
CHEQ

# Our Journey to Bot Detection with ClickHouse over Kubernetes

Daniel Amsellem, Chief Architect

# About the company.

# CHEQ is the global leader in Go-to-Market Security.

Protecting the entire Go-to-Market operation from malicious, automated and human-driven threats.

Leader
SUMMER
2022

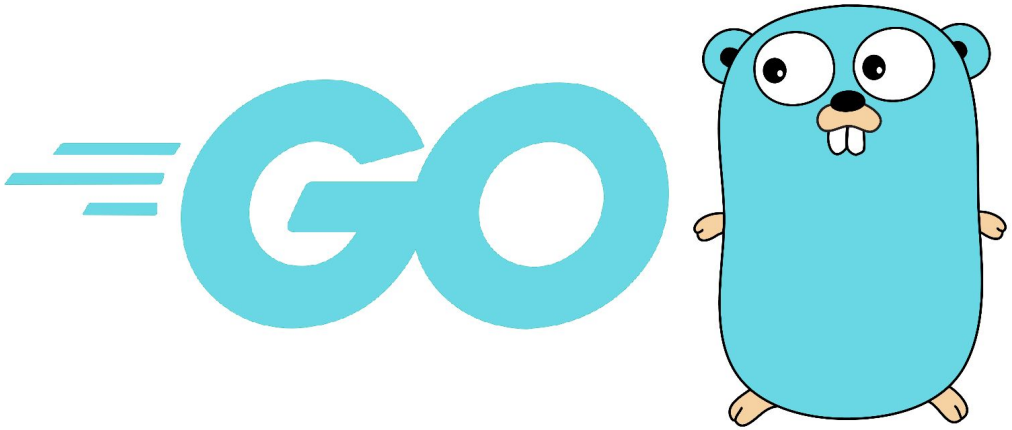
Based on 1000+ reviews

Capterra Trustpilot

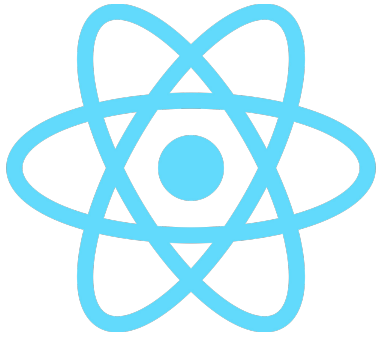
# CHEQ's Solution for GTM Security

- Proprietary fraud detection engine, applied in Real-Time
- Advanced client and server side techniques
- Analyzing server-side signals via the HTTP Protocol
- STUN / DNS for  uncovering masked users' IP addresses
- JS Techniques: Challenging users' browser data with dynamic JavaScript tests

# Tech Stack







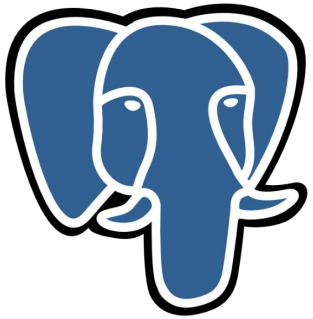

# Data, Data and Data

ClickHouse

Apache Airflow

redis

MySQL®



PostgreSQL

Amazon S3

amazon Kinesis

# Data Setup before ClickHouse

- Raw JSON files in S3
- Athena + Presto for analytical queries and reports

- Athena is very expensive and very slow in our growing scale
- 200 GB/Day (~600 GB in Parquet) , 35K RPS
- POC with ClickHouse:
  - Using CH on physical machine

# Our Journey to ClickHouse - Why Kubernetes?

Scalability

Resource Allocation

High Availability

Automation

# Our Journey to ClickHouse - Why Kubernetes?

- Disks are isolated from machines and can be easily reattached
- Cloud-native tools are used out-of-the-box - Prometheus, Grafana
- Inter-service networking - utilizing k8s built-in networking allowing all other services to talk directly to ClickHouse

**All of Kubernetes advantages + all of ClickHouse advantages in one place!**

# Our Journey to ClickHouse on Kubernetes - Take 1

- Installed open-source ClickHouse Operator
- 10 mins per pod until ready - running chmod for all files at initialization
- Very long YAML with many definitions - reverting made impossible

# Our Journey to ClickHouse on Kubernetes - Results

- Converted the YAML file to proprietary HELM chart
- Reduced readiness time to 2 minutes per-pod by using pod security context
- Rollout made better, stability increased
- Happy ClickHouse Consumers - super fast queries resulting in a far better service to customers

# Takeaways

- Complexity
- Resource requirements
- Networking
- Data durability

# Q&A

# Thank You!

# Thanks.

CHEQ