



MBL202

Taking Data to the Extreme

Gaurav Kumar, Christian Lam, David Winters

November 30, 2016



Taking data to the eXtreme

Our Speakers



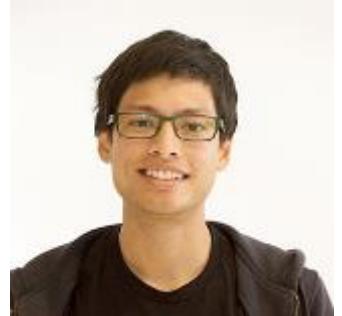
Gaurav Kumar

Product Lead,
Data Science &
Engineering, GoPro



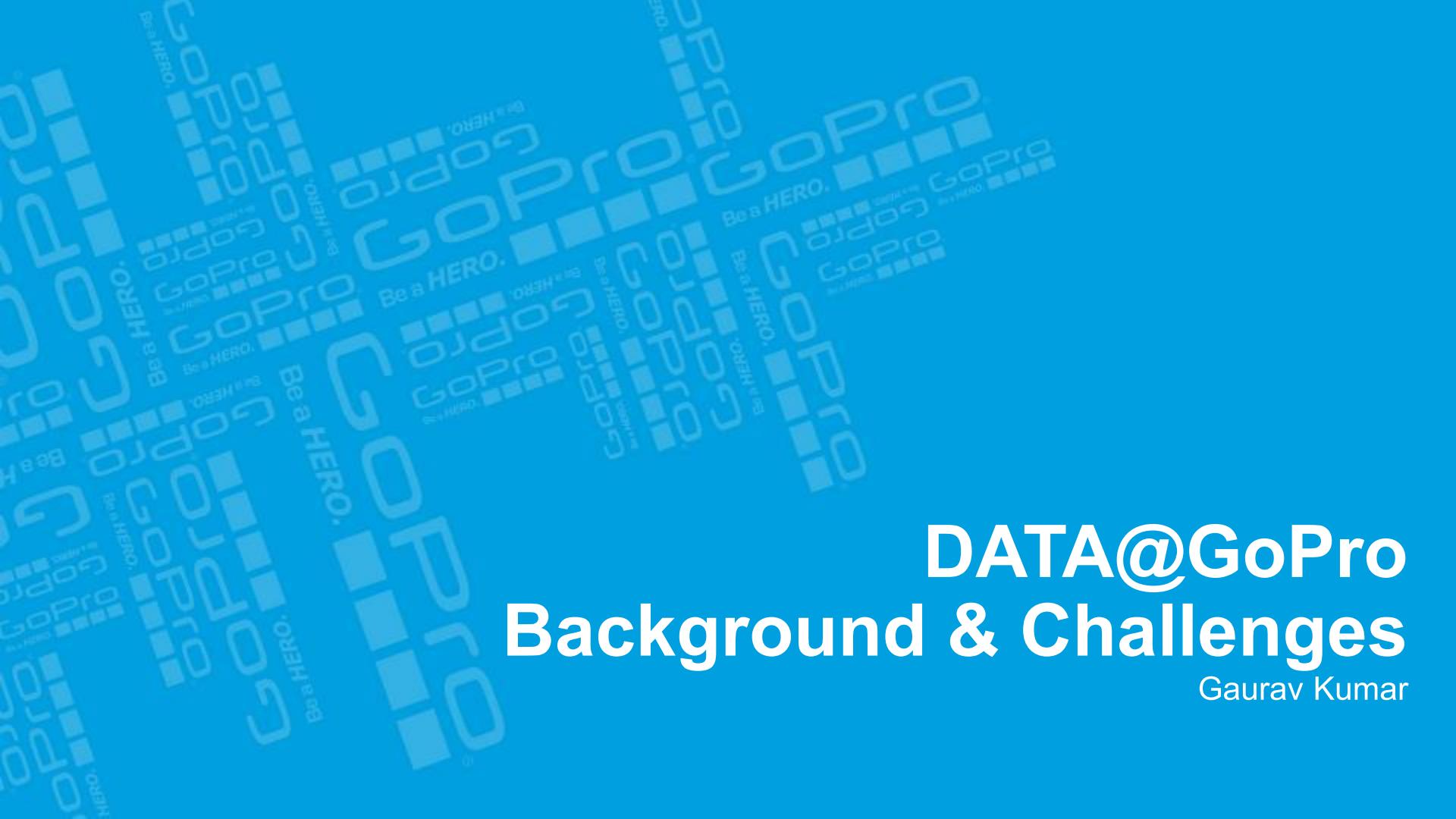
David Winters

Big Data Architect,
Data Science &
Engineering, GoPro



Christian Lam

Analytics Engineer,
Data Science &
Engineering, GoPro



DATA@GoPro

Background & Challenges

Gaurav Kumar

When we got here...

Data Analytics was based on Word of Mouth (& this guy)



Unique Challenges Data@GoPro – Growing Data Needs – X B events/day



CAPTURE



CREATE



ENJOY



DEVELOPER
PROGRAM

KARMA



Unique Challenges for Data @ GoPro – Variety of Data

- Variety of data - Hardware and Software products
 - Software - Mobile Apps
 - Hardware -> Cameras, Drones
- Variety of data ingestion mechanisms
 - Batch pipeline
 - Real-time pipeline
- Complex Transformations
 - Data often stored in binary to conserve space in cameras
- Seamless Data Aggregations
 - Blend data between different sources, hardware, and software
 - Build structures that reflect state vs. event-based
- Handle Privacy & Anonymization



How does AWS help us? (*Go Fast*)

- Faster lead times to procure and set up
 - Months vs. Days
- Quick response to elastic demand
 - Increase size of disks or compute nodes for peak demand
- Easy to try new technologies with minimum risk – Kafka, Spark



How does Tableau help us? (*Go Deep with Data*)

- Insights need to be visual
 - Data in itself is meaningless
 - **Data is Context** – Common event taxonomy
- Monitoring → Helps us monitor real-time and batch Ingestions
- Allows business users to do minimal ad hoc analysis with little training



Origin Story



Yes, this comes from the top...

