fťock safety

# Real-Time Traffic Analytics Powered by ClickHouse



Leon Kozlowski: Data Engineering Manager @ Flock Safety

# The Flock Safety Platform

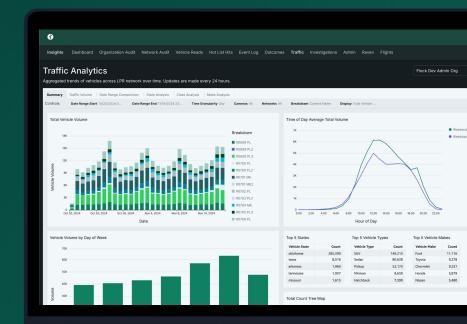


Capture more evidence. Solve more crime.

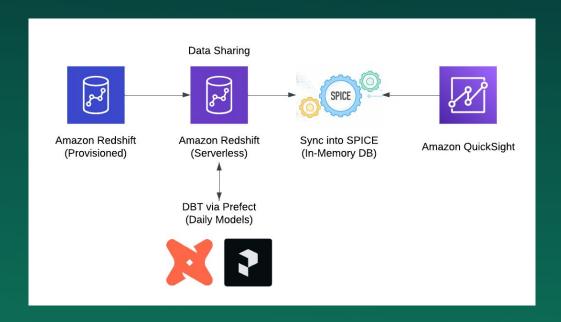
#### frock safety

# What is Traffic Analytics?

Traffic Analytics unlocks the potential of a customer's existing Flock Safety License Plate Recognition technology for critical traffic management insights. This software solution captures unique and total vehicle counts across the LPR network. Data is aggregated in 15 minute bins for fine-grained analysis



## **Pre-ClickHouse Architecture**



- Daily refresh cadence
- Long refresh times



- Data unavailable during refresh
- Hard size constraints
  - 0 1 TB or 1 billion rows\*
  - o 256 RLS keys

#### fłock safety

\* All of our reports were capped to 30 days (or less) due to size limitations of SPICE

fłock safety

# **Row Level Security and Sharing**

Data security is paramount at Flock, RLS rules were enforced using QuickSight to ensure only data owned by an organization can be viewed in analytics

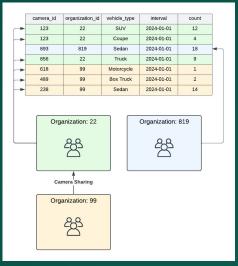
#### **Constraint:**

 Only 256 unique key are supported with Amazon QuickSight

#### **Problem:**

- Organization can share cameras with other organizations based on roles
  - In many scenarios users will have access to more than 256 cameras





frock safety

# Key Issues:

- Limited retention due to SPICE dataset size limits
- Downtime associated with datasets refreshes
- 256 key limit for row level security with SPICE
- Daily refresh cadence (gap in reporting intra-day)

#### What we tried:

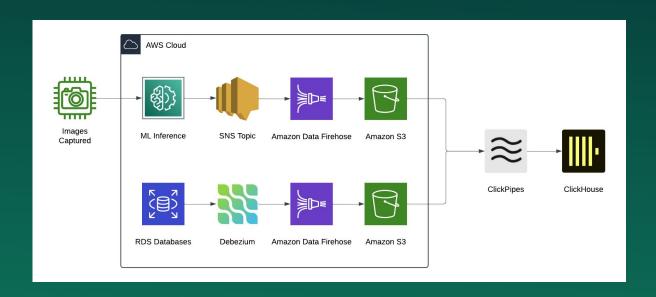
- Redshift (Prefect + DBT)
  - Daily → SPICE
- Aurora Postgres (Prefect + DBT)
  - Daily → Direct Query
- S3 + Trino (Prefect + DBT)
  - Daily → Direct Query
- Clickhouse + AMT & MV
  - Real-time → Direct Query

#### fłock safety

#### **POC Results**

- 1. Redshift (via Prefect & DBT) → SPICE
  - SPICE refreshes make data unavailable for 3+ hours
  - b. Camera sharing support blocked by 256 RLS key limit
  - c. Reporting not real time (daily)
- Aurora (via Prefect & DBT) → Direct Query
  - a. Reached maximum timeout for Amazon Quicksight
  - b. Reporting not real time (daily)
- 3. S3 + Trino (via Prefect & DBT) → Direct Query
  - a. Reached maximum timeout for Amazon Quicksight
  - b. Reporting not real time (daily)

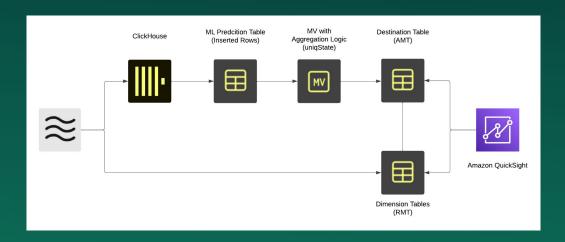
# ClickHouse: Ingestion and Data Volume







# ClickHouse Architecture Diagram

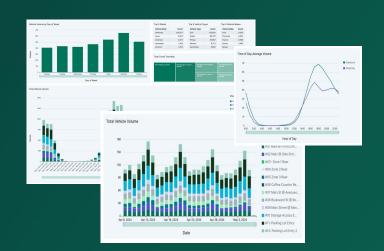


- Use uniqMerge to query destination table
- Join dimension tables to enrich reporting layer + support camera sharing via multiple RLS keys
- Use Direct Query via Amazon QuickSight to remove size constraints that we had with SPICE
  - Unlimited retention
  - Full scope of client data
  - Richer insights
  - Real-time + no refreshes

fłock safety

# Results of Traffic Analytics on ClickHouse

ClickHouse has unlocked zero downtime, enablement of shared camera data availability, and extended retention for our customers



#### 0 sec

Downtime due to removal of SPICE datasets

# Intra-day

Full scope of intra-day data available

## <5 sec

Query time for real time counts



Historical traffic counts for our customers

# Final Thoughts & Next Steps

Impact: ClickHouse has unlocked the following concepts for Flock's Traffic Analytics Product

- Real-time analytics
- Camera sharing support
- Extended retention
- Zero downtime

#### **Next Steps**

Continue moving more analytics use cases to ClickHouse