a month ago	test [sritchie]
BloomFilter.scala	a month ago	test [sritchie]
CountMinSketch.scala	3 days ago	Hotfix for CMS [johnynek]
DecayedValue.scala	a month ago	test [sritchie]
DecayedVector.scala	a month ago	test [sritchie]
Eventually.scala	a month ago	add eventually [sritchie]
Field.scala	a month ago	test [sritchie]
GeneratedAbstractAlgebra.scala	a month ago	test [sritchie]
Group.scala	18 days ago	Cleans up intTimes [johnynek]
HyperLogLog.scala	a month ago	test [sritchie]
IndexedSeq.scala	a month ago	test [sritchie]
JavaMonoids.scala	a month ago	test [sritchie]
MapAlgebra.scala	17 days ago	Adds a comment (to restart travis) [johny
Metric.scala	a month ago	test [sritchie]
MinHasher.scala	a month ago	test [sritchie]


Did you just tell me to go fuck myself?
I believe I did, Bob.

Algebird at Scale

Kurt Mehlhorn
Peter Sanders

Algorithms and Data Structures

The Basic Toolbox

Springer

MONOID COMPOSITION

```
// Views per URL Tweeted
(URL, Int)
```

```
// Views per URL Tweeted
(URL, Int)

// Unique Users per URL Tweeted
(URL, Set[UserID])
```

```
// Views per URL Tweeted
(URL, Int)

// Unique Users per URL Tweeted
(URL, Set[UserID])
```

```
// Views AND Unique Users per URL
(URL, (Int, Set[UserID]))
```

```
// Views per URL Tweeted
(URL, Int)

// Unique Users per URL Tweeted
(URL, Set[UserID])

// Views AND Unique Users per URL
(URL, (Int, Set[UserID]))

// Views, Unique Users + Top-K Users
(URL, (Int, Set[UserID], TopK[(User, Count)]))
```

ASSOCIATIVITY

```
;; 7 steps
```

a0 + a1 + a2 + a3 + a4 + a5 + a6 + a7

```
;; 7 steps
(+ a0 a1 a2 a3 a4 a5 a6 a7)
```

```
;; 5 steps
(+ (+ a0 a1)
   (+ a2 a3)
   (+ a4 a5)
   (+ a6 a7))
```

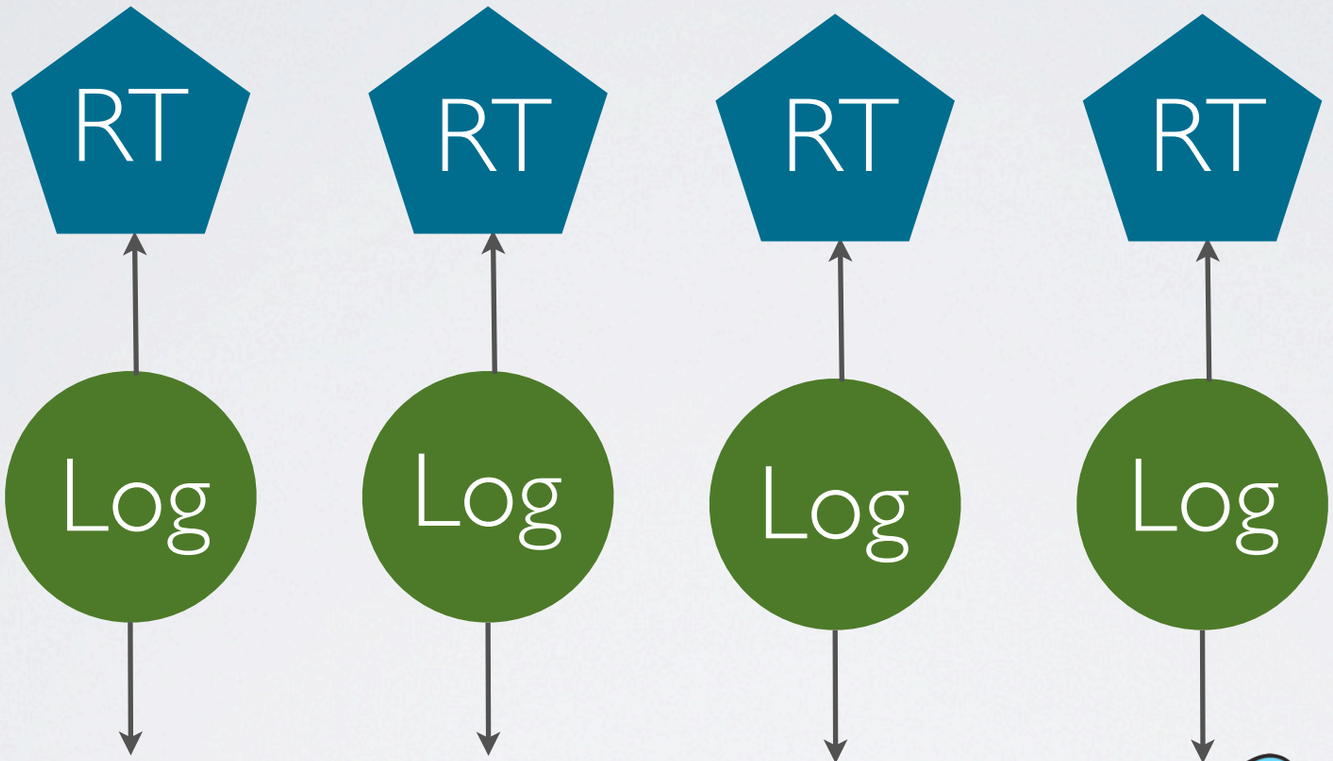
```
;; 3 steps
(+ (+ (+ a0 a1)
      (+ a2 a3))
   (+ (+ a4 a5)
      (+ a6 a7)))
```