- Make Friends
- Haul Ass
- Maintain Balance
- No Half-Assery
- Integrity. Always
- Be a *HERO*

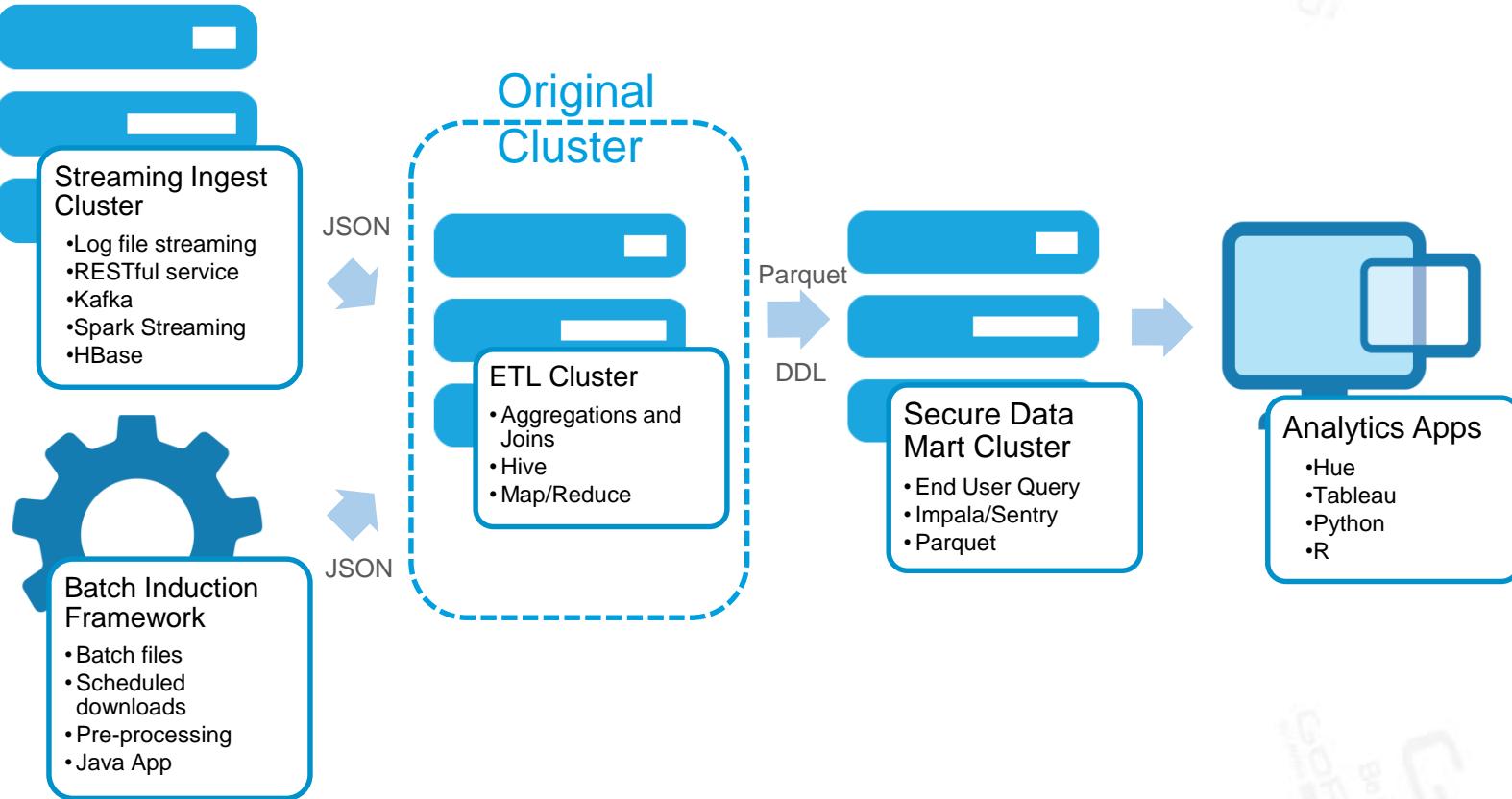




GoPro Data Platform Architecture

