Webinar will begin shortly

advanced clustering
technologies, inc.

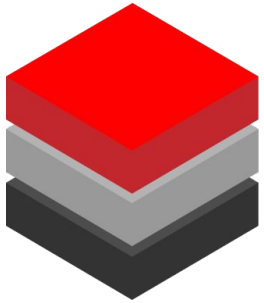
Webinar Housekeeping

General info:

- Everyone except the presenters are on mute
- Webinar will be recorded, and available for replay shortly after the end of the event

Questions:

- If you have questions use the Q&A button in Zoom to input your question
- Questions will be answered live time permitting, or followed up with after the webinar concludes



advanced clustering
technologies, inc.

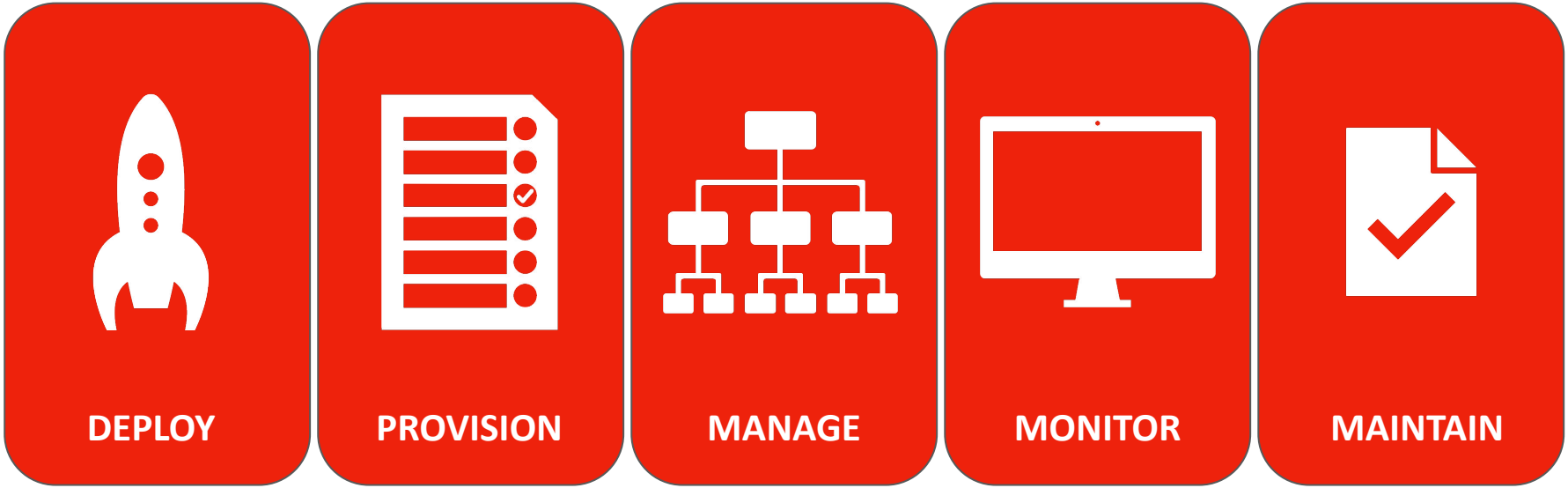
clustervisor

Introducing the first 2.x version (2.25.XX)

advanced clustering
technologies, inc.

What is ClusterVisor?

ClusterVisor provides an easy to use interface to **deploy**, **provision**, **manage**, **monitor**, and **maintain** your cluster for its lifetime.



Availability and Licensing

Availability

- ClusterVisor is available now as a complete cluster management software solution for our turn-key cluster customers and deployment on non- Advanced Clustering hardware
- There are two installation options: on the head/login node or the preferred method - installation on the dedicated ClusterVisor appliance

Licensing

- ClusterVisor is licensed by the total quantity of nodes in the system (for 1, 3 or 5 year terms)
- An active license is required to get support, updates and new versions, but software will not stop working when licenses expire

Contact us at info@advancedclustering.com to discuss an implementation plan!

advanced clustering technologies, inc.

ClusterVisor Versioning

Manual upgrade procedure to move between major versions

- Major updates to features
- Possible major changes to processes or data storage

Compatibility within the same major version number:

- Easy update for new month/builds between releases
- Limited changes needed to configuration data or images
- No major breaking changes

advanced clustering technologies, inc.

1.x to 2.x upgrade

- Great care was put into not breaking anything for existing 1.x customers wanting to upgrade to 2.x
- Migration tool exists to easily bring your system from 1.x to 2.x
- Some features have been deprecated and may be removed in the future releases, but currently left in place to make upgrade as painless as possible

If you have a valid unexpired 1.x license you can upgrade to 2.x for no additional cost

What's the status for 1.x

- The 1.x branch will continue to exist in its current state
- Major bug fixes will still be evaluated and fixed when deemed necessary
- No new features
- No new operating system version support

We encourage everyone to plan an upgrade to 2.x

advanced clustering technologies, inc.

