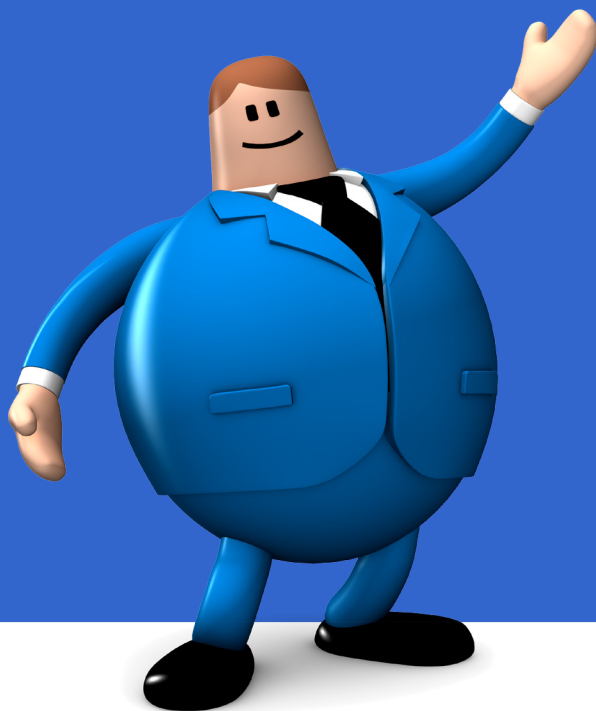


Half a year of Measuring 2.0 in production at bol.com



Agenda

- Motivation
- Design
- Production experiences
 - Measurements
 - Kafka
 - Trending products
 - Avro/Parquet
- Future plans

Ivan Budinčević

- ibudincevic@bol.com
- PhD in nuclear physics KU Leuven 2015
- Data analysis with python
- Java developer at bol.com since 2016
- Interested in big data technologies, streaming data processing



Motivation

bol.com
Kies een categorie
Verkoop Zakelijk Fotoservice Welkom Niels Bestelstatus Lijstjes Klantenservice Zoeken

30% korting op Pampers Tot 50% op bedtextiel
Gratis verzending vanaf 20 euro, gratis retourneren en 30 dagen bedenktijd*

Ik wil jou een mooi cadeau geven

Sieraden met een hartje al vanaf 9,-

Tot 60% korting op parfums

Cry Primal
Special edition met 3 extra missies
Vanaf leeftijd 7+

Bestverkochte pc games
Iedere 24 uur up-to-date

- 1 De Sims 4 - PC + MAC € 49,99
- 2 Rise of the Tomb Raider - PC € 44,99
- 3 Grand Theft Auto V (GTA 5) - PC € 42,99
- 4 Football Manager 2016 - PC € 49,99
- 5 Fallout 4 - PC € 42,99

Dagaanbieding
Je hebt nog 14 uur, 15 minuten, 2 seconden

Differnz Plato Fonteinset - 36 x 18 cm
- Keramiek
De Differnz Plato Fonteinset is door ~~149,99~~ **59,99**

ANNO 2205 - Special Edition
PC ~~€ 59,99~~ **€ 47,99**

Accessoires
Bestverkochte pc accessoires

- 1 Logitech G430 - 7.1 Virtueel Su... € 74,99
- 2 Microsoft Xbox 360 Controller - ... € 34,99
- 3 Razer Kraken USB Essential 7.1 ... € 59,99
- 4 Turtle Beach Ear Force PX22 Wir... € 68,99

Op je verlanglijstje

MINDSET
Mindset in het dagelijks leven € 15,00

MINDSET
Mindset € 17,95

Breindidactiek
€ 24,95

Just Cause 3 - Day One Edition - PC
€ 87,49

Aanbiedingen

Ik wil jou een mooi cadeau geven

Welkom Niels Bestelstatus Lijstjes Klantenservice Zoeken

bol.com

The screenshot displays the bol.com website interface with several key elements:

- Navigation:** Top navigation bar includes "Welkom Niels", "Bestelstatus", "Lijstjes", and "Klantenservice". A search bar and "Zoeken" button are also present.
- Category Menu:** A sidebar on the left lists categories such as "Boeken", "Muziek, Film & Games", "Computer & Elektronica", "Speelgoed & Hobby", "Baby & Kind", "Mooi & Gezond", "Sieraden & Accessoires", "Sport & Vrije tijd", "Wonen & Koken", "Tuin & Klussen", "Dier", and "Cadeaus & Inspiratie".
- Promotional Banners:**
 - A large banner for "Ik wil jou een mooi cadeau geven" featuring pink balloons and a woman's face.
 - A banner for "Sieraden met een hartje al vanaf 9,-".
 - A banner for "Tot 60% korting op parfums".
 - A banner for "30% korting op Pampers" with "Gratis verzending vanaf 20 euro, gratis retourneren en 30 dagen bedenktijd*".
 - A banner for "Tot 50% op bedtextiel" with "Gratis retourneren en 30 dagen bedenktijd*".
- Product Listings:**
 - Bestverkochte pc games:** A list of top-selling games including "De Sims 4 - PC + MAC" (€49,99), "Rise of the Tomb Raider - PC" (€44,99), "Grand Theft Auto V (GTA 5) - PC" (€42,99), "Football Manager 2016 - PC" (€49,99), and "Fallout 4 - PC" (€42,99).
 - Accessoires:** A list of accessories including "Logitech G430 - 7.1 Virtueel Su ..." (€74,99), "Microsoft Xbox 360 Controller ..." (€34,99), "Razer Kraken USB Essential 7.1 ..." (€59,99), and "Turtle Beach Ear Force PX22 Wir ..." (€68,99).
 - Dagaanbieding:** A section for a limited-time offer on a "Differnz Plato Fonteinset - 36 x 18 cm - Keramiek" for €59,99 (reduced from €119,99).
 - Ook interessant:** A section featuring books like "MINDSET" (€17,95) and "BREIN DIDACTIEK" (€24,95).

The image shows a screenshot of the bol.com website interface, tilted at an angle. The top navigation bar includes the bol.com logo, a search bar, and user account information (Welkom Niels, Bestelstatus, Lijstjes, Klantenservice). A blue banner at the top right offers a 30% discount on Pampers and a 50% discount on bedtextiel. Below this, a large promotional banner features pink hearts and the text "Ik wil jou een mooi cadeau geven". To the right, a "Cry Primal" game advertisement is visible. A central "Dagaanbieding" (Flash Deal) is circled in red, showing a white sink for €59,99 (reduced from €119,99) with 14 minutes and 15 seconds remaining. Below the flash deal, a "Ook interessant" (Also interesting) section lists books like "Mindset" and "Breindidactiek". On the right side, a "Bestverkochte pc games" (Best-selling PC games) list includes titles like "De Sims 4", "Rise of the Tomb Raider", "Grand Theft Auto V", "Football Manager 2016", and "Fallout 4". At the bottom right, a section for "Accessoires" (Accessories) lists items like Logitech G430 headphones, Microsoft Xbox 360 Controller, and Razer Kraken USB headset.



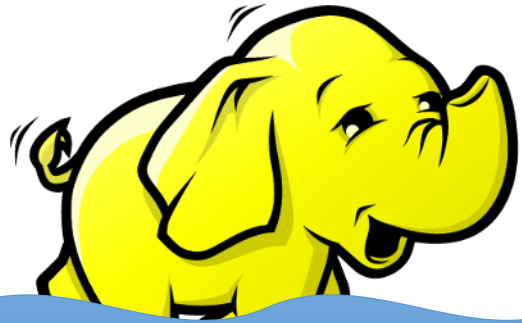
> 15 million products for sale
> 50 million in catalog



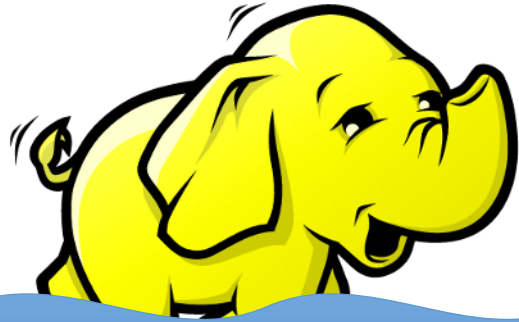
> 6 million active customers
> 27 million visits per month



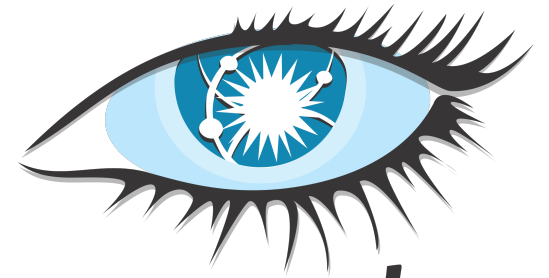
> 4 billion
pageviews/year



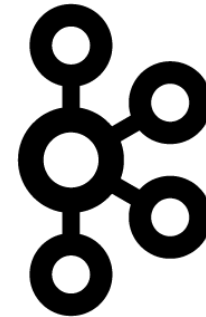
Hadoop in production
since 2010



Hadoop in production
since 2010



cassandra



kafka

APACHE
HBASE



Search suggestions

Verkopen Zakelijk Cadeaubon

Inloggen Bestelstatus Lijstjes Klantenservice 

bol.com

Kies een categorie

Dark

Alle artikelen

Zoeken



Categorieën

Outlet

Aanbiedingen

Gratis verz

Zoeksuggesties

meer je wilt met **Select** artikelen*

dark ambiënt in Muziek

dark souls 3 in Alle artikelen

dark souls 3 in Games

darksiders

dark tower stephen king

dark souls 3 ps4

dark souls

dark tower

dark matter

darksiders 2

dark knight

dark matter blake crouch

Merken

Dark Horse

Dark and Lovely

Dark Crystal

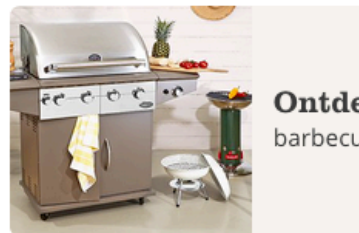
Populaire zoekresultaten



Dark Souls 3 - PS4 - Bandai Namco
PlayStation 4 | PEGI-leeftijd: 16 | 12 april 2016



Dark Souls III (3) (GOTY Edition) PS4 - Bandai Namco
PlayStation 4 | PEGI-leeftijd: 18 | 21 april 2017



Meest populair op dit mom

Dagaanbieding

bol.com

Personalization

Je bekeek

Ook interessant:



In The Mood Wicker -
Sierkussen - Paa...



In The Mood
Pailletten Klein -...

€ 18,99



In The Mood Turijn -
Sierkussen - Paa...

€ 13,99



In The Mood
Koffieboon -...

€ 18,99

> [Bekijk al je persoonlijke aanbevelingen](#)

Speciaal voor jou geselecteerd



Microsoft Office 365
Personal - Neder...

€ 69,00



12 - 99 jaar
United Entertainment
- Magic Wine Dec...

€ 17,95



Microsoft Windows 7
Professional - En...

€ 69,99



Microsoft Office 365
Home - Nederland...

€ 99,00

> [Bekijk al je persoonlijke aanbevelingen](#)

Dagaanbieding

Je hebt nog 10 uur, 40 minuten,
35 seconden



Populaire cd's en films voor 10
euro!

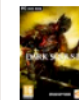
Alleen vandaag: populaire cd's
en films waar moeders dol op
zijn voor maar 10 euro! Zoals de
nieuwe John Mayer, O'G3NE en
Soof 2!

> [Naar de actie](#)

> [Ontvang elke dag de
dagaanbieding](#)

Tip voor jou

Misschien heb je interesse in dit
artikel?



Dark Souls 3 -
Windows

Windows

€ 29,99



Personalization



Je bekeek

Ook interessant:



In The Mood Wicker -
Sierkussen - Paa...



