

西遊記

Spark Summit 2016 @San Francisco

Stana He

工商時間

- 請加入gitter
<https://gitter.im/hubertfc/SparkHsinchu>
- Gitter app : <https://gitter.im/apps>
- 與meetup
<https://www.meetup.com/Apache-Spark-Hsinchu/>

Who am I ?

- Stana He
- Is-land 取經哥



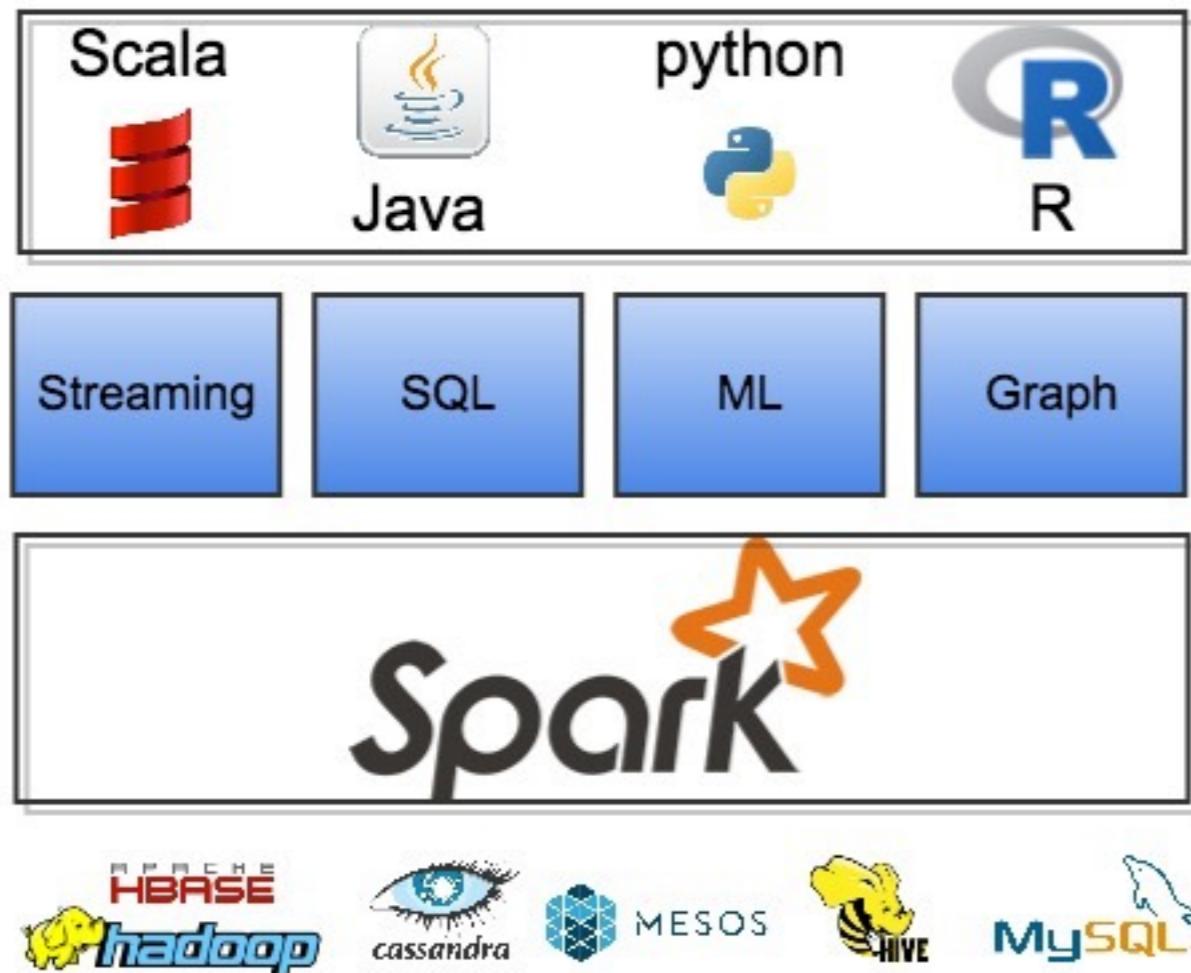
Agenda

- Something about Apache Spark
- Enterprise use case

What is Apache Spark?

- Open source cluster computing framework.
- Developed at the UC, Berkeley's AMPLab.
- Donated to the Apache Software Foundation.

