



# Snowplow at de Bijenkorf

de Bijenkorf 

# Agenda

- ✧ Introduction
- ✧ Why do event tracking?
- ✧ Architecture
- ✧ Use cases
- ✧ Lessons learned
- ✧ Questions

# Introduction

## ◆ Niels Reijmer

- Project role:
  - Data analyst

## ◆ Andrei Scorus



- Project role:
  - Main ETL developer

## De Bijenkorf

- ◆ 145 year old high-end department stores
- ◆ New course: closing 5 out of 12 stores
- ◆ Focus on premium personal service!

# Why do event tracking?

- ✦ Stay ahead of the market
- ✦ Adds flexibility to the analysis possibilities
- ✦ Makes advanced analysis possible
- ✦ Focus on premium service also online: recommendations

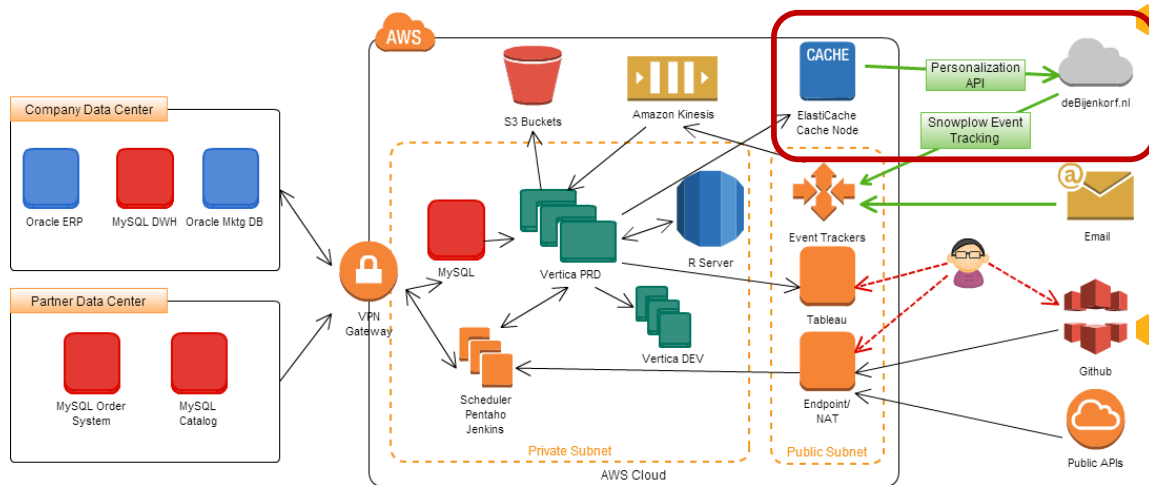
# What is Snowplow?



- The Snowplow enrichment process processes raw events from a collector and
- Cleans up the data into a format that is easier to parse / analyse
- Enriches the data (e.g. infers the location of the visitor from his / her IP address)
- Stores the cleaned, enriched data

# Architectural overview

6



## AWS

- S3
- Kinesis
- Elasticache
- Elastic Beanstalk
- EC2
- DynamoDB

## Open Source

- Snowplow Event Tracker
- Rundeck Scheduler
- Jenkins Continuous Integration
- Pentaho PDI

## Other

- HP Vertica
- Tableau
- Github
- RStudio Server

# Snowplow implementation

7

1. Tracker provided by Snowplow and fired using Google Tag Manager
2. Collector in Elastic Beanstalk
3. Enricher in Kinesis
4. Storage in the Vertica environment
5. Data modeling
6. Analytics in R, Tableau, etc.



# Uses

- ✧ Reporting
- ✧ A/B test analysis
- ✧ Personalisation on the website
- ✧ Advanced analysis (next talk)



# Snowplow table

Columns	
app_id (Varchar)	mkt_source (Varchar)
platform (Varchar)	mkt_term (Varchar)
partition_key (Integer)	mkt_content (Varchar)
collector_tstamp (Timestamp)	mkt_campaign (Varchar)
dvce_tstamp (Timestamp)	context (Varchar)
event (Varchar)	se_category (Varchar)
event_id (Varchar)	se_action (Varchar)
txn_id (Integer)	se_label (Varchar)
name_tracker (Varchar)	se_property (Varchar)
v_tracker (Varchar)	se_value (Numeric)
v_collector (Varchar)	unstruct_event (Varchar)
v_etl (Varchar)	tr_orderid (Varchar)
user_id (Varchar)	tr_affiliation (Varchar)
user_ipaddress (Varchar)	tr_total (Numeric)
user_fingerprint (Varchar)	tr_tax (Numeric)
domain_userid (Varchar)	tr_shipping (Numeric)
domain_sessionidx (Integer)	tr_city (Varchar)
network_userid (Varchar)	tr_state (Varchar)
geo_country (Varchar)	tr_country (Varchar)
geo_region (Varchar)	ti_orderid (Varchar)
geo_city (Varchar)	ti_sku (Varchar)
geo_zipcode (Varchar)	ti_name (Varchar)
geo_latitude (Numeric)	ti_category (Varchar)
geo_longitude (Numeric)	ti_price (Numeric)
geo_region_name (Varchar)	ti_quantity (Integer)
ip_isp (Varchar)	pp_xoffset_min (Integer)
ip_organization (Varchar)	pp_xoffset_max (Integer)
ip_domain (Varchar)	pp_yoffset_min (Integer)
ip_netspeed (Varchar)	pp_yoffset_max (Integer)
page_url (Varchar)	useragent (Varchar)
page_title (Varchar)	br_name (Varchar)
page_referrer (Varchar)	br_family (Varchar)
page_urlscheme (Varchar)	br_version (Varchar)
page_urlhost (Varchar)	br_type (Varchar)
page_urlport (Integer)	br_renderengine (Varchar)
page_urlpath (Varchar)	br_lang (Varchar)
page_urlquery (Varchar)	br_features_pdf (Boolean)
page_urlfragment (Varchar)	br_features_flash (Boolean)
refr_urlscheme (Varchar)	br_features_java (Boolean)
refr_urlhost (Varchar)	br_features_director (Boolean)
refr_urlport (Integer)	br_features_quicktime (Boolean)
refr_urlpath (Varchar)	br_features_realplayer (Boolean)
refr_urlquery (Varchar)	br_features_windowsmedia (Boolean)
refr_urlfragment (Varchar)	br_features_gears (Boolean)
refr_medium (Varchar)	br_features_silverlight (Boolean)
	br_cookies (Boolean)
	br_colordepth (Varchar)