In The Mood
Pailletten Klein -...
€ 18,99



In The Mood Turijn -
Sierkussen - Paa...
€ 13,99



In The Mood
Koffieboon -...
€ 18,99

[> Bekijk al je persoonlijke aanbevelingen](#)

Speciaal voor jou geselecteerd



Microsoft Office 365
Personal - Neder...
€ 69,00



12 - 99 jaar
United Entertainment
- Magic Wine Dec...
€ 17,95



Microsoft Windows 7
Professional - En...
€ 69,99



Microsoft Office 365
Home - Nederland...
€ 99,00

[> Bekijk al je persoonlijke aanbevelingen](#)



Dagaanbieding

Je hebt nog 10 uur, 40 minuten,
35 seconden



Populaire cd's en films voor 10
euro!

Alleen vandaag: populaire cd's
en films waar moeders dol op
zijn voor maar 10 euro! Zoals de
nieuwe John Mayer, O'G3NE en
Soof 2!

[> Naar de actie](#)

[> Ontvang elke dag de
dagaanbieding](#)

Tip voor jou

Misschien heb je interesse in dit
artikel?



Dark Souls 3 -
Windows
Windows
€ 29,99

How do we react to customer needs in real-time?

How do we react to customer needs in real-time?

- Measure customer behavior and take action in real-time

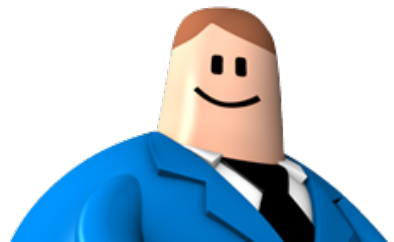
How do we react to customer needs in real-time?

- Measure customer behavior and take action in real-time
- Measure EVERYTHING

Goals of “Measuring 2.0”

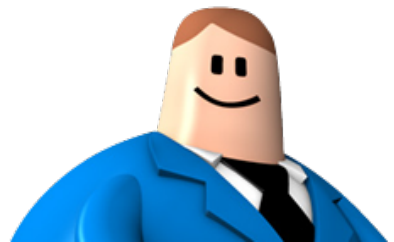
- Measure everything on our website
 - All interactions (also AJAX)
 - All channels (also mobile, email, ...)
 - All countries
 - All details
 - All visitors
- More reliable data
- Lowest possible load on the client
- Lowest possible latency (< 1 second)

MEASURE ALL THE THINGS



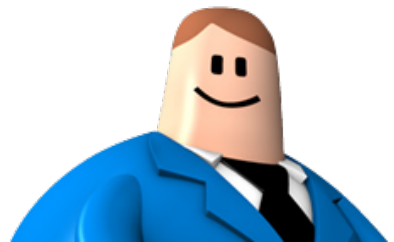
Goals of “Measuring 2.0”

- Developer
 - Easy to build
 - Easy to validate
 - Test automation
- Business
 - Always measure everything
 - Data is “independent”
 - New questions are allowed



Goals of “Measuring 2.0”

- Privacy
 - Personal Data Protection Act
 - *Wet bescherming persoonsgegevens*
 - No profiling beyond 2 years
- Security
 - Avoid storing “login” info.
- Business
 - Do “profiling” on everything
 - For many years (>2)



