

Yandex



High-Performance Distributed DBMS for Analytics

Victor Tarnavsky, Head of Analytic Systems Department

ClickHouse

Story



Yandex

- › One of the largest internet companies in Europe
- › Over 5000 employees
- › Top-1 Search in Russia
- › More than 50 different b2c and b2b products
- › Big Data, Machine Learning



The Product

Yandex Metrica

Yandex.Metrica is 2nd largest web analytics tool in the world

- › 30+ billions of events daily
- › Millions of websites
- › 100+ thousands of analysts every day

We need fast and feature-rich database capable to handle our clients data

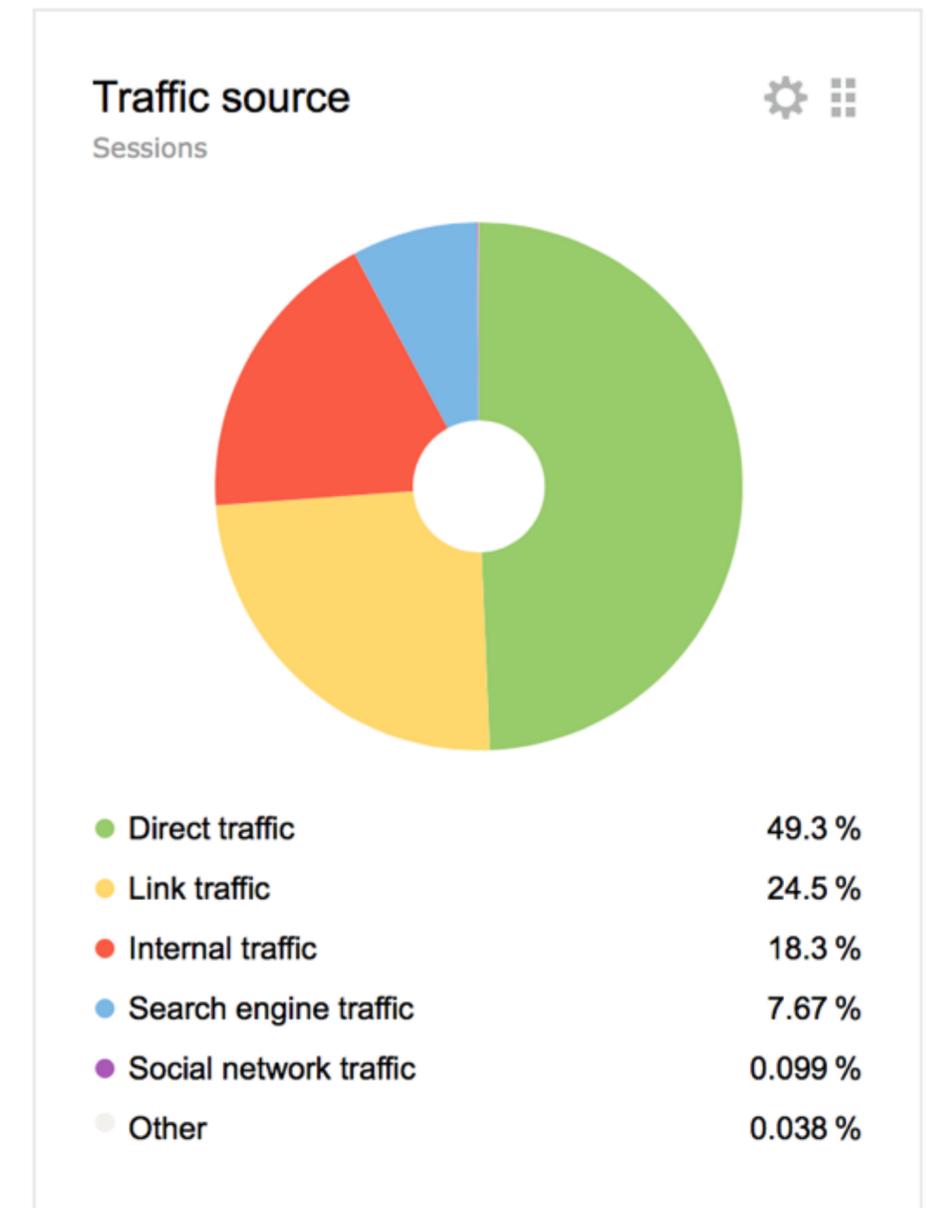
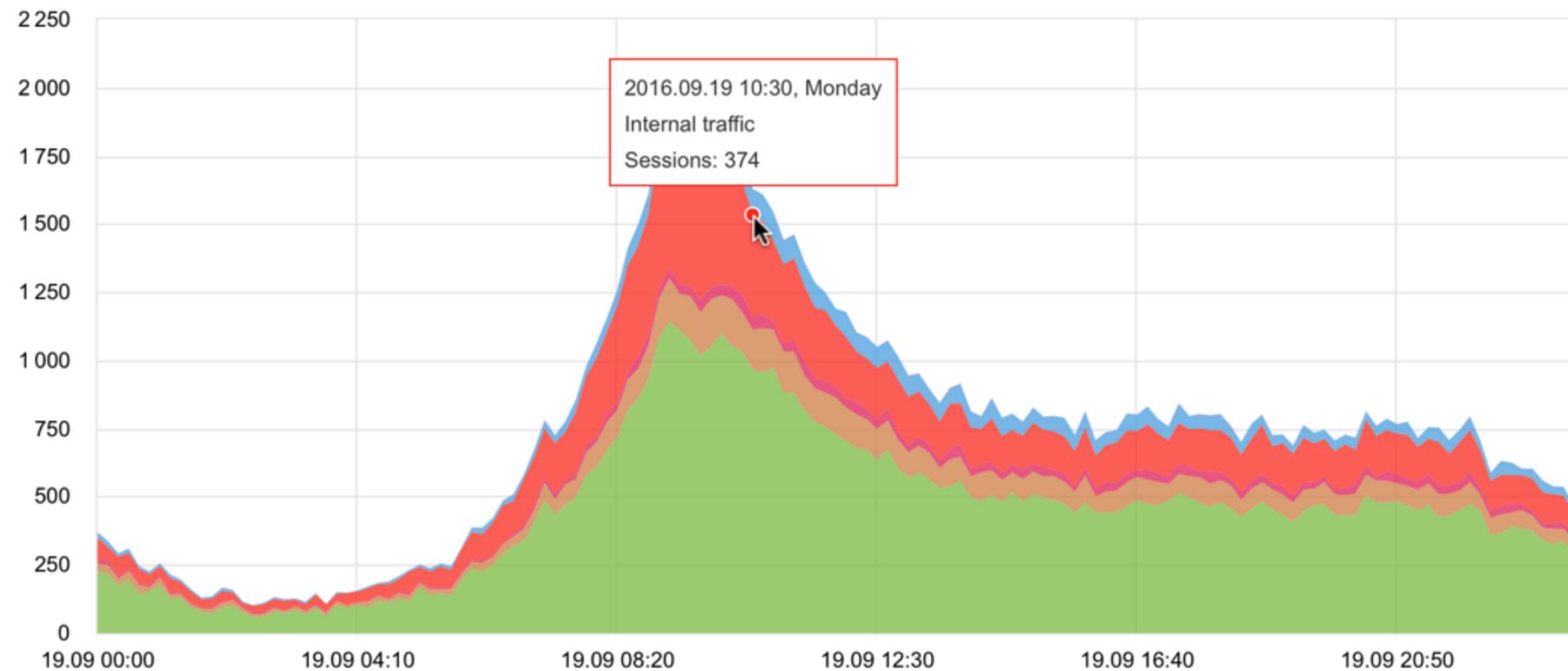
Yandex.Metrica

