ECX CROSS-MISSION GROUND &
COMMUNICATIONS ENTERPRISE

GSAW 2021

UNITED STATES
SPACE FORCE

Data Architecture & Integration within Space C2

March 2021

Captain David Knieriem
DAV PM, SMC/ECXCB

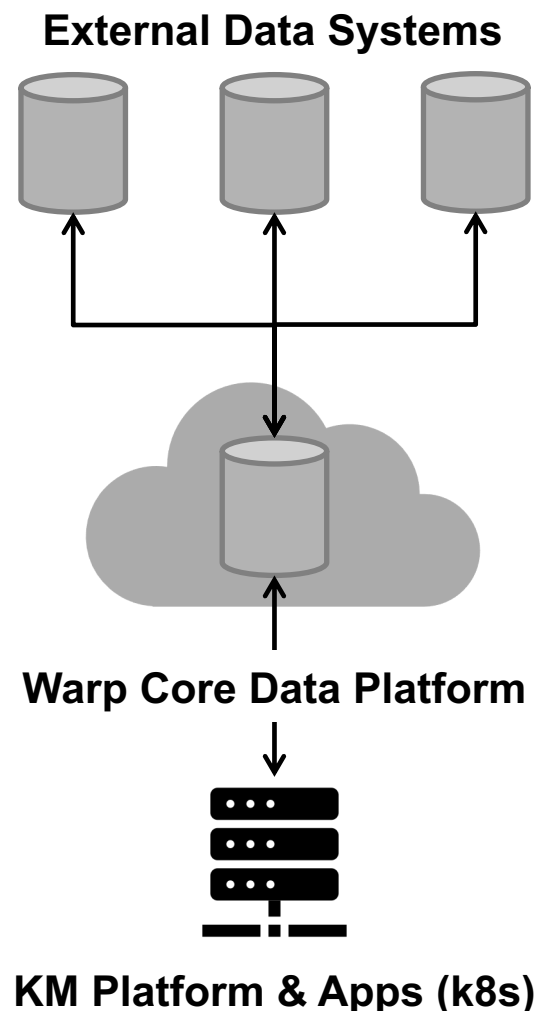**Jake Albrecht**
Data Platform Tech Lead, SMC/ECXCB

