MODERN

WHITE PAPER

DataOS®
The Modern Data Fabric

Organizations have data – lots of it.

Forbes says we're creating about 2.5 million Terabytes[1] of data each day, but just capturing it isn't enough. Businesses realize now that storing diverse data in a data lake or a big data platform won't miraculously create value out of their data. They must transform, orchestrate, and enrich the data to get to the high-value insights that a data-driven business model promises.

Becoming a Data Driven Organization

Businesses want to become data-driven, but a majority of non-technology organizations struggle to extract value out of their data. For these companies, data is not a solved problem.

Collecting data from traditional data sources may seem straightforward, but enterprises often struggle to integrate, govern, process, and syndicate data to external entities in order to generate the value they so desperately need.

Data is growing in variety (structured, unstructured, semi-structured, streaming, sensor data, etc.), and standard data integration techniques have failed to keep up in the last 5 + years. Compute limitations, connectivity constraints, rigid transformation workflows, ever-increasing data volume, and multi/hybrid cloud environment data distribution -- these complexities exacerbated data management issues and made it even more challenging for legacy platforms.

The truth everyone has grown to understand is this: most of these data platforms offer little value to customers and add needless complexity to the existing data infrastructure.

There's a better way.

Major Barriers for Organizations to Become Data-Driven

Based on our experience working with numerous data platforms, we see these things as major barriers to an organization becoming truly data-driven:

Data Integration challenges

Enterprises tend to spread data across multiple repositories from databases and data warehouses to multi-cloud environments. Integration and governance become difficult due to the heterogeneity of data across data-sources because of sheer volume and variety. The result? Data latency causing terrible business decisions.

Data Privacy Regulations

Enterprises are under constant pressure to keep up with global data privacy regulations like GDPR, CCPA, and any number of recent regulatory updates. Unless their current data platforms support advanced data governance, it's too challenging to get data governance sorted.

"Vendors now offer zero-code and low-code functionality to accelerate even large and complex fabric deployments. Look for vendors that have expanded AI/ML capabilities to automate data discovery, classification, security, ingestion, transformation."

— Forrester Research, The Forrester Wave: Enterprise Data Fabric, Q2-2020

Poor Data Quality

A report from MIT says[2] that data users waste up to 50% of their time dealing with mundane data quality issues and the cost of bad data is 15% to 25% of revenue for most companies. When your data quality is poor and data is unreliable, the consequences damage your bottom line and make it less likely that a business survives the data revolution.

Secure Real-time Data Sharing

Data should be readily available to any employee, partner, or vendor to drive business innovation. However, most data is locked up in data prisons, making access to data in realtime a major problem for enterprises. Without timely access, enterprises cannot power the near real-time business use cases they so desperately need.

In addition, enterprises must share data in real time while balancing intense security concerns surrounding that same data access. Without the right solution, enterprises cannot customize permissions, comply with internal governance, or implement structures that address any new compliance regulations efficiently from one platform.

Limited Self-Serve Abilities

Business teams require timely access to data to ensure customer satisfaction. Currently, there are too many rigid data products that require high technical expertise to operate at their greatest potential. This makes business teams dependent on IT teams for data, resulting in higher time to value.

DataOS® - The Modern Data Fabric

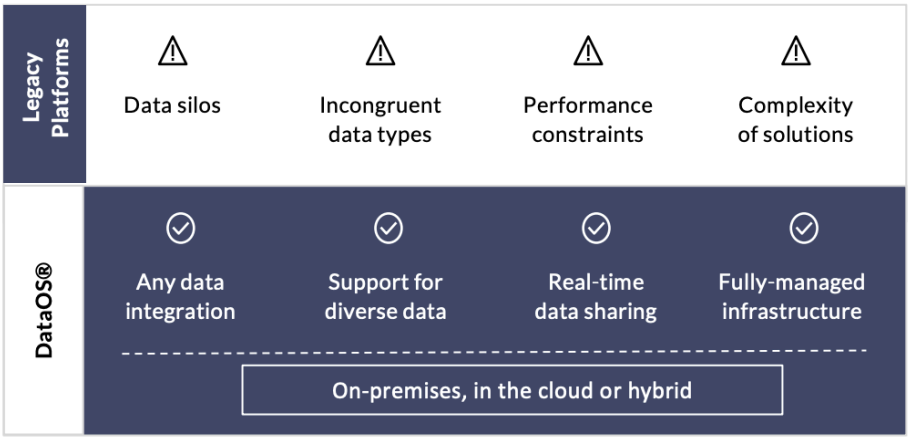
A first-of-its-kind data operating system that enables enterprises to ingest, process, transform, govern, and orchestrate data from disparate data sources to deliver a trusted and real-time view of both customer and business data. DataOS is the data fabric that frees data from constraints and enables actionable insights.