What's new in 2.x

- Provisioner system replaces cloner
- Easier assignment of new nodes
- New plugin architecture
- Connections / PDU support
- Security improvements
- Other improvements and bug fixes

Biggest ClusterVisor upgrade to date

Over 1,000 code commits, and changed over 55,000 lines of code

Provisioner

A replacement for cloner

- Why replace cloner?
- Cloner is the cause of most support tickets related to ClusterVisor
  - When it fails the errors are obtuse debugging is difficult
- Provisioner is a major re-architecting of the entire node boot process

Provisioner design goals

- Make as remote friendly as possible
  - Nobody likes to go to the data center and hook up a monitor/keyboard
  - Easily monitor the node booting progress from the CV web UI
  - Remotely connect to the console while provisioning to watch the process
  - Each boot session is recorded with a full log of all boot events
- Remove reliance on GRUB, and tools like grub2-install and /boot/loader/entries files whenever possible
  - These change between minor distro versions and caused numerous compatibility issues with cloner
  - Proved to be very difficult to debug and diagnose problems
- Make editing and setting disk layout/partition info way more user friendly
- Help alleviate drift from source image and what's installed on the node
- Stateless/diskless nodes need to behave much more like stateful ones
- Increase security / encryption during node boot/provision process

Provisioner terms to know

- **Provisioner boot environment** - the small pre-boot environment that is started before the distro OS. Responsible for getting everything setup/configured before pivoting to the real distro operating system
- **Provisioner image** - a complete copy of the operating system filesystem that will be deployed on the node
- **Disk layout -** Which of the node's disks will be used, how they will be partitioned, software raid arrays created, LVMs created, etc

Fundamental changes from cloner

Cloner (old)

- You had to manually put nodes into cloner netboot
- When cloner would run, it would always totally erase drives and re-copy full image
- When complete would change to booting from disk and reboot

Provisioner (new)

- Nodes always boot into provisioner boot environment
- Provisioner checks the disk layout and operating system to make sure they are correct
- If anything is incorrect: fixes things by partitioning disks (when necessary) and syncing the image (only changes)
- Pivots to operating system image and continues to boot as normal
- It's now a better idea to edit the image not the node

Advantages of this change

- Multiple boots not required, provisioner will set everything up and pivot to the operating system all in one boot
- On every boot provisioner checks the nodes are synchronized with the image - if not it will resync it (this is a configurable option)
- Installing updates or reconfigurations much easier:
  - Change the image by adding/updating software or making configuration modifications
  - Reboot node(s), starts the Provisioner boot environment
  - Node will not be synchronized with the image, so nprovisioner will proceed to syncing only the changes
  - Node will pivot to full operating system image and continue booting
- Stateless nodes work in the same way
  - On boot a ramdisk is created to hold the operating system
  - This newly created ramdisk will be empty, so provisioner will synchronize the image to the ramdisk
  - Pivot to operating system installed in ramdisk

Provisioner node boot flow

New standard boot process for every node

Pre-Boot

1. Node firmware set to network boot
2. Via DHCP/TFTP downloads iPXE from cv-serverd host
3. Sends its MAC address to server, creates a boot session ID
4. If MAC address known move to next step, if unknown: pauses for 30 seconds and repeats step 3

Provisioner boot

1. Uses cluster specific HTTPS client certificate embedded into iPXE server to download pieces
2. Downloads kernel for image assigned to the node, pulls down kernel modules, provisioner code, and rest of provisioner boot environment

Boot environment

1. Open a connection to cv-console-serverd process on cv-serverd host (allows remote watching of entire boot process)
2. Setup initial boot environment

Provisioner note boot flow, cont'd

New standard boot process for every node

Disk layout

1. Compare the ClusterVisor defined disk layout (partitions, mirrors, LVMs) to the node's current state
2. If the disk layout doesn't match the system - re-apply the layout (create partitions, raid arrays, format filesystems, etc)
3. Start MD devices, LVMs, and mount all filesystems

Image

1. Compare the source image stored with ClusterVisor with the operating system already on disk
2. If the operating system doesn't match start an rsync process over SSH with the cv-serverd host to get node's operating system up to date

Plugins

1. Run all ClusterVisor plugins assigned to the node
2. Configure things like networking, daemons, nfs mounts, etc

Provisioner note boot flow, cont'd

New standard boot process for every node

Boot loader

1. Install a fallback bootloader on the system
2. Only used when cv-serverd host is down - allows a node to boot off it's hard drive when network is down, or cv-serverd host isn't running

Pivot

1. Stop connection to cv-console-serverd
2. Cleanup boot environment
3. Pivot to installed operating system and call systemd init process

Distro boot (finish)

