Learning at scale on Hadoop

Olivier Toromanoff, Software engineer

Berlin Buzzwords, 2015-06-01

Prediction @ Criteo

Learning on Hadoop

Limits are lower than the sky

From Experimentation to Production:
a model lifecycle

Criteo: Performance **advertising**

« The right ad at the right time to the right user »

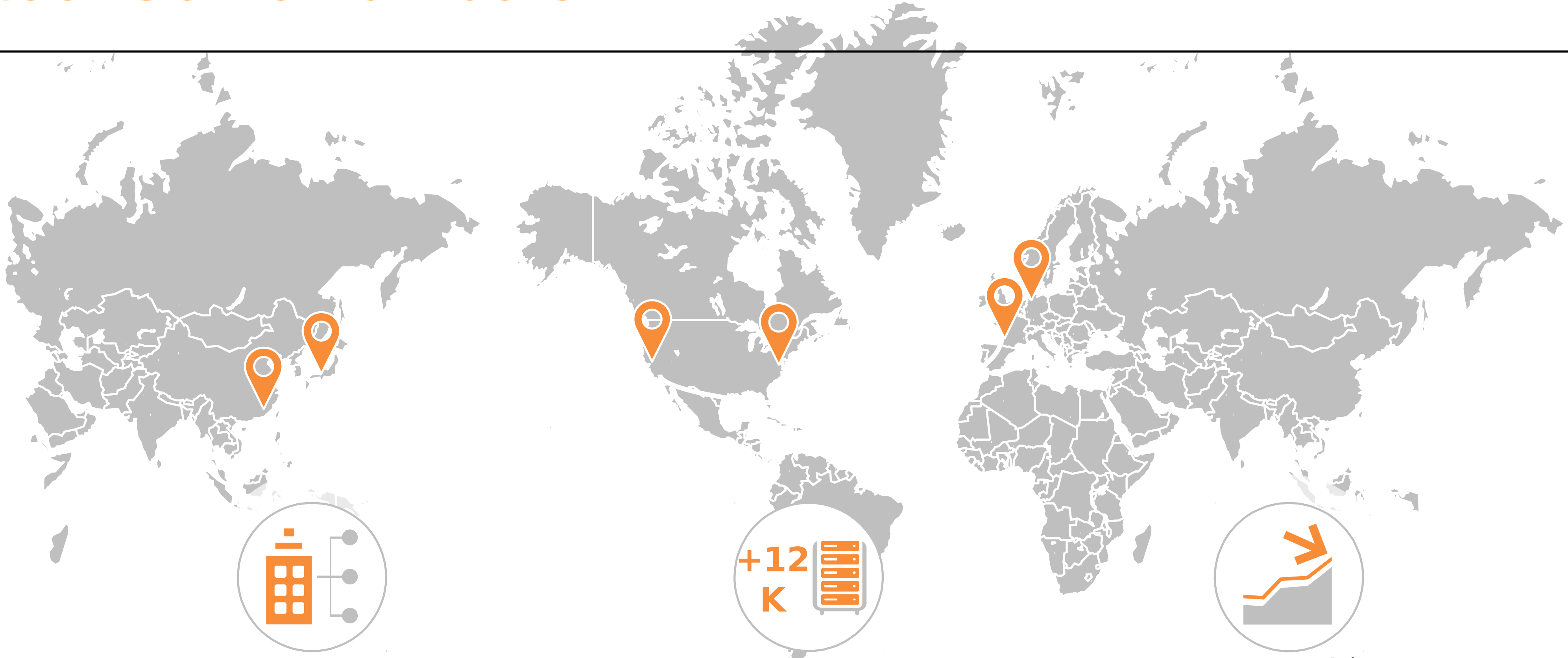