Data Democratization

DataOS automates key components of a company's data pipeline:

- Integration
- Transformation
- Preparation
- Curation
- Security
- Governance
- Orchestration

Together, these components make up data democratization, an essential part of delving into the real-time insights and true value of a business's data. **You get everything in DataOS** -- a real-time preview of to-be ingested data, automatic data profiling, data transformation, built-in governance and facilitation of secure data exchange.



Data Integration - Any Format, Any Source

DataOS integrates all of your data from different data sources (Azure EventHub, Google Analytics, RabbitMQ, RDBMS, and various other external APIs) into a single, unified view -- bringing data and users together. With a unified view of data, data complexity is reduced and enterprises can access their data faster and more efficiently.

Data Quality Management

With data increasing every day, ensuring data quality has become a major problem for enterprises. With DataOS, data set quality is auto-curated during data ingestion. This empowers data users to make smarter, more informed decisions driven by exclusively high-quality data.

Real-time data governance

The DataOS advanced governance engine ensures the availability and usability of data. *The platform:*

- manages data-access controls at granular level -- each user's access to data can be defined based on their role, credentials, and access control rights
- remediates security risks with row/column level data masking and encryption
- maintains history for each workflow -- traceable back for audit and review at any point in time

> “ By 2022, organizations utilizing active metadata to dynamically connect, optimize and automate data integration processes will reduce time to data delivery by 30%. ”
>
> — *Gartner trends*[3]

Data Sharing

Your data is more valuable when combined with other sources for deeper insights. Harmony Syndicate Technology from DataOS enables organizations to create a private data exchange hub where data providers can share live data with data consumers and set the permissions for data accessibility. This helps you create a data community, facilitating secure data sharing between peers, customers, and external partners regardless of data format or location.

Turn-key applications

DataOS comes with an inbuilt application ecosystem that delivers cost-effective data applications in a turn-key fashion. These applications are pre-built solutions to industryspecific requirements -- think Customer 360, anomaly detector, or segmentation engine.

Customers can use these applications to monitor their data ecosystem, better understand their customers, and detect data anomalies, among other critical data insights.

Isn’t that cool? Now let’s look at the architecture of DataOS to see how it can simplify existing processes and derive value out of your data.

Data Catalog

Data Ingestion

DataOS offers a huge library of connectors to support multiple data source systems and users can effortlessly ingest real-time, batch, incremental, and one-time historical data - be it structured or unstructured data.

Harmony Connect Technology

Harmony Connect Technology is an integral component of DataOS that allows users to connect to and ingest data from various data source systems like Azure EventHub, Google Analytics, RDBMS, and various other external APIs. DataOS can also read and write data in various file formats, including CSV, JSON, Avro, Parquet, Logs, and more.

Data Catalog and Discovery

DataOS Catalog is a fully managed and scalable metadata management service. It helps data users to quickly discover, understand, and manage all their data from one simple user interface. Think of it as “the Google of your data.”

Natural Language Search

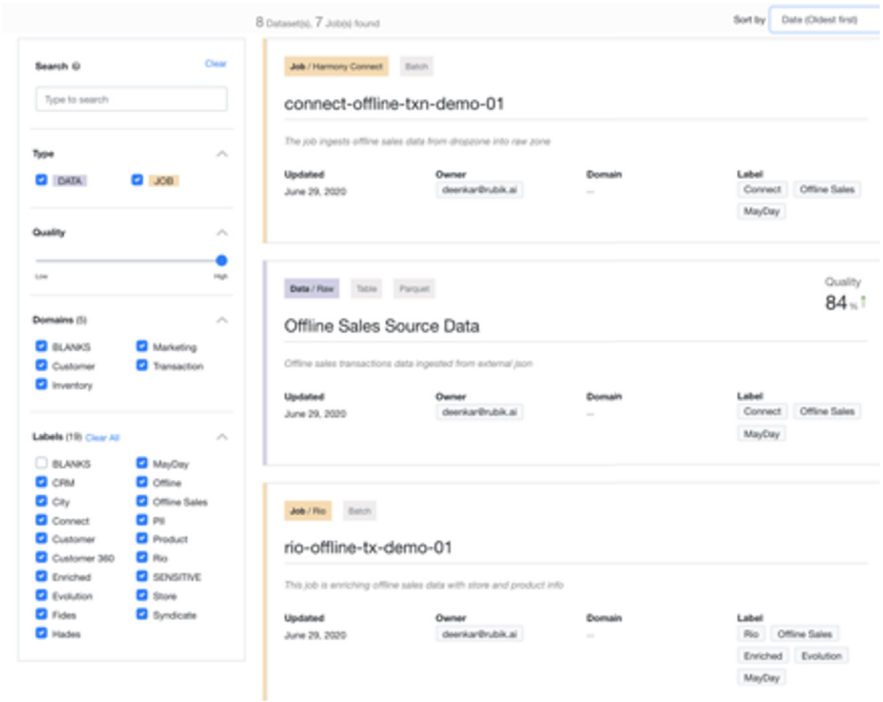
DataOS presents users with an easy to use Google-like search interface that helps users quickly search for data they require. Users can narrow down their search to a granular level and find data using table names, column names, domain-specific, data quality, etc.

