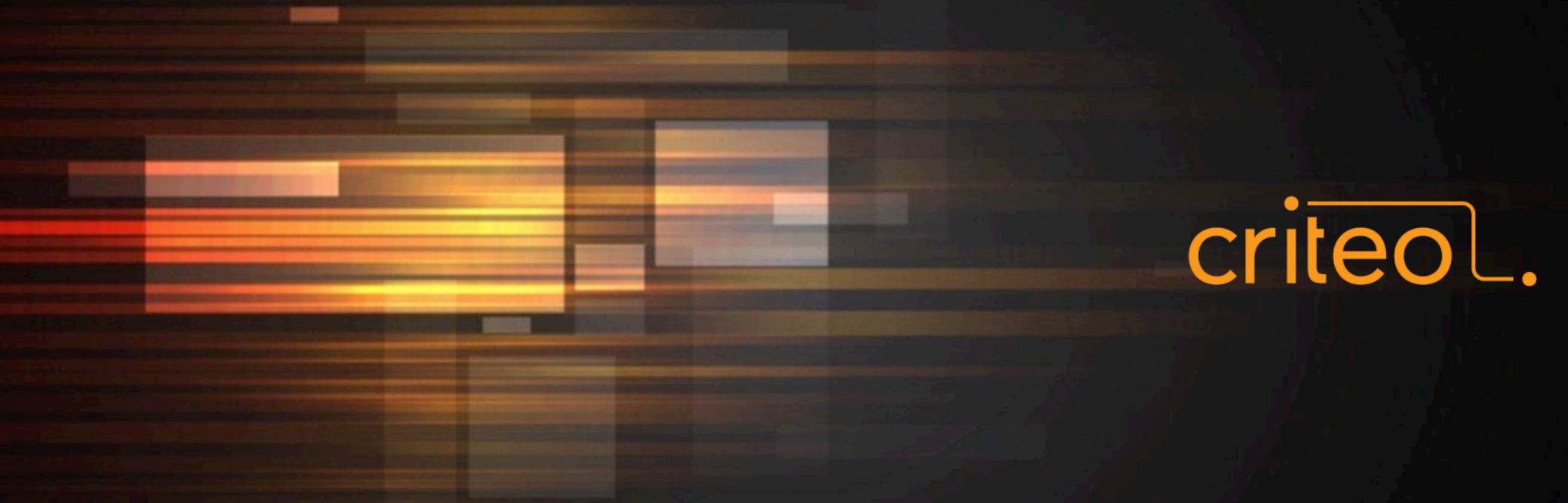
30 billion requests per day with a NoSQL architecture

USI 2013, Paris



criteo

Julien SIMON
Vice President, Engineering
j.simon@criteo.com @julsimon

PERFORMANCE DISPLAY

1

A user sees products on your website...

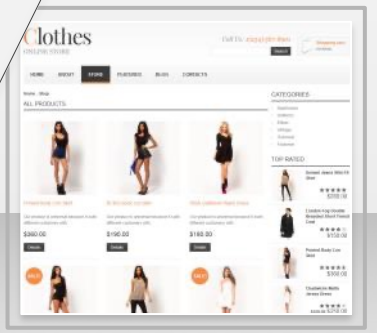
2

...then browses the internet

3

... and sees personalized banners

4



After clicking on the banner, the user goes back to the product page.