6ms

Criteo: **Performance** advertising

$$Buy\,?\;\mathbb{E}[CPM] > CPM$$

$$\mathbb{E}[CPM] = \mathbb{E}[NbClicks] * CPC$$

Criteo: Some numbers



6 DATA CENTERS

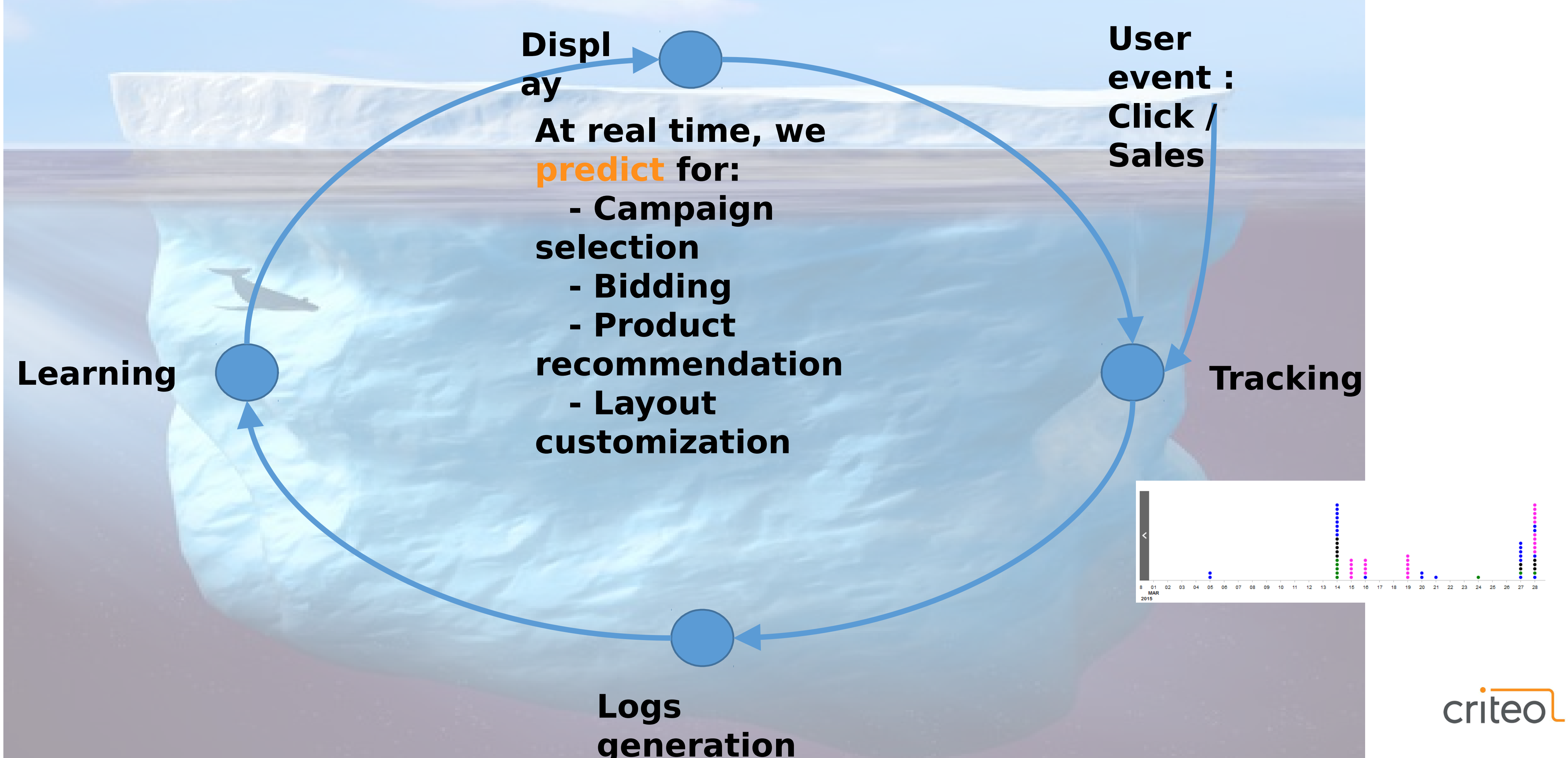
MORE THAN **12 000 SERVERS**

PEAK TRAFFIC (PER SECOND) **800K** HTTP REQUESTS

CDH4 CLUSTER OF 1200 NODES (36TB, 96GB RAM, 24 cores)

