

A Modern Data Strategy for the U.S. Department of Defense

DataOS, the right data architecture for today and tomorrow

DataOS Data Fabric maximizes the value of data and accelerates digital transformations. DataOS enables enterprises to ingest, process, transform, govern, and orchestrate data from disparate data sources to deliver a trusted and real-time view of customer and business data. Our DataOS Data Fabric solution embodies seven core principles.

Introduction

In October of 2020, the DOD published a new Data Strategy targeted at using “data at speed and scale for operational advantage and increased efficiency.” Like every other industry, government agencies are faced with the question: “Regardless of how much data you have, is your data actually usable?” From The Modern Data Company’s perspective, data usability means - **Data Quality** - Providing high-quality data for analysis, **Data Access** - Everyone can access data that they want, when they want with the right permissions, **Data Meaning** - Understanding the semantics of data and context of every data element.

Data has no value unless it is usable and accessible. By converting data into information, we add meaning and context to data, an ability to access this data securely and finally ensure high quality. We convert data (tables, columns and values) into a usable product that is treated as an asset. By managing data as a powerful network of information, enterprises can move away from point-to-point integrations via ETL (extract, transform, load) and APIs and pivot towards a more modern way of powering all their data needs from one place - ultimately, creating a Data Fabric weaving together all of their data endpoints. This provides enterprises the agility to try any data applications, ML models or 3rd party data products without the need for extensive integrations. All of the investments that agencies are placing in great data products (like Tableau, Power BI, Adobe, Snowflake, and others) can now yield the ROI they expect and need, by powering them with trusted high-quality data.

The Need for a Data fabric Solution

Traditional Data Solutions Fall Short

Once data has been converted into a powerful information network, agencies need a technology that allows them to put this data to use -- whether it is to provide business intelligence reports or applying machine learning to predict the future or powering operational systems with high-quality data. Due to rigidity and limited scalability, traditional data platforms with large price tags have left organizations with large amounts of technical debt; Technical debt in thousands of unmaintainable ETL jobs, tables, and reports that only a small group of specialized people can understand.

At The Modern Data Company, we developed a data operating system to remove complexity and future proof your data ecosystem by unifying data management under one roof. We call it DataOS® - a fully integrated Data Fabric.

DataOS, the right data architecture for today and tomorrow

DataOS Data Fabric maximizes the value of data and accelerates digital transformations. DataOS enables enterprises to ingest, process, transform, govern, and orchestrate data from disparate data sources to deliver a trusted and real-time view of customer and business data. *Our DataOS Data Fabric solution embodies seven core principles:*

Core Principle #1: **Data as an Asset**

We treat a data set the same way that an eCommerce business treats products and services. Each data set has a description, schema definition, profile of the data, quality index, tags, other datasets that are similar, top users and top queries on the data set, etc. By productizing data and attaching the semantics to the data, we can then use our technology to move this data to any user or system in the formats that they need.

Core Principle #2: **Data Quality and Profiling**

The key to extracting value from data lies in the ability to understand the quality of the data and ways to turn data into high-quality information. We do this by profiling all datasets that are ingested and provide detailed analysis on quality attributes like cardinality, completeness, missing values, uniqueness, etc. We also run business-specific data validation rules to help objectively quantify data quality.

Core Principle #3: **Data Availability and Discovery**

ccess to data, application data silos and fragmented data across the enterprise are some of the key reasons why data usability is as low as it is today. The first step to doing this is being able to catalog all things data (data sets, jobs, metrics and KPI's) and make them available with a Google-like search interface. This kind of access to data and the jobs that manipulate the data provide instant access and understanding of the data and empowers users to data-surf vs data-drilldowns that they are used to today.

Enterprises are spending heavily on a range of data products and bolting them together to get the resemblance of a Data Fabric. These data products require significant integration work and customization resulting in a lack of flexibility.

DataOS humanizes data and its access, breaks data silos and transforms companies as they take steps towards data democracy and gaining business insights in real-time.

Core Principle #4: **Data Interoperability**

The ability to support ingesting data in **any** format (from mainframes to streaming systems to IoT/5G, structured, unstructured to semi-structured, batch to real time to one-time loads), and being able to make that data available in a secure and compliant manner to be used in ANY format on the other side is the key to providing plug and play integrations and moving away from the ETL way of managing data. DataOS makes data truly interoperable.

