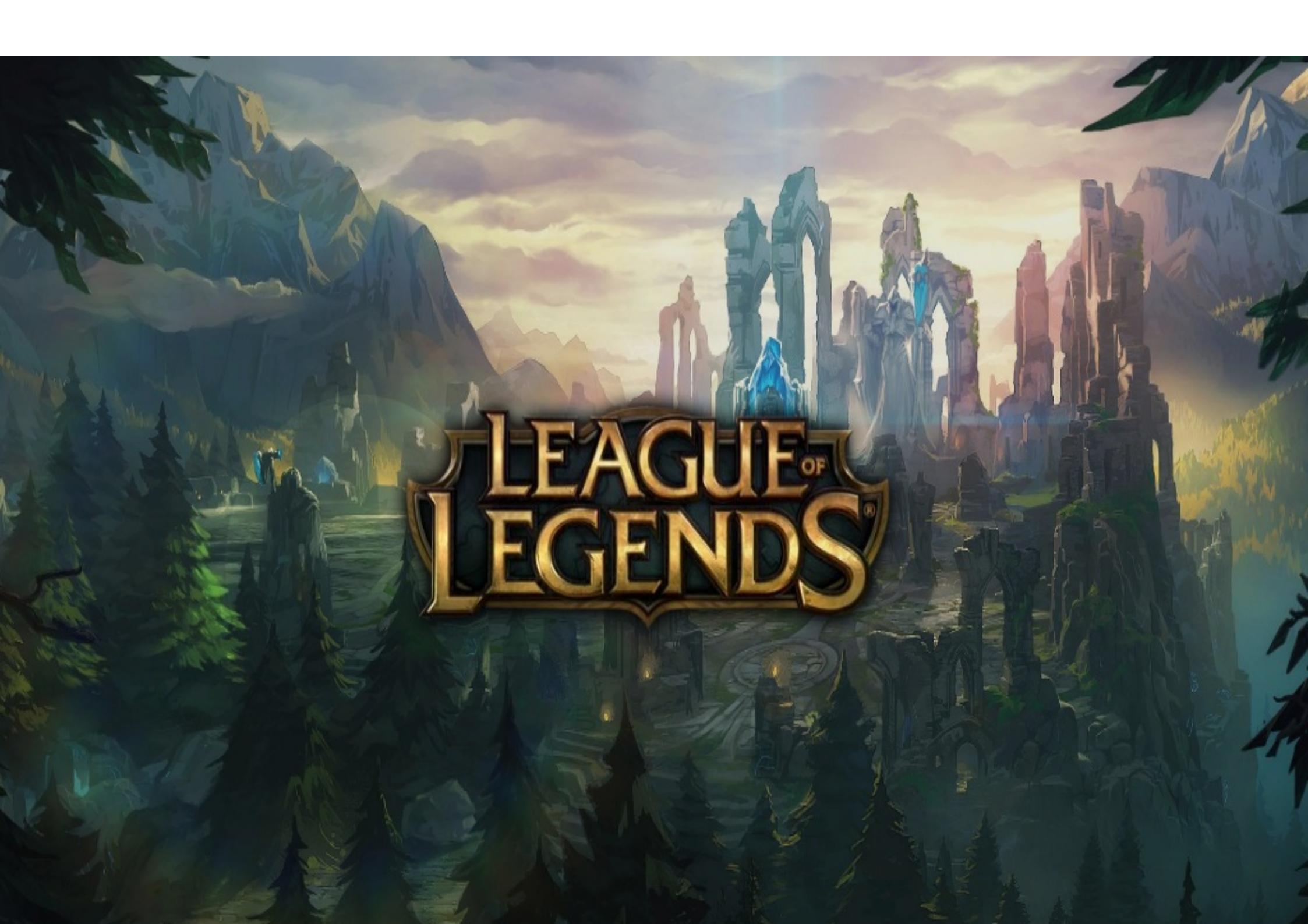




# VIDEO GAMES AT SCALE

---

Improving the  
gaming experience  
with **Spark**



# LEAGUE OF LEGENDS®

## Choose

From over 120 champions, each having a unique backstory and abilities

## Compete

With your team to complete objectives and battle the enemy team.

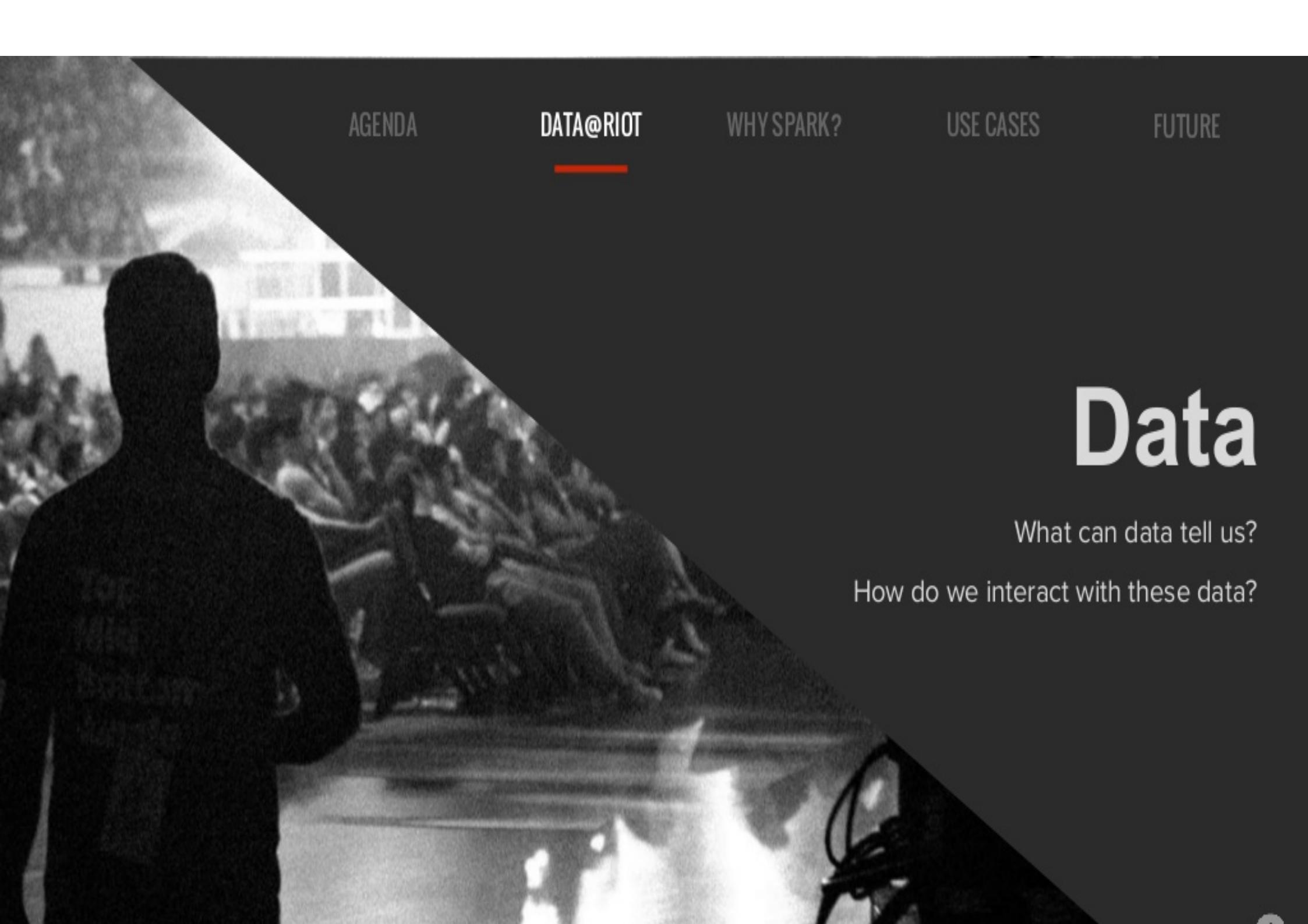
## VICTORY

## Win

Take down defenses and destroy the enemy nexus

OUR MISSION

**WE ASPIRE**  
TO BE THE MOST  
**PLAYER**  
  
**FOCUSED**  
GAME COMPANY IN THE  
**WORLD**

A large, semi-transparent black and white photograph occupies the left side of the slide. It shows the silhouette of a person from behind, looking out over a city skyline at sunset or sunrise. The sky is filled with warm, orange and yellow hues. In the foreground, the dark shape of the person's head and shoulders is visible. The city buildings are blurred in the background.

AGENDA

DATA@RIOT

WHY SPARK?

