An Introduction to Version 1.x
Version numbering

1.23.04-XXXX

- Now much easier to determine if you are using latest version
- Compatibility between major equal major version numbers
- Easy dnf updates between releases of same major version
- Limit any breaking changes
What is ClusterVisor?

ClusterVisor provides easy to use interface to **deploy**, **provision**, **manage**, **monitor**, and **maintain** your cluster for its lifetime.
ClusterVisor 1.0 is a significant upgrade from the original version

A NEW AND IMPROVED STATS ENGINE
A REVAMPED WEB UI
INTEGRATION WITH SLURM
NODE PROVISIONING ENHANCEMENTS

And many, many more ....
clustervisor

- Flexible Delivery
- Configuration Management
- Statistics & Monitoring
- Rack Diagramming
- Integration with SLURM
- Customizable Dashboards
- Provisioning
- User Management
- Command Line Tools
STATISTICS & MONITORING
FLEXIBLE DELIVERY
CONFIGURATION MANAGEMENT
RACK DIAGRAMMING
INTEGRATION WITH SLURM
PROVISIONING
CUSTOMIZABLE DASHBOARDS
USER MANAGEMENT
COMMAND LINE TOOLS

clustervisor
advanced clustering technologies, inc.
Delivery Options

Installed on head node
- Traditional approach
- Head node used for managing cluster and user logins

Dedicated ClusterVisor appliance node
- Dedicated system for managing cluster
- Provides easy disaster recovery
- Simple to manage and update
Why an appliance?

**Disaster Recovery**
The appliance holds images for all nodes including compute, login, storage, etc. Makes it easy to restore any node (including infrastructure ones) in your cluster when something goes wrong.

**Process Isolation**
Separate the functions of cluster management, and user logins. Appliance runs limited shell and regular users don’t have access. Limits issues created by users doing inappropriate things on the head node.

**Maintenance**
Based on standardized hardware and software configuration. Makes upgrades, and hardware maintenance easy. If the appliance fails, remove hard drives and install them into a replacement system, and you are back up and running again.
ClusterVisor Appliance

- Dedicated 1U server with redundant power, optimized for running ClusterVisor on systems with up to 100s of nodes
- Runs a minimal locked down version Rocky Linux 9.x with a dedicated management Web GUI.
- All user customer specific data stored on removable drives in the front of the chassis
- Much lower cost compared to standard head nodes

The ClusterVisor appliance runs all the essential services needed to make your cluster function: DHCP, DNS, PXE, LDAP, Node Provision, Monitoring and Alerting
Network design

Corporate Network

Login Node(s)

ClusterVisor Appliance

Cluster Network

Compute Node

Compute Node

Compute Node

Compute Node

Compute Node

Compute Node

Compute Node

advanced clustering technologies, inc.
clustervisor

- Flexible Delivery
- Configuration Management
- Statistics & Monitoring
- Rack Diagramming
- Integration with SLURM
- Customizable Dashboards
- Provisioning
- User Management
- Command Line Tools

advanced clustering technologies, inc.
ClusterVisor’s powerful **Configuration** tool stores your entire cluster’s configuration, making it possible to manage every part of the cluster from within this interface.
Configuring Nodes, etc.

- Everything is organized as parts of a collection. At right is the full list of collection types and a description of each.

- Each collection has unique schema of fields that store data pertinent to that type.

- Collections are not only hardware elements in the cluster but configuration elements as well.

<table>
<thead>
<tr>
<th>Collection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>config</td>
<td>Global configuration variables and settings</td>
</tr>
<tr>
<td>appliance</td>
<td>ClusterVisor appliance nodes</td>
</tr>
<tr>
<td>node</td>
<td>Nodes in your cluster (compute, login, head, etc.)</td>
</tr>
<tr>
<td>netboot</td>
<td>iPXE network boot options</td>
</tr>
<tr>
<td>cloner_image</td>
<td>A cloner image to deploy to nodes</td>
</tr>
<tr>
<td>cloner_disk_layout</td>
<td>Partition, format, RAID options for disks to be used during cloning</td>
</tr>
<tr>
<td>switch</td>
<td>Network switch devices</td>
</tr>
<tr>
<td>chassis</td>
<td>An enclosure for multi-node blade style systems</td>
</tr>
<tr>
<td>device</td>
<td>Any other device you’d like to include in your cluster</td>
</tr>
<tr>
<td>group</td>
<td>A collection of other devices (usually nodes)</td>
</tr>
<tr>
<td>monitoring_rule</td>
<td>A set of tests to apply to device stats</td>
</tr>
<tr>
<td>monitoring_action</td>
<td>Actions to perform when monitoring rules are true (email, scripts, etc.)</td>
</tr>
<tr>
<td>computed_stat</td>
<td>Take existing stats and perform additional operations on them</td>
</tr>
<tr>
<td>group_stat</td>
<td>Rollup stats for all members of a group</td>
</tr>
</tbody>
</table>
Configuration