Today Yesterday Week Month Quarter Year 19 Sep 2016 Group: by 10 minutes

Segment: 2 conditions Compare segments Accuracy: 100% Attribution: Last visit

Sessions in which Session number > 3 for people with Gender: male

Sessions



Before ClickHouse

2008-2011: MySQL

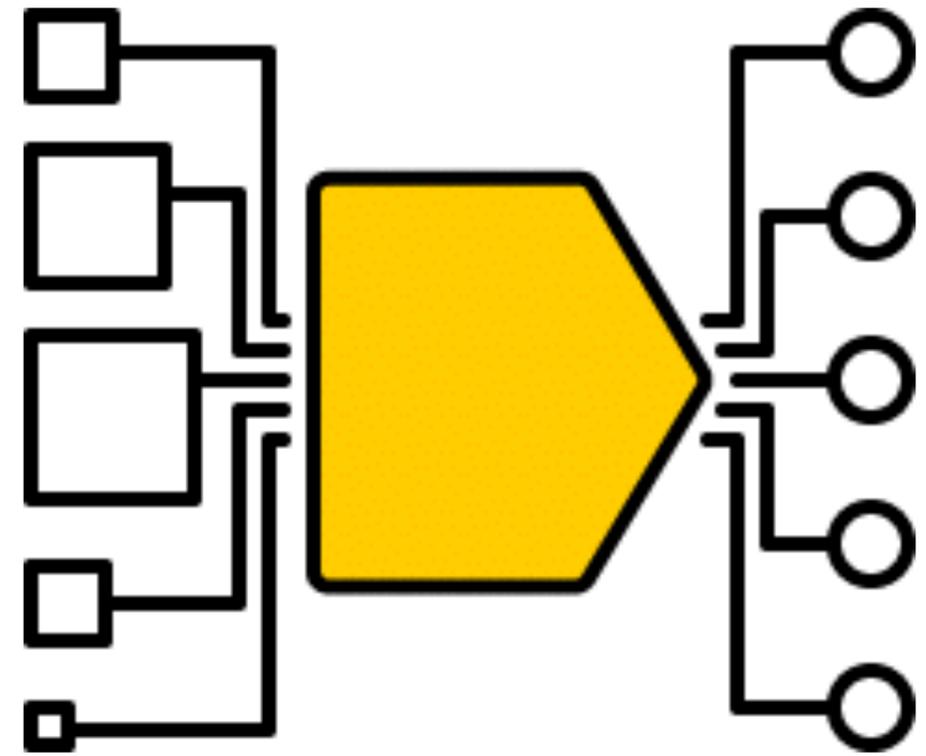
- › Slow and tricky

2010-2015: Metrage, custom aggregated data storage

- › Fast and realtime
- › Aggregated data can't be filtered

Requirements

- › Fast. Really fast
- › Data processing in real time
- › Capable of storing petabytes of data
- › Fault-tolerance in terms of datacenters
- › Flexible query language



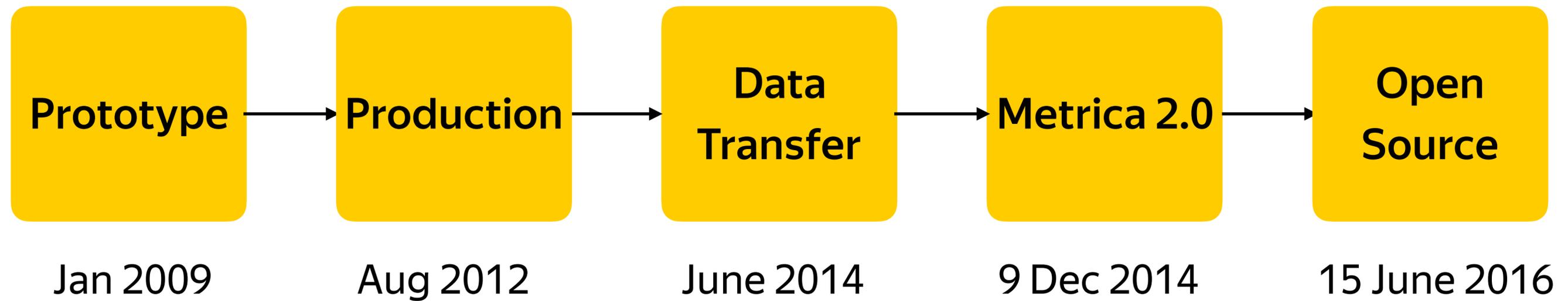
| Nothing? Oh. Well...

The main ideas behind ClickHouse

- › SQL
- › Linearly scalable
- › Focused on fast query execution
- › Realtime
- › Column-oriented



ClickHouse timeline



ClickHouse

Today



ClickHouse today

- › Open-source, Apache 2.0
- › 100+ companies outside Yandex
- › Strong community
- › Active development

Clement 'cmc' Rey @teh_cmc Follow

Yantex's #ClickHouse is, by far, the most beautifully engineered & documented data system I've ever worked with. clickhouse.yandex

RETWEETS 4 LIKES 6

4:45 PM

Benjamin Eberlei @beberlei Follow

Played around with #clickhouse by yandex today, it looks like an impressive database for timeseries clickhouse.yandex

RETWEETS 5 LIKES 9

[yandex / ClickHouse](#)

Watch ▾ 181

Unstar 2,293

Fork 285

Opensource

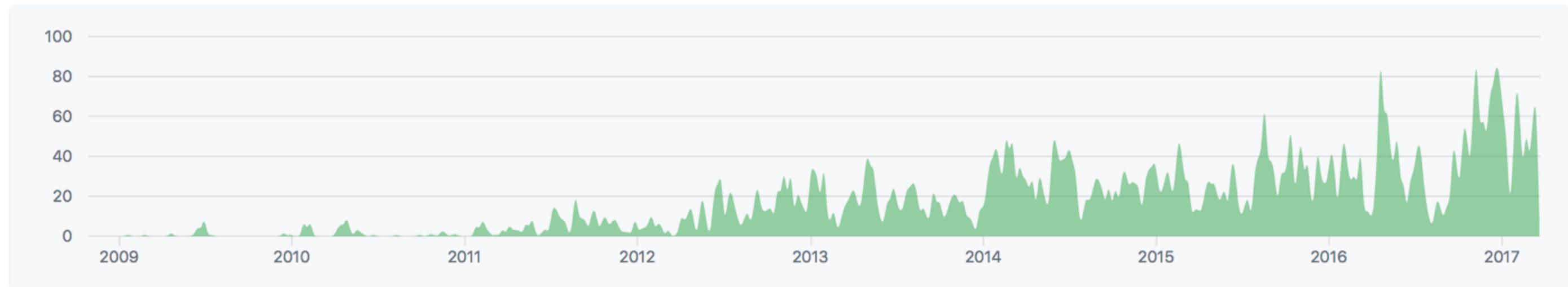
yandex / ClickHouse

Watch 181

★ Unstar 2,293

Fork 285

- › Apache 2.0
- › 53 committers on main repo
- › Committers from CloudFlare, [booking.com](https://www.booking.com)
- › 100+ companies using ClickHouse