USE CASES

FUTURE

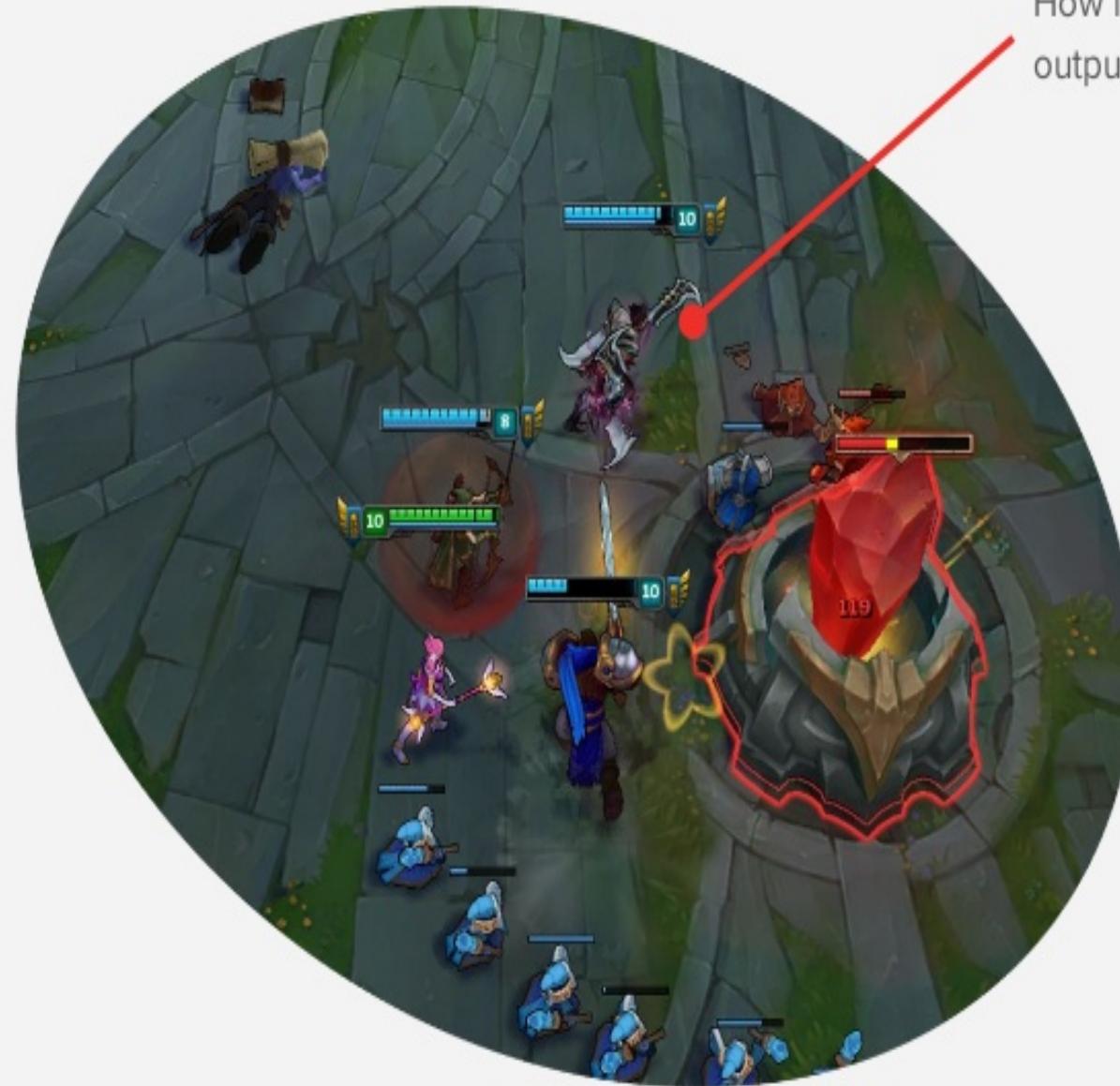
# Data

What can data tell us?

How do we interact with these data?

## Game Balance

How is Draven's damage output since the patch?



## Network Performance

Player noted lag during gameplay. Is his ISP having trouble?





## Customization

Are there customizations  
similar to Star Guardian Lux  
this player may enjoy?

# Players & Data

WORLDWIDE



**67+ million**

monthly active players



**500+ billion**

data points per day



**26 petabytes**

data collected since beta

Statistics released Jan 2014

# THE DATA SCIENCE TOOLKIT



- Scheduled ETLs
- Ad Hoc Queries
- Data pulls



- Recent
- $10^6$  events/sec
- Monitoring
- Fraud/Anomaly



- **Desktops**
- R/Python
- Tableau

# THE DATA SCIENCE TOOLKIT



- Scheduled ETLs
- Ad Hoc Queries
- Data pulls

- Recent
- $10^6$  events/sec
- Monitoring
- Fraud/Anomaly

- **Desktops**
- R/Python
- Tableau

Our data and ecosystem are scaling fine,  
but our analytic tools were *not*.

A large, semi-transparent black and white photograph occupies the left side of the slide. It shows the silhouette of a person from behind, looking out over a city skyline at sunset or sunrise. The sky is filled with warm, orange and yellow hues. In the foreground, the dark shape of the person's head and shoulders is visible. The city buildings are blurred in the background.

AGENDA

DATA@RIOT

WHY SPARK?

---

USE CASES

FUTURE

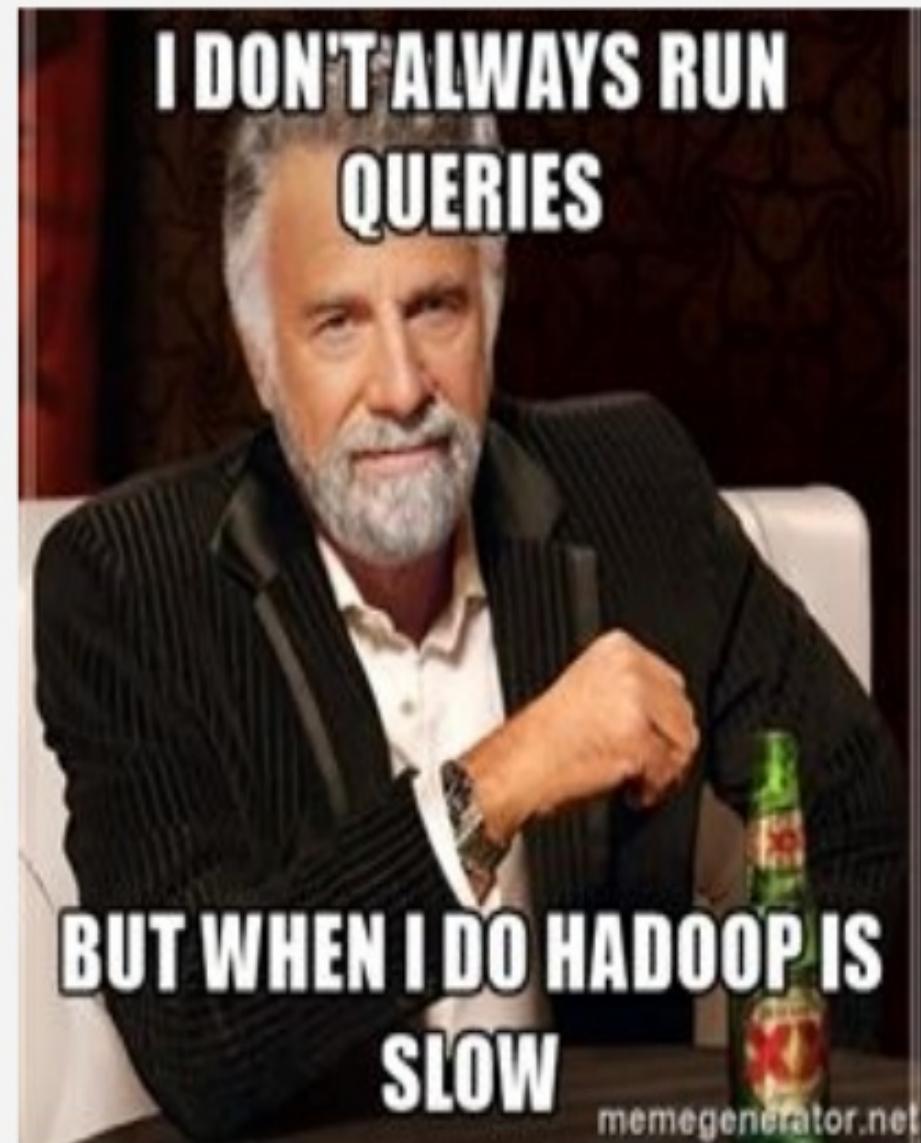
# Data Science At Scale



Why Spark?



