





促进软件开发领域知识与创新的传播



关注InfoQ官方信息
及时获取QCon软件开发者
大会演讲视频信息



[北京站] 2016年12月2日-3日
咨询热线: 010-89880682



[北京站] 2017年4月16日-18日
咨询热线: 010-64738142

EasyBI

Real-time Multi-dimensional BI Application

[Sheng Zhao](#)

Business Analytics Manager
LinkedIn

Agenda



User

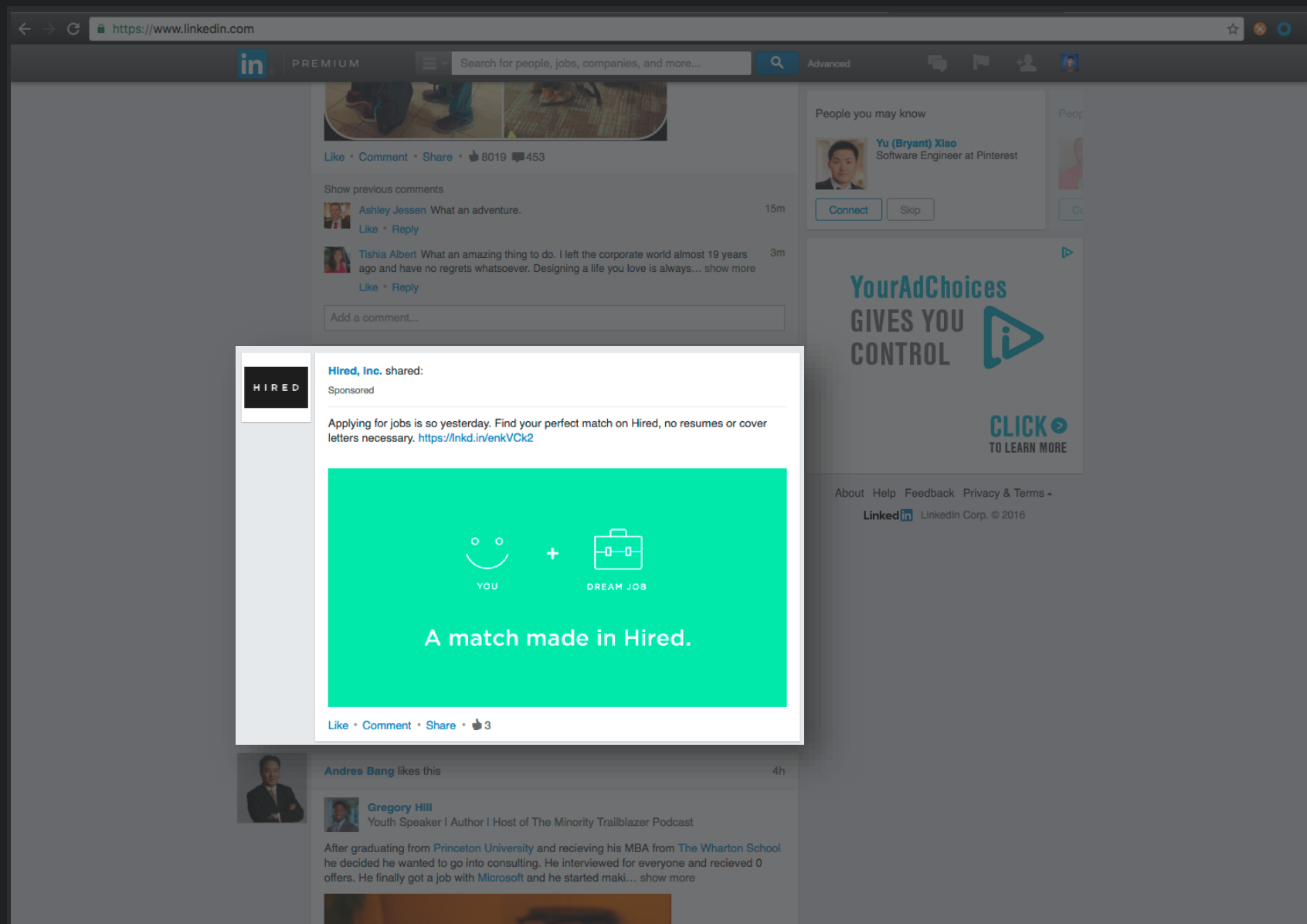


