

# Mod9 Alignment Server

## v0.8-i386-20140123\*

The Mod9 Alignment Server runs the Mod9 Alignment Service using a virtual image in Amazon's Elastic Compute Cloud infrastructure. It is intended to allow customers to operate the service on dedicated machinery within their organizations. This provides customers with additional levels of reliability, scalability, and data privacy.

### 0. Software licensing and preparation

The Mod9 Alignment Service requires the use of HTK software, owned by the University of Cambridge. Your organization must accept **HTK's license terms**:

<http://htk.eng.cam.ac.uk/docs/license.shtml>

Then download the HTK source code and inform Mod9 of a URL hosted within your organization (and possibly protected with HTTP basic access authentication) at which the following file may be downloaded:

**HTK-3.4.1.tar.gz**

The file must have the following MD5 checksum:

**b3fc12006b0af12f59cd573c07aa4c1d**

Mod9 will then prepare an EC2 Amazon Machine Image specifically for use by your organization. You should send your **Amazon Web Services Account Number** to Mod9 in order to request permission for access to the private S3-backed image. Note that the image is not EBS-backed, which is by design.

You may also request that Mod9 configure the server to include HTTP basic access authentication for selected service endpoints, although stricter security should be achieved by properly configuring a network firewall.

### 1. Installation

Launch an AWS EC2 instance from the AMI provided by Mod9. Under the AWS Console, you may need to select the filter for "Private Images".

The compatible 32-bit Linux instance types are **m1.small**, **m1.medium**, or **c1.medium**; the appropriate server to use depends on your expected usage requirements. For very light workloads, the m1.small may suffice and is the most cost-effective option. The dual-core c1.medium will fully process up to two simultaneous requests to the server at full processing speed; further scalability is possible by launching multiple server instances and appropriately load-balancing the requests. The m1.medium is not particularly cost-effective for this task.

Ensure that the instance is configured to launch with an **ephemeral storage device** (e.g. ephemeral0) that will attach to the instance at **/dev/sda2**. This should be the default setting, but is important not to unset.

The instance should be launched **without a key pair**. Login access to the server is not required, except for authorized maintenance performed by Mod9.

You should appropriately configure the instance's firewall (EC2 security group) to allow access to the standard **HTTP port 80**. You may wish to restrict access to only within your organization. For support, it would be helpful if access to ports 80 and 22 (SSH) were also provided to Mod9. At the current time, these are IP address ranges from which Mod9 would typically try to access the service:

```
50.18.56.18
192.150.186.0/24
192.150.187.0/24
```

For convenience, most customers typically leave ports 80 and 22 open to all IP address ranges. As a basic level of security, the Web-based server logging, status, and maintenance facilities can be password-protected.

You may also wish to assign a new IP address and/or associate the current instance with a more convenient DNS hostname.

If you have properly informed Mod9 of the URL for the HTK software, the instantiated server will download, configure, build, and install the HTK software automatically upon its initial boot. It may take about 5 minutes for this compilation and build process to complete. After this, the HTK source URL is no longer required to be accessible, unless a new instance is to be launched.

## 2. Supported Network Monitoring Software

SNMP has been installed and configured as follows:

```
# Basic info
syslocation "AWS EC2"
syscontact support@mod9.com

# Exclude process names, but not CPU time
view all included .1
view all excluded .1.3.6.1.2.1.25.4
#view all excluded .1.3.6.1.2.1.25.5.1

# VACM configuration for "logicmonitor" community string
com2sec readonly default logicmonitor
group MyROGroup v1 readonly
group MyROGroup v2c readonly
group MyROGroup usm readonly
access MyROGroup "" any noauth exact all none none
```

To use SNMP, your instance must belong to an EC2 security group that has allowed UDP access to port 161. Note that this SNMP service is not particularly secure, so individual process reporting has been excluded.

### 3. Customized Status and Resource Monitoring

Although direct login access to the server is not provided, there are several facilities that enable users to monitor the current state of the instance:

`http://YOURSERVER/mod9/stat/ping`

Returns an HTTP message with the content body of `pong\r\n`

This can be useful for automated monitoring.

`http://YOURSERVER/mod9/stat/top`

Displays the currently active processes and their resource usage.

`http://YOURSERVER/mod9/stat/free`

Displays the current allocation of memory resources.

`http://YOURSERVER/mod9/stat/df`

Displays the currently available disk space.

### 4. Alignment Server Maintenance

Several facilities are provided to alter the state of a running server:

`http://YOURSERVER/mod9/stat/kill?pid=N`

Attempt to kill the process specified by the required argument `pid=N`. The process ID can be determined by inspecting logs and messages output by the Alignment API, or by monitoring `top` via the Alignment Server's status-reporting endpoint. A message is returned to indicate whether the process was successfully killed.

`http://YOURSERVER/mod9/stat/renice?pid=N&priority=N`

This is useful if your server's CPU resources are over-committed and you need to have certain jobs finish sooner than others. The required argument `pid=N` specifies the process ID number, while `priority=N` specifies a valid niceness priority value.

`http://YOURSERVER/mod9/clean/delete`

Clean up disk space taken up by temporary files. It is recommended that *recent* temporary files be retained on the server, as they are useful to aid debugging. These files are also conveniently used as a cache if you pass the same data through the server (e.g. to realign an old audio recording with a new transcription).

Optional arguments:

`atime=N`: only files at least N days or older (accessed)

`sizek=N`: only files at least N kB in size

`dryrun=1`: don't delete files, just report the total size that would be cleared

## 5. Troubleshooting

Some problems that could be encountered:

- The server is unresponsive (e.g. to `/mod9/stat/ping`). This can happen upon initial launch of an instance, due to known EC2 issues. Currently, the only solution is to reboot, which should be required only once in the lifetime of the instance.
- The disk is full, despite attempts to clean up using `/mod9/clean/delete`. Mod9 might be able to investigate; or just launch a new instance.
- CPU usage is maxed by “rogue” processes, as reported by `/mod9/stat/top`. You should try to kill these processes using `/mod9/stat/kill`, or reduce their CPU consumption by adjusting their execution priorities using `/mod9/stat/renice`. If that fails, please contact Mod9 for investigation.
- The server is otherwise in some unpleasant state. The simplest solution is to start a new instance. This is the beauty of working with virtual images: you have an exact representation of a server's initial state, from which you may launch a pristine new instance. Although Mod9 will respond to support requests as immediately as possible, it is generally easier for a user to simply launch a new instance – which would then be ready in a matter of minutes.

Some recent updates:

- When things start to go haywire, the server's disk now fills up less rapidly than before. Previously, the internal error logging was overly verbose.
- Rogue processes should be less common, thanks to the restriction of a 36-hr process timeout (wall clock, not CPU clock). If you have really huge inputs that require this much time, you're probably doing it wrong...

## 6. Contact

This server was built and configured by Arlo Faria:

[support@mod9.com](mailto:support@mod9.com)

+1 (510) 705-3223