# Example of the data

10

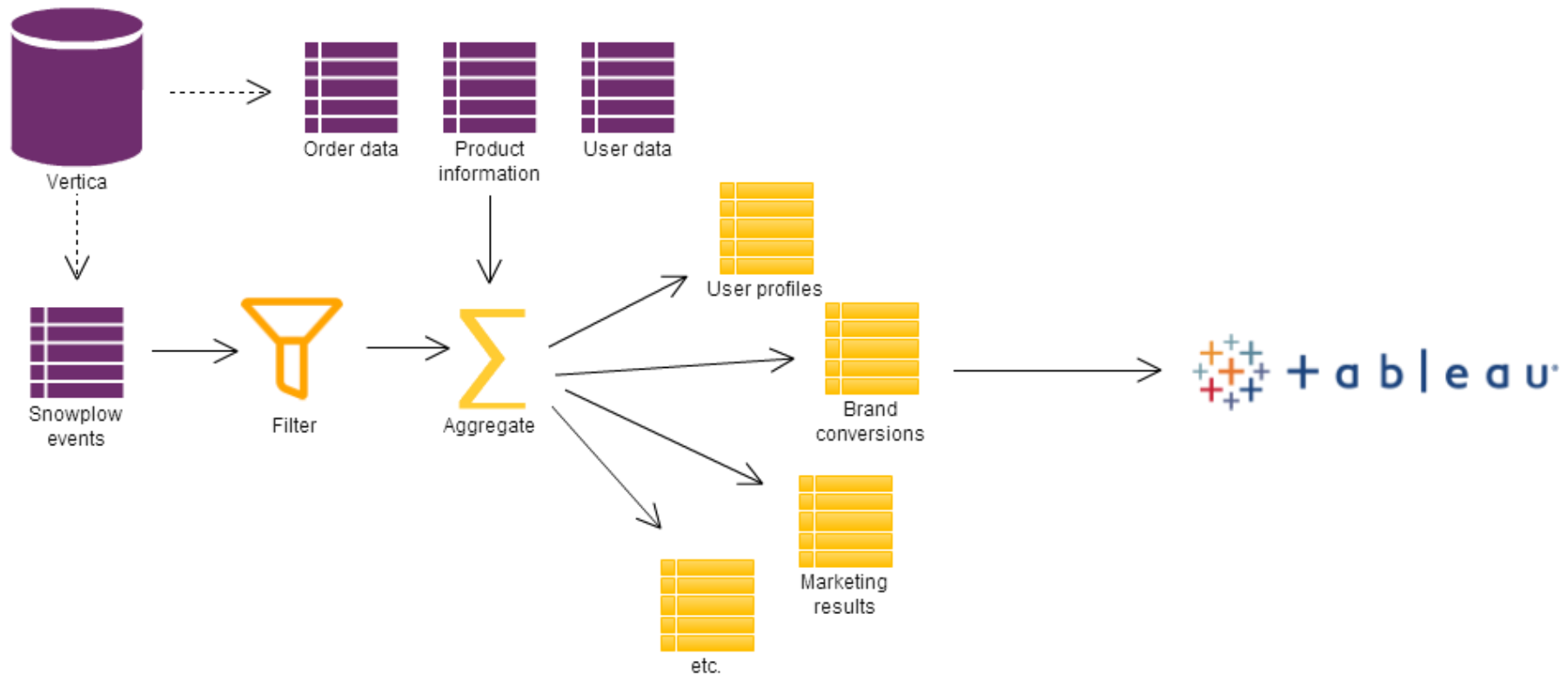
app_id	event	collector_tstamp	domain_userid	domain_sessionidx	user_id	page_url	page_title	refr_source	refr_medium	mkt_source	mkt_medium	se_category	se_action	se_label
desktop	struct	2015-05-13 09:18:51	bba125339460f243		2 (null)	http://www.debijenkorf.nl/Home/Dames/Jassen_ILI	(null)	(null)	(null)	(null)	(null)	Navigation	Click	("ty":"FlyOut", "w":"Heren", "ty": "...
desktop	page_view	2015-05-13 09:18:51	184f768e8f6acba3		6 p0przJk02YE=	http://www.debijenkorf.nl/Home	de Bijenkorf   Ontdek de nieuwe collectie	(null)	(null)	(null)	(null)	(null)	(null)	(null)
desktop	struct	2015-05-13 09:18:51	2f674f791b615097		1 (null)	http://www.debijenkorf.nl/Home/Outlet_ILI	(null)	(null)	(null)	(null)	(null)	Navigation	Click	("ty":"FlyOut", "w":"Heren", "ty": "...
desktop	transaction_item	2015-05-13 09:18:51	704dcf5c63a0e063		10YC3dwUrniA=	https://www.debijenkorf.nl/INTERSHOP/web/WFS/dbk-shop-...	(null)	(null)	(null)	(null)	(null)	(null)	(null)	(null)
desktop	struct	2015-05-13 09:18:51	5b9c2c65bce6521b		2 (null)	http://www.debijenkorf.nl/Home/Dames/Lingerie & nachtmod...	(null)	(null)	(null)	(null)	(null)	Navigation	Click	("ty":"Facet", "sw":"Off-Top", "it": "...
desktop	transaction_item	2015-05-13 09:18:51	704dcf5c63a0e063		10YC3dwUrniA=	https://www.debijenkorf.nl/INTERSHOP/web/WFS/dbk-shop-...	(null)	(null)	(null)	(null)	(null)	(null)	(null)	(null)
desktop	struct	2015-05-13 09:18:51	2f674f791b615097		1 (null)	http://www.debijenkorf.nl/Home/Outlet_ILI	(null)	(null)	(null)	(null)	(null)	Navigation	Click	("ty":"FlyOut", "w":"Heren", "ty": "...
desktop	page_view	2015-05-13 09:18:51	8af58ce039a82ad6		2 (null)	http://www.debijenkorf.nl/Home/Dames/Sieraden_ILI	Sieraden   Shop Sieraden bij de Bijenkorf	(null)	internal	(null)	(null)	(null)	(null)	(null)
desktop	transaction	2015-05-13 09:18:51	704dcf5c63a0e063		10YC3dwUrniA=	https://www.debijenkorf.nl/INTERSHOP/web/WFS/dbk-shop-...	(null)	(null)	(null)	(null)	(null)	(null)	(null)	(null)
desktop	struct	2015-05-13 09:18:51	2f674f791b615097		1 (null)	http://www.debijenkorf.nl/Home/Outlet_ILI	(null)	(null)	(null)	(null)	(null)	Navigation	Click	("ty":"FlyOut", "w":"Heren", "ty": "...
desktop	page_view	2015-05-13 09:18:51	704dcf5c63a0e063		10YC3dwUrniA=	https://www.debijenkorf.nl/INTERSHOP/web/WFS/dbk-shop-...	Shop	(null)	unknown	(null)	(null)	(null)	(null)	(null)
desktop	page_view	2015-05-13 09:18:51	7aee2f3086f855a		9 (null)	http://www.debijenkorf.nl/Home/Wonen/Woonaccessoires/O...	Areaware Cubebot • de Bijenkorf	Google	search	(null)	(null)	(null)	(null)	(null)
desktop	struct	2015-05-13 09:18:51	e6e78471599bf046		1 (null)	http://www.debijenkorf.nl/Home/Heren/Pakken & colberts ILI	(null)	(null)	(null)	(null)	(null)	CTAs	Click	("tune":"ChooseSize", "sku":"4497...