criteo

The display cycle

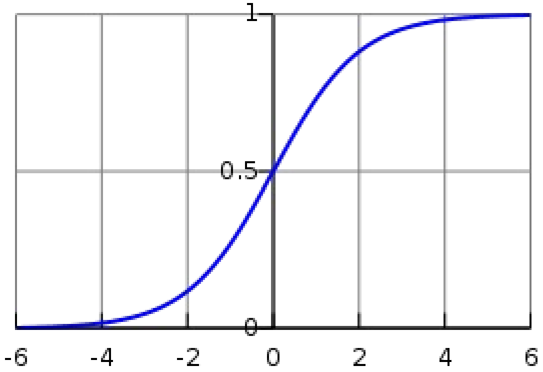


Click prediction modelling

$$\mathbb{E}[NbClicks] = \mathbb{P}\{Click\}\mathbb{E}[NbClicks|Click]$$

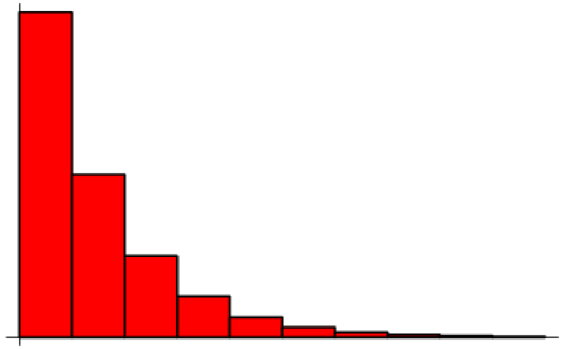
$$= \frac{1}{1+e^{-<w.x>}}$$

(logistic)

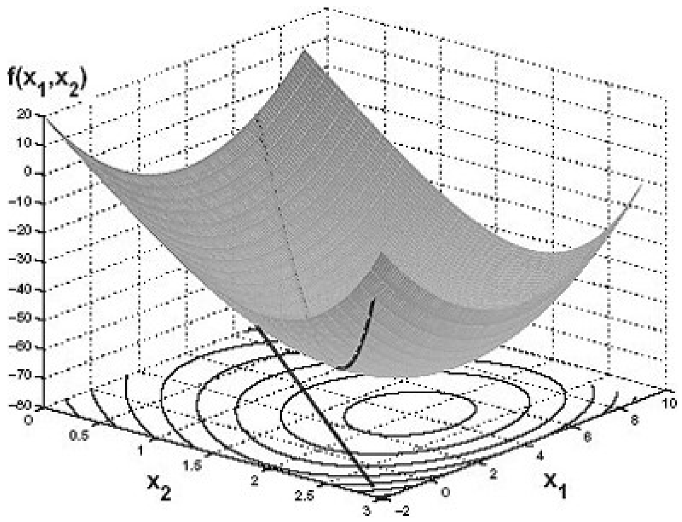


$$= 1 + e^{-<w.x>}$$

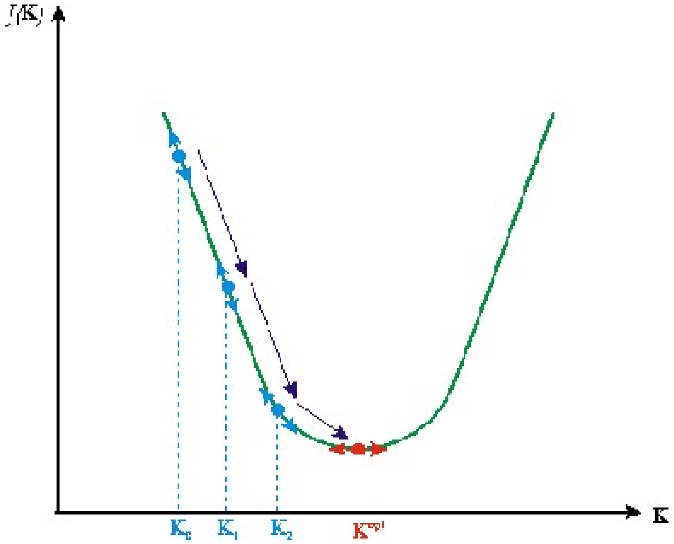
(geometric)



The nice sides of Logistic Regression



Convex Optimisation



Solvable with iterative Gradient Descent Algorithms (L-BFGS)

Fast Prediction at runtime

criteo

Prediction @ Criteo

Learning on Hadoop

Limits are lower than the sky

From Experimentation to Production:
a model lifecycle

Need for Scale

Learning a mo

- 7 days
- 30 000
- ~ 7 TB
- 190 000

... and we ha
(click/sales/

.. and we want ...uch as possible..