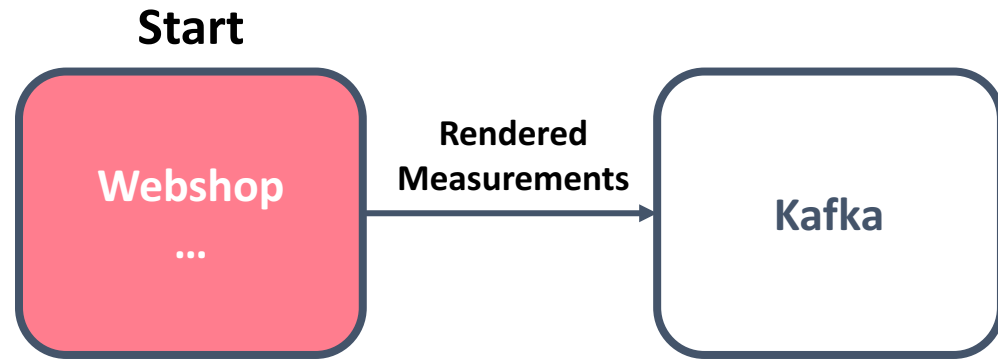
Design

Design

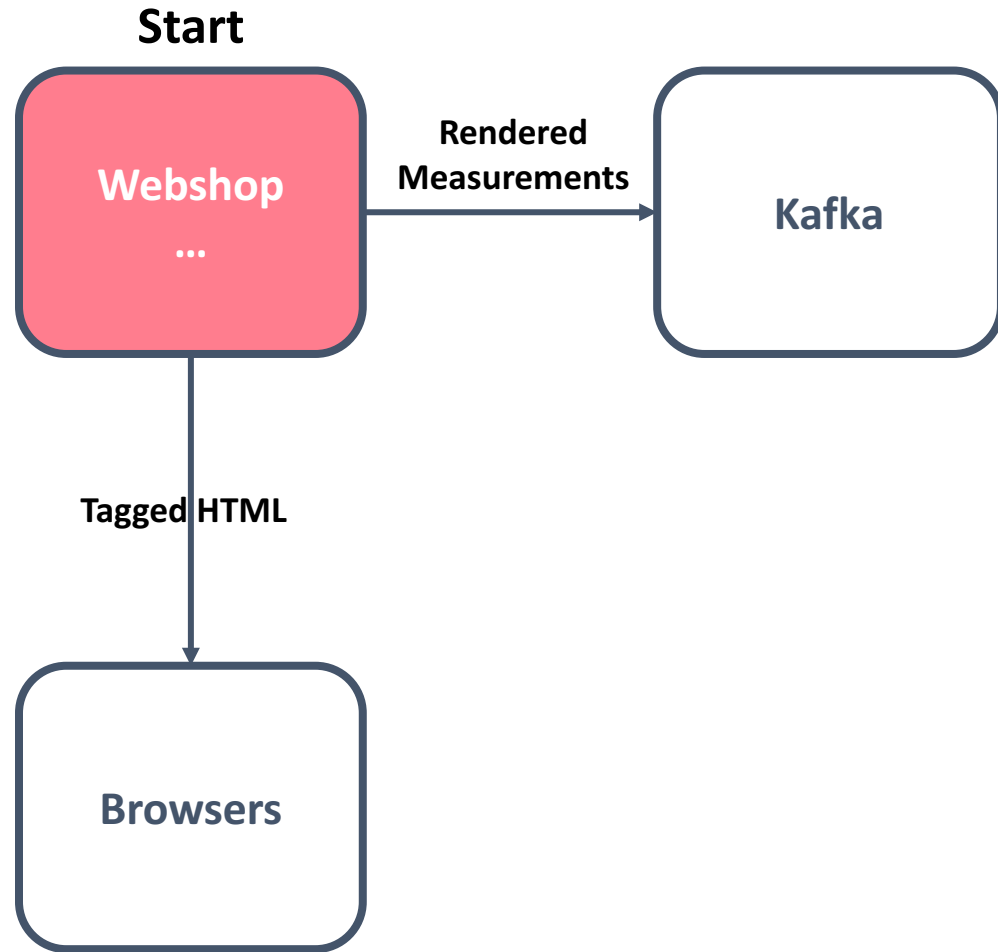
Start



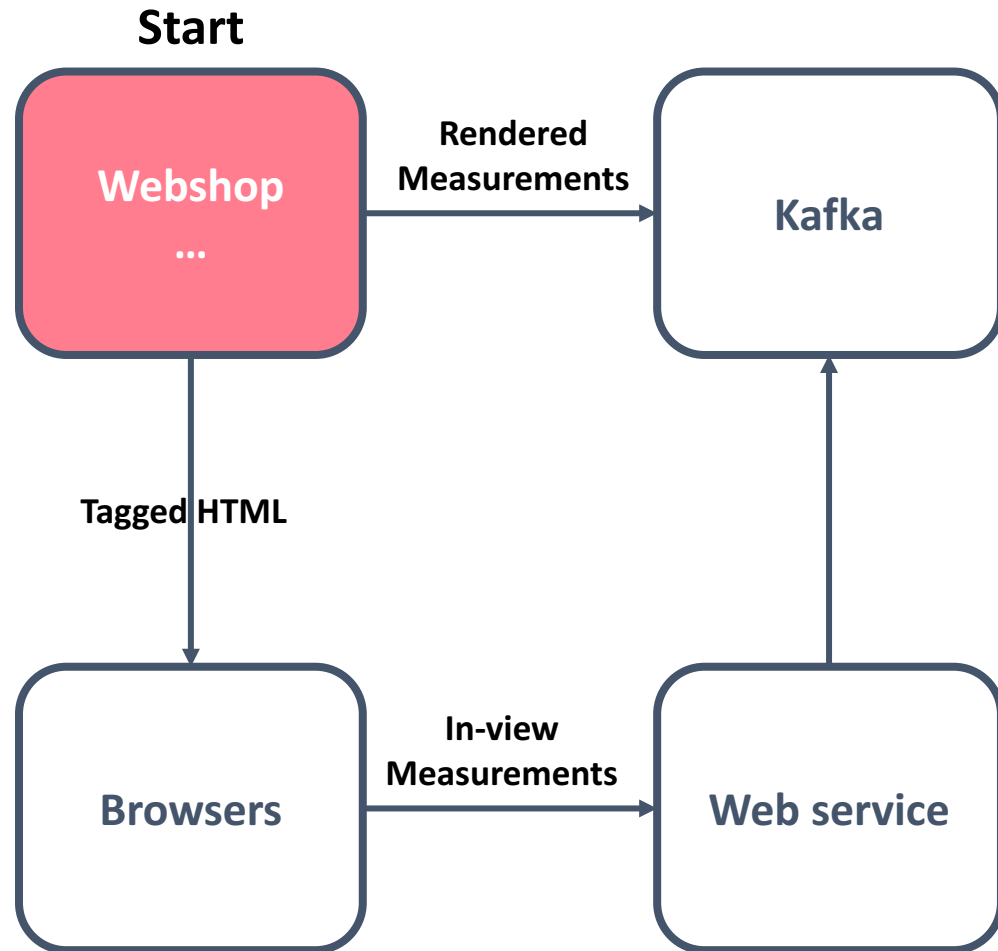
Design



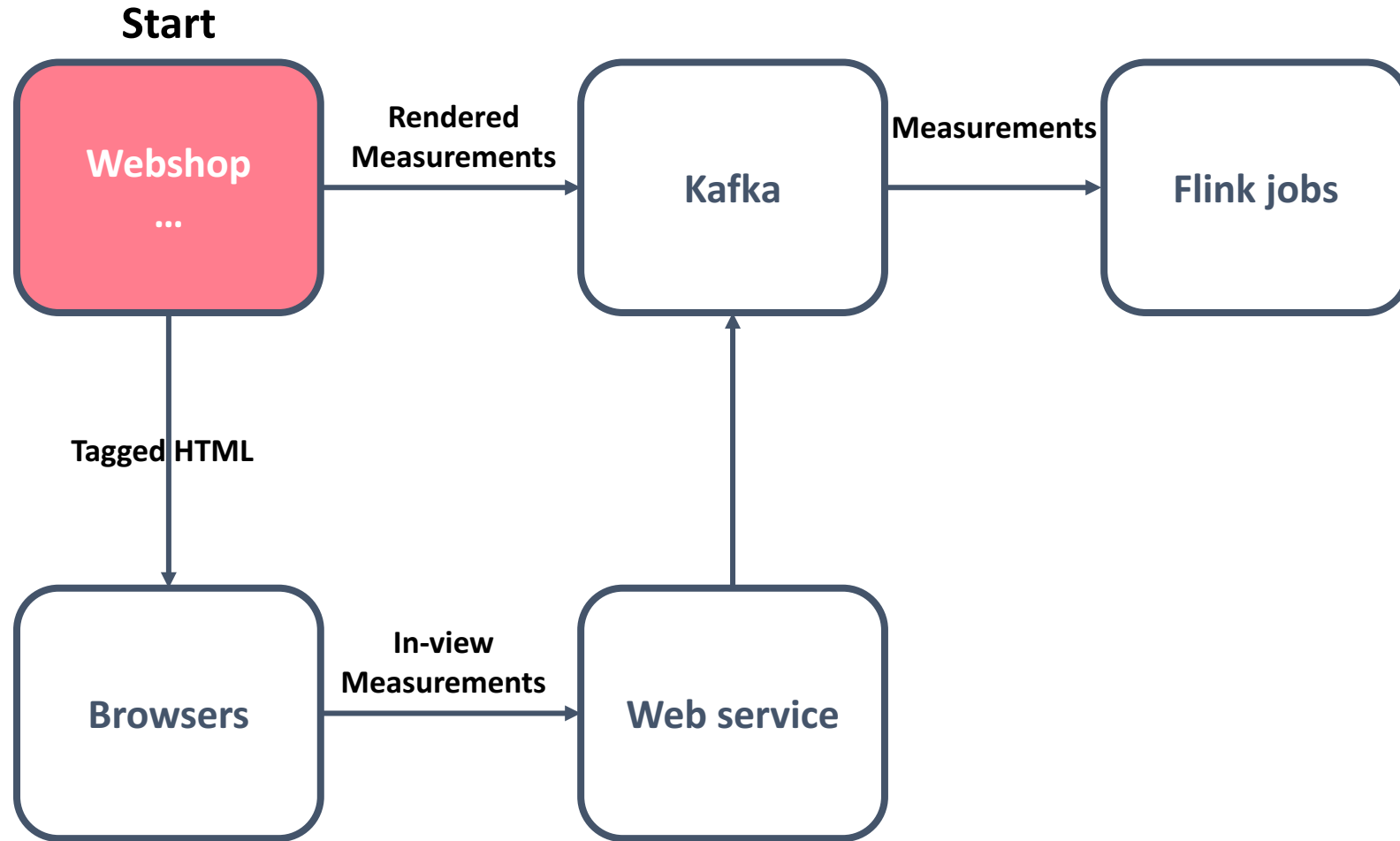
Design



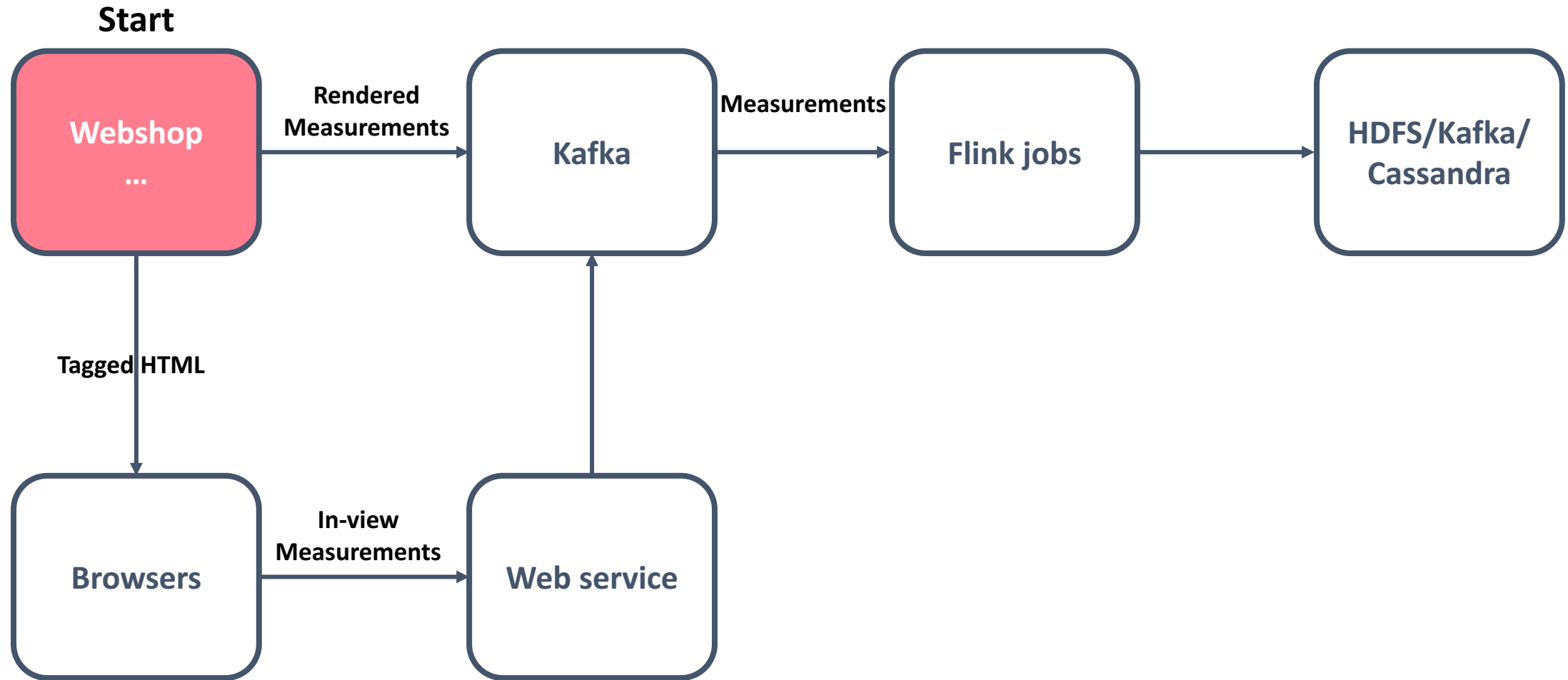
Design



Design



Design



Data structure

Verkopen Zakelijk Cadeaubon


Inloggen Bestelstatus Lijstjes Klantenservice 

bol.com 

Waar ben je naar op zoek?

Alle artikelen ▾

Zoeken


 2 Bestellen >

Kies een categorie ▾

1+1 gratis op Nivea Sun Korting op tuinsets & parasols

Outlet

Aanbiedingen

Gratis verzending vanaf 20 euro, gratis retourneren, bezorging waar en wanneer je wilt met **Select**  artikelen*

 **BulkAlert** >

Winkelwagentje

Bestel nu en je bestelling wordt gratis verzonden!



The Elder Scrolls Online: Morrowind - Collector's Edition - Windows
Bethesda
Windows

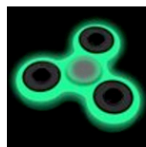
Nog niet verschenen - reserveer een exemplaar Beschikbaar op 06-06-2017

Aantal

€ 99,99

 [Verplaats naar verlanglijstje](#)

 [Verwijder](#)



Hand Fidget Spinner Glow in the Dark - met hoogwaardige keramische lagers

Op voorraad. Voor 16:00 uur besteld, morgen in huis

Verkoper: BestGadgets4U

Aantal

€ 11,99


 [Verplaats naar verlanglijstje](#)

 [Verwijder](#)

Cadeaukaart- of kortingscode

> [Code invoeren](#)

Totaal artikelen (2) € 111,98

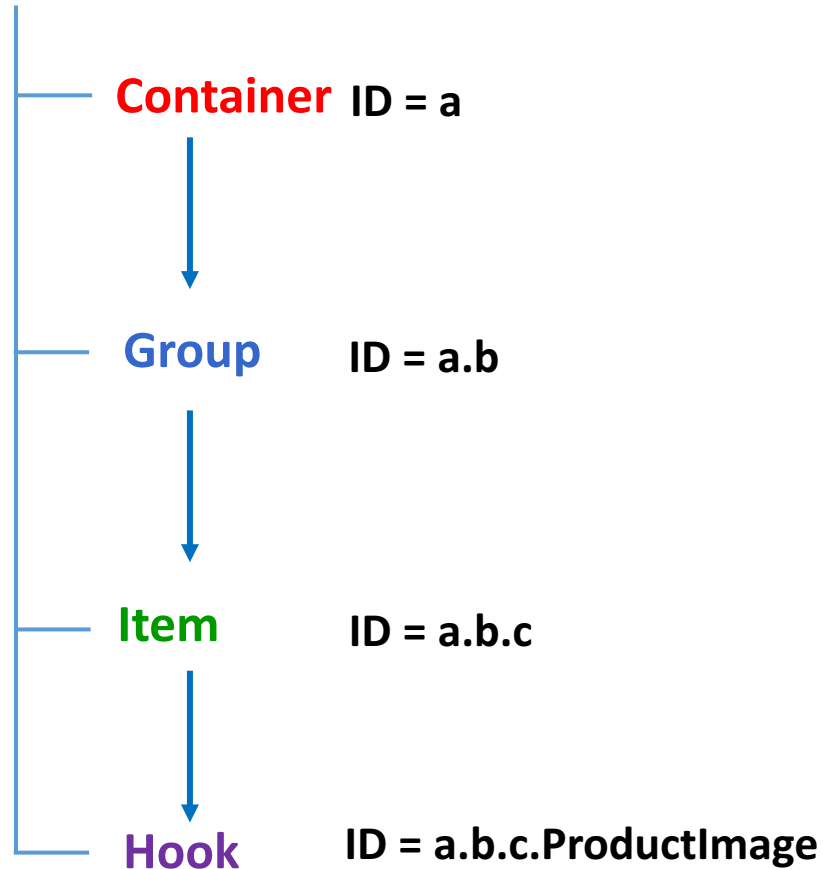
Verzendkosten  **Gratis**

Totaal € 111,98

> Verder naar bestellen

Data structure

Measurements



Verkopen Zakelijk Cadeaubon Inloggen Bestelstatus Lijstjes Klantenservice


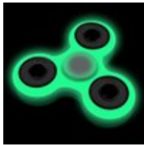
bol.com Alle artikelen v Zoeken 2 Bestellen >

Kies een categorie v 1+1 gratis op Nivea Sun Korting op tuinsets & parasols Outlet Aanbiedingen

Gratis verzending vanaf 20 euro, gratis retourneren, bezorging waar en wanneer je wilt met **Select** artikelen* **BulkAlert** >

Winkelwagentje

Bestel nu en je bestelling wordt gratis verzonden!