Benefits of Apache Spark



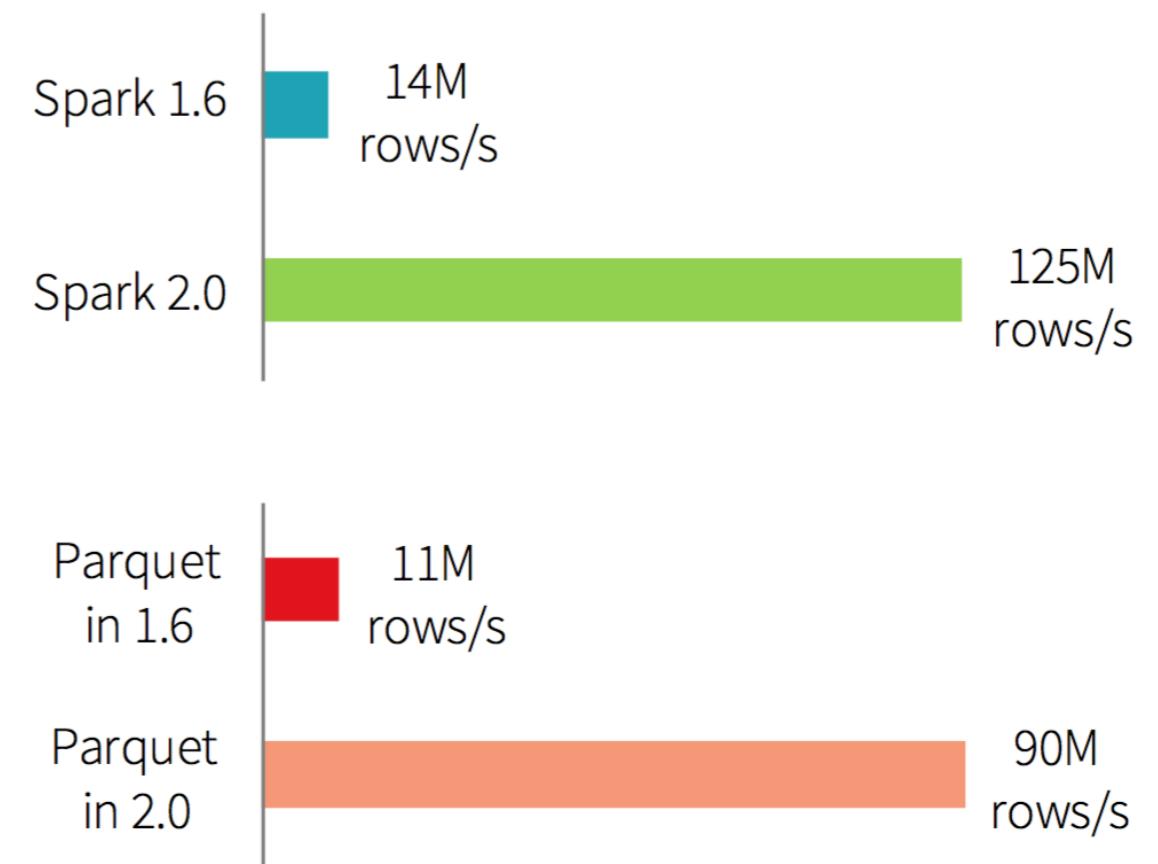
- Speed
 - 100x faster than Hadoop for large scale data processing.
- Ease of Use
 - Easy-to-use APIs.
- Unified Engine
 - Packaged with higher-level libraries, including streaming data, SQL queries, machine learning and graph processing.

What's New in 2.0 ?

- Structured API improvements
 - SQL, DataFrames, Datasets
- Structured Streaming
- MLlib model export
- MLlib R bindings
- SQL 2003 support
- Scala 2.12 support

What's New in 2.0 ?

- Whole-stage code generation
 - Fuse across multiple operators
- Optimized input / output
 - Apache Parquet + built-in cache



Enterprise use case



Winning the game with Spark!

Unfortunately, it doesn't!



reference:<http://www.slideshare.net/SparkSummit/video-games-at-scale-improving-the-gaming-experience-with-apache-spark>

Players and Data

- 67+ million monthly active players
- 500+ billion data points per day
- 26 petabytes data collected since beta