criteo

But… a strong C# Microsoft .net « legacy »

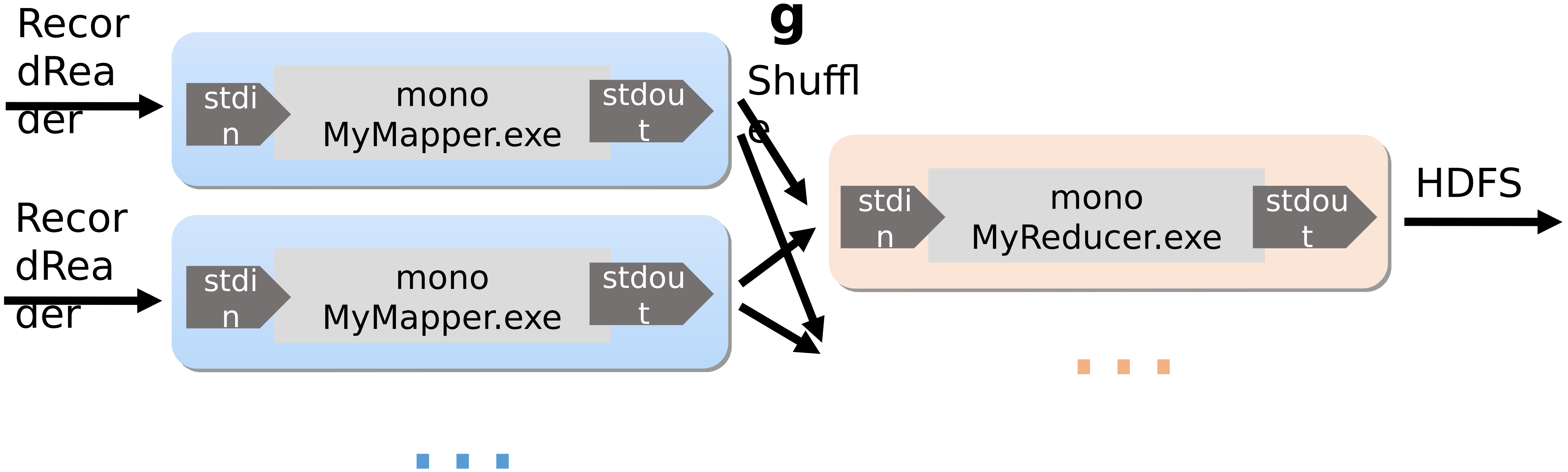
- Existing in-house … brary … C#
- Front-end code in …
- Do we really want …
  - Re-implement miss… …en source library?
  - Rewrite all the lear… …ratch?
  - Take the risk to intr… …cross language duplication?

➢ hmmm.. let's try to run C# on Hadoop!

criteo

Let’s meet mono hadoop Streaming

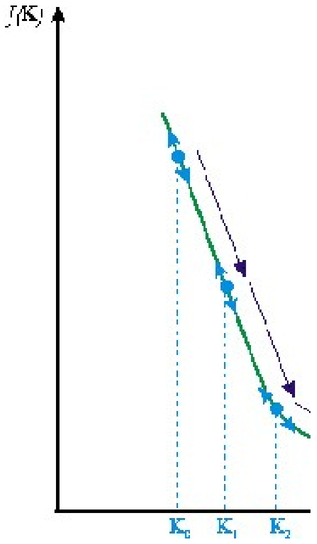
»



Taming the beast

- Limitation on arguments length
- Fitting Mono VM + JVM into container memory
- Additional “mini” sub-processes in Java to interact with HDFS
- Stumbling upon Mono NotImplementedException
- Still (rare) issues with Mono:
  - Crashes when Garbage Collector under heavy load
  - Hanging threads

