

Linux Cluster tehnologije



Nikola Pavković

`<nix@irb.hr>`

Što je to Linux cluster?



⌘ skupina PC računala koja rade pod Linux OS-om, na neki način spregnuta u jedinstveni sustav

Funkcionalna podjela



⌘ High Availability clusteri (HAC)

⌘ High Performance clusteri (HPC) ili
Beowulf clusteri

Malo povijesti



- ⌘ Donald Becker CESDIS @ NASA's Goddard Space Flight Center
- ⌘ ljeta 1994. prvi Beowulf cluster sa 16 čvorova (ESS)
- ⌘ Scyld Computing Corporation

High performance clusteri

⌘ primjena:

☑ znanstvene aplikacije

- ☒ simulacije fizikalnih ili kemijskih procesa
- ☒ meteorologija
- ☒ genetski algoritmi
- ☒ ...

☑ skalabilni poslužitelji

- ☒ web poslužitelji
- ☒ mail poslužitelji sa antivirusnom zaštitom

Zašto cluster a ne superračunalo?



- ⌘ jeftinije rješenje koje nudi iste ili bolje performanse
- ⌘ gradi se od komponenti dostupnih u svakom computer-shopu
- ⌘ skalabilniji od superračunala
- ⌘ koristi free software
- ⌘ neovisan o proizvođaču

Topologija mreže u Linux clusteru



Načini kontrole i dodjeljivanja resursa



⌘ MOSIX clusteri

⌘ clusteri sa batch sustavom

MOSIX (<http://www.mosix.org/>)



- ⌘ sustav za korisnika djeluje kao jedno računalo velikih performansi
- ⌘ jednostavna implementacija
 - ☑ MOSIX kernel patch
 - ☑ konfiguracija `/etc/mosix.conf`
- ⌘ omogućava dinamičku migraciju procesa između radnih čvorova
 - ☑ proces se seli na čvor koji je trenutno slabije opterećen

MOSIX (<http://www.mosix.org/>)



⌘ decentralizirani rad

- ☑ svi čvorovi imaju istu važnost

- ☑ korisnik se prijavljuje na bilo koji čvor

⌘ omogućuje koegzistenciju računala različitih resursa u clusteru, a bez dodatne konfiguracije parametara

⌘ primjena:

- ☑ skalabilni web serveri

- ☑ multi-user sustavi sa 'time-sharing' okolinom

Što je batch sustav?

- ⌘ poslužiteljski servis na frontalnom čvoru
- ⌘ nadgleda potrošnju resursa na svakom pojedinom čvoru u clusteru
- ⌘ prima korisničke zahtjeve za pokretanjem procesa
 - ☑ korisnik, kod podnošenja zahtjeva batch sustavu, mora definirati niz parametara:
 - predviđeno vrijeme obrade
 - zahtjevi za memorijom i brojem procesora (i/ili čvorova)
 - naziv izvršne datoteke koja se pokreće

Što je batch sustav?

- ⌘ konstantno izvještava scheduler o trenutnoj opterećenosti radnih čvorova
- ⌘ pokreće korisničku aplikaciju tj. proces na radnom čvoru kada se za to stvore uvjeti:
 - ☑ zatraženi resurs je dostupan
 - ☑ scheduler odobrava da se pokrene obrada baš tog zahtjeva
- ⌘ PBS (<http://www.openpbs.org/>)
 - ☑ LSF, NQE, DQS, CONDOR

scheduler



⌘ FIFO

- ⊞ najprimitivniji

- ⊞ iskorištenost resursa nije optimalna

⌘ 'backfill' algoritam

⌘ definiranje grupa korisnika sa pravima

⌘ 'fairshare' algoritam

Ostali servisi u batch clusteru



⌘ NFS poslužitelj

- ☑ korisnički direktoriji dostupni na svim čvorovima

- ☑ SSI clusteri

⌘ NIS poslužitelj

⌘ rsh ili ssh poslužitelji

- ☑ autentikacija na radnim čvorovima

⌘ NTP poslužitelj

- ☑ sinkronizacija vremena unutar clustera

Ostali servisi u batch clusteru



⌘ DHCP poslužitelj

⌘ SQL poslužitelj

☑ informacije o čvorovima, korisnicima i sl.

