

ALGORITHMIA ENTERPRISE

TECHNICAL FAQ

Algorithmia Enterprise is the foundation layer for intelligent software. It turns complex services and machine learning models into REST APIs, centralizes them for ease of discoverability, and monitors them from a single dashboard. Companies use Algorithmia Enterprise to reduce duplication of effort between siloed teams and accelerate go-to-market for AI-driven products.

This is a working document that is meant to answer commonly asked technical questions about Algorithmia Enterprise. If you have additional questions, don't hesitate to reach out directly to info@algorithmia.com.

Sections:

- Cluster Hosting
- Cluster Infrastructure
- Cluster Scaling
- Cluster Monitoring
- Security
- Algorithm Deployment
- Data Connectors
- External Algorithms
- Support

Cluster Hosting

1. What type of installation options are available for Algorithmia Enterprise?

Algorithmia Enterprise is cloud-agnostic and can be deployed on any private infrastructure behind strict firewalls. This includes public cloud (ie: Amazon Web Services, Google Compute Platform, Microsoft Azure) and on-premises.

All options are designed to meet commercial-level security and data lockdown. Both data and compute will never leave your private network. Your DevOps team will be trained to deploy, diagnose, and manage the cluster. Algorithmia engineers can be made available 24x7 for hands-on support.

2. Do you support hybrid clouds?

Yes. A single deployment may include compute nodes (called “Workers”, see next section) in different locations. Workers can be in different zones, different regions, or completely different clouds.

3. Do you offer a managed service?

Yes. Clients may opt for a managed service whereby Algorithmia engineers take on the role of first-line DevOps.

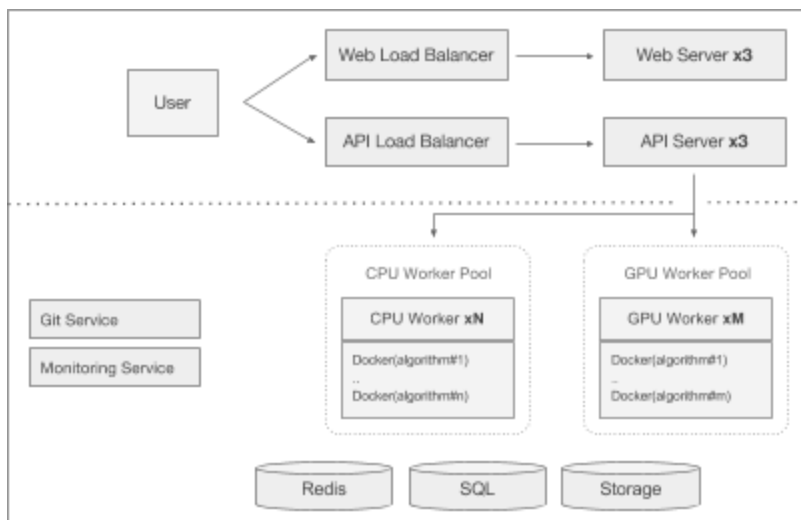
4. How are updates applied for different solutions?

A managed solution should expect bi-weekly updates, or as needed, after at least one week of internal testing. A self-managed solution is updated less frequently. All updates are opt-in and require the approval of system administrators.

Cluster Infrastructure

1. What are the major parts of the Algorithmia Enterprise infrastructure?

Algorithmia Enterprise infrastructure is made up of three types of servers: web servers, API servers, and worker nodes. The simplified illustration below does not include permanent storage and key-value store nodes.



Algorithmia Enterprise is designed for production throughput, and therefore a typical installation will have at least three instances for each type (triple resiliency) to ensure graceful recovery from any failure.

2. What are “worker nodes”?

A worker node is a server instance hosting a container cluster. Each container, in this case a Docker container, hosts a single algorithm in a sandboxed environment. The Docker container is configured to load the relevant runtime interpreter (such as Java or Python), load external libraries, and process requests from the API servers.

3. Are there different types of worker nodes?

There are two types of worker nodes: CPU nodes and GPU nodes. An algorithm developer specifies the type of architecture they choose their algorithm to run on. GPU nodes are meant to run deep learning models more efficiently, although these same models can be ran on CPU nodes with lower performance. Algorithmia Enterprise does not require GPU worker nodes for on-premises solutions if it is not already available.

Cluster Scaling

1. Can Algorithmia Enterprise scale horizontally?

Yes. A worker node is automatically added or removed to proportionally meet the throughput requirements at any given point. This operation is automated using Algorithmia Enterprise's built-in auto-scaler, however administrators can always interfere to manually control this behavior from the Admin Panel.

2. What is the observed latency for an API call?

Observed latency, defined as the duration of routing an API call from the load balancer to reaching an already initialized algorithm process within a container, is under 20 milliseconds for out-of-the-box CPU algorithm on a low-load AWS cluster.

3. What happens when the cluster reaches capacity?

API calls will be queued until a slot is made available. Administrators will have real-time visibility to the queue chart from the Admin Panel. This metric is also one of the inputs to the built-in Algorithmia Enterprise auto-scaler, which (if enabled) will provision a new worker instance depending on the queue sensitivity.

