

StarGazer:

Data-Driven Insights in StarCraft II

Guang Yang

Motivation

Real-time strategy game, developed by Blizzard Entertainment

~40 sponsored tournaments in 2014, prize pool ranging from \$25,000 to \$250,000

Tournaments are televised (Twitch, Korean TV networks, ESPN)

300,000 active players

Most games are 1v1 matches



Data

- Data scraped from ggtracker.com/api
- Both historical and real-time updated (averaging twice per minute)
- For each match, there are two files (both JSON):
 - simple details: static information invariant of the match (map name, winner)
 - extended details: 10s snapshots of resources/supplies & events by frame
- Example: <http://api.ggtracker.com/api/v1/matches/3529593.json>

Ingestion

Before

- 2 Separate JSON files for each match (simple and extended details)

After

- Multiple serialized Avro files
- One Avro file for each second that an event occurred (a snapshot or frame-specific event)

Master Dataset

- A sea of serialized Avro files
- Multiple Avro files for a match
- Collection of matches

Batch Processing

- Executed via Spark, saved directly to Cassandra
- Example query: average game duration on different maps

Real-time Processing

- Under construction :(

Demo

Guang Yang

- B.A. in Computational and Applied Mathematics (Rice University)
- M.S. in Industrial Engineering & Operations Research (UC Berkeley)
- Got into Diamond League as Terran without making any Tanks
- Email: gy8@berkeley.edu
- GitHub: github.com/gy8

Shetland, Scotland, UK