David Winters

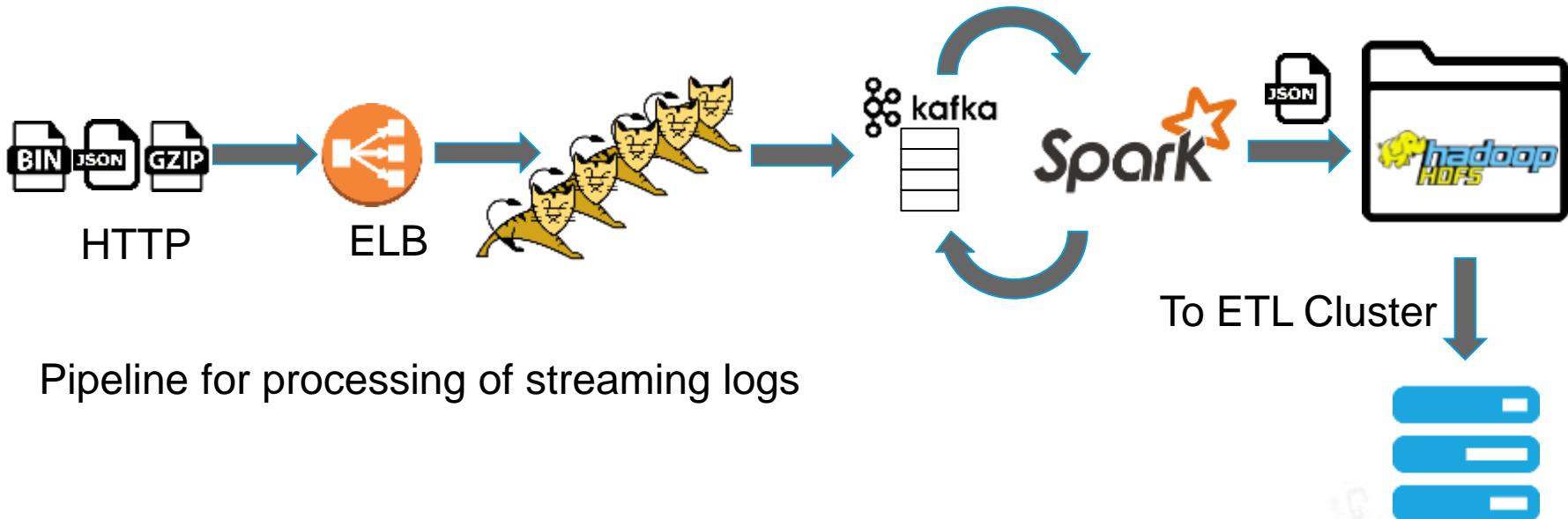
High-Level Architecture



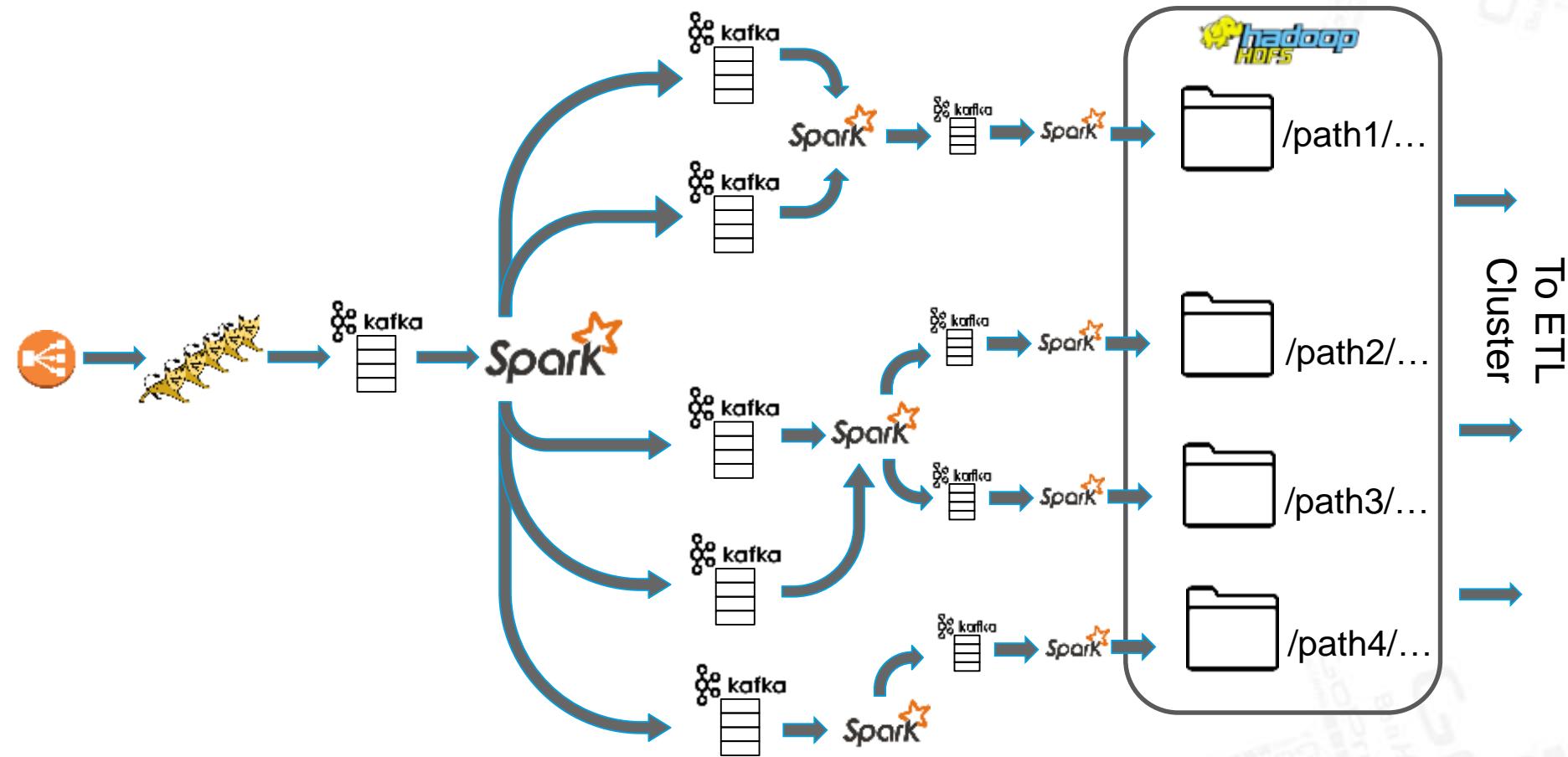
Data Pipeline