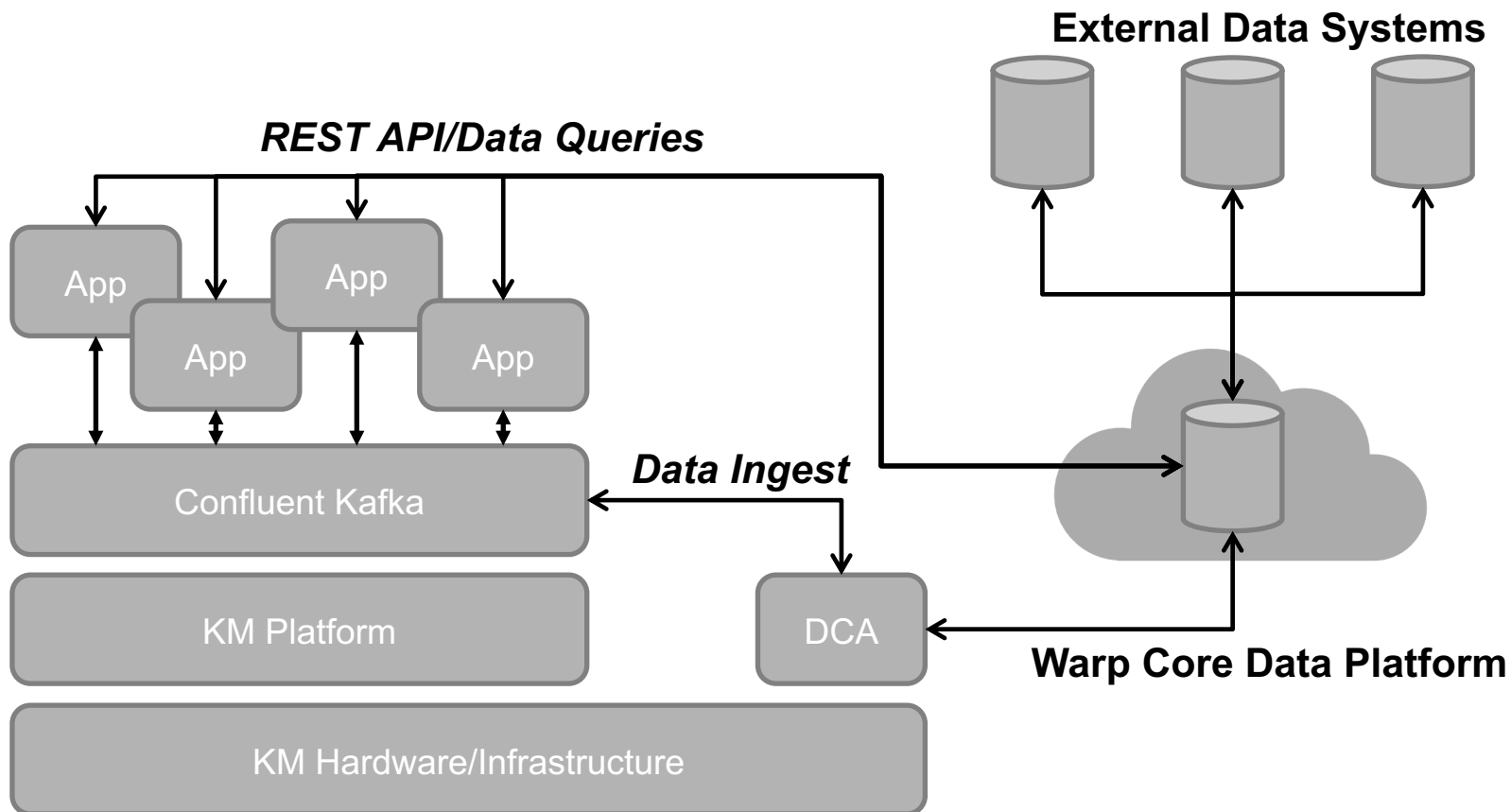# *Overview*

- **Space C2 Kobayashi Maru is responsible for building the Operational C2 layer within the USSF space enterprise**

- **Many apps and capabilities exist within Space C2 with different data needs**

- **These needs are consolidated into a centralized Data Management and Analytics Platform**

