Fully Managed Data Orchestration

How Astro Securely Manages the Flow of Data

ASTRONOMER

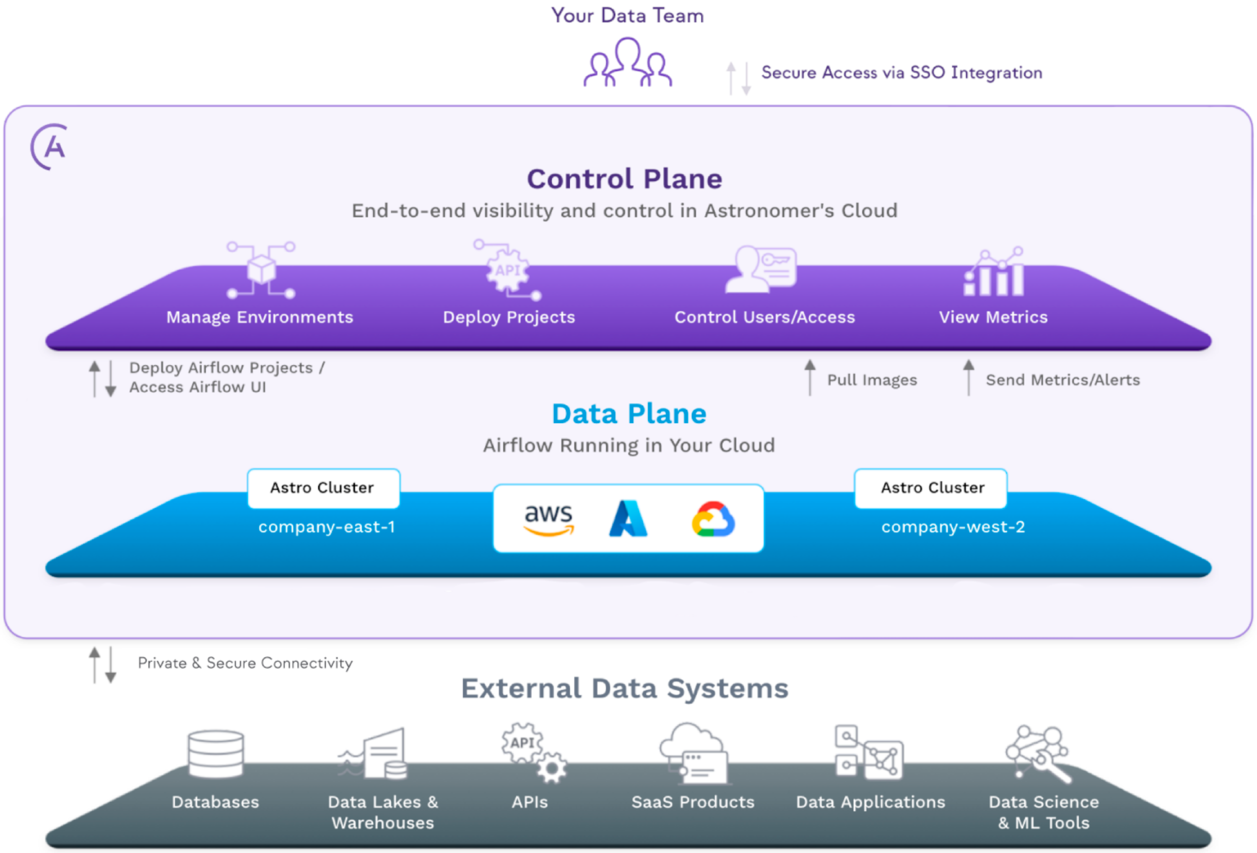
Astronomer recognizes the criticality your data pipelines have in the success of your business. To that end, we offer Astro – a managed software service that enables a next-generation experience for modern data teams running Apache Airflow, the open source industry standard for data orchestration. This document provides an overview of the Astro architecture, access controls, platform security, and compliance details built into the service providing security and peace of mind for you and your data teams.

Architecture

Astro boasts a hybrid deployment model founded on a multi-cloud Control Plane hosted by Astronomer and a single-tenant Data Plane that is hosted in your public cloud environment (AWS, Google Cloud, or Microsoft Azure). Both are fully managed by Astronomer.