criteo

CRITEO

PHASE 1 : 2005-2008

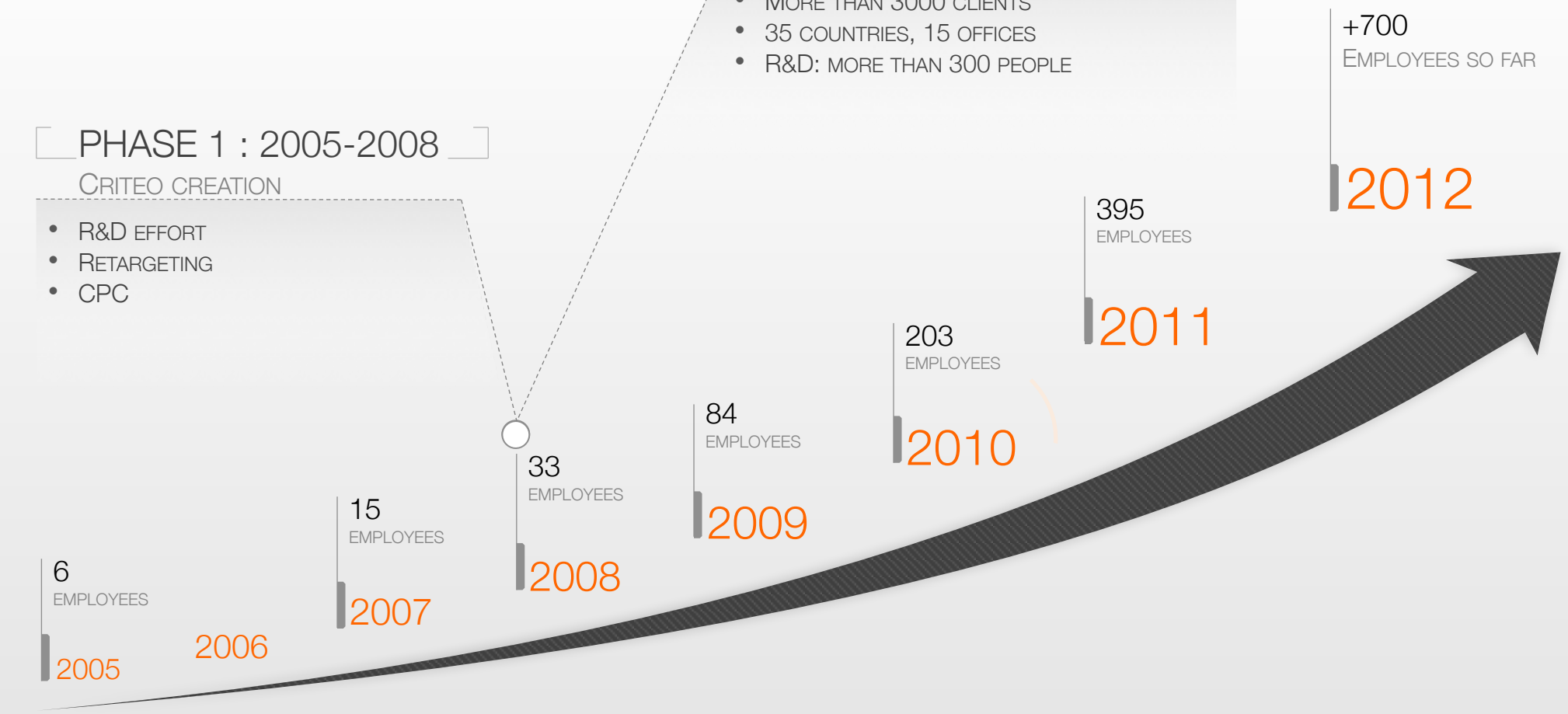
CRITEO CREATION

- R&D EFFORT
- RETARGETING
- CPC

PHASE 2 : 2008-2012

GLOBAL LEADER : + 700 EMPLOYEES!

- MORE THAN 3000 CLIENTS
- 35 COUNTRIES, 15 OFFICES
- R&D: MORE THAN 300 PEOPLE



INFRASTRUCTURE

- 7 DATA CENTERS
- SET UP AND MANAGED IN-HOUSE
- AVAILABILITY > 99.95%

- DAILY TRAFFIC
  - HTTP REQUESTS: 30+ BILLION
  - BANNERS SERVED: 1+ BILLION
- PEAK TRAFFIC (PER SECOND)
  - HTTP REQUESTS: 500,000+
  - BANNERS: 25,000+

HIGH PERFORMANCE COMPUTING

FETCH, STORE, CRUNCH, QUERY 20 additional TB EVERY DAY ?

