

Red Hat Cluster Suite

Building high-available Applications

Thomas Neidhart

Grazer Linuxtage 2008

Outline

- 1 Einführung
 - Clusterarten
- 2 RH Cluster Suite
 - Technische Details
 - Komponenten
- 3 Anwendungsbeispiel
 - Architektur
 - Konfiguration
- 4 Tipps/Tricks

Einführung

- Was ist eigentlich ein Cluster?
- Wozu braucht man so etwas?
- Welche Probleme kann ich damit lösen?
- Wie kann ich so einen Cluster aufbauen?
- Welche Komponenten werden dazu benötigt?

Arten von Clustern

- Storage cluster - data access
- High availability - failover
- Load balancing - scalability
- High performance - parallel computation

Was bietet jetzt die RH Cluster Suite?

- kann bis zu 128 Knoten verwalten
- shared data storage (NFS, CIFS, GFS)
- bietet automatisches failover
- bietet load-balancing

Komponenten der Cluster Suite

Die Cluster Suite besteht aus mehreren Komponenten, die einzelne Bereiche abdecken:

- Cluster infrastructure (config-file management, membership management, locking, fencing)
- HA Service Management
- Administration Tools
- Linux Virtual Server (LVS - IP Load Balancing)

Grundbegriffe

Quorum

- zur Überprüfung der Funktionstüchtigkeit des Clusters
- Minimum an nötigen Stimmen
- jeder Knoten hat eine Stimme
- $\text{Quorum} = 50\% \text{ Knoten} + 1 \text{ Stimme}$
- verhindert “split-brain” Situationen

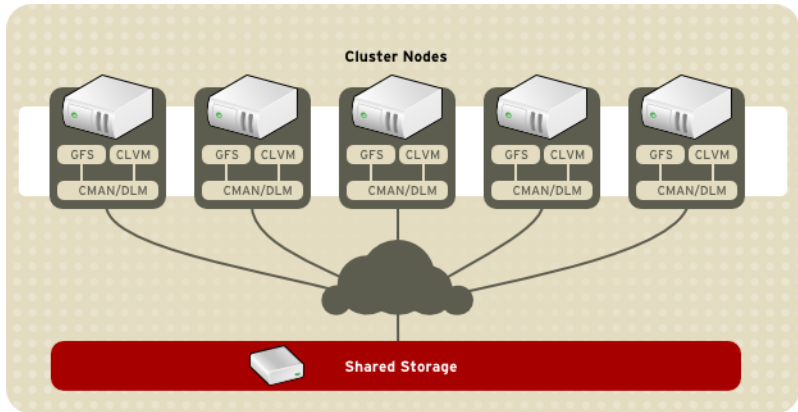
Cluster und Lock Management

- Überwacht Quorum und Cluster Membership mit Hilfe von Heartbeat Nachrichten (Ethernet)
- Synchronisiert Zugriff auf gemeinsame Ressourcen

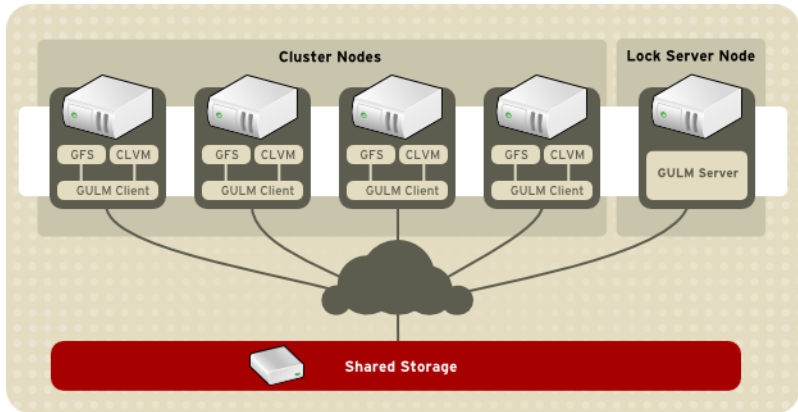
Locking Arten

- distributed: CMAN/DLM (Cluster Manager, Distributed Lock Manager)
- client-server: GULM (Grand Unified Lock Manager)

CMAN Variante



GULM Variante



Locking

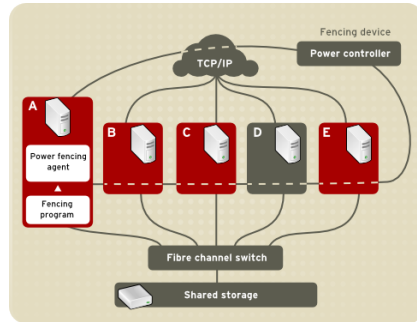
- Erlaubt Zugriff auf gemeinsame Ressourcen (GFS-Partition)
- Gewährleistet Datenintegrität
- Knoten die sich nicht im Cluster befinden, können nicht auf gemeinsame Ressourcen zugreifen

Fencing

Notwendig für die Datenintegrität des Clusters
Ist ein Knoten ausgefallen (kein Heartbeat), wird sämtliche Aktivität eingefroren, bis der Knoten erfolgreich gefenced wurde.