Streaming Ingest Cluster

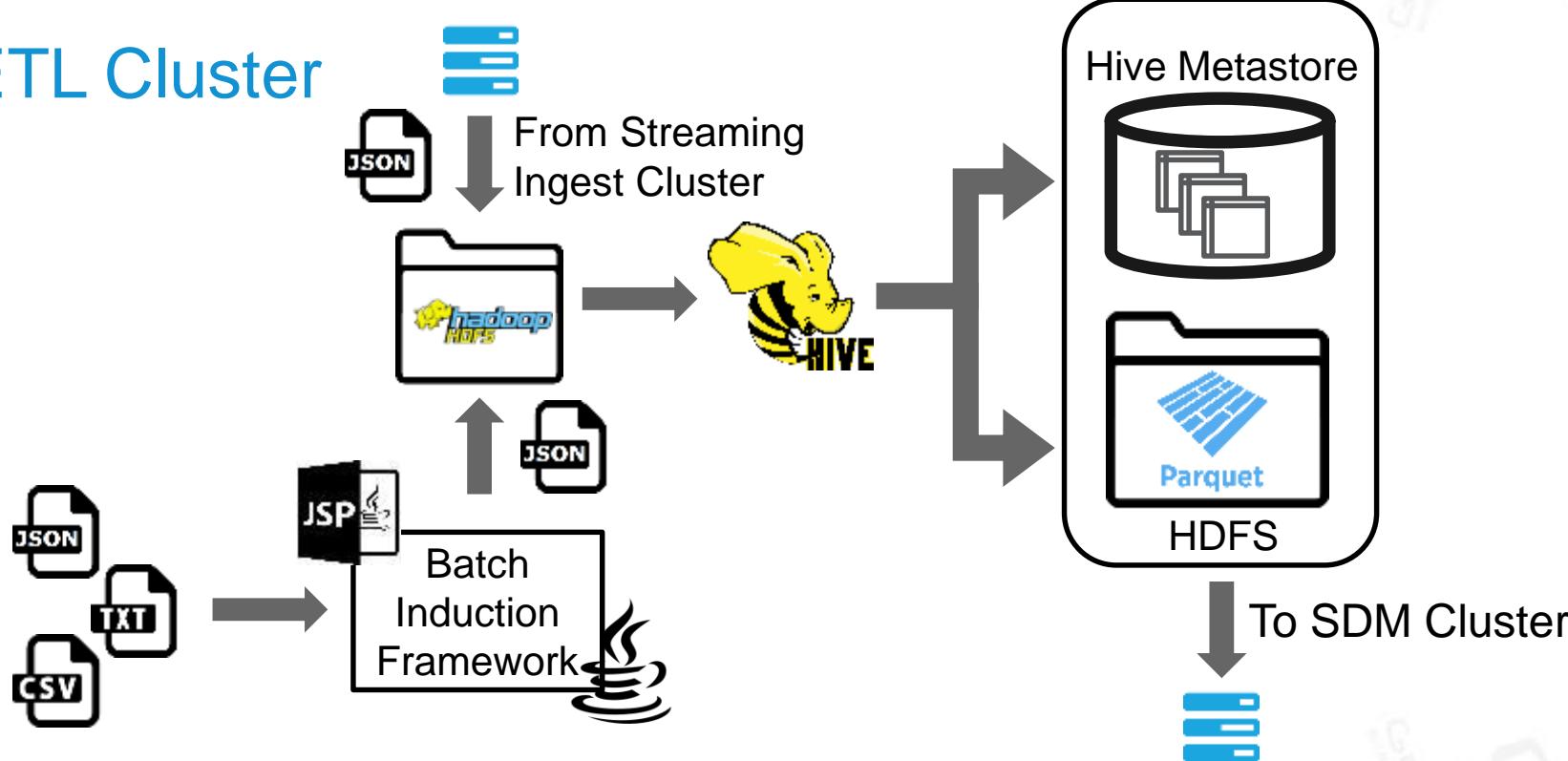


Data Pipeline

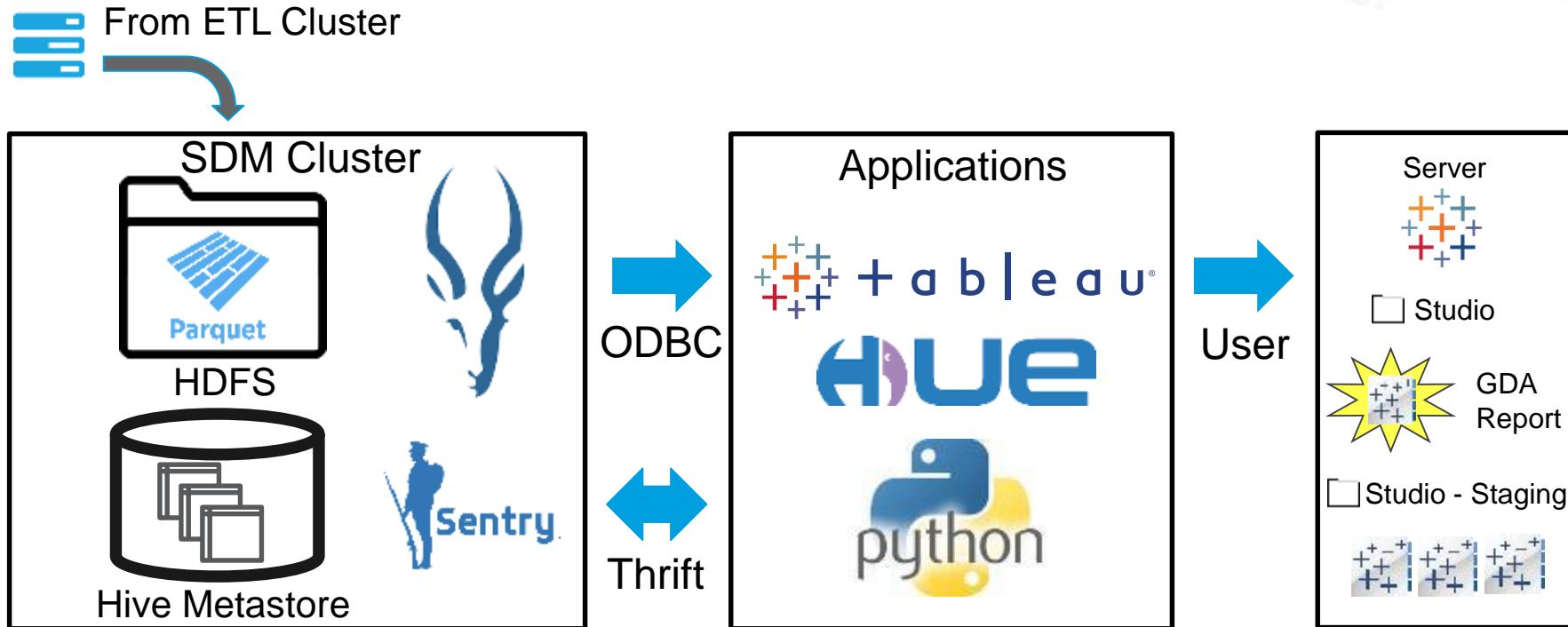


Data Pipeline

ETL Cluster



Data Delivery!



