MODERN

WHITE PAPER

Data Security and Governance in DataOS

MODERN

Table of Contents

DataOS supports Microsoft Azure, Google Cloud Platform, and Amazon Web Services cloud providers.

At TMDC, we are fully committed to the security of our customers and their data. While designing DataOS, we kept in mind all the features required for customers to protect their data.

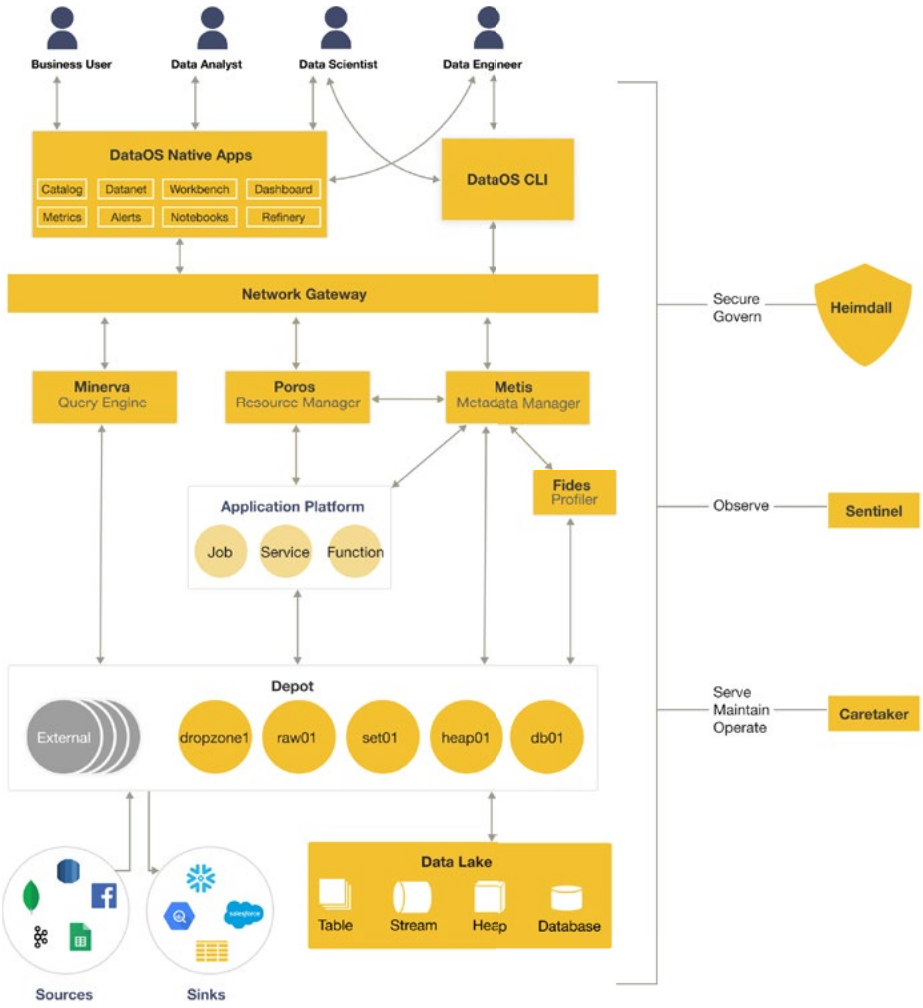
Introduction

The Modern Data Company's® (TMDC) DataOS® is a platform as a service (PaaS) solution that can be deployed within your cloud infrastructure. DataOS, the modern data fabric, protects your data regardless of where you move, store, or analyze your data. To achieve this TMDC created a security framework that guarantees availability, integrity, and confidentiality of your data at all times.

Security overview of cloud service providers

Every cloud provider has a robust set of security features that act as the backbone to the overall DataOS security implementation. Security features offered by the major cloud service providers may differ from one another but they all provide the key measures required to protect their customers' data. These security measures generally include network and infrastructure security, host and endpoint security, data encryption by default, logging and monitoring, identity and access control, governance, and risk and data compliance.

Fig.1 *- Achitectural context of DataOS - The Modern Data Fabric*



Secure by Design

DataOS includes a multitude of security features like the protection of data at rest or in transit, authentication and authorization of users, creation of audit feeds and capturing logs, and more. DataOS ensures data is always fit for purpose—we carefully consider all data concerns involved in data collection, sharing, and use, as well as rapid data integration and restriction of access.