- Configuration elements can be managed via the command line or the Web GUI.
- The Web GUI is easier for discovery to see all the possible fields and values available.
- The command line can be faster to edit once you are familiar with the syntax and options.
- When editing via the command line, the element’s entire config will be present to you as YAML.
Configuration Plugins

- Certain collections support plugins (currently node and appliance)

- Plugins are used to take data stored in ClusterVisor and write that configuration to nodes.

- For example, the networking plugin takes all the networking configuration for the nodes and configures those interfaces for you.

- Plugins can be enabled or disabled depending on if you want ClusterVisor to manage that part of the configuration for you.

- For example, it may make sense to disable the chrony plugin if you have a very advanced time server setup and would prefer to manage that configuration on your own.

- Certain node types will have different plugins enabled (compute nodes will not need all the plugins as a head or login node)

- If no plugins are enabled on nodes, no configuration changes will ever be done.
<table>
<thead>
<tr>
<th>Collection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>chronyd</td>
<td>Setup/configure the chrony daemon to synchronize a system’s date/time</td>
</tr>
<tr>
<td>cloner_server</td>
<td>Manage the cloner server (rsyncd)</td>
</tr>
<tr>
<td>dhcp_dns_server</td>
<td>Manage the dnsmasq DHCP/DNS server</td>
</tr>
<tr>
<td>efiboot</td>
<td>Set the EFI boot order of the system</td>
</tr>
<tr>
<td>filesystemsync</td>
<td>Synchronizes files to other nodes in the cluster</td>
</tr>
<tr>
<td>fstab</td>
<td>Write the extra mount entries into the fstab on a node</td>
</tr>
<tr>
<td>hosts</td>
<td>Create the /etc/hosts file on the node</td>
</tr>
<tr>
<td>ipmisettings</td>
<td>Configure the IPMI interface and/or authentication on the node</td>
</tr>
<tr>
<td>ldapauth</td>
<td>Configure LDAP authentication on the node</td>
</tr>
<tr>
<td>networking</td>
<td>Create all the networking configuration files on the node</td>
</tr>
<tr>
<td>nfsserver</td>
<td>Configure the NFS exports for this node</td>
</tr>
<tr>
<td>resolv_conf</td>
<td>Set the nameserver in the /etc/resolv.conf for a node</td>
</tr>
<tr>
<td>rsyslog</td>
<td>Configure rsyslog on the node</td>
</tr>
<tr>
<td>slurmclient</td>
<td>Manage the SLURM scheduling client on a node</td>
</tr>
<tr>
<td>slurmserver</td>
<td>Manage the SLURM scheduling server on a management node</td>
</tr>
<tr>
<td>ssh</td>
<td>Manage the /etc/ssh/ssh_known_hosts file on a node</td>
</tr>
</tbody>
</table>
## Plugins

<table>
<thead>
<tr>
<th>Collection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sudoers</td>
<td>Handle the sudoers settings on the node</td>
</tr>
<tr>
<td>timezone</td>
<td>Set a node's timezone</td>
</tr>
<tr>
<td>yumrepo</td>
<td>Manage the yum repositories on a node</td>
</tr>
<tr>
<td>yumrepo_server</td>
<td>Run a local yum repository on the node</td>
</tr>
<tr>
<td>limits</td>
<td>Set memlock and stack limits needed for IB networks</td>
</tr>
<tr>
<td>lmod</td>
<td>Setup and configure LMod</td>
</tr>
<tr>
<td>selinux</td>
<td>Configure Security Enhanced Linux (SELinux)</td>
</tr>
<tr>
<td>serialconsole</td>
<td>Configure Linux's serial console redirection</td>
</tr>
<tr>
<td>bootoptions</td>
<td>Setup various boot configuration options</td>
</tr>
<tr>
<td>tuned</td>
<td>Set the node's tuned profile</td>
</tr>
</tbody>
</table>
Stat/monitoring system

- The entire stats and monitoring subsystem was totally revamped in the 1.0 release

- Fixes many bugs and offers a lot more features

- Better retention and rollover of older stats

- Easy to view history and of stats per node, group, by job, etc

- Monitoring rules much more powerful with a full logic engine
Stats and monitoring

• Stats are collected on the node via a plugin architecture

• Any stat collected can be used by the monitoring rules engine and perform actions

  • Rule example: node is down, temperature is too high, RAID array degraded, etc.