app_id	event	collector_tstamp	domain_userid	domain_sessionidx	user_id	page_url
desktop	struct	2015-05-13 09:18:51	bba125339460f243		2 (null)	http://www.debijenkorf.nl/Home/Dames/Jassen_ILI
desktop	page_view	2015-05-13 09:18:51	184f768e8f6acba3		6 p0przJk02YE=	http://www.debijenkorf.nl/Home
desktop	struct	2015-05-13 09:18:51	2f674f791b615097		1 (null)	http://www.debijenkorf.nl/Home/Outlet_ILI
desktop	transaction_item	2015-05-13 09:18:51	704dcf5c63a0e063		10YC3dwUrniA=	https://www.debijenkorf.nl/INTERSHOP/web/WFS/dbk-shop-...
desktop	struct	2015-05-13 09:18:51	5b9c2c65bce6521b		2 (null)	http://www.debijenkorf.nl/Home/Dames/Lingerie & nachtmod...
desktop	transaction_item	2015-05-13 09:18:51	704dcf5c63a0e063		10YC3dwUrniA=	https://www.debijenkorf.nl/INTERSHOP/web/WFS/dbk-shop-...
desktop	struct	2015-05-13 09:18:51	2f674f791b615097		1 (null)	http://www.debijenkorf.nl/Home/Outlet_ILI
desktop	page_view	2015-05-13 09:18:51	8af58ce039a82ad6		2 (null)	http://www.debijenkorf.nl/Home/Dames/Sieraden_ILI
desktop	transaction	2015-05-13 09:18:51	704dcf5c63a0e063		10YC3dwUrniA=	https://www.debijenkorf.nl/INTERSHOP/web/WFS/dbk-shop-...
desktop	struct	2015-05-13 09:18:51	2f674f791b615097		1 (null)	http://www.debijenkorf.nl/Home/Outlet_ILI
desktop	page_view	2015-05-13 09:18:51	704dcf5c63a0e063		10YC3dwUrniA=	https://www.debijenkorf.nl/INTERSHOP/web/WFS/dbk-shop-...
desktop	page_view	2015-05-13 09:18:51	7aee2f3086f855a		9 (null)	http://www.debijenkorf.nl/Home/Wonen/Woonaccessoires/O...