Areas for Improvement

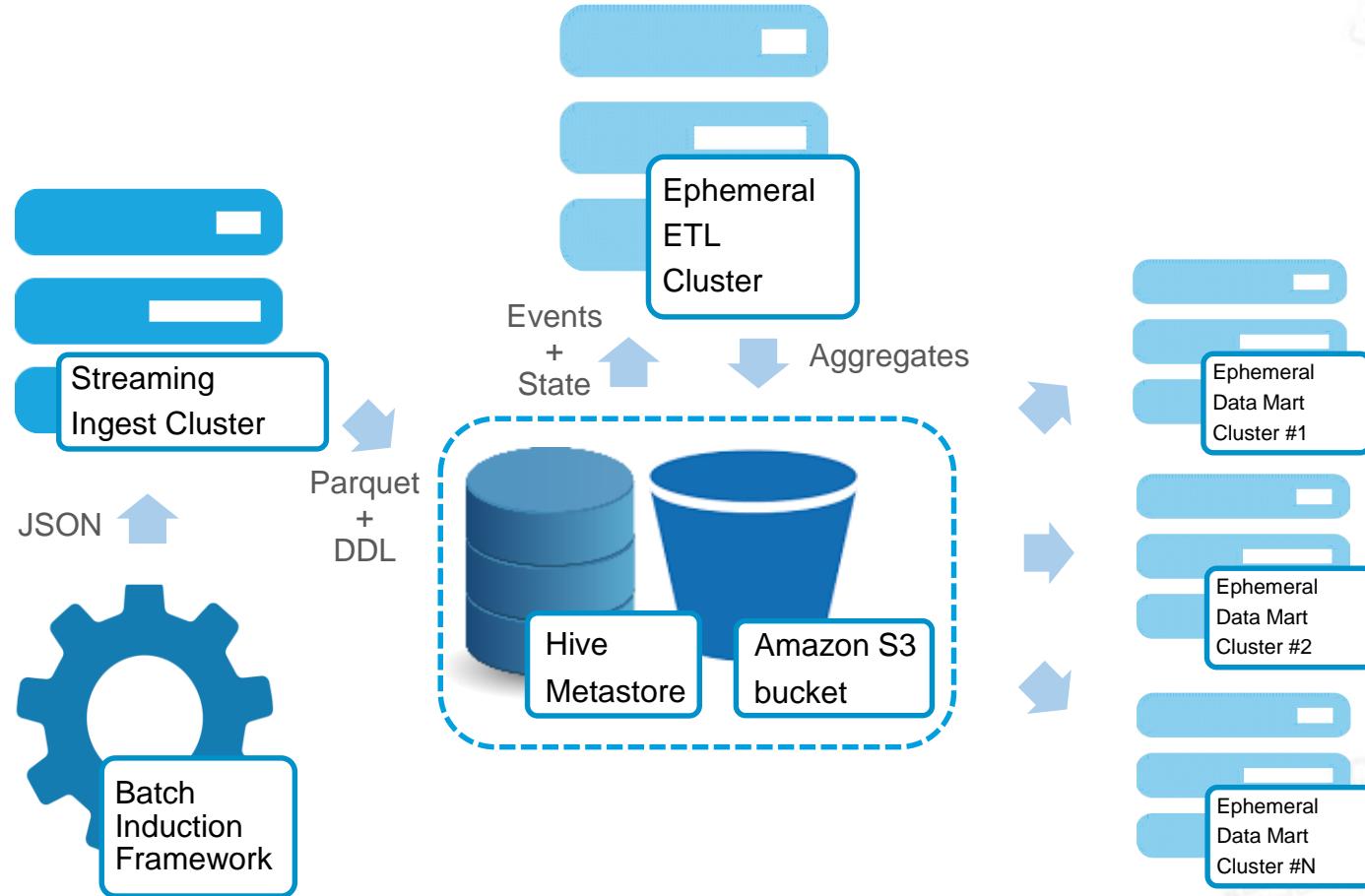


- Isolation of workloads
- Fast ingest
- Secure
- Fast delivery/queries
- Loosely coupled clusters



- Multiple copies of data
- Tightly coupled storage and compute
- Lack of elasticity
- Operational overhead of multiple clusters

Future Architecture



Insights

Christian Lam

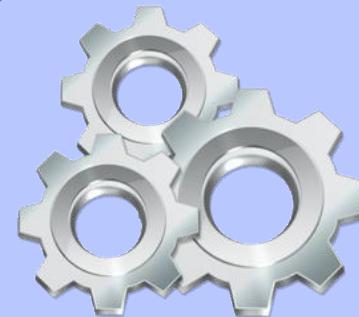


Visualization Categories

3 categories for all visualizations...



PreETL



Operations



Analytics

Data Ops

GoPro
Be a HERO.