  • Action examples: send emails, run scripts

• You can view history

• You can compare stats among nodes

<table>
<thead>
<tr>
<th>interval</th>
<th>retention period</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 seconds</td>
<td>last 2 weeks</td>
</tr>
<tr>
<td>5 minutes</td>
<td>last 2 months</td>
</tr>
<tr>
<td>15 minutes</td>
<td>last 3 months</td>
</tr>
<tr>
<td>1 hour</td>
<td>last 5 years</td>
</tr>
<tr>
<td>Collection</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>cpu</td>
<td>CPU usage and load information</td>
</tr>
<tr>
<td>cvclient</td>
<td>ClusterVisor version installed on the node</td>
</tr>
<tr>
<td>disks</td>
<td>Details about each disk and filesystem</td>
</tr>
<tr>
<td>firmware</td>
<td>BIOS and BMC firmware versions</td>
</tr>
<tr>
<td>infiniband</td>
<td>InfiniBand or Omni-Path fabric device performance and error counters</td>
</tr>
<tr>
<td>intelssd</td>
<td>Health info specific to intel branded datacenter SSDs</td>
</tr>
<tr>
<td>ipmi</td>
<td>Temperature, voltage, fans from IPMI</td>
</tr>
<tr>
<td>md</td>
<td>Software RAID status including if any arrays are degraded</td>
</tr>
<tr>
<td>megaraid</td>
<td>Hardware RAID status from LSI/Avago/Broadcom RAID controllers</td>
</tr>
<tr>
<td>mem</td>
<td>Memory utilization and ECC error counts</td>
</tr>
<tr>
<td>net</td>
<td>Network stats by Ethernet device</td>
</tr>
<tr>
<td>nfs</td>
<td>Status of NFS mounts and exports</td>
</tr>
<tr>
<td>ntp</td>
<td>Date / time info</td>
</tr>
<tr>
<td>nvidia</td>
<td>GPU information including power, ECC, utilization, memory</td>
</tr>
<tr>
<td>power</td>
<td>Power consumption on the node</td>
</tr>
<tr>
<td>system</td>
<td>Kernel version, uptime, etc.</td>
</tr>
</tbody>
</table>
Stats and Monitoring

ClusterVisor has a full statistics and monitoring engine.
Stats and Monitoring

ClusterVisor has a full statistics and monitoring engine.
ClusterVisor has a full statistics and monitoring engine.
Cluster Health Checks

• One of ClusterVisor’s most important features is its ability to act as a health monitor for all aspects of your cluster via our monitoring rule system.

• Any stat collected can be turned into a monitoring rule. Examples:
  • Node not responding, InfiniBand down, RAID arrays degraded, ECC errors, InfiniBand down, Low disk space, etc.
Monitoring rules

- Templates for many common rules are available to make adding rules for your system easier.
- Rules are run at 30 second intervals; rules can be defined to run once, always, or on clear.
- The monitoring rule can be as simple or as complicated as you like. The Web GUI has a full GUI-based rule editor.
- Customizable actions including emails, or running scripts on your system.
• Full queryable history of all monitoring rules
• Stat snapshot taking of all values when rule fails
• Helpful for diagnosing problems on your system
• Can acknowledge rules, so you won’t get notified in the future
Coming soon

• Ability to query and collect stats for non-node devices

• Currently planned:

  • PDUs, UPSs, Switches, Storage arrays

  • Command line tools and API so you can inject your own stats into ClusterVisor
Rack Diagrams

- Old versions of ClusterVisor required editing text values to create diagrams.
- With the 1.x release, rack editing is all drag and drop.
• Stats and monitoring rules can be applied to rack diagrams to create visual heatmaps

• Multiple racks and rows can be created to simulate the real world setup of your datacenter
ClusterVisor now can pull data from your SLURM installation.

- No data duplication, uses the existing SLURM REST API to pull data from slurmctld and slurmdbd.
SLURM integration

- Easily query job history information
- Query by time and see jobs on an interactive timescale graph to see how the system is operating
SLURM integration

- Connect job information with ClusterVisor stats
- Easily query any collected stat during a job run
- CPU utilization, memory usage, temperatures, power consumption, etc.
- Very helpful for diagnosing problems like “my job is slow”
Dashboards in ClusterVisor are customizable by the admin to show the information that they find most important on their cluster.

