

# Scalable Analytics

F\*ck Sampling

John Adams, CTO

Bringhub

# Introduction – A little about me

- CTO at Bringhub
- Family man
- Lover of data
- Hater of PHP
- Lover of Go



*Beautiful*

# Obligatory Golang Plug

The right language to use... pretty much always



(Join us at our Golang-StPete meetup, also sponsored by Snowflake <3!)

bringhub

# History of Analytics at Bringhub

# Syslog -> SQL

A simple solution, and a great place to start



# But it wasn't perfect

- Log rotation crapping out... oops
- Fluentd memory usage... yikes
- SQL cluster... \$\$\$
- Throughput under scale was an issue



# DynamoDB-> Lambda-> Mongo/Keen-> Custom Dashboards

- A better solution for us
- It allowed us to partition our data dynamically, and really scaled with our needs.
- Written in Node & Go (read: not PHP)
- Writing to Keen made CSV dumps easy



mongoDB®



Keen IO



bringhub

# But, it wasn't perfect.

- Mongo Cluster for raw data ... \$\$\$\$\$\$\$\$\$\$
- Lambda debugging was sad at best
- Most BI solutions (that are affordable) don't support MongoDB
- Custom pipelines are expensive in engineering resources
- Turn around time for custom dashboards was too long
- Keen's Dashboard solutions failed to meet business needs

No! No you didn't!

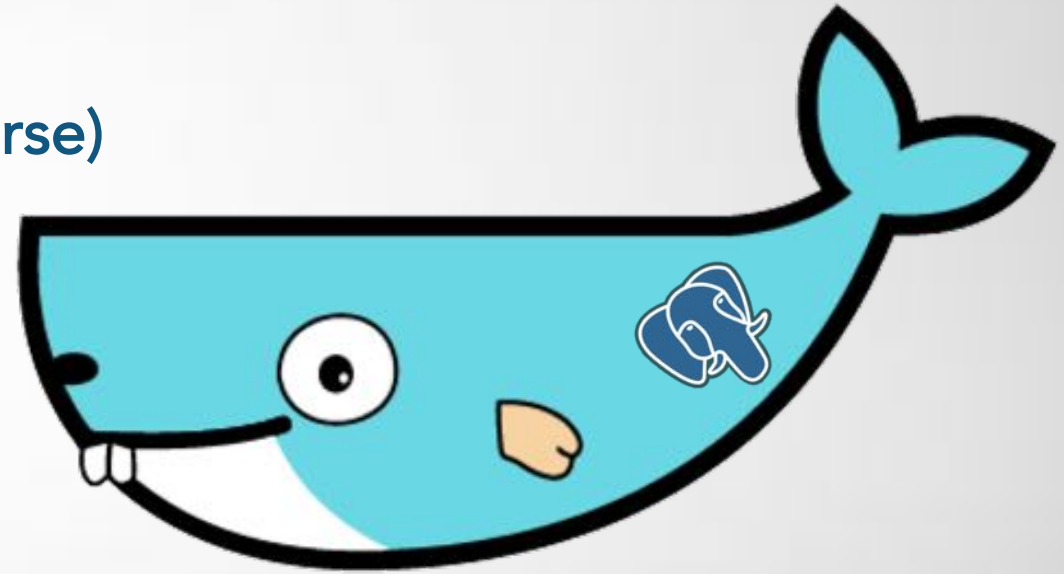




# Docker Cloud -> Go API -> Postgres -> Periscope

Since we were finally able to cast off the shackles of our oppressors (PHP), and had decided a containerized Go based environment was the ultimate solution to our stack... we naturally wanted to write our analytics there.

- Single language for all the things! (Go of course)
- A familiar stack
- Data is easily queryable
- Dashboard providers galore!
- Client Success could develop their own reports!!!



# But, it wasn't perfect.

- Docker Cloud's networking had some scaling issues... (k8s plug here)
- Ingestion was cheap and fast, but required devops
- Cost of our growing clickstream data was extremely high
- We aren't DBAs, and our query performance suffered
- Rigid relational data structures... egad



# The Problem(s)

# Winter is coming

- Partnerships are on the upswing
- Analytics outages aren't acceptable
- We're a cash starved startup
- We need a stack that scales
- We need a DB that we can query with SQL, but has the wiggly-ness of NoSQL



Platform adoption leads to traffic  
Traffic leads to click events  
Events lead to empty wallets

- Yoda

# What we needed from our stack

- Real-time\* processing of clickstream data
- Scalable fast ingestion
- Laissez faire database management
- Affordable petabyte warehouse
- Highly concurrent access to all the bits and bytes
- Flexible analytics data models