Data Protection

- All data is always encrypted
- Encryption keys are automatically managed

Authentication

- Embedded multi-factor authentication
- Federated authentication

Authorization

- Role based access control method
- Tag and attribute based access controls

Auditing

- Capture and Syndicate Audit Feed
- Complete audit trail of system events and user access logs

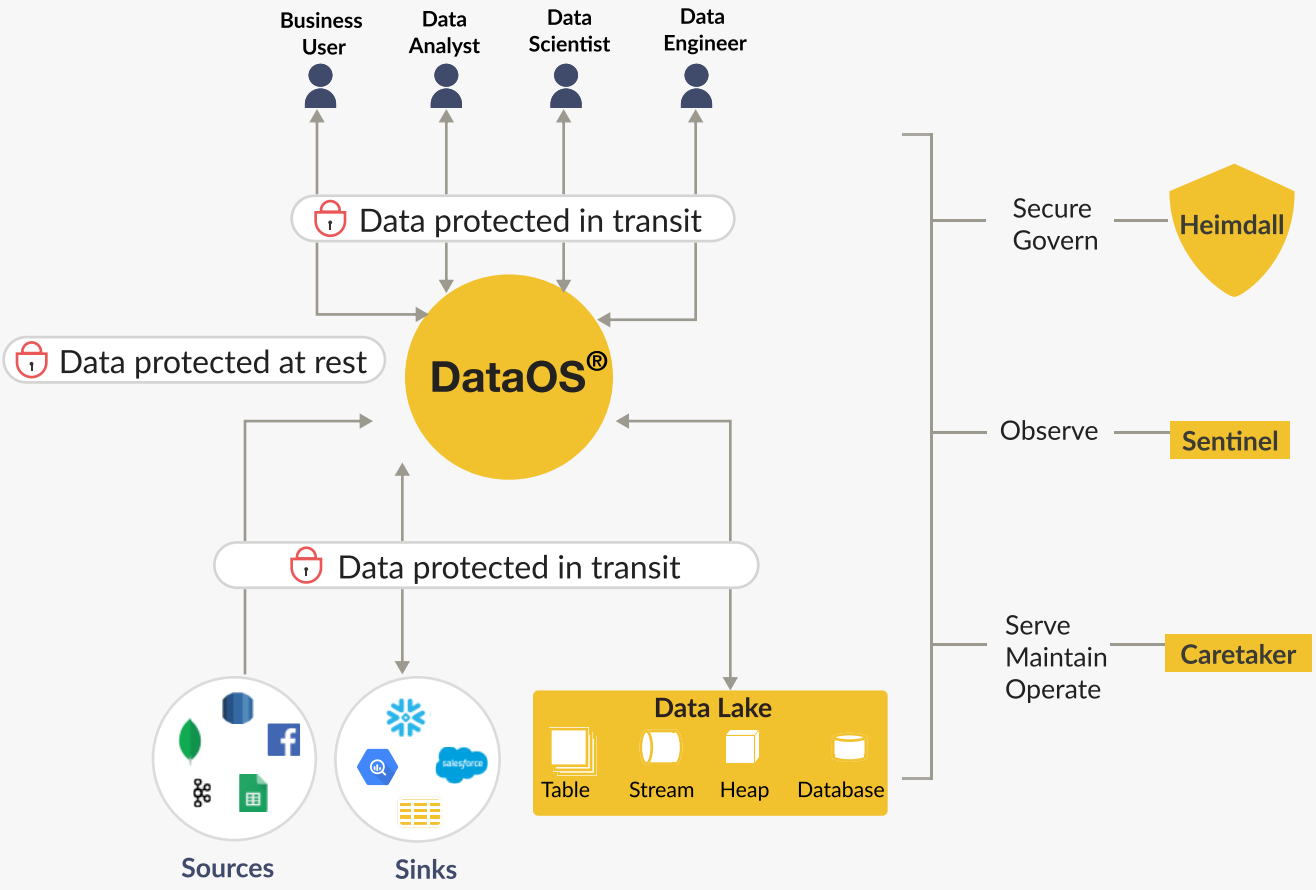
Data Protection

DataOS utilizes encryption at various points to protect your data during its life cycle from origin to destination, including encryption of data at rest and data in transit and ensuring high data availability and redundancy.

Encryption of Data at Rest

DataOS considers all customer data as confidential and encrypts data at rest, leveraging the cloud service provider's encryption algorithm to protect against physical security breaches. Every cloud provider has a slightly different way of encrypting, but storage is generally encrypted using encryption keys that are generated and rotated automatically by the cloud provider. Encryption keys can also be explicitly defined and managed as determined by the customer's security policies. Data stored in the default DataOS depots (e.g., Table, Stream, Heap, and Database) is encrypted at rest at the storage level using AES-256, leveraging the cloud service provider's default storage encryption.. This enables customers to guard against unauthorized access, illegal copying, and other security threats, ensuring data and firmware integrity at all times.

