goto; conference

Please evaluate this talk via the mobile app!

The multifactor recommender system @bol.com

Barrie Kersbergen
expert software engineer
bkersbergen@bol.com

bol.com

Topics

- **develop an operating recommender system**
- **crucial customer behavioral factors involved**
- **impact of visual presentation of item recommended.**

bol.com

Barrie Kersbergen

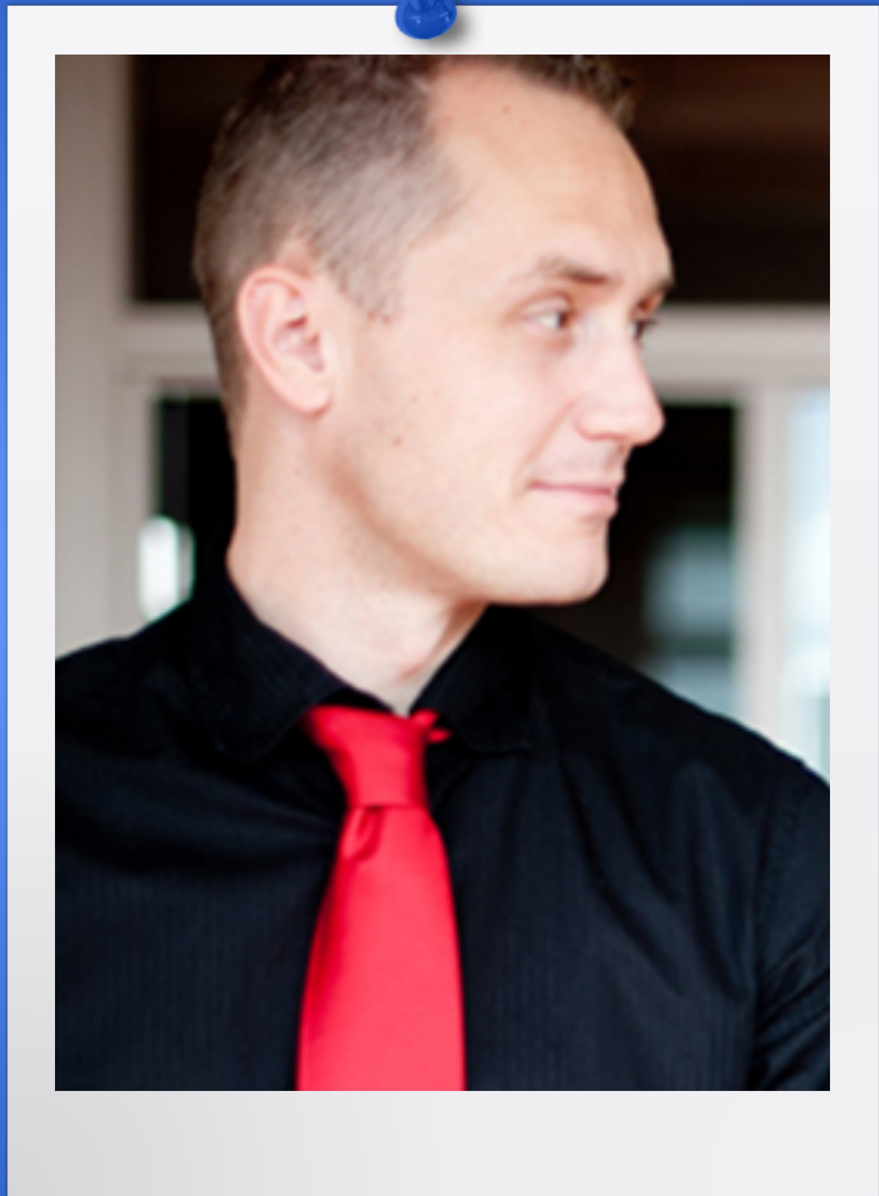
bkersbergen@bol.com

@krsbrgn

Computer Science

12 Yrs University Utrecht, CS/ Mathematics

Since 2010 @ Bol.com



bol.com

This is no hype

Thursday March 27th 2014

Thuiswinkel awards

Bol.com awarded
'Best webshop in the Netherlands'

"Bol.com has put major steps forward in the past year, regarding **personal recommendations and the use of big data**.

