TRIFACTA
The Data Engineering Cloud

# Enabling Shared Data Services at Scale for the Public Sector

## Background

Driven by public policy (the DATA Act, Cloud Smart, National Defense Strategy, etc.), many public sector agencies in the United States are actively working to build and deliver shared service programs focused on leveraging data as a strategic asset. A key part of these implementations has been a movement to more agile DataOps operating models, which empower users across the organization with "self-service" access to data. This shift to self-service data analytics has helped to reduce operational barriers, eliminate bottlenecks in data pipelines, and expedite data-driven modernization.

As more agencies seek to build their own shared data platforms, they can accelerate their progress and "skip ahead to success" by employing the architectural and operational models of successful platforms in the public sector that have achieved similar results.

## The Path to Success: Building a Self-Service Architecture

Traditionally, data operations in the public sector have been led by IT teams. This approach is time-consuming and cumbersome, as operations are dependent on technical resources to prepare data pipelines for analysis and decision making.

There has been a shift for the better in recent times, with public-sector agencies in the U.S. beginning to leverage cloud-based solutions for self-service data analytics. This is primarily driven by the scale, flexibility, and operational efficiencies of the cloud. Using cloud-native architectures, many shared data platforms have found success by empowering business and technical users with specialized tools for data preparation, cataloging, governance, and analysis. These tools allow users of various technical skill levels to operationalize data or AI for their own specific mission needs.

## ADVANA: A Shared Data Services Blueprint

A great example of an evolving shared data services platform is the Office of the Secretary of Defense's ADVANA (Advanced Analytics) platform. This program was initially established to handle the Department of Defense's congressionally mandated financial audit. Since its inception, the platform has become a successful shared service powering data analysis for a wide range of use cases for the Department of Defense (DoD) from financial reporting to force readiness.

ADVANA's success wasn't fully realized right out of the gate. A big shift came when self-service tools were implemented that provided enhanced capabilities in the areas of data preparation, visualization, and reporting. These self-service tools eliminated the need to write code or use highly technical tools to work with data, enabling "less technical" users to operationalize data assets for their own mission needs and significantly expanding ADVANA's target user base beyond technical data professionals.

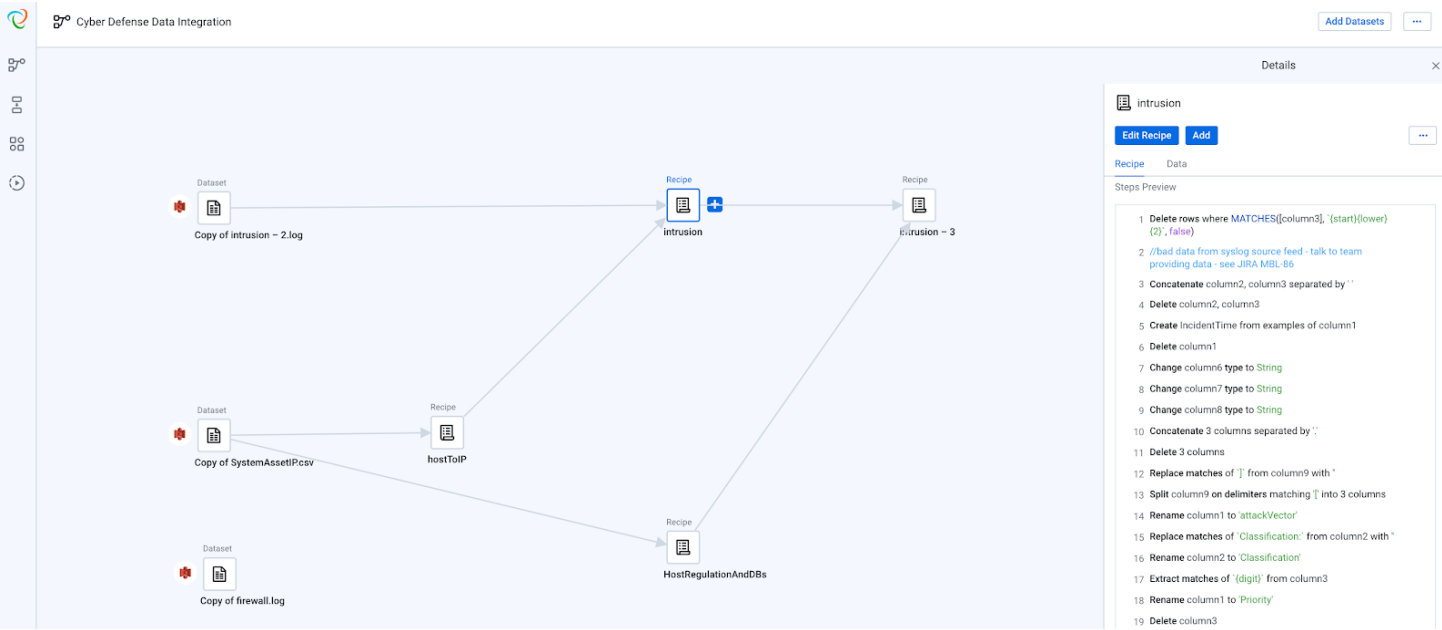
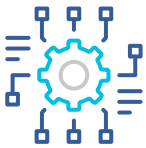



Figure 1. Operationalize data workflows with ease.