Core Principle #5: **Data Fit for Purpose**

Making data available for various uses like Analytics, Data Science, Automation, etc. in the formats and security classifications relevant for that particular use is one of the core aspects of a data fabric architecture and is a foundational capability of our products. This provides an ability for business teams to use the same data for various needs without the need for use case specific data pipelines, transformations, data copies, etc.

Core Principle #6: **Data Ownership**

Data ownership is a construct that has hindered data usability in many organizations. This approach to data governance leads to many data prisons across the enterprise which are tough to govern and manage. This also creates high risk of data breaches. We shifted the data ownership paradigm to a data access paradigm. In this new approach, data is owned by the enterprise and not individual teams. Any user or system that needs access to the data can do so in a compliant manner by leveraging foundational capabilities such as row and column level redactions, data abstraction, data masking, etc. These capabilities, along with data observability, ensures that every user or system in your company has access to the data that they need, in the format that they need with the right governance applied to it.

Core Principle #7: **Security and Compliance**

We are a post GDPR company. As such, we don't have the baggage of rigid architectures or old technologies that limit how large systems can be compliant with standards such as GDPR, CCPA, etc. We use tagging extensively to manage security and priority classifications at the most atomic level, which makes new regulatory compliance needs a business logic solve rather than and re-architecture solve that is often prohibitively expensive and takes months or years. These approaches allowed us to innovate on how we manage governance. Our tag-based governance engine provides customers with the flexibility to provide role based, attribute based as well as tag-based access controls. This empowers teams to set up conditional access controls like the ability to access a data set on a certain network or during a certain period of time or other such rules.

The Modern Advantage



Plug and Play Data Products

Our approach to data, allows customers to use value-driving data products like Snowflake, Google Tensor Flow, Azure ML, C3.ai, etc. in a plug and play fashion without the need for extensive integrations. Gone are the days of managing 100's of thousands of ETLs without knowing where all of the data and jobs are. This new world allows you to bring on products that you think will drive value without having to think about architecture changes, implementation challenges, professional services costs, etc.



Control in Your Hands, Instead of Vendors

This approach of centralized data management, data quality control, governance, etc. ensures that you control the data in your organization and make the front-end products replaceable without big vendor lock ins. We are creating a data future that won't be limited by legacy infrastructure that is expensive to maintain and hard to replace.



Collaboration

In this new age where most of the teams are working remotely, having an ability to collaborate on data workloads in the same fashion that teams commonly collaborate using tools like Slack, Asana, Jira, and Google Docs becomes essential. We have a collaboration layer across our entire product that allows teams to collaborate on everything from data prep, data blending, queries, dashboarding, charting, data quality and data syndication.



Business Agility

In order for businesses to succeed in the post Covid new normal, having an ability to understand changes happening to your business in real time and being able to respond to those changes and innovate becomes essential. You need a data infrastructure that makes technology transparent while you focus on how to put data to use to grow your business and delight your customers.



Real Time

Our system not only identifies changes in data patterns and anomalies in data values in a real time fashion, but also provides multiple ways to act on these insights by empowering other systems like SAP Hybris and Salesforce to react to these real-time alerts.



Experimentation

The business agility that the modern approach brings to data management creates an environment that makes it easy for our customers to launch multiple vendor PoC's at the same time, share data with vendors in a scalable and compliant fashion. We enable data experimentation and make it easy for companies to launch these experiments in a matter of days versus months that it takes today.

DataOS features for government agencies

01

Data Discoverability

Locate all of the needed data with a simple search.

02

Data Accessibility

Retrieve the data with a simple interface.

03

Data Semantics

Examine descriptions of data to recognize the content, context, and applicability.

04

Data Lineage

Explore all of the steps that influenced the data.

05

Data Relationships

Leverage complementary data elements through innate relationships.

06

Data Quality

Consumers can be confident in all aspects of data for decision-making.

