# matriposo

# Nov 7, Douwe Osinga, @dosinga







# What is what is





# Smart Travel Guides

# Algorithm based.





I PET D	Bodl
Name of Street, or other Designation of the local division of the	_





# Covering the entire world









# Suggestions





# Nearby

# Start with the web



dmoz



facebook.











# Put it back together



# Push it to the users







# Stuff nearby





# Weather based

# Weather data





# Usage on a sunny day





# Pictures on a rainy day







# Weather suggestions





![](_page_17_Picture_2.jpeg)

# Time based

# Users keep time

![](_page_18_Picture_1.jpeg)

![](_page_18_Picture_2.jpeg)

![](_page_18_Picture_3.jpeg)

# Spider the web at large

![](_page_19_Picture_1.jpeg)

![](_page_19_Picture_2.jpeg)

# Opinion mining

![](_page_20_Picture_1.jpeg)

![](_page_20_Picture_2.jpeg)

![](_page_21_Picture_0.jpeg)

![](_page_21_Picture_2.jpeg)

# Time based

![](_page_22_Picture_0.jpeg)

![](_page_22_Picture_1.jpeg)

## Done!

![](_page_23_Picture_0.jpeg)

### adosinga

![](_page_24_Picture_0.jpeg)

### GameAnalytics Building data-intensive services (aka. immutability and idempotence)

@knutin GameAnalytics

![](_page_25_Figure_0.jpeg)

Instrument your game to send events on user action, such as log in, purchase, level up etc.

with UI.

Analyse game performance

Improve game.

![](_page_26_Figure_0.jpeg)

![](_page_27_Picture_0.jpeg)

- 15M devices daily
- 3B events per day (35k per second)
- 750 GB uncompressed

![](_page_28_Picture_3.jpeg)

### Store events in a log (immutability)

![](_page_29_Picture_3.jpeg)

![](_page_30_Figure_0.jpeg)

Log: immutable, write by appending 

**G** 

- Split producer & consumers
- High-availability write path (S3)

- 3 4

![](_page_31_Figure_0.jpeg)

**G**\

- Log: immutable, write by appending
- Split producer & consumers
- High-availability write path (S3)

![](_page_32_Figure_1.jpeg)

- Log: immutable, write by appending
- Split producer & consumers
- High-availability write path (S3)

![](_page_32_Picture_5.jpeg)

![](_page_32_Figure_6.jpeg)

![](_page_33_Picture_1.jpeg)

### If you mess up, redo it (idempotency)

![](_page_33_Picture_3.jpeg)

G

### get\_checkpoint() return "2014-10-01"

![](_page_35_Picture_2.jpeg)

- get\_checkpoint() return "2014-10-01"
- Process events from log offset "2014-10-01" to log offset "2014-10-02"

![](_page_36_Picture_3.jpeg)

- get\_checkpoint() return "2014-10-01"
- Process events from log offset "2014-10-01" to log offset "2014-10-02"
- When all messages for 2014-10-01 are processed, write to DB, overwrite any existing data (idempotence)

![](_page_37_Picture_4.jpeg)

- get\_checkpoint() return "2014-10-01"
- Process events from log offset "2014-10-01" to log offset "2014-10-02"
- When all messages for 2014-10-01 are processed, write to DB, overwrite any existing data (idempotence)
- set\_checkpoint("2014-10-02")

![](_page_38_Picture_5.jpeg)

### Where can I get one?

- Apacha Samza!
- Does everything we do and *much more*
- Released after we went live ... :/

![](_page_39_Picture_4.jpeg)

# Thank you

![](_page_40_Picture_1.jpeg)

![](_page_40_Picture_2.jpeg)

Q&A

![](_page_41_Picture_1.jpeg)

![](_page_41_Picture_2.jpeg)

![](_page_42_Picture_0.jpeg)

### GameAnalytics Building data-intensive services (aka. immutability and idempotence)

@knutin GameAnalytics

### Please ask questions via the mobile app!

![](_page_43_Picture_1.jpeg)

Engage

![](_page_43_Picture_3.jpeg)

![](_page_43_Picture_4.jpeg)