	The Elder Scrolls Online: Morrowind - Collector's Edition - Windows Bethesda Windows	Nog niet verschenen - reserveer een exemplaar Beschikbaar op 06-06-2017	Aantal <input type="text" value="1"/>	€ 99,99
		Verplaats naar verlanglijstje		
		Verwijder		
	Hand Fidget Spinner Glow in the Dark - met hoogwaardige keramische lagers	Op voorraad. Voor 16:00 uur besteld, morgen in huis Verkoper: BestGadgets4U	Aantal <input type="text" value="1"/>	€ 11,99
		Verplaats naar verlanglijstje		
		Verwijder		

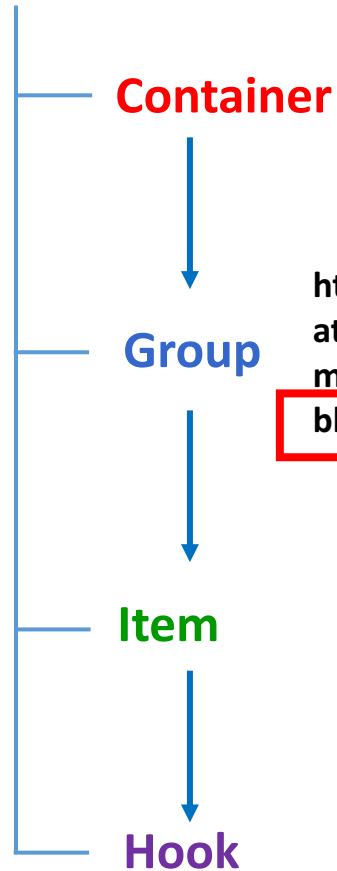
Cadeaukaart- of kortingscode Totaal artikelen (2) € 111,98
> Code invoeren Verzendkosten [?] Gratis

Totaal € 111,98

> Verder naar bestellen

Data structure

Measurements



<https://www.bol.com/nl/p/daten-is-meten/1001004010720589/?bltgh=ItemID.ProductImage>

Verkopen Zakelijk Cadeaubon Inloggen Bestelstatus Lijstjes Klantenservice


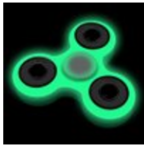
bol.com Alle artikelen v Zoeken 2 Bestellen >

Kies een categorie v 1+1 gratis op Nivea Sun Korting op tuinsets & parasols Outlet Aanbiedingen

Gratis verzending vanaf 20 euro, gratis retourneren, bezorging waar en wanneer je wilt met **Select** artikelen* **BulkAlert** >

Winkelwagentje

Bestel nu en je bestelling wordt gratis verzonden!

	The Elder Scrolls Online: Morrowind - Collector's Edition - Windows Bethesda Windows	Nog niet verschenen - reserveer een exemplaar Beschikbaar op 06-06-2017	Aantal <input type="text" value="1"/>	€ 99,99
		Verplaats naar verlanglijstje		
		Verwijder		
	Hand Fidget Spinner Glow in the Dark - met hoogwaardige keramische lagere	Op voorraad. Voor 16:00 uur besteld, morgen in huis Verkoper: BestGadgets4U	Aantal <input type="text" value="1"/>	€ 11,99
		Verplaats naar verlanglijstje		
		Verwijder		

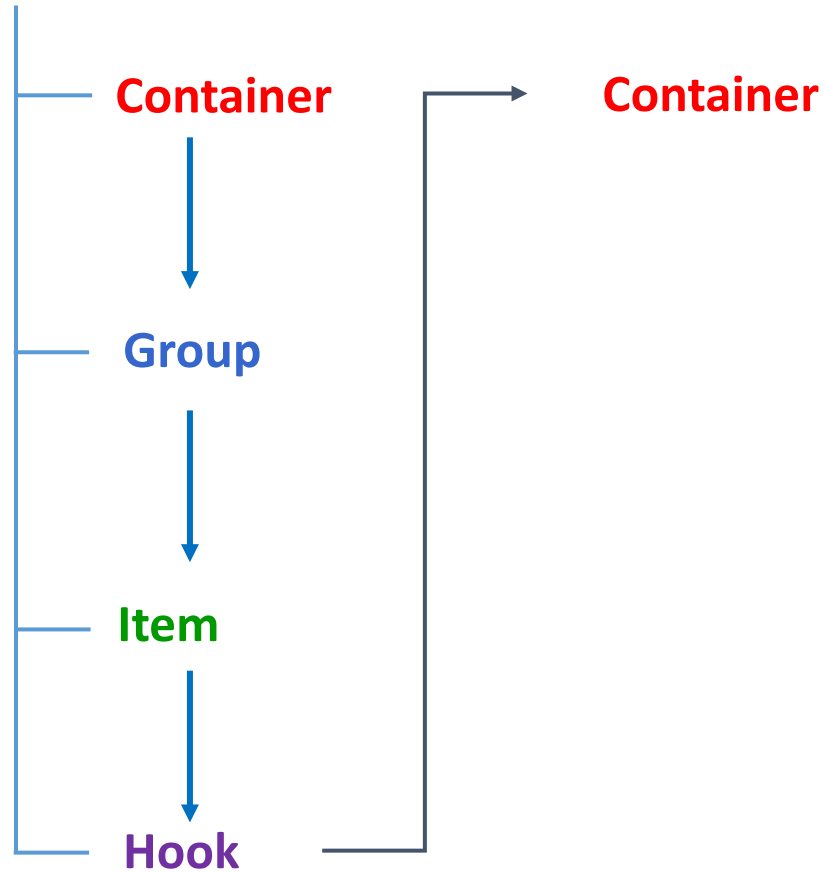
Cadeaukaart- of kortingscode Totaal artikelen (2) € 111,98
> Code invoeren Verzendkosten ? **Gratis**

Totaal € 111,98

> Verder naar bestellen

Data structure

Measurements



Bethesda videogames Xbox One videogames Windows videogames Actie videogames Top 10 videogames

< Terug > Games > Videogames

The Elder Scrolls Online: Morrowind - Collector's Edition - Xbox One

☆☆☆☆☆ Schrijf een review | Stel een vraag | E-mail deze pagina

Kies je platform

Xbox One PlayStation 4 Windows

98,99

3-6 werkdagen

Verkoop door bol.com

+ In winkelwagen Op verlanglijstje

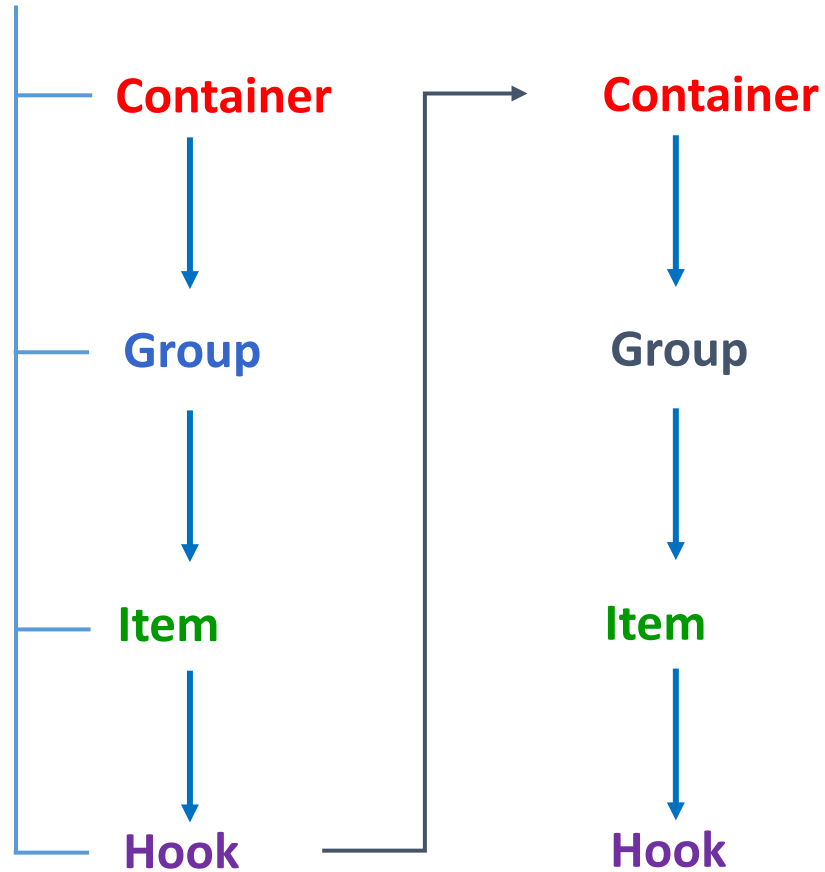
- ✓ Gratis verzending
- ✓ 30 dagen bedenktijd en gratis retourneren
- ✓ Dag en nacht klantenservice

Advertentie

Gratis ontpluizer
bij Philips stoom-
strijkers

Data structure

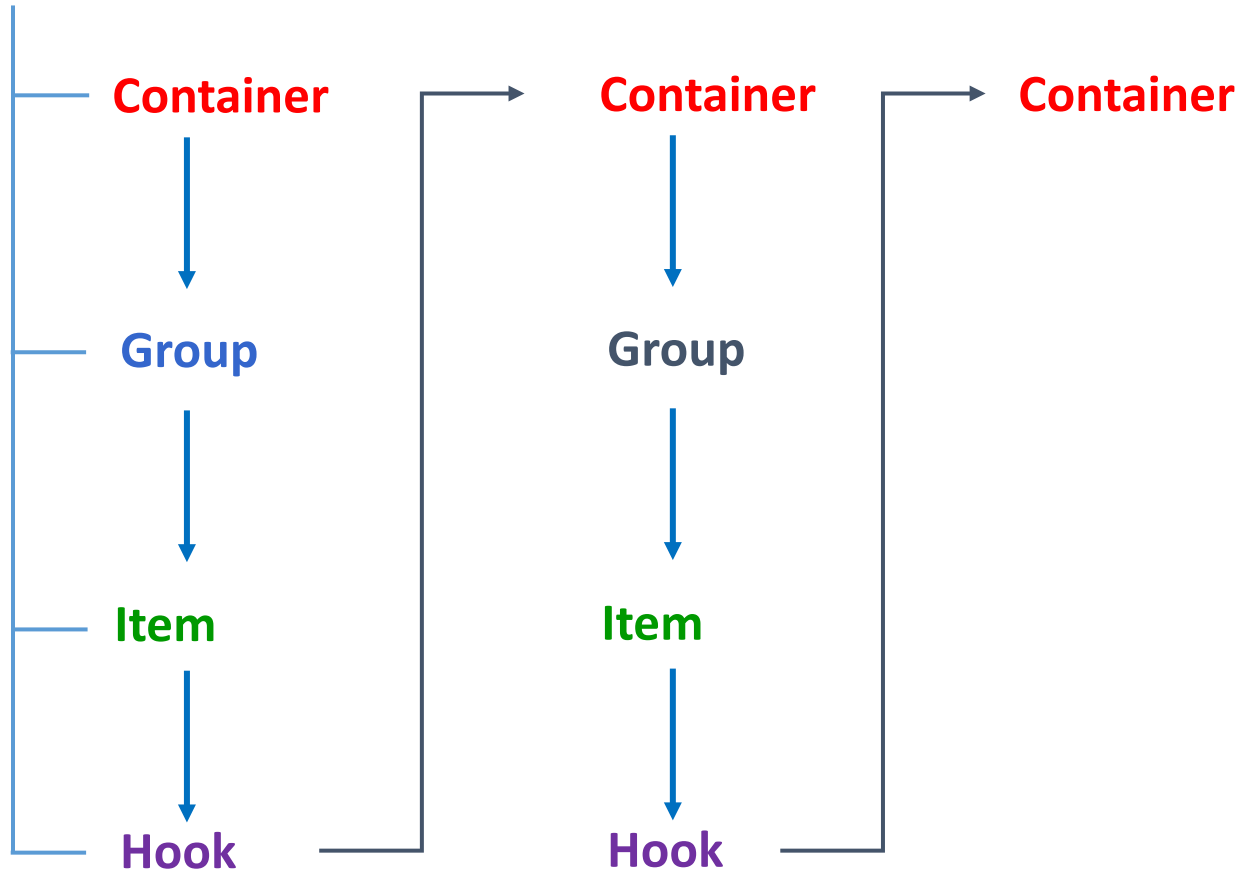
Measurements



The screenshot shows a product page for 'The Elder Scrolls Online: Morrowind - Collector's Edition - Xbox One'. The page is framed with a red border. At the top, there are navigation links: 'Bethesda videogames', 'Xbox One videogames', 'Windows videogames', 'Actie videogames', and 'Top 10 videogames'. Below the navigation is a breadcrumb trail: '< Terug > Games > Videogames'. The main title is 'The Elder Scrolls Online: Morrowind - Collector's Edition - Xbox One'. Below the title are options to 'Schrijf een review', 'Stel een vraag', and 'E-mail deze pagina'. The main image shows the game box set for 'The Elder Scrolls Online: Morrowind Collector's Edition' for Xbox One. To the right of the main image is a section for 'Kies je platform' with buttons for 'Xbox One', 'PlayStation 4', and 'Windows'. Below this is the price '98,99' and '3-6 werkdagen'. There is a 'Verkoop door bol.com' label and a '+ In winkelwagen' button. Below the button are three checkmarks: 'Gratis verzending', '30 dagen bedenktijd en gratis retourneren', and 'Dag en nacht klantenservice'. At the bottom of the page, there is an advertisement for a 'Gratis ontpluizer bij Philips stoomstrijkers' and the 'bol.com' logo.