Not very useful this way

# Data and dashboards

11



## Why snowplow?

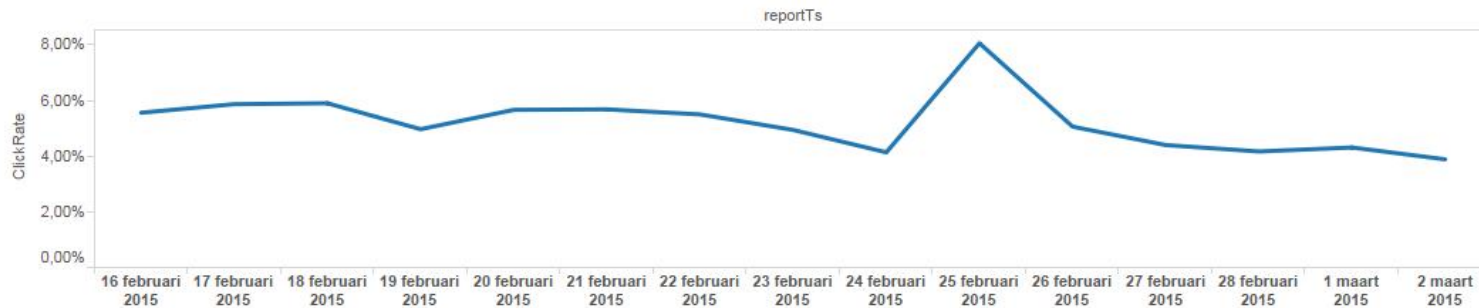
- Data can be combined in many ways due to the granularity
- Any question can be answered

## We do not rebuild Google analytics

# Example of a dashboard

12

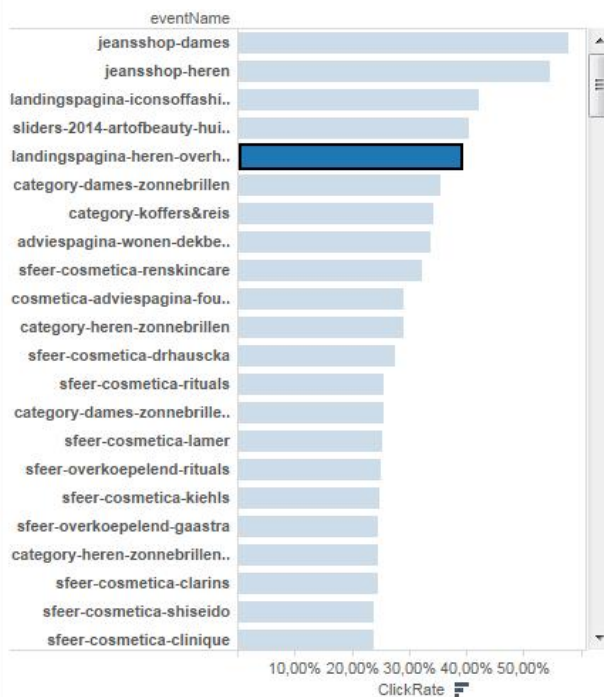
## Overview



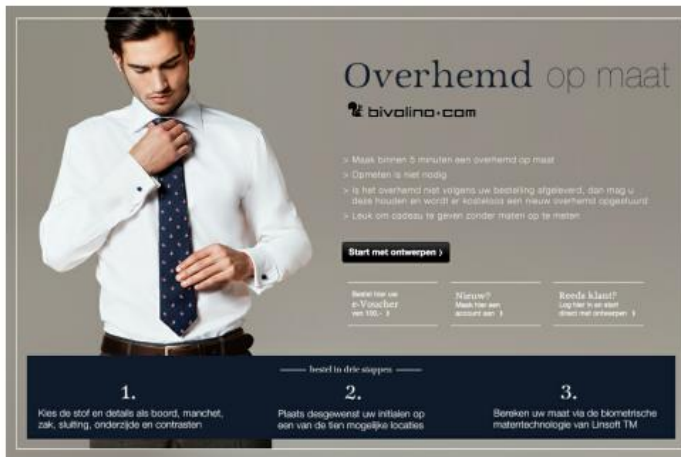
Month, Day, Year of r..

- ☒ (All)
- ☒ 16 februari 2015
- ☒ 17 februari 2015
- ☒ 18 februari 2015
- ☒ 19 februari 2015
- ☒ 20 februari 2015
- ☒ 21 februari 2015
- ☒ 22 februari 2015
- ☒ 23 februari 2015
- ☒ 24 februari 2015
- ☒ 25 februari 2015
- ☒ 26 februari 2015
- ☒ 27 februari 2015
- ☒ 28 februari 2015
- ☒ 1 maart 2015
- ☒ 2 maart 2015

## Top campaigns



## Campaign image



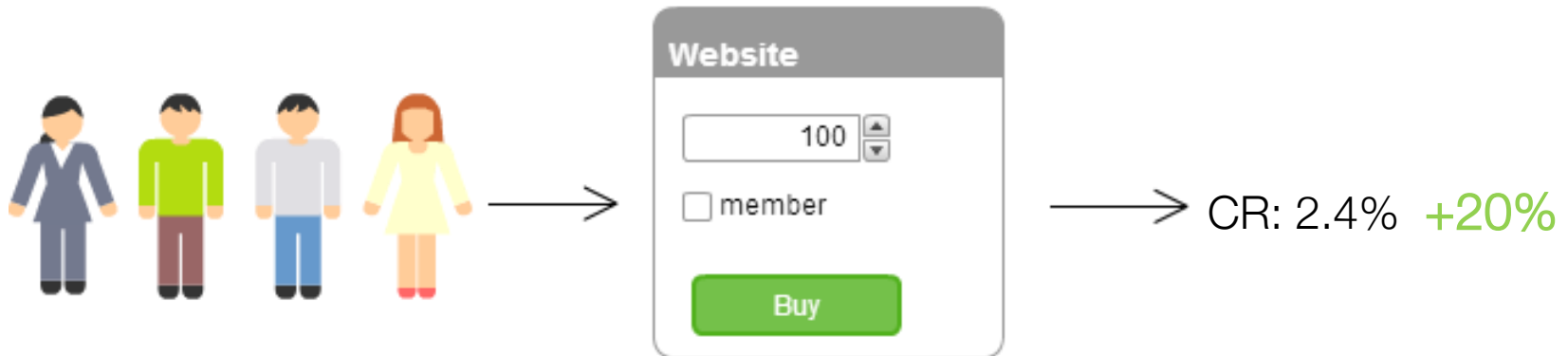
## Metrics for landingspagina-heren-overhemdopmaat-bivolino