Application



OLAP Server

Use Case Example



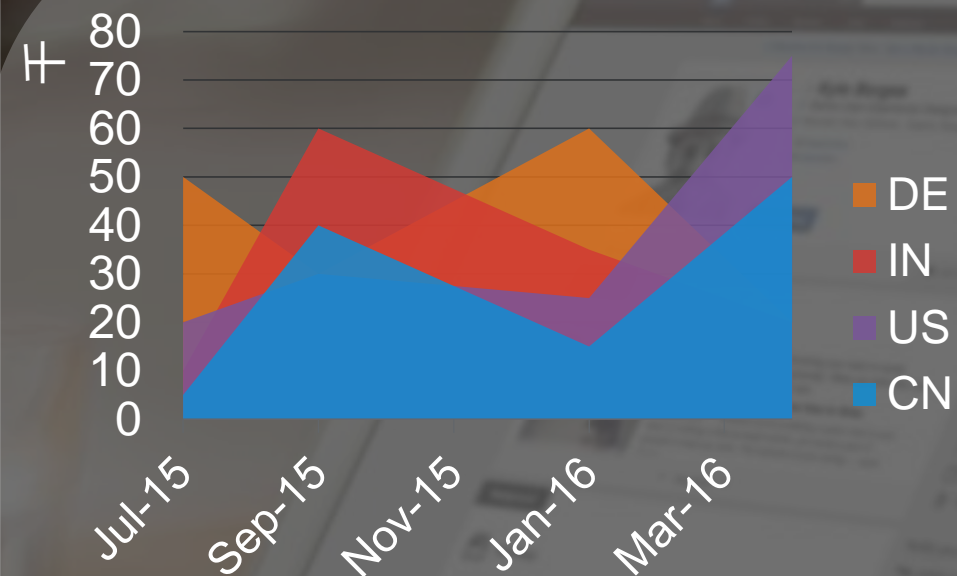
Dashboard Example

07/01/2015 – 06/30/2016

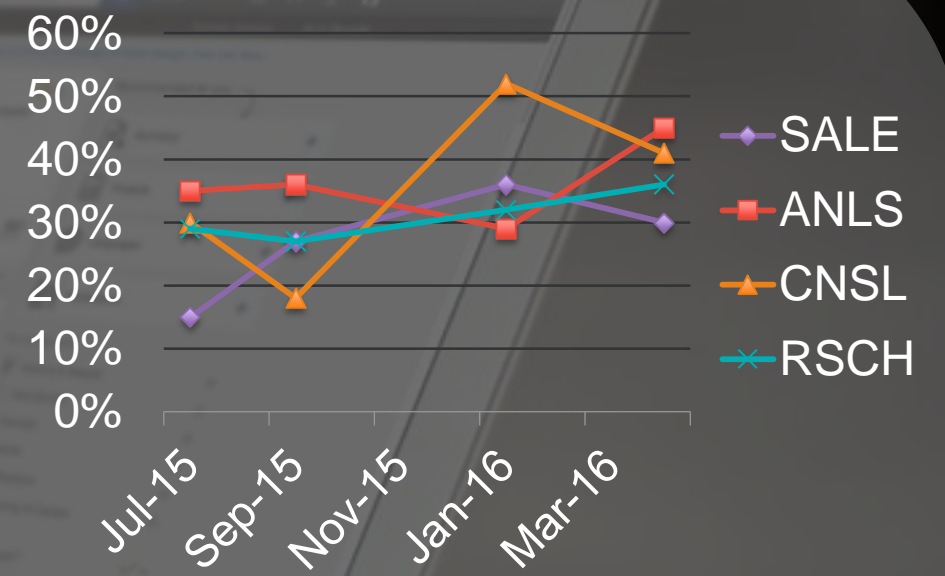
Dimension Filter

- Platform
- Function
- Geo
 - Country
 - City
- Industry
- Seniority
- Account Level
- Etc.

