Department of the Air Force

Integrity - Service - Excellence

Data Analytics with VAULT Tech Stack

ICEAA May 2023

Disclaimer

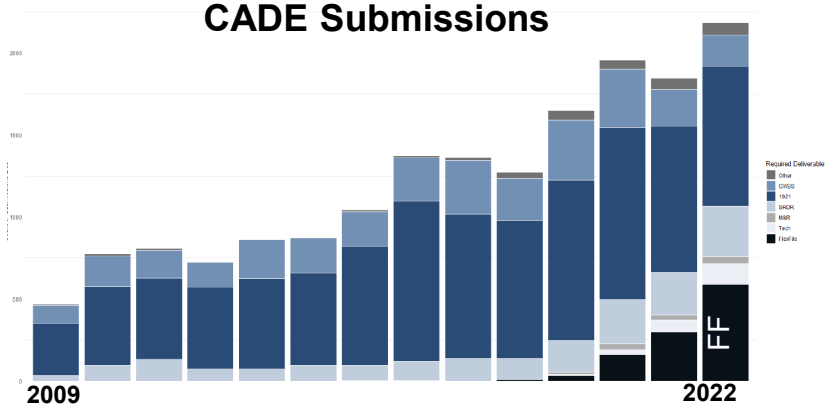
- **Any reference to contractors, companies, or any other non-government entity in this briefing is for informational purposes only and is not to be construed as an endorsement**

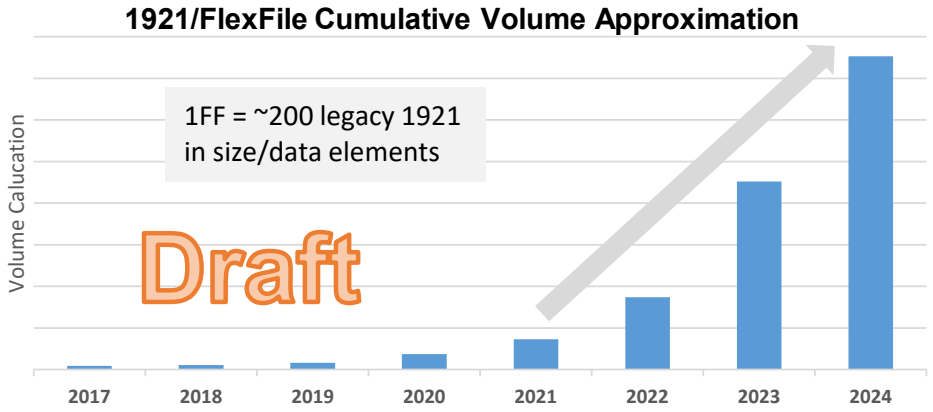
Overview

- **Largely the cost community operates on desktop-based models and toolsets today- AFCAA is undergoing a digital transformation as part of the SAF/FM Strategic plan**
- **AFCAA is utilizing state-of-the-art toolsets on VAULT platform in order to build data pipelines as well as analytic products such as models/dashboards/machine learning/ etc.**