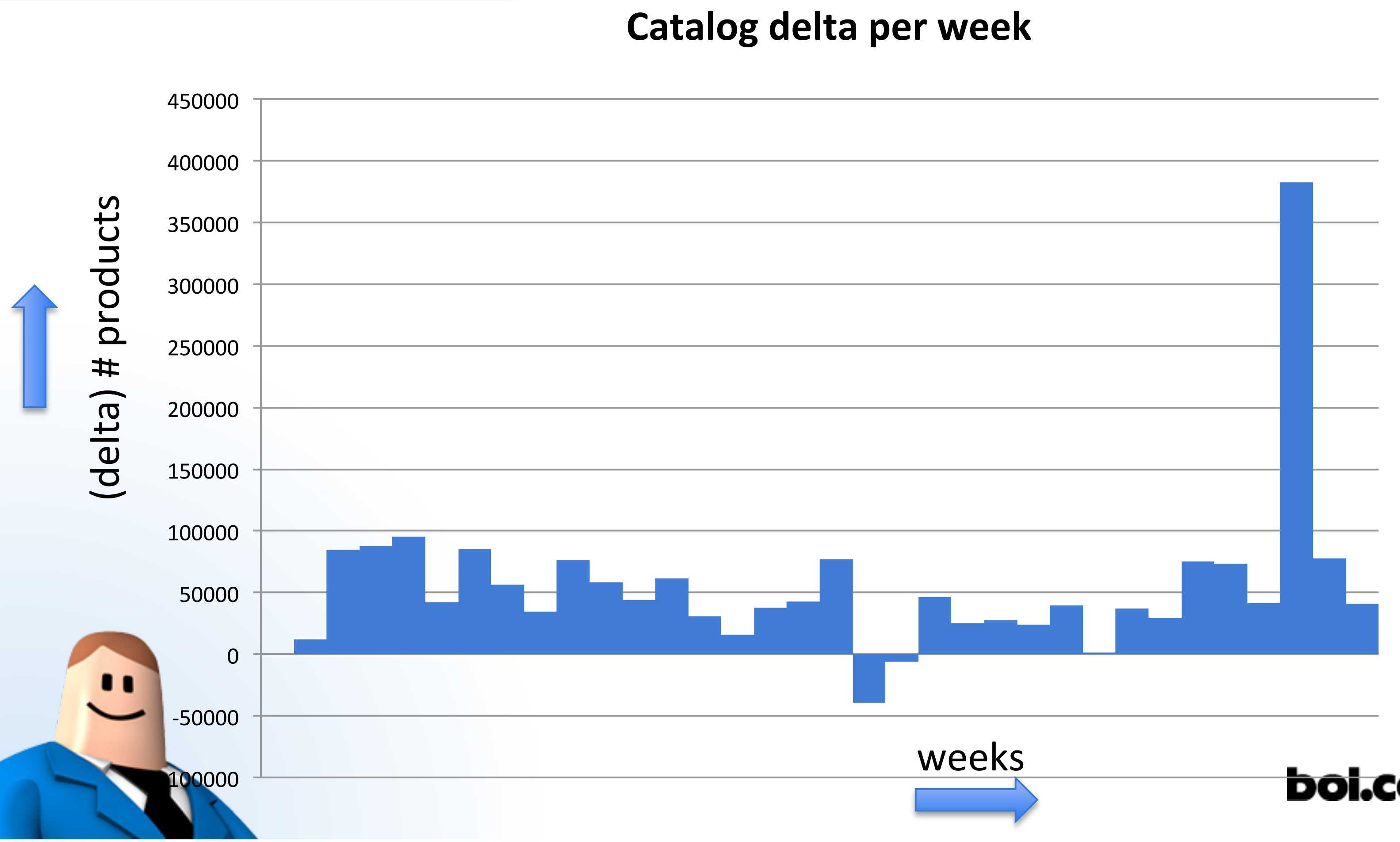
bol.com

Unique products per store

- **Average supermarket**
- **XL supermarket**

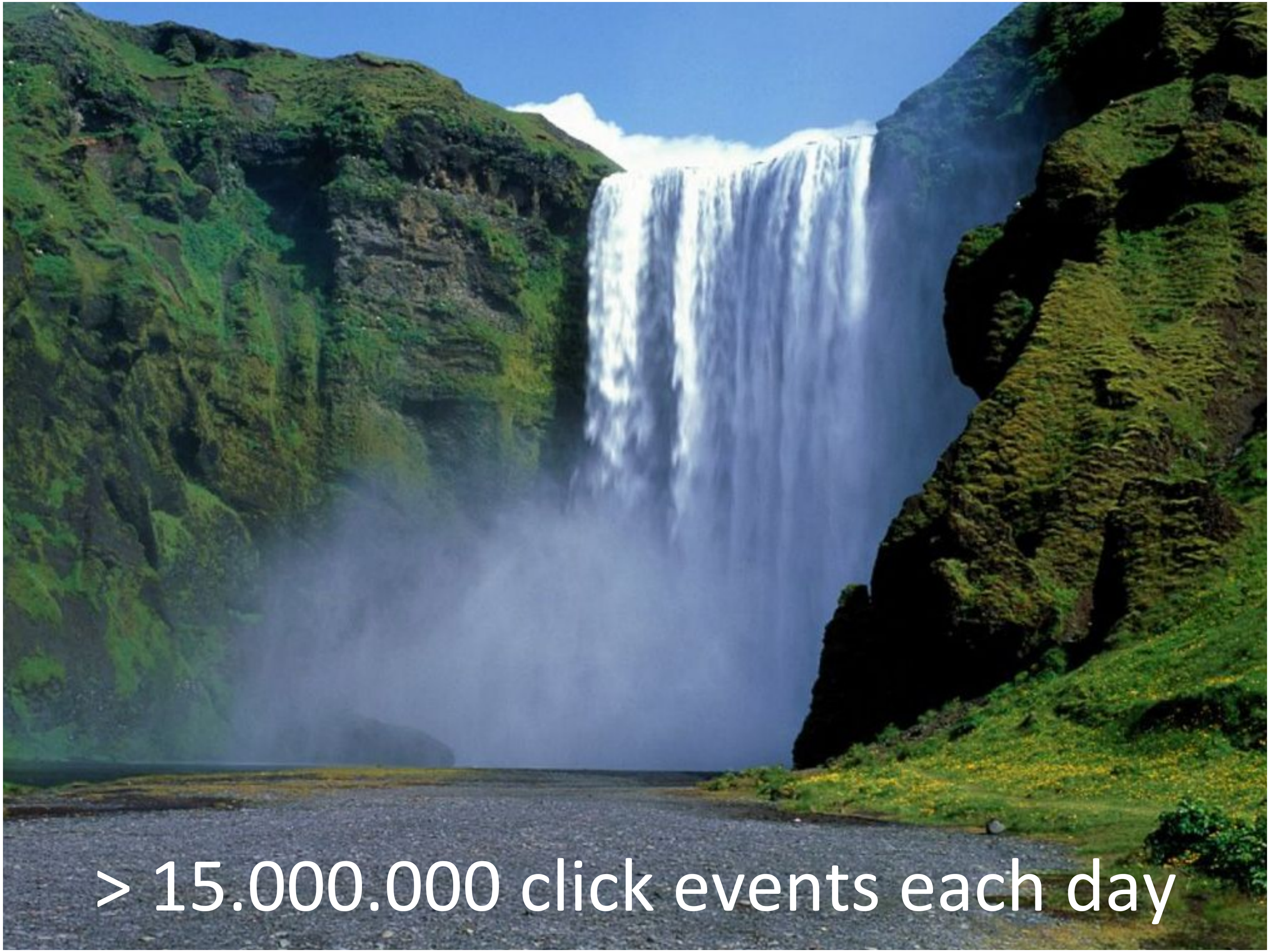
added products per week

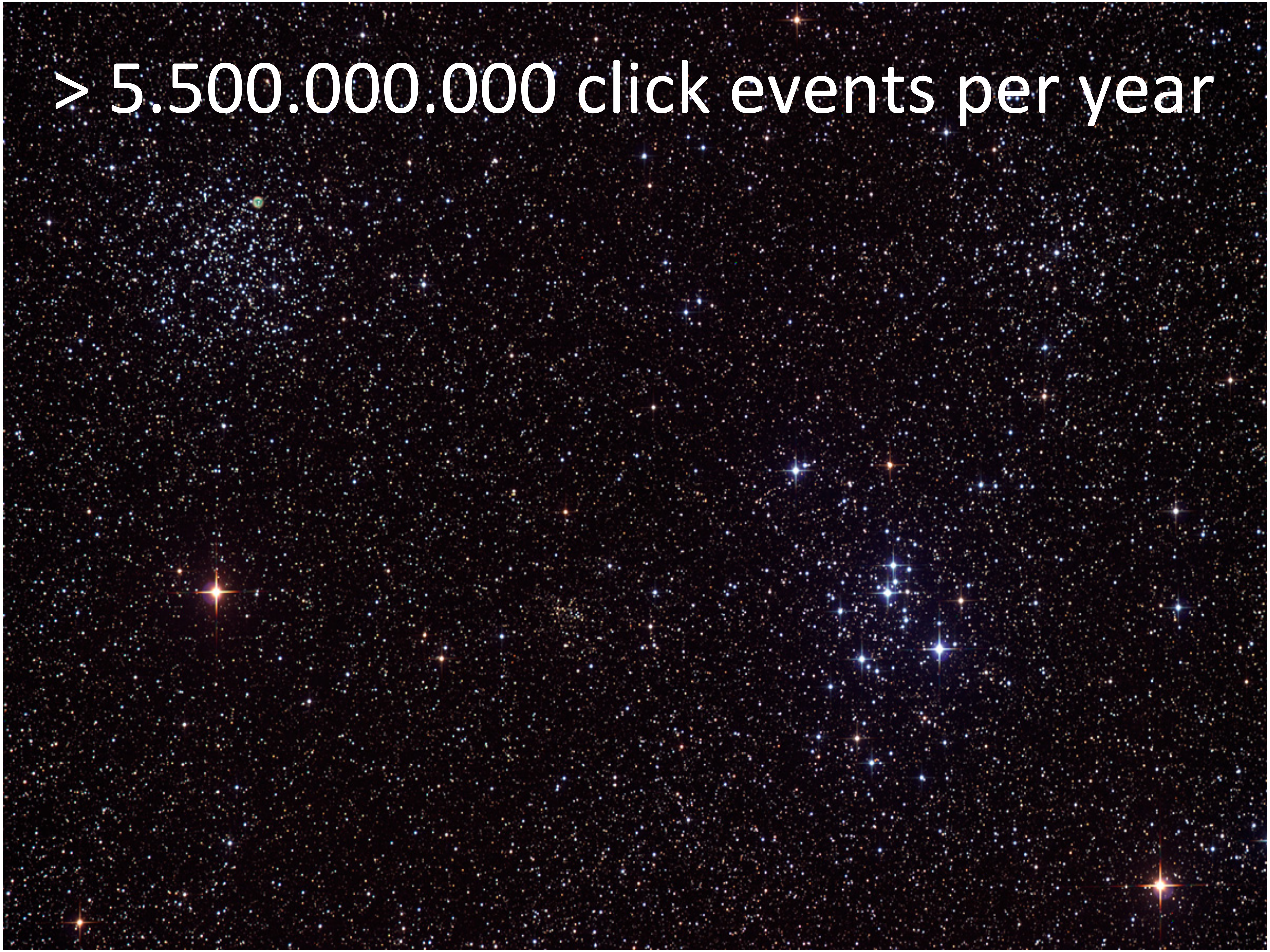
+ 1 XL supermarket every 4 days




13
12
11
09
> 8.000.000 products online

> 20.000.000 products total


> 15.000.000 click events each day


> 5.500.000.000 click events per year

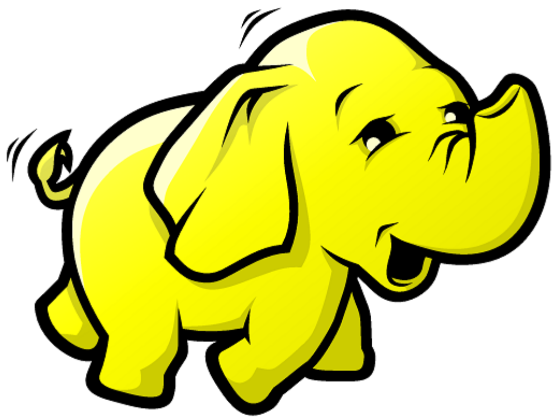
Behavioral factors give us clues

- **How long does he look at the product?**
- **Does he read the product reviews?**
- **Does he share the product on social media?**
- **Add item to wishlist**
- **What products does the visitor click?**
- **Was the clicked product a recommendation?**

bol.com

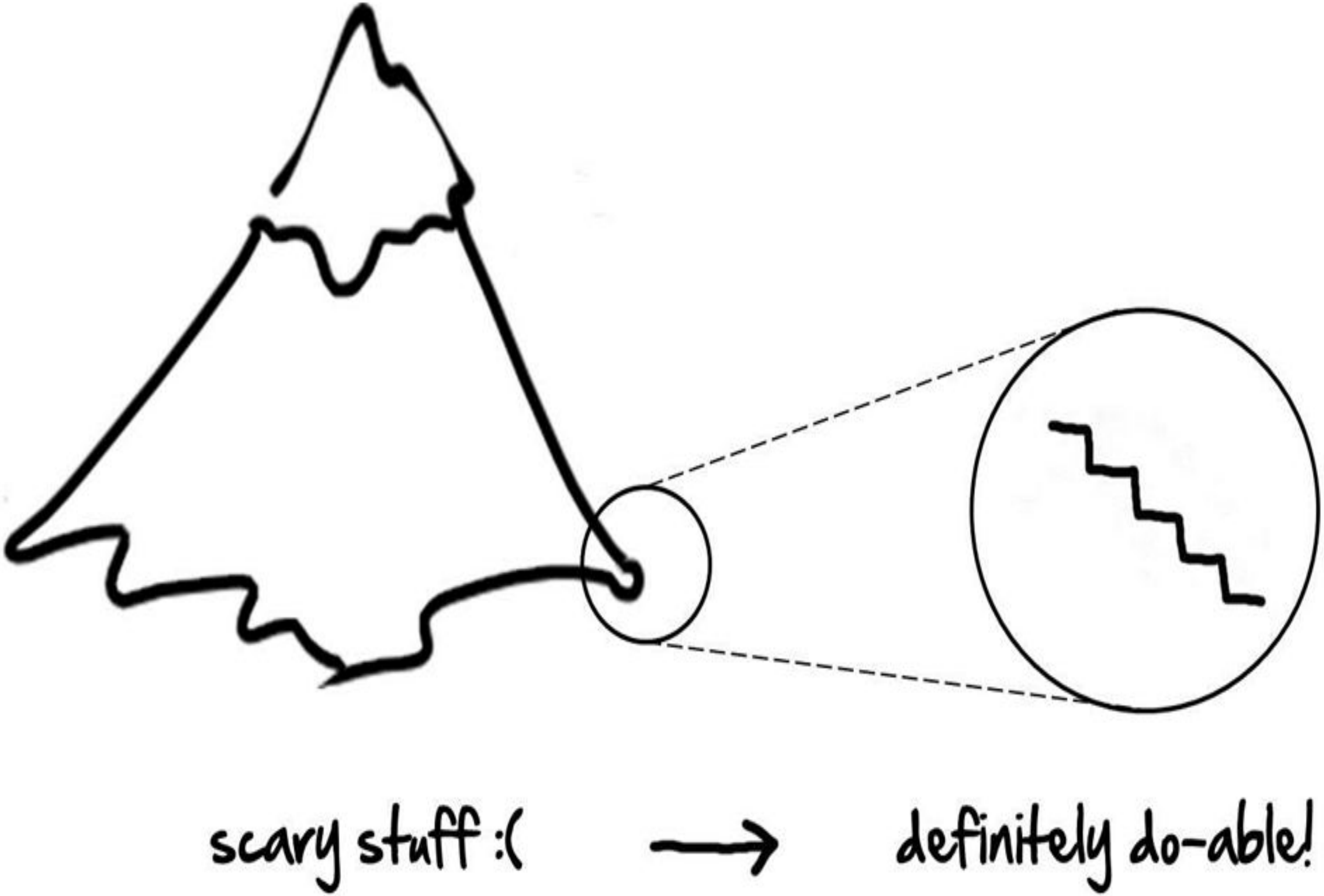
hadoop @bol.com

- **Development in 2009**
- **First applications in production (2010)**

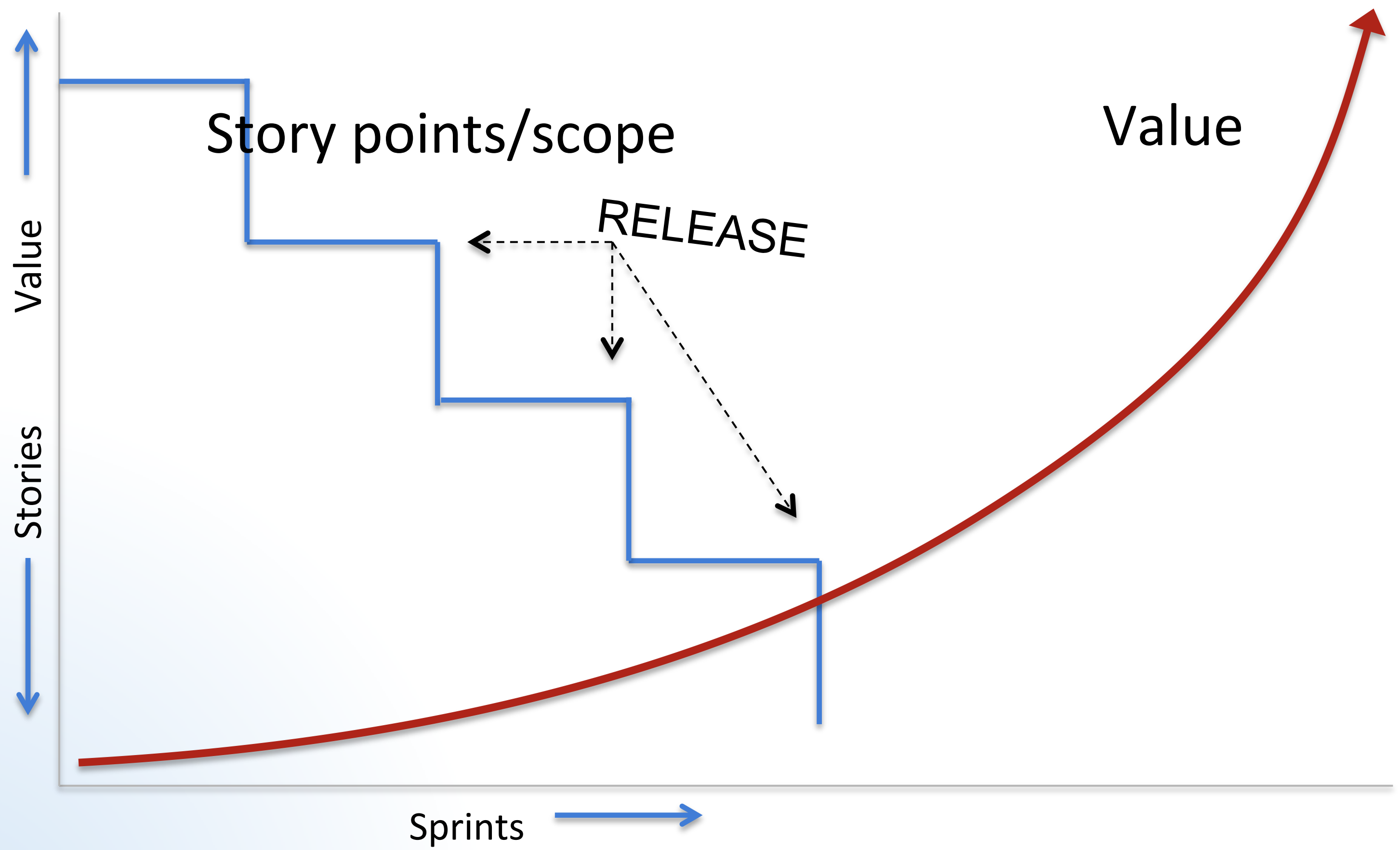


Why building from scratch?

- **Horizontal scalable**
- **Perform in “realtime”**
- **Handle our volume of data**
- **We can explain the workings of our system to our stakeholders. Together we can realize every business requirement that our stakeholders have.**

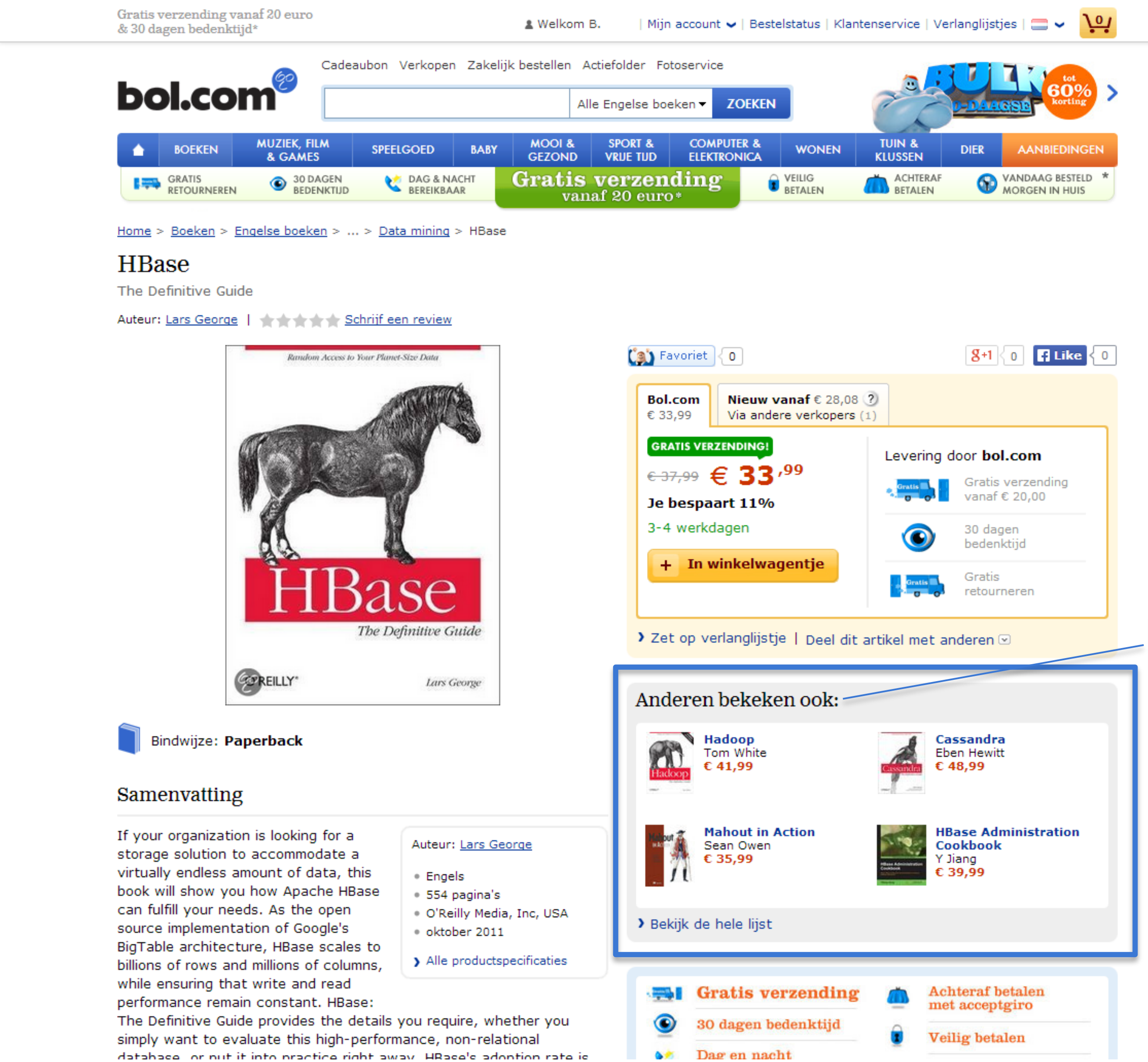

scary stuff :(
definitely do-able!