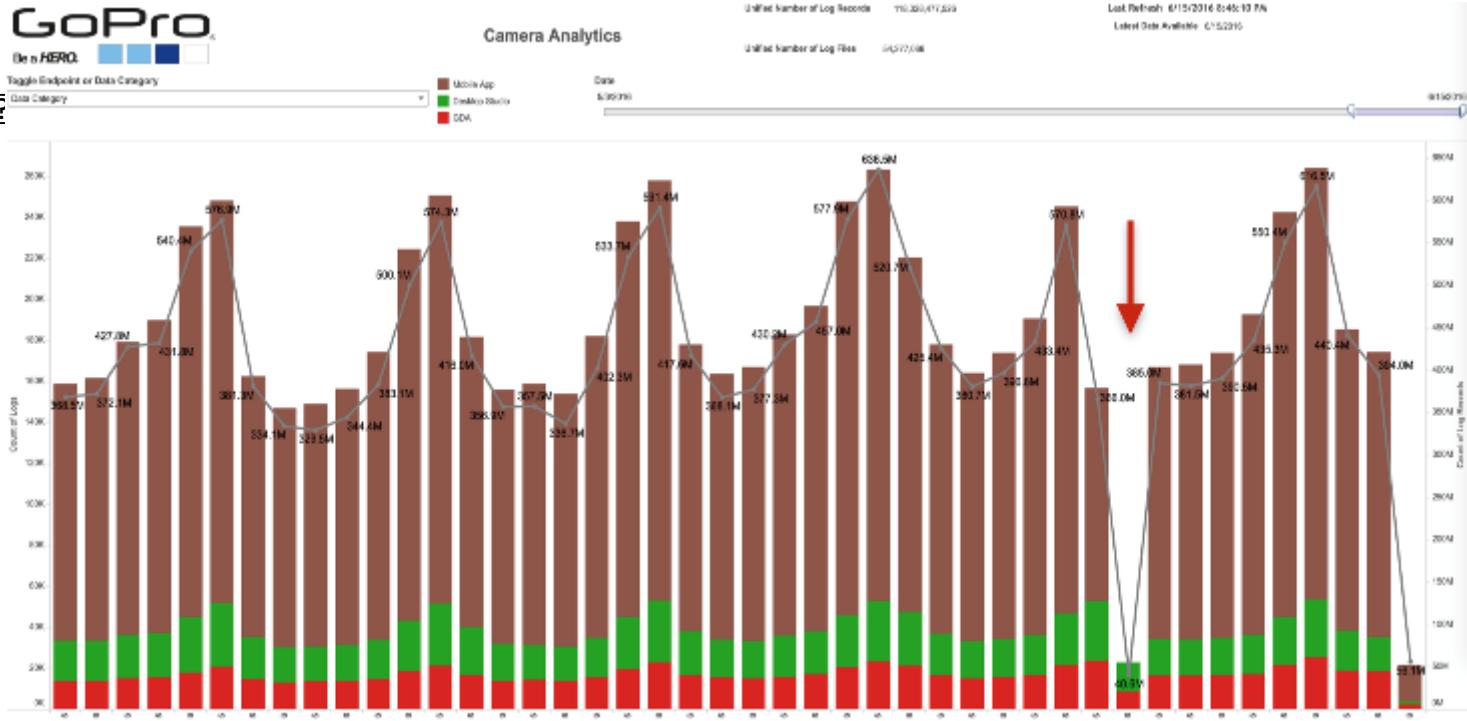
OPERATIONS



Data Ops



DataOps dashboards allow us to monitor the health of data streams and detect anomalies, as well as Tableau Server itself.



Analytic Dashboards



The wow...

ANALYTIC DASHBOARDS
Custom Visualization/
Interaction



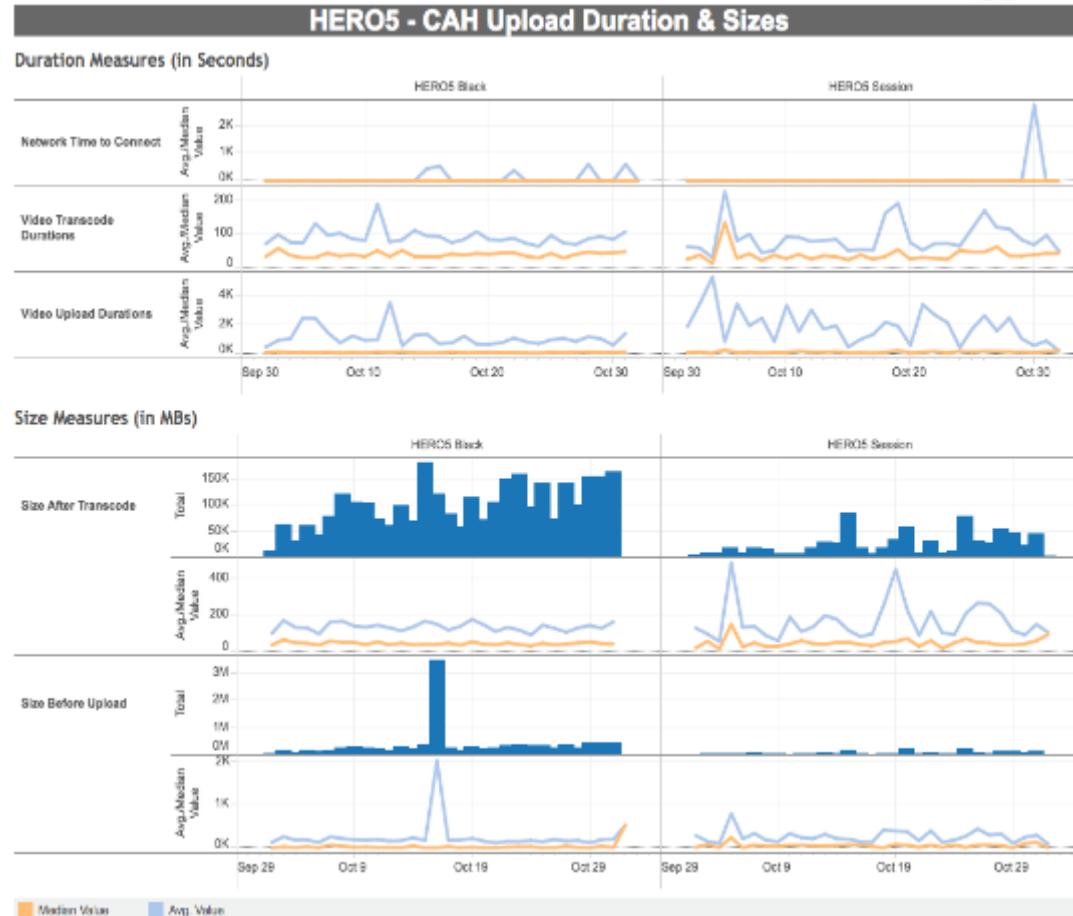
PRE BUILT AGGREGATIONS

- Aggregates are important for the successful adoption of Tableau
- Example: Karma flight table