Load	DisplayRate	ClickRate
994,0	97,69%	39,03%

## Averages

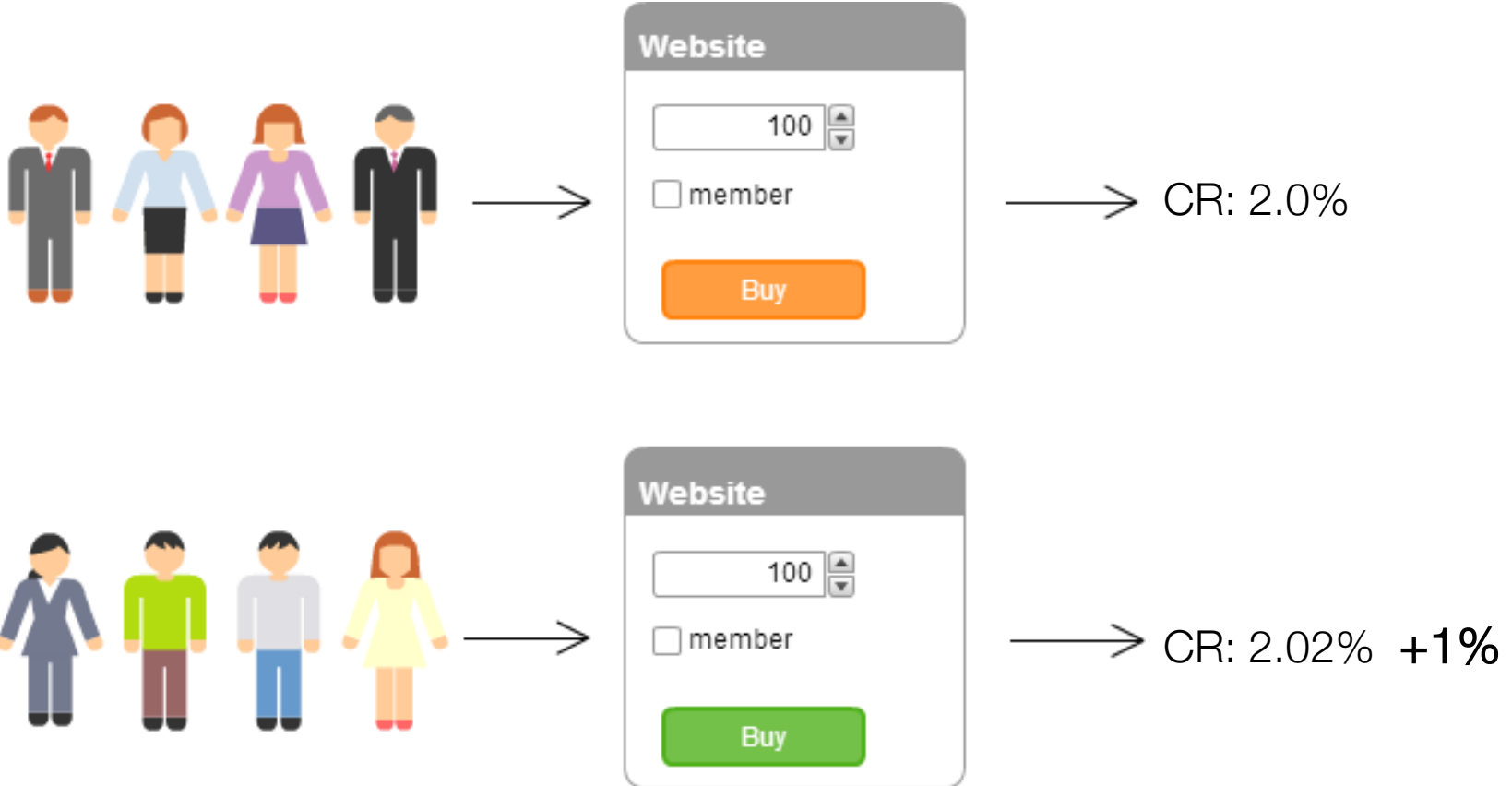
Average # of loads	DisplayRate	ClickRate
3.157	79,17%	5,27%