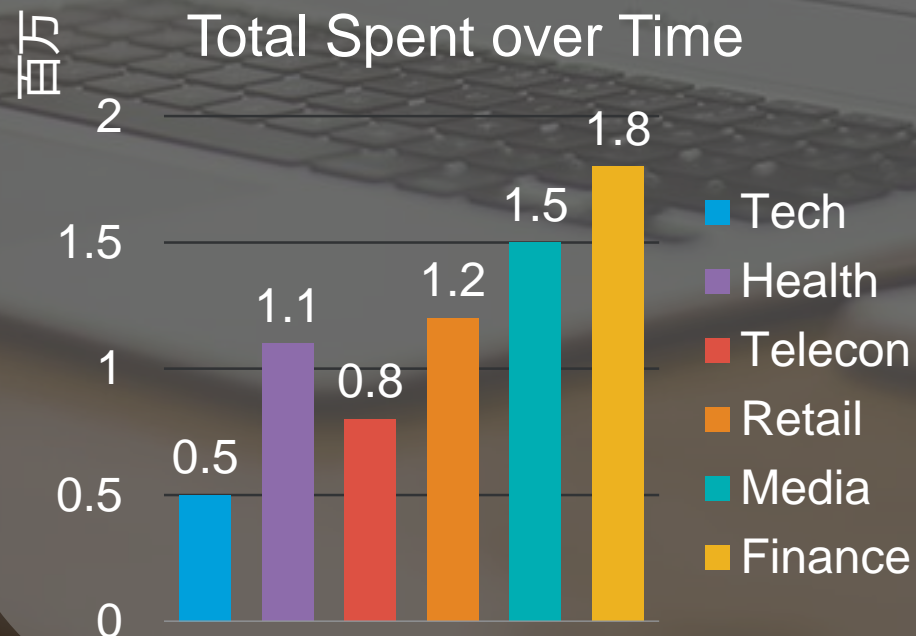
Total Click by Country



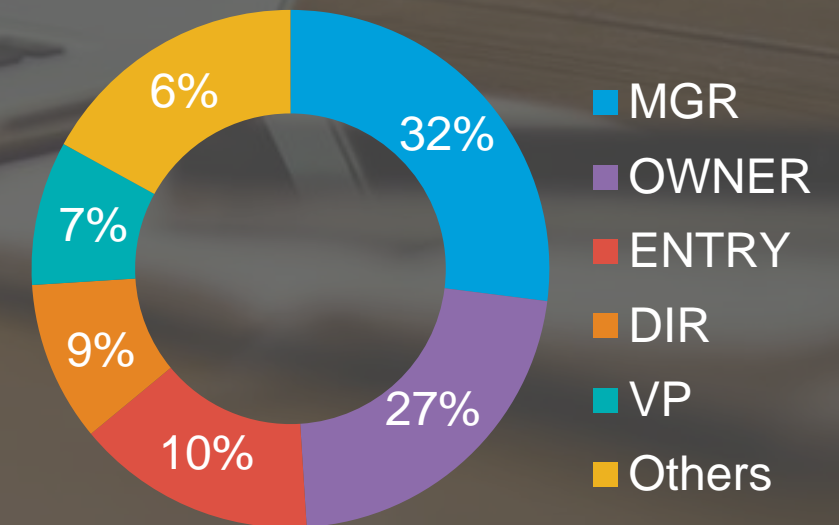
CTR by Function



Total Spent over Time



Total Engagement by Level



Source: As mocked up by presenter, don't take it seriously

263 * 148 * 12 * 12 * 168 * N
countries industries functions seniorities browsers dashboards

> 110B

Problem to Resolve



User



Application Server



OLAP Server

Pinot

A near-realtime distributed OLAP datastore



- A column-oriented database
- Pluggable indexing technologies
- Horizontally scalable and fault tolerant
- SQL like language