Spark SQL



memegenerator.net



Spark SQL

Spark Streaming

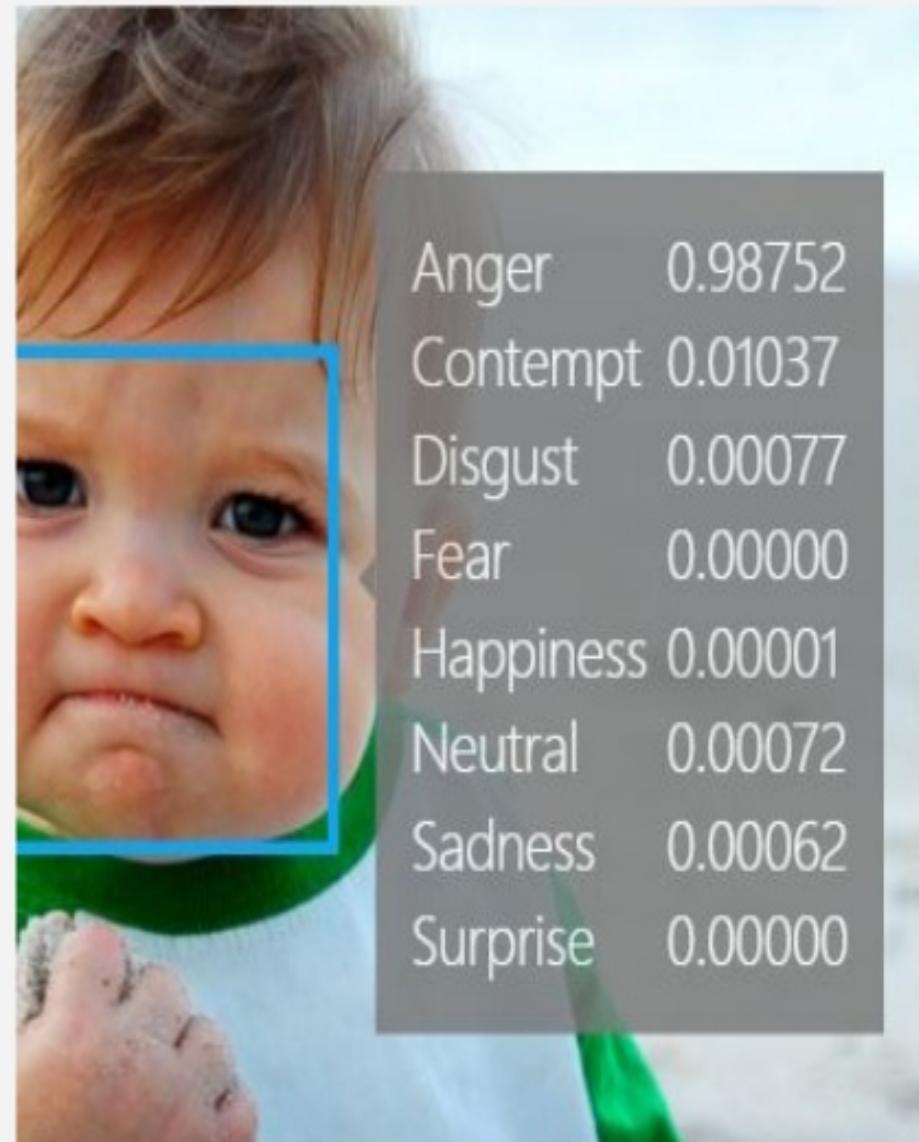




Spark SQL

Spark Streaming

**Spark ML**





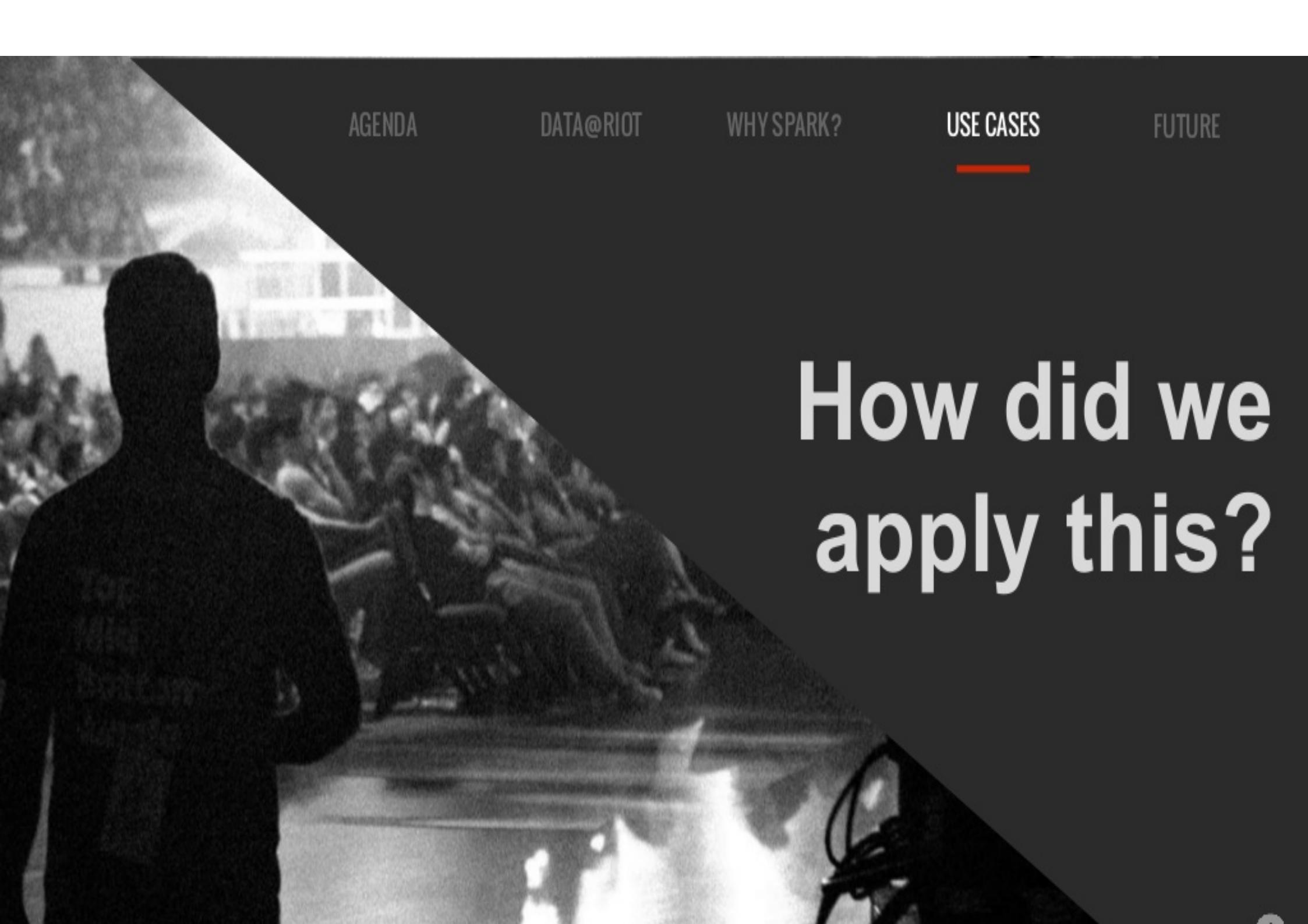
Spark SQL

Spark Streaming

Spark ML

Empowerment





AGENDA

DATA@RIOT

WHY SPARK?

USE CASES

---

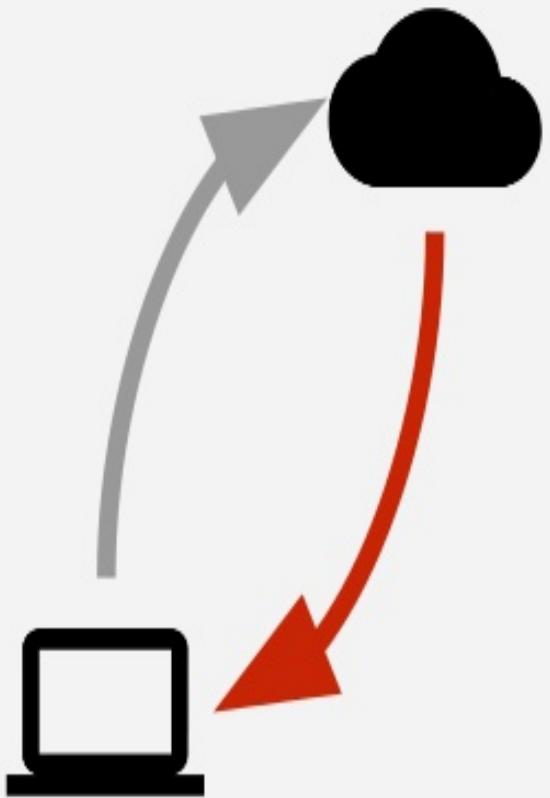
FUTURE

# How did we apply this?

# How to Play (with data)

---

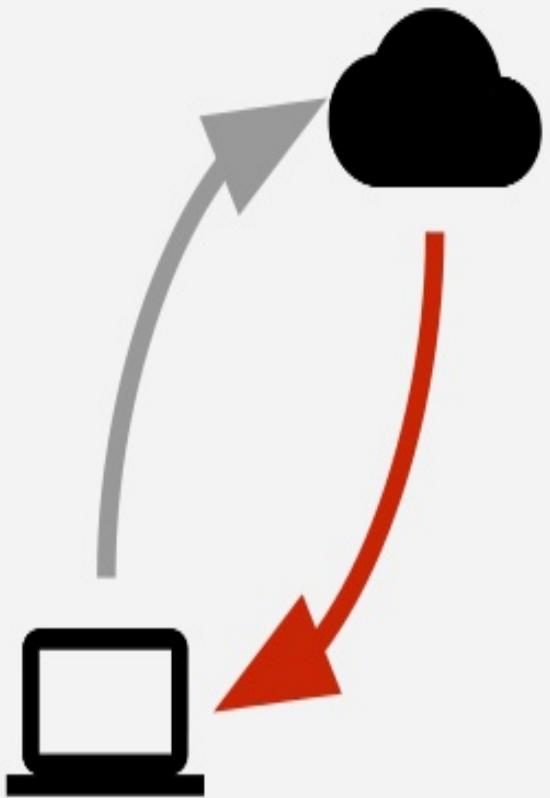
SparkSQL



# Pain Points

FOR DATA EXPLORATION + REPORTING

Iterative query development



# Pain Points

FOR DATA EXPLORATION + REPORTING

Iterative query development

Data pulls are slow

**Workflow is inefficient.**  
**Data is separated from tools.**

report.nvrs

Attached: sql crash and burn test · View Code · File · Permissions · Run All · Schedule · Comments · Revision history

## Network performance from off-shard west coast players

This notebook explores the game experience for west coast players playing on EU game servers.