Stories burn down, value increases



bol.com

Recommendations



Other people also looked at:

How can we tell the success of these recommendations?

- **Visualize output data**
- **Offline recommender evaluation**
  - Run algorithms that express the numerical errors of the recommender output.

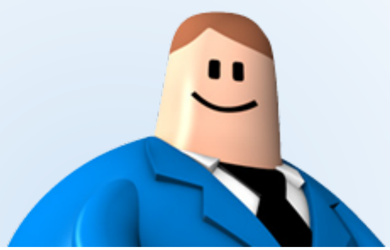
Root Mean Squared Error (RMSE)

The square root of the mean/average of the square of all of the error.

The use of RMSE is very common and it makes an excellent general purpose error metric for numerical predictions.

Compared to the similar Mean Absolute Error, RMSE amplifies and severely punishes large errors.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$



bol.com

How can we tell the success of these recommendations?

- **Is that all we can do?**

The gasoline recommender system

The gasoline recommender system

Requirements:

- **Support anonymous customers**
- **Realtime: response 100ms**
- **Low numerical error value**
- **Recommendation is successful if customer buys the item after seeing it first as a recommendation.**

The gasoline recommender system

- **Customer: new customer**
- **Cartype: Tow Truck**
- **Recommendation: Diesel**
- **Does customer buy item after seeing this as recommendation? : yes**

bol.com

The gasoline recommender system

- **Customer: new customer**
- **Cartype: sportscar**
- **Recommendation: Petrol (gasoline)**
- **Does customer buy item after seeing this as recommendation? : yes**

bol.com

The gasoline recommender system

- **Offline evaluation:**
  - Visual data inspection
  - Low error value

- **Online evaluation:**
  - Accurately predict the recommendations
  - Almost everything we recommend gets bought

The gasoline recommender system

Is this a good recommender system?

bol.com

What is a good recommendation?

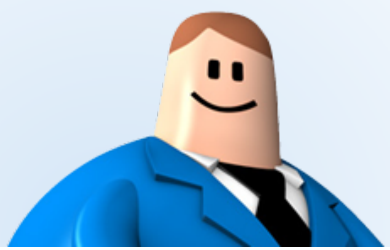
Common error metrics	Commerce
The accuracy to predict	Add value to customer experience

How can we tell the success of this recommender?

- **Visualize output data**
- **Offline recommender evaluation**
  - Run algorithms that express the error of the recommender output in a numerical form.

- **Long Term analyses**
  - Analyze the performance of the recommender over time.
    - Do visitors return after buying a recommendation?
  - Analyze behavior to learn more from our visitors.

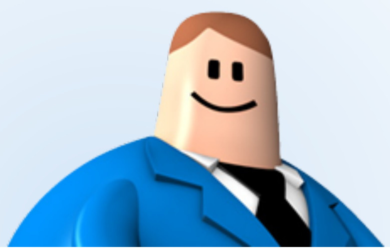
- **Online recommender evaluation**
  - Live user experiment ’s



Live user experiments?

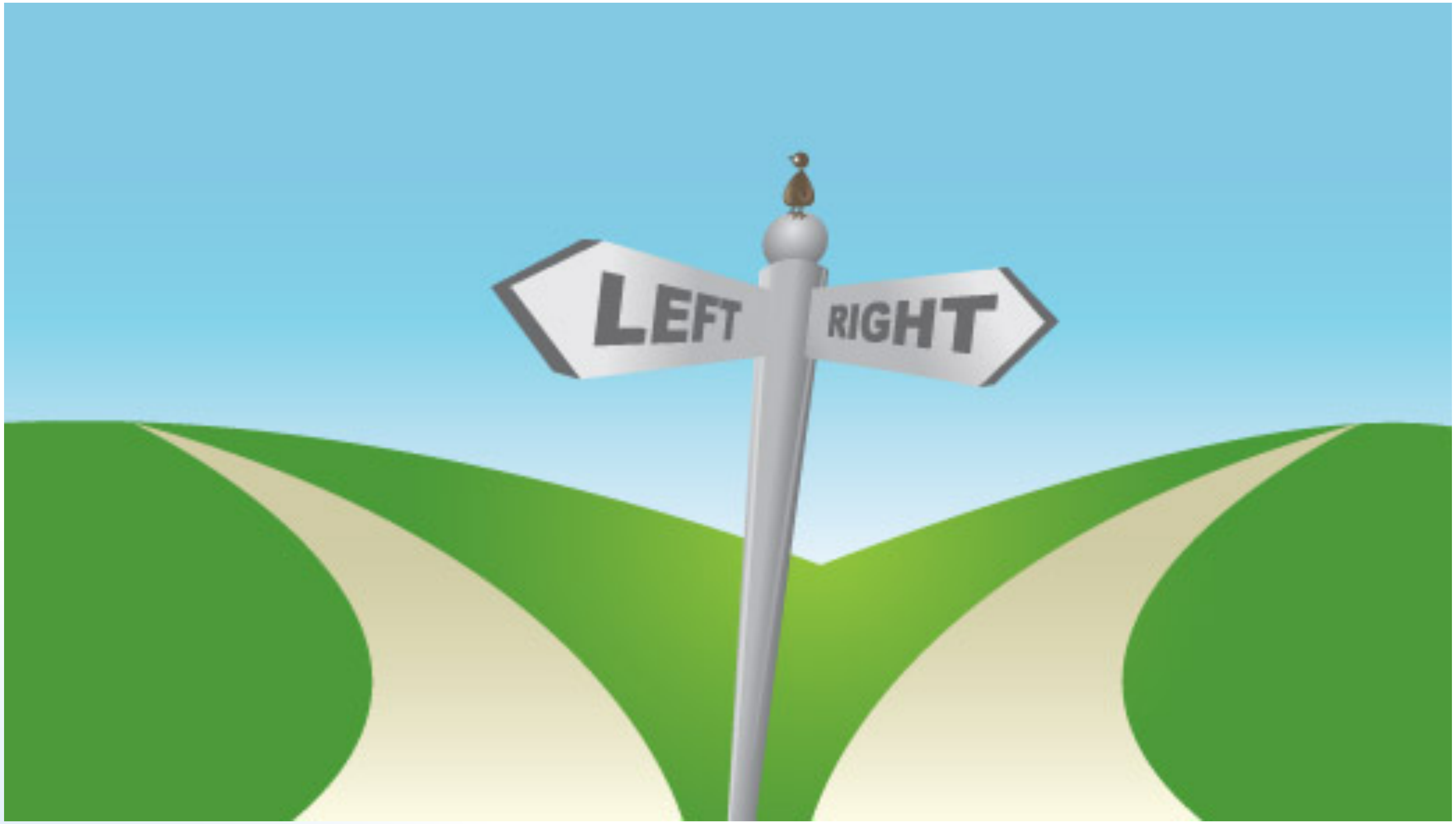
Recommendations = Data-driven decisions

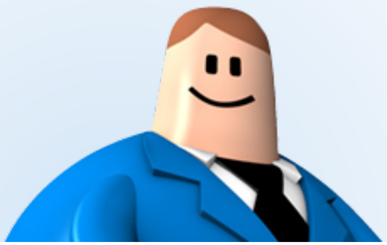
- **Optimize combinations of recommendation algorithms**
- **User Interface**
- **User Interaction Flows**



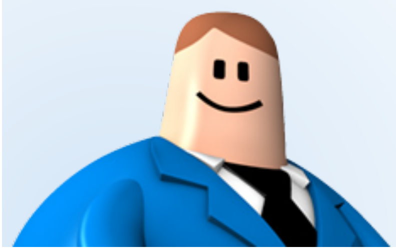
bol.com

What do visitors really prefer?





Experimentation framework

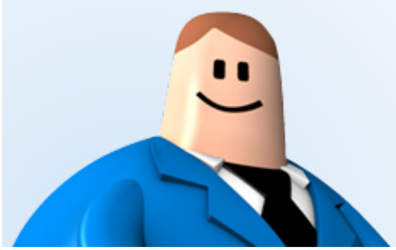


Experimentation results

Impact of presentation

Impact of presentation

Recommendation interaction



bol.com

Recommendations

General personal reco's

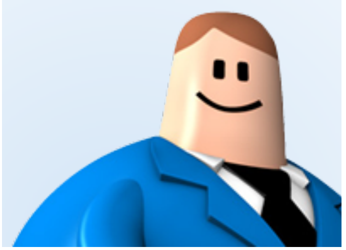
Recommend a deal

Recommend wishlist items

Recommend price change

Recommend search queries

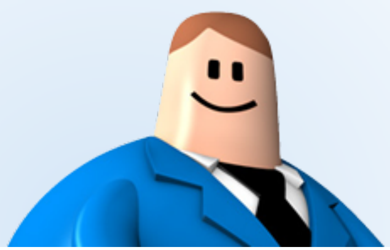
Recommend a category to the visitor



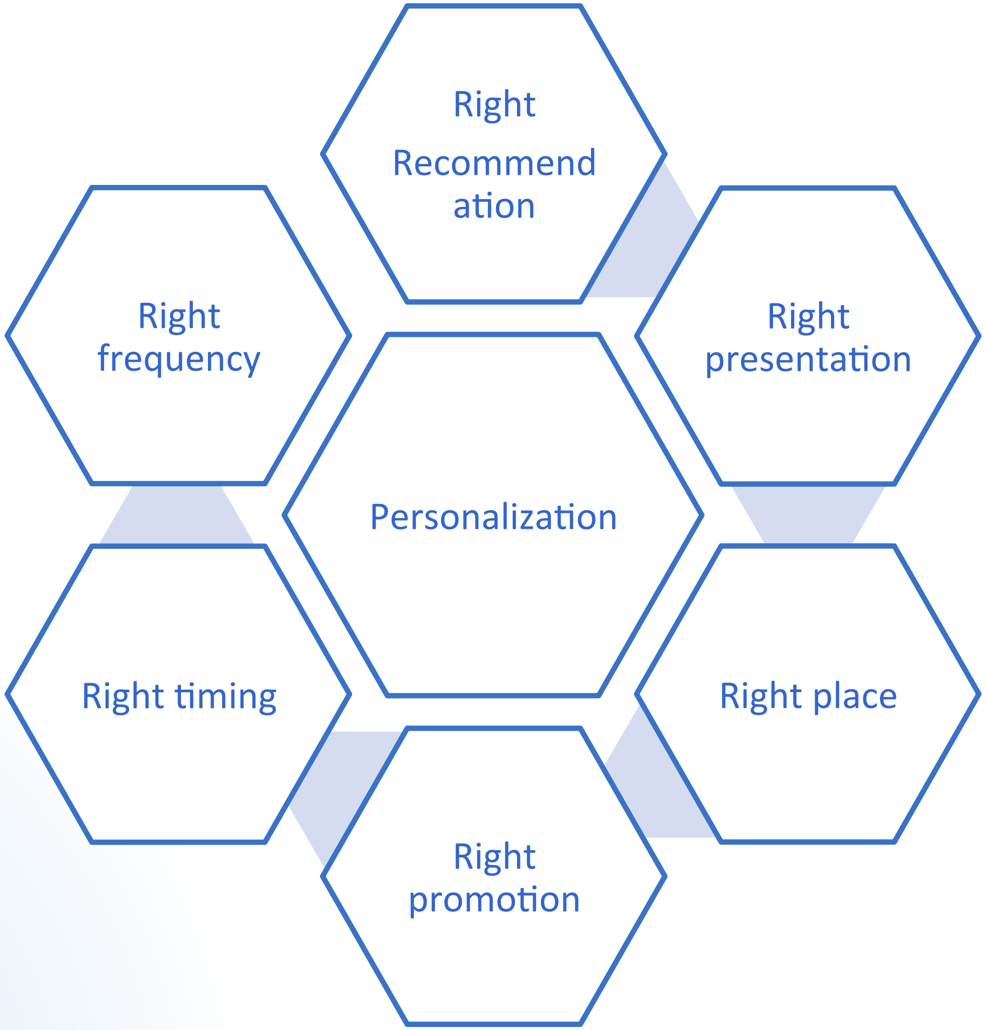
bol.com

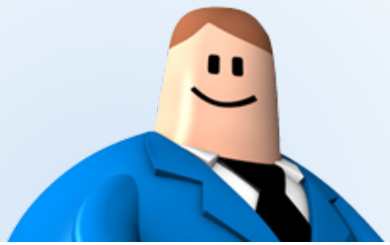
What did we just recommend?