Fig.2 - Showing data protection at various stages of data life cycle in DataOS



Encryption of Data in Transit

DataOS employs all necessary security measures to help ensure customer data is protected during transit. It does this by encrypting the data before transmission, authenticating the endpoints, and decrypting and verifying the data upon arrival.

Internal data transfer within DataOS: Default DataOS depots utilize secure and updated versions of protocols like TLS 1.2 instead of SSL for data transfer. All data accessed from applications occurs over https using TLS 1.2.

External data transfer outside DataOS: From data ingestion and syndication to external depots, source and sink systems define the protocols and security. Source systems are usually the input data source systems and the sink systems are the data destination systems. DataOS utilizes current versions of connector libraries and chooses the most secure option available for connecting to these external depots.

Because DataOS's innovative implementation of attribute-based access control uses tags and conditions to define policy, you can implement very coarse-grained access controls, like role-based access controls or very fine-grained access controls like AWS, IAM, and beyond. In DataOS, policies are defined, and users and their tags are managed in Heimdall.

Data Availability and Redundancy

Data availability and redundancy are system features designed to provide a consistent level of uptime for prolonged periods. Availability and redundancy features vary across cloud service providers, but blob storage and disk storage are highly available and redundant based on provisioning configurations. As you ingest data, it can be synchronously and transparently replicated across availability zones. Storage provisioning configurations are tuned to suit your business requirements during DataOS deployment. Disk backups are performed nightly and seven days of these snapshots are retained.

Authentication

DataOS asks every user accessing it the question, "Who are you?" and verifies their response as a process of user authentication.

Federated Identity Management

Federated identity management is essentially a set of agreements and standards that help enterprises and applications share user identities—this enables, for example, a user to log into Facebook, LinkedIn, and Spotify using their Google ID. DataOS leverages external identity management solutions for login and user creation like LDAP, OIDC, Google, Microsoft AD, and SAML. The identity provider is configured during DataOS deployment. In DataOS®, the Heimdall (Access Control Gatekeeper) service acts as a portal to external identity providers through connectors. Heimdall provides an OpenID Connect interface to all DataOS applications and services, regardless of the external identity management solution's protocol. Once a user successfully logs into the external identity provider, Heimdall will proxy or issue a JSON web token that is then used for identity verification in DataOS applications and services.

User Management

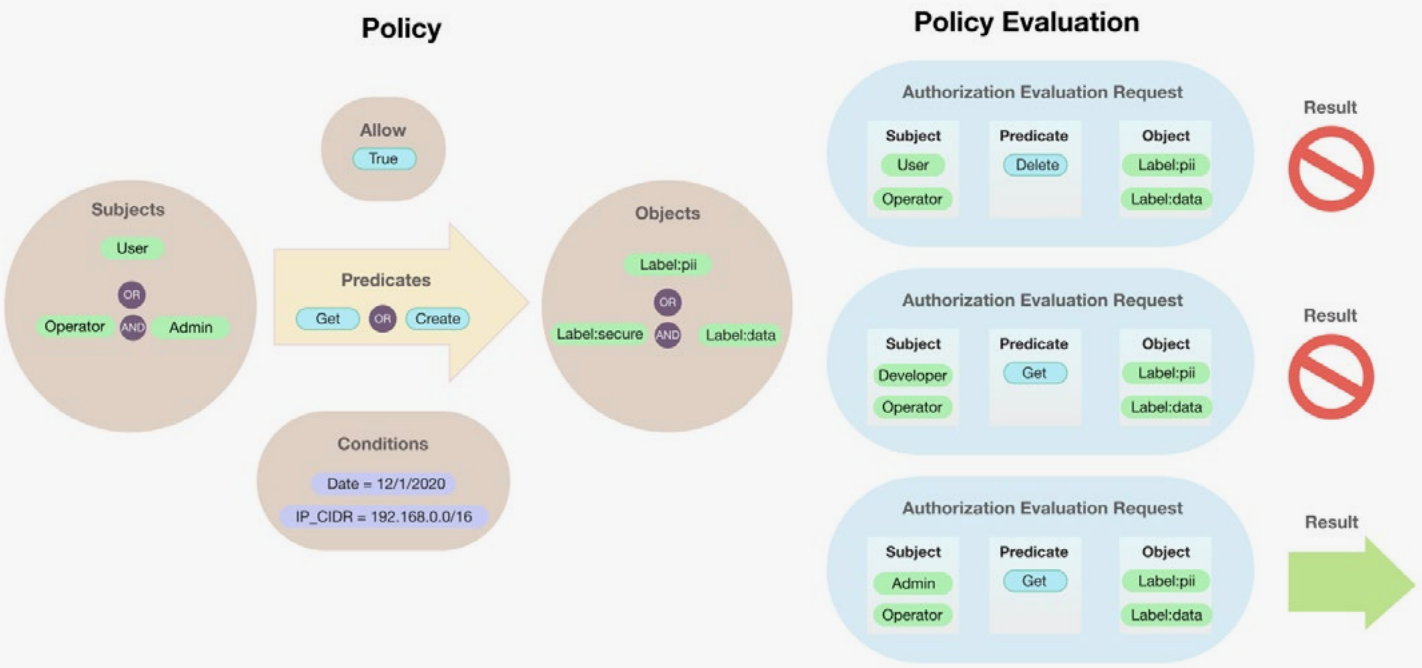
When an authorized user logs into DataOS for the first time, a user entity is created in DataOS through the federated identity provider and Heimdall. The user entity is tagged according to the configured external group, defined during DataOS deployment. Admins can easily manage users, their attributes, and tags in the DataOS Operations Center.