Problem to Resolve



User



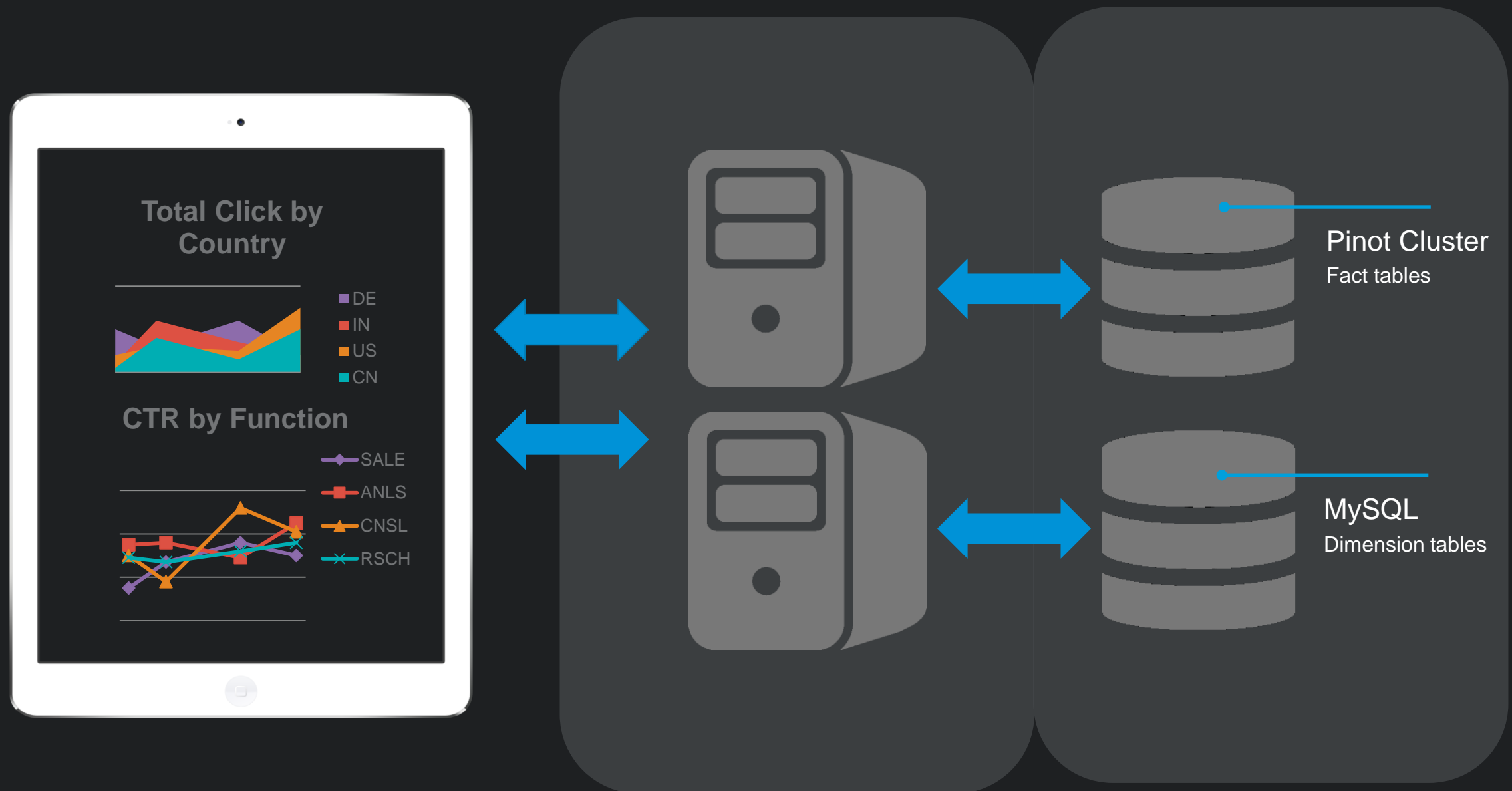
Application Server



OLAP Server

Take a Glance

A tiny move



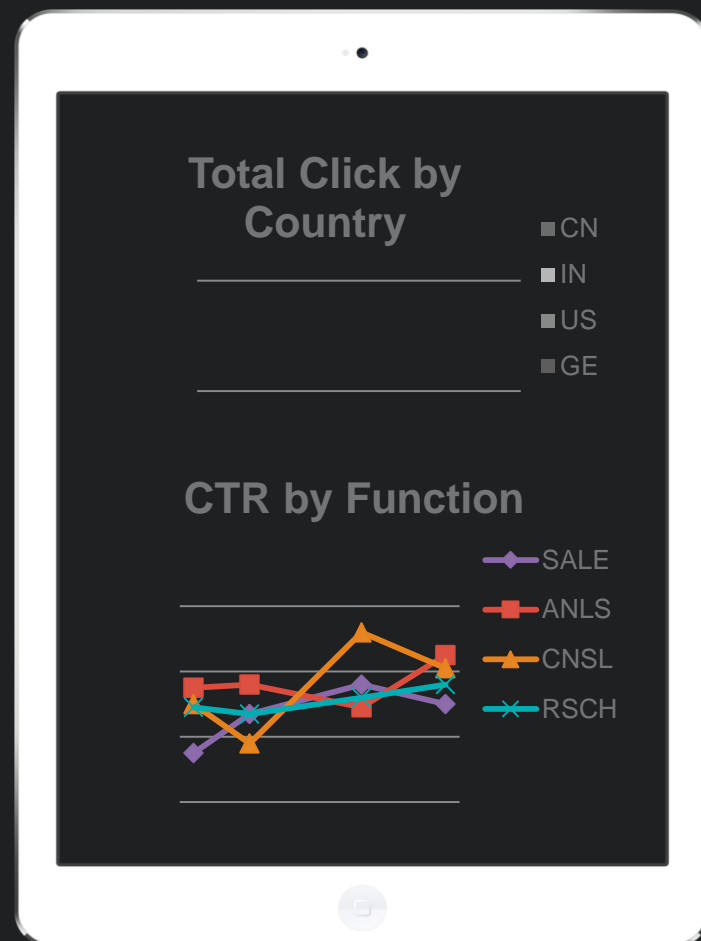
Dimension Tables

Why on MySQL?

- Requires frequent modification
