The Evolution of High Availability

What is ‘M3’?
At a Glance

• Intelligent Availability™
  • Keep your servers running, everything else is secondary.
  • Plug-able, automatic hardware discovery / monitoring
  • N-database redundancy with automatic split-brain recovery

• Built for the modern HA stack
  • RHEL / CentOS 8
  • Pacemaker 2 (finally!), DRBD 9, Corosync 3 / Kronosnet 1

• Ground-up rewrite
  • Database / “jobs” based system, more secure
  • Built around systemd, RPM based, Web based
  • Templated, translatable on the fly / full UTF8 support
Overview

- Software platform
  - Hardware agnostic
  - Infrastructure blueprints provided for full stack redundancy

- Designed for the “Edge of the network”
  - Hands-off, offline operation. Fully self-contained.
  - Integrated DHCP/PXE on isolated LAN for offline machine rebuild.
  - Intelligent availability logic designed with the understanding that faults may not be repaired for some time.
  - Alerts admin, doesn’t expect a response.

- Cares about the hosted servers, everything else is secondary.
What it does

- *Keeps hosted servers running as much as possible.*
  - Everything is secondary to this goal.
  - Automatic load shed, emergency shut down and restart on power and temperature events.
  - Adaptive hosting based on overall node health scores.
  - Allows *any* hardware component to be removed and replaced without needing a maintenance window.
  - Isolated networking for bandwidth and security management.
Features

- Entirely web-based UI with live-update (jQuery/AJAX)
  - It’s an appliance
- Third-party plugin support via simple PostgreSQL DB connection.
- Live-Migration
  - Custom Pacemaker resource agent, single pacemaker config line per VM.
  - DRBD resource per VM run in single-primary mode;
    - RA switches resources to dual-primary, VM migrates, resource(s) reset to single-primary
- Minimizes CIB changes
  - All migration logic / DRBD management contained in custom RA.
  - DRBD fencing is traditional stonith, no location constraints.
- NO DLM! (sorry, but not sorry)
  - File management handled in the Anvil! database / daemon based syncing.
Integrated DR

- Optional “Third Node” for disaster recovery.
- Two operating modes, two sync methods.
  - Can run continuously connected or automatic, timed “boot, sync, shutdown” mode.
  - Can run asynchronous or synchronous.
    - BSS mode allows for protection against crypto-locker attacks.
    - Async allows for more physical distance (higher latency) without performance cost to HA nodes.
- Optional support for DRBD Proxy possible, not yet implemented
Network Features

- Matches MAC addresses to IPs for easier VM tracking.
  - Automatic download/parse/update of OUI data.
  - Automatic, periodic network ping sweeps to create a list of IP address to MAC addresses.
- Future plan to use same data to automatically find and configure "foundation pack" equipment. IE: Swap a failed UPS and have the replacement discovered and configured.
What’s Done

- All components needed for the cluster; Pacemaker RA, DRBD Fence handler.
- Database replication / resync.
- All infrastructure needed for the web interface handling.
- All parts of Striker dashboard management.
- Anvil! daemon and job handling.
- All components needed for offline management and machine rebuild (all pxe/dhcp/http/tftp, auto-update of RPM repos, etc)
What’s In Progress

• Install Manifest
  • Nodes and DR hosts can already be “initialized” (stage-1 install + anvil-node or anvil-dr RPMs installed via Striker WebUI).
  • Fence device discovery / auto-generated web configuration complete.
  • Install Manifest proper creation underway

• File Management
  • Files (ISOs, scripts) can already be uploaded via Striker and tracked in the Anvil! database)
  • Virtual Server definitions stored in the Anvil! DB.
  • Pushing out of “master” files to peers pending.

• Email Alert system (postfix backend) in progress
What’s Pending

- Needed for 3.0 release;
  - Server provisioning and management web interface.
  - Live WebUI display of cluster / server states / hardware health.
    - Back-end for this is complete, front-end jquery/ajax is pending.
  - ScanCore IA engine to be ported from M2
    - This was the last component added to M2 and will be almost a direct port.
- Targeting spring/summer 2020
The Future

- 3.1+
  - DR Management in Striker Web-UI
  - Adding front-end support for vertical scale-out over N-number logical Anvil! “pairs”.
The Futurer

- 4.0
  - ScanCore AI (already in active development)
    - Partnered with York University for two+ years now
    - SimEngine complete; Full, granular emulation of real hardware in qemu; UPSes, PDUs, IPMI (fans, temps, voltages).
    - Fail a PDU, peer UPS load spikes.
    - Fail a fan, cooled components “heat up”
    - Now able to emulate dozens / hundreds of Anvil! platforms for AI training.
    - ScanCore AI will run in parallel with M3 ScanCore to test/validate AI logic.
Questions?

https://alteeve.com
mkelly@alteeve.ca

Code; https://github.com/digimer/anvil

(Will be migrated to Clusterlabs when 3.0 is released.)