PARALLELISM

ASSOCIATIVITY

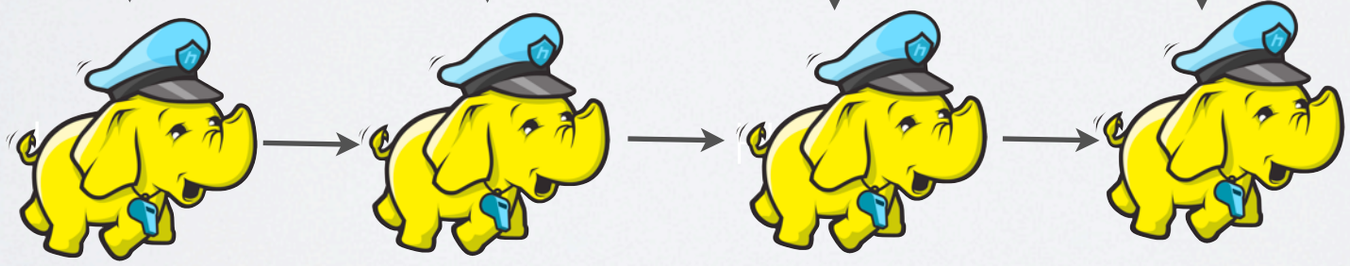
BATCH / REALTIME

Noisy:



Realtime sums from 0, each batch

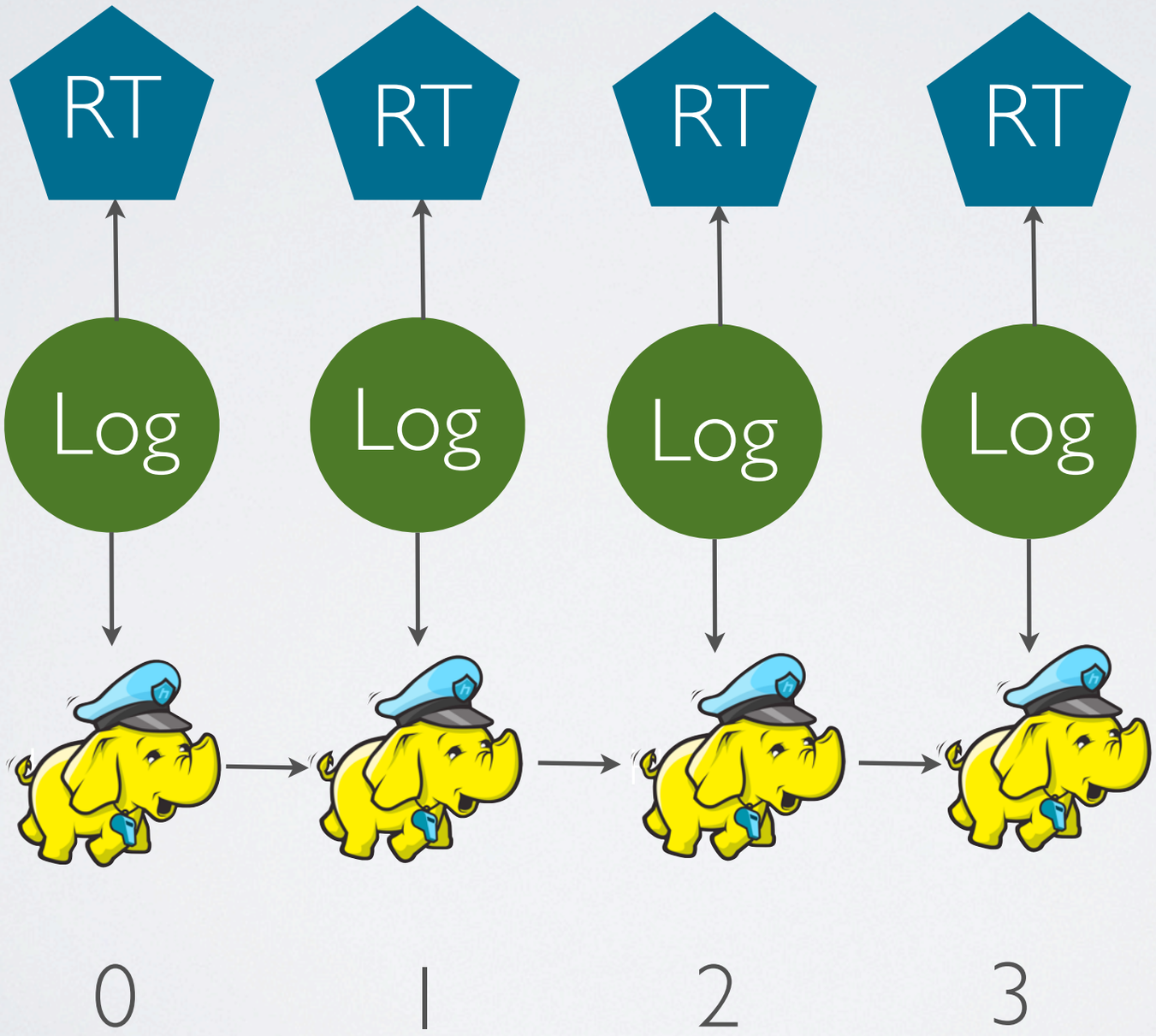
fault tolerant:



BatchID: 0 1 2 3

BATCH / REALTIME

Noisy:



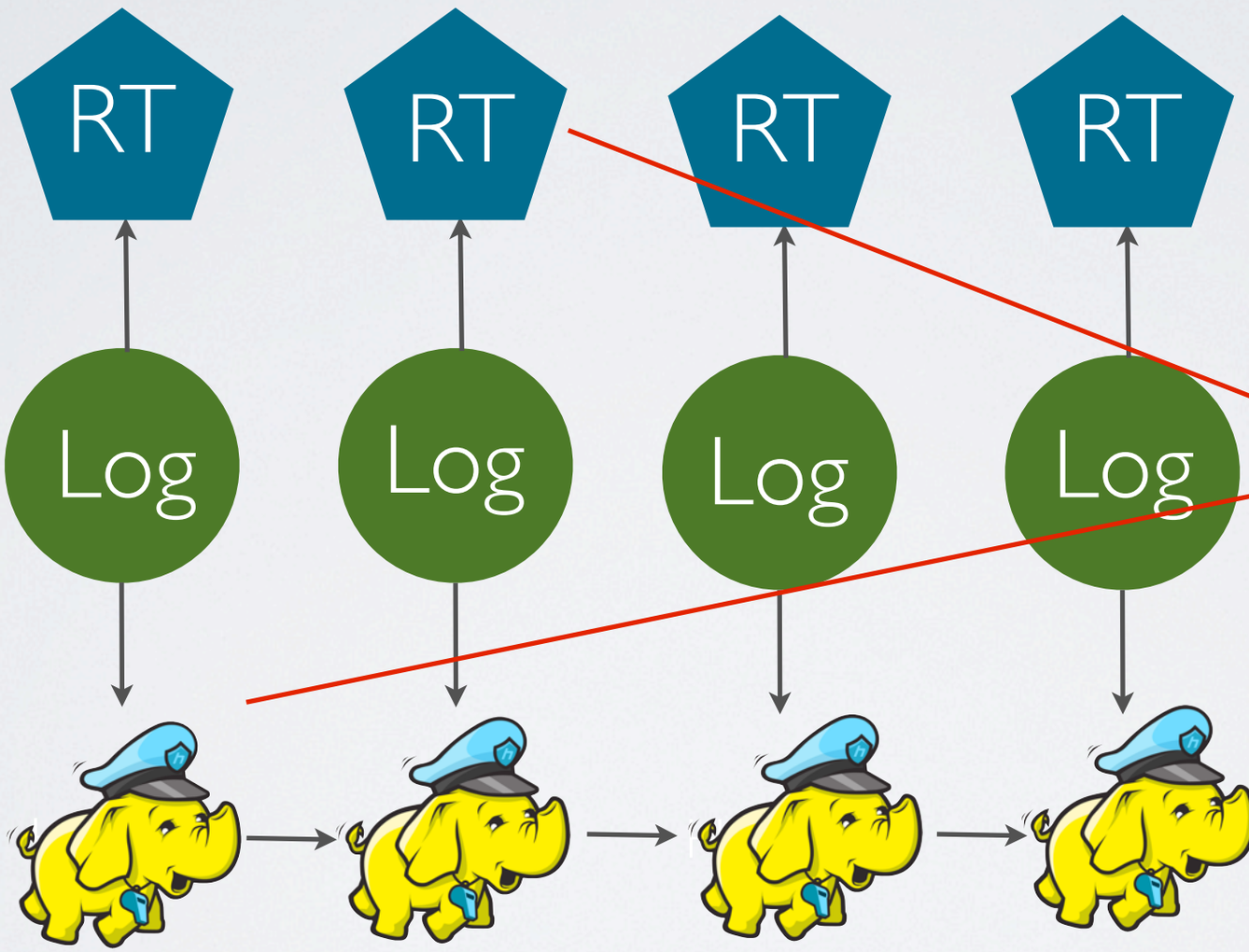
fault tolerant:

Hadoop keeps a total sum (reliably)

BatchID: 0 1 2 3

BATCH / REALTIME

Noisy:



Sum of RT Batch(i) + Hadoop Batch(i-1) has bounded noise, bounded read/write size

fault tolerant:

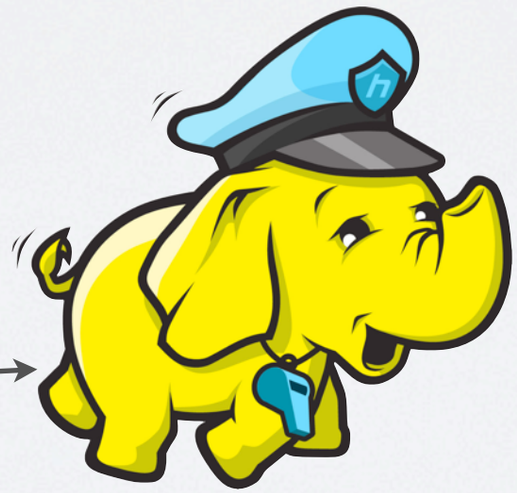
BatchID: 0 1 2 3

Your App

Github

Mandrill

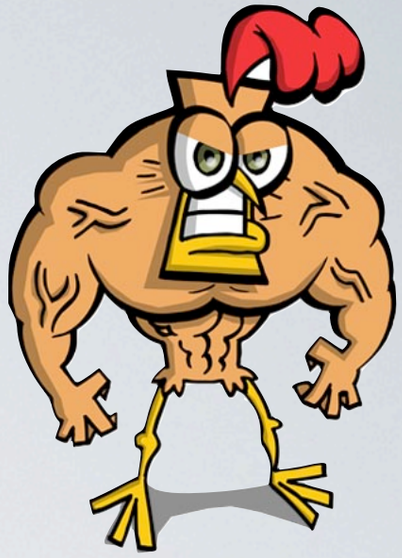
S3



Memcached

ElephantDB

Clients



Four more years. pic.twitter.com/bAJE6Vom

8:16 PM - 6 Nov 2012

800,833 RETWEETS 299,776 FAVORITES

TWEET EMBED COUNTS

How to Embed a Tweet on your Website

Every Tweet on twitter.com and Tweetdeck has a set of Tweet actions at the bottom, including Reply, Retweet, Favorite, and More. Click the "More" Tweet action and select "Embed Tweet":

Popular Latest

Acquires Mobile Advertising
ge MoPub

hn Is Excited About Twitter.