## The solution

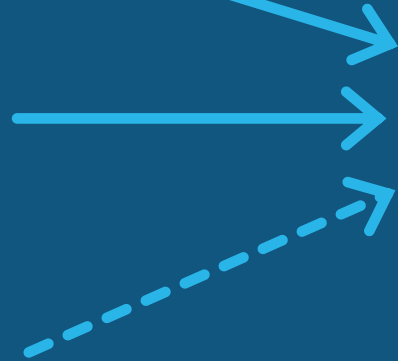




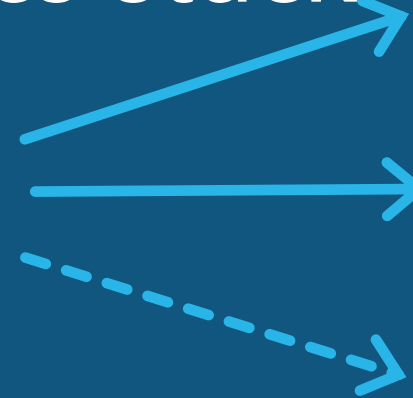
alooma



# Our new analytics stack



Snowflake



Periscope



looker





# Problems solved

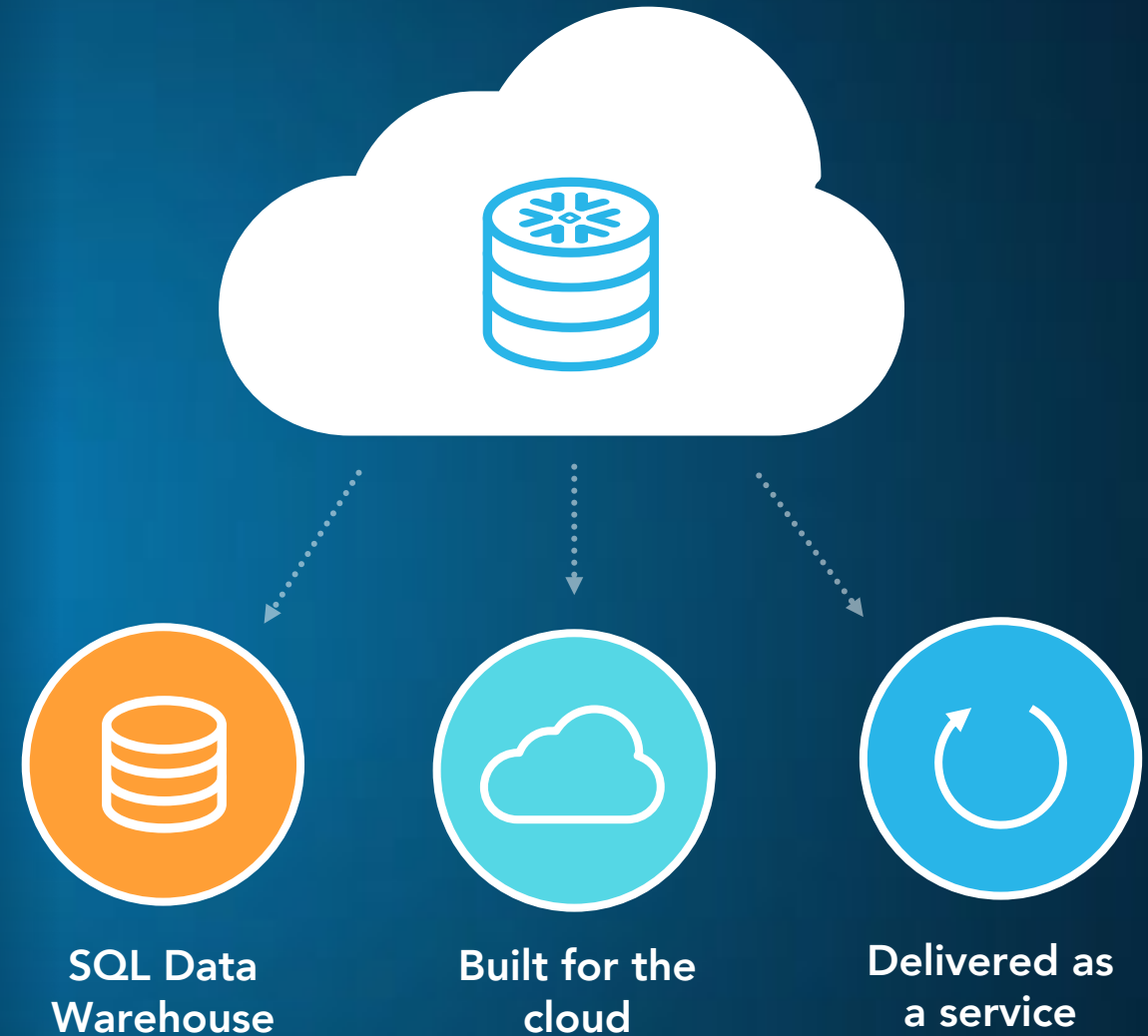
- As fast as we can write to S3 (or an ETL vendor of our choice), we can load into Snowflake
- Vendors galore, we can pick the ones that makes the most sense for us
- Loading data has zero impact on our read performance



# Why Snowflake?

## Snowflake Solution

All-new data warehouse built from the ground up for the cloud and for today's data and analytics

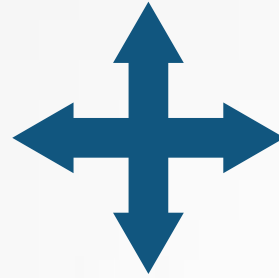




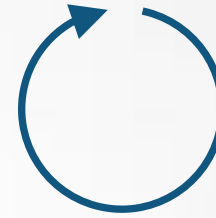
# What makes Snowflake different



Structured + semi-structured data



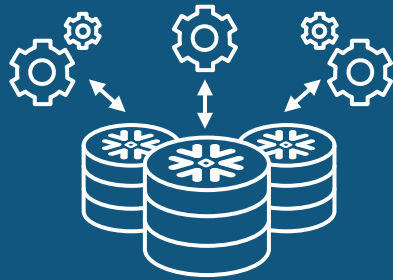
Scalability



No management



Concurrency



Unique architecture:  
Multi-cluster, shared data

# Problems solved with Snowflake



Laissez faire database management



Affordable petabyte warehouse



Highly concurrent access to all the bits and bytes



Flexible analytics data models



Thank you.