1. Node boots as normal using all standard systemd units, configuration, etc

Provisioner images

- Similar to cloner images but with some added features
  - Tools available to migrate a cloner image to a provisioner image
- Image names no longer tied to directory on filesystem, so can include spaces, special characters, etc.
- Multiple branches (copies) can be made under the same image
  - Great for testing new versions of software or configuration changes
  - Can switch between branches and decide which is the current “live” branch
  - Should prevent names like `cloner_image.node_pre-2022-06_ver1`
- Same image can be used for stateful and stateless, no rebuilding of stateless images after changes
- Full audit logs recorded on image edits and changes

Disk layouts

- Complete revamp into “simple” and “advanced”
- Strongly encourage simple disk layout for most customers
  - Supports single disks and software RAID mirrors
  - Everything is LVM behind the scenes but you don’t need to set that up
  - Easy to use graphical interface for configure partitions and sizes
- Advanced mode = full featured, but more difficult to create (only use when simple mode doesn’t fit requirements)
- Disk detection is now by vendor, device type, size, S/N or path
  - Disks are defined in CV as disk[0-x] and will be translated in realtime to the matching device
  - One boot disk0 could be nvme0n1 and another nvme4n1 - this will work correctly based on the other detection fields (vendor, device type, size, etc)
- Size matching is fuzzy with a user defined percent variance (120GB +/- 5%)

Provisioner demo

advanced clustering
technologies, inc.

New nodes

Simplifying deploying node replacements or expansions

- When an unknown node is booted it will be in a loop waiting to be identified
- Sends in metadata to server (like MAC address) to help identify the node
- A simple UI interface allows you to assign it to a defined node
- Node will then provision itself to the node you assigned

Unassigned node demo

advanced clustering
technologies, inc.

New config plugin system

Architecture change

- Complete re-write of how the config plugin system works
- Offers much more flexibility in what a config plugin can do
- Better handling of systemd units
- Can be smarter and not just always overwrite config files
- Works with provisioner in the provisioner boot environment to run all plugins before operating system boot

advanced clustering
technologies, inc.

Connections

Defining the connections between devices

- Devices can now have connections to Switches and PDUs
- Switch connections are more informational for now
- PDU connections allow for managed PDU control

advanced clustering technologies, inc.

PDU connections

- With PDU connections, we can now support power control via managed or switched PDUs in addition to just IPMI
- Changes to all devices (nodes, switches, etc)
  - Now have a power supply definition (how many power supplies and what their names are)
  - Has a default power method (IPMI, PDU)
- New PDU device type
  - Has a list of outlets
- Connects with external-stat-collector plugins to allow power control of outlets
  - Currently we support Eaton G4 PDUs via their REST API
  - More PDUs to come based on customer feedback
- UI changes to make connections
- `cv-power` command re-created to fully support PDU power control

Connections demo

advanced clustering
technologies, inc.

Security improvements

- Networking booting done all over HTTPs with client certification authentication
- Some PXE options only available if MAC address is registered in CV
- No more rsyncd, all rsync is now done over SSH
- Audit database now part of ClusterVisor, each audit event contains
  - Timestamp
  - Username & UID
  - Host (where command was run)
  - Kind (type of audit event; e.g. config, provisioner, cli, etc)
  - Method (a subtype of kind; e.g. add, replace, delete, cv-exec, etc)
  - References (which devices, images, etc)
  - Message (a text stream of what happened)
  - Extras (optional recording, details of changes)
- `cv-audit` command line tool to view audit entries
- `cv-record` command to record all cli activity and store in the audit log

Other changes and bugfixes

- Web UI improvements
  - Additional configuration widgets (radio lists, dynamic options, disk partitioner)
  - New sections: cloudflare, logs / auditing
  - Power control includes PDU support
  - System cockpit modules in sub-menu
- Cloudflare support
  - Easy setup of cloudflare zero trust connection to allow ACT support to help with your system
- New configuration plugins
  - limits (custom ulimits)
  - Provisioner_server (required for provisioner)
- New stat plugins
  - SolidigmSSD and MicronSSD specific data collection
  - Systemd Units (get alerts when certain units aren't running)
- Many tweaks and bugfixes

Ready to upgrade?

advanced clustering
technologies, inc.

1.x to 2.x upgrade

- Migration tool will be used that will step you through the process and automate everything
- Available from https://repo.advancedclustering.com/clustervisor/cv-migrate.tgz

```
# curl -o cv-migrate.tgz \
https://repo.advancedclustering.com/clustervisor/cv-migrate.tar.gz

# tar xfz cv-migrate.tgz
# cd cv-migrate-1-to-2-2.25.09-6022/
# ./cv-migrate-1-to-2

[ … follow on screen instructions … ]
```

Mailing List

Find out about new versions

Announce mailing list has limited posts (1-2 per month)

Only announcements about new versions of CV with changelogs

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

Q&A

advanced clustering
technologies, inc.