Zajištění vysoké dostupnosti služeb

[pomocí Novell Cluster Services]

Bc. Jakub Talaš
systémový inženýr – Linux
Unis Computers spol. s r.o.
Jundrovská 33, 624 00 Brno
jtalas@uniscomp.cz

ABSTRACT

Cílem přednášky je vysvětlit principy používání clusterů v prostředích, kde je nutná vysoká dostupnost síťových služeb. Posluchači se dozví o používaných technologiích, postupech, správě služeb a řešení výpadků. Podrobněji se budem věnovat clusterování nad linuxovou platformou Novell Open Enterprise Server i s praktickou ukázkou.

Keywords

Novell Cluster services, high availability, Linux clustering, open source, Linux HA project, Hearbeat

1. ÚVOD

Informační technologie dnes tvoří srdce každé instituce a i krátký výpadek informačních služeb může znamenat velkou finanční ztrátu. Základním vzorem zachovávání vysoké dostupnosti síťových služeb je redundance používaných aktivních prvků, serverů, datových spojů a datových úložišť. Proto se v prostředích jako jsou nemocnice, firmy s provozem 24/7 apod. investuje mimo jiné do provozu škálovatelných clusterovaných řešení.

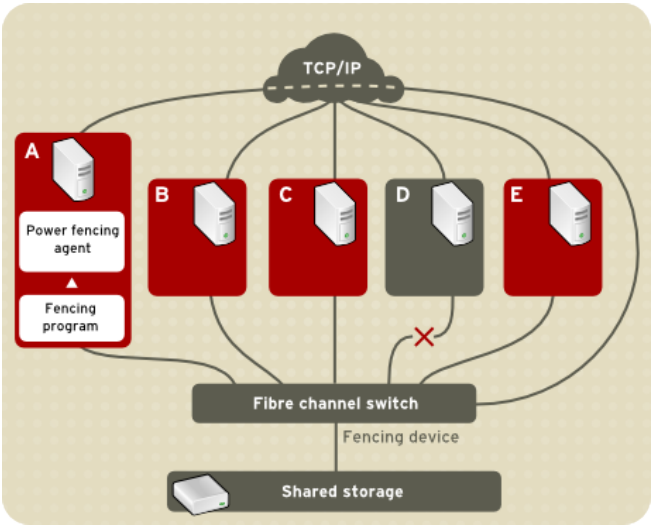
2. CLUSTERY A JEJICH VÝHODY

Clustery patří do skupiny volně vázaných (anglicky loosely coupled) distribuovaných systémů. Ty se vyznačují propojením LAN, distribuovanou pamětí a vyšší latencí komunikace mezi procesory. Cluster je tvořen dvěma nebo více fyzickými stroji (tzv. *uzly*, angl. nodes), které jsou od uživatelů abstrahovány a navenek se celý cluster tváří jako jediný výkonný počítač. Rozlišujeme několik základních typů clusterů. Podle požadovaného účelu jsou používány:

- clustery pro rozložení zátěže (load balancing)
- výpočetní clustery (high performance computing, HPC)
- clustery s vysokou dostupností (high availability, HA)

Pokud počet uzlů přesahuje několik set nebo tisíc, mluvíme o takzvaných *gridech*. Ty se často používají k akademickým a vědeckým výpočtům. V komerčním prostředí jsou však nejpoužívanější právě HA (někdy se můžete setkat také s výrazem *failover*) clustery, které zajišťují souborové, síťové a tiskové služby.

Pro zajištění spolehlivého chodu je třeba používat kvalitní, avšak často drahé, technologie – Fibre Channel infrastrukturu SAN (Storage Area Networks) apod. Použitím různých technologií pro ukládání dat (např. FC SAN, iSCSI, DRBD – real-time replikace dat na zařízení, apod.) a jejich sdílení mezi uzly se liší i schémata zapojení clusteru. Většinou však bývají uzly připojeny více kanály k několika Fibre Channel switchům, které jsou opět více cestami propojeny s diskovým polem. Cílem tohoto zapojení je eliminovat single point of failure. Tok dat (data mohou proudit různými cestami) řídí a kontroluje ovladač pro multipathing. Zjednodušené schéma zapojení i s fencingem (viz dále) ukazuje následující ilustrace převzaná z http://tinyurl.com/redhatdoc