☑ informacije o postavkama na radnim čvorovima

☑ uvelike olakšava administraciju sustava

☒ generiranje instalacijskih skripti za radne čvorove

☒ generiranje konfiguracijskih datoteka na frontalnom čvoru

Administracija clustera



⌘ ručna administracija svakog čvora pojedinačno

- ☑ teoretski moguća

- ☑ komplicirana

- ☑ uzima previše vremena

- ☑ razvoj vlastitih alata za olakšanje administracije

Administracija clustera

⌘ SSI (Single System Image) sustavi

- ☑ bazirani na 'root over NFS' konceptu

 - ☒ <http://www.linuxdoc.org/HOWTO/Diskless-root-NFS-HOWTO.html>

 - ☒ cluster-nfs (<http://clusternfs.sourceforge.net>)

- ☑ glavna prednost:

 - ☒ vrlo jednostavna administracija

- ☑ glavni nedostatak:

 - ☒ generira veliki mrežni promet

Administracija clustera

⌘ cluster management alati

- ⊞ automatizacija administracije radnih čvorova
 - ⊞ instalacija i konfiguracija operacijskog sustava
 - ⊞ instalacija i konfiguracija korisničkih aplikacija
- ⊞ automatizacija konfiguracije servisa na frontalnom čvoru
 - ⊞ batch sustav
 - ⊞ NIS
 - ⊞ NFS

Stvarni primjeri alata za cluster management

⌘ komercijalni alati:

☑ Scyld (<http://www.scyld.com>)

☑ ...

⌘ open source alati

☑ MSC (<http://msclinux.com/>)

☑ OSCAR (<http://oscar.sourceforge.net/>)

☑ ROCKS (<http://rocks.npaci.edu/>)

☑ ...

ROCKS



⌘ RedHat Linux kao osnova

☑ RedHat kickstart datoteka

☑ konfiguracija sustava

- particijska tablica, vremenska zona, raspored na tastaturi i sl.

☑ popis paketa za instalaciju

☑ pre- i postinstalacijske skripte

- dodatna prilagodba konfiguracije programskih paketa lokalnim potrebama

ROCKS



⌘ MySQL baza na frontalnom čvoru

- ☑ podaci o čvorovima u clusteru
 - ☒ broj procesora
 - ☒ raspoloživa memorija
 - ☒ fizička lokacija čvora...
- ☑ generiranje konfiguracijskih datoteka za servise na frontalnom čvoru

ROCKS



⌘ sustav za konfiguraciju programske podrške na radnim čvorovima

- ☑ dinamički generira kickstart datoteku
- ☑ kreiranje distribucije paketa prema potrebi
- ☑ mogućnost dodatne konfiguracije gotovih paketa (bez repaketiranja)
- ☑ XML konfiguracija
- ☑ omogućava kreiranje različitih distribucija za čvorove različitih namjena

ROCKS



⌘ instalacija radnih čvorova:

- ☑ network boot

- ☑ dodjeljivanje IP adrese

- ☑ dohvat kickstart datoteke i paketa preko mreže (NFS-om sa frontalnog čvora)

 - ☒ kickstart datoteka se generira prema namjeni čvora

ROCKS



- ⌘ NFS poslužitelj na frontalnom čvoru
 - ☑ distribucija paketa prilikom instalacije
 - ☑ distribucija korisničkih home direktorija

- ⌘ DHCP poslužitelj na frontalnom čvoru
 - ☑ dodjeljuje IP adrese radnim čvorovima

ROCKS



⌘ NIS poslužitelj na frontalnom čvoru

☑ informacije o korisnicima

☑ informacije o imena čvorova

ROCKS... zaključak



- ⌘ u velikoj mjeri automatizirana administracija radnih čvorova kao i servisa na frontalnom čvoru
- ⌘ velika fleksibilnost pri konfiguraciji distribucije programkih paketa
- ⌘ trenutno najnapredniji free alat

Budućnost



⌘ razvoj alata za cluster management

☑ jednostavnost korištenja

⌘ ubrzanje propusnosti mreže

☑ shareanje memorije preko mreže

☑ brža komunikacija među procesima

⌘ optimizacija scheduling algoritama

☑ poboljšanje iskoristivosti raspoloživih resursa

Budućnost



⌘ Sa <http://clusters.top500.org/>

☑ Dr. Thomas Sterling: " It's quite possible that by the middle of this decade clusters in their myriad forms will be the dominant high-end computing architecture. "

Pitanja...