Evolution of Cost Analysts' Data

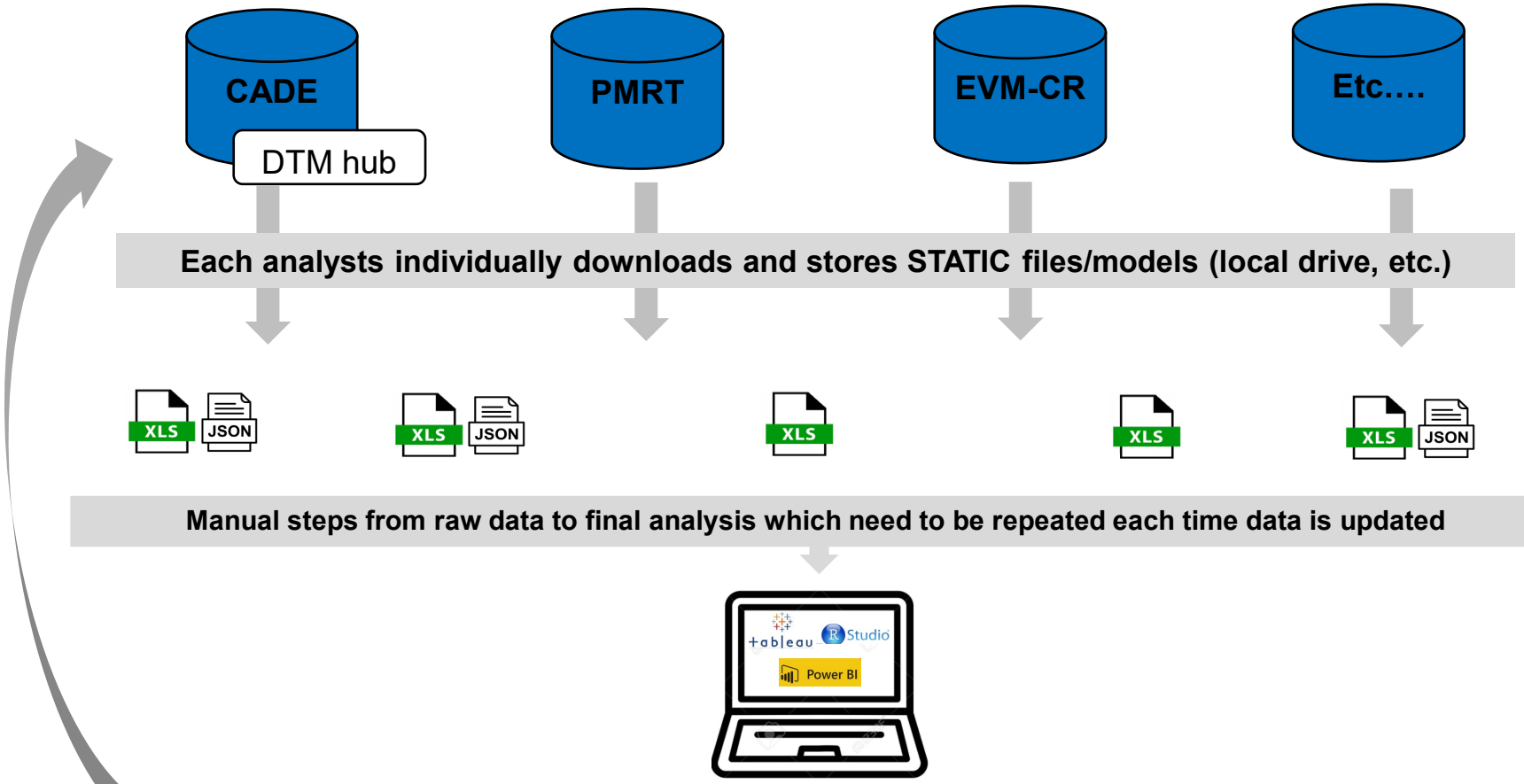
- **Cost community has significantly more data now than ever before - projected to continue to grow**
  - New reporting captures technical/ software data as well as cost
  - FlexFiles are over ~200X size of legacy cost report
  - Statutory requirement change
    - Legacy ACAT I only
    - Current >100M programs
  - Size of additional datasets (EVM, KDB, AFTOC, etc. also growing)
- **Traditional process/toolsets becoming challenging & increasingly not viable**





Last 5 years were dedicated to improving data collection
Next 5 years will be focused on evolving our processes/products/toolsets

Legacy


CADE
DTM hub
PMRT
EVM-CR
Etc….
Each analysts individually downloads and stores STATIC files/models (local drive, etc.)
XLS
JSON
XLS
JSON
XLS
XLS
XLS
JSON
Manual steps from raw data to final analysis which need to be repeated each time data is updated
tableau
Studio
Power BI
(Multiply by frequency to keep data current & Multiply by # of analysts)

Data Operating Model for Cost End State

SMEs/Analysts*

Data Scientists

Analysis

Custom Applications

Dashboards

Curated Data (SME): Analysts' specific transformations/mappings/etc.