Data structure

Measurements

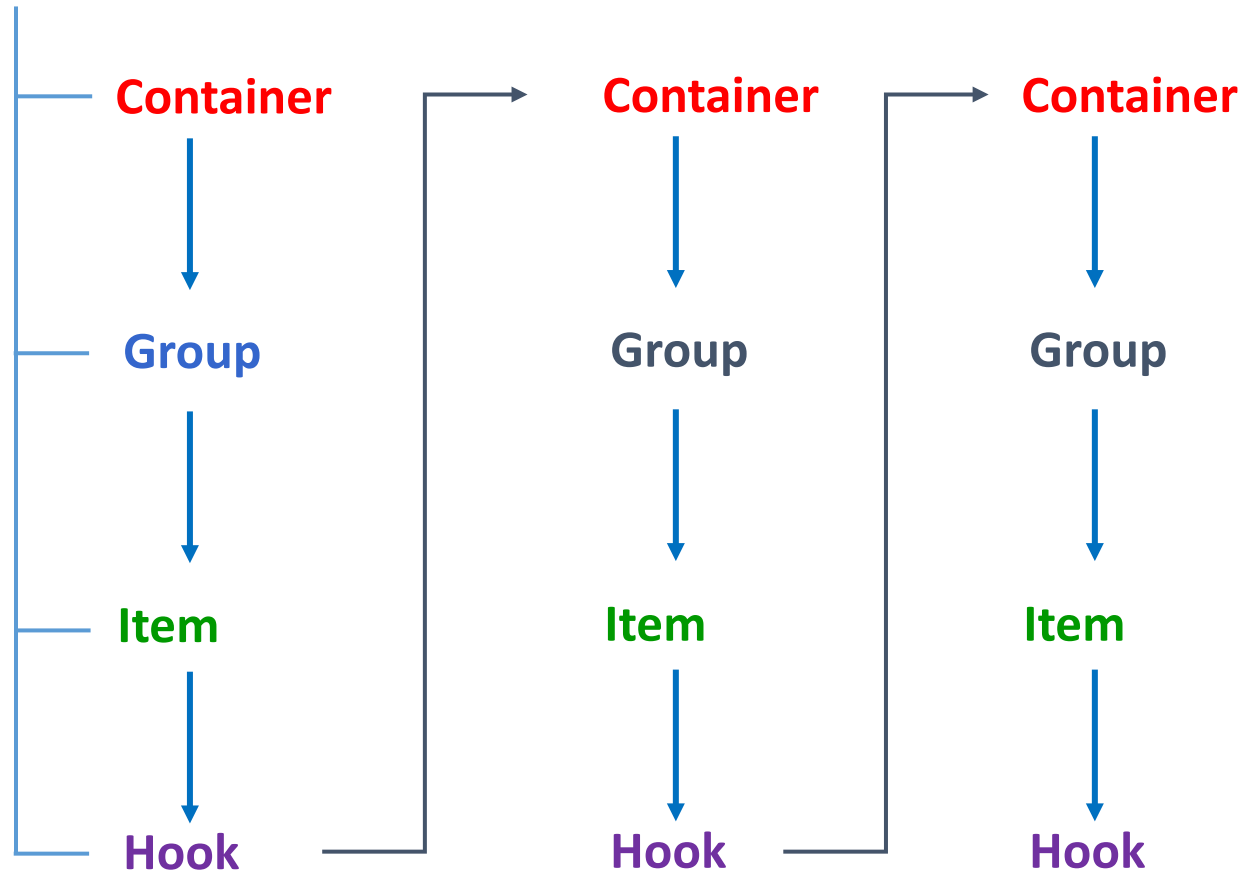


The screenshot shows a product page for 'Grand Theft Auto V (GTA 5) - PS4' on the bol.com website. The page is framed with a red border. On the left, there is a sidebar with filters for 'Games platform' (listing PlayStation 3, PlayStation 2, and PlayStation 4), 'Games genre' (listing Action, Adventure, and Sport), 'Aanraders' (listing 'Bestbeoordeeld', 'Te reserveren', and 'Nieuw verschenen'), 'Serie' (listing 'Games exclusief voor Xbox One', 'Pokemon - Games', and 'Games exclusief voor Playstation 4'), and 'PEGI-leeftijd' (listing '3+').

The main content area displays the game's title, platform (PlayStation 4), release date (18 november 2014), and a 5-star rating. It includes a price of €39,99, a 33% discount, and a 'Je bespaart 33%' badge. A 'Tweakers Awards 2015' badge is also visible. Below the main product, there are options for different platforms: PlayStation 3 (€20,99), Windows (€39,99), Xbox 360 (€20,99), Xbox One (€32,99), and Xbox One download (€69,99). A 'Sponsord' section for 'Pokemon X - 2DS + 3DS' is also shown with a price of €42,99 and a 14% discount. The page includes a search bar, a 'Verkoop door: bol.com' note, and a '+ In winkelwagen' button.

Data structure

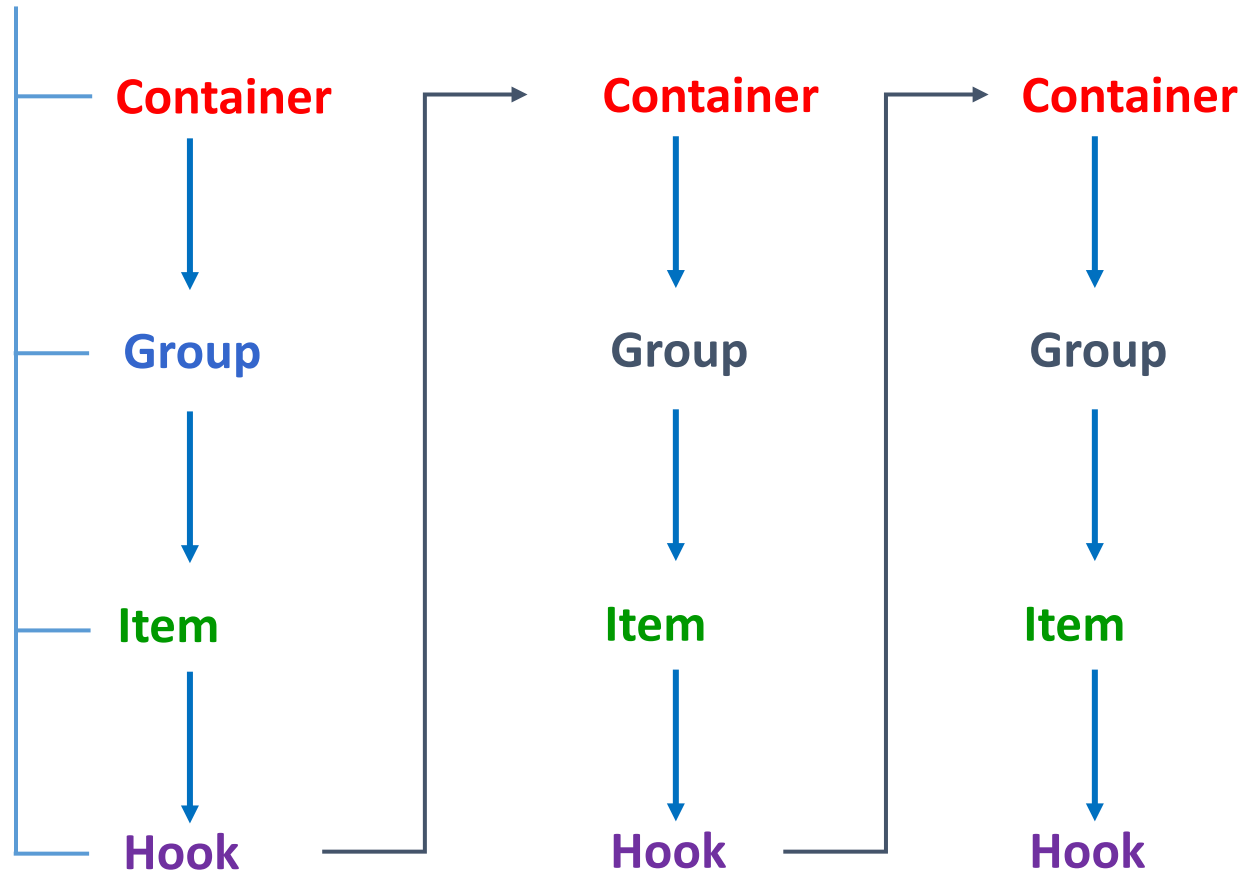
Measurements



The screenshot shows a website interface for video games. The main content area displays two game listings. The first listing is for "Grand Theft Auto V (GTA 5) - PS4" by TakeTwo, priced at €39,99. The second listing is for "Pokemon X - 2DS + 3DS" by Nintendo, priced at €42,99. Both listings include details such as platform, PEGI rating, and availability. The interface also features a sidebar with filters for Games platform, Games genre, Aanraders, Serie, and PEGI-leeftijd. The bottom right corner of the screenshot is highlighted with a purple box, showing a "In winkelwagen" button.