As non-technical business users utilized the platform at an increasing pace, ADVANA experienced a significant spike in demand, causing them to outgrow their on-premises data center and move to AWS' Govcloud. Today, ADVANA has thousands of users from dozens of programs leveraging the shared environment on the cloud for their specific mission needs.

## Following in the Footsteps of Success: Keys to a Successful Shared Data Services Program



As shared-service data and analytics programs are launched across various Federal and State departments and agencies, including the US Army, US Air Force, US Transportation Command, Department of Homeland Security, Centers for Disease Control and Prevention, the US Agency for International Development, and the State Department, ADVANA's story serves as a key proof-of-concept and a blueprint for success. The most successful shared services programs have maintained an emphasis on enabling users through self-service.

For those wishing to follow in ADVANA's footsteps, success is best achieved by implementing scalable, cloud-based solutions that can deliver mission impact to a broad and diverse user base. Users of all technical skill levels need to be enabled to explore, transform, and process available data so that they can operationalize it effectively for their own unique needs without relying on IT. This approach eliminates bottlenecks and enables the agility that is needed to achieve rapid mission success.

Self-service, of course, does not happen by accident. Instead, data democratization is the result of a series of intentional decisions and the intentional adoption of the advanced technology that makes it possible. Public sector agencies building shared services platforms, especially those that are still in early stages, can save a significant amount of time and resources by considering the need to build a self-service architecture from the beginning of their shared services project. This process can be accelerated by looking to successful platforms like ADVANA for a blueprint of technologies and vendors to choose from.

# Trifacta's Role in Shared Data Services and Agile DataOps

As the public sector moves towards Shared Data Services and more agile data operations, one industry-leading tool that plays a key role in enabling self-service is Trifacta.

Trifacta provides an open, secure, governed, and collaborative environment to operationalize data for both low-code and no-code requirements. This democratization allows relevant data professionals to focus on higher value work by offloading and automating time-consuming data prep work.