- **Algorithms with outcome**
- **Personalized content**
  - Products
  - Authors
  - Artists
  - Deals
  - Categories
- **Personalized User Interface**
- **Determine the priority of algorithm outcome on a page**
- **Over channels (webshop, mobile, email)**



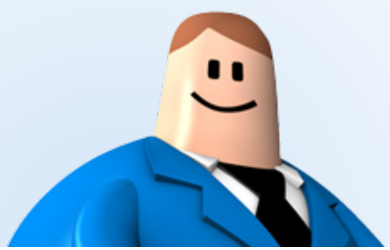
Level of personalization?



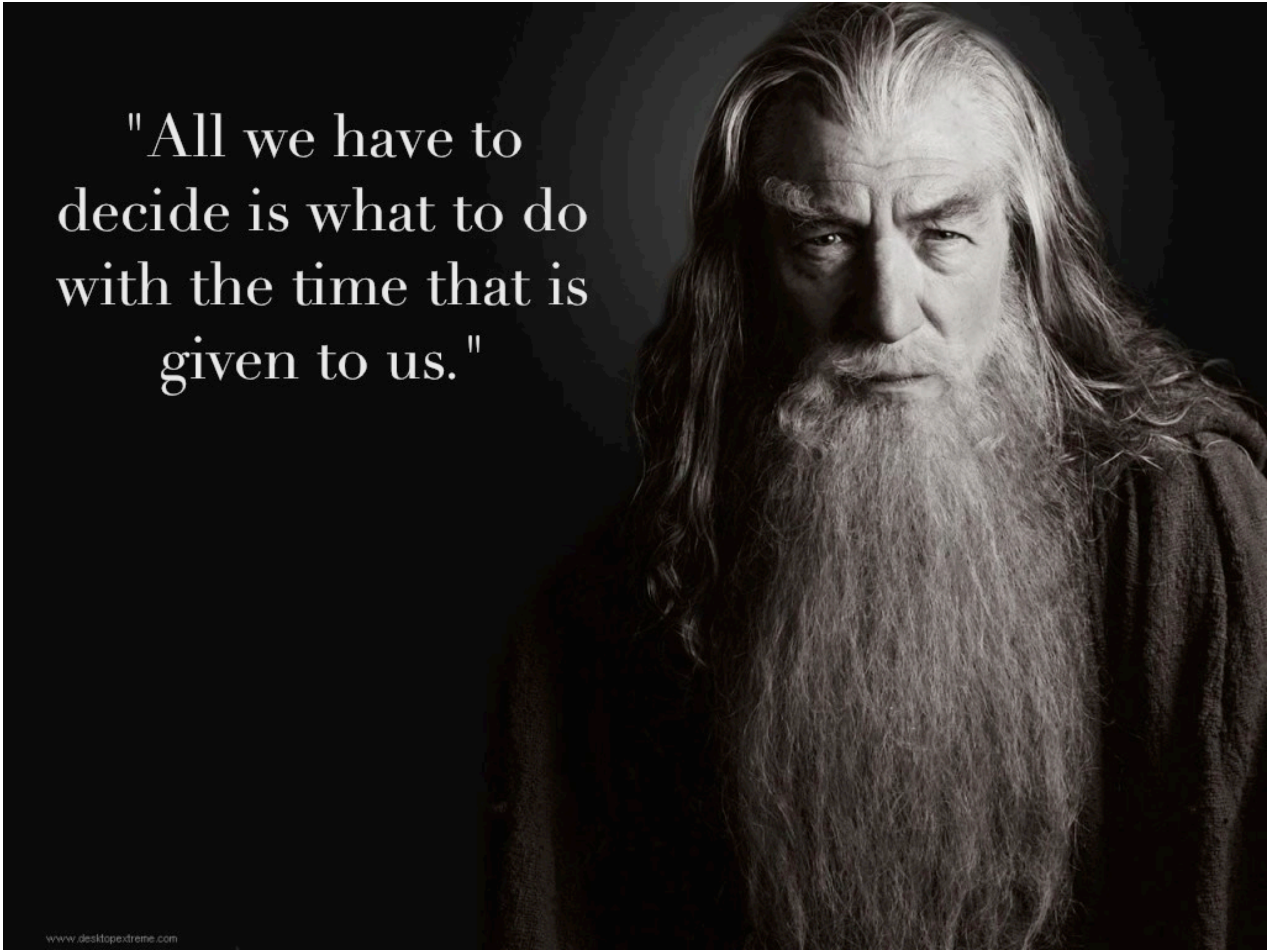


Realtime recommender

- **What would add value to the customer given his current context?**
- **What do we already know about the visitor? (previous behavior)**
- **Determine the level of personalization**

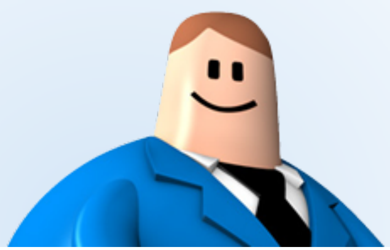


bol.com

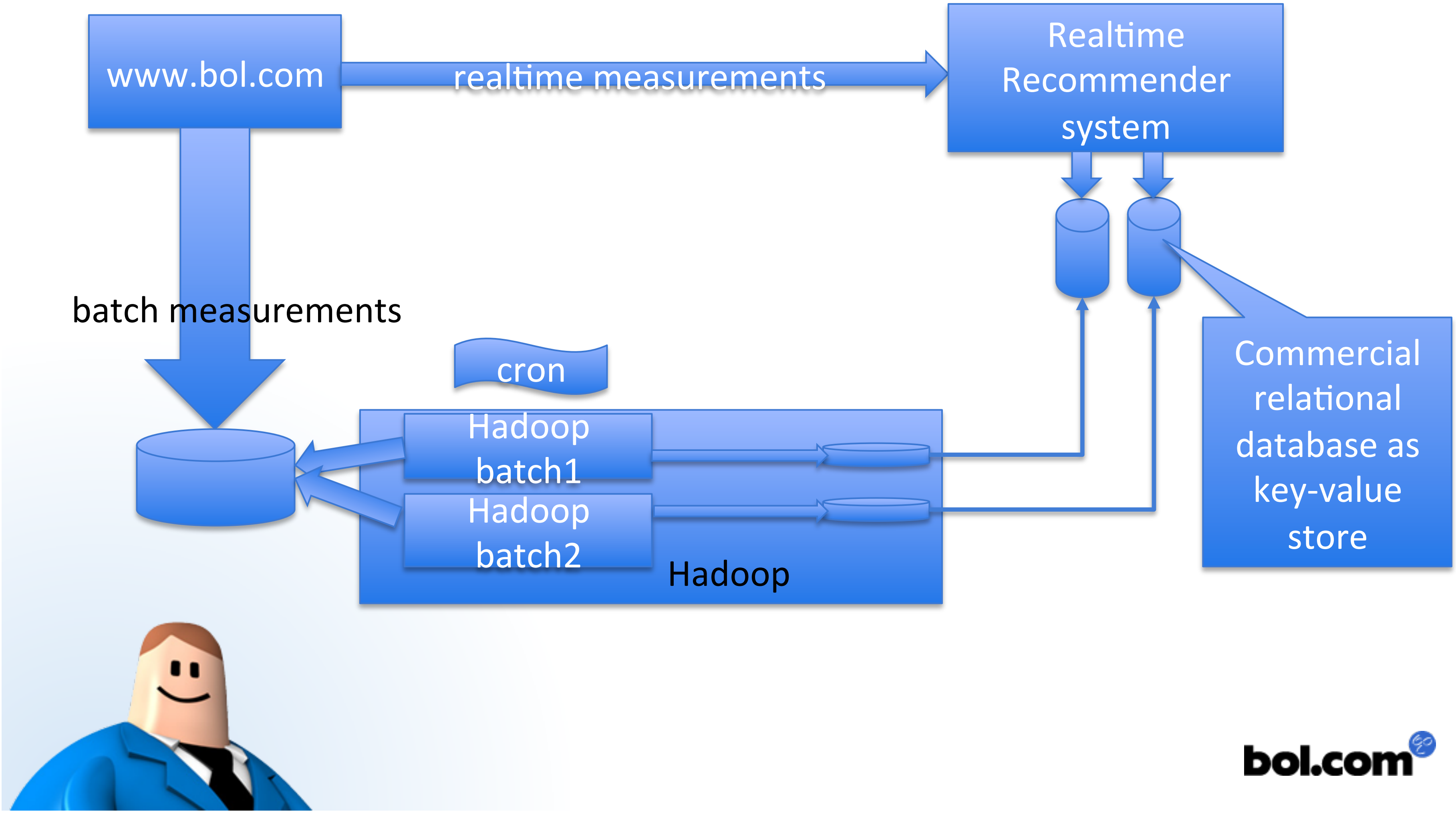

"All we have to decide is what to do with the time that is given to us."
www.desktopextreme.com

What to do with given time

- **Cache ‘expensive’ data**
- **Tune/(try different) jdbc connectivity components**
- **Profile your code (Java VisualVM)**



Architecture (2011)

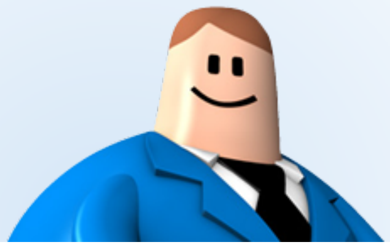


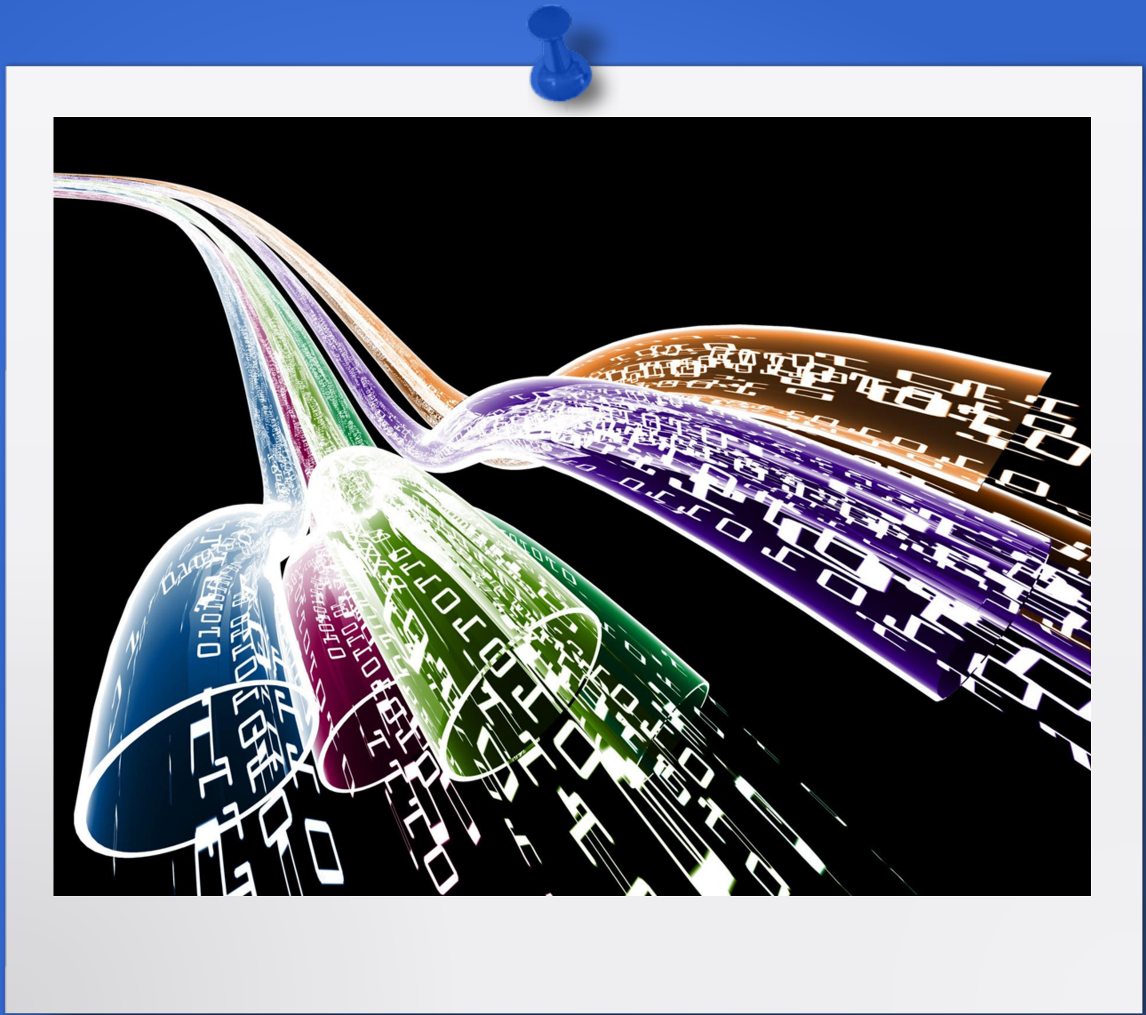
bol.com

Limitations

- **Current implementation**
  - ‘Limited’ realtime analysis
    - Memory footprint too large
  - No detailed ‘long history’
    - Too much data for ‘relational’ database
  - React fast to ‘all onsite visitors’
    - *Did a famous person die?*
  - Realtime and Batch have different data

- **We need a ‘new’ way of handling this.**





Sessionized Lambda Architecture

Sessionized Lambda Architecture

- **Who am I**
- **Usecases & Requirements**
- **Lambda Architecture**
- **Sessionized Lambda Architecture**
  - Bounded Event Streams
  - Queries take time
  - Service Oriented
- **Project status**