Data structure

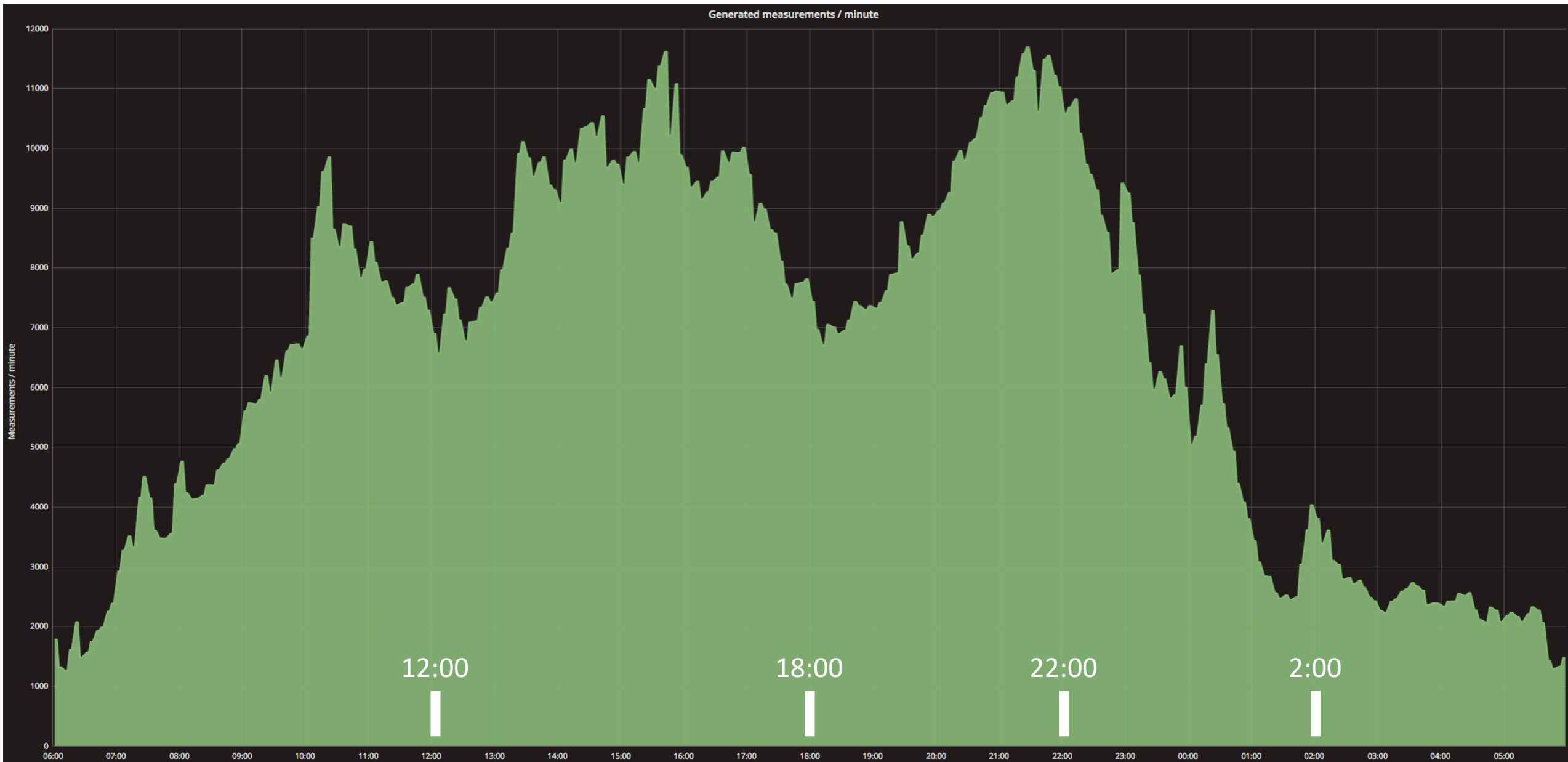
Measurements



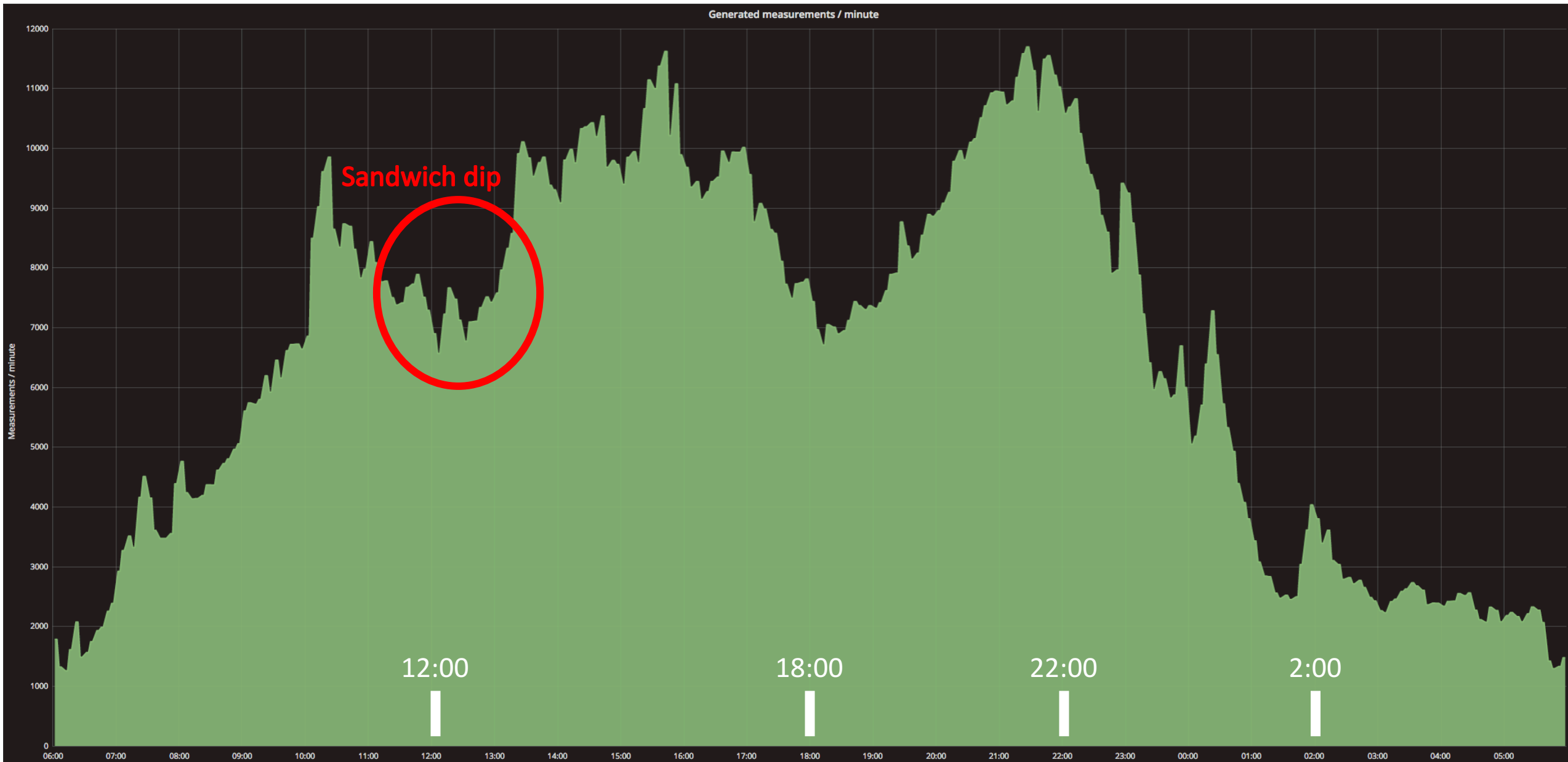
Customer's journey with cause and effect