- Join between dimension and fact handles by application
- Limited on the data volume
- A MySQL is needed to serve all application metadata

Something Challenging

Oops, random failure



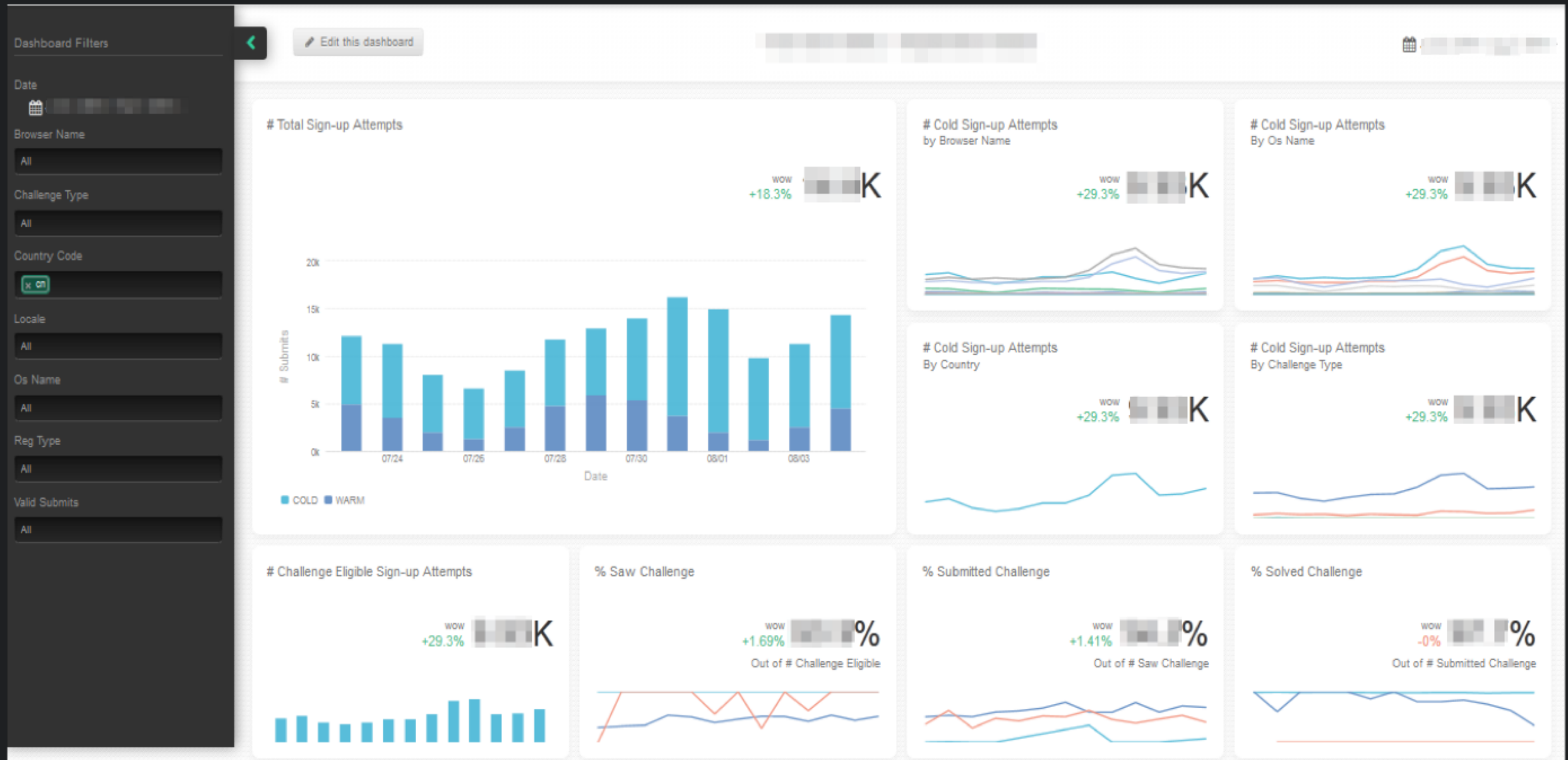
- Partial failure
- Performance drop

What's Wrong?