# A/B test data



# A/B test data

14

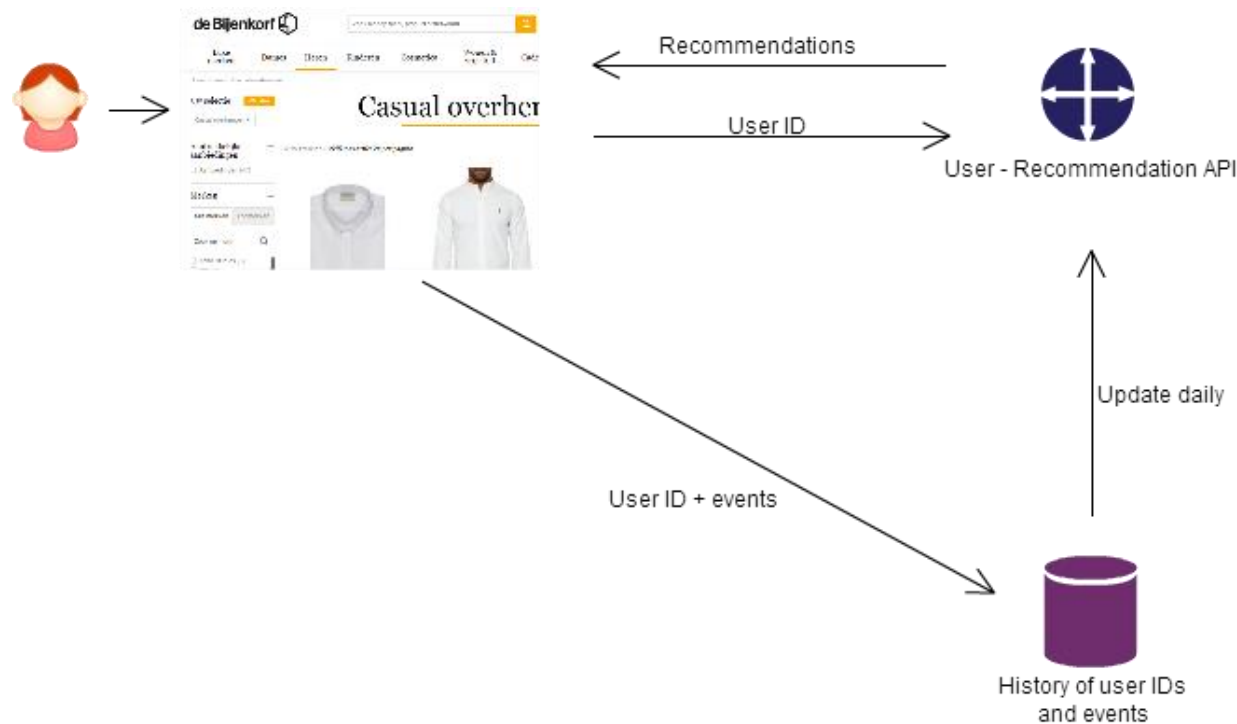


🟡 Further analysis in snowplow to determine what happened

# Recommendations

# Overview

16



- Something to recommend: 50.000 products
- Ability to make models: R, python, sql
- Historical data on each individual user
- Recognize users over sessions



# Data availability



Aggregated



Individual data 😊

Limited rows per user

Major cold-start problem



Individual data 😊

A lot of rows per user 😊

Minor cold-start problem 😊

# What data for example

18

Merken

Alle merken Topmerken

Zoek op merk

☐ Acne Studios (4)  
☐ Anerkjendt (3)  
☐ Antony Morato (6)  
☐ Armani (43)  
☐ Brax (14)  
☐ Calvin Klein (13)  
☐ Cheap Monday (6)  
☐ Denham (12)  
☐ Diesel (53)  
☐ Edwin (10)  
☐ Filippa K (2)


Tallenaat

☐ 27 (1)  
☐ 28 (83)  
☐ 29 (68)  
☐ 30 (312)  
☐ 31 (276)  
☐ 32 (402)  
☐ 33 (347)  
☐ 34 (393)  
☐ 35 (16)  
☐ 36 (363)

tapered ? slim ? straight ? relaxed ?

(440 artikelen) [Bekijk 240 artikelen per pagina](#)

Sorteer op Nieuwe collectie



Armani  
Armani Jeans  
J28 slim fit jeans faded mediumblauw  
139,95

Armani  
Armani Jeans  
J06 slim fit jeans met destroyed details  
159,95

Armani  
Armani Jeans  
J28 low rise slim fit jeans in donkerblauw  
119,95

de Bijenkorf

# Recognize the user in multiple sessions

- For historical data processing:
  - Snowplow has two user identifiers:

*	app_id	event	collector_tstamp	domain_userid	user_id	u
1	desktop	struct	2015-04-04 00:00:00	33de7a904af1d450	(null)	(n
2	desktop	page_view	2015-04-04 00:00:00	5fc32e4efaeaa3a2	(null)	Mi
3	desktop	page_ping	2015-04-04 00:00:00	b224594bac711a1f	BEE/kRWmA/A=	(n
4	desktop	page_view	2015-04-04 00:00:00	33de7a904af1d450	(null)	Mi
5	desktop	page_ping	2015-04-04 00:00:01	5fc32e4efaeaa3a2	(null)	(n
6	desktop	struct	2015-04-04 00:00:01	9f32ffc54fdabce2	(null)	(n
7	desktop	page_ping	2015-04-04 00:00:01	cc19ce855069c45b	(null)	(n
8	desktop	struct	2015-04-04 00:00:02	bce0fc3024e0be1f	(null)	(n
9	desktop	page_view	2015-04-04 00:00:03	5e6b2d98d7d07a37	wjwT0AgJ+RAMST9k7A4tDQ==	Mi
10	desktop	page_ping	2015-04-04 00:00:03	867975fcbadd32a	(null)	(n

- User\_id is entered on (soft) login

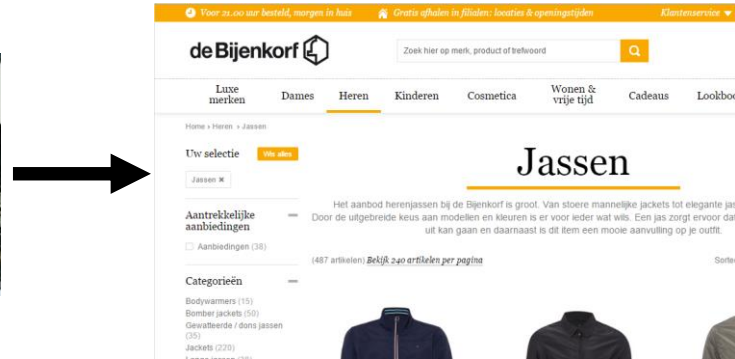
The challenge is to link the domain\_userid to the userid

# Recognize the user

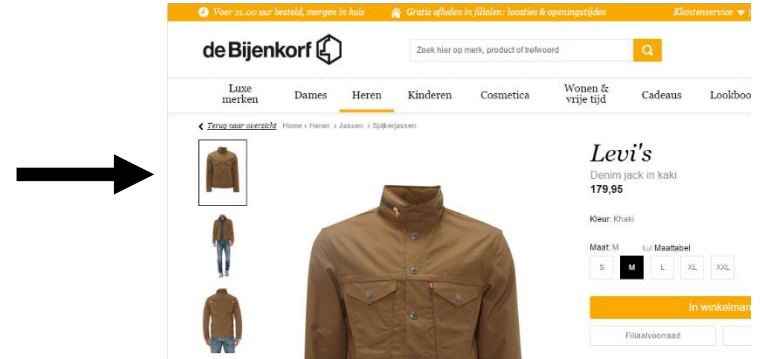
20



Yali S.