Community

- › 700+ people in Telegram chats, active every day
- › 102 side projects on GitHub: drivers, clients, interfaces etc.
- › Tabix: web interface over ClickHouse

› Integrations:

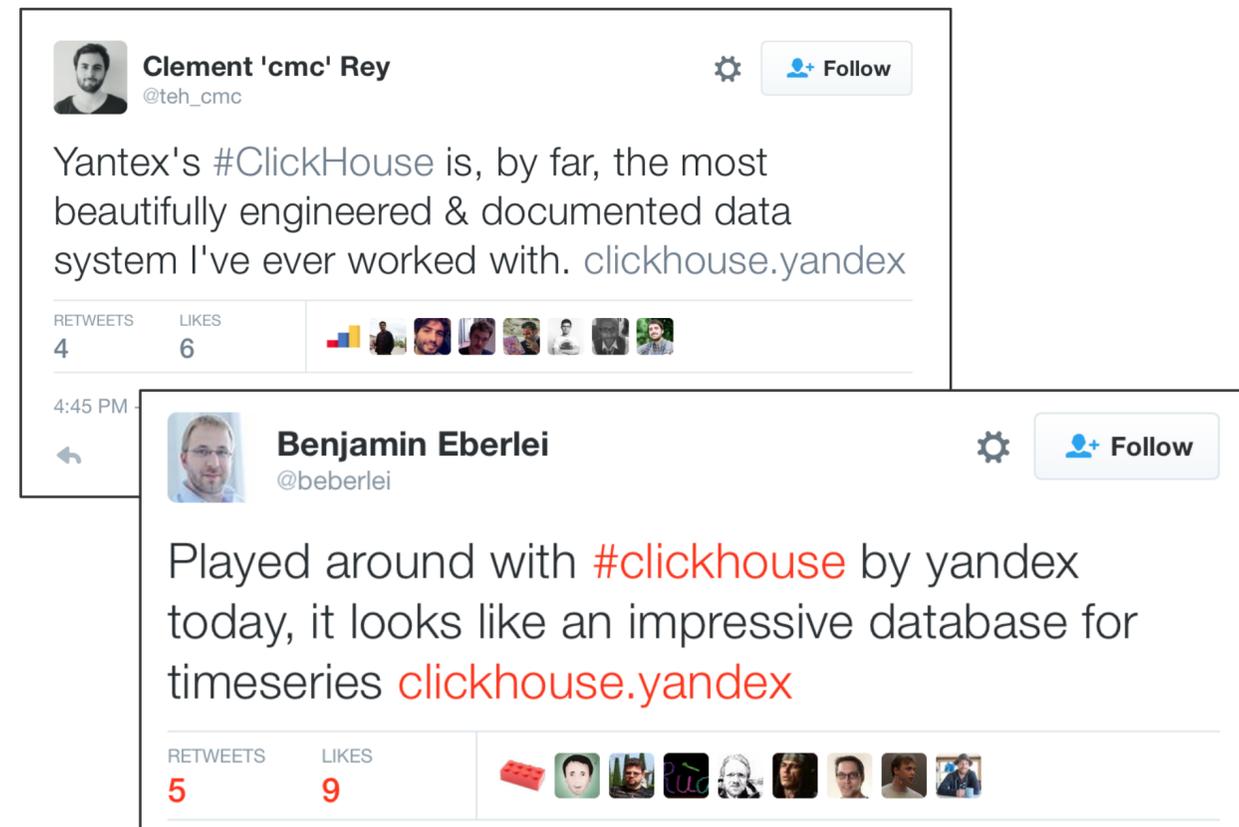
Grafana

Redash

Apache Zeppelin

Superset

Power BI



ClickHouse

Features



Scalable

- › Petabytes of data
- › Cross-datacenter
- › High availability
- › Data compression



Metрика Cluster

- › 20+ trillions of rows
- › 3 Pb
- › 450 Servers
- › 6 Datacenters
- › Up to 2 terabytes per second on query processing