**Tables used:**

- derived.network\_data which is derived from raw.game\_network\_data via this [ETL job](#)

**SQL**

```
select
    region_code
    , avg(latency) as avg_latency
from derived.network_data
where dt>='2016-03-01' and dt<'2016-04-01'
and region_code in ('CA','WA','OR') and shard = 'EUW1'
group by region_code
```

(0) Spark Jobs

A bar chart titled 'avg\_latency' on the y-axis (ranging from 0 to 500) and 'region\_code' on the x-axis (with categories OR, CA, and WA). The bars are blue. The approximate values are: OR ~350, CA ~200, and WA ~450.

region_code	avg_latency
OR	~350
CA	~200
WA	~450

Plot Options... ▾

Command took 85.34s

# Data, easier.



Integrated



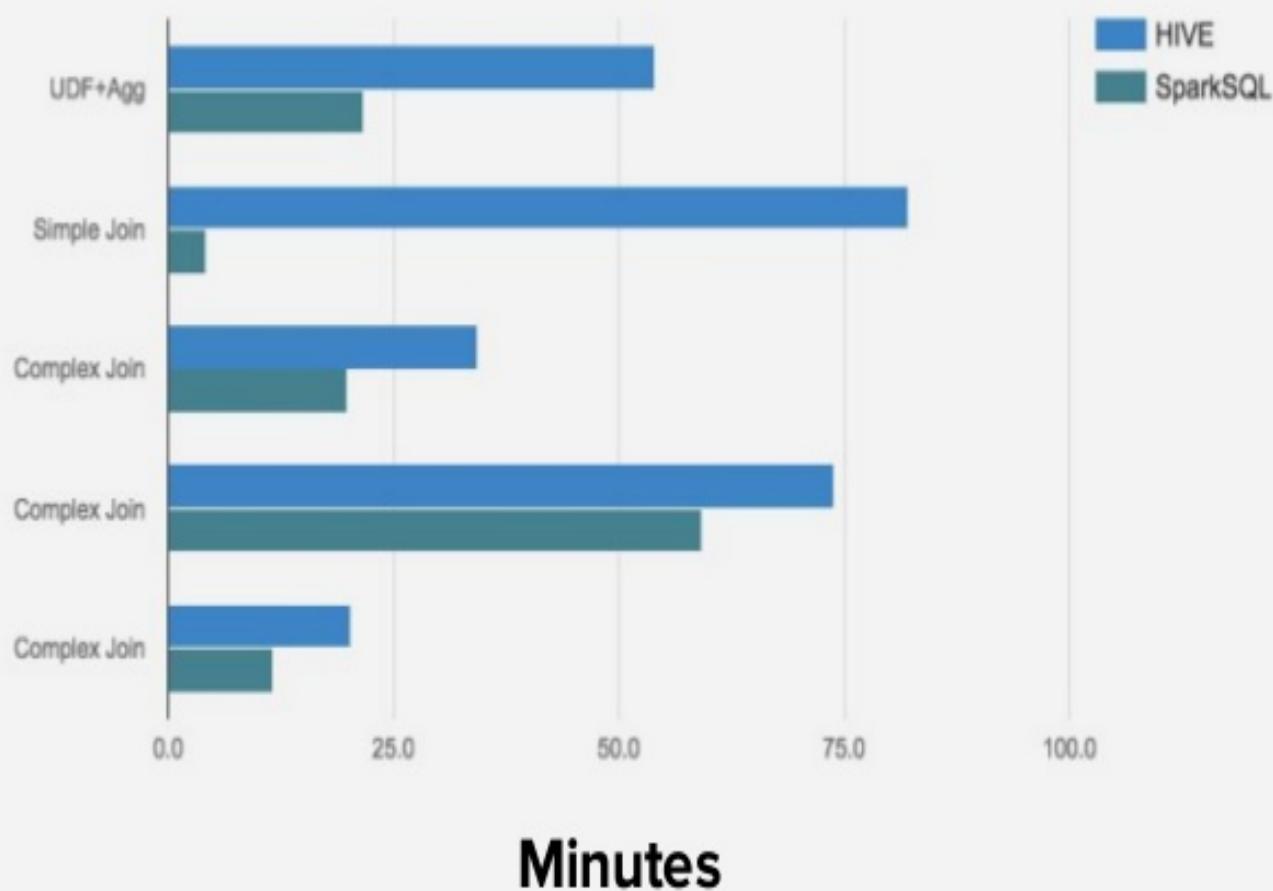
Interactive



Faster

# Head-to-head tests between **EMR** and **SparkSQL** .

(equal cost clusters)