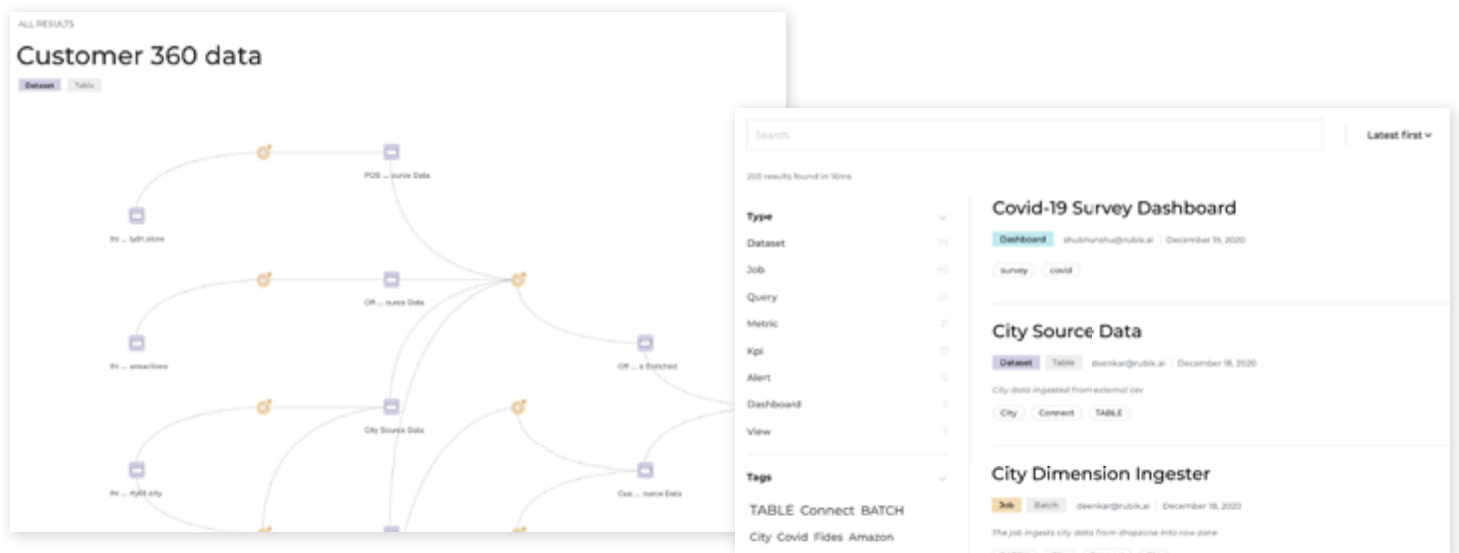
07

Data Interoperability

Consumers and producers have a common representation and comprehension of data.



Fig.1 Data graph rendered in DataOS



The Modern Approach

You own your data destiny. We enable you to be self-sufficient in order to manage your data needs in a future-proof manner. To ensure success, we have a three-pronged approach to engaging with customers.



Summary

DataOS is a fully connected enterprise Data Fabric that reacts to your changing business needs in real-time. It includes everything you need to enable your data and each component can be swapped out with existing solutions you may already have. It is completely modular. Data Fabric is easy to learn and operate. It provides government institutions with end-to-end data management capabilities that can empower these government institutions to become data-driven by transforming their data landscape.

Conclusion

DataOS is tailor made for the post-COVID data world

Data and Analytics are essential in assisting faster and better decision making. However, data capitalization is not easy. The extreme disruption in the aftermath of COVID-19 has invalidated many prediction models that are based on historical data. Organizations using machine learning to build their recommendation engines will now have to rethink their approach to keep pace with changing customer behavior.

In any case, the new normal is still emerging and hence validation of predictive models is a challenge.

100x

DataOS provides 10x the value at a 1/10th of the cost.

A Modern Data Fabric + Simplified Access + Secure Data Exchange in one product.

To make accurate business decisions in an uncertain and complex environment like now, organizations need to bring experimentation to the forefront. With DataOS and its applications, you can quickly test your hypotheses and see what is working for you in these turbulent times.

A recent trends report⁴ from Gartner says, “By 2021, proof-of-concept analytic projects using quantum computing infrastructure will have outperformed traditional analytic approaches in multiple domains by at least a factor of 10”. With its capability to integrate and support end-to-end data pipelines along with providing real-time data governance and sharing abilities, DataOS gets you there.

Appendix

1. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read

<https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#2ecd335760ba>

2. Seizing Opportunity in Data Quality

<https://sloanreview.mit.edu/article/seizing-opportunity-in-data-quality/>

3,4. Our Top Data and Analytics Predicts for 2019

https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/