- The dashboards are useful to show you cluster-wide information and to compare multiple nodes or devices against each other.

- You can create as many dashboards as you wish, with a different focus on each dashboard.
Dashboards

- Dashboards contains multiple widgets that are configurable with the stats you care about:
  - Stat tables
  - Graphs
  - Gauges
  - Rack diagrams
  - Scheduler info
- You can adjust the dashboard to view different time windows

- The Dashboard section also lets you run commands across multiple nodes, and power cycle selected devices
clustervisor

- FLEXIBLE DELIVERY
- CONFIGURATION MANAGEMENT
- STATISTICS & MONITORING
- RACK DIAGRAMMING
- INTEGRATION WITH SLURM
- CUSTOMIZABLE DASHBOARDS
- USER MANAGEMENT
- COMMAND LINE TOOLS

advanced clustering technologies, inc.
Provisioning

- ClusterVisor’s provisioning subsystem is called “cloner”
- Full suite of tools to create images and take images from existing nodes
- Images can be updated from existing running systems
- Image data separate from disk configuration
- Image can be multicast to multiple nodes simultaneously
- From bare-metal to working node in just minutes

<table>
<thead>
<tr>
<th>Image</th>
<th>Cloner Image</th>
<th>Path</th>
<th>Detect...</th>
<th>Source...</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>node</td>
<td>cloner_image.node</td>
<td>/clustervisor/images/node</td>
<td>rocky 8.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>node-rocky8</td>
<td>cloner_image.node-rocky8</td>
<td>/clustervisor/images/node-rocky8</td>
<td>rocky 8.6</td>
<td>rocky-8.6</td>
<td>node</td>
</tr>
<tr>
<td>node-rocky9</td>
<td>cloner_image.node-rocky9</td>
<td>/clustervisor/images/node-rocky9</td>
<td>rocky 9.1</td>
<td>rocky-9.1</td>
<td>node</td>
</tr>
<tr>
<td>test</td>
<td>cloner_image.test</td>
<td>/clustervisor/images/test</td>
<td>rocky 9.1</td>
<td>rocky-9.1</td>
<td>node</td>
</tr>
</tbody>
</table>
Provisioning

• New in the 1.x release:
  • Updated to support EL8 and EL9 distros: Rocky, RedHat, Alma, etc
  • Upload distro ISOs and then create new images from the uploaded distro
  • No need to install a node first
  • Templates for: node, login, storage, etc

Create a new image from distro

- Image name: my node
- Distro: rocky-9.1
- Role: node - Create a standard ACT compute node
- Root password
- Client: Install clustervisor-client
- Updates: Install software updates from the Internet (this could increment the minor version number of the distro - i.e. 8.1 to 8.3)
- Cleanup: Cleanup the image if fails to create

Create image  Cancel
Provisioning

• Cloner combined with a ClusterVisor Appliance node:
  • Install your login node, and storage nodes directly from the appliance
  • Special supported added for non-internet connected systems for highly secure environments
  • Take backup images of all your systems in case of a hardware failure or software issue
  • Easily restore not only compute nodes but login and storage nodes too
clustervisor

- Flexible Delivery
- Configuration Management
- Statistics & Monitoring
- Rack Diagramming
- Integration with SLURM
- Customizable Dashboards
- Provisioning
- User Management
- Command Line Tools

Advanced Clustering Technologies, Inc.
Users and groups

- ClusterVisor can run and manage an LDAP server for use inside your cluster.
- Web and command line tools for making users, home directories, SSH keys, and Slurm accounting users.
- Internal LDAP support is optional, and can use external authentication if desired (Active Directory, etc).

**LDAP Users/Groups**

<table>
<thead>
<tr>
<th>User name</th>
<th>Full name</th>
<th>UID number</th>
<th>Primary</th>
</tr>
</thead>
<tbody>
<tr>
<td>demo_user</td>
<td>Demo User</td>
<td>99998</td>
<td>99998</td>
</tr>
<tr>
<td>test_user</td>
<td>Test User</td>
<td>199998</td>
<td>1000</td>
</tr>
<tr>
<td>cv-admin</td>
<td>cv-admin</td>
<td>199999</td>
<td>10000</td>
</tr>
</tbody>
</table>
Access Control