Managed/SparkSQL

## TAKEAWAY

Data exploration efficiency increased

Performant SQL queries

Self managed cluster deploys

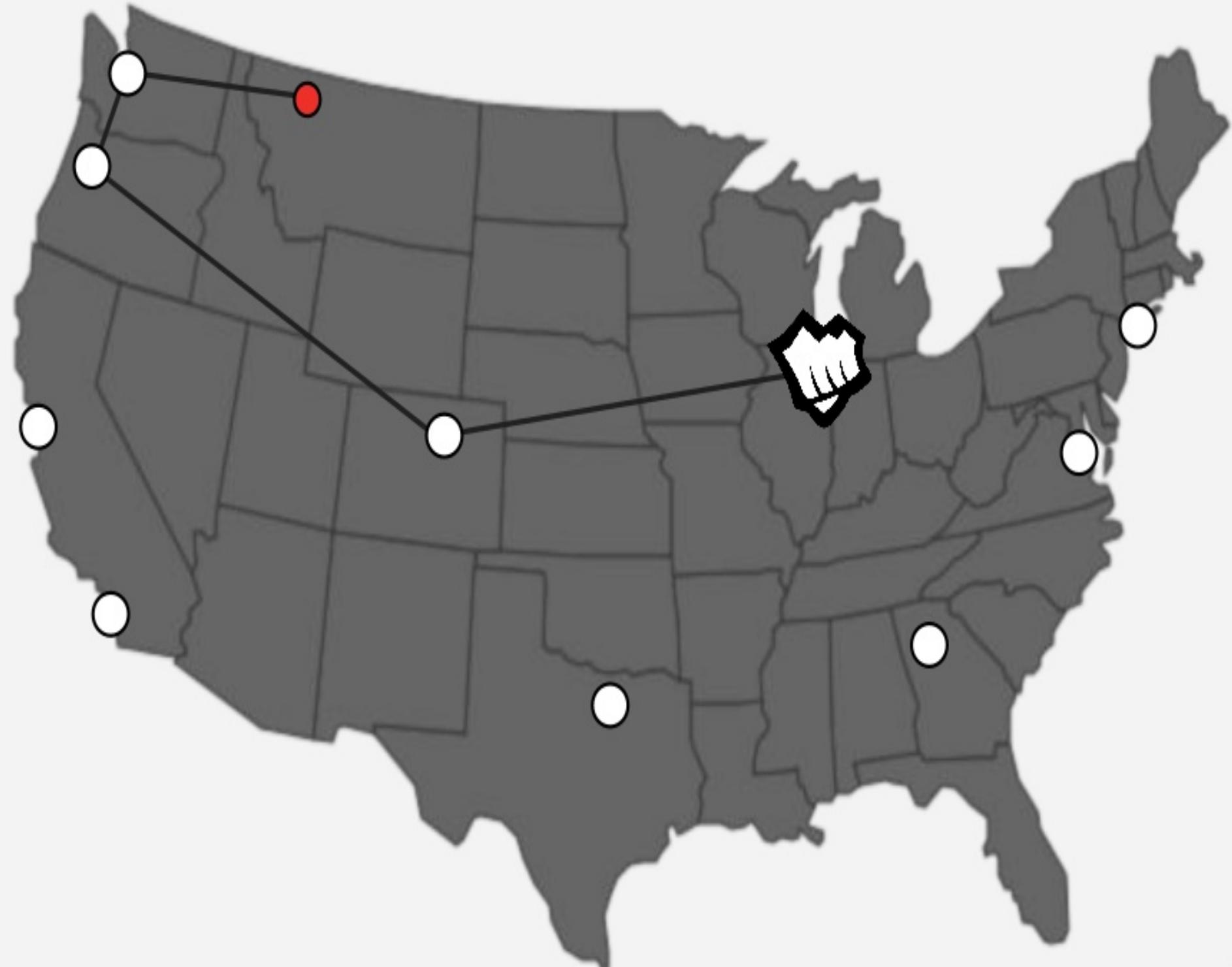


# Winning the war on lag

---

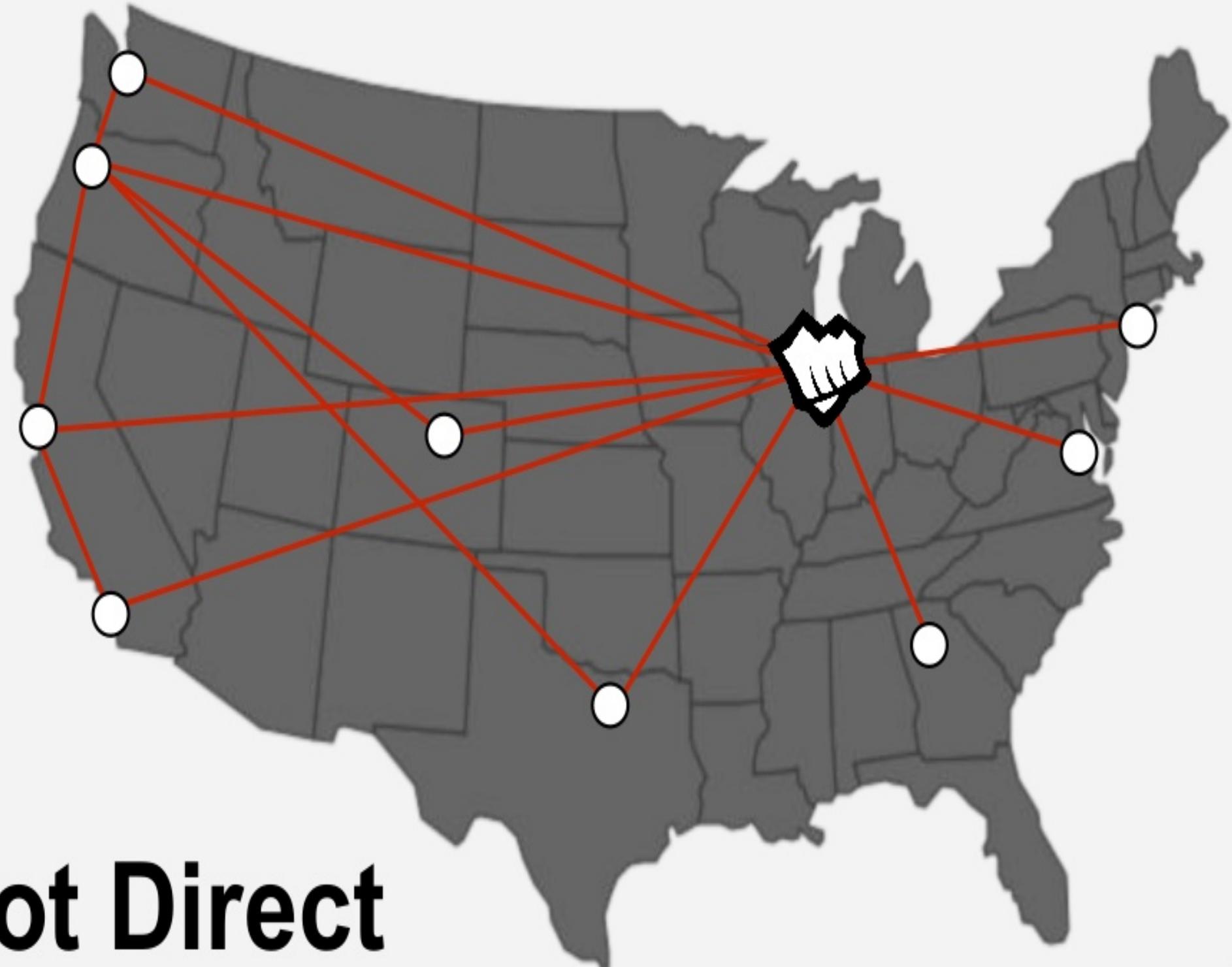
With Spark Streaming

**Gameplay highly dependent on  
network connection.**

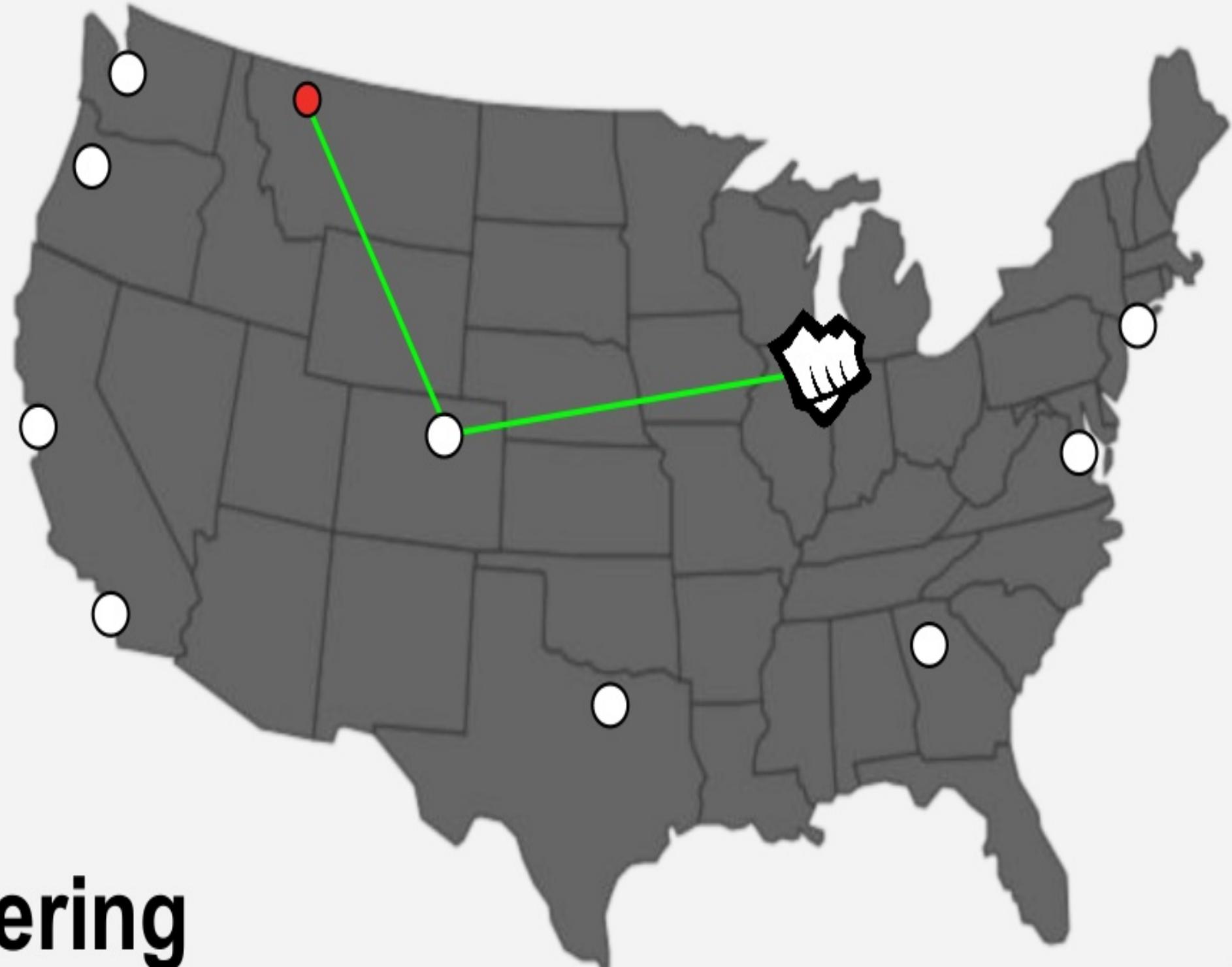


**The existing network limits our players.**

**So we built our own!**



# Riot Direct



# Peering

We have this awesome tool to fix problems.