- **The Warp Core data integration platform enables a data-centric architecture for the Space C2 program**
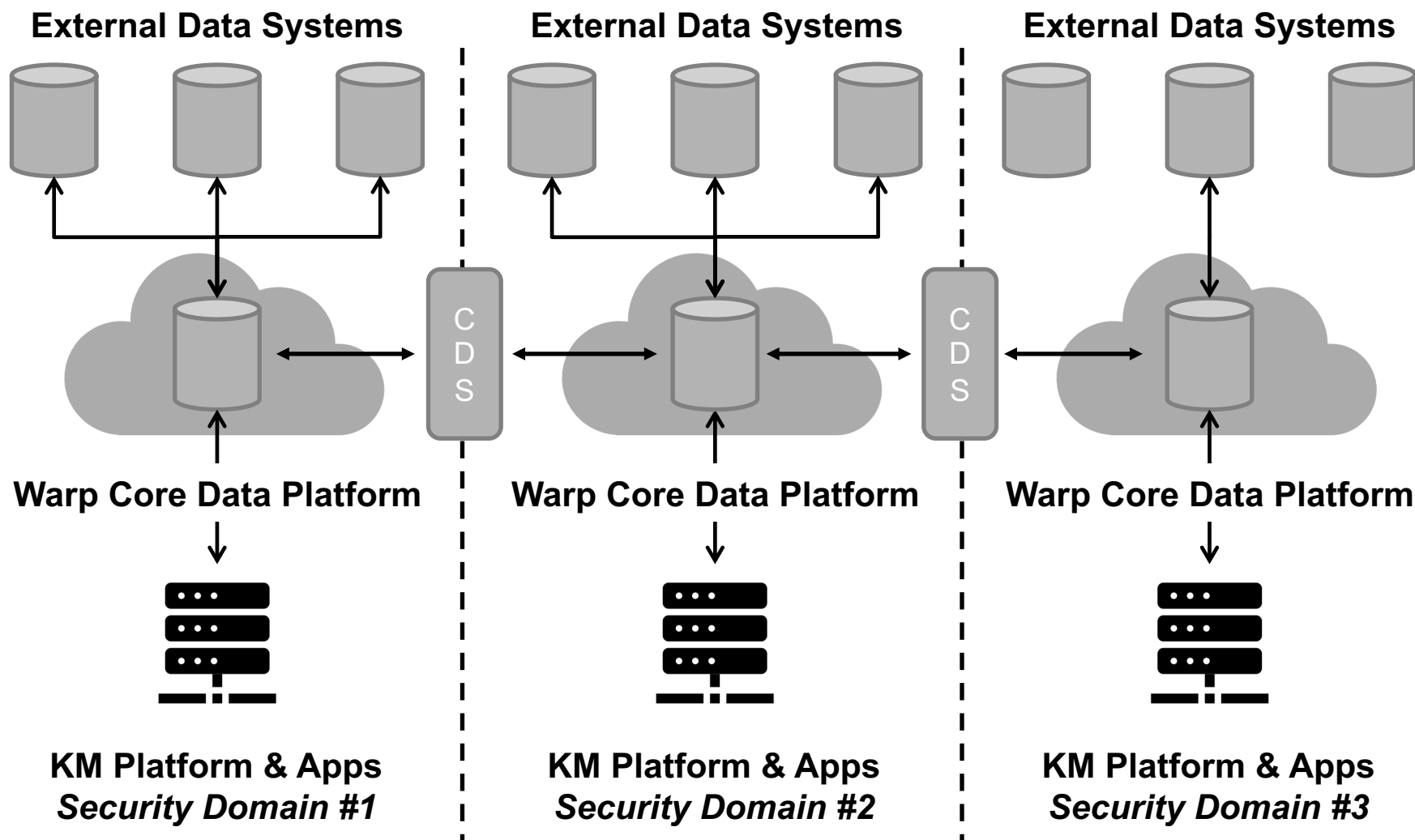
External Data Systems

Warp Core Data Platform

KM Platform & Apps (k8s)

Data Architecture (Single Domain)

External Data Systems

REST API/Data Queries

App

App

App

App

Data Ingest

Confluent Kafka

KM Platform

DCA

Warp Core Data Platform

KM Hardware/Infrastructure

Data Architecture (All Domains)

External Data Systems

External Data Systems

External Data Systems

C D S

C D S

Warp Core Data Platform

Warp Core Data Platform

Warp Core Data Platform

KM Platform & Apps
Security Domain #1

KM Platform & Apps
Security Domain #2

KM Platform & Apps
*Security Domain #3*

Why a Data Platform?

- **Data is valuable, but technology must exist to exploit it**

- **Many techniques exist to try to fully realize the value of data and turn it into *actionable intelligence***
  - **Standards, normalization, metadata, catalogs, etc.**

- **This data must also be *usable* by different people with different roles**
  - **Not everyone is a data scientist!**
  - **Our platform supports everyone from General Officers to software developers**

- **For Space C2, the choice came down to *build vs. buy* for a data platform**
  - **Build: Stand up a team of a few engineers and give them a few months**
  - **Buy: Finding a turnkey, integrated COTS-based solution, tailorable to Space C2 needs**

Warp Core Functional Overview

- **Space C2 Warp Core is an instantiation of the COTS-based Palantir Foundry data platform**

- **Warp Core provides end-to-end data management capabilities for all data within Space C2**

