

GOVERNMENT

USDA Data Modernization and Confluent

The Data Challenge

Data is the by-product of every digital action taken by USDA mission areas, partners and by the public communities that it supports. Across all mission areas from food safety to rural development or farm production, every system, process and sensor generates data. Information technology makes it easier for organizations to collect, store and share data. Too often though, data doesn't get to the people at the "point of contact" who need it to make better decisions and create more tailored experiences for end users.

Most Government organizations possess a complex landscape of data stores and point to point, often batch, integrations. While this has served reasonably well in the past, it tends to be expensive and often inefficient. It is also challenging to field new capabilities and the limitations of data sitting at rest often result in inconsistencies in the user or customer experience. To provide richer, more complete and satisfying "customer" experiences, organizations such as USDA must adapt legacy data stores to bring data to the point of action.

Data-in-motion

Data-in-motion has emerged as a new paradigm to achieve this objective, enabling data to flow across the enterprise as a continuous stream of events, making for highly personalized interactions and best possible decisions and actions.

Real-time Insights with Confluent

Customer Staging Area

Load customer data into dedicated repositories and stream using change data capture with **Connect** or connect directly to a customer database

Ingest & Process

Capture changes with a consistent data structure using **Schema Registry**, develop real-time ETL pipelines using **ksqldb** and **Streams**; and stream data to downstream services and storage using **Connect**

Store

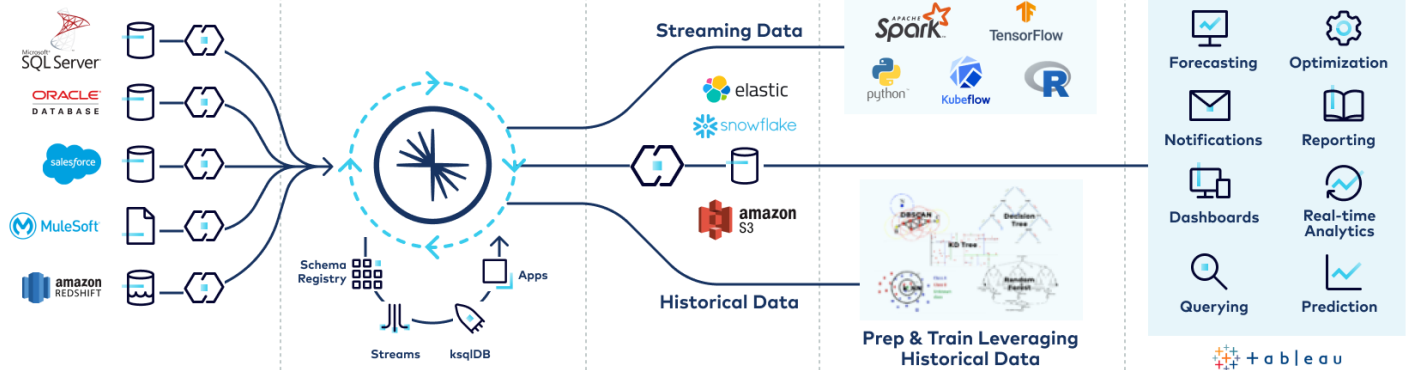
Keep your events in a storage service using **Tiered Storage** and query engine

Prep, Train & Deploy

Analytics and ML
Leverage modern ML toolkits for model prep, training, and deployment; gain deeper insights with real-time inference

Customer Insights

Gain real-time insights into streaming data; visualize information and analytics using intuitive dashboards; generate customer specific reports



USDA Strategy

As part of the USDA Data Strategy, there is a stated plan to “develop and leverage technology, infrastructure, and analytics tools to enable shared access and use of data to achieve our mission and drive innovation.” Each Mission Area, Program and the Enterprise needs the ability to access and share data to produce information products at their organization level. Ongoing modernization efforts to shift workloads to the cloud can help but integrating on-prem and cloud-based systems so that data flows in a timely manner across the hybrid IT landscape requires unique software architecture and capabilities. This approach also needs to work with existing data silos while making it possible to rationalize and retire them in the future as appropriate.

A data in motion platform will enable the USDA to create data pipelines that stream data as it is created to the apps, systems and users that need it most. This ensures a consistent view of inventories, eligibility, balances, cyber threats, etc. and ensures that systems are in sync, providing a consistent experience for users and customers across channels and enabling action as events occur.

Introducing Confluent

Confluent Platform was designed specifically to meet the requirements of complex enterprises like the USDA's. It is the leading data-in-motion software, from the creators of Apache Kafka® in response to the explosive growth and adoption of Kafka, a highly-scalable, event-driven open source software project capable of processing trillions of events per day.

With Confluent, USDA can deploy across on-prem and cloud environments so the different Mission Areas can continue to manage their data as they need. It creates a flexible and future-ready data infrastructure by decoupling producers and consumers, simplifying how data can be combined in new ways to bring new applications online or retire end-of-life data sources. As DISC/OCIO work to stand up new capabilities, Confluent easily integrates new data sources and existing systems without impacting existing data workflows. Once data is moving through Confluent, future modernization efforts become more efficient and much less costly.

Why Confluent

Confluent completes Apache Kafka by adding a number of essential Enterprise-grade capabilities:

Enterprise Security

Role Based Access Control, Encryption, structured audit logs and integration with directory services

Resilience

Cluster linking and multi-region cluster data replication to minimize unplanned downtime and avoid data loss

Enterprise Manageability

Streamline operations with performance tuning, Kubernetes Operator and Ansible Playbooks for ease of deployment, and management/governance interface

Data Connectors

Out of the box for 160+ data sources and sinks to speed development and limit maintenance efforts, e.g. Oracle, Splunk, Snowflake, S3, Databricks, Salesforce etc.

Developer Productivity

Multi language support and schema registry to ensure compatibility across schema and application versions

Enterprise Support

Best in class 24/7 end-to-end expert technical support, with emergency patching for both Confluent & Apache Kafka

Private companies such as Netflix, Tivo, FedEx, and Epic Games rely on Confluent to run their business. Federal Government customers relying on Confluent for streaming data management, data sharing and advanced analytics include IRS, USPS, VA, CMS, DISA, US Army and many other agencies and services.

Ready to get started? Contact a Confluent expert today

Email us at publicsector@confluent.io | Or visit www.confluent.io/government/ for more details