CEO Talks About How To Be
r

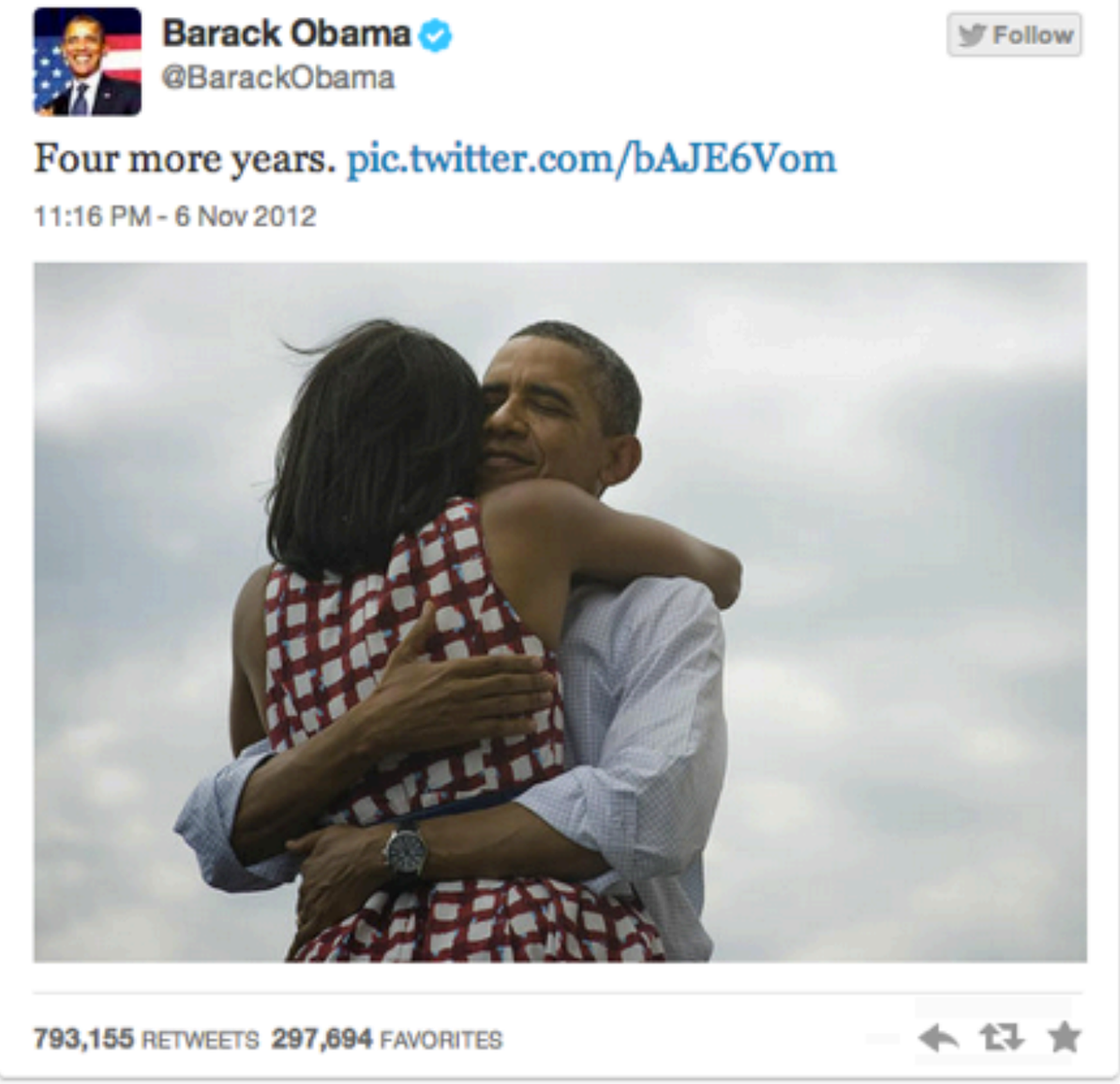
low Lets You Access ALL
ITTERS

ur Twitter Profile May be an

Amplify Signs Its First Video
ship in Europe

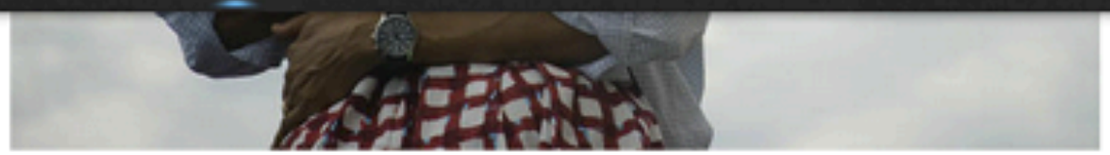
Load More

This one pretty much speaks for itself. A simple message that acknowledges what everyone reading the tweet will already know.



President Obama's message became the most shared tweet of all time within a day of it going out.