Production experiences

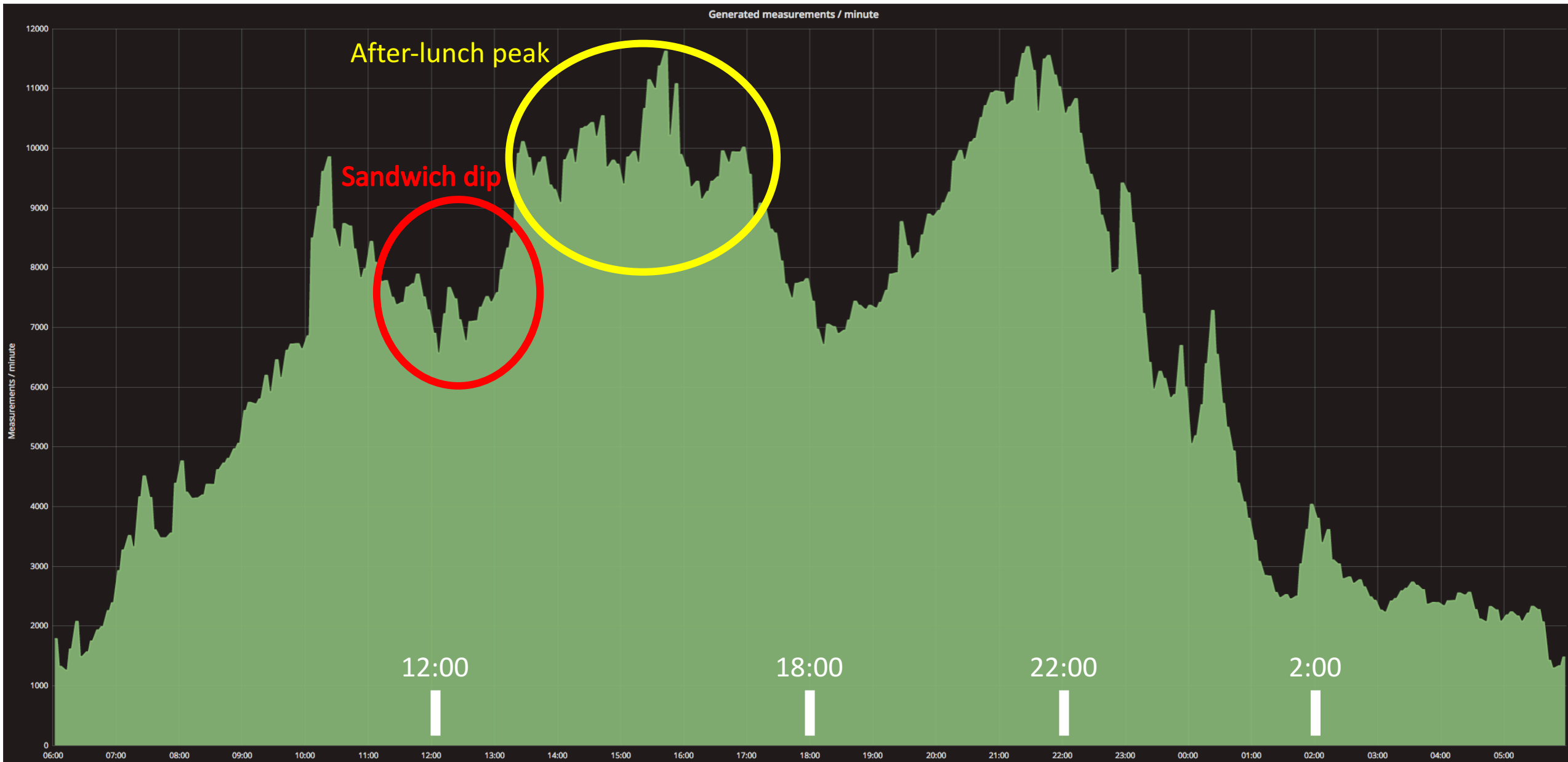
Production measurements



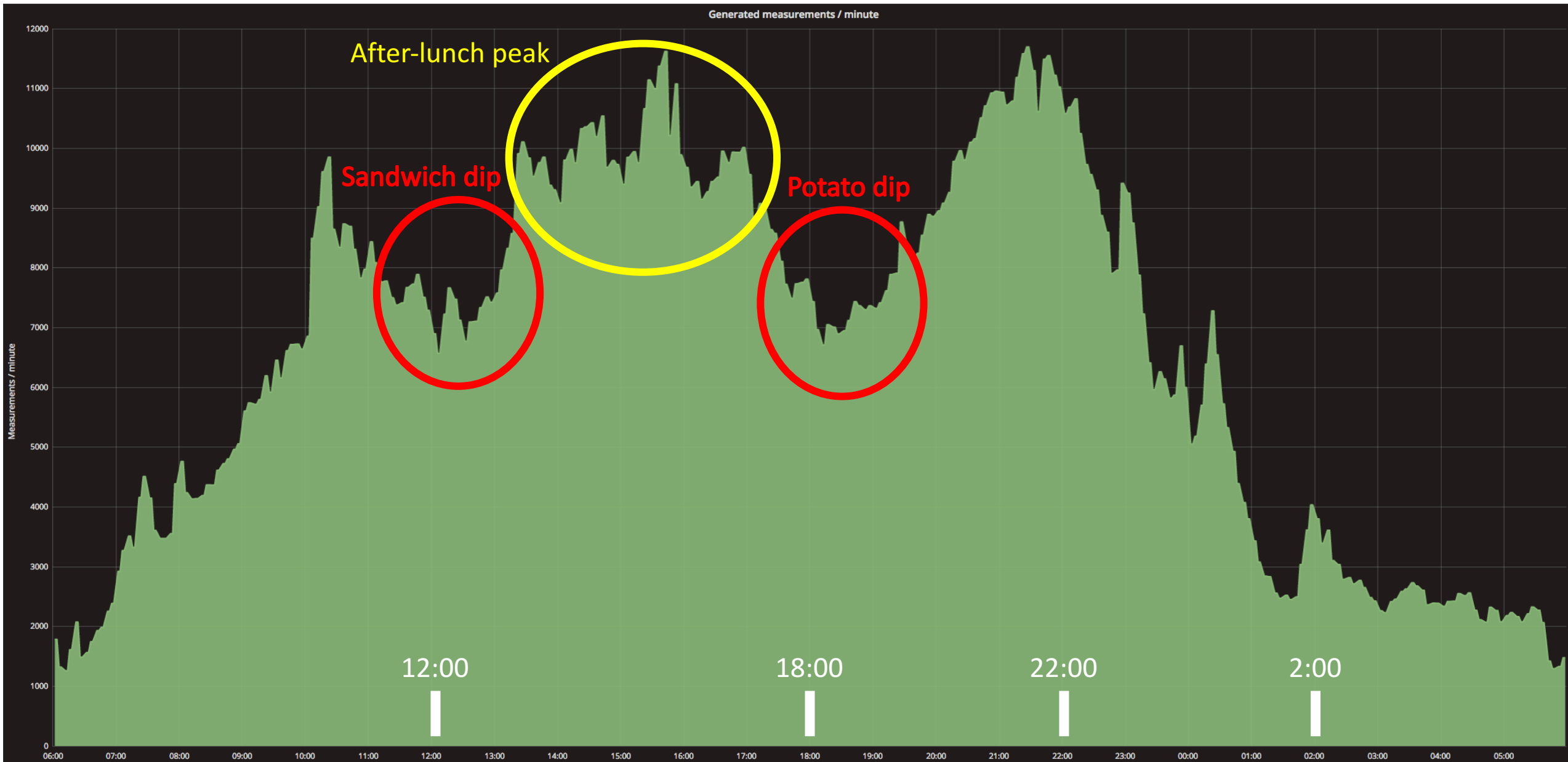
Production measurements



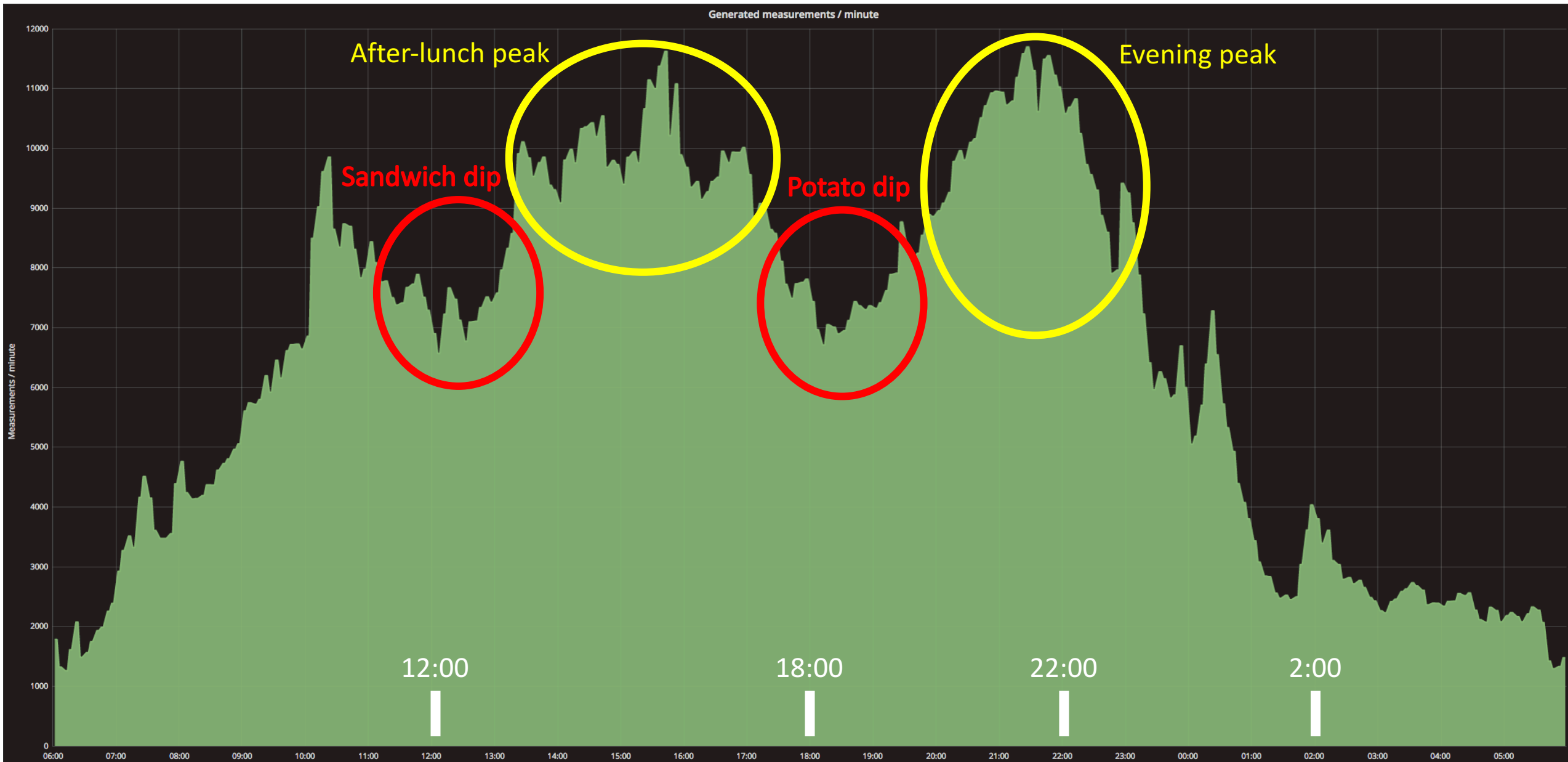
Production measurements



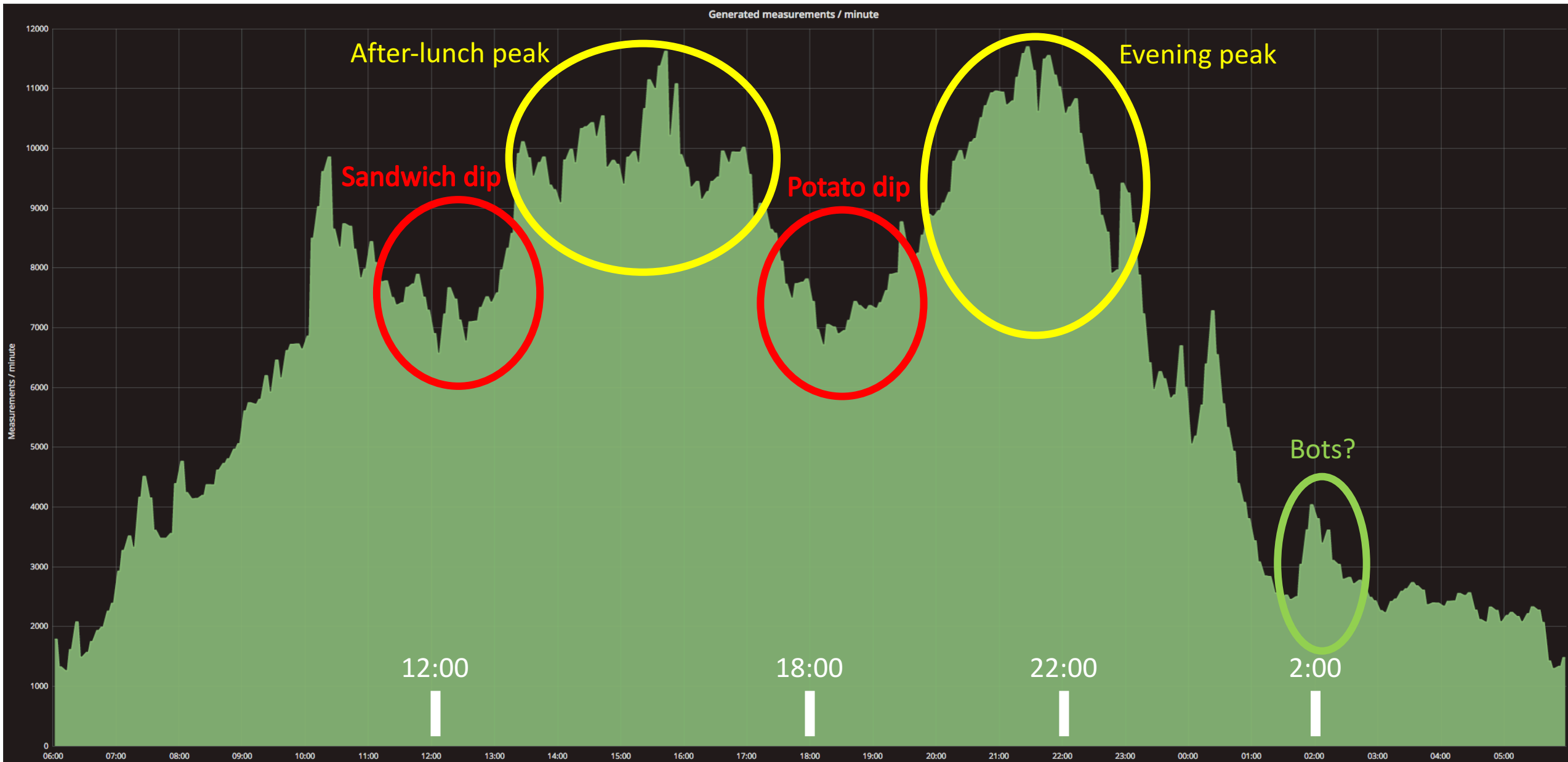
Production measurements



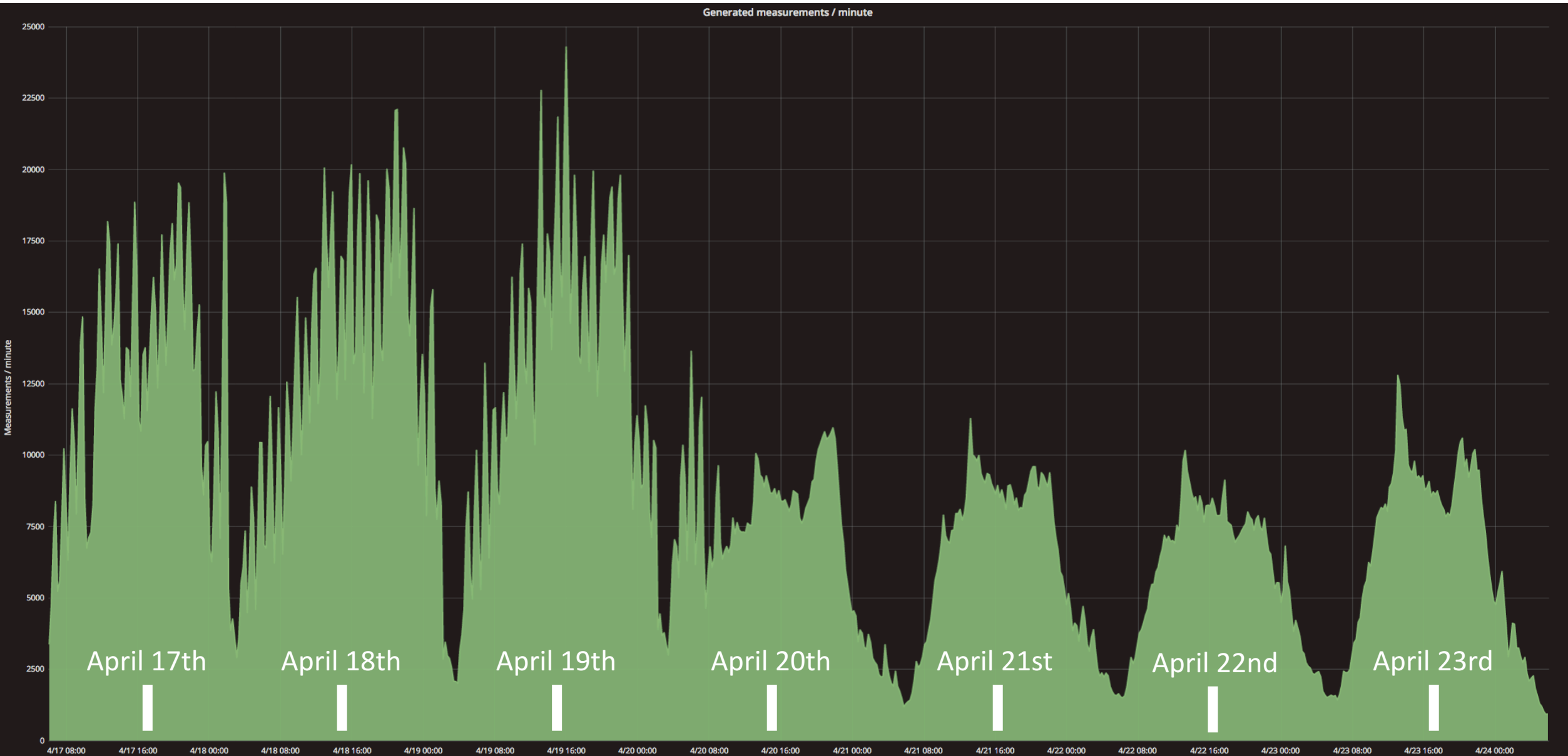
Production measurements



Production measurements

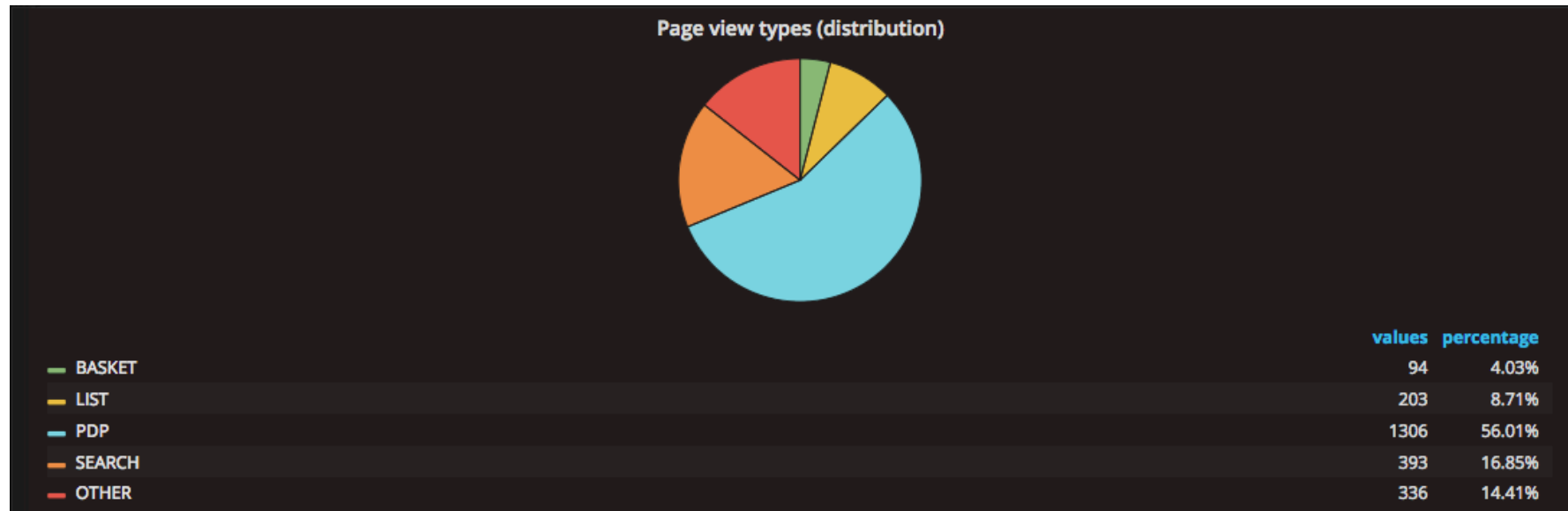


Filtering bot traffic



Some numbers

- Currently generated measurements/minute: ~15 500
- Expected peak load with website fully tagged: 6 000 000 measurements/minute
- Average message size: 500 bytes
- Most visited page: product page with 75% of views



Kafka – cluster specs

- 3 nodes
 - 15x 126 GB disks
 - 3.6 TB RAM
 - 40 cores (2.3 GHz)
- First Kafka users at bol.com

Kafka – lessons learned

- Problem: disks spent a very long time on I/O operations

Kafka – lessons learned

- **Problem:** disks spent a very long time on I/O operations
- **Cause:** property “flush.messages” set to 0

Kafka – lessons learned

- **Problem:** disks spent a very long time on I/O operations
- **Cause:** property “flush.messages” set to 0

flush.messages

This setting allows specifying an interval at which we will force an fsync of data written to the log. For example if this was set to 1 we would fsync after every message; if it were 5 we would fsync after every five messages. In general we recommend you not set this and use replication for durability and allow the operating system's background flush capabilities as it is more efficient. This setting can be overridden on a per-topic basis (see [the per-topic configuration section](#)).

Kafka – lessons learned

- **Problem:** disks spent a very long time on I/O operations
- **Cause:** property “flush.messages” set to 0

flush.messages

This setting allows specifying an interval at which we will force an fsync of data written to the log. For example if this was set to 1 we would fsync after every message; if it were 5 we would fsync after every five messages. In general we recommend you not set this and use replication for durability and allow the operating system's background flush capabilities as it is more efficient. This setting can be overridden on a per-topic basis (see [the per-topic configuration section](#)).

