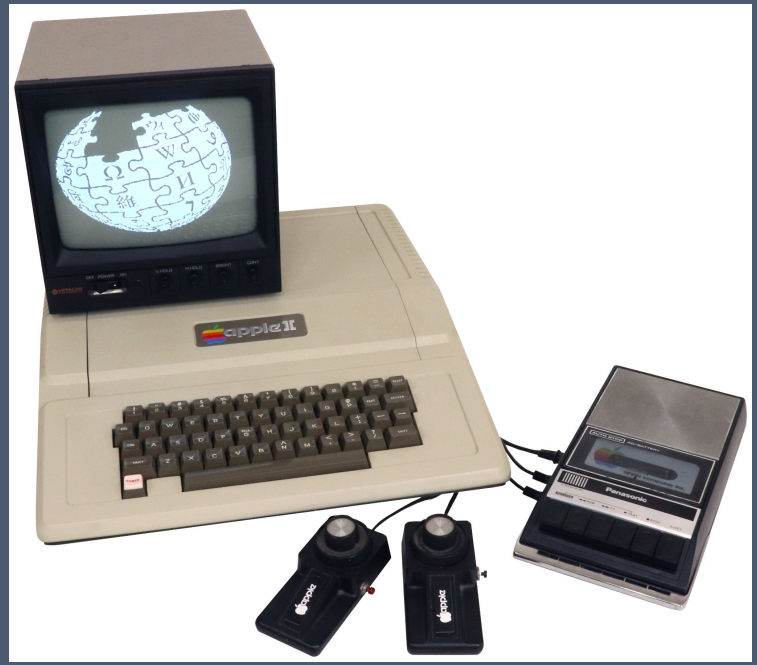
# ClickHouse At Scale

## My Story w/ TWC & Comcast

# whoami?

- 20+ years as a lead engineer
- A decade in project-based consulting
- Skeptical of the latest magic bullet.






ClickHouse

# Why Am I Speaking?

- As a technical lead and ‘expert’ it was my responsibility – both before and now – to get stuff done
- ClickHouse was the magic ingredient to ‘get stuff done’
- I liked it so much…I joined.



ClickHouse

# Big Data at TWC

- Initially Scala -> Kafka -> Flume -> Hadoop
- Slow Hive Queries to populate a MySQL database to actually be usable for some limited operations dashboards
- Eventually developed into a Frankenstein of Druid/Spark/Parquet/Kafka

# The Holy Grail

## Ingest speed above all the things

- CDN is used to deliver streaming IP video to most of Comcast's cable customers.
- I was hired based on my cable and big data experience at TWC.
- The CDN was consuming more than 80% of Splunk license.
- "How would you handle 2 million records per second?". "I don't know, buy a lot of hardware?
- I tried ClickHouse on a few of indexers
- Kafka engine on one reading our ElasticSearch JSON, and in one weekend of playing around with CickHouse we matched that 400k ingest speed on 1 server instead of 10

# A Summary

- Ingest speed was the critical factor, but the other benefits of ClickHouse quickly became evident

- Query speed.  When you have an operations team that is used to waiting minutes to hours for Splunk queries, seeing the real time results  from hundreds of CDN cache servers in a second or two makes you very, very popular.

- Query speed enables REAL TIME alerting

ClickHouse

Source: Wall Street Journal.
Note: Lorem ipsum dolor sit amet.
(1) Consectetur adipiscing elit.
(2) Consectetur adipiscing elit.

# The Future???

**A logical question.**

# Cloud

(that’s it. that’s the future)