Niels Basjes

Niels@Basjes.nl
@nielsbasjes
https://github.com/nielsbasjes

TU-Delft Computer Science
Nyenrode Business School

Software developer
Research Scientist (NLR)
Infra Architect (NLR)
WebAnalytics Architect
IT Architect (Bol.com)

Since 2008 @ Bol.com

bol.com

Our usecases

- **Banner optimization**
  - Look / Search → Next page better banner
- **A/B testing**
  - Show feature → Use → Buy product
- **Search Suggestions**
  - Search → Find → Choose → Buy product
- **Attribution modeling**
  - Show ad → Click → Buy product
- **...**

bol.com

Behavioral analytics

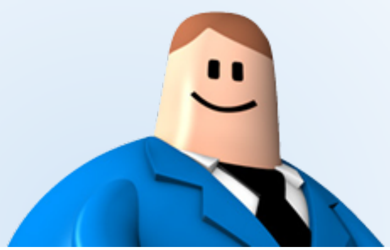
- **Cause and effect**
  - **Action:** We show something
  - **Reaction:** To click or not to click

bol.com

Realtime BigData

- **We need**
  - Near real time
  - Long history
  - Behavioral analytics

- **This is a ‘Realtime BigData’ problem!**
  - What solutions already exist?



bol.com

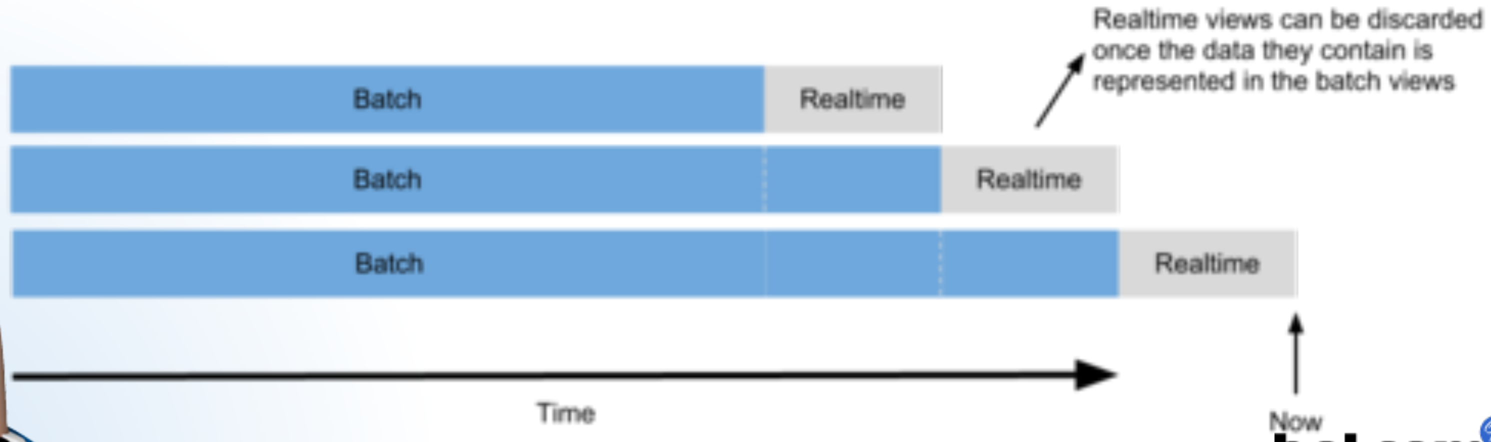
Lambda Architecture

by Nathan Marz (Twitter)

bol.com

Lambda Architecture

- **Basic idea:**
  - Batch tools are good at ‘a lot of data’
  - Realtime tools are good at ‘low latency’
- **Lambda Architecture:**
  - Combine the two at query time.
  - Result: from ‘long ago’ to ‘just now’.



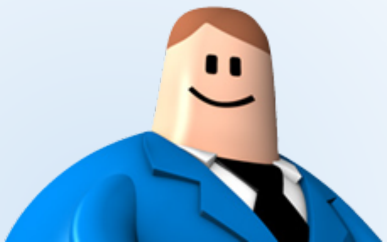
bol.com

Lambda Architecture

- **Fault-tolerant,**
  - Hardware failures
  - Human mistakes
- **Wide range of use cases.**
- **Horizontal scalability.**
- **Extensible, easily debuggable, minimal maintenance.**

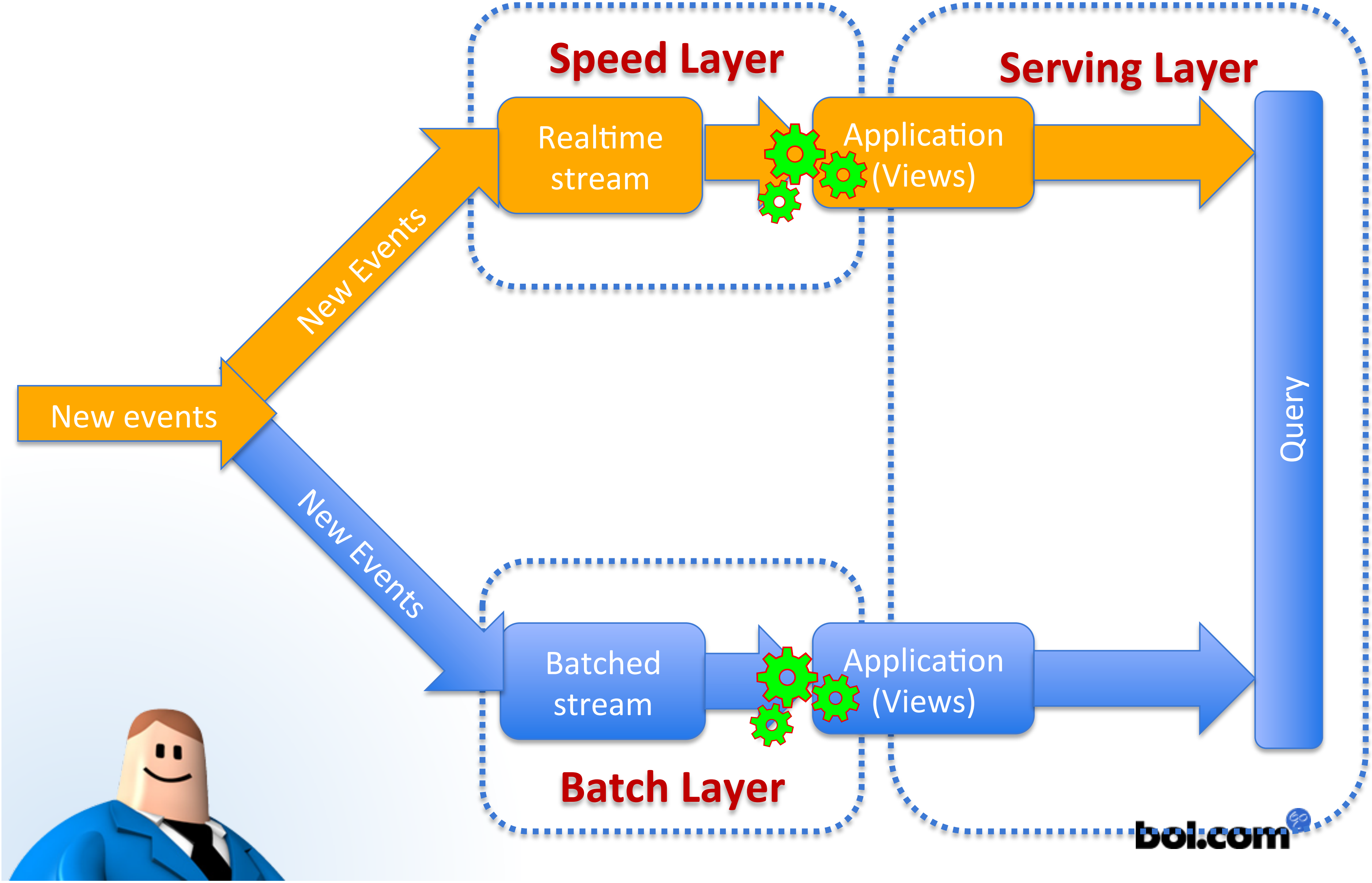
Lambda Architecture

- **Immutable data**
  - You never change the original input.
  - Then you can always recover from failures



bol.com

Lambda Architecture



However …

- **Most presentations**
  - Cut boundary at a millisecond
    - Fine for single events (twitter)
  - Queries take ‘zero’ time
    - Perhaps acceptable assumption
- **Real usecases (for bol.com)**
  - Cause and effect analysis on streams
    - Cannot cut at any millisecond
  - Queries take a ‘long’ time
    - Several seconds for BI dashboard

bol.com

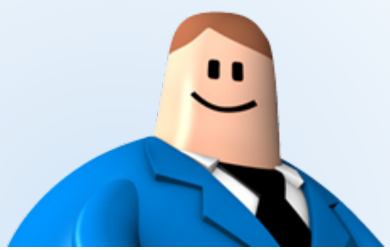
Our clicks

Cutting on the millisecond

We want complete visits

Requirements

- **Online:**
  - Near real-time (< 2 seconds)
  - Have long history available
  - Seamless integration history + real-time
- **Offline:**
  - Incremental batch jobs during the day
  - Have only complete visits in a file
  - Have long history available



Sessionized Lambda Architecture

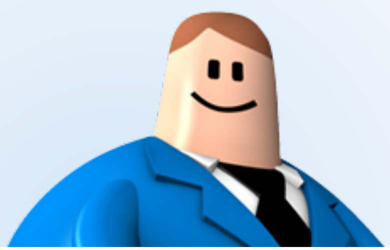
by Niels Basjes (Bol.com)

bol.com

Sessionized Lambda Architecture

Extension of the Lambda Architecture

1. **Bounded event streams**
   related events stay together
2. **Queries take time**
   and multiple can overlap
3. **Service orientation**
   because it is part of a bigger thing



bol.com

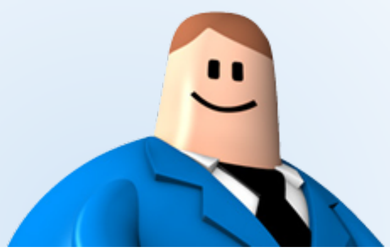
Bounded Event Streams

Because events belong together

bol.com

Sessions and Visits

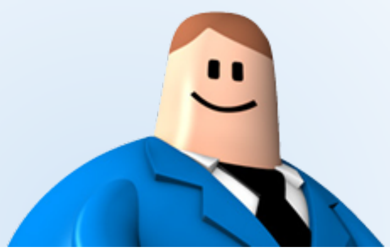
- **Browsers**
  - The software installed on a computer
- **Sessions**
  - Start: Open browser + visit website
  - End: Close browser
- **Visits**
  - Start: Visit website
  - End:
    - 30 minutes idle
    - max 12 hours active



bol.com

Focus on visits!

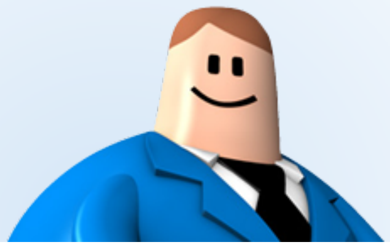
- **Sessions**
  - Can last for days … weeks … !
    - Device is ‘suspended’
      - Browser doesn’t close.
      - Session doesn’t end.
- **Visits**
  - Bounded under our control.



bol.com

Sessionizer

- **Create predictable visit id**
  - Only needed if the source does not have such an id
- **Route events to**
  - ‘Speed’ layer
  - ‘Batch’ layer



bol.com

Queries take time

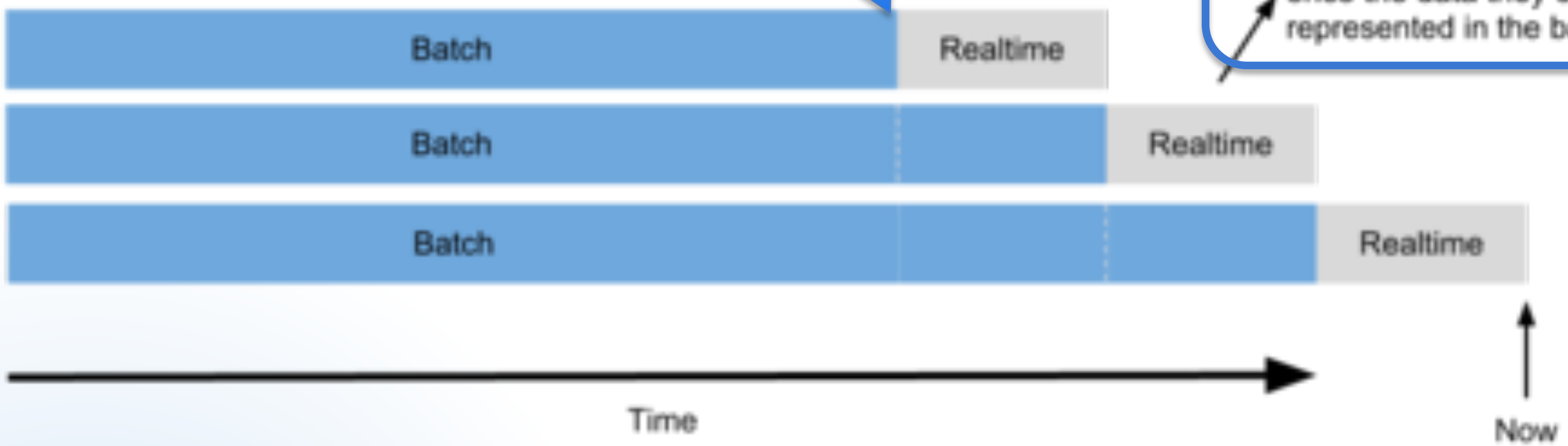
Simply not atomic

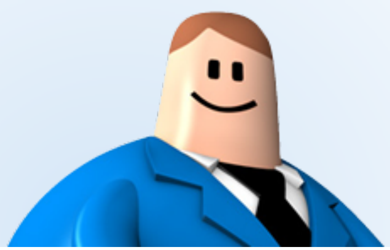
Query Boundary & Cleanup

The edge between speed and batch

Cleanup !!