Querying

- › SQL dialect + extensions
- › Additional features: approximate functions, URI functions and more
- › Arrays, nested data types
- › Distributed out of the box
- › Pluggable external key-value dictionaries



```
# Weekly traffic and audience
```

```
SELECT
```

```
    count() as visits,
```

```
    sum(PageViews) as hits,
```

```
    uniq(UserID) as users
```

```
FROM visits_all
```

```
WHERE StartDate > today() - 7
```

```
# Using dictionary for regions
```

```
SELECT
```

```
    count() as visits,
```

```
    regionToName(regionToCountry(RegionID), 'en') as country
```

```
FROM visits_all
```

```
WHERE StartDate > today() - 7
```

```
GROUP BY country
```

```
ORDER BY visits DESC
```

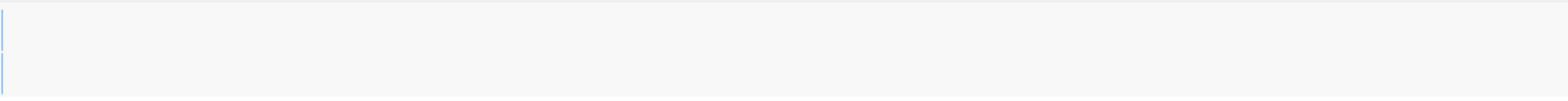
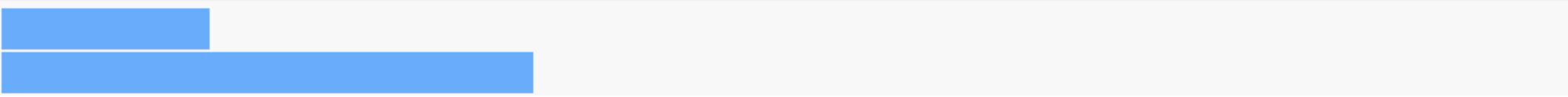
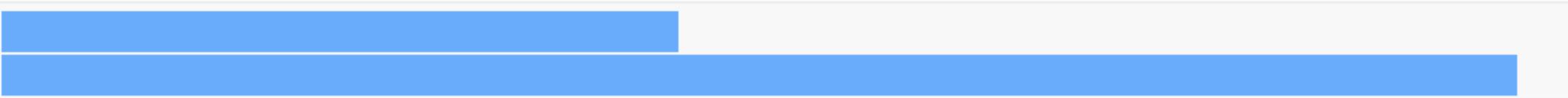
```
LIMIT 10
```

Performance

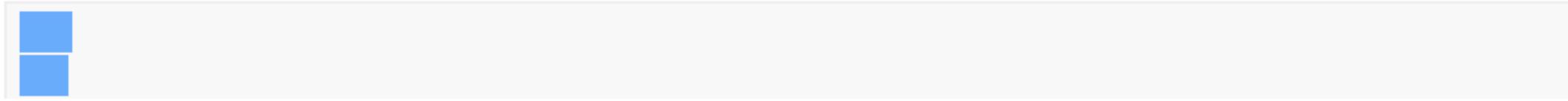
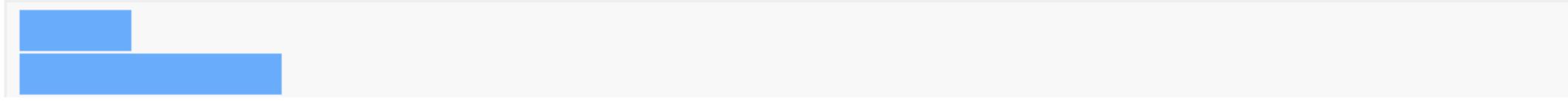
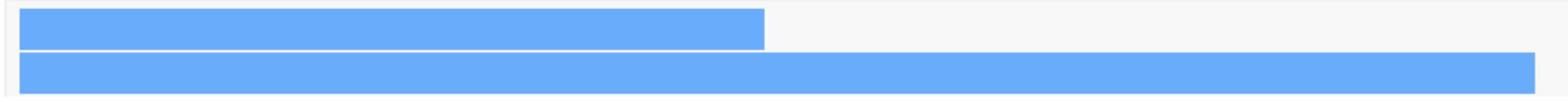
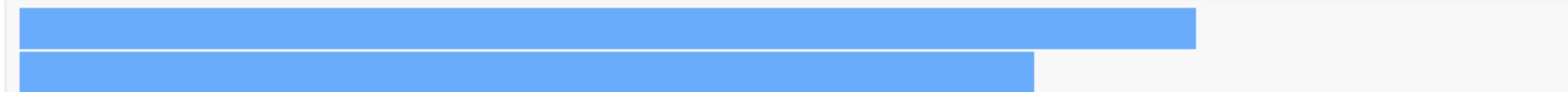
- › Sub-second query latency
- › >100x faster than Hadoop,
>100x faster than typical DBMS
- › Up to a few billion rows/second
per single node
- › Up to 2 terabytes per second
on clustered setup of 400 nodes



