Introductions

Jared Gordon

Google Customer Engineer

Cloud Technologies, Distributed Systems, App-Dev, Java/Spring

Agenda: Real-Time Data Ingest: Traffic Telemetry (Chicago Buses)

- Problem Statement
- Initial Design
- Demo
- Future Design
- Pricing Estimates
- Recommendations/Next Steps

Bus Events: Problem Statement

Real-Time Data Ingest: Traffic Telemetry:

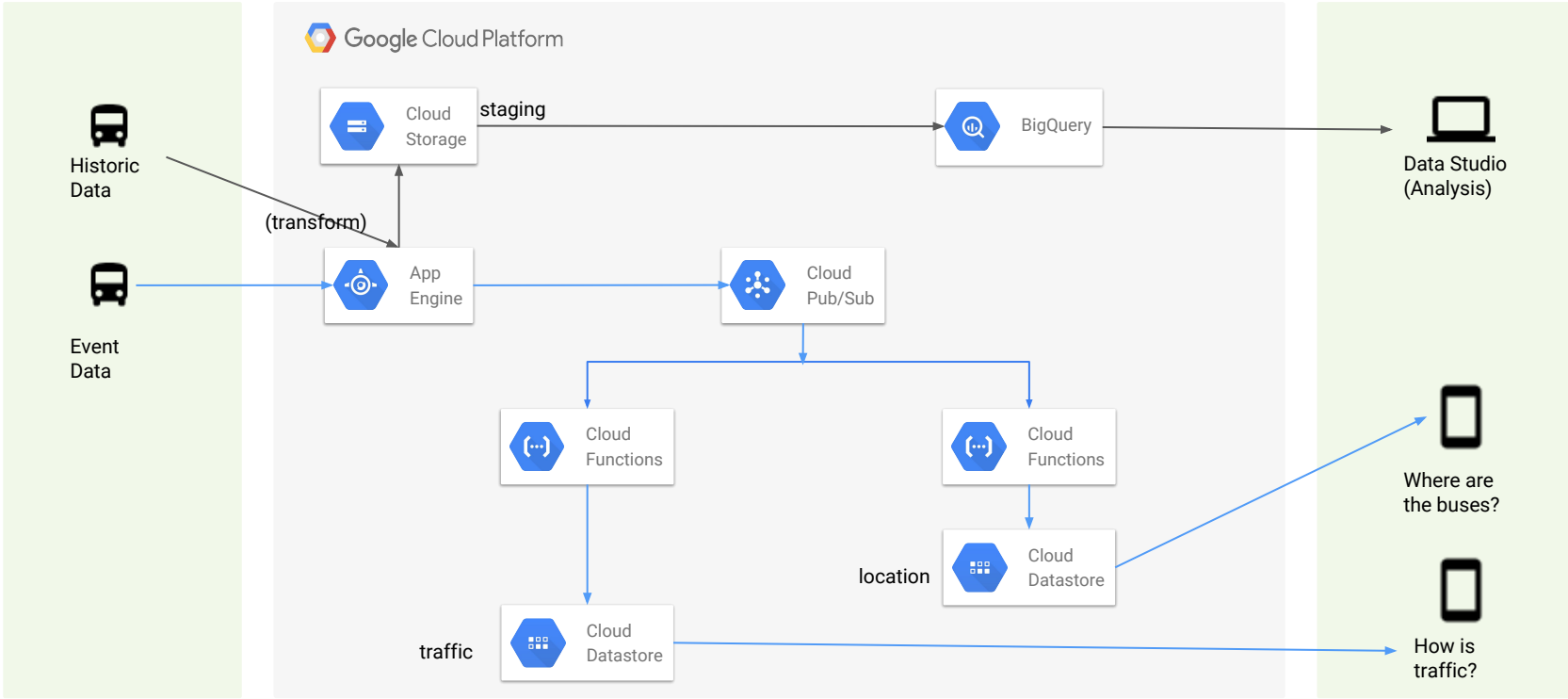
City of Chicago Bus data

- 150,000 near real-time events daily (and growing)
- plus several years of historic data

Goals:

- Provide a means of collecting and transforming the data
- Show ways to organize and analyze the data
- Explore ways to visualize and present the data to end users