FEED PRODUCT INSIGHTS

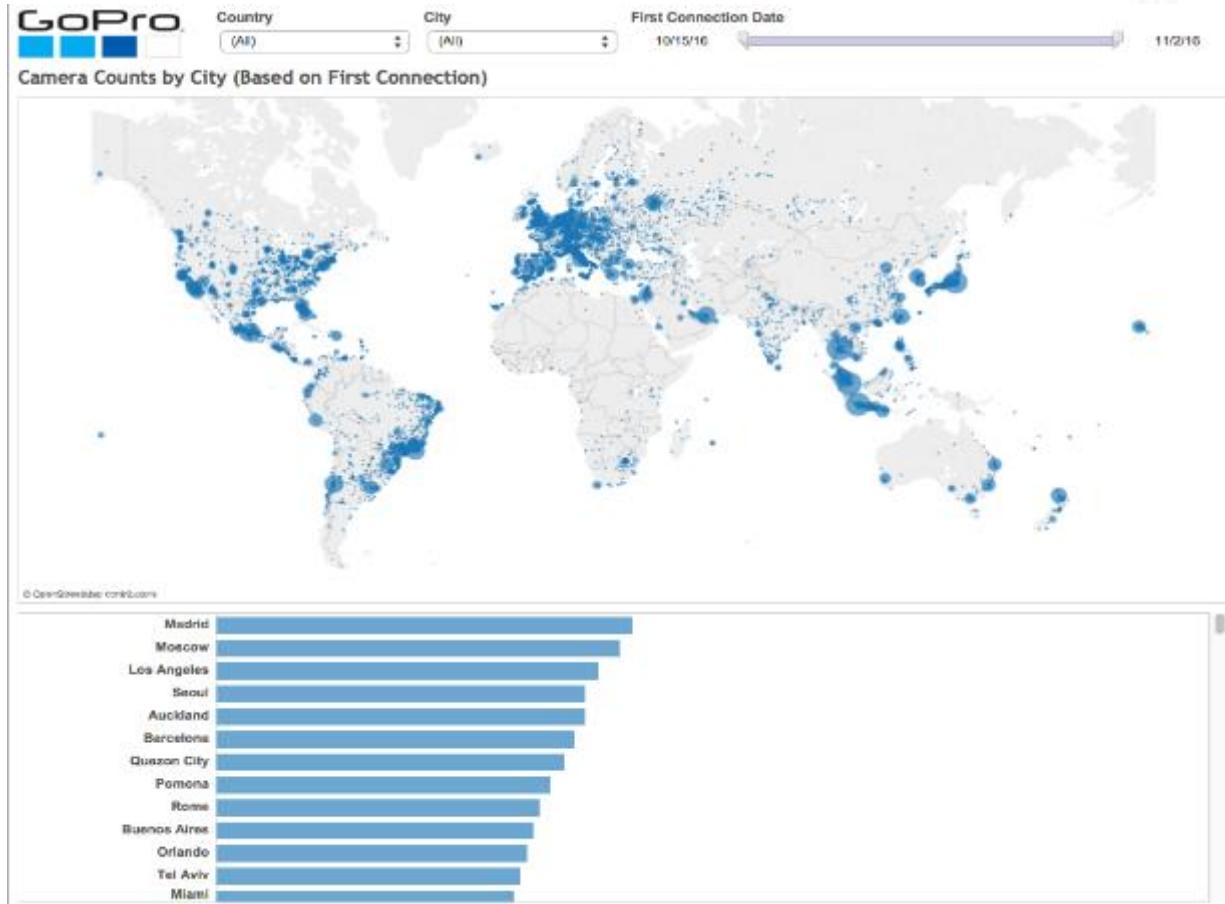
- Product insights can result in product design changes
- How we observed Camera as A Hub upload behaviors



Reporting

CAMERA CONNECTS...

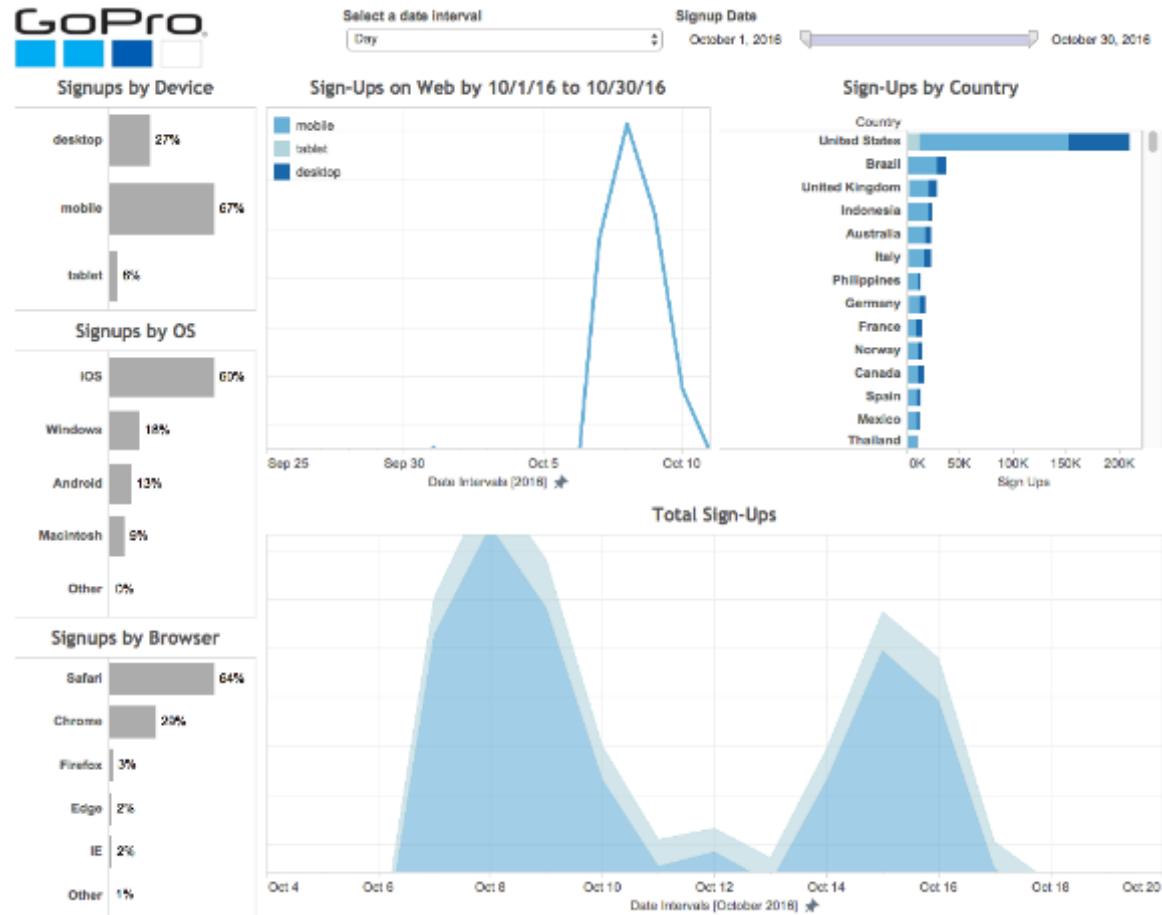
- Help us with figuring out stolen/smuggled cameras
- Where we should put our new marketing dollars for new products