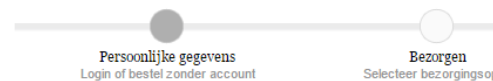


app_id	event	collector_tstamp	domain_userid	user_id
desktop	page_view	2015-04-04 00:04:10	4n4ly71c5m4573r	(null)



app_id	event	collector_tstamp	domain_userid	user_id
desktop	page_view	2015-04-04 00:04:14	4n4ly71c5m4573r	(null)

2 weeks later



1. Persoonlijke gegevens

Inloggen

Vul uw e-mailadres en wachtwoord in en uw bewaarde gegevens worden opgehaald.

E-mailadres

Wachtwoord [Wachtwoord vergeten?](#)

[Naar bezorging](#)

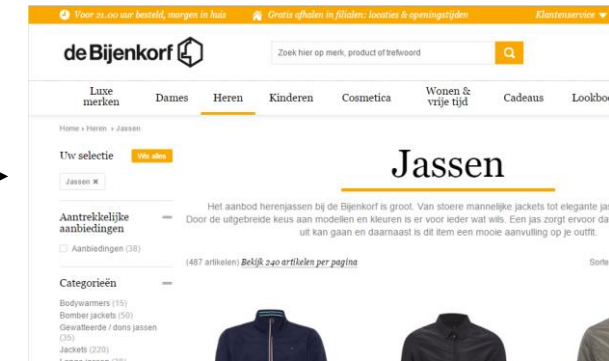
app_id	event	collector_tstamp	domain_userid	user_id
desktop	page_view	2015-04-04 00:04:23	4n4ly71c5m4573r	qQWLierVcxuvm7ybabULeA==

# Identify user from email

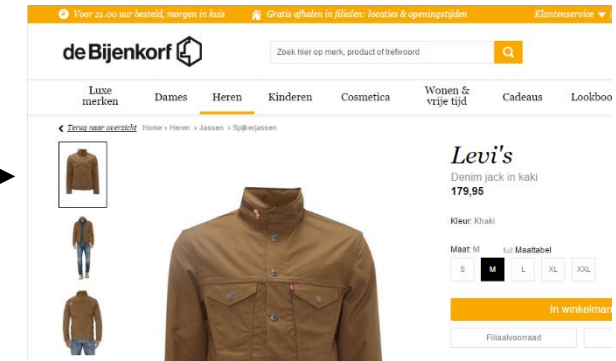
21



Yali S.



app_id	event	collector_tstamp	domain_userid	user_id
desktop	page_view	2015-04-04 00:04:10	4n4ly71c5m4573r	(null)

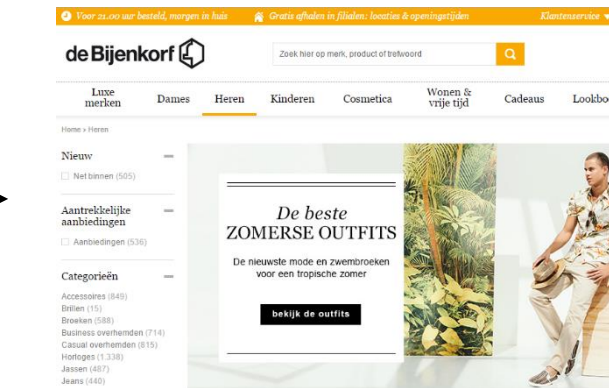


app_id	event	collector_tstamp	domain_userid	user_id
desktop	page_view	2015-04-04 00:04:14	4n4ly71c5m4573r	(null)

2 weeks later



Newsletter  
with id



app_id	event	collector_tstamp	domain_userid	user_id	url_parameter
desktop	page_view	2015-05-04 00:00:25	4n4ly71c5m4573r	(null)	emailID=xyz123