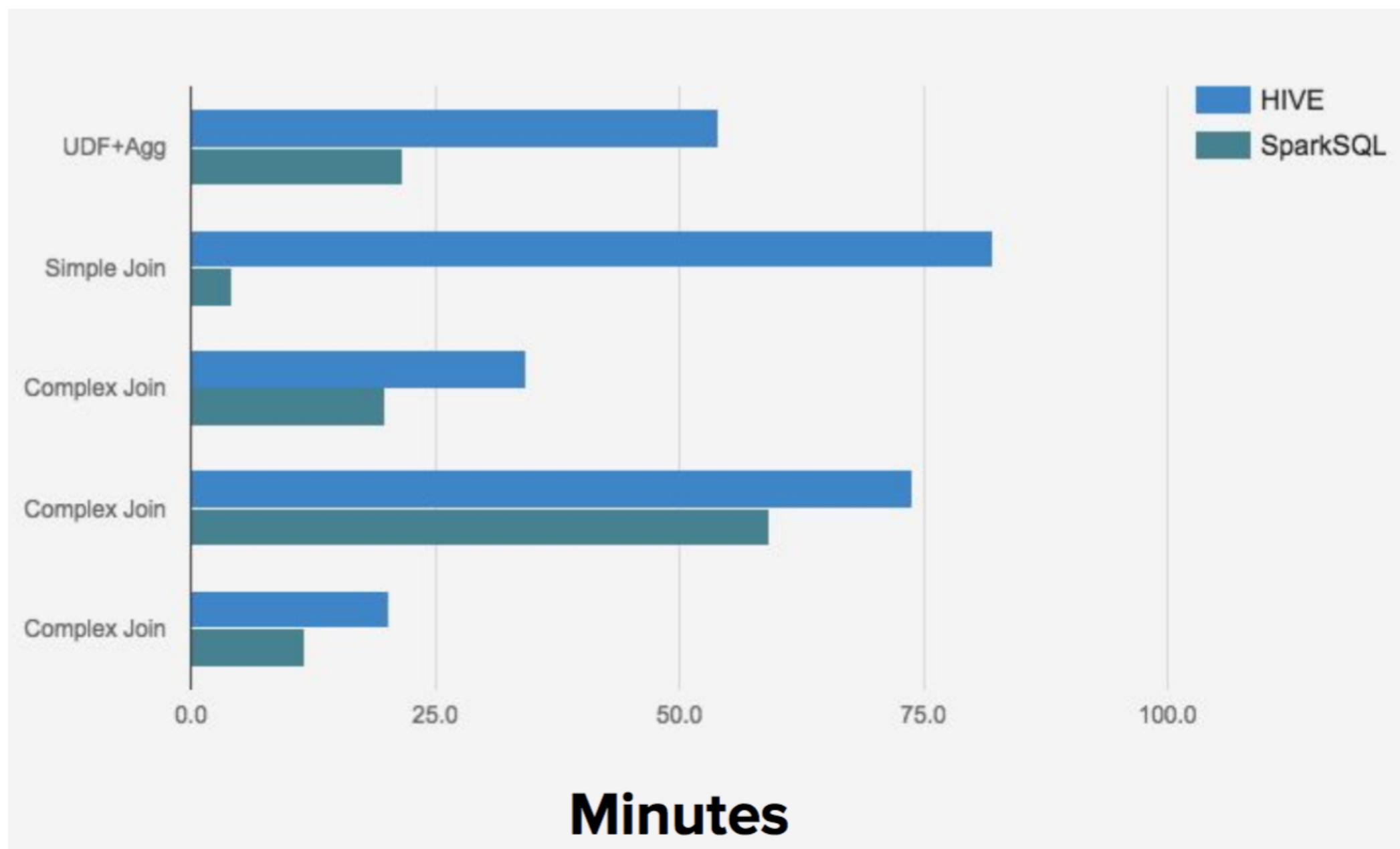
What does Spark do ?

- Spark SQL
Data exploration and reporting
- Spark Streaming
Network performance
- Spark MLlib
Recommendation system