Governance Layer

Curated Data: Cleaned/Transformed/Mapped/Normalized- consistent logic

Raw Data (Databricks)

Feedback loop to incorporate consistent logic that is reusable by other analysts

Draft VAULT Architecture

Tableau

Dataiku

Governance Layer

Databricks

IMMUTA

Data Lake (S3 storage)

KDB

SAR

FPRA

EVM-CR

Budget

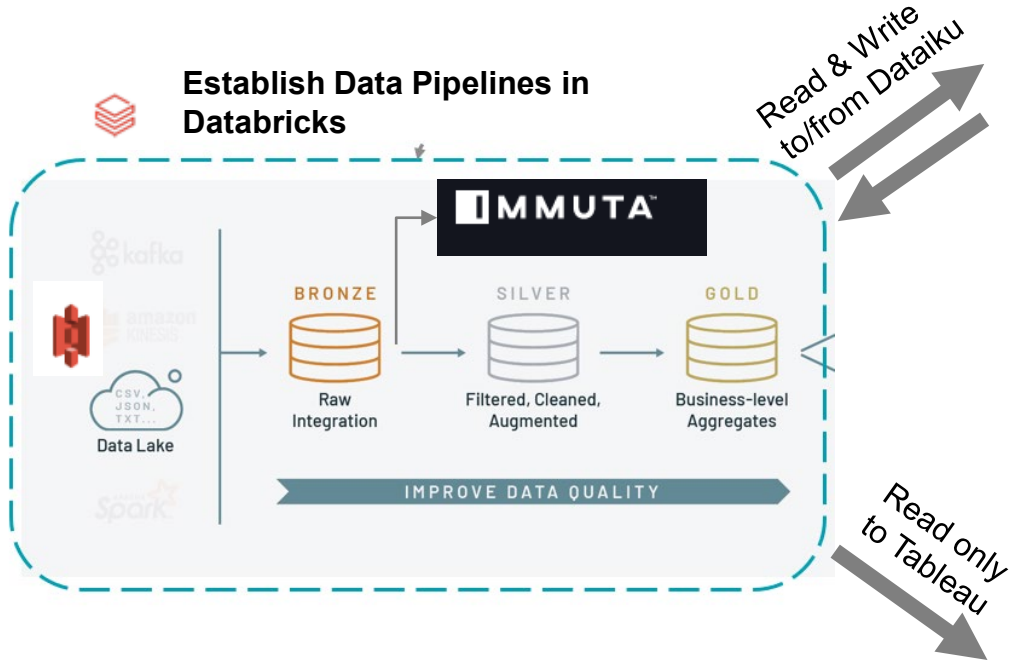
CADE

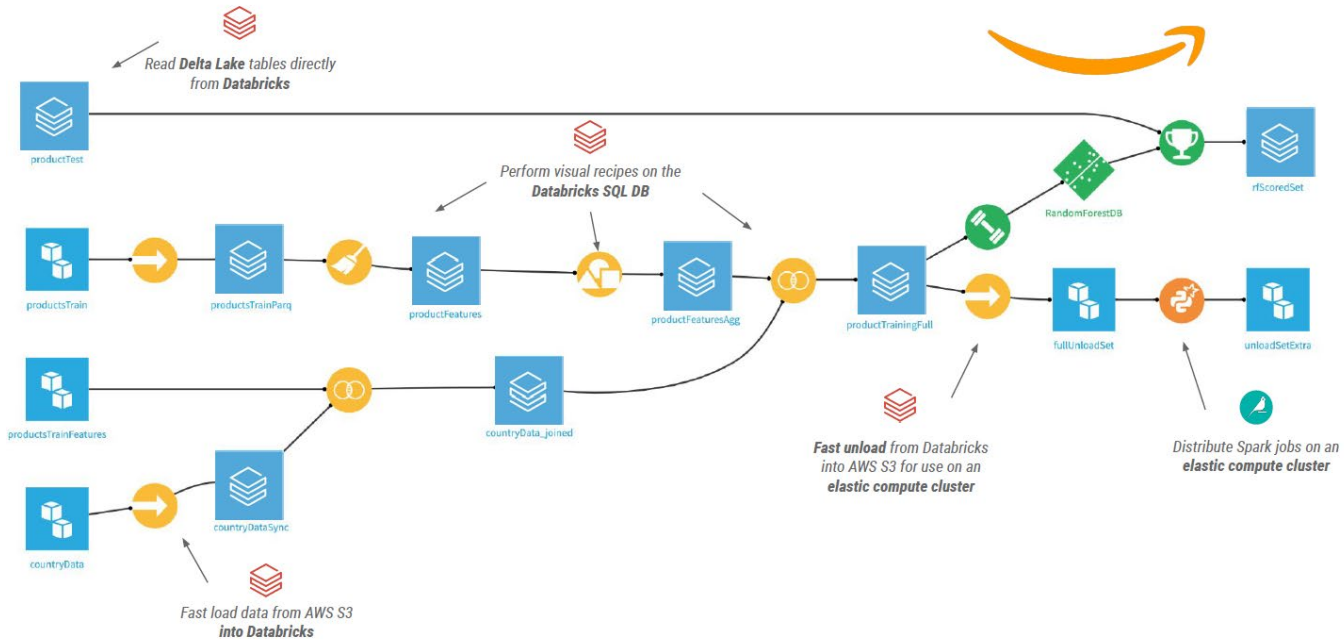
AFTOC

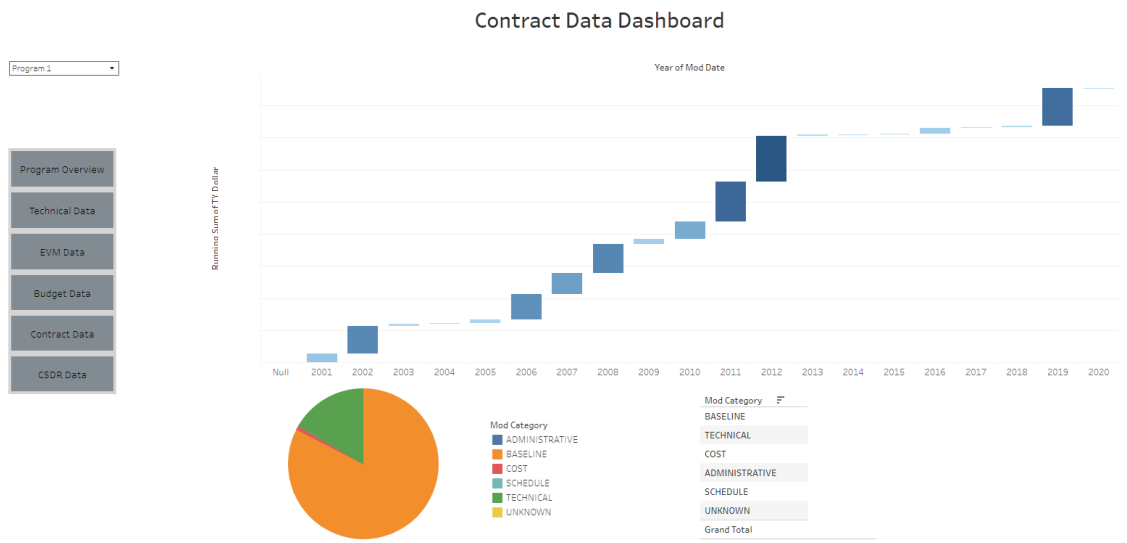
Other Sources

*Data Scientists also potentially involved with the ***development*** of the dashboards and custom applications