# Why you don't want your realtime analytics to be exact

### Mikio Braun, TU Berlin/streamdrill @mikiobraun

GOTO Berlin, Nov 7, 2014

Mikio Braun Why real-time analytics don't have to be exact

### Analyzing User Interaction

![](_page_45_Picture_1.jpeg)

Mikio Braun Why real-time analytics don't have to be exact

### What can we do besides scaling?

### Approximate?

### But is that ok? Do we really want our analytics to be exact?

Mikio Braun Why real-time analytics don't have to be exact

### Why you don't want your real-time analytics to be exact

- 1. Results are changing all the time anyway.
- 2. You can't have exactness, real-time, and big data at the same time (or it costs a lot).
- 3. Exactness is often not necessary.
- 4. You probably already have a batch system in place.

Reason 2: You can't have exactness, real-time, and big data at the same time (or it costs a lot)

![](_page_48_Figure_1.jpeg)

http://www.slideshare.net/acunu/realtime-analytics-with-casaandra Why real-time analytics don't have to be exact (c) 2014 streamdrill

Mikio Braun

### Why you don't want your real-time analytics to be exact

- 1. Results are changing all the time anyway.
- 2. You can't have exactness, real-time, and big data at the same time (or it costs a lot).
- 3. Exactness is often not necessary.
- 4. You probably already have a batch system in place.

### Reason 3: Exactness is often not necessary

![](_page_50_Figure_1.jpeg)

Mikio Braun

Why real-time analytics don't have to be exact

### Why you don't want your real-time analytics to be exact

- 1. Results are changing all the time anyway.
- 2. You can't have exactness, real-time, and big data at the same time (or it costs a lot).
- 3. Exactness is often not necessary.
- 4. You probably already have a batch system in place.

### streamdrill

- Core Engine
  - approximative counting and trends
  - rolling time windows
    based on exponential
    decay
  - secondary indices

- Features
  - true real-time, low latency (ms)
  - Dashboard & REST interface
  - about 20 events/sec, track 1M objects/1GB RAM
- Applications
  - real-time user profiling
  - recommendation

### Dashboard

C C A D play.streamdrill.com/dashboard/#!/							
stream <b>drill</b> <sup>β</sup> Dashboard					196MB 78MB 98MB	Documentation	
E	6 active trends	306.1k analyzed entities	<b>0.0k</b> updates/s				
#	trend	fill		capacity	timescales	updates/sec	
1	symbol-combinations	6%		100,000	day,hour,minute		
2	symbol-hashtag	3%		100,000	day,hour,minute	0.0	
3	symbol-keywords	69%		100,000	day,hour,minute	· · · · · · · 2.1	
4	symbol-mentions	14%		100,000	day,hour,minute	0.3	
5	symbol-trend	39%		1,000	day,hour,minute		
6	symbol-url	8%		100,000	day,hour,minute	0.3	

### Mikio Braun Why real-time analytics don't have to be exact

### Trend view

OOO Bstreamdrill - Dashboard ×							
🗧 🔆 C 👬 🗋 play.streamdrill.com/dashboard/#!/view/symbol-keywords 🖒 🔽 🦓 🗢 🌣 β 📭 🚍							
stream <b>drill</b> <sup>β</sup> Das	hboard	🚍 6 🗳 306.3k 🔌 0.0k	170MB	103MB 98MB	Documentation	L demo@streamdrill.com ▼	
« Dashboard / symbo	ol-keywords						
symbol-keywords	69%						
© ⊙ day ) hour ) mir	nute <b>T</b> filter by \$FB	X keyword	×				
#	symbol	keyword			score		
1	\$FB	facebook			275.9		
2	\$FB	more			119.4		
3	\$FB	people			112.9		
4	\$FB	than			112.0		
5	\$FB	engineer			110.8		
6	\$FB	push			110.2		
7	\$FB	most			109.9		
8	\$FB	lifetime			108.4		
9	\$FB	code			108.4		
10	\$FB	pacific			106.1		
11	\$FB	goog			67.4		
12	\$FB	aapl			61.9		
13	\$FB	home			60.8		
14	\$FB	current			58.1		
15	\$FB	value			56.6		
16	\$FB	utc			56.0		

Mikio Braun

Why real-time analytics don't have to be exact

### Real-time Recommendation at serienjunkies.de

![](_page_55_Picture_1.jpeg)

Serie News Spoiler Stream Darsteller Episoden Reviews FAQ Download Video Shop Links

### Forum

 $\geq$ 

NCIS

NCIS (Naval Criminal Investigative Service) verfolgt Verbrechen in der United States Navy und im United States Marine Corps

<u>Über die Serie</u>
 Trailer

Hauptdarsteller

FAQ

Aktuelle Meldungen

8+1 11 Empfehlen 108

**NCIS** steht für (Naval Criminal Investigative Service) und ist ein Spinoff der beliebten Anwaltserie "JAG" bei dem aber die strafrechtlichen Ermittlungen im Zentrum stehen.