9. Bill Gates' tribute to Steve Jobs



793,155 RETWEETS **297,694** FAVORITES

8:16 PM - 6 Nov 12 Flag media

Related headlines

Overview Embedded Tweets make it possible for you take …
Twitter API @twitterapi

Key statistics, most-followed users, the most popular twee…
Yahoo News @YahooNews

Nach „Hashtake"-Versprecher von ZDF-Reporter: Welche Be…
BILD @BILD

Du premier tweet de Jack Dorsey à celui du pape François, …
Nouvel Observateur @LeNouvelObs

Twitter swelled to 200 million monthly active users in 2012. …
Mashable @mashable

Twitter is synonymous with many things. It's where people …
The Next Web @TheNextWeb

Au terme d'une longue campagne, Barack Obama a été réél…
Le Monde @lemondefr

Andy Murray's big Wimbledon win made history in sports a…
GigaOM @gigaom

It may already be social media's most-shared image ever --…
CNN @CNN

What is the most tweeted about event of all time? If you we…
Metro @MetroUK

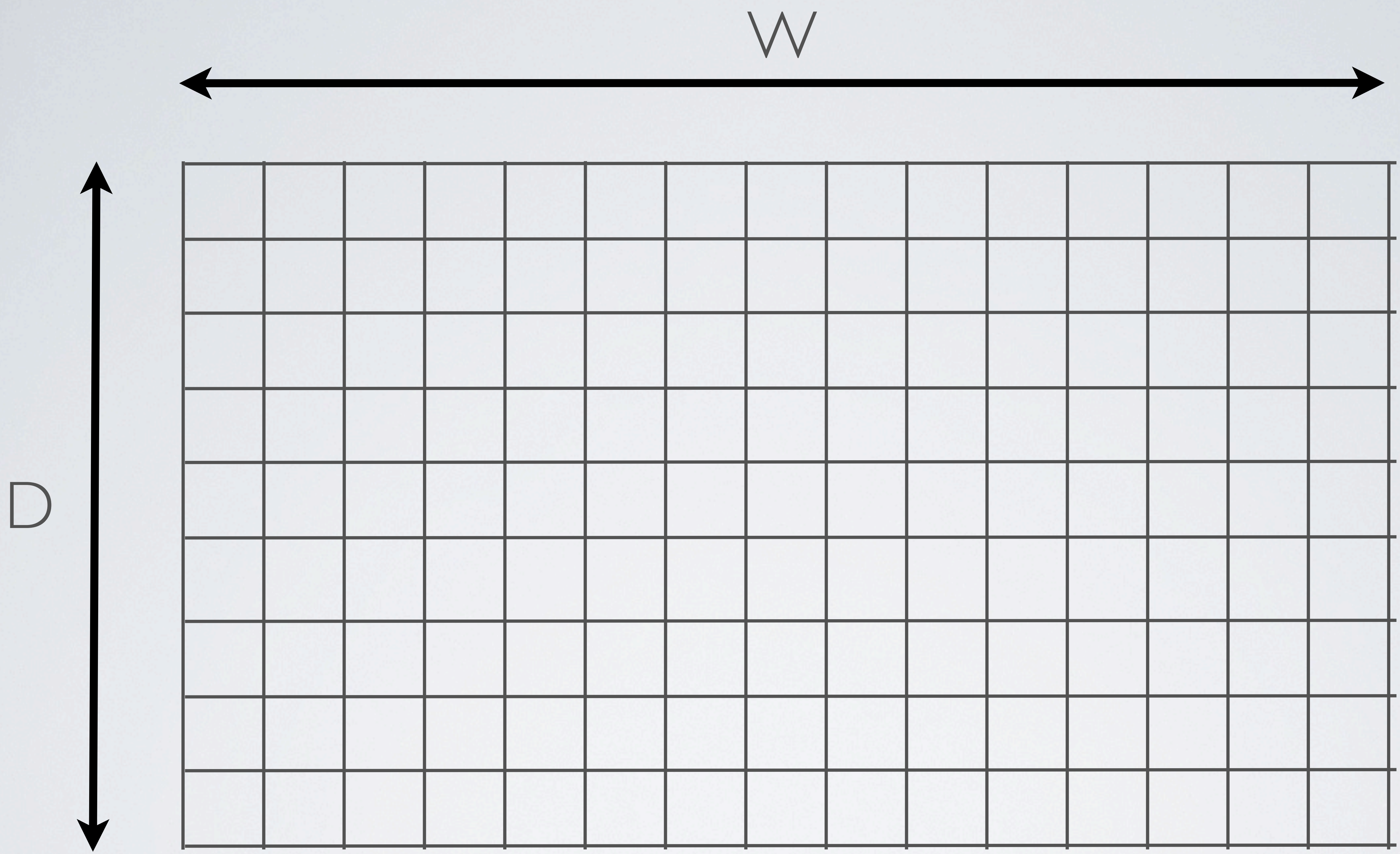
HOLLANDE - On a beau dire, ces petits instants volés, dans…
Le HuffPost @LeHuffPost