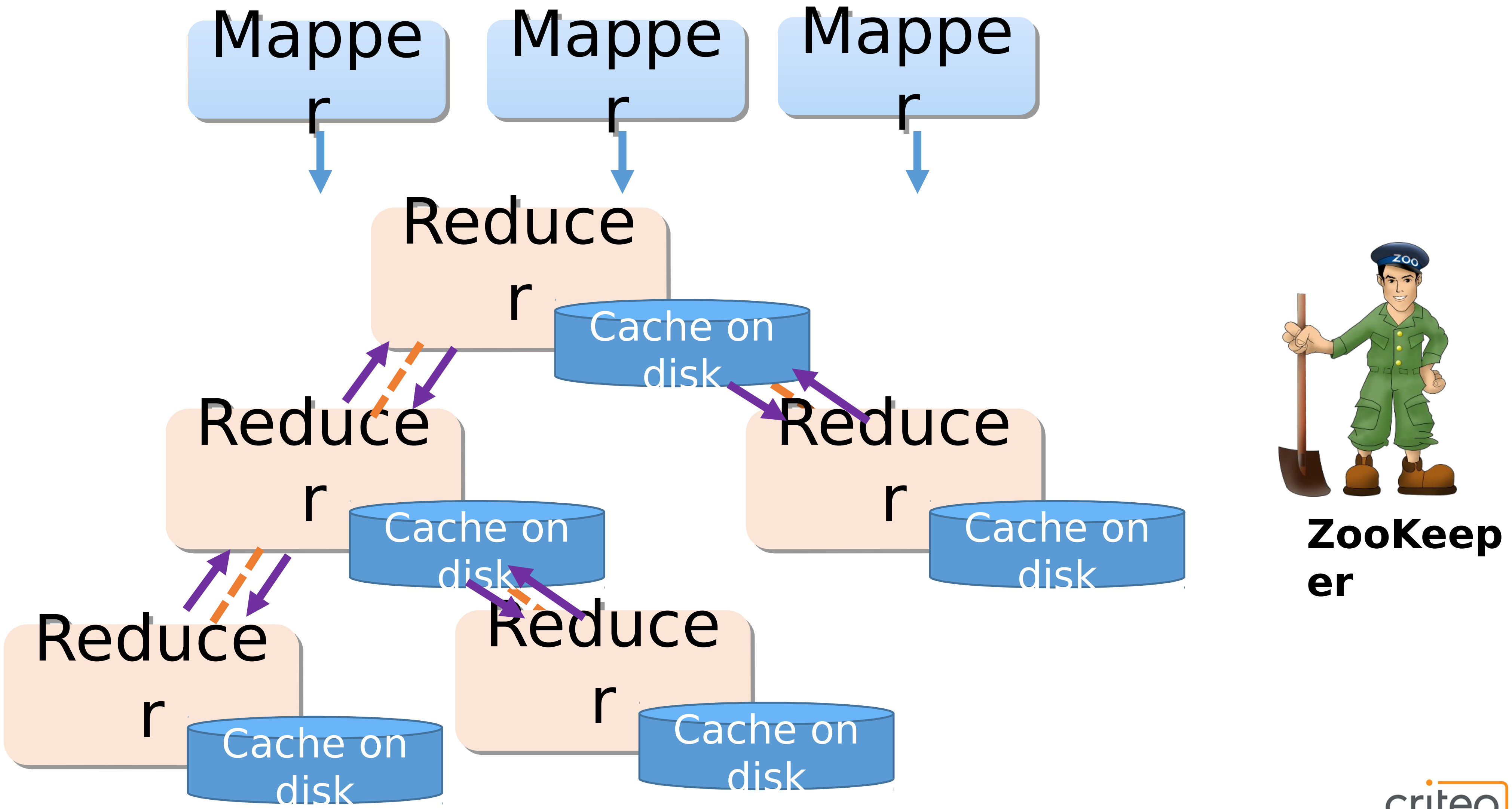
But… iterative? You said iterative?