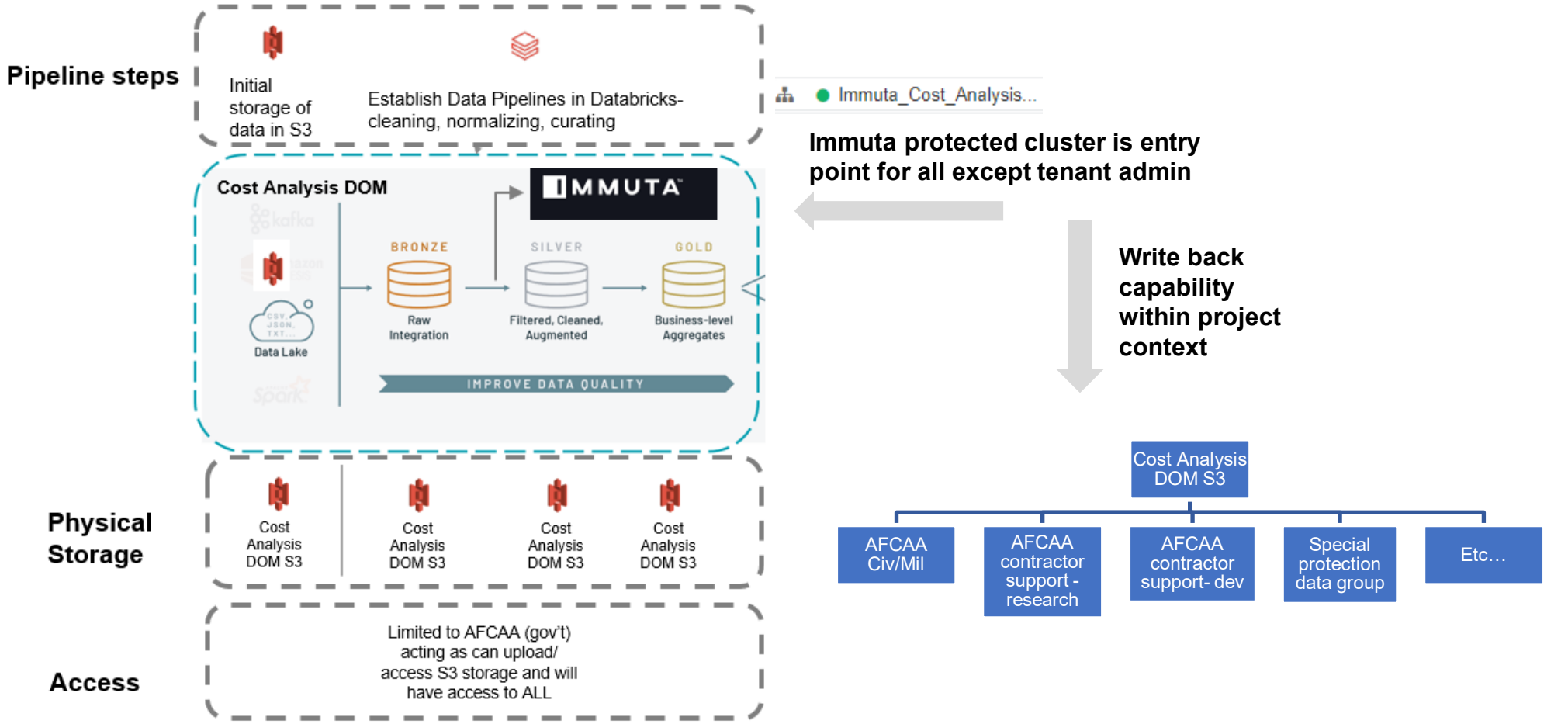
Databricks-Dataiku-Tableau Tech stack





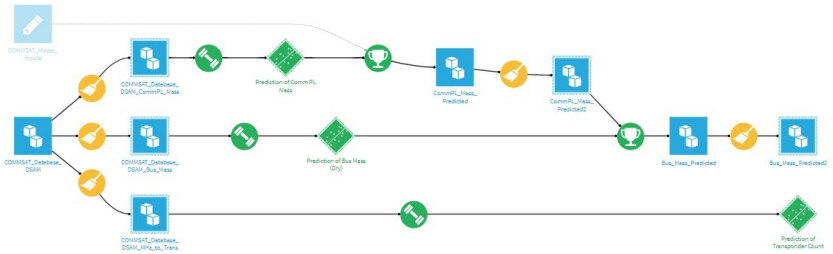


Architecture with Immuta



Live Demo of Data Science in VAULT

Dataiku



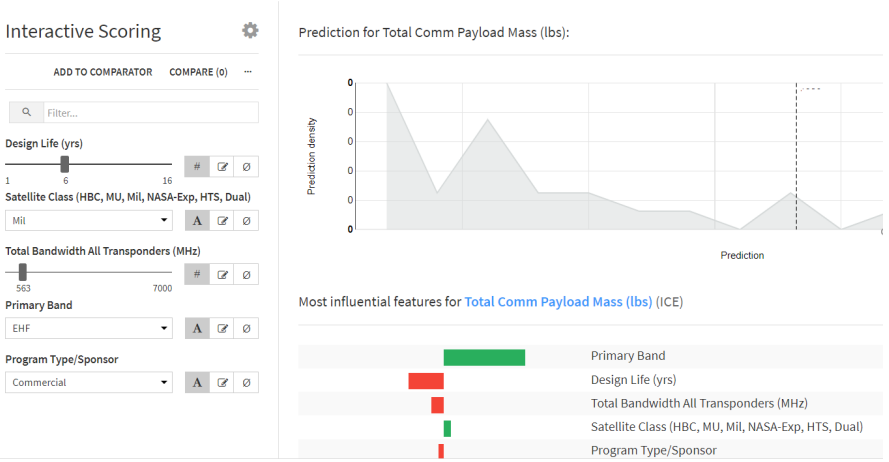
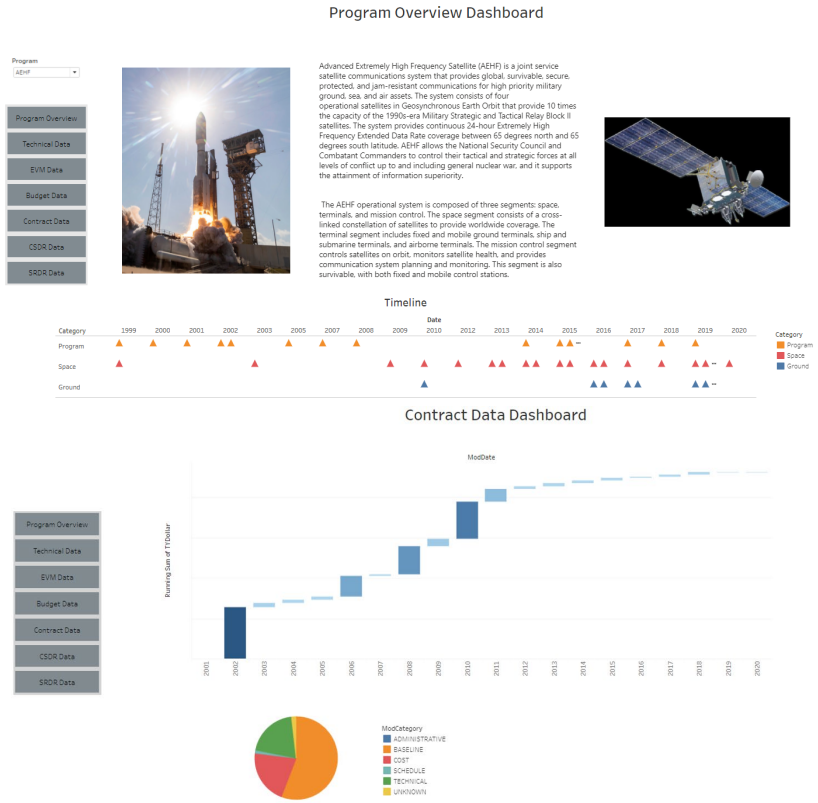


Tableau Server



IMMUTA

Databricks

Practical Steps

- **What can you do to embrace data science processes, toolsets, etc. as a part of your day-to-day cost analysis?**
  - **What can you do as an ORGANIZATION?**
  - **What can you do as a COMPANY?**
  - **What can you do as an INDIVIDUAL?**