Architecture: Initial Design



Demo

Recent events endpoint: https://celestial-brand-212615.appspot.com/

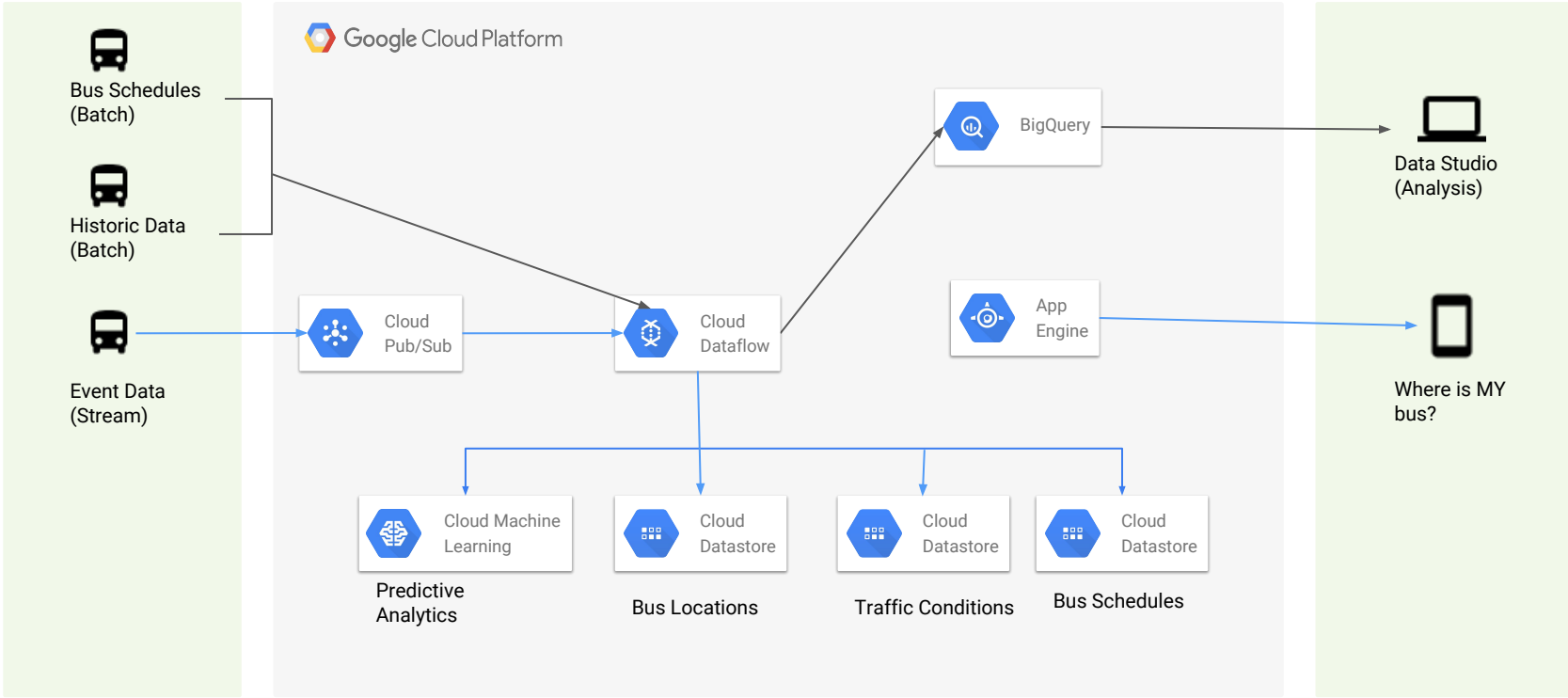
Recent bus sightings: https://celestial-brand-212615.appspot.com/buses

Current traffic conditions: https://celestial-brand-212615.appspot.com/traffic

Trends: https://datastudio.google.com/u/0/reporting/1_QbpHl7NGotvjhQArS6B1UFPP8EOMu3y/page/l2sX

Code: https://github.com/JaredGordon/buses

Architecture: Future Design



Estimated Monthly Pricing (Current Architecture)

AppEngine	1500 hours CPU 1500G mem	$90
Cloud Storage	16 GB regional	negligible
BigQuery	4G added monthly	negligible
Pub/Sub	4M per month	negligible
Functions	3x 250MB, 4M invocations	$80
Datastore	2 x 5M bytes, 3.6M writes	$110
Total		$280

https://cloud.google.com/products/calculator/#id=a39a4e5a-38c3-4e87-9522-27442c15ccd4

Estimated Monthly Pricing (Future Architecture)

AppEngine	1500 hours CPU 1500G mem, 3 instances	$260
Cloud Storage	25 GB regional (after one year)	negligible
BigQuery	4G added monthly	negligible
Pub/Sub	4M per month	negligible
Cloud Dataflow	1G data per month, 5 workers	$10
Cloud Machine Learning	Standard, 1000 hours	$50
Datastore	2 x 5M bytes, 3.6M writes	$110
Total		$430

https://cloud.google.com/products/calculator/#id=21c040be-0450-435f-8c57-2fa30b4d670c

Recommendations

Next steps:

- Build upon the current initial architecture (robustness, error handling).
- If further cities are added, extend via the use of additional publishers, subscribers, and transformers.
- If real-time data becomes available, establish a better, non-polling data ingestion mechanism.
- Continue to look for ways to add value to customers via the use of mapping and other visualizations.

Questions?

Thanks