3. CLUSTERING POD LINUXEM

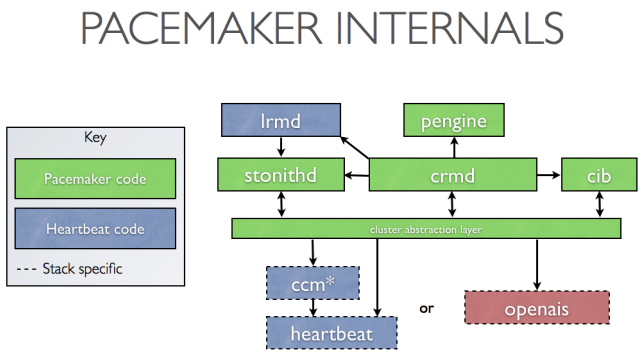
V Linuxu má clusterování dlouhou historii. Proto není divu, že pro tyto služby existuje několik různých nástrojů. Základem clusteru je *cluster stack* – vrstva zajišťující komunikaci a členství uzlů, management selhání a podobně. V současnosti se nejčastěji používá jeden z těchto projektů.

- Linux HA Project
- OpenAIS
- Red Hat Cluster Suite

Linux HA Project vyvíjí cluster stack Heartbeat. Časem se od něj oddělil jeho manažer zdrojů (*cluster resource manager – CRM*) Pacemaker, který už pokračuje samostatným vývojem. Pacemaker nyní umožňuje mimo Heartbeatu užít jako cluster stack i OpenAIS a jeho úkolem je spouštění, zastavování a migrování služeb podle definovaných politik.

3.1 Fencing

Jakmile se stane, že kvůli nějakému selhání, např. v síti, spolu přestanou některé skupiny uzlů komunikovat, přichází ke slovu další část clusteru, a to software zajišťující tzv. *fencing*. Jeho úkolem je odpojit neodpovídající uzel (uzly) od zdroje energie nebo od sdíleného úložiště. Pokud by ke sdílenému úložišti přistupovaly uzly nesynchronizovaně, hrozilo by poškození dat kvůli jejich současnému zápisu. Fencing řeší většinou tzv. STONITH (Shoot The Other Node In The Head) daemon, který vypne neodpovídající uzel. Cluster information base (CIB) představuje centrální úložiště informací, fyzicky v XML formátu. Všechny tyto základní součásti přehledně ukazuje následující ilustrace (byla převzata z http://clusterlabs.org/wiki/Image:Stack.png).



Red Hat Cluster Suite funguje na podobných principech jako řešení postavné na Linux HA nebo OpenAIS. Jako souborové systémy jsou většinou používány OCFS2 nebo GFS.

4. NOVELL CLUSTER SERVICES

Americká firma Novell[1] podporuje vývoj otevřených clusterových řešení především prostřednictvím projektů Heartbeat a Pacemaker. Součástí SUSE Linux Enterprise Serveru 11 (který je také součástí portfolia Novellu) je High Availability Extension postavená na OpenAIS, který by se do budoucna měl stát hlavním linuxovým cluster frameworkem pro Novell i Red Hat.

[1]http://www.novell.com

Novell Cluster Services (dále NCS) je nástroj, který je sice proprietární, ale je postaven na Heartbeatu a mnoha dalších otevřených komponentách. Je součástí Novell Open Enterprise Serveru (dále OES; nadstavba SLES a pokračovatel klasického Netware) a poskytuje množství základních funkcí, které se od cluster frameworku očekávají (zakoupením OES získáte v ceně licenci pro dvouuzlový cluster). Jeho hlavní výhoda je jednoduchost, webové administrační rozhraní, podpora pro služby OES, integrace s eDirectory.

4.1 Jak fungují Novell Cluster Services

Novell Cluster Services má, podobně jako většina služeb v OES, svůj základ v eDirectory - používá ji jako společné úložiště konfigurace. Tak jako ostatní služby používá pro svoji konfiguraci nástroj YaST a modul ve webovém konfiguračním nástroji iManager. V YaSTu se provádí prvotní vytvoření clusteru a přidávání nových uzlů, další správa už je prováděna v iManageru a nastavení jsou zapisována do LDAPu.

4.1.1 Správa služeb

Cluster resource manager je oproti Heartbeatu a Pacemakeru nebo jiným nástrojům chudší. Neumožňuje vytvářet složitější politiky obsahující failover domény, závislosti služeb, pluginy a další pokročilejší vlastnosti. Přesto umožňuje kvalitně řídit dodávku souborových, tiskových a síťových služeb s vysokou dostupností ve většině nejčastějších případů.