Image: Palantir
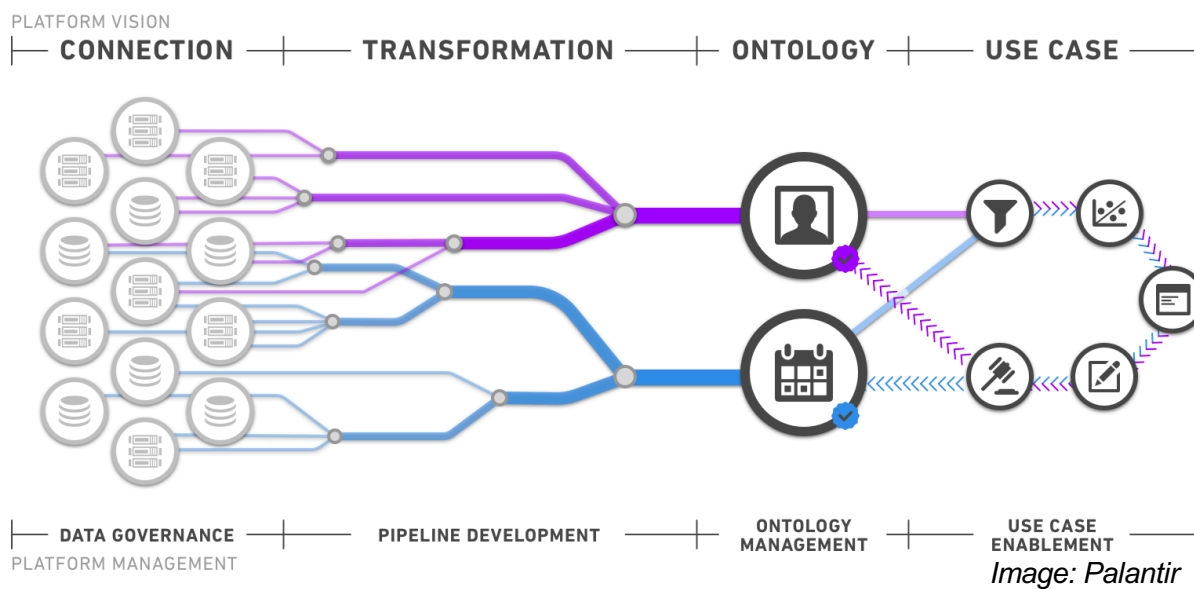
Data Connection Agent (Ingest)

- **Warp Core retrieves data from source systems through a Data Connection Agent (DCA)**

- **The DCA is a lightweight process that initiates a secure connection back to the platform in AWS**

- **DCA retrieves config and commands from Warp Core for retrieving data from the source system**

- **All DCA-to-Warp Core traffic is initiated by DCA and is TLS-encrypted over HTTPS/443**
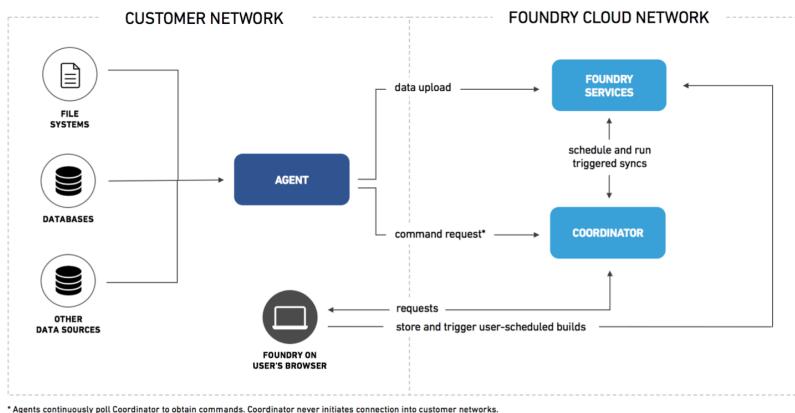