Data Profiling

DataOS picks up every dataset within the platform and presents you with a summary of the data. It helps you to quickly understand the accuracy, completeness, and validity of your data. DataOS detects many different data types, including personally identifiable information (PII), credit card numbers, social security numbers, phone numbers etc.

Data Catalog in action

View all datasets and transformation jobs under one roof. You can discover data sets with a simple search and narrow down using tags or domains to sort them as you like.



The world of data compliance is constantly changing

Ever since Europe's General Data Protection Regulation (GDPR) entered public consciousness, tech companies are under the radar of the rapid change in the field of data compliance.

Staying one step ahead of these compliance regulations and protecting their customer data is critical to the long-term success of these companies.

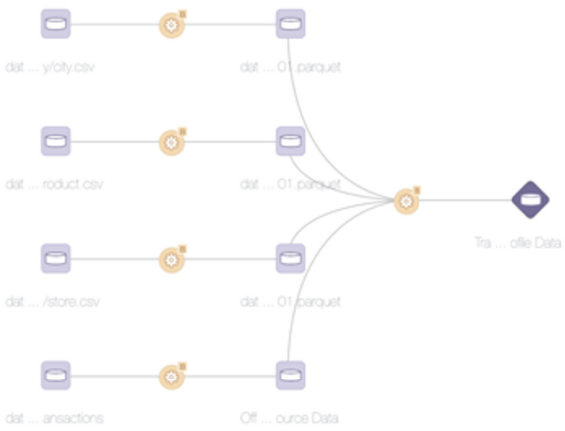
Governance & Compliance

Sensitive Data Encryption, Data Access & Policies

GDPR, CCPA Compliant, Data-loss prevention

Data Lineage

Understand the origin and destination of any data asset, along with data transformations applied, in the data catalog with an intuitive data lineage graph constructed automatically.



Data Governance and Compliance

Data governance in DataOS is built on three key pillars - data quality, ownership, and data accessibility. DataOS presents an omniscient view of data owners, data users, and data consumers at once, helping Infosec teams to track both data and its usage.

Data Quality

Insights from data are only as good as the data quality. After a dataset in the platform, DataOS® automatically curates the data quality of that dataset and presents you with a quality score. Users can define data quality rules based on their requirements and arrive at a reconciled data quality score.

Data Accessibility and ownership

Infosec teams can control data accessibility at a very granular level and provide access based on the user's role and function. Users who deploy jobs to bring in any data will automatically have the ownership rights of that data. Ownership can also be manually curated via DataOS user interface.

Data Compliance built in

DataOS translates laws and regulations from authorities and presents customers with an inbuilt data compliance mechanism. DataOS is a post GDPR platform and is designed to be GDPR, CCPA ready.

Data-out and Application ecosystem

Data Syndication

Your data is of more value when it is combined with other sources for deeper insights. Harmony Syndicate Technology of DataOS enables organizations to create a private data exchange hub where data providers can share live data with data consumers and set the permissions for data accessibility. This helps you create a data community by enabling secure data sharing between peers, customers, and external partners regardless of the data format and its location.

Self-serve data and analytics

Business teams are dependent on IT teams to get hold of data they want to share with external parties. This resulted in a huge time delay between hypothesis initiation and insight generation. With self-serve data capabilities of DataOS, Business teams can self-serve and fetch the data themselves with few clicks and carry out secure data sharing at their own pace.

Modern App Store

DataOS App Store is the first-of-its-kind feature in the data ecosystem world. App Store contains turn-key fashioned applications that solve industry-specific use cases. *A few of them are listed below:*

- **Customer 360** helps you with a collection of all of your customer data at one place, from the basic info, to all their past and present interactions with the company, as well as their social media behavior
- **The notification engine** helps you with alerts on anomalies, spikes, or other patterns of interest from data.
- **The segmentation engine** is an app that lets you easily segment customers according to key variables, and get deeper insights into their behavior and value, etc.
- **DataOS Reports** helps you create customized reports on data in DataOS with visualizations for different stakeholders within an organization.
- **Anomaly** identifies outliers, novelties, noise, deviations, and exceptions in real-time and takes action.

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.

Contact us for a demo

The Modern Data Fabric Advantage

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. 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 report[4] 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.**

Sources

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-mi nd-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/

ADDENDUM

DataOS® - The Modern Data Fabric