But we need to **find them!**

# Scope of Data

WORLDWIDE



**17,000+**

Unique ISPs



**171,000+**

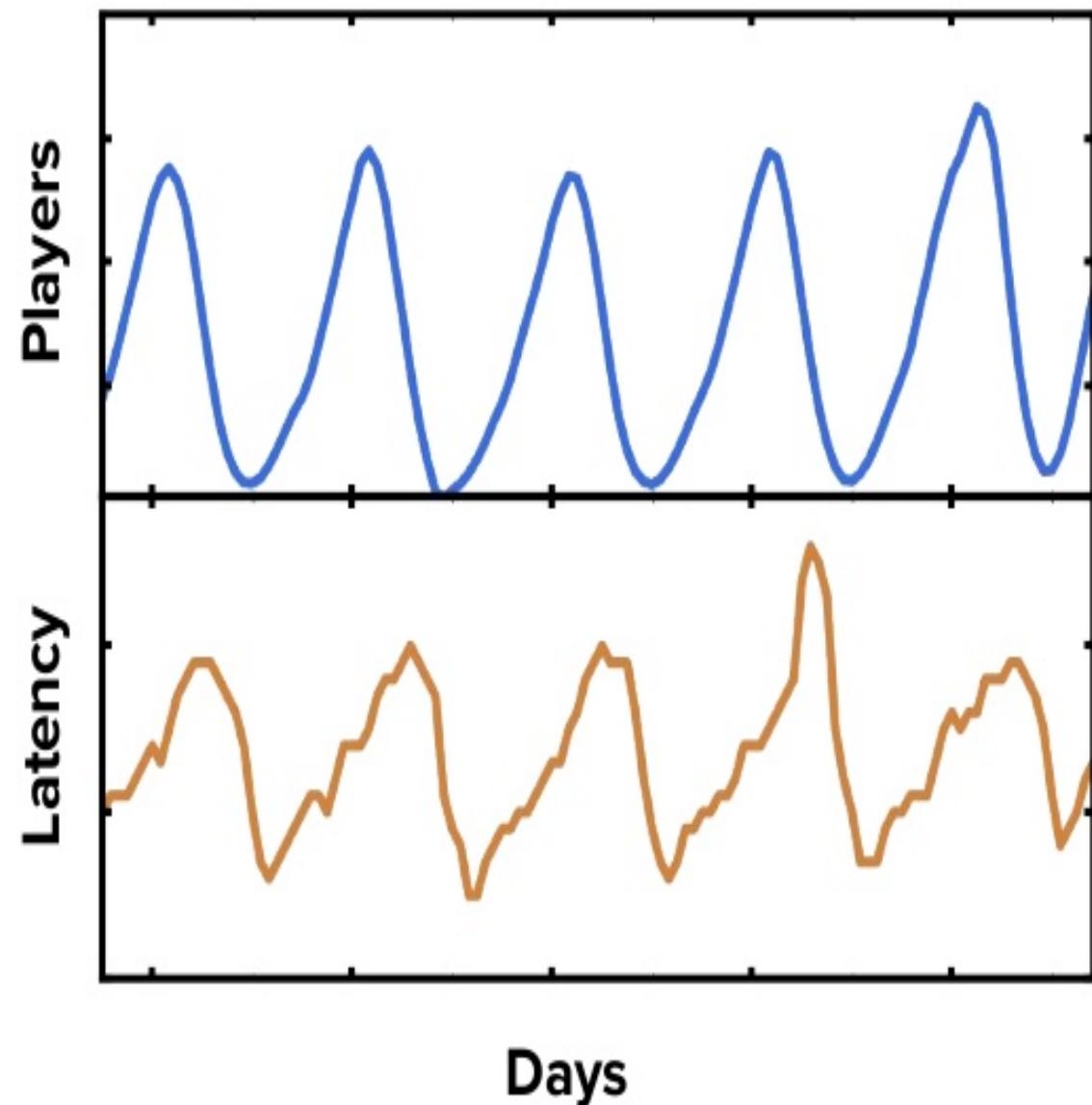
City/ISP combinations



**250,000+**

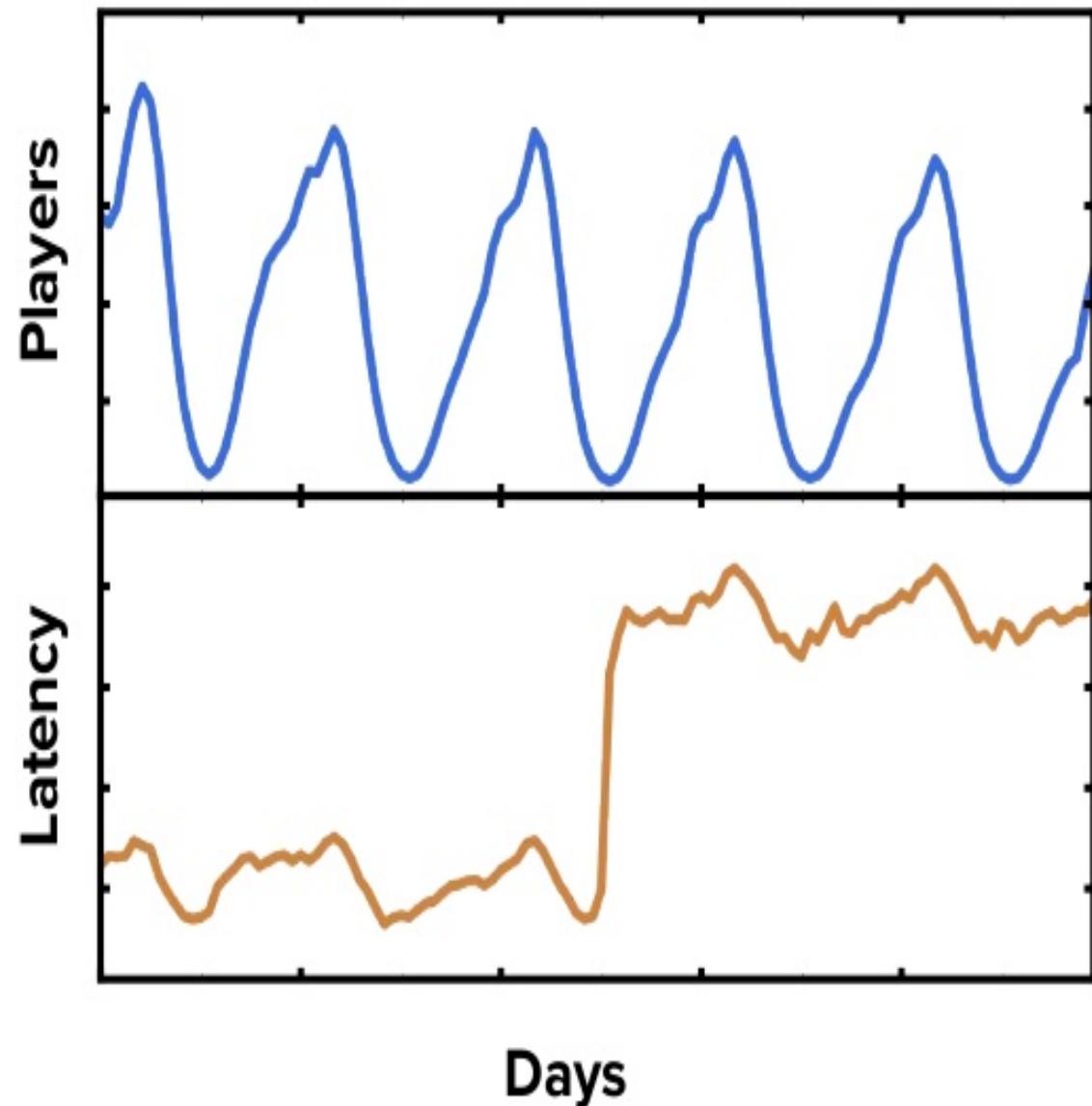
Network stat messages per second

# Model this



**Model this**

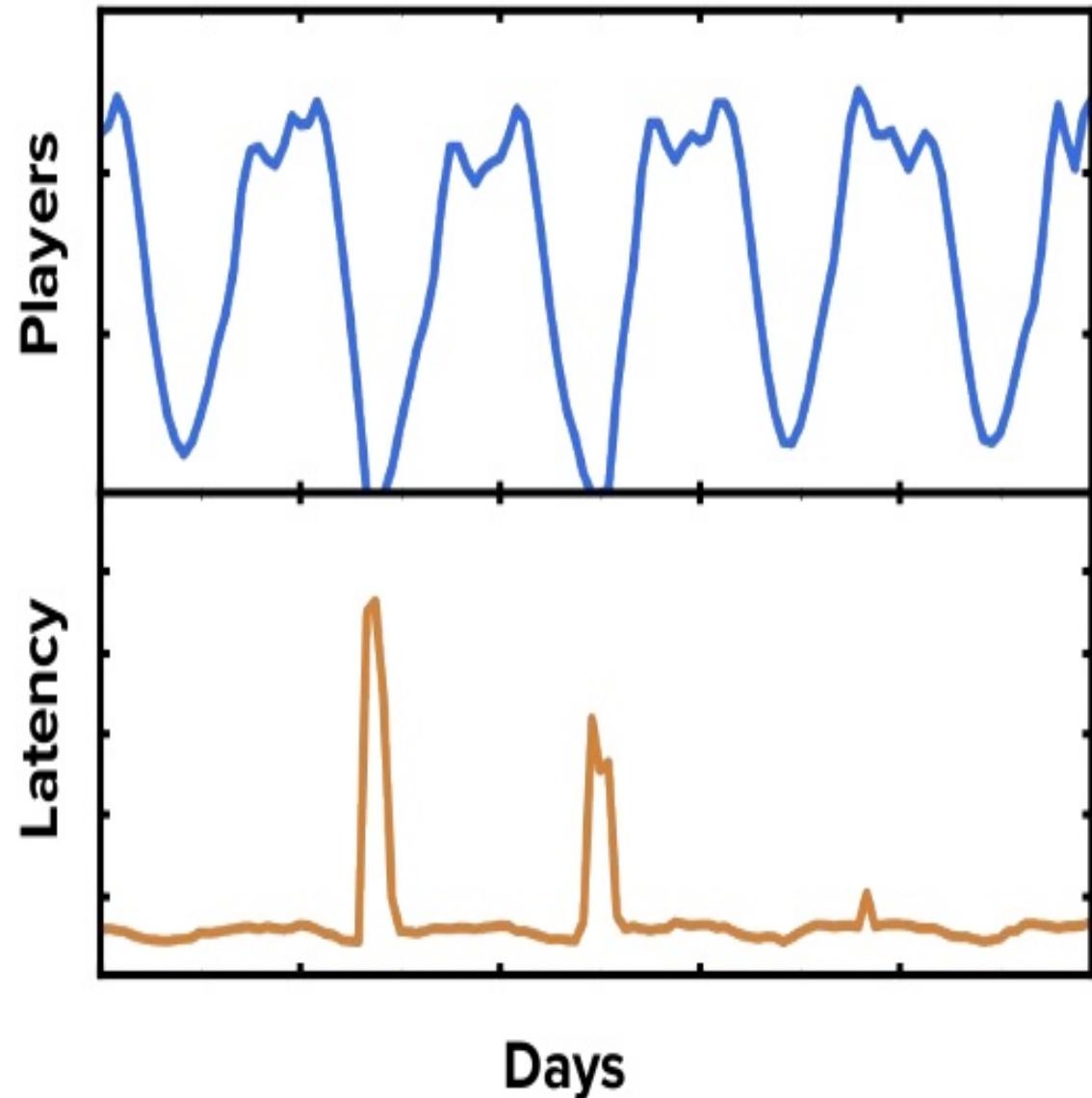
**...so we can  
detect *this***

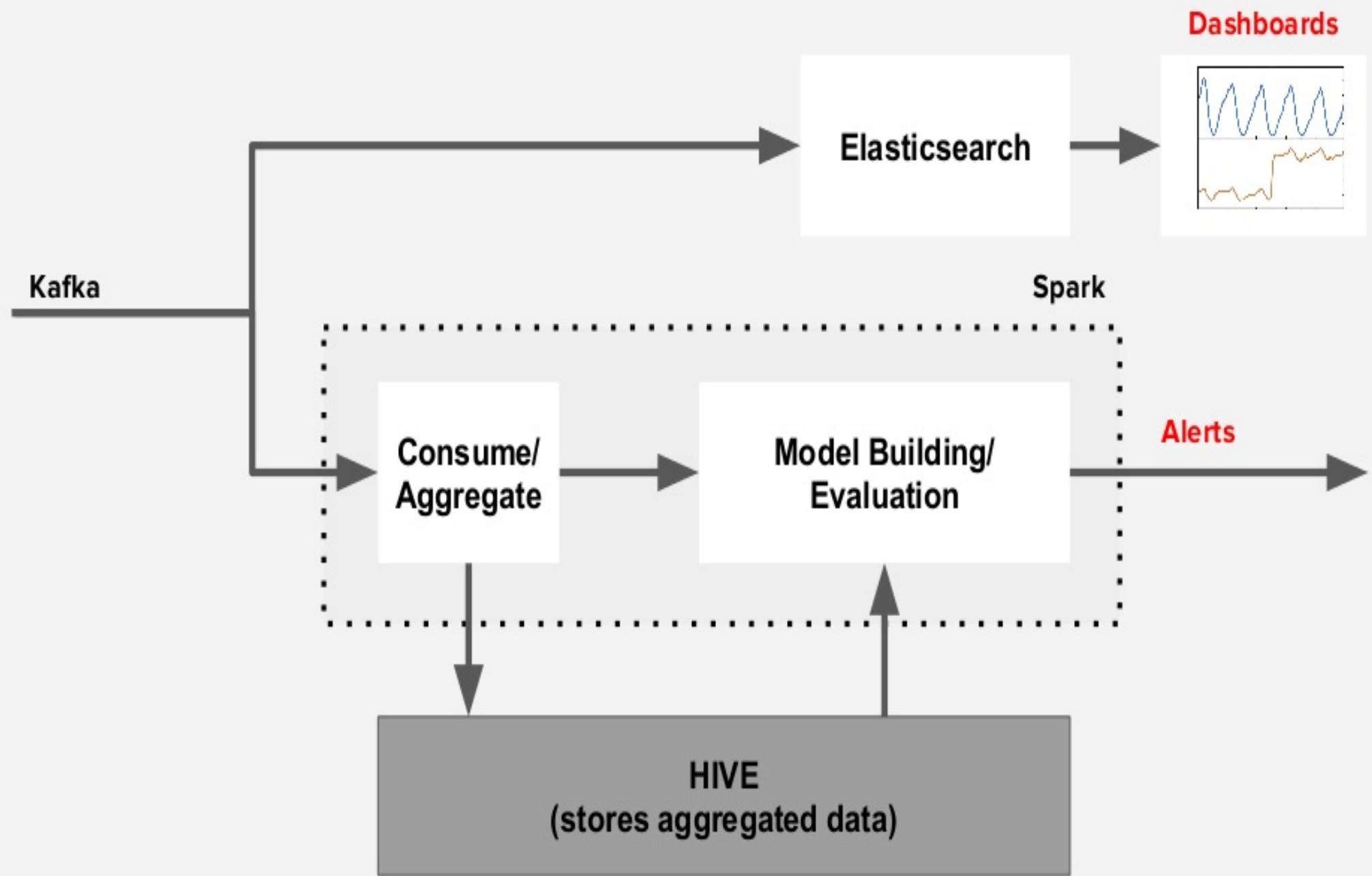


**Model this**

**...so we can  
detect *this***

**...or *this*.**







**Spark Streaming**

# TAKEAWAY

Works

Intuitive

Will be a lot easier in 2.0!

# Secret Store

---

The Recommendation System



# LEAGUE OF LEGENDS





# LEAGUE of LEGENDS

## CATEGORIES

- Skins
- Chroma Packs

 Show Owned

Sort By: Release Date ▾

## OWNERSHIP

 Champion Owned

## TYPES

 Limited Availability Legendary Ultimate

## SALE STATUS

 On Sale

Void Bringer Illaoi

1350



Dragon Trainer Tristana

OWNED



Spirit Fire Brand

1350



Demon Vi

1350



Cosmic Reaver Kassadin