- OLAP Server Timeout
 - High Concurrency
 - Large Aggregation Volume

A Dashboard



Combine Queries

Browser

Challenge
Type

Country

Locale

OS Name

Reg Type

Valid
Submits

M: Sign-up

D: cold/warm

M: Cold Sign-up

D: Browser Name

M: Cold Sign-up

D: OS Name

M: Challenge

D:

M: Cold Sign-up

D: Country

M: Cold Sign-up

D: Challenge Type

M: % Saw
Challenge

D: Country

M: Submitted
Challenges

D: Country

M: Solved
Challenges

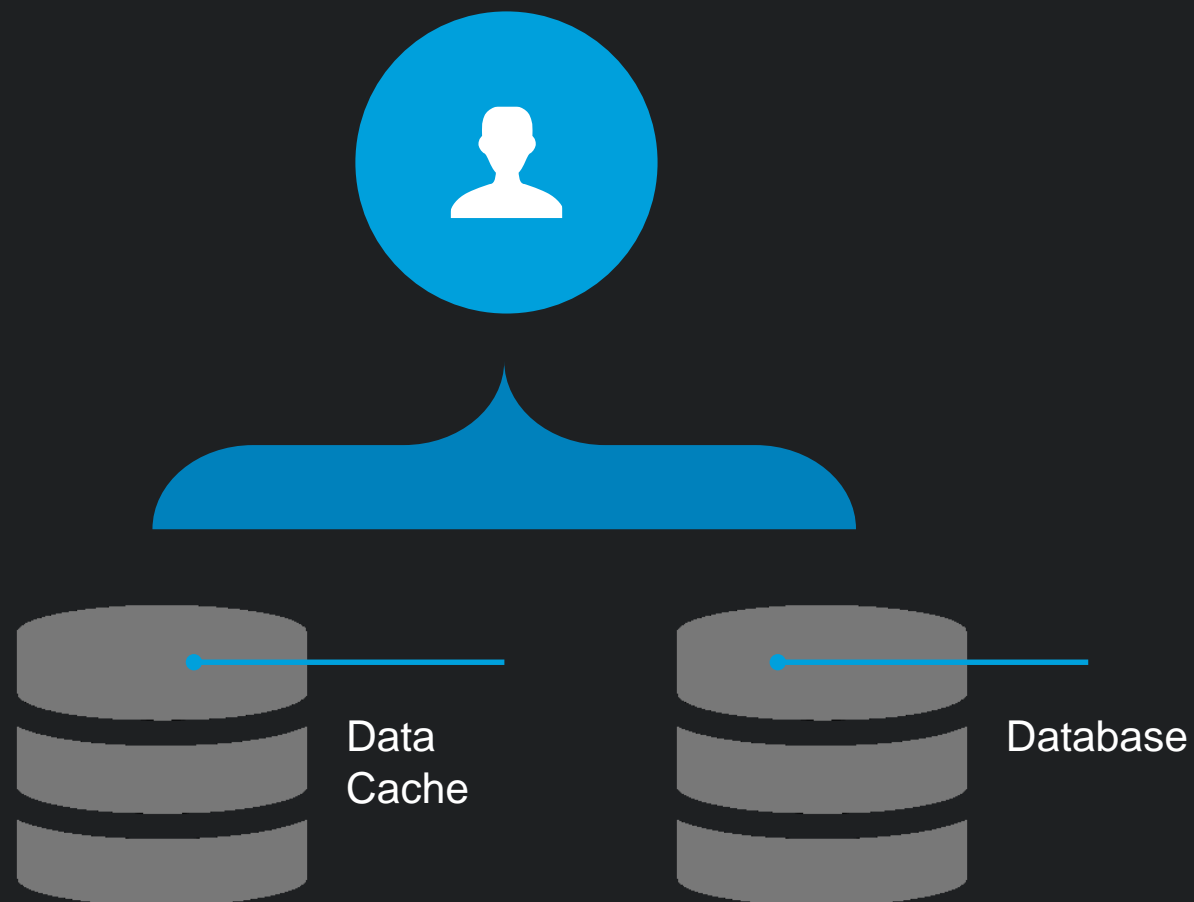
D: Country

Dimension Filter Compression

| | |
|----------------|------|
| Browser | > 10 |
| <hr/> | |
| Challenge Type | < 10 |
| <hr/> | |
| Country | > 10 |
| <hr/> | |
| Locale | > 10 |
| <hr/> | |
| OS Name | > 10 |
| <hr/> | |
| Reg Type | < 10 |
| <hr/> | |
| Valid Submits | < 10 |

- Dimensional filters have different cardinalities
- Limit on high cardinality filters
 - Default N
 - Top N

Hybrid Query with Data Cache



- Cache server response in browser
- Client side aggregation
- Request data only necessary