DASHBOARDS (GoPro Plus SUBSCRIPTIONS)



- Real-time subscriber growth
- Utilizes Hive External Tables and JSON SERDE



Pre-ETL

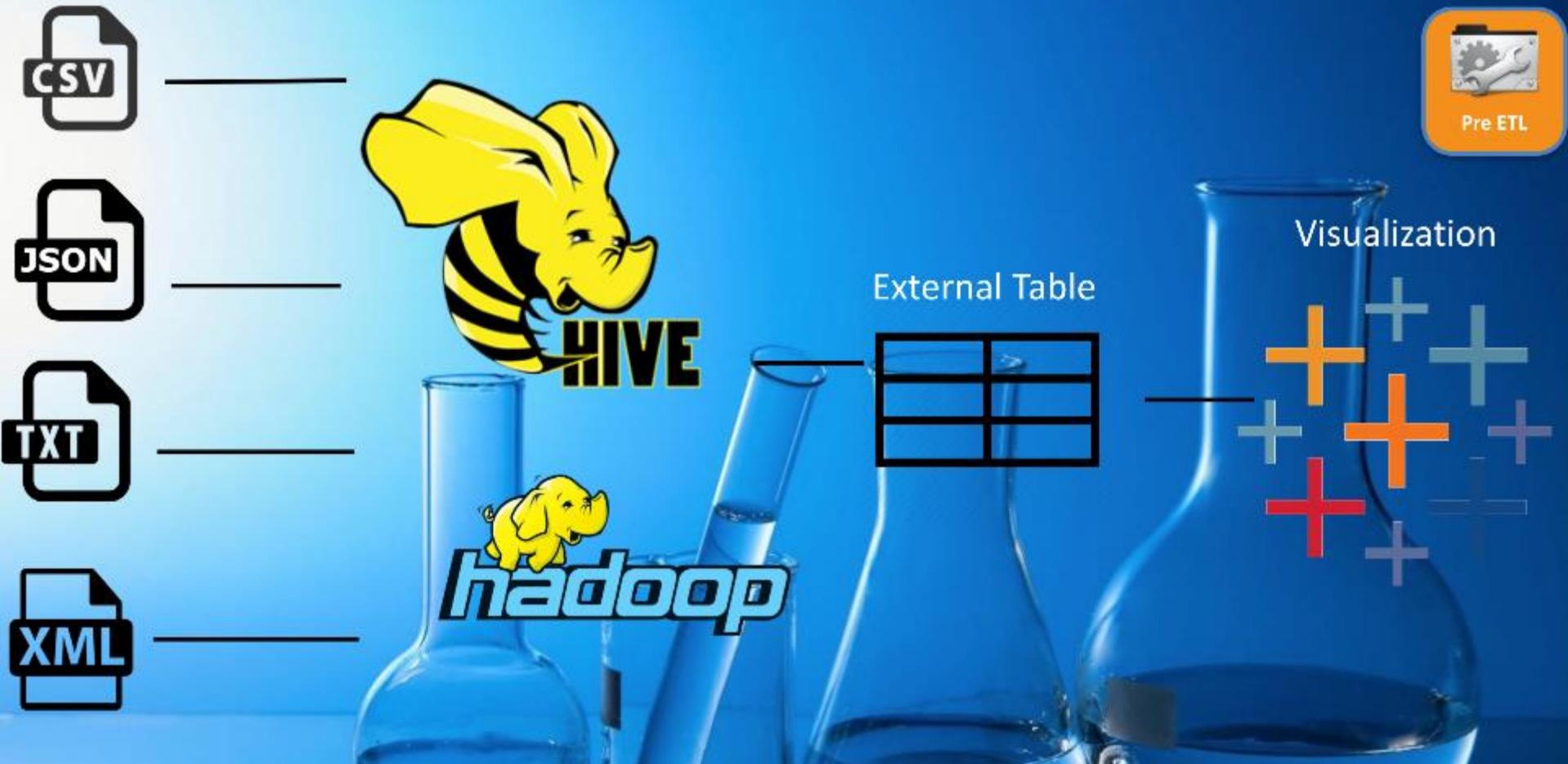


DATA PROFILING

QUALITY ASSURANCE



Pre ETL Process



Quik - Pre-ETL

Demonstration



Questions?

GoPro[®]
Be a *HERO.* 



AWS
re:Invent

Thank you!



**Remember to complete
your evaluations!**