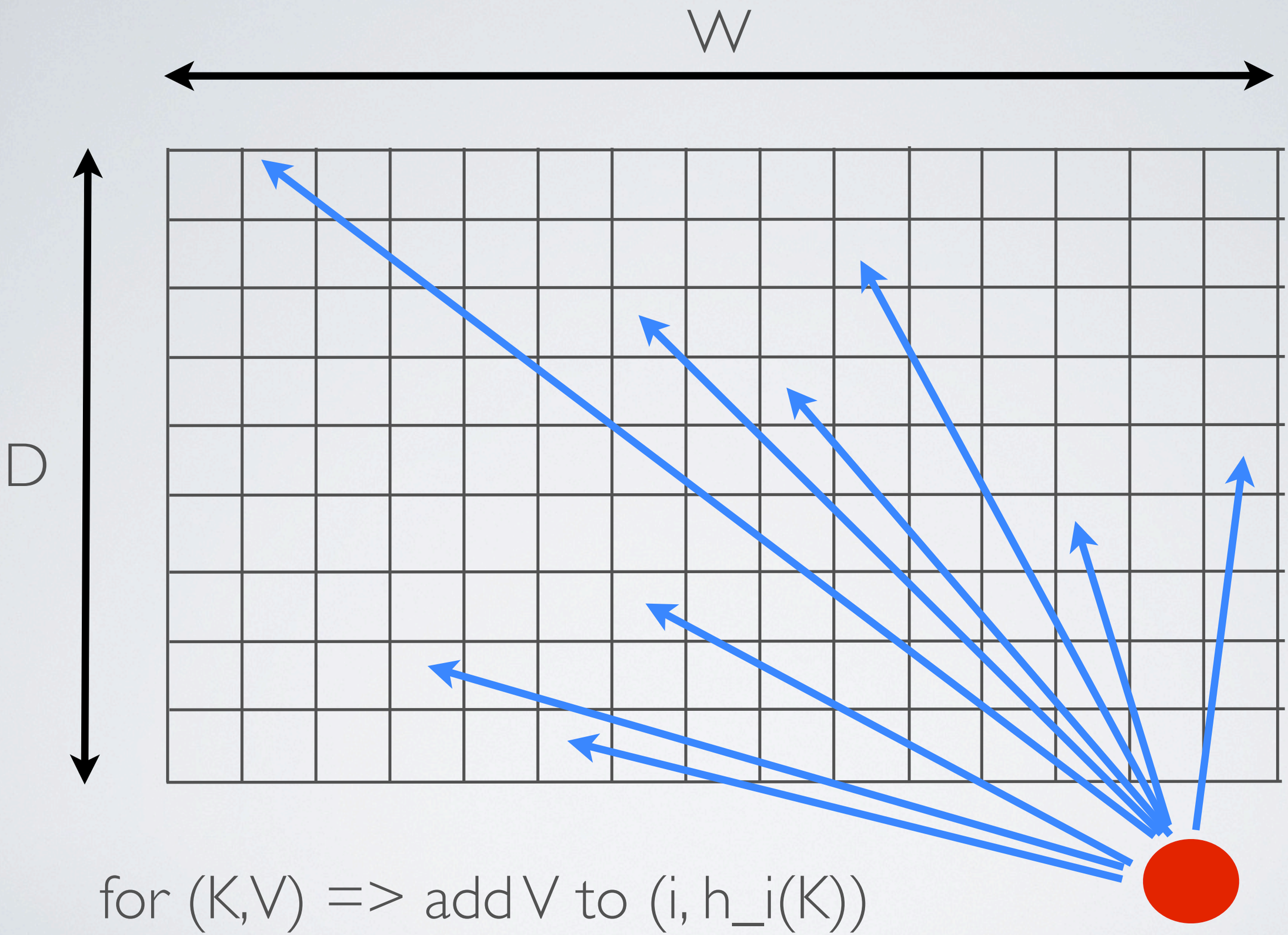
In 55 wide-ranging lists, TIME surveys the highs and lows, t…

Approximate Maps

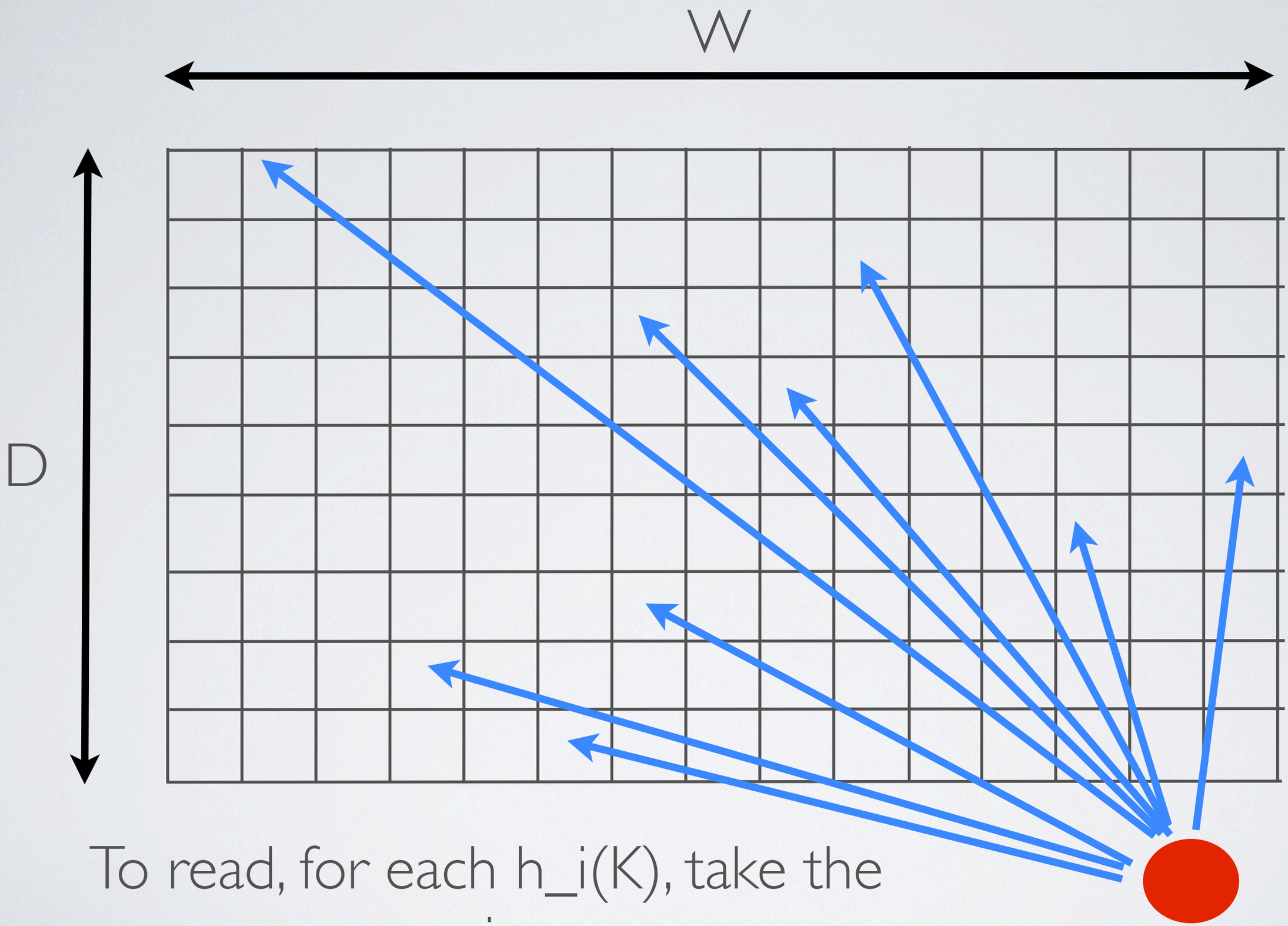
- We would probably be okay if for each Key we could get an **approximate** Value.