Authorization

Authorization allows or restricts actions that authenticated users can perform on the DataOS platform.

Our tag-based governance mechanism provides you with the flexibility to provide access control based on roles, attributes, and tags. This empowers teams to set up conditional access controls like the ability to access data sets on a specific network or during a specific time window.

Fig.3 - Showing a sample policy evaluation methodology during authorization in DataOS



Network Communication

You may need DataOS to communicate with external third-party data solutions to deliver a wide range of services. All communications between DataOS and external data solutions must be authorized and initiated within DataOS. Network communication gives you visibility into the network traffic flow between your services, applications, and availability zones.

Ingress Network Security

Virtual private cloud offerings vary across cloud providers, however the DataOS compute tier can use virtual machines that are placed in a private network. The virtual machines have their own private IP addresses.

The private network has a firewall with highly restrictive ingress rules from public IP addresses on ports 80, 443, and 7432. Port 80 is required for public certificate creation and rotation from Letsencrypt. Port 443 is for all applications to talk with services using HTTPS. Port 7432 is for the Minerva query engine to expose its streaming services which use HTTPS.

Additionally, all ingress traffic goes through a policy enforcement point which validates with Heimdall whether a request is allowed.

DataOS can help your organization build a fully compliant data solution, enabling you to process sensitive and personal data fairly and in accordance with the law.

Cluster Network Security

DataOS uses a zero-trust network perspective to implement internal security measures. All communication between applications and services is explicitly defined as a communication policy, reducing the ability of bad actors to call internal services and ensuring that dependencies are known.

DataOS also implements a service mesh for applications and services. The mesh requires every application and service to have an envoy proxy sidecar. All communication is done through the envoy proxy which makes sure there is an explicit allowed communication policy and creates a mutual TLS connection between envoys.

Auditing and Logging

DataOS's data observing capabilities allow it to pull logs at the application level, ensuring no sensitive data is captured in logs. All the logs are captured whenever an event occurs in the system and can be used for audit purposes.

Audit Data Set

DataOS by default comes with the functionality to create a dedicated data set for auditing events happening across DataOS applications and services. To add visibility and auditability, this data set contains the who, what, where, and when of activity that occurs across your DataOS environment in the form of audit logs. You can consume and analyze this audit feed as needed.

Logging Depot

DataOS has a depot dedicated for logs from applications and services running within the platform. All applications and services within the DataOS stream compress their logs into the Logging Depot which utilizes cloud blob storage by default. Log data sets can be consumed and analyzed as needed, using standard DataOS primitives.

Audit data sets and logs can also be syndicated to external systems, if defined by your operations and security processes, to analyze and deeply contextualize events.

Compliance

DataOS can help your organization build a fully compliant data solution, enabling you to process sensitive and personal data fairly and in accordance with the law. Being a post-GDPR company, we have provided all the primitives needed for organizations to be data compliant in DataOS. It can be used to implement many data regulatory and privacy compliant standards like GDPR, CCPA, and PCI DSS.

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.

Conclusion – The Modern Way to Data

DataOS's comprehensive security framework guarantees availability, integrity, and protection of your data at all times. Whether you have data fragmented across on-premise, remotely, or in the cloud, the DataOS platform can store, encrypt, move, protect, and recover your data. Regardless of where you move, store, and analyze your data, DataOS ensures the privacy and safety of your applications and data.

About DataOS®

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. 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.

Learn more →

About The Modern Data Company®

Founded in 2018, The Modern Data Company® (TMDC) began with the realization that enterprise-wide data access has been siloed. Data engineers and database administrators have been the longstanding data gatekeepers who funneled data to analysts and data scientists. We aim to change that by freeing enterprises to make better data driven decisions by democratizing access to data. When all employees, irrespective of their technical skills or background, can easily explore and analyze enterprise data, then both productivity and market expansion are realized at a faster pace.

Learn more →