Spark SQL

-Data exploration and reporting

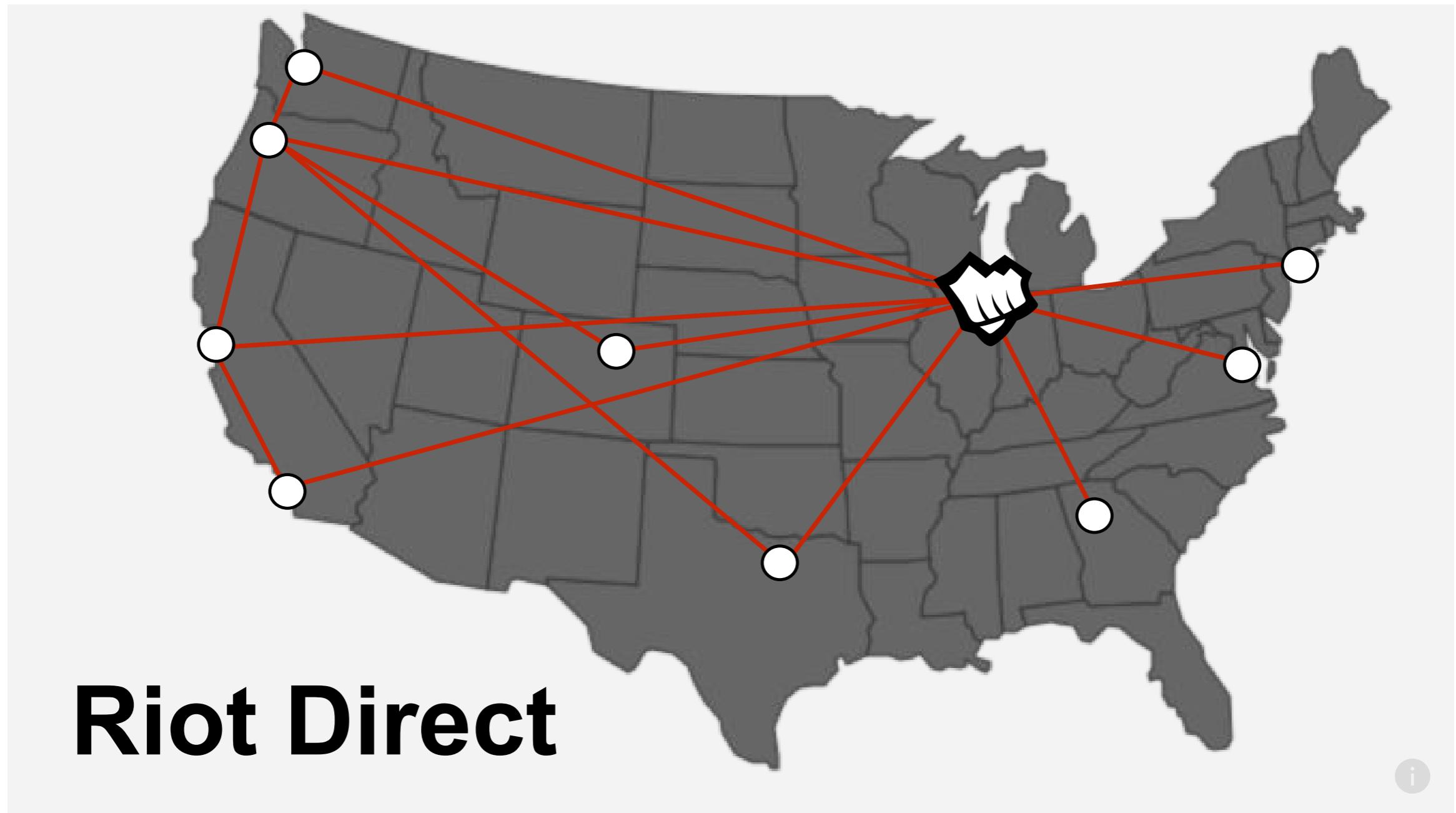
Performance



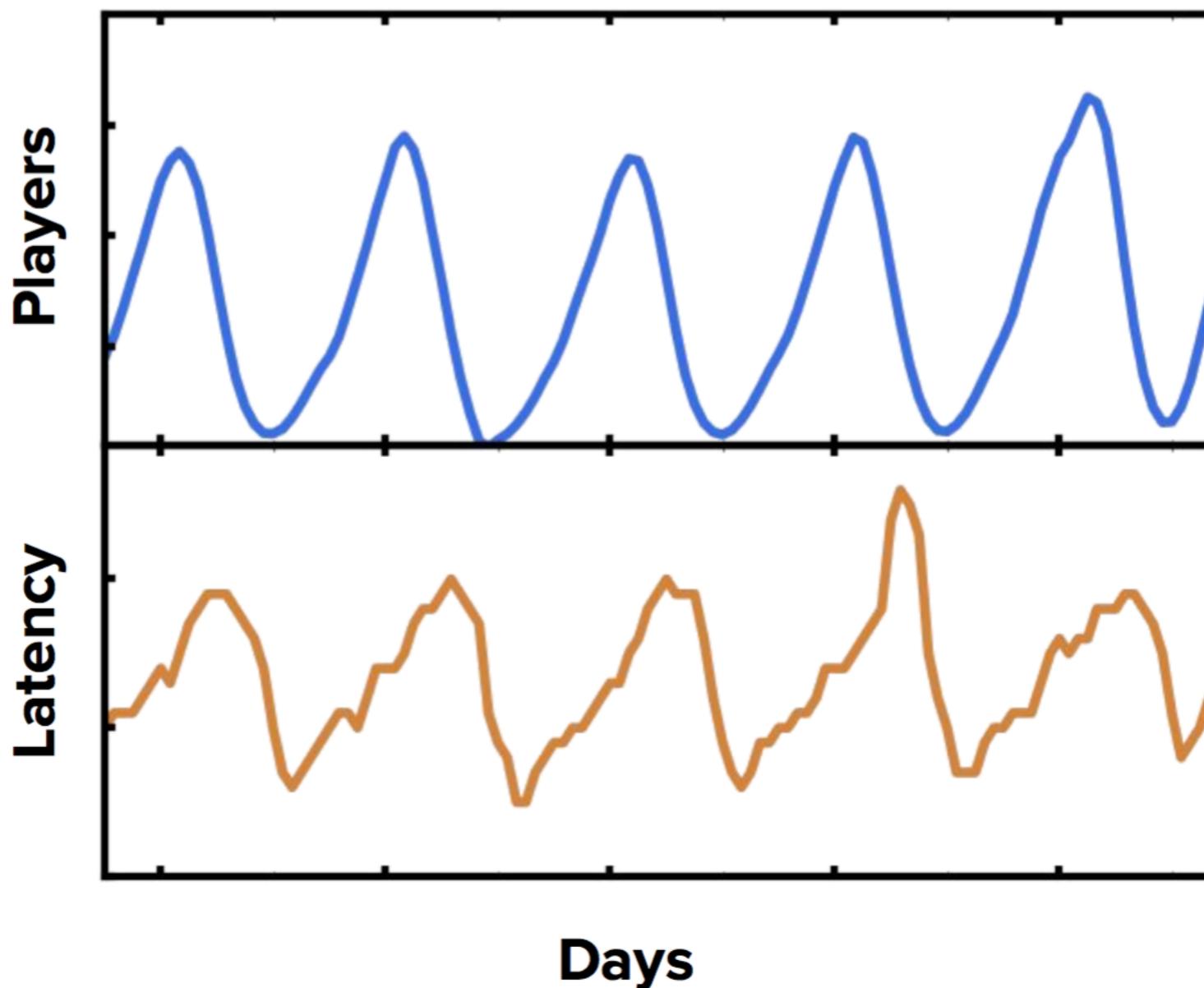