Converge in iterations

Solution n+1

Let’s « all-reduce »



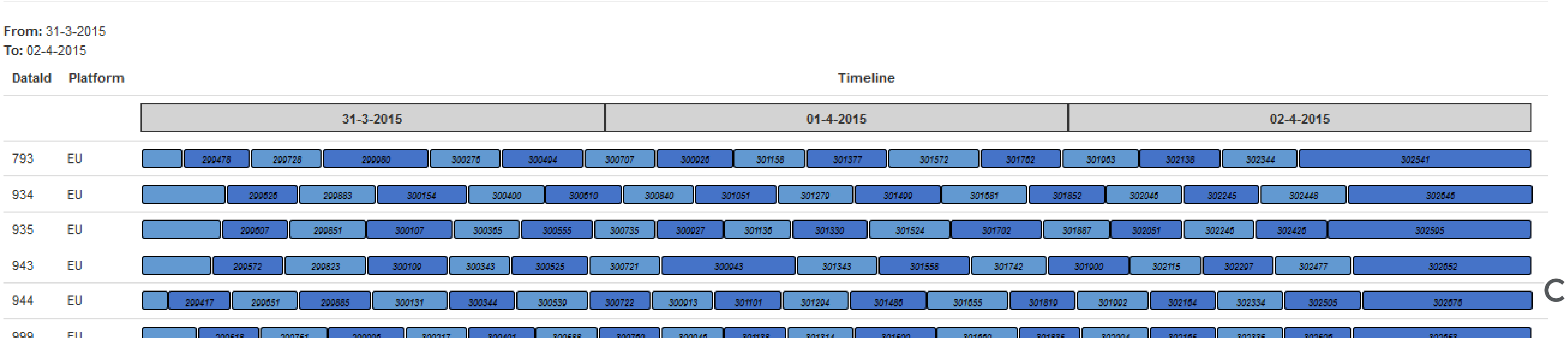
criteo

Distributing ain't easy

- No resilience to reducer crashes
- Keeping reducers for up to 3h
- Processing will start only when all reduc are provisioned

One (production) year later …

- ~ 1300 models/day
- Ingesting 596 TB/day
- Consuming 6310 CPU day/day
- Learning time: [10min; 3h]
- Refresh rate: [3h; 6h]
- Last two weeks success rate: 97.4%



Prediction @ Criteo

Learning on Hadoop

Limits are lower than the sky

From Experimentation to Production: a model lifecycle

Big cluster, small gateway

- Jobs are all launched from a single machine

➢ Asynchronous jobs in Yarn containers

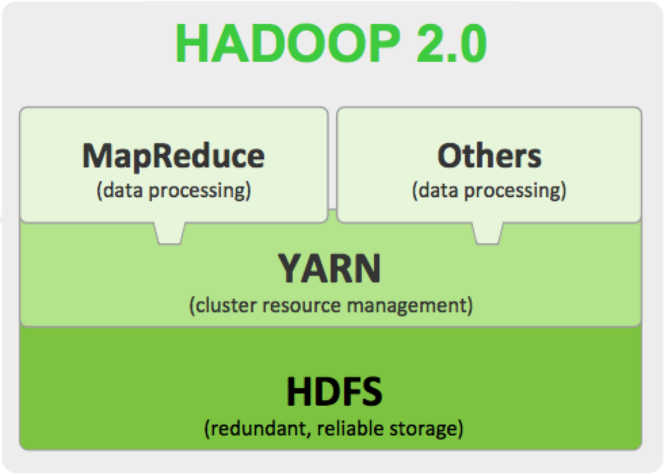
Wasting resources is easy (…and painful)

- We started to see jobs waiting containers for hours
- Loooots of small mappers

➢ CombineFileInputFormat

What about further scaling?

- 98% success rate is good for now
- What about 2x more models, more data, more reducers, ...



Prediction @ Criteo

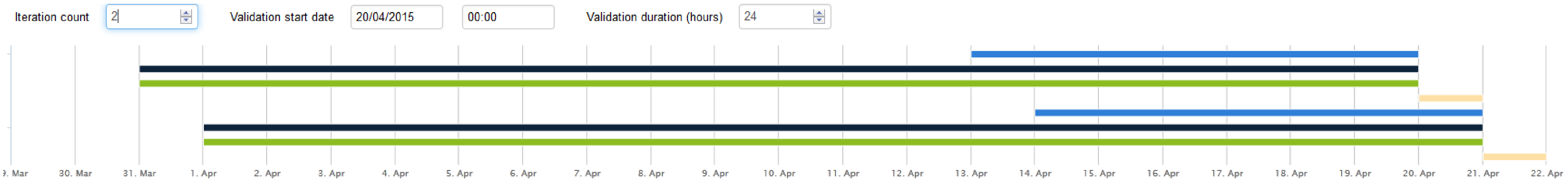
Learning on Hadoop

Limits are lower than the sky

From Experimentation to Production: a model lifecycle

Hello « TestFramework »!