Kafka – lessons learned

- **Problem:** disks spent a very long time on I/O operations
- **Cause:** property “flush.messages” set to 0
- **Lesson:** do not change a default property without a concrete justification

flush.messages

This setting allows specifying an interval at which we will force an fsync of data written to the log. For example if this was set to 1 we would fsync after every message; if it were 5 we would fsync after every five messages. In general we recommend you not set this and use replication for durability and allow the operating system's background flush capabilities as it is more efficient. This setting can be overridden on a per-topic basis (see [the per-topic configuration section](#)).

Applications – Real-time trending products

- Calculating products which are trending per 10 minute intervals compared to a reference period of 4 hours

Meest populair op dit moment



NIVEA SUN Protect &
Hydrate Hydratere...

€ 14,91



6 - 99 jaar
Intex Frame Pool
Zwembad - 220 x 150 ...

€-69,99 € 59,99



Bestron AFT760Z -
Tower Ventilator - ...

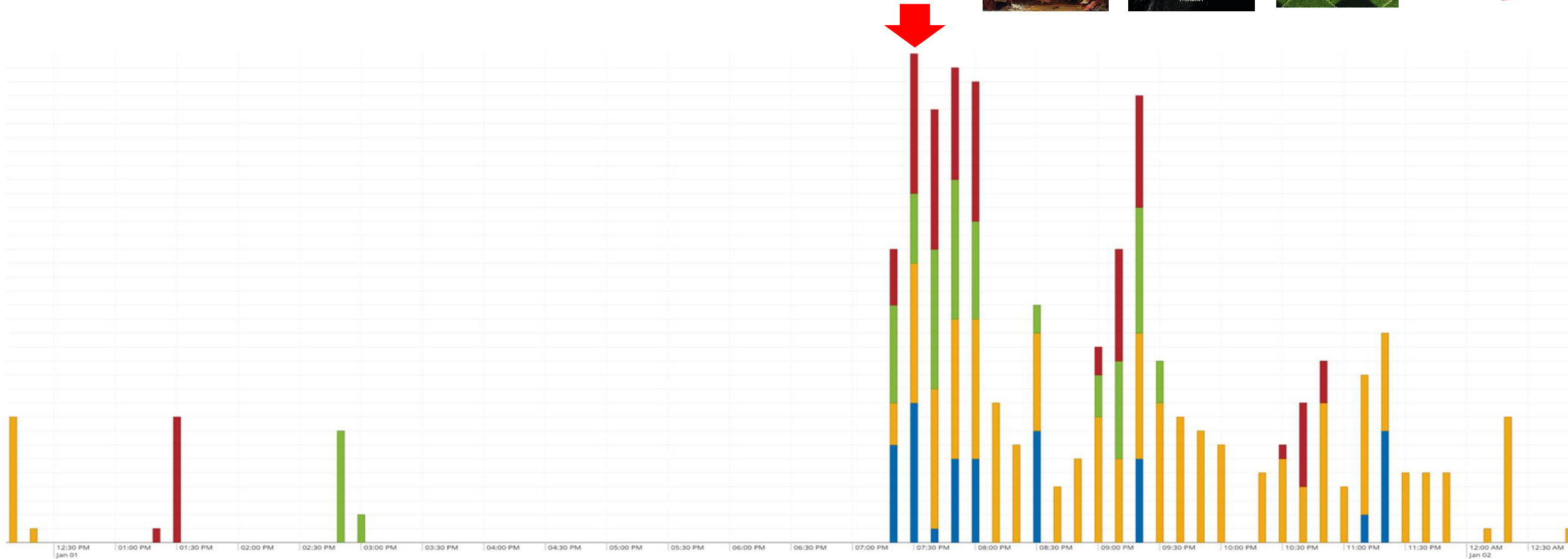
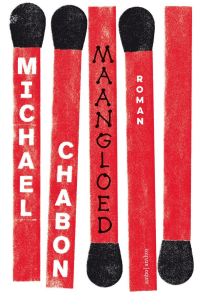
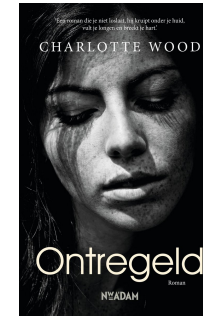
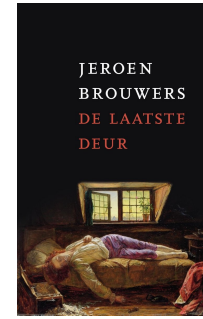
€-47,99 € 38,95



Jon Bernthal
The Accountant (Blu-
ray)

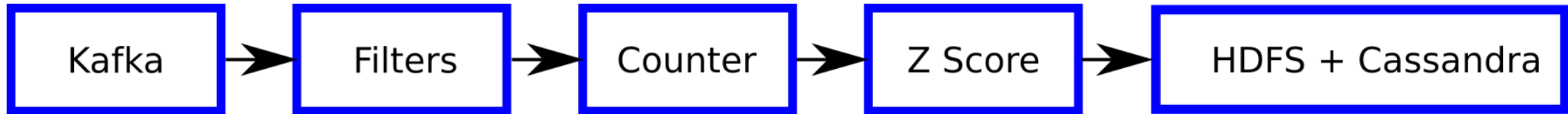
€ 22,99

Real time trending



19:00 19:30

How do we do it?



Persisting data

- Define Schema in: Apache AVRO
 - Nice schema language (IDL)
 - Generates Java classes
 - Supports schema evolution



- Persist in files: Apache Parquet
 - Compresses well
 - Great on read use cases
 - Write straight from Avro classes.
 - See `o.a.parquet.avro.AvroParquetWriter`



Persisting data

- Apache AVRO
 - Standard byte[] for the record.
 - Needs the original Schema on read
- What if the Schema changes?
 - Kafka keeps the events for weeks
- AVRO-1704
 - Serialize record to a “Message” byte[]
 - schema fingerprint (hash)
 - pluggable schema ‘database’

Future plans

Future plans

- Tag the entire webshop
- Tag the mobile app/ the seller dashboard
- Fully enable in-view measurements (currently at 1%)
- Provide tooling for working with the data
- Generate live reporting in real-time
- ...

Thanks for listening

Questions?

Ivan Budinčević

ibudincevic@bol.com

bol.com

de winkel van ons allemaal