Big Data with DotA2

How do we engineer a Big Data pipeline for the game of DotA2

Outline

What is DotA2?

What is DotA2?

DotA2 stands for **Defence of the Ancients 2**

- A multiplayer online battle arena video game
- It is played in matches between two teams of five players
- Each team occupies and defends their base
- A team wins by being the first to destroy the other team's "Ancient"

The Business Questions

Easy

1. What is the longest game played?
2. What is the average duration of a game?
3. What was the average first blood time?
4. Has a match been abandoned?
5. How many games were played in a day?

Medium

1. How many games has a particular hero been played in the last week?
2. What is the win rate of a hero in the last week?
3. What is the overall win-rate of a hero?
4. What are the most effective items for a hero?

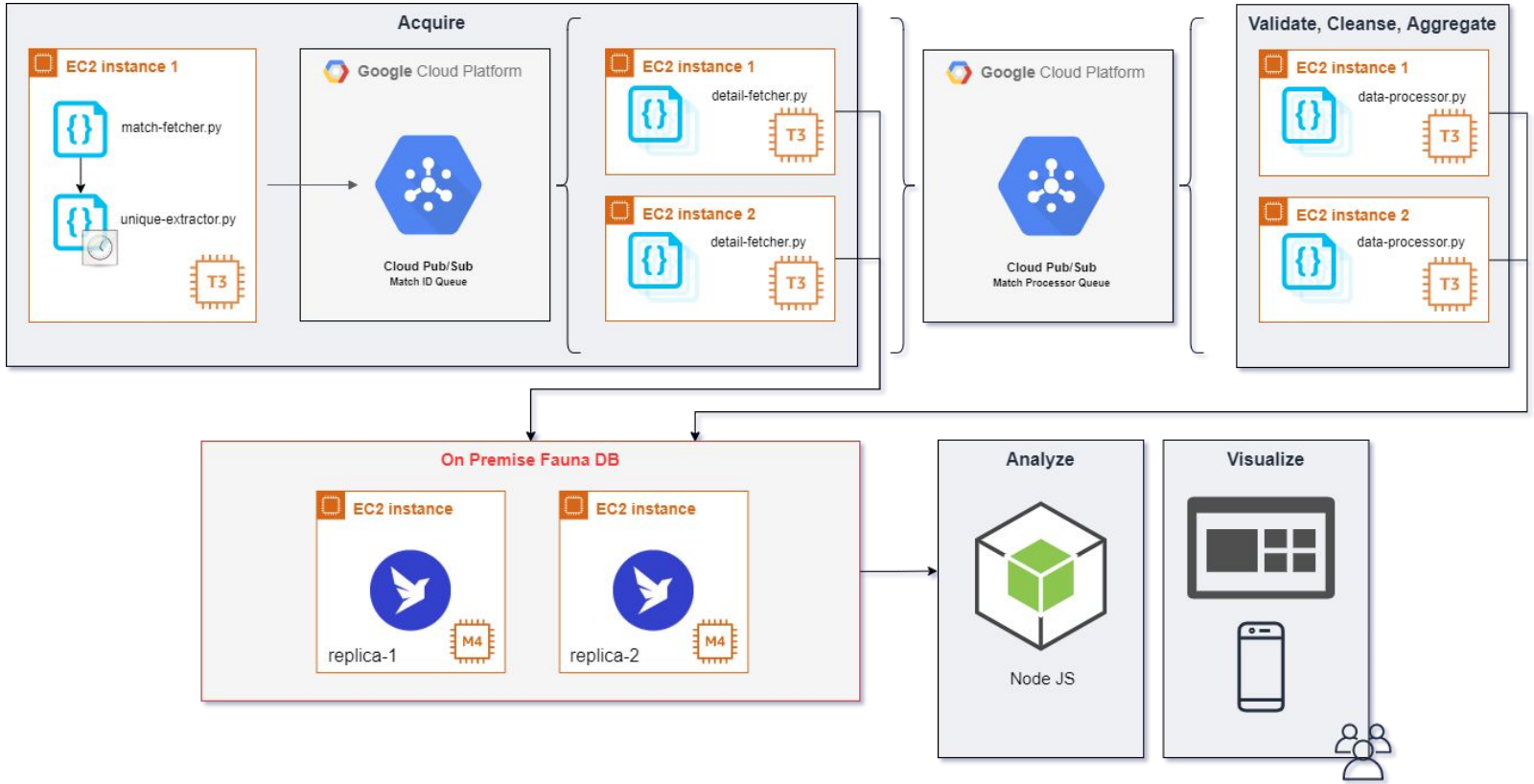
Hard

1. Identify pairs of heroes that work well together. Quantify the synergy.
2. Predict the outcome of a game, given the heroes playing

3.4 seconds

Thats the average time it takes our pipeline to fetch, filter, denormalize, and make available a record

Architecture



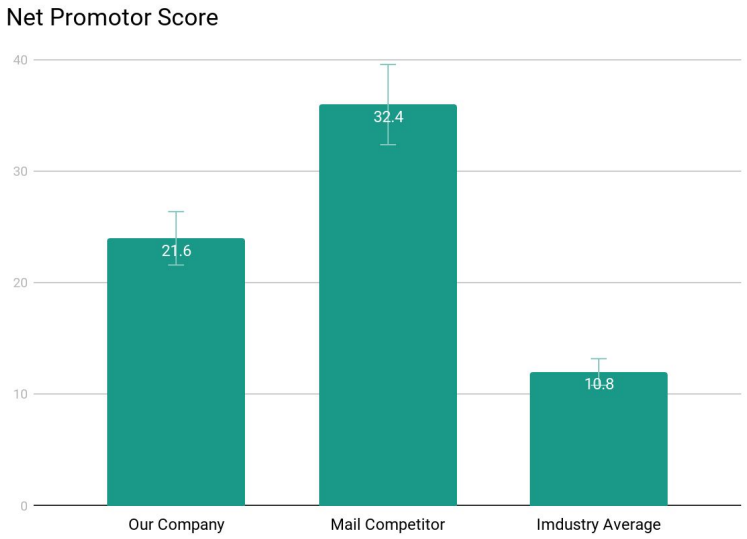
The Data Pipeline

Results

Supporting information

01

List any research or data you have to support the need for a solution.



Challenges

1. Fauna Docker container terminated (with data) last week
2. Duplication issues with data (5 billion -> 250k)
3. Steam API data changes half way through the project
4. Learning FQL and its functions

Next Steps

What next?

- ➔ Implement a replay parser to complement existing data
- ➔ Go Serverless!

Questions?

Big Data with DotA2

Bhaargav Sriraman

Prabjoth Singh Rai

Suresh Siddharth

Vishwesh Mishra