The control plane provides end-to-end visibility, control, and management of users, workspaces, Astro Runtime (an enterprise-grade distribution of Apache Airflow) deployments, metrics, and logs. The data plane is the single-tenant environment in your cloud for data orchestration; it forms the foundation from which Runtime deployments run and your data pipelines orchestrate other data services.

This model offers the self-service convenience of a fully managed service while respecting the need to keep data private, secure, and within corporate boundaries. It optimizes for security while relieving your team of operational overhead.



Access Control

Astro is accessible over the internet through secure connections only. Customer users can connect to Astro using our web-based user interface or programmatically through our command-line interface (CLI). All communication between users and Astro is secured and encrypted using strong TLS 1.2 ciphers. Astro also employs threat intelligence signals to protect the platform and your users from automated attacks:

1. Brute-Force Protection with Account Lockout: Safeguards against brute-force attacks that target a single user account.

2. Suspicious IP Throttling: Protects against high-velocity attacks that target multiple user accounts from a single IP address.

Authentication

A variety of options are available to securely authenticate to Astro. Customers can take advantage of our default options, providing a secure, out-of-the box authentication experience.

Email + Password

- Passwords are securely stored — never in cleartext
- Passwords are hashed and salted
- Strong password policy is enforced
- Previous passwords are disallowed from reuse
- Dictionary words are disallowed in passwords
- Personal data is disallowed in passwords
- Email verification is enforced

Google or GitHub Social Login

- Single sign-on experience
- Emails are verified by the provider
- Up-to-date user information

For customers that are looking for more control of their users' authentication mechanism via federated authentication, Astro supports integration with Okta and Azure AD, enabling a common single sign-on experience for users and central management of identities. Customers can also take advantage of advanced access policies to enforce trusted IP ranges or authorized devices via Adaptive Authentication from Okta, and Conditional Access from Azure AD.

Support for federated identity management is available with Astro Standard and above. Astro Premium provides an ability to enforce specific authentication methods for an organization.

User Authorization

Astro provides a reliable and secure role-based access control (RBAC) authorization system to ensure your data, data pipelines, metadata, and users can be accessed and managed only by authorized users within your Astro Organization.

Isolation by team, department, or use case can be achieved using roles across your Workspace(s). Astro Runtime deployment isolation can be achieved through Deployment API Keys, which are unique per deployment, and can be used to programmatically deploy pipelines.

An Organization Role grants a user a base level of access throughout an Astro Organization, while a Workspace Role grants a user additional access to a specific Workspace. In an upcoming release, customers may also choose to organize their users using Teams. A Team may be granted access to a Workspace in place of or in addition to individual users. Privileged roles inherit the permissions of the less-privileged roles.

An Astro Organization belongs to one customer and includes all of their users, data, data pipelines, metrics, and logs. Astro validates that every user and API key have the necessary privileges to access these assets.

Pipeline/Task Access

When pipelines run in Astro, tasks are executed that may include connection to external data services. These tasks are defined with standard connections, which may use defined environment variables for credentials to securely access the external data service. Alternatively, Astro integrates natively with common secrets management platforms to fetch credentials at runtime.

On Google Cloud, each Astro Runtime deployment runs with a unique identity in the form of an IAM Service Account that can be granted access within the external data service for a credential-free option. A similar capability is planned for AWS and Microsoft Azure.

Astronomer Access

Astro is a fully managed modern data orchestration service, composed of a multi-tenant Control Plane and single-tenant Data Plane implemented and supported by Astronomer. Astronomer's access to your Data Plane is limited to the public cloud service APIs used by our automation, and support tooling for the cluster.

The Control Plane automation leverages cloud-specific APIs and accounts to provision and manage Data Plane cloud accounts and infrastructure, implemented with cloud-specific constructs:

- AWS cross-account IAM role for Data Plane AWS account
- Google Cloud service account for Data Plane Google Cloud project
- Microsoft Azure service principal for Data Plane Azure subscription

Astronomer leverages a remote access control platform to proactively respond to Data Plane cluster alerts and to triage support cases. Access to Data Plane clusters is limited to authorized employees and is time-bound. All sessions require multi-factor authentication, and have their issued commands logged and audited, and where relevant screen recorded; both logs and screen recordings are stored centrally and regularly reviewed by Astronomer.