1350



Shadowfire Kindred

1350



Ironside Malphite

975



Marauder Alistar

750



Marauder Olaf

750



Warden Jax

750



Warden Karma

750



PROJECT: Flora

1350



PROJECT: Leona

1350



PROJECT: Lucian

1350



PROJECT: Zed

1350

## CATEGORIES

 Skins Chroma Packs Show Owned

Sort By: Release Date ▾

## OWNERSHIP

 Champion Owned

## TYPES

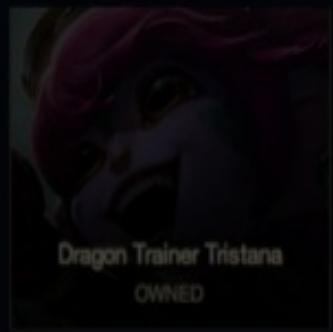
 Limited Availability Legendary Ultimate

## SALE STATUS

 On Sale

Void Bringer Ilaal

1350



Dragon Trainer Tristana

OWNED



Spirit Fire Brand

1350



Demon Vi

1350



Cosmic Reaver Kassadin

1350



Shadowfire Kindred

1350



Ironside Malphite

975



Marauder Alistar

750



Marauder Olaf

750



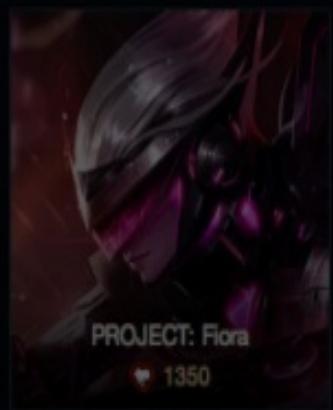
Warden Jax

750



Warden Karma

750



PROJECT: Flora

1350



PROJECT: Leona

1350



PROJECT: Lucian

1350



PROJECT: Zed

1350

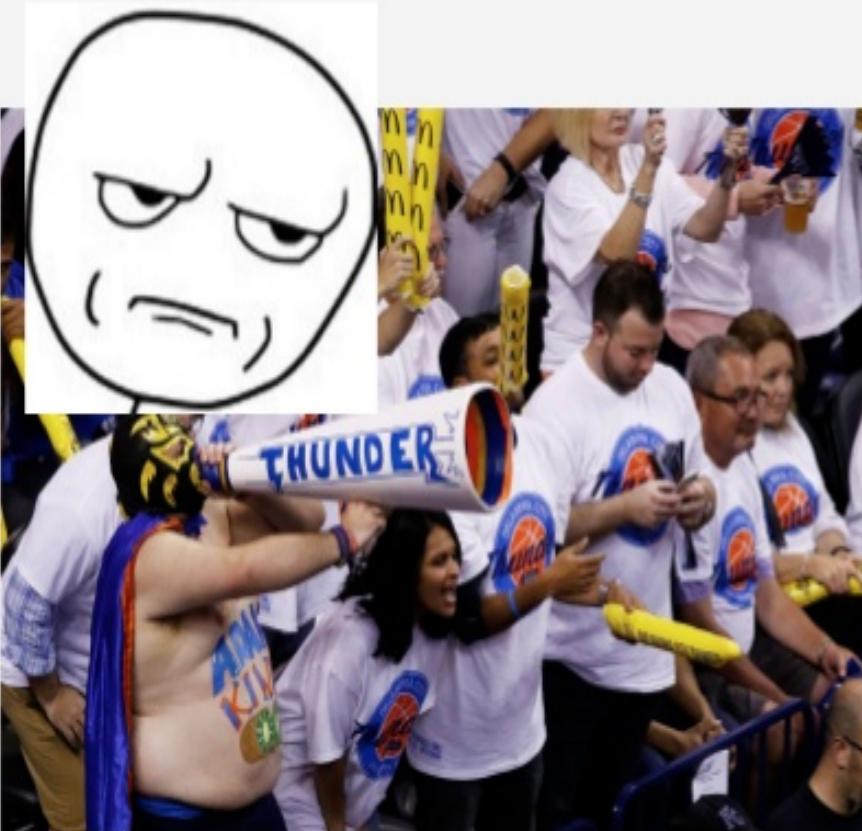
# Why personalized recommendations?