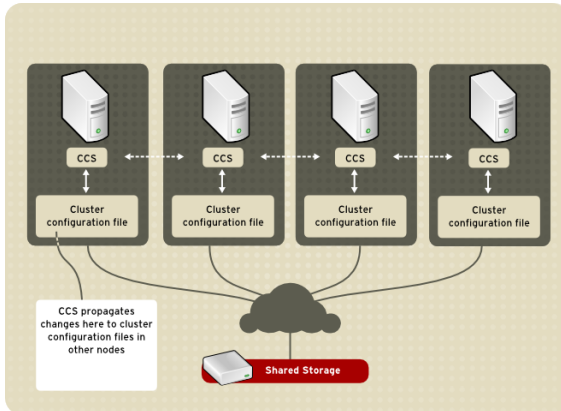
Fencing Arten

- Power fencing
- Fibre Channel switch fencing
- DRAC, IPMI, ...
- Achtung bei redundanten
Power Supplies



Cluster Configuration System

Verwaltet die Cluster Konfiguration.



Cluster Configuration System

Im Cluster Configuration File (/etc/cluster/cluster.conf) stehen folgende Infos:

- Allgemeine Clusterinfo (Name, Locking Type, Fence Timing)
- Cluster Knoten (Name, Fence Methode, Quorum Votes)
- Fence Devices
- Managed Resources (Failover Domains, IP addresses)

HA Service Management

Definition von Services die ausfallsicher im Cluster laufen sollen.

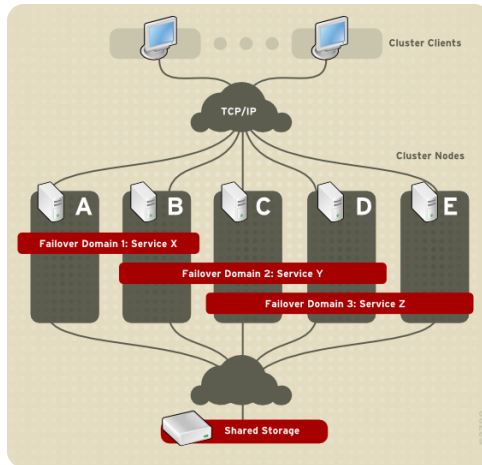
Wird verwaltet vom rgmanager (Resource Group Manager)

Service Definition

- Service Name
- Init Script (z.B. /etc/init.d/httpd)
- Verwendete Ressourcen (floating IP Address)
- Failover Domain

Failover Domains

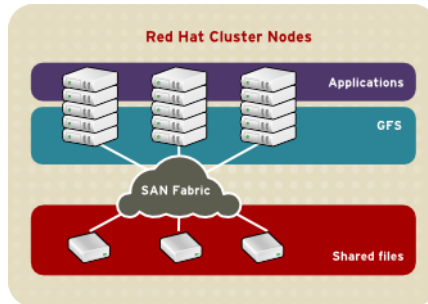
Wo soll ein Dienst laufen?



Shared Resources

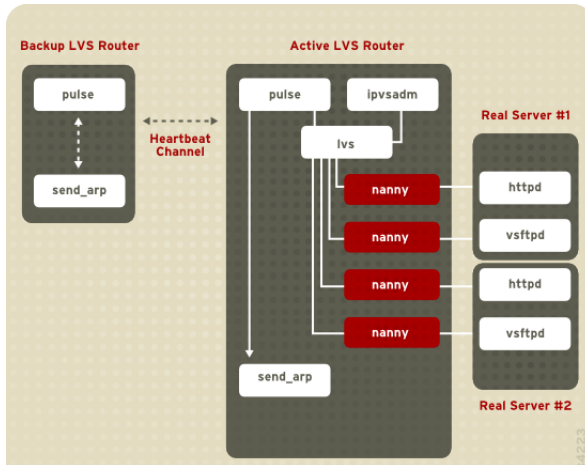
Cluster FS GFS

- Gemeinsame Datenquelle für alle Cluster Knoten
- Anbindung z.B. an ein SAN
- Vereinfacht Deployment und Backup von Applikationen



Linux Virtual Server

Verteilt die Last auf mehrere “echte” Server:



LVS

Load-Balancing Algorithmen

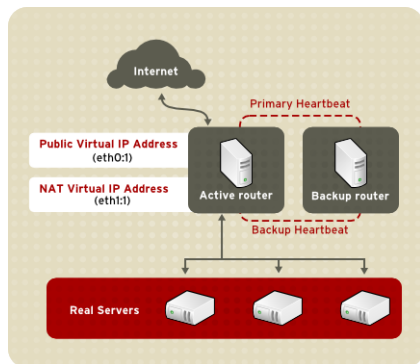
- (weighted) Round-Robin
- (weighted) Least-Connections
- Source Hash Scheduling
- ...