Spark Streaming

-Network performance

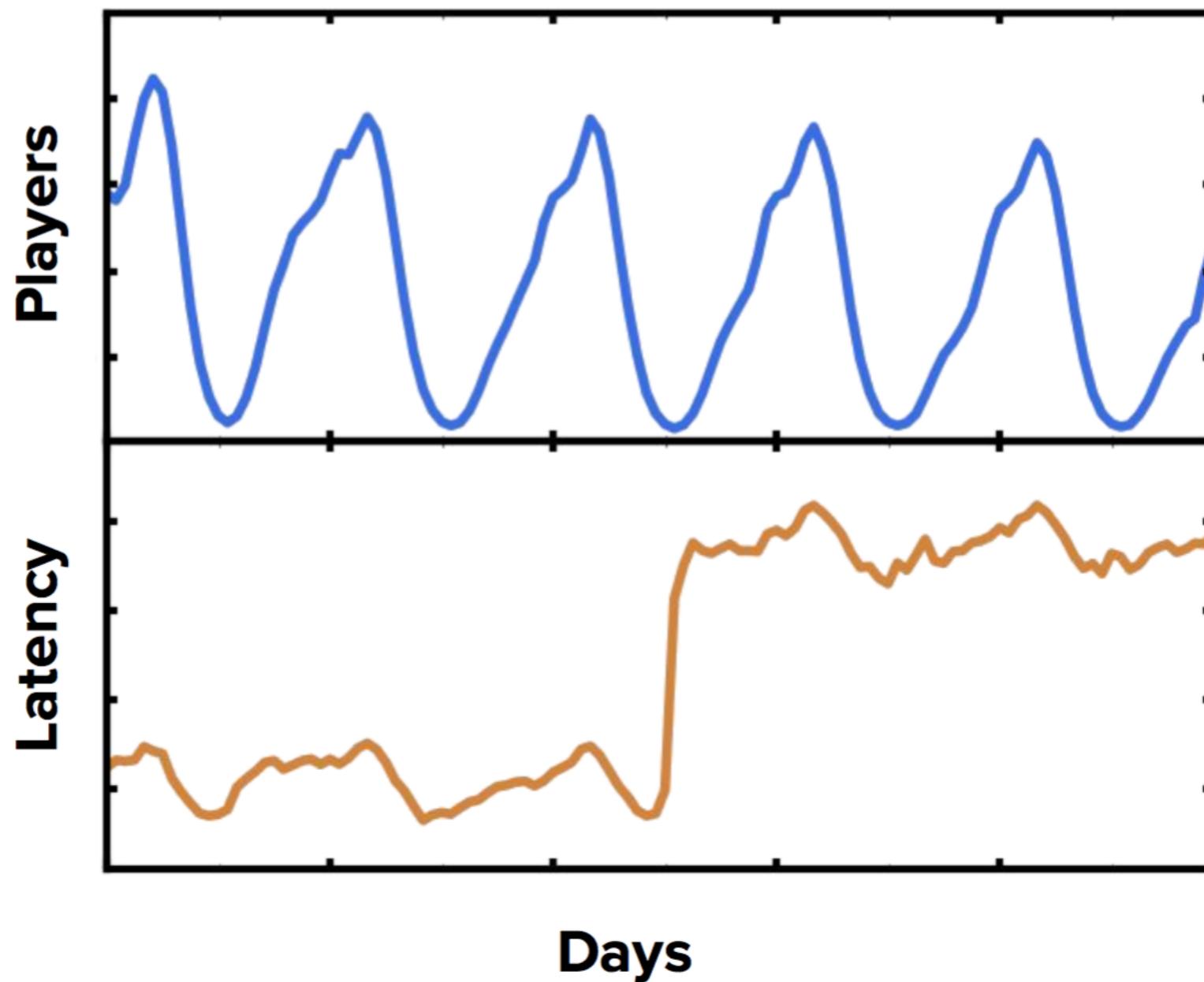
Build network



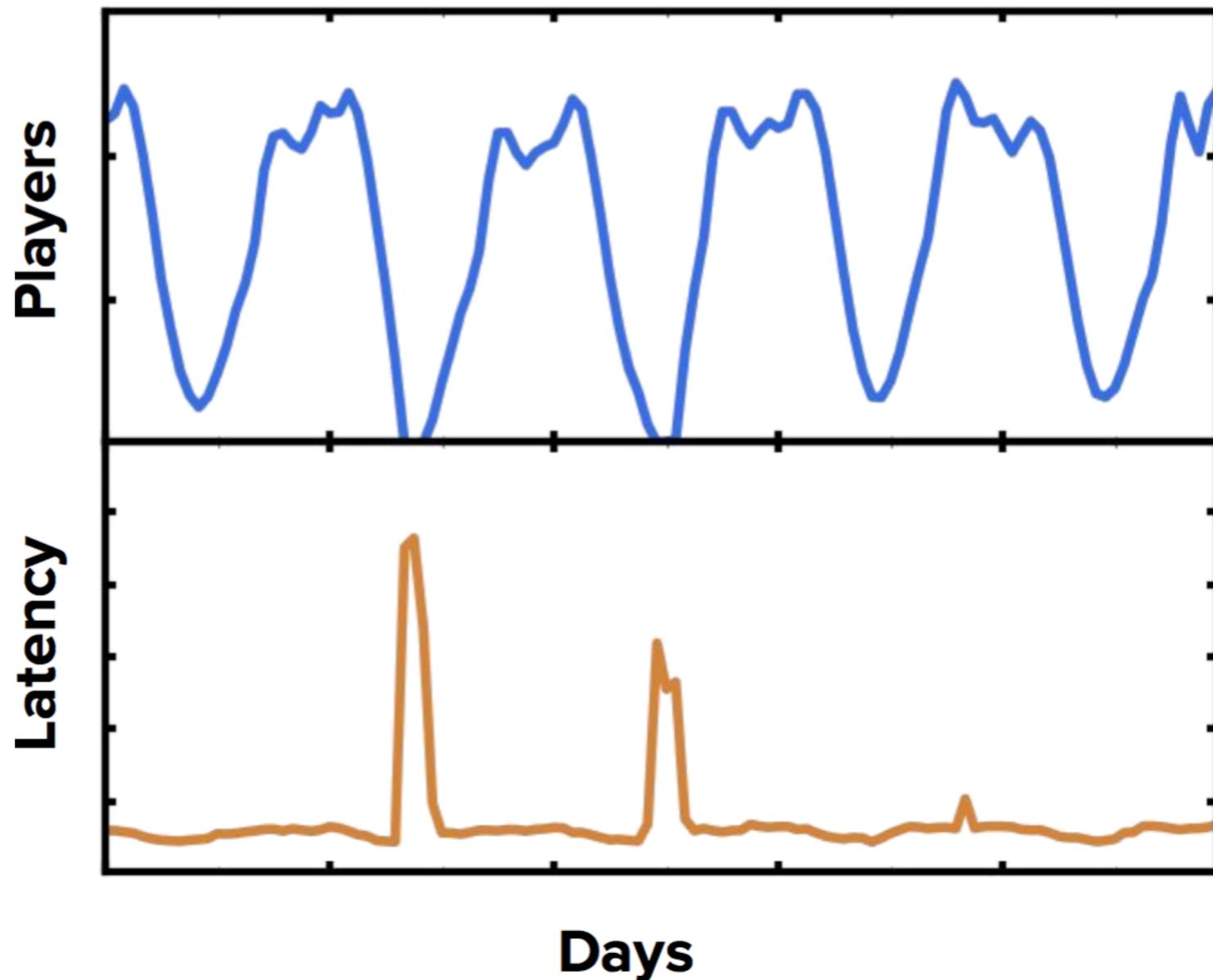
Normal Network Model