Email ID  
xyz123



User\_id  
qWier586\_kasd==



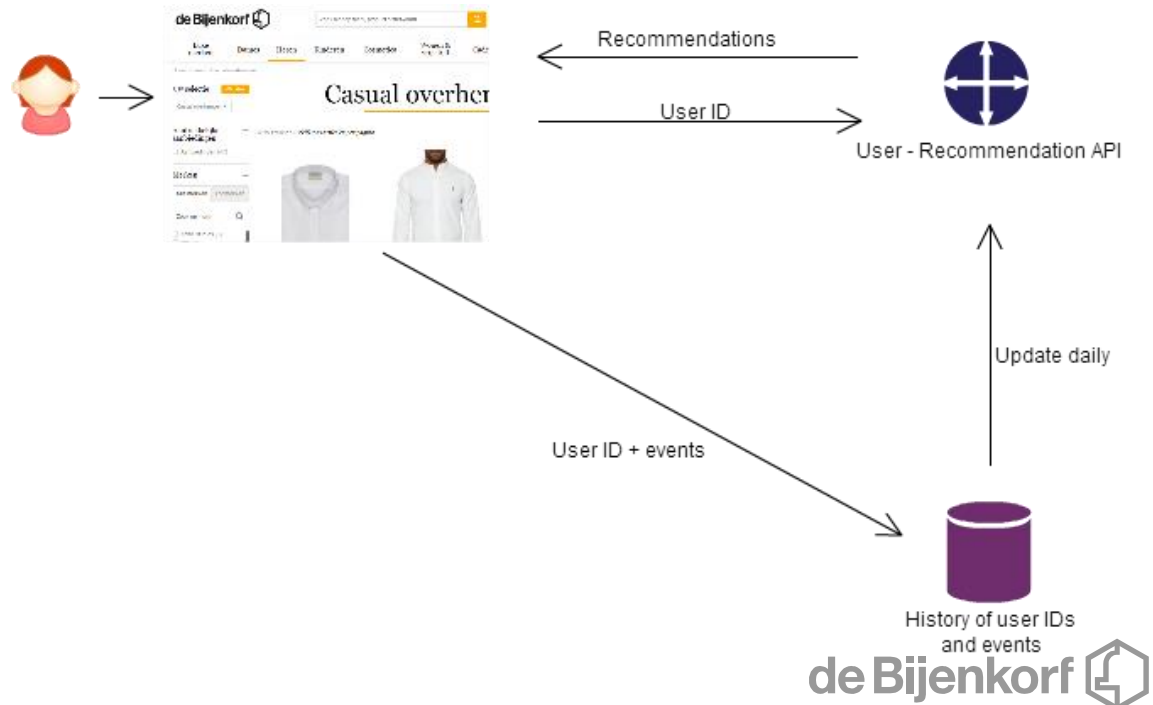
DWH

# Recognize user during the session

We can link the domain\_user to the userid

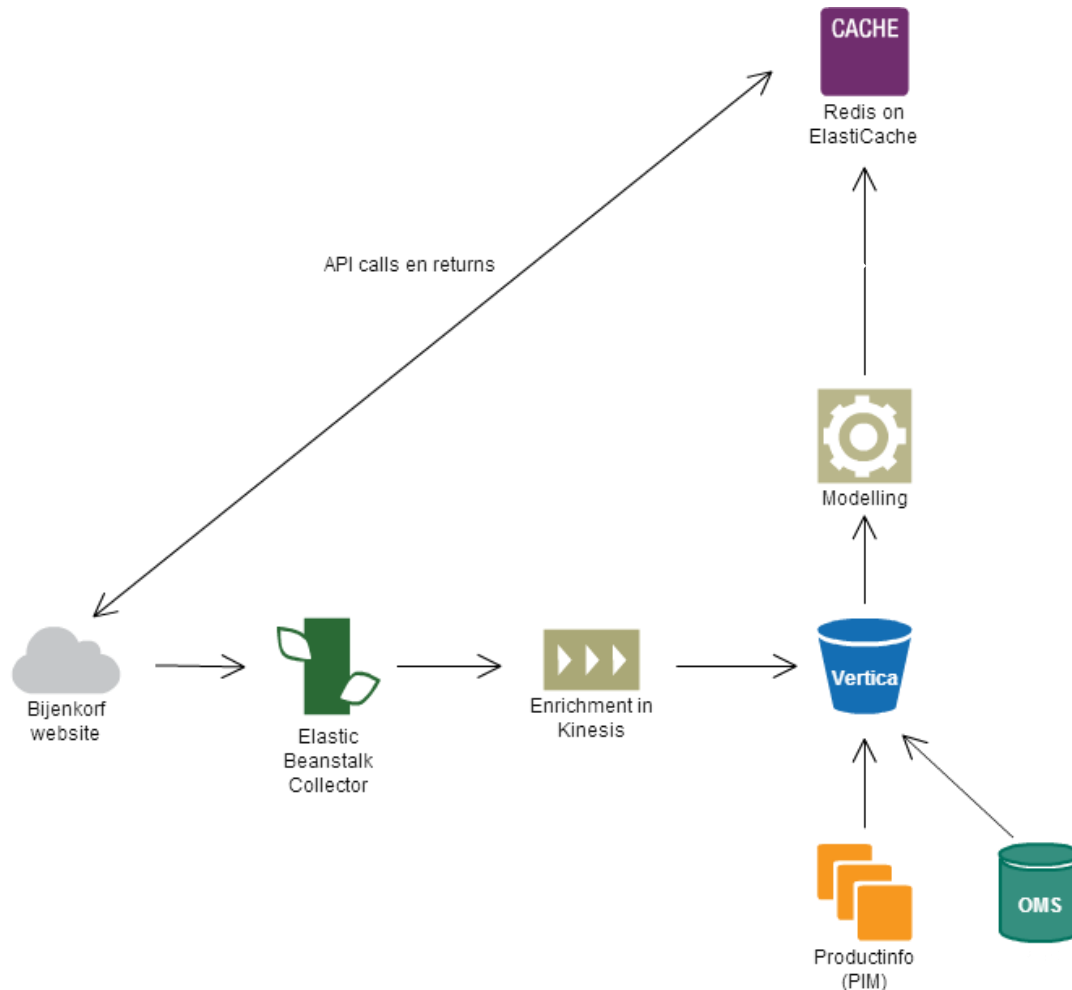
But we need to have a user ID to use for the API call:

- User is known from their account login
- User is already known from their snowplow id
- User is unknown
  - Nothing.....?



# Recommender dataflow in detail

23



- User is known from their account login
- User is already known from their snowplow id
- User is unknown
  - Nothing.....?

# Lessons learned

## Implementation

- ✦ Fairly easy to start, but highly customizable
- ✦ Documentation of your implementation should be in sync with the actual implementation.

## Reporting, A/B testing

- ✦ More insights possible, adds flexibility
- ✦ Need additional tools/resources to make it usable for non-technical people
- ✦ Google Analytics and Snowplow can complement each other

## Recommender

- ✦ Just start and keep it simple 😊



# Questions?