Při vytváření zclusterovaných služeb (tzv. zdrojů, resources) lze efektivně využít předem připravených šablon. Ty obsahují předvyplněné bash skripty, které jsou srdcem celého koloběhu služby – každá služba má spouštěcí, vypínací a případně ještě monitorovací skript. NCS podporuje (mimo nějaké ošklivé hacky:-)) pouze mód active-standby, tedy jednotlivá služba běží v momentě pouze na jednom uzlu. Není teda třeba používat speciální clusterové souborové systémy. V případě filesystem zdroje (pro poskytování souborových služeb) je možné použít jak tradiční linuxový ext3 nebo reiser, tak NSS – Novell Storage Services filesystem, který má také zabudovanou podporu pro přístup z více uzlů NCS clusteru.

4.1.2 NCS startovací skript

Ukažme si příklad startovacího skriptu pro NSS svazek. Tento skript se vyplňuje v textovém poli webového administračního rozhraní. Poté se uloží do eDirectory a tím automaticky rozdistribuuje k jednotlivým uzlům.

```
1.  #!/bin/bash
2.  . /opt/novell/ncs/lib/ncsfuncs
3.  exit_on_error nss /poolact=UZIVATELE
4.  exit_on_error ncpcon mount UZIVATELE=251
5.  exit_on_error add_secondary_ipaddress 172.16.0.21
6.  exit_on_error ncpcon bind \
     --ncpservername=CLUSTER_UZIVATELE_SERVER \
     --ipaddress=172.16.0.21
7.  exit 0
```

Co tento skript dělá? Řekněme že máme na diskovém poli NSS oddíl obsahující cluster-enabled *pool* UZIVATELE. Ten nese svazek jménem UZIVATELE s uživatelskými daty. Při

spouštění tohoto zdroje náš skript na třetím řádku daný pool aktivuje. Následující řádek připojí požadovaný svazek (i s číslem připojení 0-255). Poté přidáme fyzickému stroji další sekundární IP adresu, kterou konečně na šestém řádku přiřadíme našemu virtuálnímu serveru s uživatelskými daty. Tedy – at' už poběží náš virtuální NSS souborový server na jakémkoliv uzlu v clusteru, uživatelům stačí, že se k němu vždy pod jemu přiřazenou IP adresou 172.16.0.21 nebo jejím hostname připojí. Do skriptů je pochopitelně možné vkládat i jakékoliv další bash příkazy, například pro spuštění zálohovacích nástrojů apod.

4.1.3 Failover

NCS se liší od jiných linuxových cluster řešení také způsobem, jakým se vyrovnává udržováním a ztrátou komunikace. Jeden z uzlů je vždy tzv. *master*, ten posílá multicastem malé heartbeat pakety ostatním. Ti mu unicastově odpovídají. NCS má na sdíleném úložišti speciální malý oddíl nazývaný SBD – Split Brain Detector. Když přestane uzel na heartbeat pakety odpovídat, je ostatními uzly vyjmut z clusteru a zbývajcí uzly zvýší číslo epochy. SBD podle neodpovídajících čísel epochy zjistí, že cluster je rozdělen na dvě části (split brain). Poté menší část (popř. ta neobsahující master uzel nebo bez spojení na LAN) "spolkne" tzv. *pilulku s jedem* (poison pill), která všechny příslušné uzly shodí, aby se zajistila konzistence dat. Zdroje původně běžící na vypnuté části clusteru se podle politik přemigrují na běžící stroje nebo zůstanou nedostupné dokud jejich uzly znovu nenajedou. Kvůli systému s SBD a síť'ovým heartbeat protokolem není třeba používat STONITH daemon.

5. ZÁVĚR

Novell Cluster Services jsou příjemným doplňkem platformy OES v prosředích, kde je nutné zaručit vysokou dostupnost služeb, ale není nutné použít řešení poskytující některé pokročilé funkce. S výhodou lze využít souborového systému NSS a šablon pro většinu součástí Open Enterprise Serveru (Novell DNS, Novell DHCP, iPrint,...). Tradice novellovských HA řešení sahá hluboko ještě do vod Netware a jsou ověřena mnoha produkčními nasazeními.

5.1 Zdroje

Pro zájemce o hlubší studium clusterování a vysoké dostupnosti doporučuji začít na v této sekci zmíněných stránkách. Z nich jsem čerpal i pro tento článek.

OpenAIS http://www.openais.org/doku.php

Pacemaker http://clusterlabs.org/wiki/

NCS http://www.novell.com/documentation/oes2

Red Hat CS http://sources.redhat.com/cluster/wiki/

SBD Fencing http://www.linux-ha.org/SBD_Fencing