- An internal tool that replays Prod traffic from logs
- Used to:
  - Train models
  - Exercise models
  - Compute metrics



criteo

Prediction analytics

- Observations:
  Basic (clicks,…)
  Prediction (Observed Ctr, Logged PCtr, Simulated PCtr)

- Metrics:
  Basic (count,…)
  ML (MSE, LLH, …)
  ML+Business (MSE_weightedBy*, …)
  Business (Advertiser added value, Criteo gross, …)

Prediction analytics: MR job

Mapper s

Parsing/Filtering/Transformation

`dim1=mod1; … dimN=modN`

Prediction

`dim1=mod1; … dimN=modN; Pred1=p1`

PreAggregation

`SELECT metrics(observations)`

`GROUP BY dimensions`

Historic al Prod Logs

PreAggrega ted Result

Reduc er

Final Aggregation

Resul t

…

criteo

Offline ABTesting

My model improved this fancy MSE_weightedBy430… yeah!..

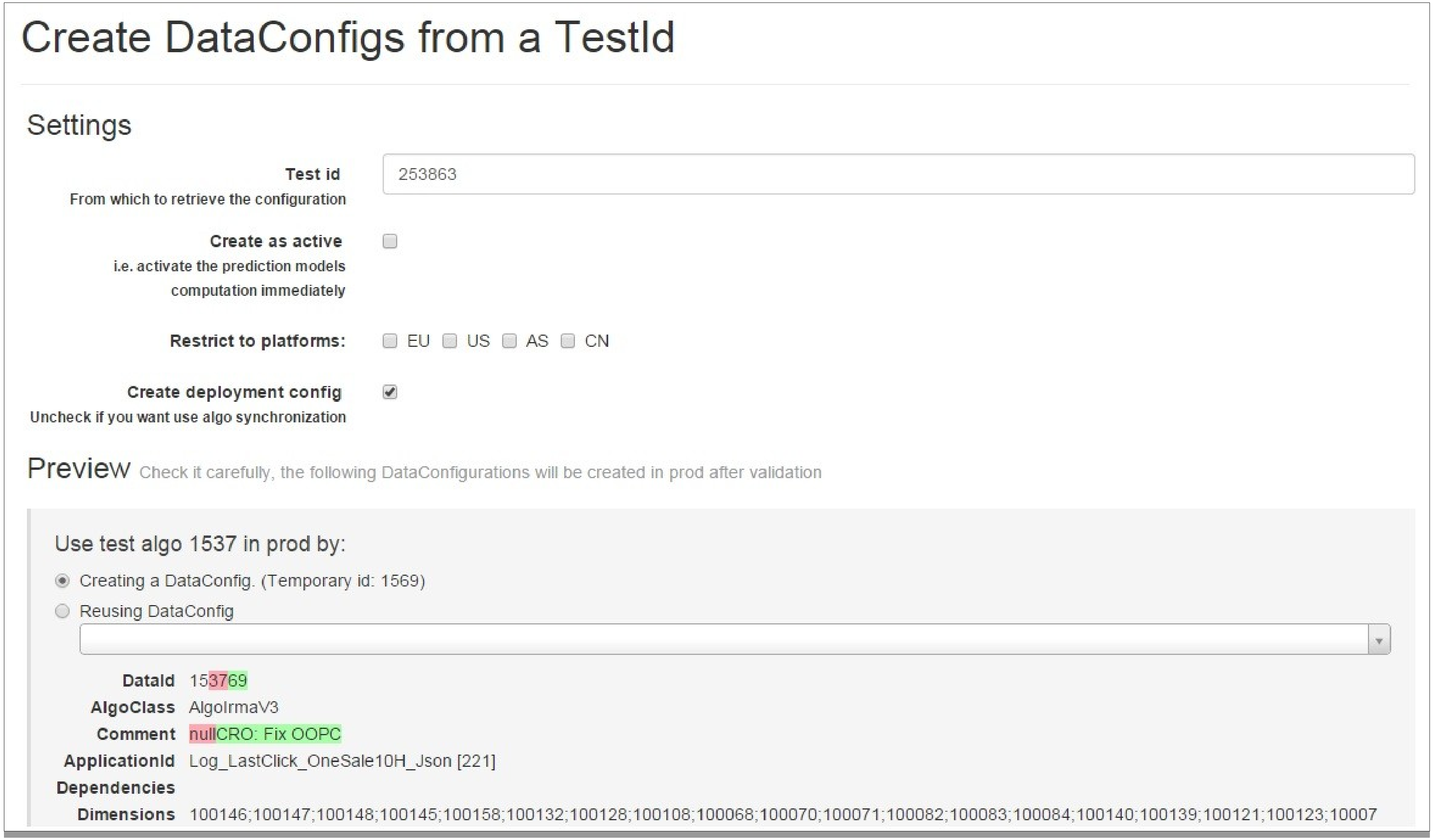
… weeks of productification work latter, IRL it under-performed :(