Image: Palantir

Data Storage

- **Data is extracted from the source, then loaded into Warp Core in its original format**

- **Data is stored AWS S3 buckets, either in its raw format or as Apache Parquet (column-based) files for downstream analytics**

- **Warp Core *Catalog Service* maps S3 files (raw or otherwise) into a single, logical *materialized view* of data (we call this a *dataset*)**
  - **Thousands of files with data can map to a single dataset with billions of records**

- **The *dataset* is the fundamental resource within Warp Core, all data transformations, data model object types, analytics, and visualizations operate on datasets**

- **Using S3/Parquet, datasets with low entropy (i.e. most values are relatively the same) encode and compress nicely (e.g. 15GB for 150M records)**


Amazon S3
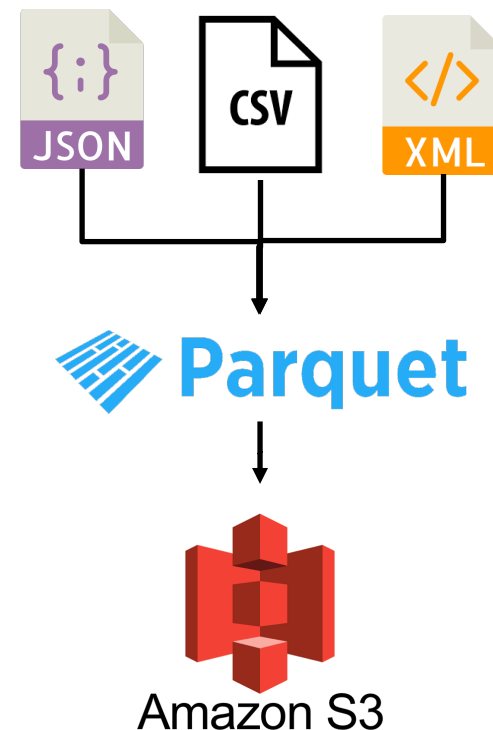
Data Transformation

- **Warp Core applies a *transform* to a dataset to filter, join, aggregate, or perform some arbitrary computation to records within one or more datasets**

- **If a data source is at rest (e.g. a database), transforms are done using Apache Spark**

- **If a data source is in motion (e.g. a message bus), transforms are done using Apache Flink**

- **Transforms are submitted as jobs to Spark/Flink clusters by the *Build Service*, which are executed as resources are available**

- **Compute limitation policies can be applied ensure SLAs**

Parquet ⟶ **Spark** ⟶ **Parquet**

Dataset A   **Data Transform (Your Code)**   **Dataset B**

- **DataOps is the application of DevOps principles to data management**

- **Warp Core *Catalog Service* understands *branching* of data using code branches in Git**

- **Allows safe modification, experimentation, and merging of data into production baseline (master branch)**

- **Collapses data platform footprint to a single Warp Core stack**
   - **No need for separate Dev, Test, Staging stacks**

Master
Test
Dev

Data Branches
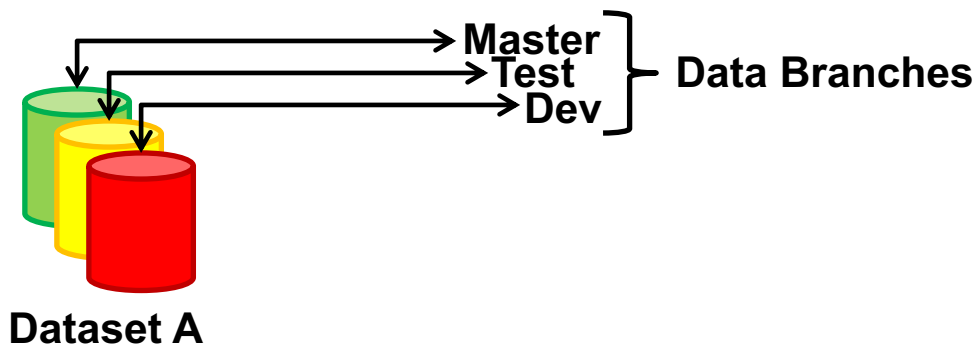
Dataset A

Data Lineage and Provenance

- **We define hundreds of transforms across hundreds of datasets to build *data pipelines***