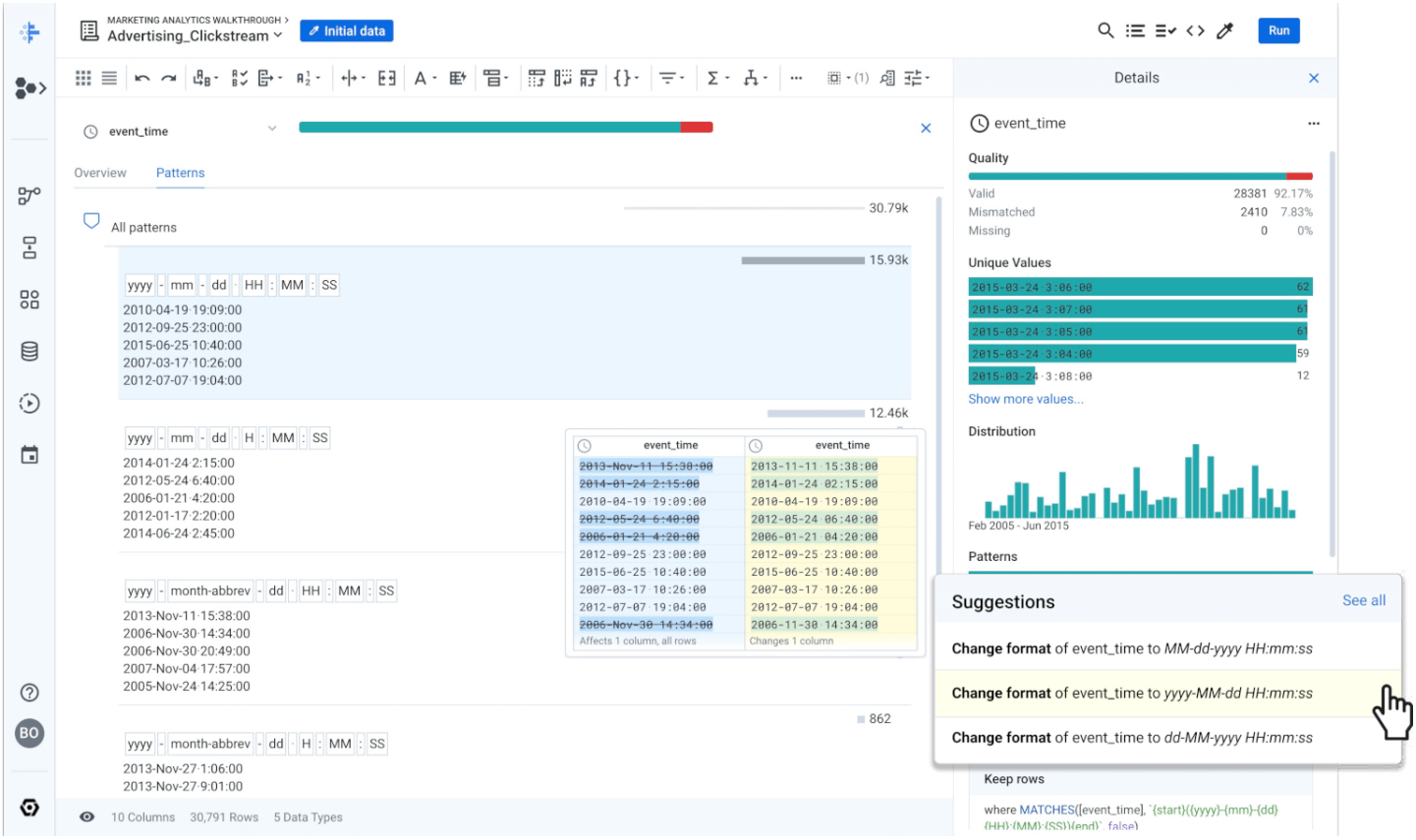



Figure 2. Trifacta leverages machine learning to make understanding and resolving data transformation challenges intuitive to users of all backgrounds.

*"It's impossible to overstress this: 80% of the work in any data project is in cleaning the data."*

- DJ Patil Former Chief Data Scientist of the United States

Trifacta works by providing users with an interactive platform to profile, prepare, and pipeline their data. Using Trifacta's visual user interface, users of any technical skill level can easily transform their data with clicks, not code, backed by intuitive product features such as visual data profiling, transformation by example, and predictive transformation suggestions.

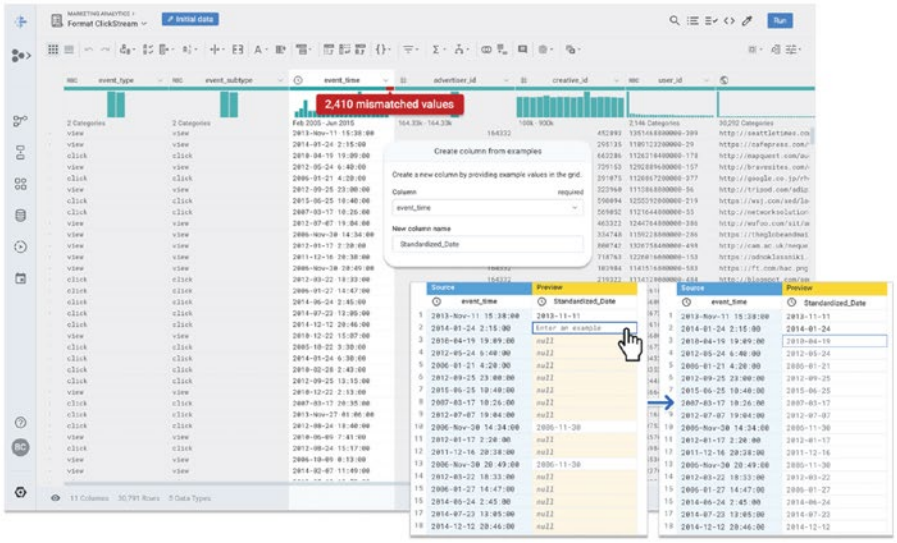



Figure 3. Trifacta's transform by example feature makes it easy for users of any technical skill level to standarize their data into a desired format.

Once data is prepared, Trifacta makes it easy to schedule and automate data pipelines. The end result is democratized data engineering and a drastic reduction in the time required to accomplish data preparation tasks.

Trifacta is a cloud-native, cloud-agnostic tool that provides the accessibility, security and scale required for enterprise data analytics as identified in the Federal Data Strategy Action Plan. For many Shared Data Services initiatives, including ADVANA and the US Air Force's VAULT data platform, Trifacta's data engineering cloud is the data preparation platform of choice.

To learn more about how Trifacta can help enable your agency's Shared Data Services success, schedule a demo today.

## Sources

https://comptroller.defense.gov/Portals/45/Documents/defbudget/FY2017/budget_justification/pdfs/2017MarchAmended/03_RDT_and_E/OSD_FY17_RDTE_ABS_20170314.pdf

https://strategy.data.gov/action-plan/ https://cloud.cio.gov/

https://en.wikipedia.org/wiki/Digital_Accountability_and_Transparency_Act_of_2014