Relative query processing time (lower is better):

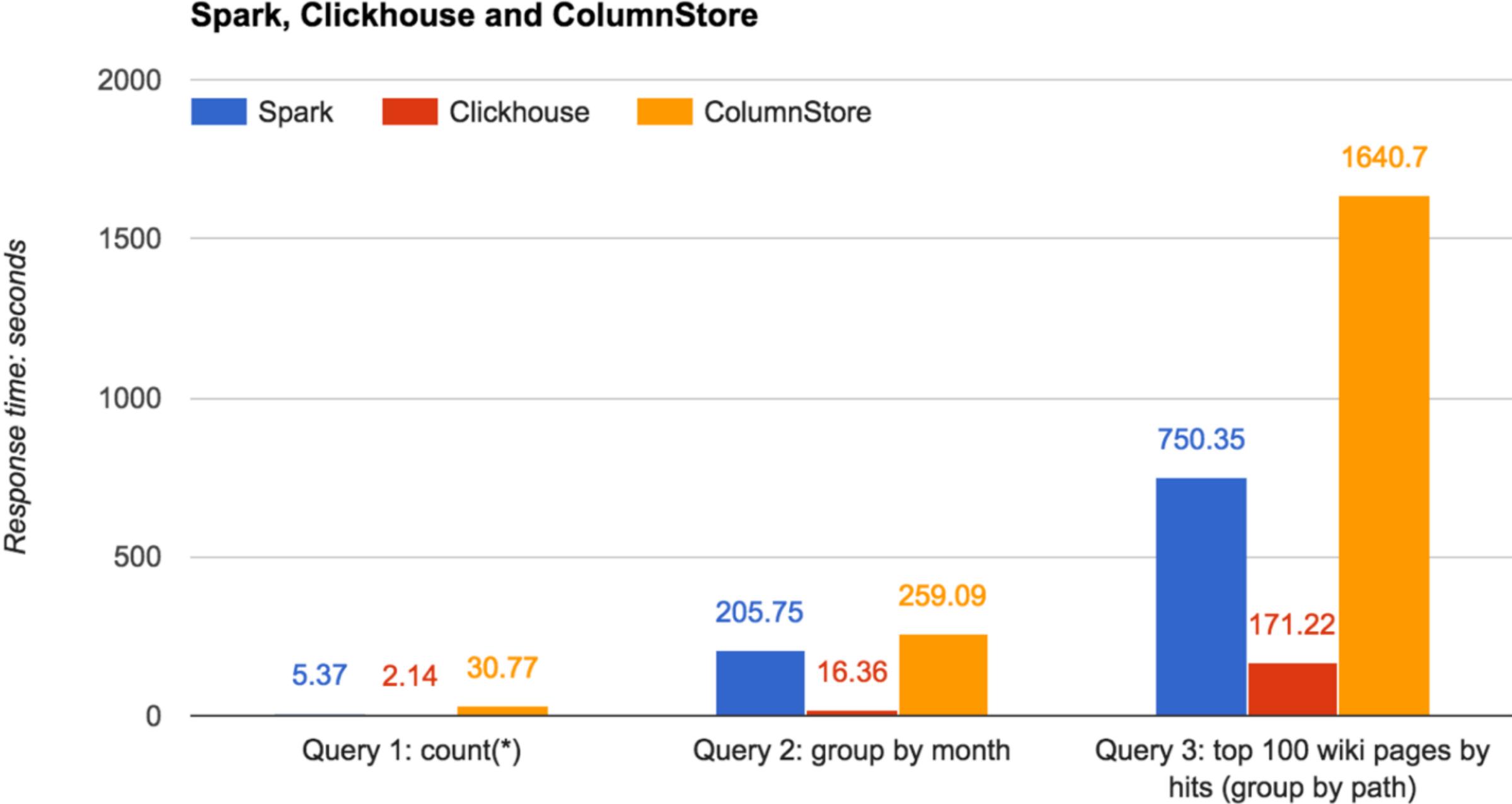
ClickHouse (1.1.53960)		1.00
Hive (0.11, ORC File)		168.19
MySQL (5.5.32, MyISAM)		511.57

Relative query processing time (lower is better):

ClickHouse (1.1.53960)		1.00
Vertica (7.1.1)		3.38
InfiniDB (Enterprise 3.6.23)		20.91
MonetDB		21.51

More info: <https://clickhouse.yandex/benchmark.html>

ClickHouse vs. Spark and MariaDB ColumnStore



More benchmarks

- › 1.1 Billion Taxi Rides on ClickHouse by Mark Litwintschik
<http://tech.marksblogg.com/billion-nyc-taxi-clickhouse.html>
- › ClickHouse vs. other Open-source Databases by Percona
<http://bit.ly/2pfd9aF>

Interfaces

- › Console client
- › HTTP
- › JDBC Driver, ODBC Driver in beta
- › Clients for:
Python, PHP, Go, Node.js
Perl, Ruby, R, C++
.NET, Scala, Julia



ClickHouse

Inside



Why is it so fast?