- We might not need to enumerate all resulting keys; perhaps only keys with large values would do.


W
D



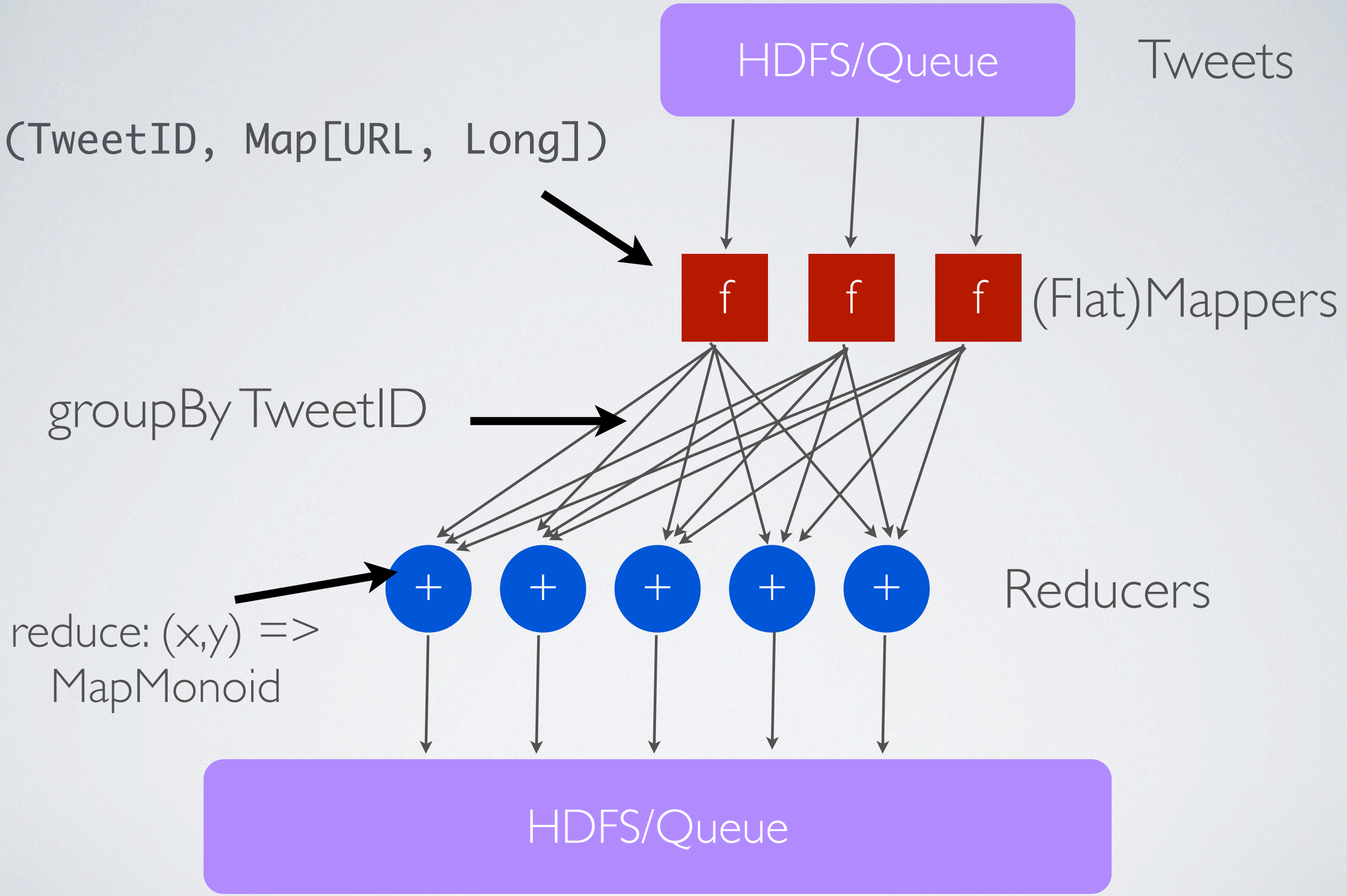
for (K,V) => add V to (i, h_i(K))



To read, for each h_i(K), take the min.