Realtime views can be discarded once the data they contain is represented in the batch views

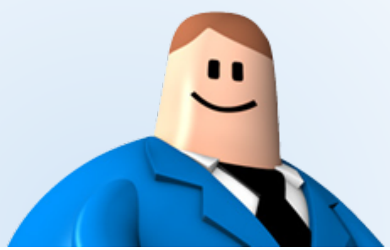




bol.com

Query Boundary

- **Lambda Architecture:**
  - Millisecond
    - Simple administration
    - One value
- **Sessionized Lambda Architecture:**
  - Per visit
    - Only complete visits
    - Many (thousands) values

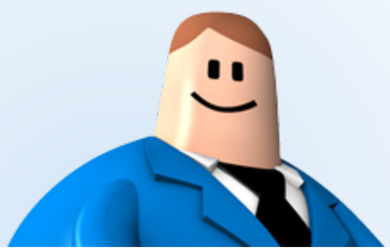


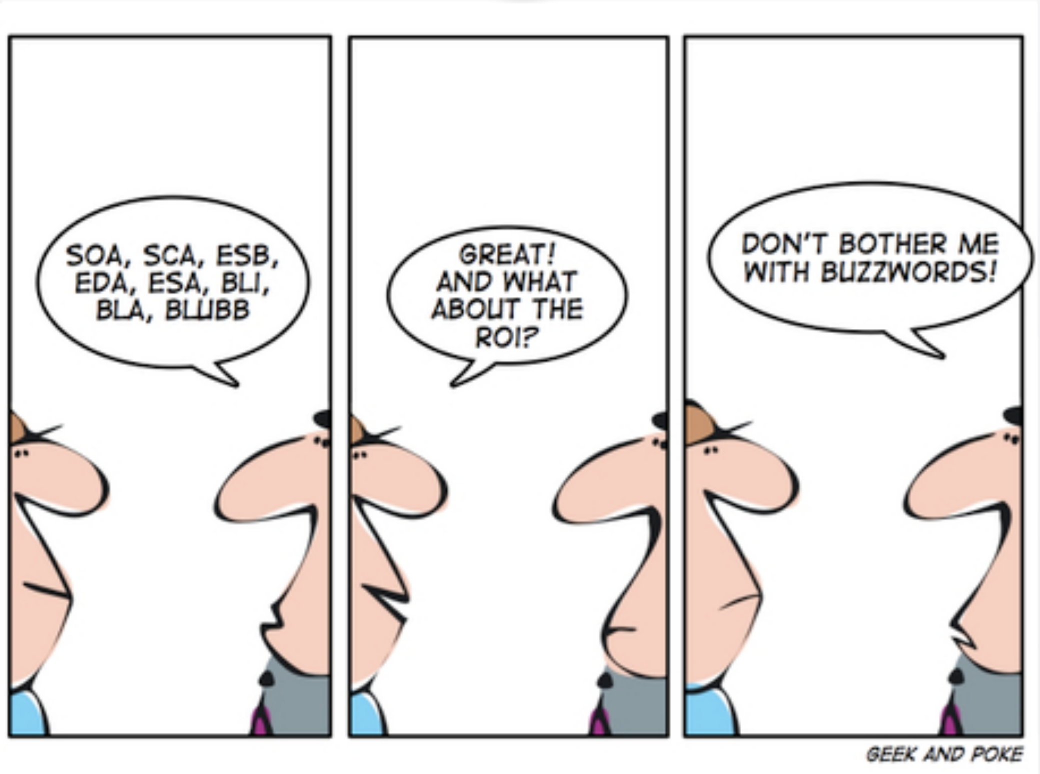
Queries take time

- **Query boundary**
  - can change during the query
- **Cleanup**
  - should wait for queries to finish

Running a query

- **Query**
  - Get current boundary + lock
- **Cleanup of speed views**
  - Only when ‘unlocked’





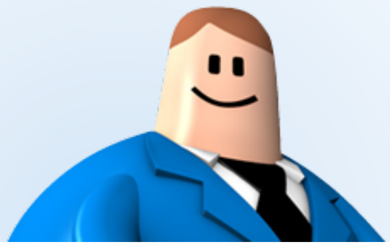
Service Orientated Architecture

Making it part of the whole

bol.com

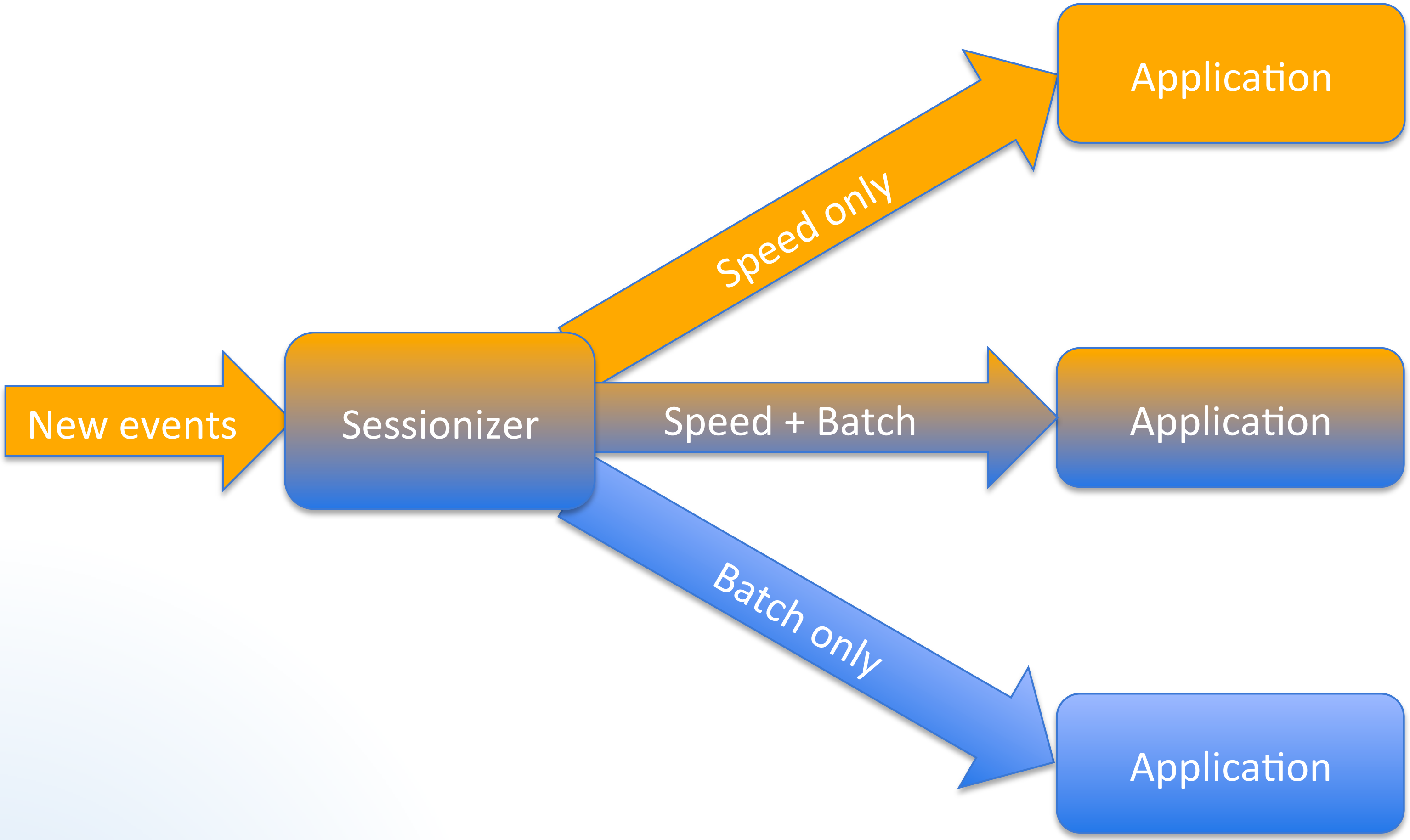
One 'interface'

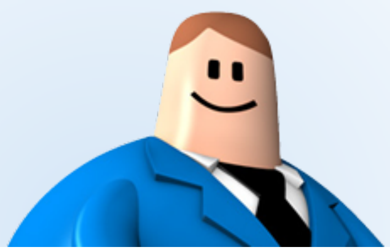
- **One service interface**
  - *All clicks so far*
- **Two technologies**
  - Speed: Kafka
  - Batch: Files on HDFS / Hbase / …