Routing Methoden

- NAT
- direct routing

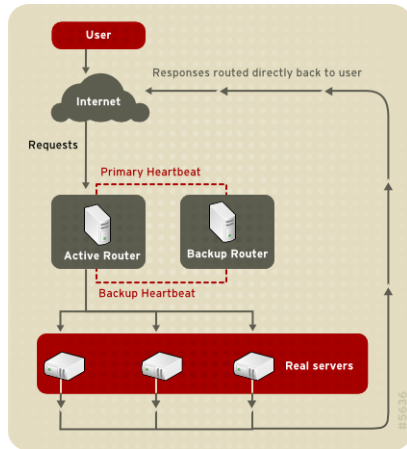
NAT Routing

- Sowohl intern als auch extern wird eine “floating IP” verwendet
- Backend-Server haben interne floating-IP als default Route eingetragen



Direct Routing

- LVS Router verarbeitet nur eingehende Pakete
- Antworten werden von den Servern direkt an die Clients geschickt



Wichtige Kommandos

Befehle

- `ccs_tool` update der Cluster Configuration
- `cman_tool` steuert Cluster Manager
- `clustat` zeigt Status des Clusters an
- `clusvcadm` verwaltet Services

Wir wollen eine kleine, feine Internet Applikation bauen, die dynamische Inhalte ausliefert:

Beschreibung

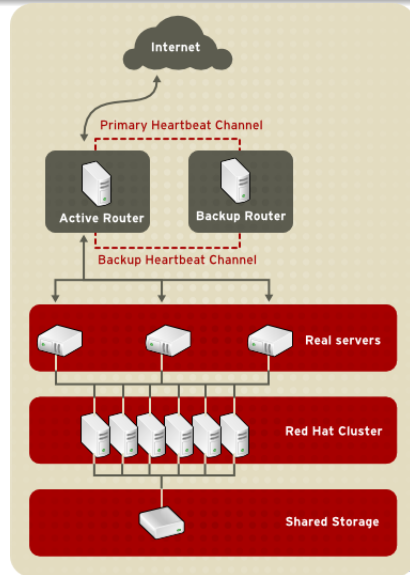
- Webserver (z.B. apache) liefert die Seiten aus welche z.B. mit php generiert wurden
- Datenbankserver hält den dynamischen Content bereit

Anforderungen

- Applikation muss skalierbar sein, d.h. zusätzliche Last soll durch weitere Knoten abgedeckt werden können.
- Datenbank muss ausfallsicher laufen.
- Konsistentes Backup der Daten muss möglich sein.

Stufen

- Stufe 1
Load-Balancing
Server
- Stufe 2
Application
Servers
- Stufe 3
Database Server
- Stufe 4 GFS
Data Source



Knotenkonfiguration

```
<cluster config_version="1" name="DEMO">
  <fence_daemon post_fail_delay="0" post_join_delay="3"/>
  <clusternodes>
    <clusternode name="cl-node-70" votes="1">
      <fence>
        <method name="1">
          <device name="apc-71" port="1"/>
          <device name="apc-73" port="1"/>
        </method>
        <method name="2">
          <device name="ipmi-71"/>
        </method>
      </fence>
    </clusternode>
    <clusternode name="cl-node-72" votes="1">
      ....
    </clusternode>
  </clusternodes>
  <fencedevices>
    <fencedevice agent="fence_apc" ipaddr="172.10.73.71" .../>
  </fencedevices>
```

Servicekonfiguration

```
<rm>
  <failoverdomains>
    <failoverdomain name="PgSQL_Domain" ordered="1" restricted="1">
      <failoverdomainnode name="cl-node-70" priority="1"/>
      <failoverdomainnode name="cl-node-72" priority="2"/>
      <failoverdomainnode name="cl-node-74" priority="3"/>
    </failoverdomain>
  </failoverdomains>
  <resources/>
    <service autostart="1" domain="PgSQL_Domain" name="PgSQL" recovery="relocate">
      <ip address="172.10.73.90" monitor_link="1">
        <script file="/etc/init.d/cl_pgsql" name="initpgsql"/>
      </ip>
    </service>
    <service autostart="1" domain="PgSQL_Domain" name="Clus_Cron" recovery="relocate">
      <script file="/etc/init.d/clus_cron" name="initclus_cron"/>
    </service>
  </rm>
```

Tipps

- Separates Netzwerk für Cluster und Applikation
- Interface-bonding für internes Cluster-LAN
- Anpassung tcp send und receive Buffer
- Genügend Journals in der GFS Partition anlegen (pro zugreifendem Node muss ein Journal existieren)

Resourcen

- http://www.redhat.com/docs/manuals/csgfs/browse/4.6/Cluster_Suite_Overview/
- <http://sourceware.org/cluster/faq.html>
- <http://sources.redhat.com/cluster/wiki/FAQ>

Ende

Vielen Danke für die Aufmerksamkeit!
Fragen?