Combine Queries Stratagem

1. Always combine queries with same dimension
2. Expend combined queries to filters
3. Combine queries with different dimensions when possible
4. Repeat until hit the dashboard is processed

Problem to Resolve



User

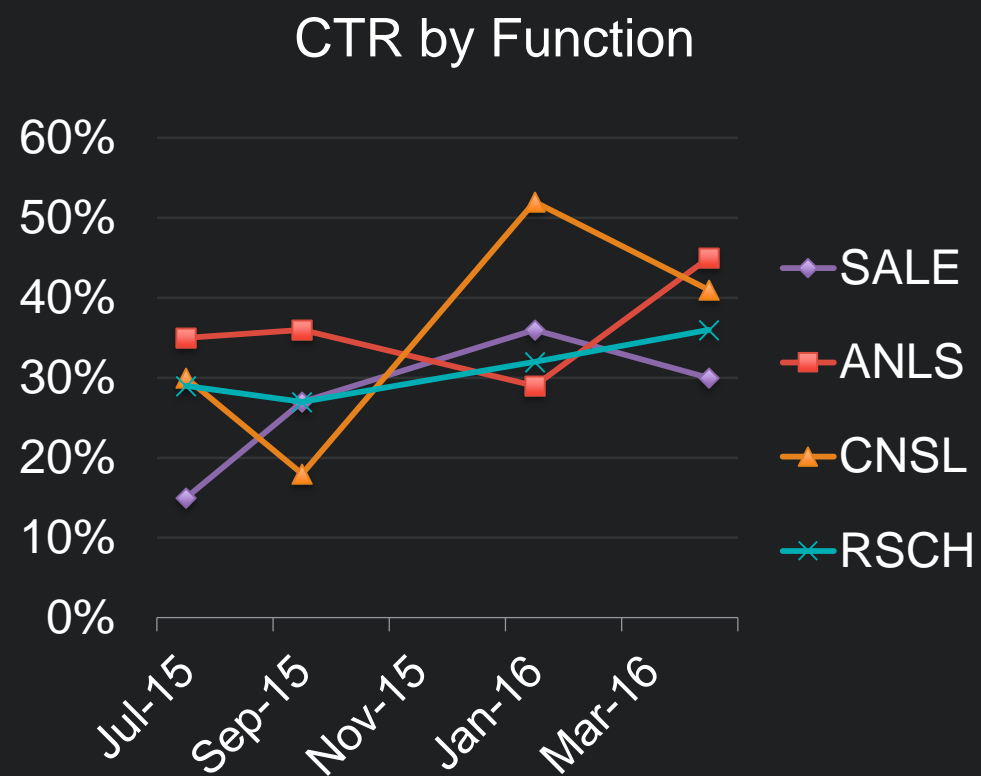


Application Server




OLAP Server

Optimizations



- Split queries by date
- Asynchronized queries

A photograph of a workspace with a laptop, a glass of water, and a smartphone, all under a dark overlay with white text.

45%

fewer queries after
combination

Future Tasks

1. Automatic dimension filter optimization
2. Adjustable combination of threshold
3. Enable real-time data

Appendix

Open Source Projects

Pinot

<https://github.com/linkedin/pinot>

Crossfilter

<http://square.github.io/crossfilter/>



*Our mission is to connect the world's
professionals to make them more
productive and successful.*

Jeff Weiner



*Our vision is to create economic
opportunity for every member
of the global workforce.*

Jeff Weiner