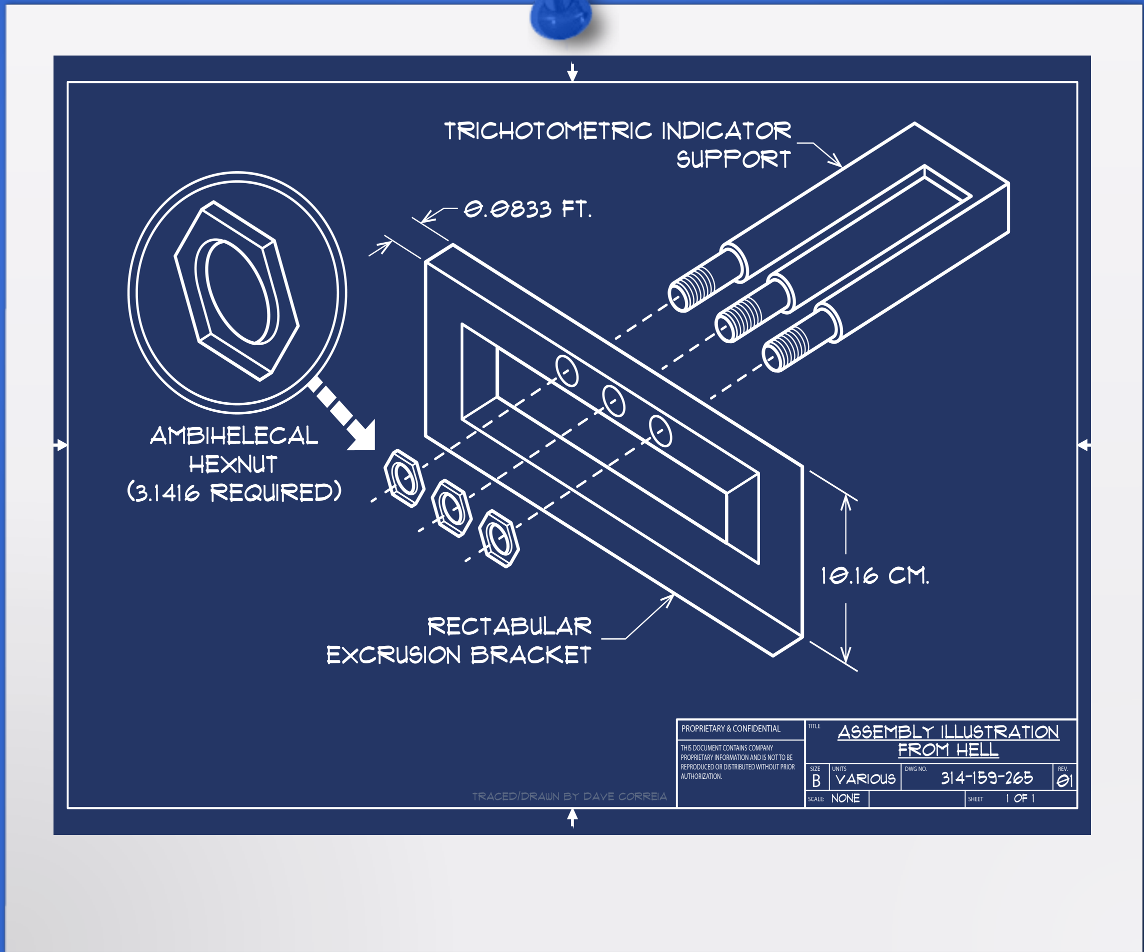


bol.com

Applications choose





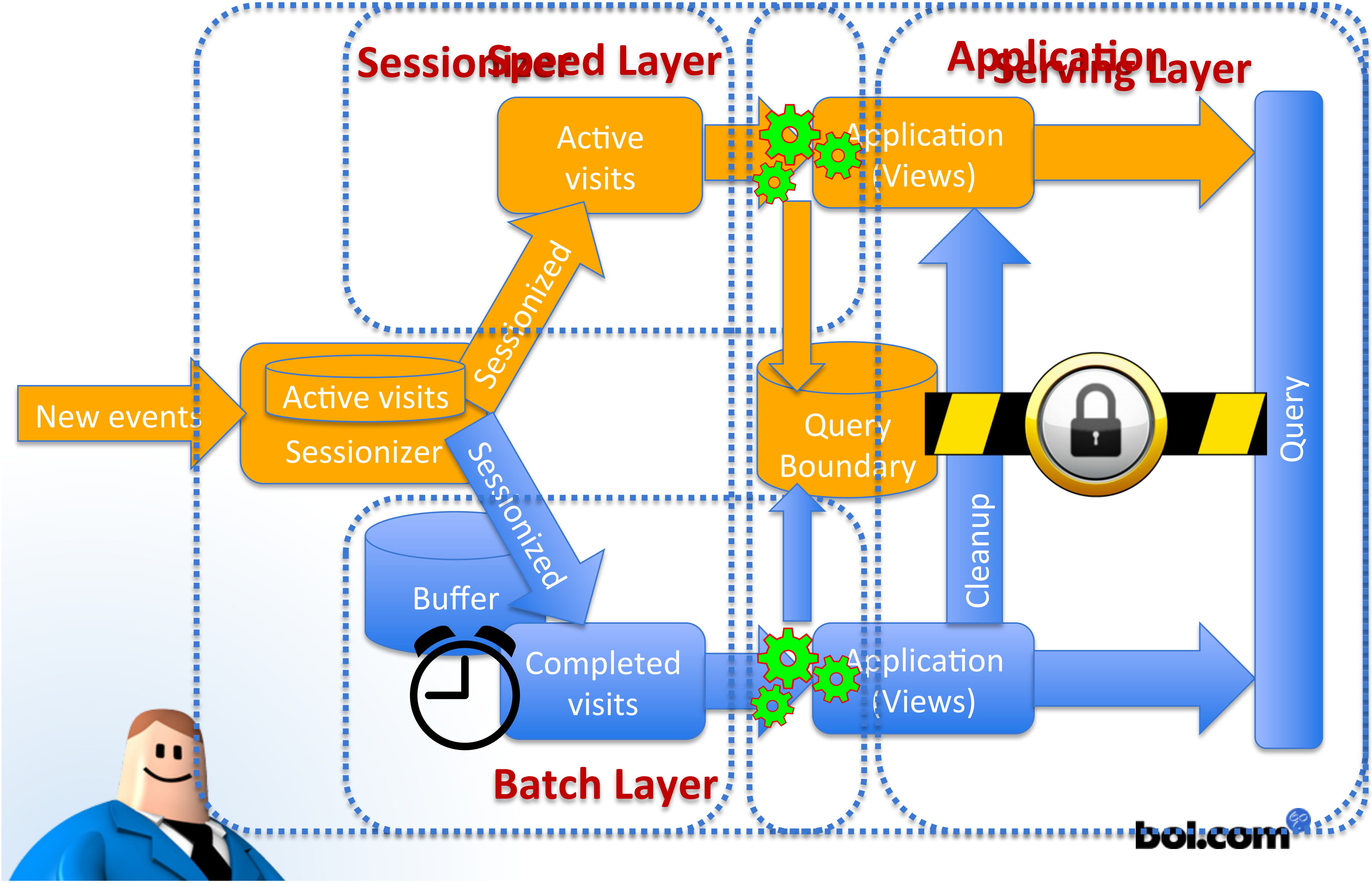


Overview

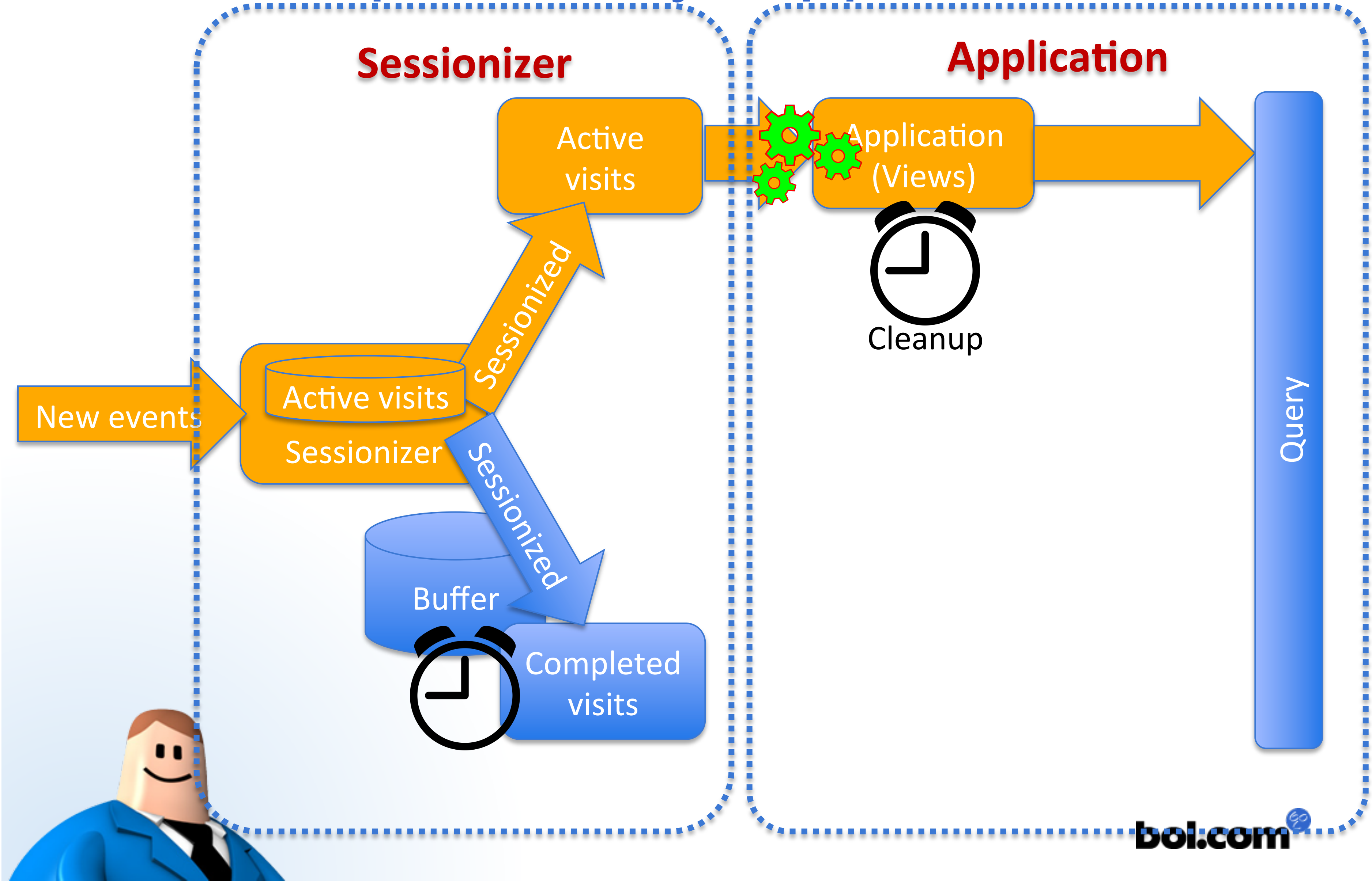
How it all fits together

bol.com

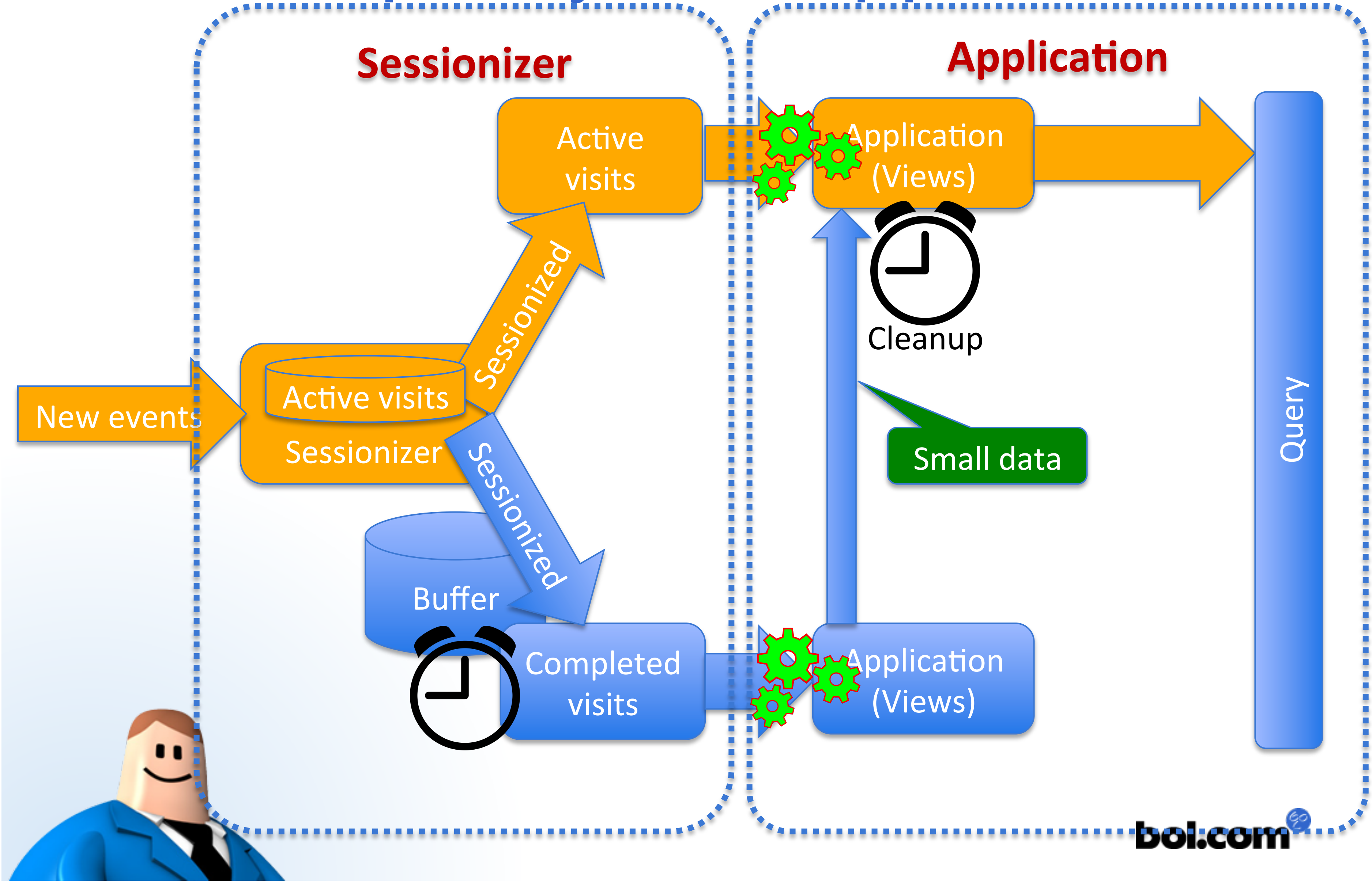
Sessionized Lambda



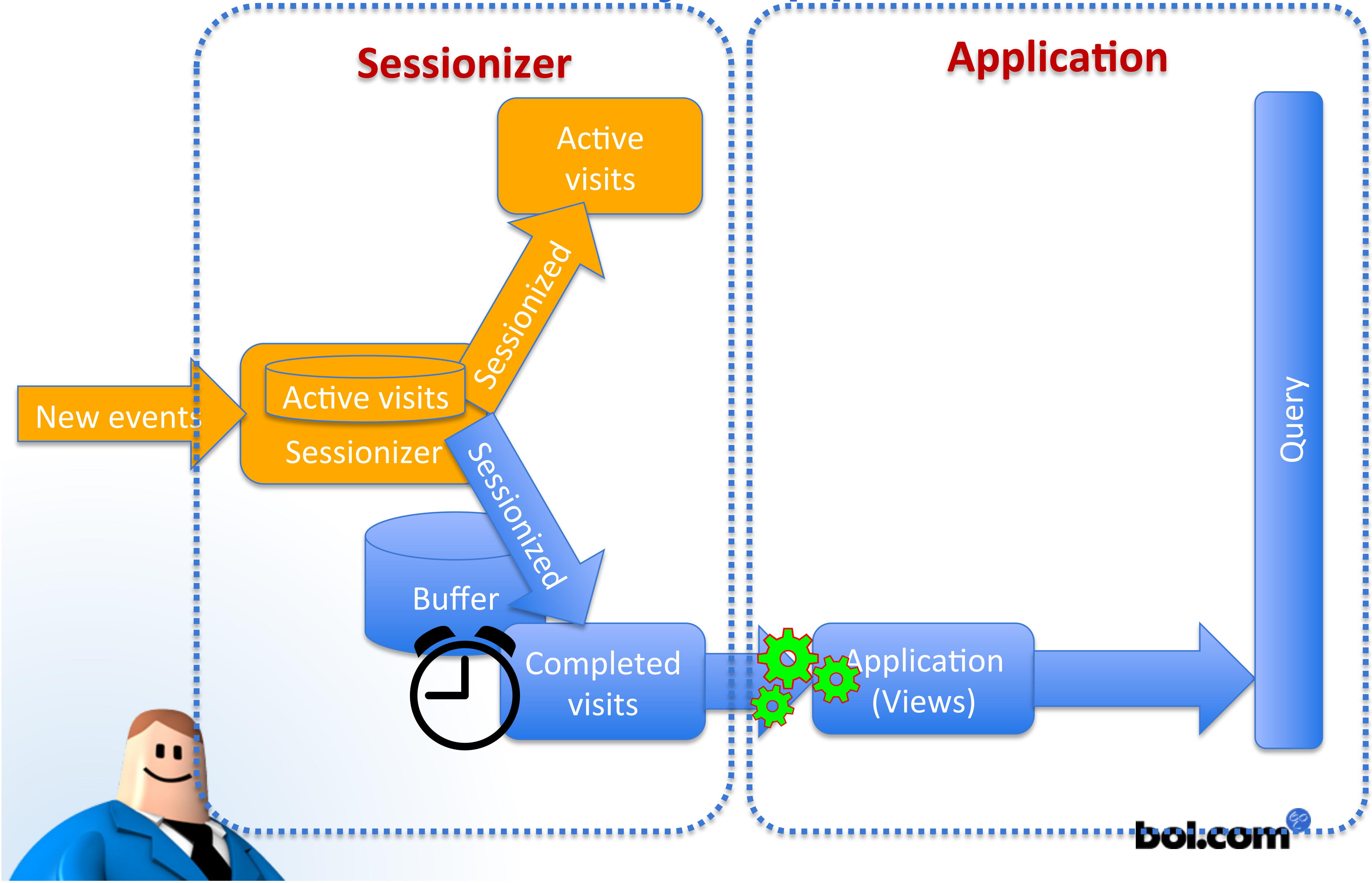
“Speed only” application



“Simple hybrid” application

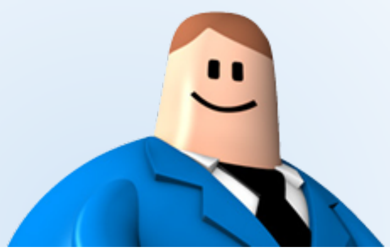


“Batch only” application

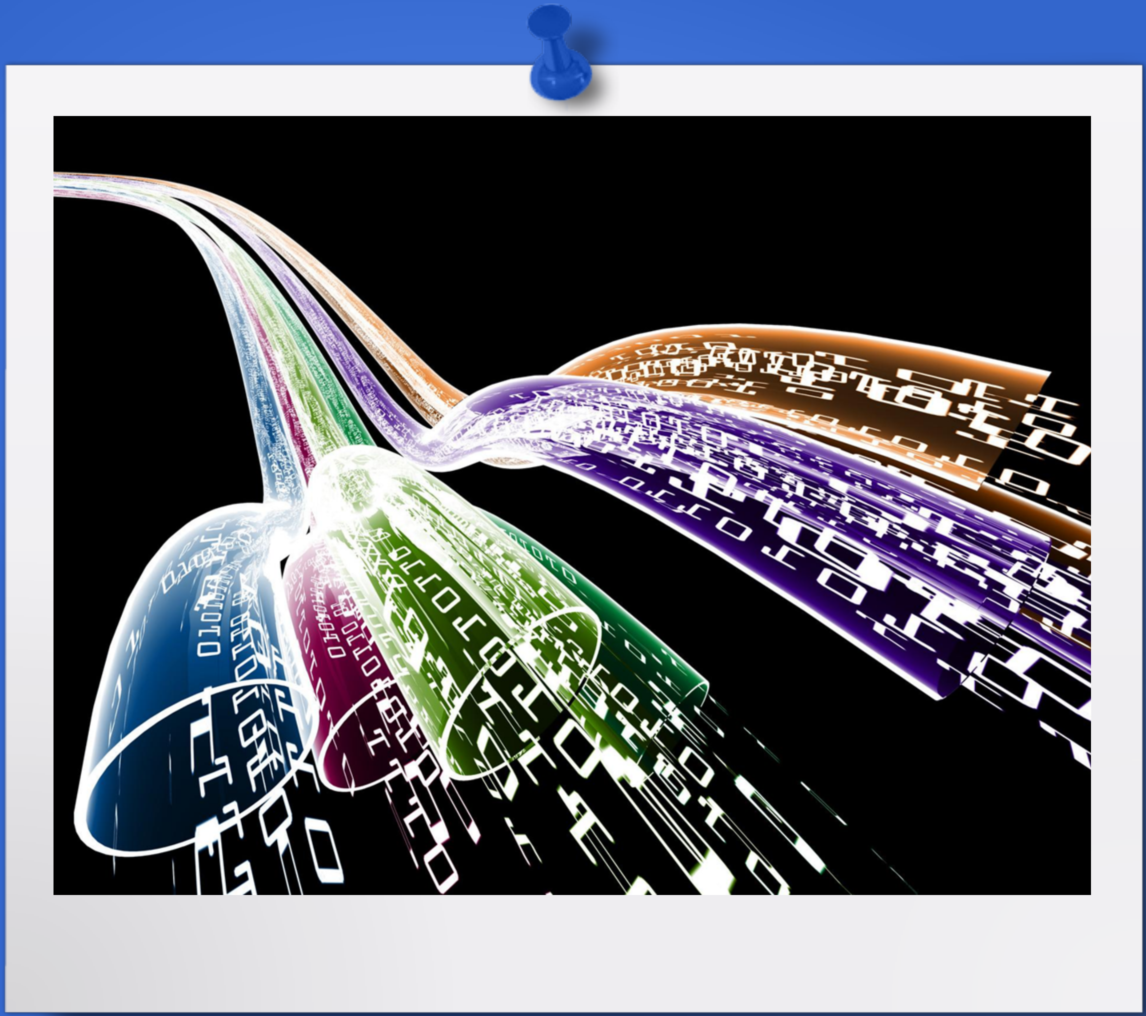


Sessionized Lambda

- **Advantages**
  - Data is available in complete visits
  - Separation of concerns
  - Easier behavioral analytics code
- **Disadvantages**
  - Batch layer data is available but incomplete for multiple hours.
