

Linux

High Availability Overview



Harald Milz
SuSE Munich GmbH

for
Usenix Annual Technical Conference
June 06–11, 1999
Monterey, CA

HA Logo © 1998, 1999 Gerard Germann

Some Basics



- Project started early 1997
- Linux High Availability HOWTO
- inspired by a commercial product
- not very bazaar like
- Information sources
 - <http://www.henge.com/~alanr/ha/>
 - <http://metalab.unc.edu/pub/Linux/ALPHA/linux-ha/High-Availability-HOWTO.htm>
- Today, more then 10 projects or solutions exist, some free (under whatever license), some commercial
- more commercial ports to follow
- Few solutions are general purpose
 - most have only web servers or firewalls in mind
- Clustering: Distributed Computing vs. High Availability

Heartbeat



- Info on <http://www.henge.com/~alanr/ha/>
- Available, usable, incomplete
- Can Do:
 - General purpose intranet communications
 - Heartbeat over multiple media (serial ring, UDP broadcast)
 - IP takeover, application API,
 - information for load balancing, easy to add heartbeat media
- Can't do > 2 nodes, multiple network interfaces
- To-Do: lots (i.a.w. Alan)

Heart



- <http://www.lemuria.org/Heart/>
- basic framework for HA implementation
- designed with web and DB servers in mind
- IPAT is standard action
- tries to be as general as possible
- keeps an distributed database of each nodes' status
- calls external scripts to take appropriate action
- no real documentation available yet i.a.w. Web page

Failover



- Info on <http://failover.othello.ch/>
- ported from Solaris
- works service oriented not server oriented
- has a concept of masters and slaves
- tcl based
- is under the GPL but may not be ported to any M\$ OS
 - this is a violation of the GPL
- can do IPAT (on virtual interface)
- issues SNMP traps (private enterprise MIB available)

Fake Redundant Server Switch



- designed to switch in backup servers on a LAN (Mail, Web and Proxy servers)
- does IPAT and ARP spoofing
- additional interface can be either a physical interface or an IP alias.
- Designed to cooperate with Heartbeat or Heart
- Can't do disk failover
- Presented at Linux Expo May 98
- Info on <http://linux.zipworld.com.au/fake/>

**REDUNDANT
LINUX**

Failoverd



- <http://www.ps-ax.com/failoverd/>
- provides rudimentary failover capability
- not 100% Linux specific, written in Perl
- launches scripts with a start/stop argument
- uses ping as HB mechanism
- uses PLIP interface as secondary HB
- GPL'd
- apparently needs more effort to mature i.a.w. Web page

Linux Virtual Server Project



- Info on <http://proxy.iinchina.net/~wensong/ipptvs/>
- Primary Goal: load balancing and HA across servers
- represents a single virtual server by a cluster of real machines
- uses single load balancer (use Fake for second load balancer & redundancy)
- uses NAT, tunneling or Direct Routing
- can't do disk failover right now.



Eddieware



- Primary Goal: Web Server Scalability
- Info on <http://www.eddiware.org>
- Availability: Erlang Public License (source code downloadable)
- derivative of Mozilla PL
- consists of
 - a load balancing HTTP server (Intelligent HTTP gateway) and
 - a Enhanced DNS server
 - two-tier server structure (Frontend & Backend)
- does IPAT and traffic rerouting after a server failure
- available for Solaris, Linux and FreeBSD (NT coming soon)



TurboLinux Cluster



- Provides HA to routers and servers (only HTTP officially supported)
- Info on <http://community.turbolinux.com/cluster/>
- based on Linux Virtual Server plus some more available bits & pieces, plus some PHT proprietary »glue code«
- kernel patch plus some more ideas from LVS project
- Management tools for the installation, configuration and management of Turbo Linux Cluster Web Server
- uses ping checks and protocol specific checks
- uses tunneling devices on the server
- price not set – likely around 1000 USD per node
- kernel patch will be GPL, tclusterd and monitoring/configuration tools licensing not yet announced

RSF-1

(Resilient Server Facility)



- Not positioned as a specialized solution –> General Purpose
- <http://www.rsi.co.uk/>
- Available commercially and via free download for non-commercial use – service contracts available
- supports 2-node, 2-service clusters (V2.0: 16 / 240)
- IP- and non-IP (RS232 and disk) heartbeat
- Can do automatic or manual failover incl. Disk
- Multi-platform capable
- supports load balancing by moving applications around
- supports Disaster Recovery by shadowing geographically dispersed sites
- ported from Solaris
- more platforms to come: AIX, NT, FreeBSD

RSF-1

Wizard Watchdog

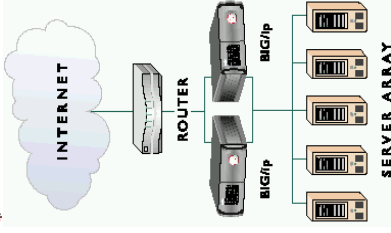


- <http://www.wizard.de>
- General Purpose, service (not server) oriented
- Freely dividable resources, such as disks, network devices, and applications to a service
- 64K services / nodes theoretical limit
- cross-platform failovers (e.g. Solaris / Linux)
- uses Network Attached Storage
- API's for
 - object specific status messages
 - monitoring and elimination of errors of any software objects
 - the support of any RAID and disk management software
 - service monitor with Java GUI & SNMP capable service agent
 - ported from Solaris – has its success stories (service cluster at popular German Sports TV channel (DSF))
 - quite costly but still cheap compared to HACMP, MC/ServiceGuard... – Linux version discounted at 40%



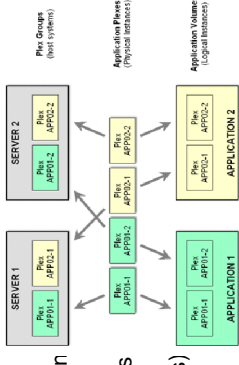
BIG/ip

- Primary goal: server load balancing
- Combined hardware/software solution (2 tier)
- Info on <http://www.f5.com>
- 3 products: BIG/ip LB, BIG/ip HA and BIG/ip HA+ (offers different service levels)
- uses NAT to protect well-known ports (like 80, 22, 23 ...)
- SSL session ID persistence
- BIG/ip HA + DNS ~ 100-120KUSD



Net/Equater

- From BSCsoft (<http://www.bscsoft.com>)
- load balancing and HA software solution
- shared nothing approach (doesn't do disk failover)
- uses the terminology
 - logical partitions (between which load is balanced)
 - plexes (application plexes distribute the load between physical servers)
- price and licensing UNK



IBM WebSphere Performance Pack

- Load Balancing across physical web servers within a WebSphere setup
- <http://www.software.ibm.com/websevers/perfpack/about.html>
- integrates
 - SecureWay Network Dispatcher (load balancing)
 - Web Traffic Express (cache & proxy)
 - Andrew File System (Transarc AFS)
- currently on Solaris, NT, AIX – Linux is due RSN.



What's missing ?

- Journalled Filesystem
 - guarantees file system meta data integrity
 - does not guarantee data integrity – application's task
 - provides short fsck times after a failure
 - should go in kernel 2.3 (Stephen Tweedie)
- Raw I/O
- marketing issue?
- Solves the fsck issue (some databases only)
- <ftp://ftp.uk.linux.org/pub/linux/sct/fs/>
- Solutions:
 - CODA (<http://coda.cs.cmu.edu/>), handles disconnected sites properly
 - GFS (<http://gfs.lcse.umn.edu/>), handles Dlock (Seagate, Mylex, ...) to implement a Distributed Lock Manager
 - SGI is porting and open sourcing XFS