Code

- › Vectorized query execution
- › Low-level optimisations and specialisations
- › Every piece of code is tested in terms of performance



Why is it so fast?

Data

- › Column-oriented
- › Merge Tree
minimal number of seeks
- › All processing as close to data as possible



Why is it so fast?

Features

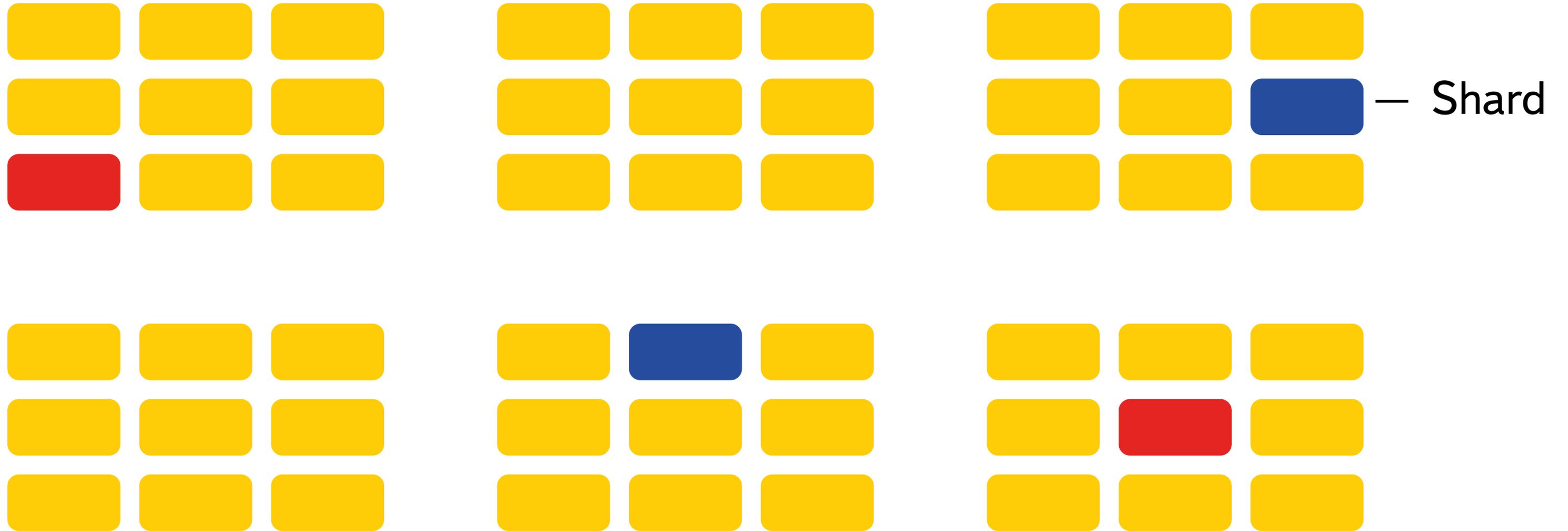
- › Sampling
- › Approximate functions
- › Performance tuning even on a request level



Scalability and fault-tolerance

- | Hours of downtime on Metrica cluster for years
 - › Cross-datacenter
 - › Asynchronous replication
 - › Eventual consistency