- ClusterVisor supports setting users as “admins” so they have control to make changes on your cluster.
- With “admin” privileges they can login to ClusterVisor and edit configuration, dashboards, make users, etc.
- ClusterVisor also has the ability manage your sudo setup for privilege based access on login and/or compute nodes
Clustervisor

flexible delivery
configuration management
statistics & monitoring
rack diagramming
integration with slurm
customizable dashboards
provisioning
user management
command line tools

advanced clustering technologies, inc.
<table>
<thead>
<tr>
<th>command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cv-authsync</td>
<td>Sync password / group files across nodes (deprecated: use LDAP system instead)</td>
</tr>
<tr>
<td>cv-cloner</td>
<td>Create / update cloner images and disk layouts</td>
</tr>
<tr>
<td>cv-commit</td>
<td>Commit changes in the ClusterVisor database to the nodes</td>
</tr>
<tr>
<td>cv-conf</td>
<td>Edit ClusterVisor configuration settings</td>
</tr>
<tr>
<td>cv-console</td>
<td>Connect to a node’s IPMI serial console for remote debugging</td>
</tr>
<tr>
<td>cv-cp</td>
<td>Copy files in parallel to multiple nodes in the cluster</td>
</tr>
<tr>
<td>cv-exec</td>
<td>Execute commands in parallel across the cluster</td>
</tr>
<tr>
<td>cv-identify</td>
<td>Turn on the identifier LED to help find systems in the rack / datacenter</td>
</tr>
<tr>
<td>cv-ipmitool</td>
<td>Issue arbitrary IPMI commands to devices on the cluster</td>
</tr>
<tr>
<td>cv-netboot</td>
<td>Change how the node will start up on it’s next boot</td>
</tr>
<tr>
<td>cv-nodenames</td>
<td>Output node names (useful for scripting)</td>
</tr>
<tr>
<td>cv-power</td>
<td>Power control a node via IPMI (power on, off, reboot, etc.)</td>
</tr>
<tr>
<td>cv-reconfigure</td>
<td>Take all ClusterVisor configuration for this node and reapply it to the node</td>
</tr>
<tr>
<td>cv-sel</td>
<td>View the nodes IPMI event log</td>
</tr>
<tr>
<td>cv-sensor</td>
<td>Query the nodes IPMI sensors</td>
</tr>
<tr>
<td>cv-sshkey</td>
<td>Help create user authorized keys</td>
</tr>
<tr>
<td>command</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>cv-stats</td>
<td>Query ClusterVisor collected stats</td>
</tr>
<tr>
<td>cv-useradm</td>
<td>Add / edit LDAP users</td>
</tr>
<tr>
<td>cv-db-image</td>
<td>Backup and restore ClusterVisor configuration</td>
</tr>
<tr>
<td>cv-distro</td>
<td>Upload Linux distribution to ClusterVisor</td>
</tr>
<tr>
<td>cv-image</td>
<td>Create / edit cloner images from distros</td>
</tr>
<tr>
<td>cv-cp</td>
<td>Copy files in parallel to multiple nodes in the cluster</td>
</tr>
<tr>
<td>cv-statsadm</td>
<td>Manage the stats database</td>
</tr>
</tbody>
</table>
A research team at Caltech is among the first customers to utilize ClusterVisor 1.0 in their HPC cluster.

The team started using features in February such as:

- Creating custom dashboards
- Relying on user management to add new accounts and recover user passwords
- Using rack layout to see how each job is distributed across the cluster
Case Study

"The one feature I really wanted to mention and say how it’s really important to us is for a particular job that we are running on the cluster, we are able to track the RAM memory as a function of time. This is wonderful, not only to see how our software runs on the cluster, but also to test our software. We are talking to our peers in the U.S. and Europe, and nobody among them has these tools. We share our excitement with other people and say we have this feature. This is great."

Ivan Maliyov
Postdoctoral Research
Caltech
Next Steps

Visit us at https://www.advancedclustering.com/products/software/clustervisor/ to:

• Stay tuned for our next webinar that features a deep dive into the technical aspects of ClusterVisor 1.0.

• Download our PDF on How to Upgrade from ClusterVisor to ClusterVisor 1.0.

• Contact us to ask questions or talk about adding ClusterVisor to your HPC cluster.
Stay Up to Date

Join our clustervisor-announce mailing list to receive updates and announcements about ClusterVisor by Advanced Clustering.

https://lists.advancedclustering.com/mailman/listinfo/clustervisor-announce

You can also look back at the archives to see what updates have already been shared.
Contact Us

866-802-8222

info@advancedclustering.com

advancedclustering.com