- Counterfactual analysis:

Online exploration allows us to do offline evaluation of how the tested model would have performed

criteo

From Offline Test to ABTest



ABTest monitoring

Populations

Pop-ID	Name	Size	A = Reference (0.50)	B = Test (0.50)	Excluded (0.00)
0	RefSmall	0.05	◉	○	○
1	RefBig	0.45	◉	○	○
2	Test	0.50	○	◉	○

Recommended Conclusion

Test vs Ref :

Positive

Show/Hide metrics

Global

Displays	Clicked Display	Cost	Revenue	RevExTac	Margin	Advertiser Value	Advertiser Criteo ROI
Δ(%)	Δ(%)	Δ(%)	Δ(%)	Δ(%)	Δ(%)	Δ(%)	Δ(%)
↑ +0.41%	↑ +1.51%	↑ +0.26%	↑ +1.66%	↑ +3.43%	↑ +1.74%	→ +0.87%	→ +0.61%

-0.55% ~ +2.55%

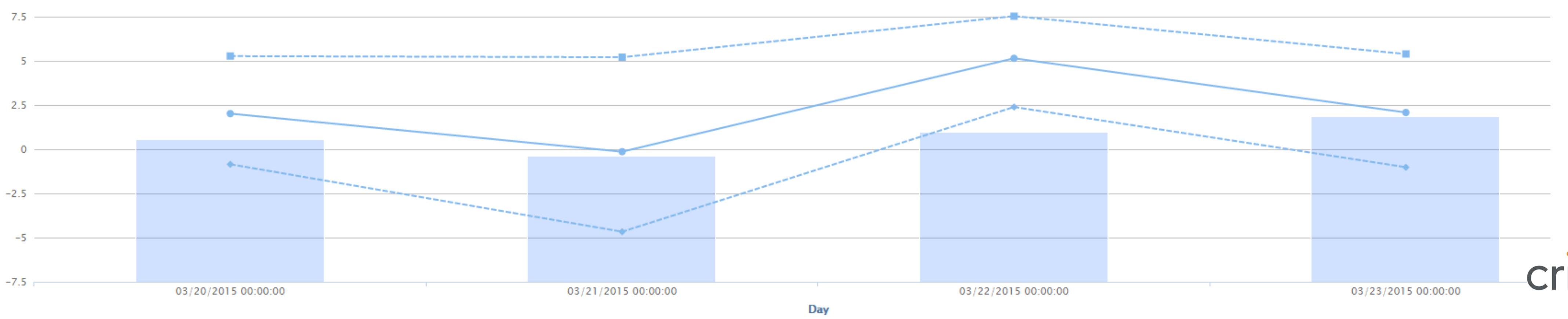
In metrics we trust

My ABTest: +0.3% on the first 2 hours … y

… but 2 days after: -1% … was it just noise

➢ Computing confidence interval using bootstrapping:

By computing metrics from several ‘instances’ of the dataset generated with random sampling, we get accuracy measures



Wrapping-up

- Prediction at the core of Criteo's platform
- Thanks to Hadoop, we could greatly distribute our learnings:
  - 1300 models learnt from 600TB daily
- Even with somewhat unorthodox implementation: 97.4% success rate
  - mono + Hadoop Streaming
  - all-reduce
- Some resources optimization needed to scale
- An integrated testbed from Offline ABTest to monitoring metrics