Count-Min Sketch is an Approximate Map

- Each K is hashed to d values from [0 to w-1]
- sum into those buckets
- Result is min of all buckets.
- Result is an upper bound on true value.
- With prob > (1 - delta), error is at most eps * Total Count
- w = 1 / eps, d = log(1/delta)
- total cost in memory O(w * d)


HDFS/Queue
Tweets
(TweetID, Map[URL, Long])
f
f
f
(Flat)Mappers
groupBy TweetID
+
+
+
+
+
Reducers
reduce: (x,y) =>
MapMonoid
HDFS/Queue

Brief Explanation

This job creates two types of keys:

1: ((TweetId, TimeBucket) => CMS[URL, Impressions])
2: TimeBucket => CMS[TweetId, Impressions]

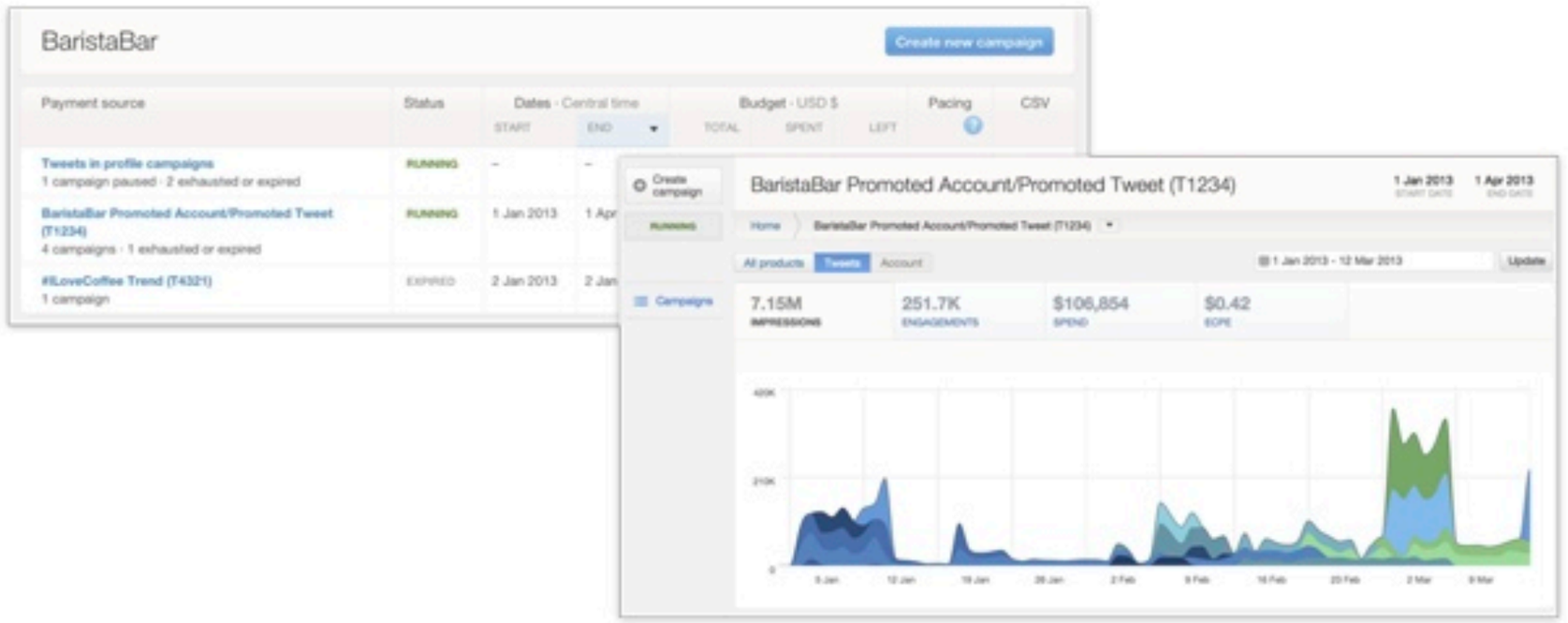
WHAT ELSE?

Twitter Advertising

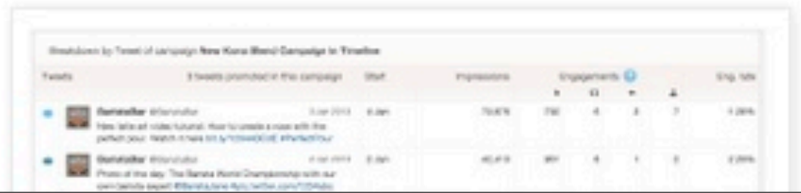
WEDNESDAY, MARCH 13, 2013

The new Twitter Ads center

Today, we're excited to share some changes we've made to the Twitter Ads center. Based on feedback from our advertisers, we've created a revamped experience that improves campaign reporting, provides more visibility into campaign performance analytics and spend, and also makes it easier to manage campaigns in real time.



A major focus of ours is improving campaign analytics. With this in mind, we are now reporting all engagements that Promote Tweets receive — not just engagements that advertisers pay for, but earned media as well. This change gives marketers more complete insight into the impact Promoted Tweets have in driving engagement and exposure on Twitter.



WHAT'S NEXT?

Future Plans

- Akka, Spark, Tez Platforms
- More Monoids
- Pluggable graph optimizations
- Auto-tuning Realtime Topologies

TAKEAWAYS

TAKEAWAYS

- Scale - Fake it ‘til you Make It

TAKEAWAYS

- Scale - Fake it ‘til you Make It
- Structured Logging

TAKEAWAYS

- Scale - Fake it ‘til you Make It
- Structured Logging
- Include timestamps EVERYWHERE

TAKEAWAYS

- Scale - Fake it ‘til you Make It
- Structured Logging
- Include timestamps EVERYWHERE
- Record your Schemas

Questions?

Sam Ritchie :: @sritchie :: Data Day Texas 2014