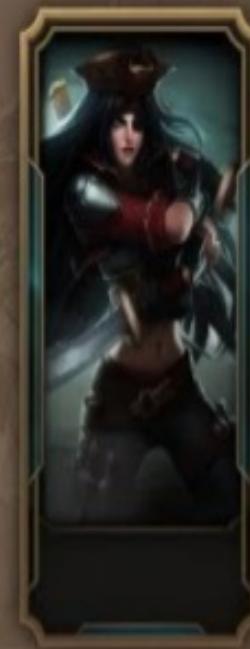


# Why personalized recommendations?



# Why personalized recommendations?



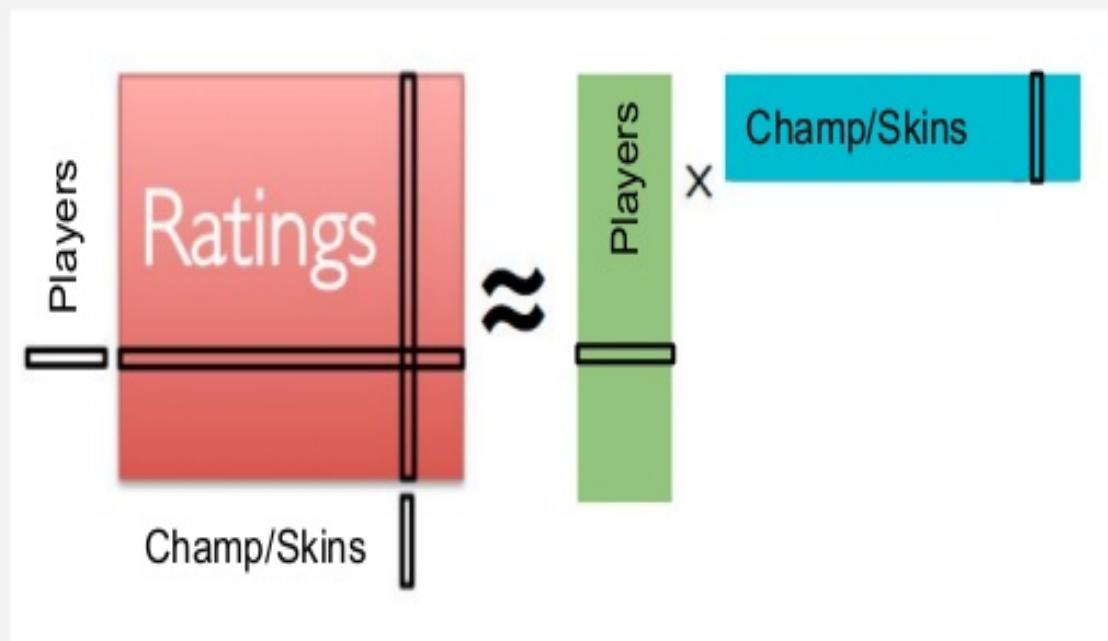




**67+ million**  
monthly active players

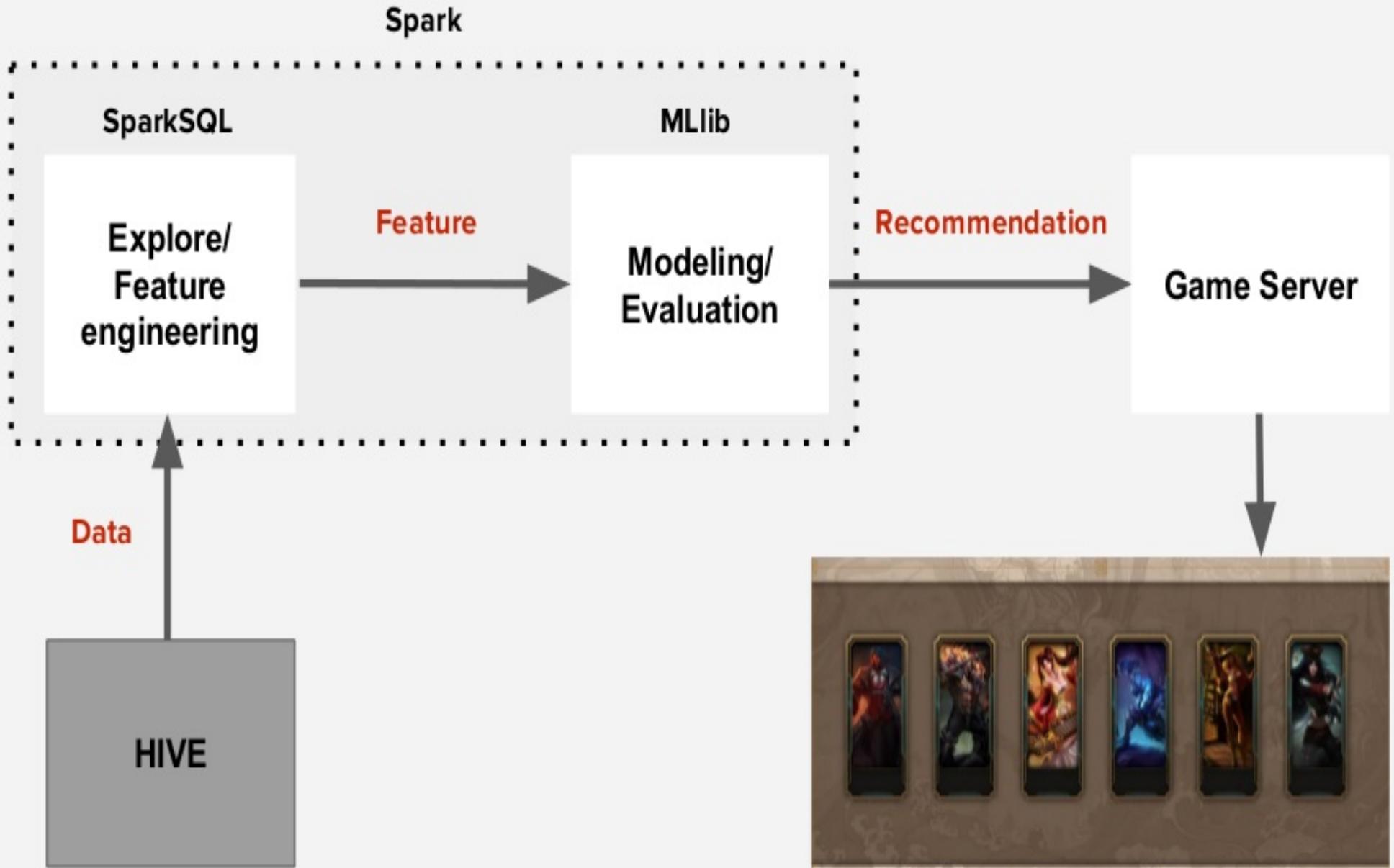


**500+ billion**  
data points per day



**~1000**  
champion/skin combo







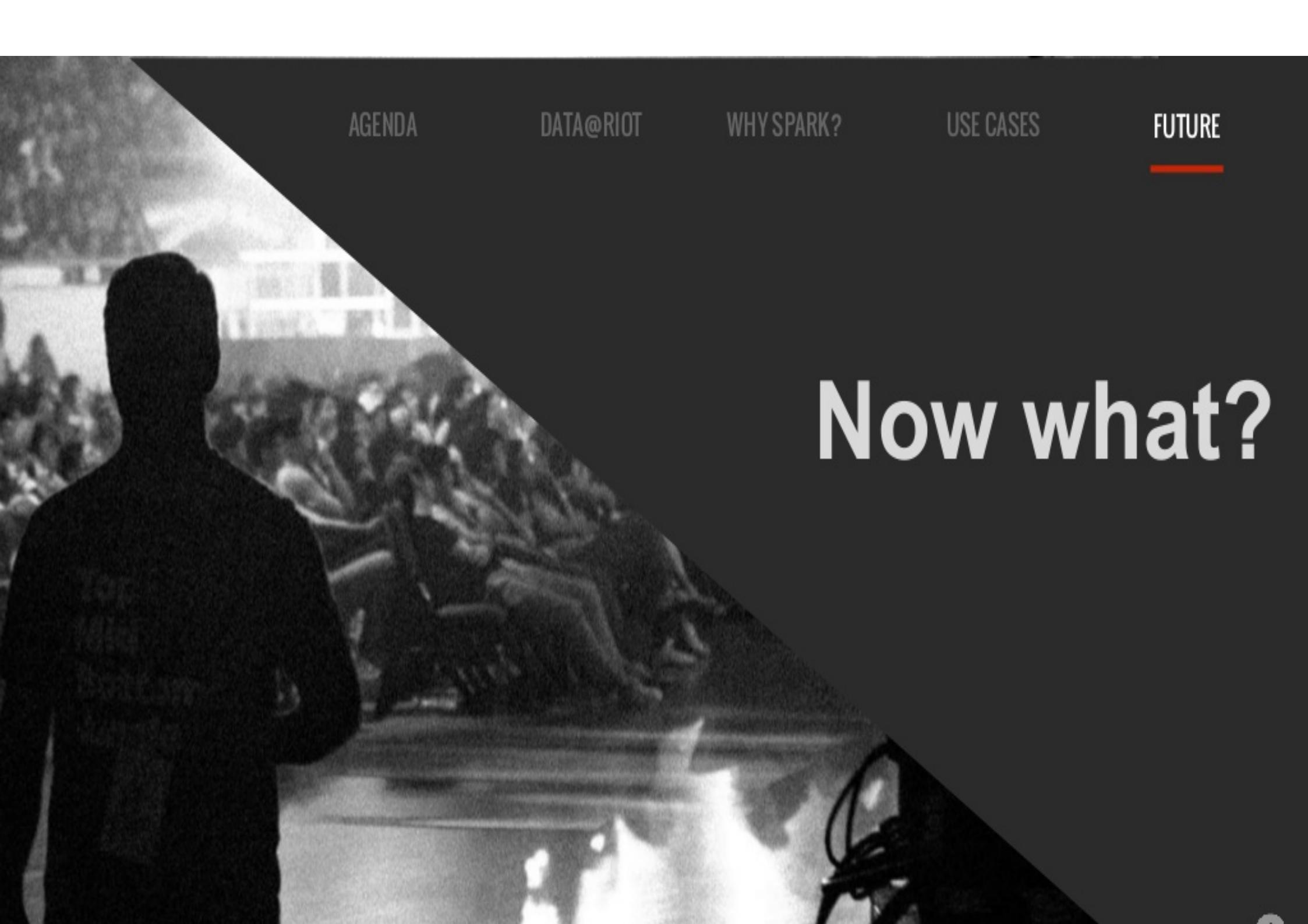
**Recommendation System**

# TAKEAWAY

Fast prototyping

Works on big matrix!

Easy automation

A large, semi-transparent black and white photograph occupies the left side of the slide. It shows the silhouette of a person from behind, looking out over a city skyline at sunset or sunrise. The sky is filled with warm, orange and yellow hues. In the foreground, the dark shape of the person's head and shoulders is visible. The city buildings are blurred in the background.

AGENDA

DATA@RIOT

WHY SPARK?

USE CASES

FUTURE

---

# Now what?

**With Spark, we think that we're  
just getting started.**

# Wanna learn more?

(we're hiring!)



**Colin Borys**

**CBORYS @ RIOTGAMES.COM**

**Xiaoyang Yang**

**XYANG @ RIOTGAMES.COM**