Data Security

Astro uses both encryption in transit and encryption at rest to protect data across and within the Control Plane and Data Plane.

All communication between Control and Data Planes is encrypted in transit using TLS 1.2 with strong ciphers. This includes end-user traffic proxied via the Control Plane to the Airflow UI and API in the Data Plane, and all system-generated egress traffic initiated from the Data Plane to the Control Plane for the purposes of pulling state and updates, sending metrics, logs, and metadata, and support operations.

All data at rest–OS and data disks, object storage, and DB tables, DB temporary files, and DB backups–across Control and Data Planes is encrypted with the industry-standard 256-bit Advanced Encryption Standard (AES-256) encryption algorithm, one of the strongest block ciphers available. This is implemented using native cloud provider key management technologies and envelope encryption.

Infrastructure and Network Security

Networking

The vast majority of the network traffic between the two planes is egress-oriented from the Data Plane. All customer data flows within the Control Plane cluster transit through an mTLS mesh, enforcing TLS 1.2 with strong ciphers. All internal service communication within the Data Plane is transmitted using TLS 1.2 with strong ciphers.

Within the Data Plane, Astro enforces network isolation between Astro Runtime deployment namespaces, ensuring that communication between deployments is denied, and unintended communications and attempted data exchanges are blocked.

Most importantly, Astro allows you to securely connect to external data services from your Astro Runtime deployments using a variety of mechanisms:

- Public Endpoints to third party cloud services, with a fixed pair of egress IPs
- PrivateLink / Private Service Connect, to connect to cloud PaaS Services and cloud hosted customer-owned/partner services
- VPC / VNet Peering, to connect to services running in private IP space

Additional cloud-specific details are available in our product documentation.

Cloud Computing Security

Astro supports customers orchestrating and processing their most highly sensitive and confidential data within the Data Plane, by employing deep hardware-enabled encryption of compute across all three public cloud providers.

- AWS: hardware-enabled encryption at rest and in transit with EC2 Nitro instances
- Google Cloud: Shielded GKE nodes leveraging built-in encryption at rest and in transit cluster features
- Microsoft Azure: AKS-managed Virtual Machine Scale Sets (VMSS) leveraging built in encryption at rest and in transit cluster features

In combination with regular cluster and node upgrades, Astro’s built-in node auto-scaling results in nodes that are regularly repaved. Clusters on Google Cloud and Azure support scaling Astro Runtime worker nodes to zero when data pipeline workloads have completed, ensuring compute and storage are released and permanently wiped.

Threat Detection

Astro uses a combination of advanced threat detection and attack protection tools to monitor and protect the platform. All security events and logs are centrally logged, monitored, and alerted on. In addition to the protections described in the Access Control section, our tooling detects threats at runtime by observing the behavior of the Control and Data Plane containers, including but not limited to:

- Unauthorized or abnormal changes to critical system files, directories, file permissions, and listening ports, all detected by file integrity monitoring (FIM)

- Unexpected network connections or socket mutations
- Spawning of processes and attempted privilege escalation

Physical and Environmental Security

Astro leverages all three major public cloud providers (AWS, Google Cloud, Microsoft Azure), thus physical and environmental security is delegated to these providers. Each cloud service provider provides an extensive list of compliance and regulatory assurances that they are rigorously tested against, including SOC 1/2-3, PCI DSS, and ISO27001.

Compliance

Astronomer is committed to maintaining existing compliance and pursuing additional well-established security and privacy industry standards.

Astro is compliant with AICPA SOC 2 controls with respect to the security, availability, and confidentiality Trust Service Categories. If you are interested in obtaining our SOC 2 Report and Penetration Test report under a non-disclosure agreement, contact sales@astronomer.io.

Astronomer is also both GDPR and HIPAA compliant as an organization, and the Astro platform is GDPR and HIPAA ready. Astronomer offers a Data Processing Agreement (DPA) for GDPR and a Business Associate Agreement for HIPAA, to satisfy the requirements between all parties as imposed by both regulations.

Additionally, for organizations processing payment card information, Astro is certified as compliant with PCI DSS security standards.

Astronomer continues to invest in security on an ongoing basis. If there are questions about any of the content above, or suggestions on further documentation and product enhancements, please reach out to security@astronomer.io.