    - Long running visits



bol.com



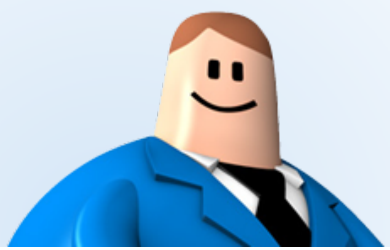
Project status

bol.com

Building prototype

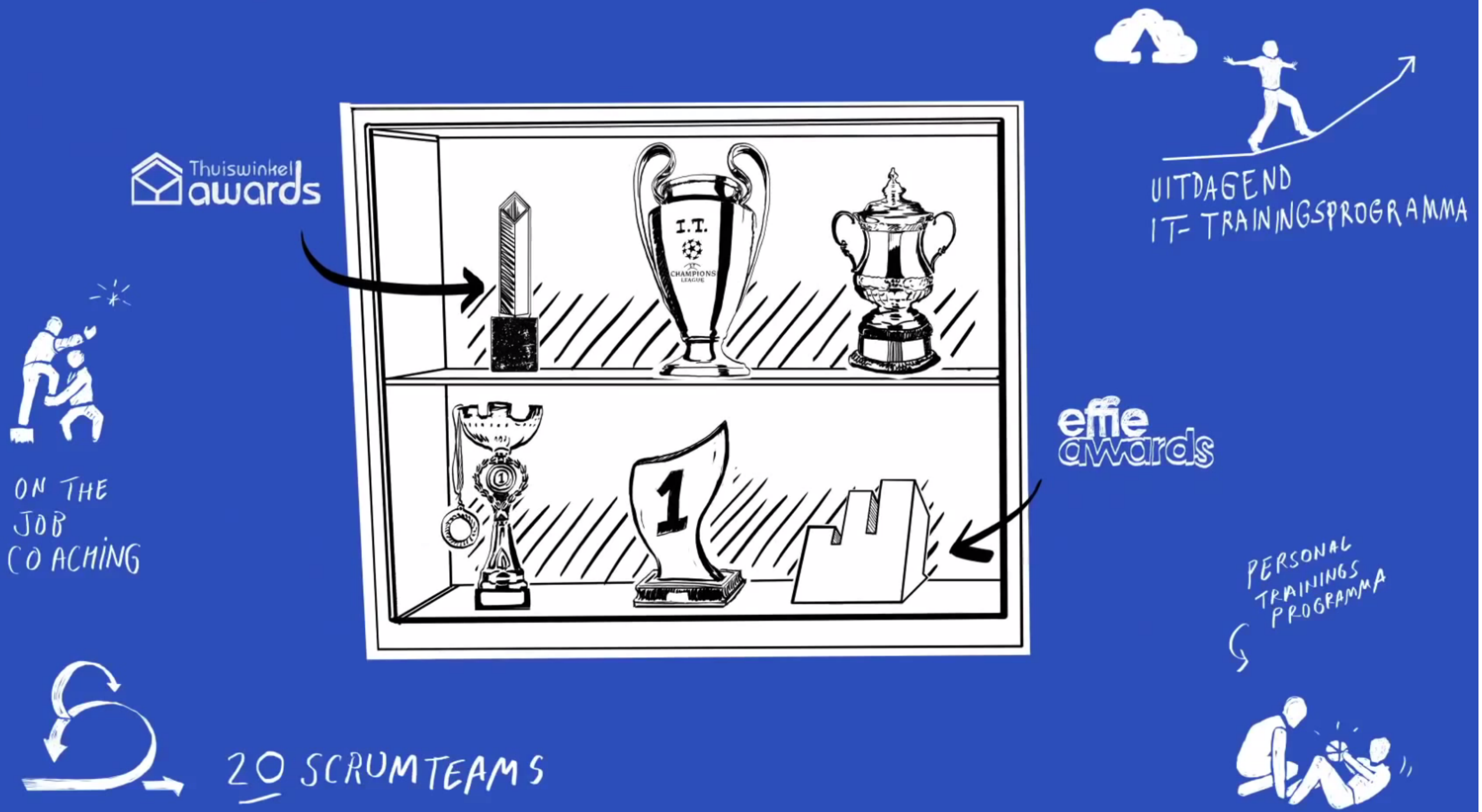
Challenges

- **Scalability design**
  - Query boundary
  - Query / Cleanup locking
- **Recovery design**
  - Because humans make mistakes
- **Handling real events**
  - No session
  - Events are out of order
  - Robots / Hackers / …



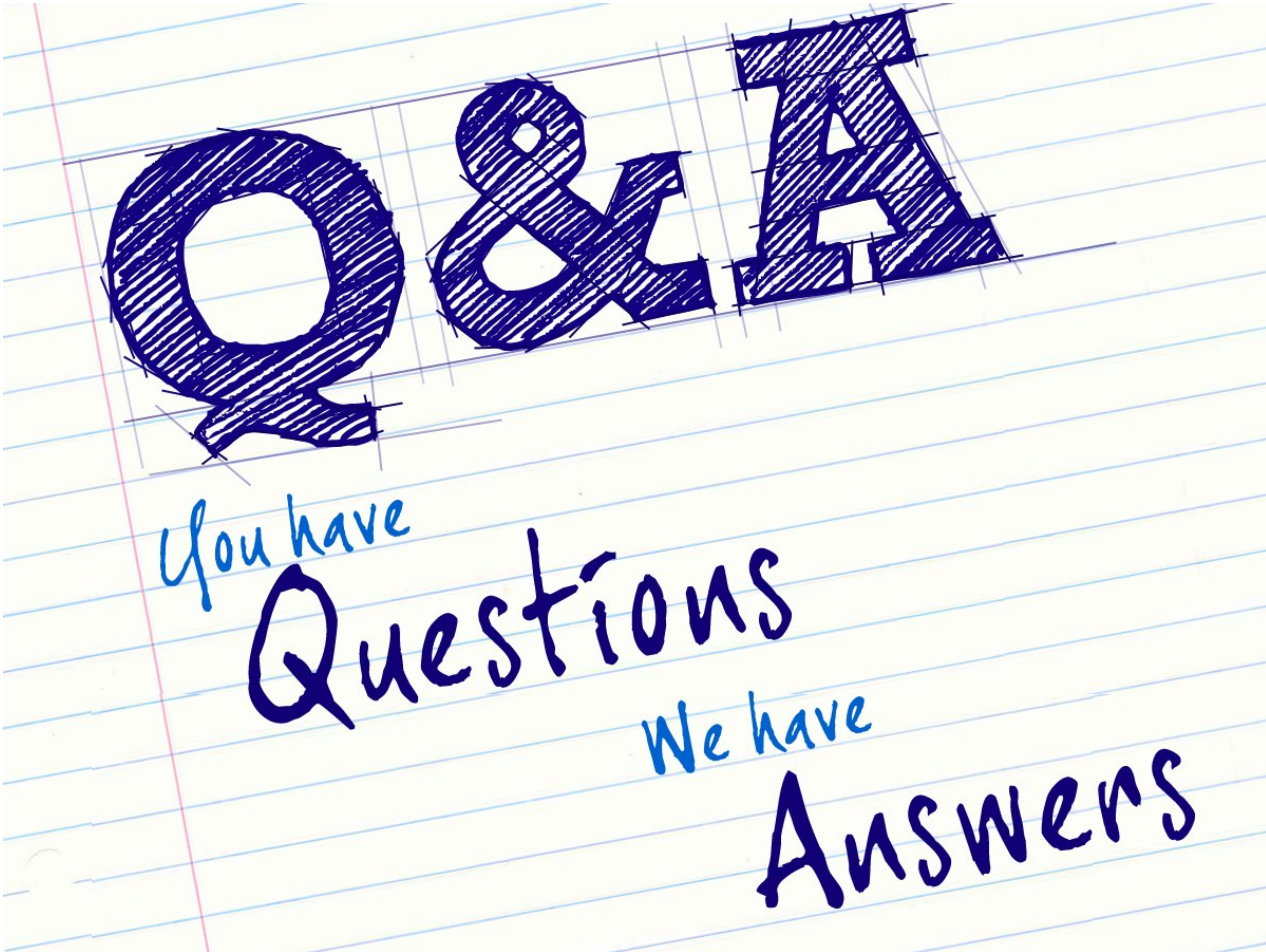

20
bol.com

Join us



Thuiswinkel awards
UITDAGEND
IT-TRAININGSPROGRAMMA
ON THE
JOB
COACHING
effie awards
PERSONAL
TRAININGS
PROGRAMMA
20 SCRUMTEAMS

www.bol.com/banen


Q&A
You have
Questions
We have
Answers

goto; conference

Please evaluate this talk via the mobile app!