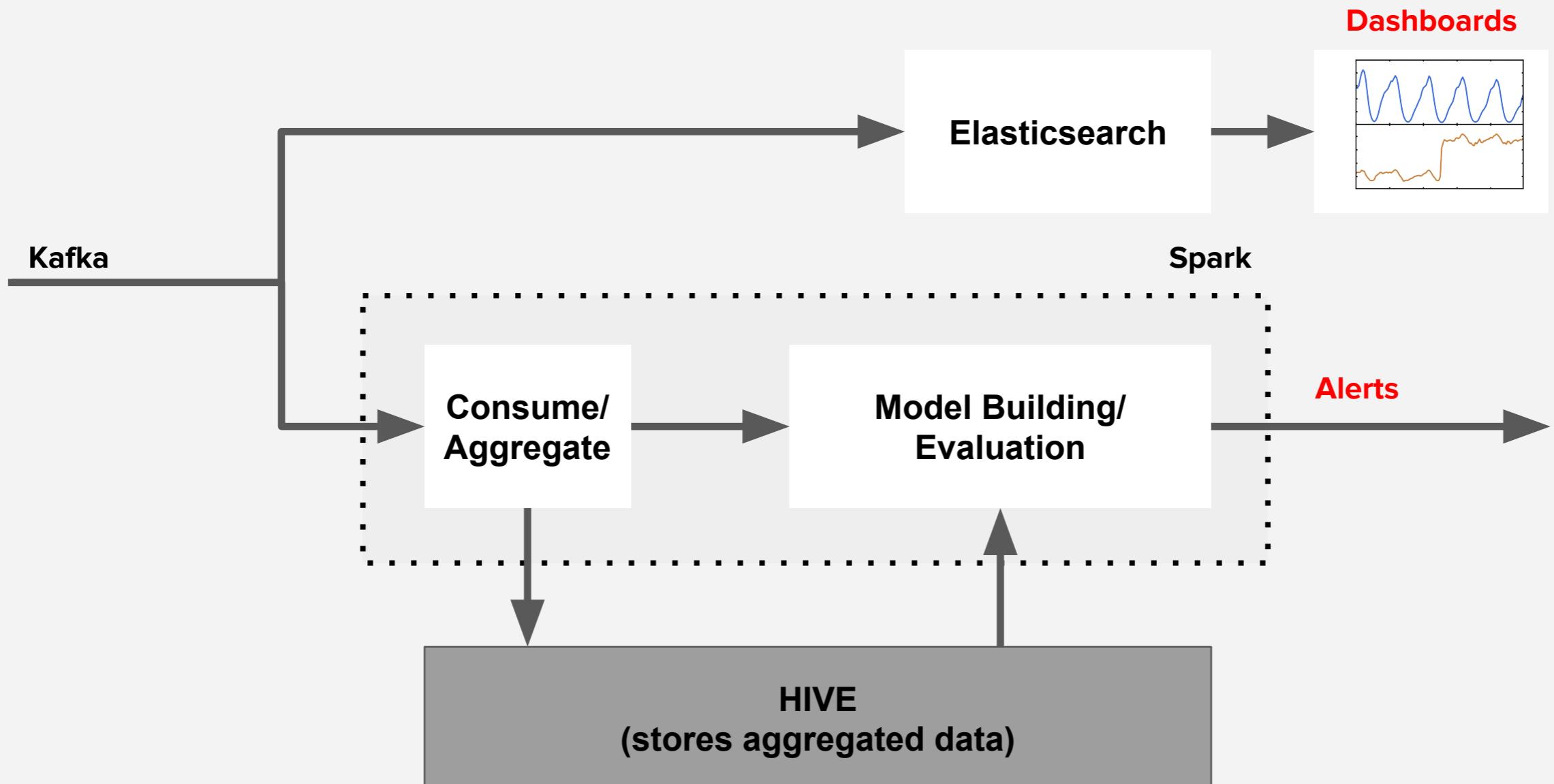


Detect model



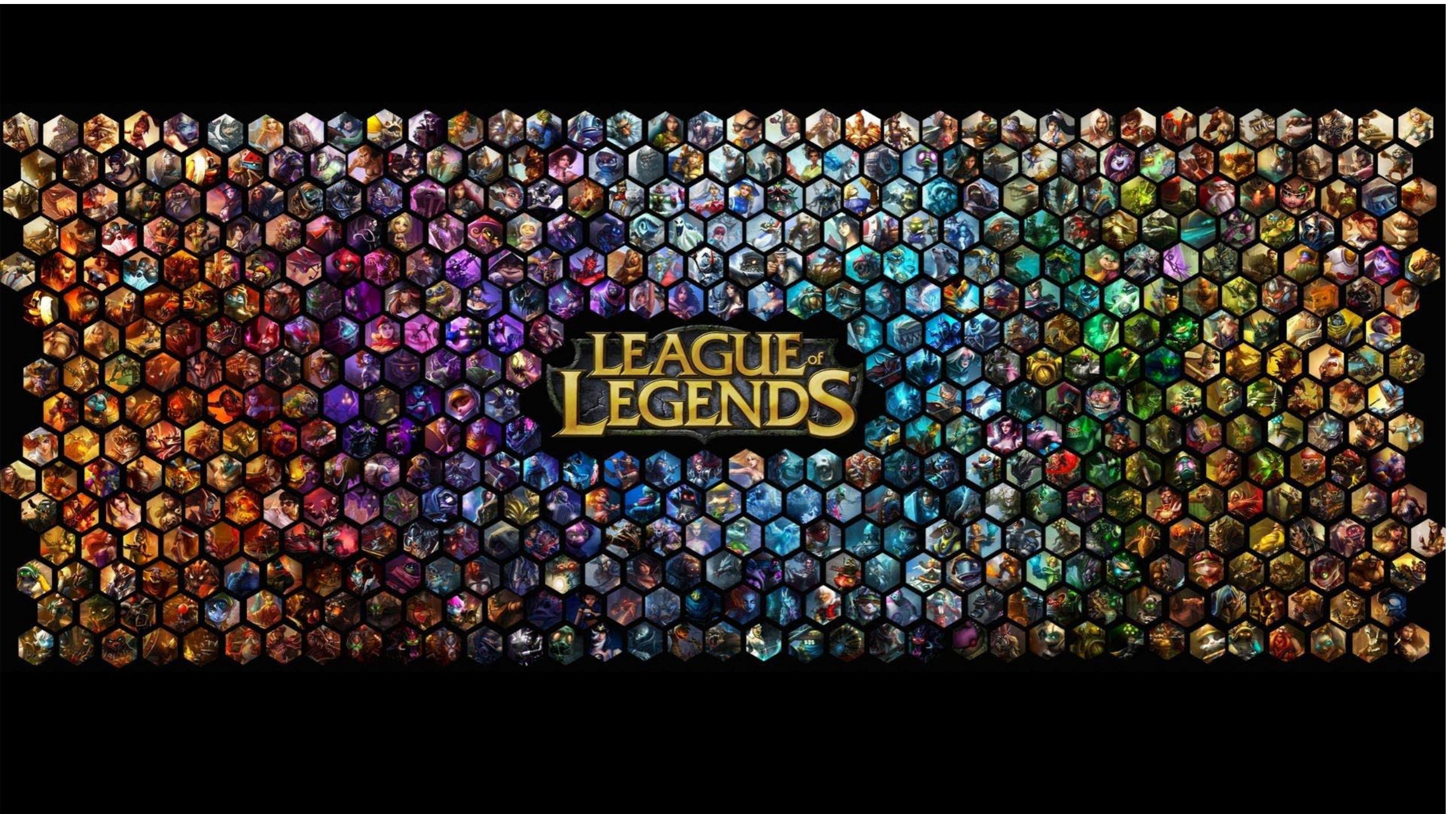
Another detect model





Spark MLlib

-Recommendation system



reference:<http://www.slideshare.net/SparkSummit/video-games-at-scale-improving-the-gaming-experience-with-apache-spark>