Cluster Scheme



Datacenter

ClickHouse

Use cases



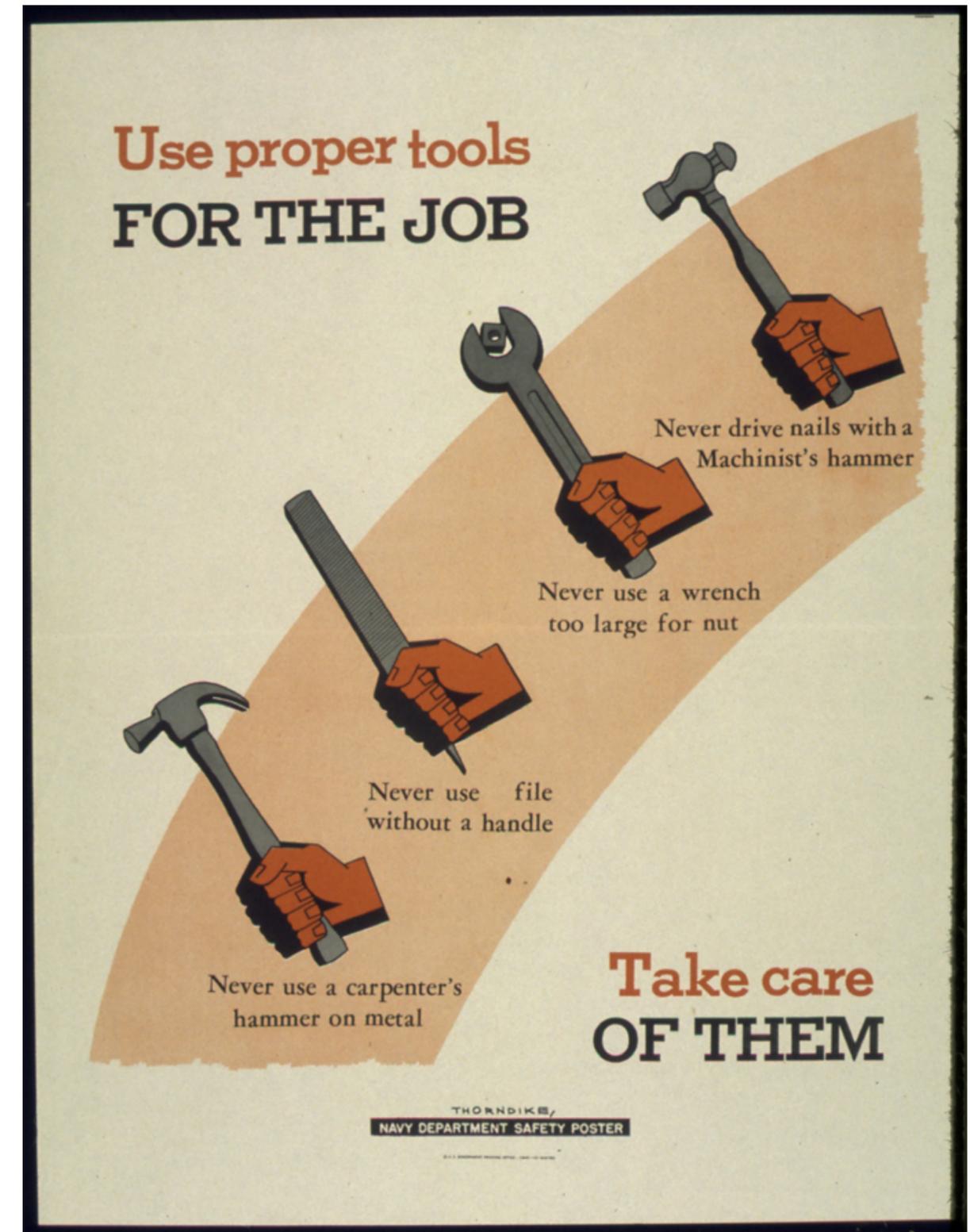
ClickHouse: wrong cases

- › Not an OLTP
- › Not a key-value store
- › Not a document store
- › Do not modify your data
(you don't need that)



ClickHouse: best practices

- › A few wide tables with a lot of columns
- › Structured data
- › QPS is relatively low but data usage per request is high
- › Huge amounts of data incoming
- › Petabytes of data



Typical use cases

Adv networks data
RTB
Web/App analytics

Ecommerce
Telecom logs
Online games

Sensor data
Monitoring
Messengers

Unusual applications

- › **Blockchain search and analytics engine**
<https://blockchair.com/>
- › **Evolutionary genetics and genomics (analyzing BLAST database)**
<https://github.com/msestak/FindOrigin>
- › **Nuclear research: CERN's LHCb experiment**
https://www.yandex.com/company/press_center/press_releases/2012/2012-04-10/

Case: server log analysis

Common first case for new ClickHouse users.

Estimated time: few hours

- › Insert access logs into ClickHouse
- › Analyze incidents with instant queries
- › Monitoring reports: error rates, response timings and more

Case: in-house analytics database

Build your own data warehouse and dig your data in seconds.

- › Take your Hadoop or other 'not so fast' storage and want to do things faster
- › Copy all your data to ClickHouse
- › Build internal dashboards/metrics
- › Do realtime analysis of your business process

ClickHouse

Wrap up



ClickHouse briefly

- › Open-source column-oriented DBMS
- › Linearly scalable
- › Blazingly fast
- › SQL dialect with extensions
- › Strong community



ClickHouse

How to start?



- › **Try our tutorial:** <https://clickhouse.yandex/tutorial.html>
- › **Ask anything:** clickhouse-feedback@yandex-team.com
- › **GitHub:** <https://github.com/yandex/ClickHouse>
- › **Telegram:** https://t.me/clickhouse_en
- › **More info:** <https://clickhouse.yandex>

Contacts

Viktor Tarnavskiy

Head of Analytic Products Department in Yandex



jkee@yandex-team.ru



clickhouse-feedback@yandex-team.ru