Die Spezialeinheit des NCIS, die sich primär mit der Strafverfolgungs- Spionageabwehr der Navy und des Marine Corps. befasst, ist in Washington D.C. angesiedelt. Das NCIS untersucht alle Straftaten und Verbrechen die vor dem Militärgericht innerhalb der Navy

![](_page_55_Picture_13.jpeg)

![](_page_55_Picture_14.jpeg)

Vampire Diaries Dir gefällt auch Game of Thrones

The Walking Dead Beliebte Serie

### Mikio Braun Why real-time analytics don't have to be exact

### **Realtime User Profiles**

![](_page_56_Figure_1.jpeg)

Mikio Braun Why real-time analytics don't have to be exact

### **Realtime User Profiles**

stream <b>drill</b> Dashboard	📰 1 🚨 1720.1m 🔌 4.3k	13GB	90 Documentation & Anonymous -
«Dashboard / nuggad			Tue Apr 01 2014 16:42:10 GMT+0200 (CEST)
Real-Time User Profile nuggad			99%
Global Fingerprints more details »			
			week
			day
			hour

### O Activity by week day hour

### Top 5 Sites more »

Mikio Braun

# Site	Activity Category Fingerprint	Top Categories
1 18/56/869	91,969,740.6	market place,football,lottery
2	69,572,802.1	regional,humour,ecommerce
3	59,012,708.1	car accesories,funbikes,lift truck
4	48,277,773.9	classified ad,market place
5	41,376,343.4	internal policy,obligatory,real

### Top 5 Networks more »

# Network	Activity Category Fingerprint	Top Categories
1.1479001	94,357,907.6	jewellery,international,dating
2	76,881,622.0	high-performance cars,fashion (men),beer
3	69,572,934.9	regional,humour,ecommerce
4	59,012,818.0	car accesories,funbikes,lift truck
5	48,277,837.6	market place, classified ad

![](_page_57_Figure_7.jpeg)

#	User	Activity Category Fingerprint	Top Categories
		by week low high	
1	10000-000-000-000	56,726.5	
2	to-people people	39,142.5	travel planning, red.content
3	FORE ITERMENCO	38,179.7	news,red.content
4	NETHEORIES/NOACE	37,251.3	news,red.content
5	1.06/1.06/100/107	34,492.0	classified ad,job offering
6	POT THE GAST INCOME.	27,197.3	
7	1000007691000	24,856.3	lifestyle and leisure,red.content
8	COBINED ARRENGE	23,868.5	news,red.content
0	100070060400700001	00.036.0	markat place elevelfied ed

Why real-time analytics don't have to be exact

### **Realtime User Profiles**

Category	Activity						
	week	ы.	day	.al	hour	al.	
chat	2.0		1.9	0.45% 🕇			
information	12.5		9.9	0.38% 🕇	0.9		
society				1.14% 🕇	0.5		
entertainment media	64.7	0.13% 🕇	53.2	2.67% 🕇	7.1		
television	174.5	0.15% 🕹	141.2	3.12% 🕹	14.1		
video gallery	253.7		206.0	0.59% 🕇	23.0		
communication boards	2.0		1.9	0.45% 🕹			

Mikio Braun Why real-time analytics don't have to be exact

### Summary

- real-time doesn't have to be exact
- streamdrill: real-time analytics plattform
- Contact us at info@streamdrill.com if you're interested in
  - real-time profiling
  - real-time recommendation
  - anything else real-time related!

### Please evaluate this talk via the mobile app!

![](_page_60_Picture_1.jpeg)

Engage

![](_page_60_Picture_3.jpeg)

![](_page_60_Picture_4.jpeg)