- **The *Data Lineage* service shows transparent provenance and lineage of data as it's being used within the platform**

- **Also describes build history and metadata for each dataset**

**Each box is a dataset composed of many physical files in AWS S3**
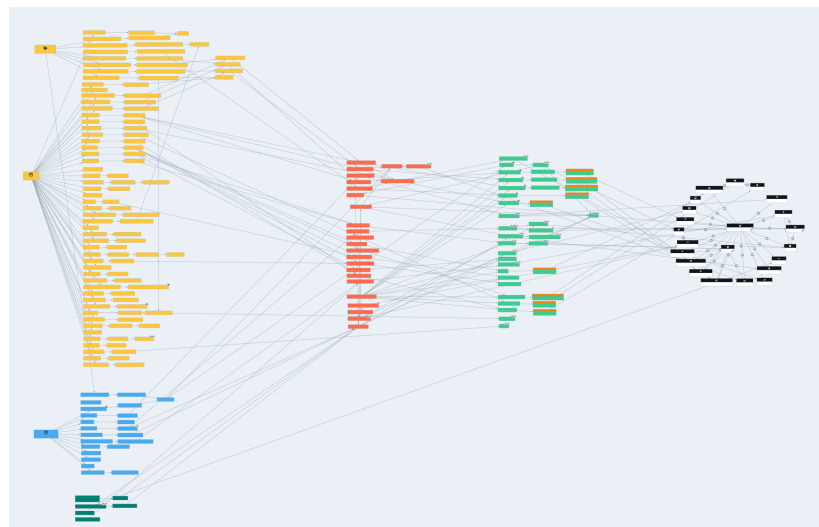
Warp Core Data Lineage Graph

Data Ontology and Object Explorer

- **Datasets are mapped to *object types* using the Data Ontology application**

- **Useful for a natural understanding of the underlying data, *not* just rows and columns in a table**

- **Object types are indexed into an Elasticsearch cluster and searchable using the *Object Explorer* service**

- **Each object type has a *view template*, customized to the user's preference**
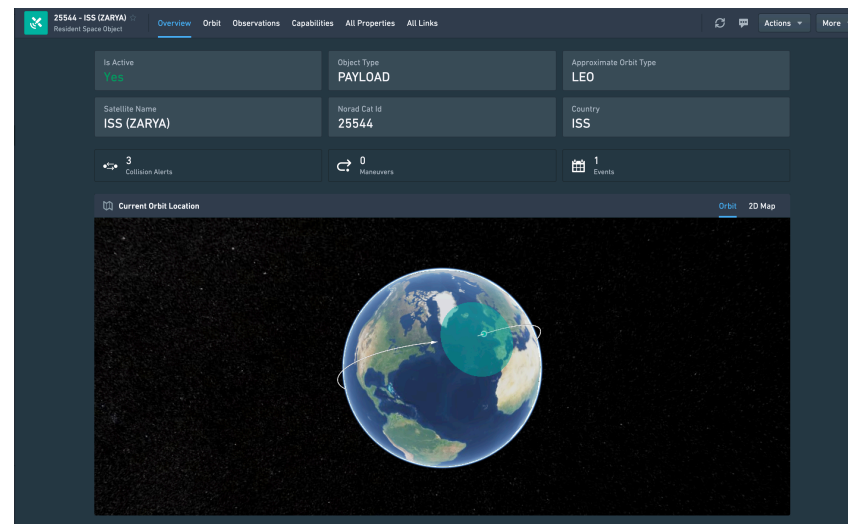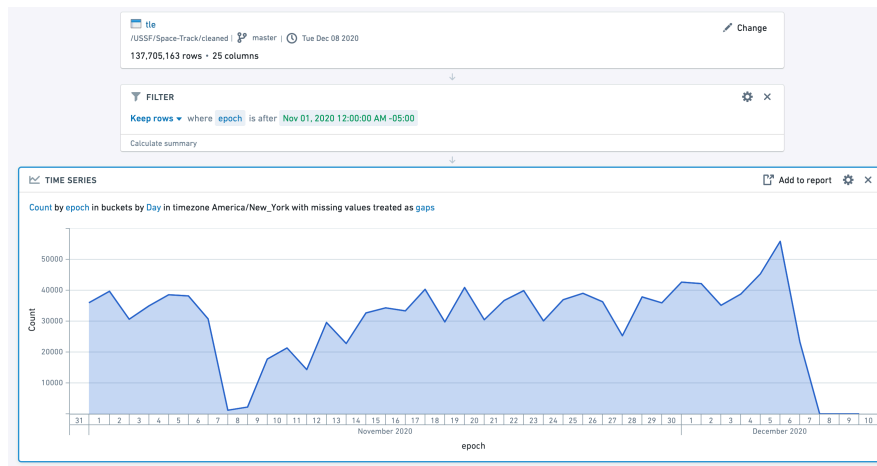