…SUBTITLED « HOW I LEARNED TO STOP WORRYING AND LOVE HPC »

mongoDB HIVE hadoop cloudera CouchBase RabbitMQ Open Source Enterprise Messaging Storm Kafka

criteo

CASE STUDY #1: PRODUCT CATALOGUES

- Catalogue = product feed provided by advertisers (product id, description, category, price, URL, etc)
- 3000+ catalogues, ranging from a few MB to several tens of GB
- About 50% of products change every day

- Imported at least once a day by an in-house application
- Data replicated within a geographical zone
- Accessed through a cache layer by web servers
- Microsoft SQL Server used from day 1
- Running fine in Europe, but…
  - Number of databases (1 per advertiser)… and servers
  - Size of databases
  - SQL Server issues hard to debug and understand
- Running kind of fine in the US, until dead end in Q1 2011
  - transactional replication over high latency links

FROM SQL SERVER TO MONGODB

- Ah, database migrations… everyone loves them ☺
- 1st step: solve replication issue
  - Import and replicate catalogues in MongoDB
  - Push content to SQL Server, still queried by web servers
- 2nd step: prove that MongoDB can survive our web traffic
  - Modify web applications to query MongoDB
  - C-a-r-e-f-u-l-l-y switch web queries to MongoDB for a small set of catalogues
  - Observe, measure, A/B test… and generally make sure that the system still works
- 3rd step: scale !
  - Migrate thousands of catalogues away from SQL Server
  - Monitor and tweak the MongoDB clusters
  - Add more MongoDB servers… and more shards
  - Update ops processes (monitoring, backups, etc)
- About 150 MongoDB servers live today (EU/US/APAC)
  - Europe: 800M products, 1TB of data, 1 billion requests / day
  - Started with 2.0 (+ Criteo patches) → 2.2 → 2.4.3

MONGODB, 2.5 YEARS LATER

- Stable (2.4.3 much better)
- Easy to (re)install and administer
- Great for small *datasets* (i.e. smaller than server RAM)
- Good performance if read/write ratio is high
- *Failover* and inter-DC replication work (but shard early!)

- Performance suffers when :
  - *dataset* much larger than RAM
  - read/write ratio is low
  - Multiple applications coexist on the same cluster

- Some scalability issues remain (master-slave, connections)
- Criteo is very interested in the 10gen *roadmap* ☺

criteo

CASE STUDY #2: HADOOP

1st cluster live in June 2011
(2 Petabytes)

« Express » launch required
by brutal growth of traffic

Traditional processing
(in-house tools + SQL Server)
completely replaced by Hadoop

Dual-use: production
(prediction, recommandation, etc.)
and Business Intelligence
(reporting, traffic analysis)

Visible ROI :
Increase of CTR and CR

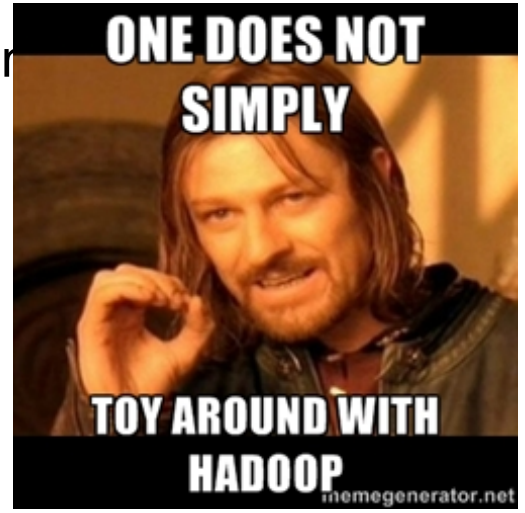
2nd cluster live in April 2013
(6 Petabytes→?)

criteo

HADOOP IS AWESOME… BUT *CAVEAT EMPTOR!*

- *Batch processing* architecture, not real-time → hence our work on Storm
- Beware how data is organized and presented to jobs (LZO, RCFile, etc) → hence our work on Parquet with Twitter & Cloudera
- *namenode* = SPOF → backup + HA in CDH4
- Understand HDFS replication (*under-replicated blocks*)
- Have a stack of extra hard drives ready

- Lack of ops / prod tools: data import/export, monitoring, r
- Lots of work needed for an efficient multi-user setup: scheduling, quotas, etc.
- At scale, infrastructure skills are mandatory
  - Server selection, CPU/storage ration
  - Linux & Java tuning
  - LAN architecture: watch your switches!



THANKS A LOT FOR YOUR ATTENTION!

www.criteo.com
engineering.criteo.com

criteo