FEATURED CHAMPIONS SKINS GAMEPLAY ACCESSORIES BUNDLES + RP 🎁

CATEGORIES

- Skins
- Chroma Packs

Search 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



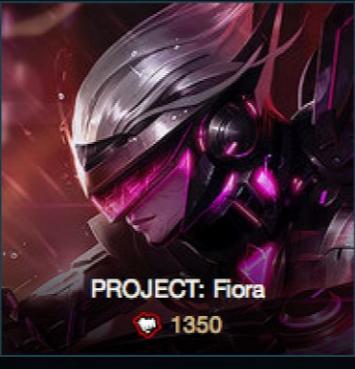
Marauder Olaf 750



Warden Jax 750



Warden Karma 750



PROJECT: Fiora 1350



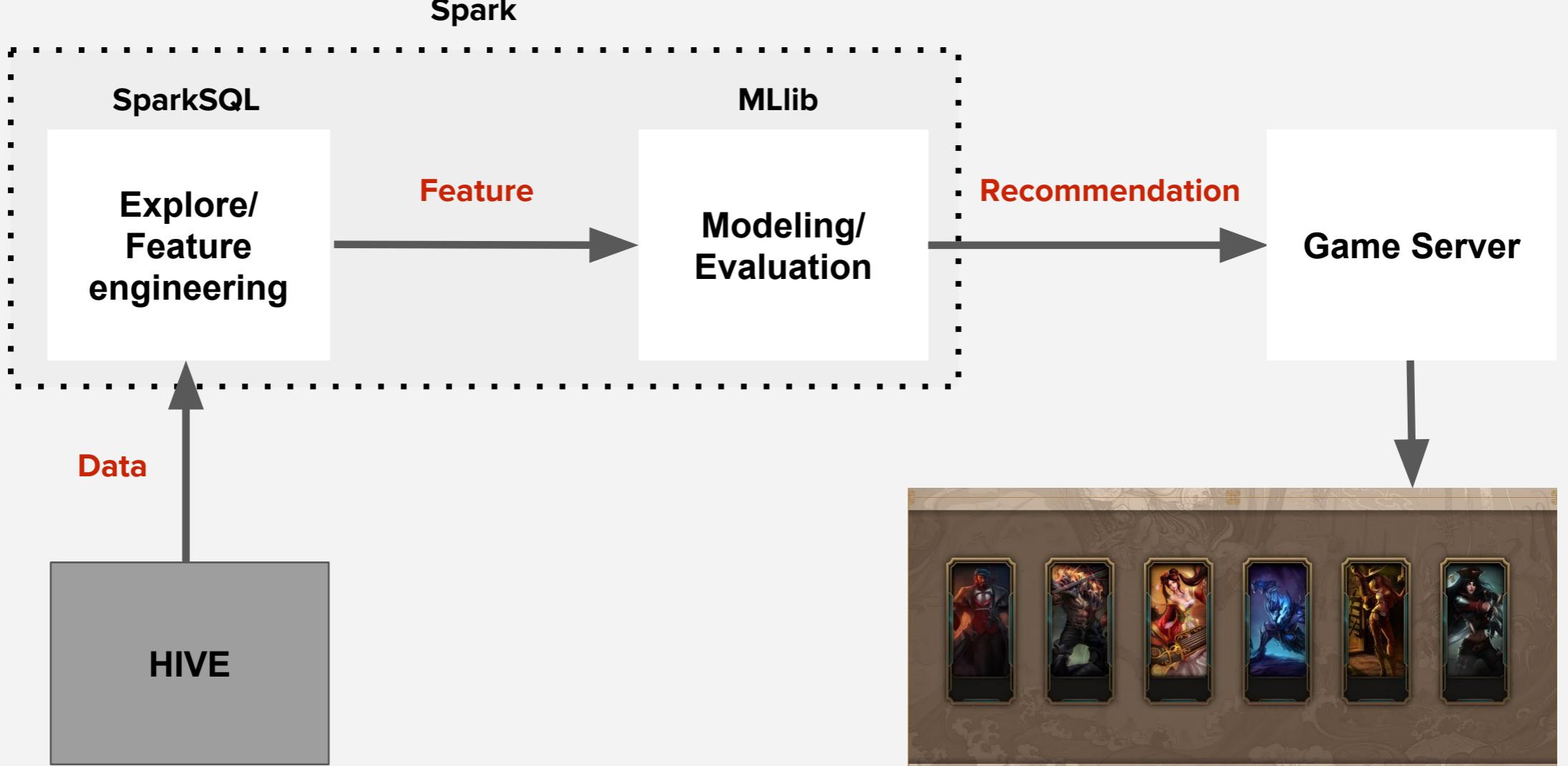
PROJECT: Leona 1350



PROJECT: Lucian 1350



PROJECT: Zed 1350



Q & A

Spark Summit

- <https://spark-summit.org/2016/>
- Slides and video

Spark Cookbook-讀書會新招募

- Ch1. Getting Started with Apache Spark (Chunhung Huang) (4月份)
- Ch2. Developing Applications with Spark (大家自己讀)
- Spark RDD (Allen 加映場)
- Ch3. External Data Sources (昌璐) (8月份)
- Ch4. Spark SQL (何昇道)
- Ch5. Spark Streaming (又晨)
- Ch6. Getting Started with Machine Learning Using MLlib (峰睿)
- Ch7. Supervised Learning with MLlib - Regression (Dean Du)
- Ch8. Supervised Learning with MLlib - Classification (開放認領)
- Ch9. Unsupervised Learning with MLlib (Vito)
- Ch10. Recommender System (Leorick)
- Ch11. Graph Processing using GraphX (開放認領)
- Ch12. Optimizations and Performance Tuning (開放認領)