Object View for a Resident Space Object (RSO)

Data Analytics and Visualization

- **Datasets can be analyzed and visualized using many different services**

- **Analytics includes low-code options to quickly join, filter, and aggregate data as ad-hoc Spark jobs**

- **Results can be rendered in a number of visualization widgets, specified by the user**

Time Series Aggregation Example (~140M Records)

Data-Driven Decision Making

- **The Warp Core *Decision Suite* allows users to create complex action-based workflows using real-world data**

- **Granular permissions enforce policy rule sets to control data access and modifications**

- **Actions performed using Decision Suite can automatically propagate to external systems for live notifications and events**


Example Decision-Making App

Programmatic Access

- **Warp Core is intended to serve as the centralized access point for data within Space C2**

- **All services within Warp Core are developed as REST-based services, including data query services**

- **Data can be queried using industry standard formats (JSON, GraphQL, etc.)**

- **API and data access is authorized based on the users underlying credentials (token or OAuth-based)**
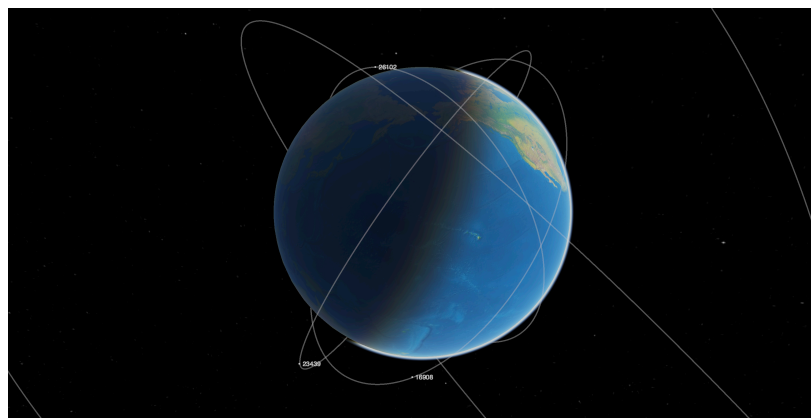
External App Rendering TLEs using Warp Core APIs

- **Space C2 Warp Core is deployed within AWS in multiple security domains to serve apps and users within Space C2**

- **Warp Core is designed for scalability and flexibility using standard APIs and open-source technologies**

- **Services allow users to exploit data regardless of skill set**

- **Warp Core uses modern techniques to continuously deploy new capabilities to users with no downtime**

- **Operational at multiple security levels with connections to legacy space systems**

Jake Albrecht
Data Platform Tech Lead, SMC/ECXCB
jacob.l.albrecht@aero.org

ECX CROSS-MISSION GROUND &
COMMUNICATIONS ENTERPRISE