Cluster Monitoring

1. Are there out of the box dashboards?

Yes. Metrics includes (per minute) number of API calls, API latency, worker nodes workload, etc. A dashboard is also available to list latest user actions and system-wide events. See "*Admin Experience*" document for more info.

2. Can you integrate with external dashboards?

Algorithmia Enterprise is bundled with [Grafana](#). All metrics and data are available as APIs for integration with external dashboards.

3. Do you log errors?

Yes. Administrators will have real-time access to any errors resulted from user algorithms or Algorithmia Enterprise infrastructure. Error logs include full error messages and a trendline chart.

4. Do you keep an audit trail?

Yes. All user activity and API calls, whether successful or not, are logged and accessible to administrators.

5. Do users have unique API keys?

Yes. Each user is given a default API and have the ability to issue new API keys. Users are strongly encouraged to create a API key for every production integration or experiment, and continuously rotate those keys when it makes sense. Dedicated API keys allow for better monitoring and risk isolation.

An API key can be configured to run on specific algorithm(s), specific organization(s), or exclusive client types. API keys also support data-access types: no data, read only, and read-write.

6. Do you support teams and organizations?

Yes. By default, any member can create a organization and invite other members. Members of an organization behave as collective owners of their work. Therefore, a private (or unpublished) algorithm is made available to all members within an organization.

Security

1. Do you support Single Sign On (SSO)?

Yes. The authentication and authorization mechanism is customizable when needed, including support for OpenID Connect, Okta, LDAP, and other systems.

2. Do you support multi-tenancy?

Yes. A job (or API call) on AE will always operate in its own memory space and will never share or leak memory to other jobs. Each API call instantiates a dedicated Docker container which is destroyed after execution, resulting in perfect isolation at a peta-scale performance.

Algorithm Deployments

1. What technology stacks does Algorithmia Enterprise support?

Algorithmia Enterprise supports algorithms written in Java, Scala, Python, R, Ruby, NodeJS, and Rust. Built-in support is also available for external libraries and package managers, such as Maven, Python Package Index (PIP), Ruby Gems, NPM, and others.

Algorithmia Enterprise also supports custom runtimes, enabling administrators to support additional programming languages or environments. This can be coordinated with your Account Manager.

2. How are algorithms accessible?

Every algorithm/model is made available through a secure REST API, with JSON input and output, including support for binary JSON. To streamline integrations further, an Algorithmia Enterprise client is available for most programming languages that is meant to abstract away the REST API with more friendly syntax. Algorithms are also accessible from a command line interface (CLI).

3. Do you support method chaining?

Yes, this is a core principle of Algorithmia Enterprise. Any algorithm can depend on or call another algorithm written by a different user as long as their API-key settings allow it. In fact, Algorithmia Enterprise will attempt to slot chained algorithms in a way that maximizes cluster utilization and minimizes runtime latency.

4. Do you support private and staging-deployments of models?

Yes. Any change to an algorithm is made available as a private API endpoint that is only consumable by the author's API-key. The author of the algorithm has the choice to continue testing an algorithm or publish it, which makes it available to other users. Publishing an algorithm can be accomplished through the web interface or through a command-line interface.

5. Do you support source code management / git?

Yes. Every algorithm has its own git repository URI. Authors can pull and push from that repository similar to traditional workflows. Every change to an algorithm is saved as a new git-commit. See [link](#).

6. Are models/APIs versioned?

Yes. Every git-push is made available as a versioned, scalable API endpoint. Algorithmia Enterprise follows a semantic versioning system (major.minor) which allows users to build dependencies on certain versions and avoid breaking changes from authors. Authors have control over the version increments when they publish a change.

7. Can deployments be achieved with continuous integration pipelines?

Yes. Algorithmia Enterprise supports a REST API and command-line interface to facilitate deployments of algorithms and data sources.

Data Connectors

1. What data sources do you support?

We provide native support for Algorithmia Hosted Data, AWS S3, OpenStack Swift, Dropbox, and HDFS through the Algorithmia Data API, which enables fine-grained settings to setup these data sources with role-based access controls (per user/organization/API key).

Algorithm authors still have access to traditional data sources (ones that are not listed here) through their framework APIs, such as SQL and Graph connectors.

2. As a multi-instance infrastructure, how is state shared between parallel API calls?

Algorithm containers are stateless, however, algorithm authors can use the Data API (or native data source clients) to persist and communicate state between calls.

External Algorithms

1. What are Marketplace Algorithms?

These are algorithms authored by users from the public Algorithmia.com Marketplace; users not part of your organization or company.

The public marketplace already includes thousands of algorithms written by researchers and application developers. Some of these algorithms are open source and others are closed source with royalty fees. Algorithms span every field of computer science, including machine learning, deep learning, text analysis, computer vision, time series, etc.

See <http://algorithmia.com/algorithms>.

2. Can Algorithmia Enterprise access external algorithms from the Algorithmia Marketplace?

Yes. A standard Algorithmia Enterprise installation does not include any marketplace algorithms by default. However, system administrators can cherry pick algorithms from the marketplace at any time. These algorithms will be replicated (or imported) into the Algorithmia Enterprise installation and will always run within the private cloud of the customer. License fees might apply.

Support

1. Do you provide SLA?

Yes. We tailor our SLA to cover uptime and level of support.