Linux High Availability Cluster

Pacemaker ir Corosync

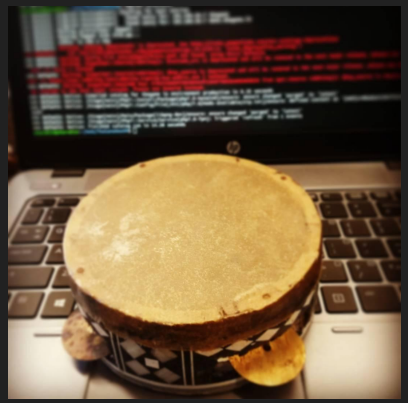
Sergej Kurakin

Sergej Kurakin

Amžius: 36

Dirbu: NFQ Technologies

Pareigos: Programuotojas



NFQ Offline Day 2015 (2015-09-04)

Pranešimas “Kodėl aš skaitau pranešimus?”:

- Pasidalinkime savo eksperimentais.

Manau, jau atėjo laikas papasakoti apie vieną.

Bet tai tik eksperimentas, labiau teorinis, kurio nesu pritaikęs realiam projekte.

2007 - 2012

Dirbau prie vieno didesnio projekto.

Ne viskas veikė puikiai.

Ne viskas buvo automatizuota.

Daugiausiai problemų kildavo su “failover” sprendimais.

Situacija

Truputį “legacy” projektas.

Mes negalime keisti DNS įrašų.

Mes negalime migruoti į “Cloud”.

Yra limituotas kiekis techninės įrangos resursų.

Esant reikalui techninė įranga per tam tikrą laiką gali būti nupirkta, sumontuota ir pajungta.


Round Robin DNS


NGINX1


NGINX 2


PHP-FPM


PHP-FPM


MySQL


File Storage

Prieš 4+ metus

Jau dirbant prie kitų projektų kitoje vietoje aš radau laiko ir noro kažką išbandyti.

https://twitter.com/zaza_lt/status/287636375992422400

CentOS + Linux-HA

Tikslas: rasti sprendimą, kuris leis RHEL/CentOS OS pagrindu užtikrinti “High Availability” NGINX serveriui.

Bet pradėkime nuo pradžių.

Kas gali sulūžti?

Tinklo šakotuvas (Switch)

Tinklo laidas

Tinklo korta (NIC)

Maitinimo šakotuvas (PDU)

Maitinimo blokas (PSU)

Maitinimo laidas

Diskas

Pagrindinė plokštė (Main board / Motherboard)

Diskų valdiklis

Ventiliatoriai

Atmintis (RAM)

Mikroprocesoriai (CPU)

Kiti specifiniai komponentai

Kas yra “High Availability”?

Sistemos savybė, kurios pagalba siekiama užtikrinti sutartą eksploatacinių savybių lygį.

Dažniausiai siekiama užtikrinti didesnį nenutrūkstamą sistemos veikimo laiką negu įprastai.

Klasika be kurios nebūna HA prezentacijų

99%	3.65 dienas
99.9%	9 valandas
99.99%	52 minutes
99.999%	5 minutes
99.9999%	30 sekundžių

Nenutrūkstamas sistemos veikimas

Įjungtas serveris?

Veikianti operacinė sistema?

Atsakimas į “ping”?

Tam tikro TCP/UDP porto veikimas?

Tam tikro URL atidarymas?

Tam tikro funkcionalumo veikimo užtikrinimas?

Nustatyti silpniausias sistemos dalis

Kas gali lūžti?

Ar sustos veikti visa sistema?

Ar nustos veikti svarbus sistemos funkcionalumas?

Nustatyti “Single Point of Failure”.

Kas yra tas “Cluster”?

Tai grupė kompiuterių, sujungtų tinklu, skirtų atlikti tą pačią funkciją ar užduotį.

Dažniausiai “iš išorės” jie atrodo kaip vienas kompiuteris.

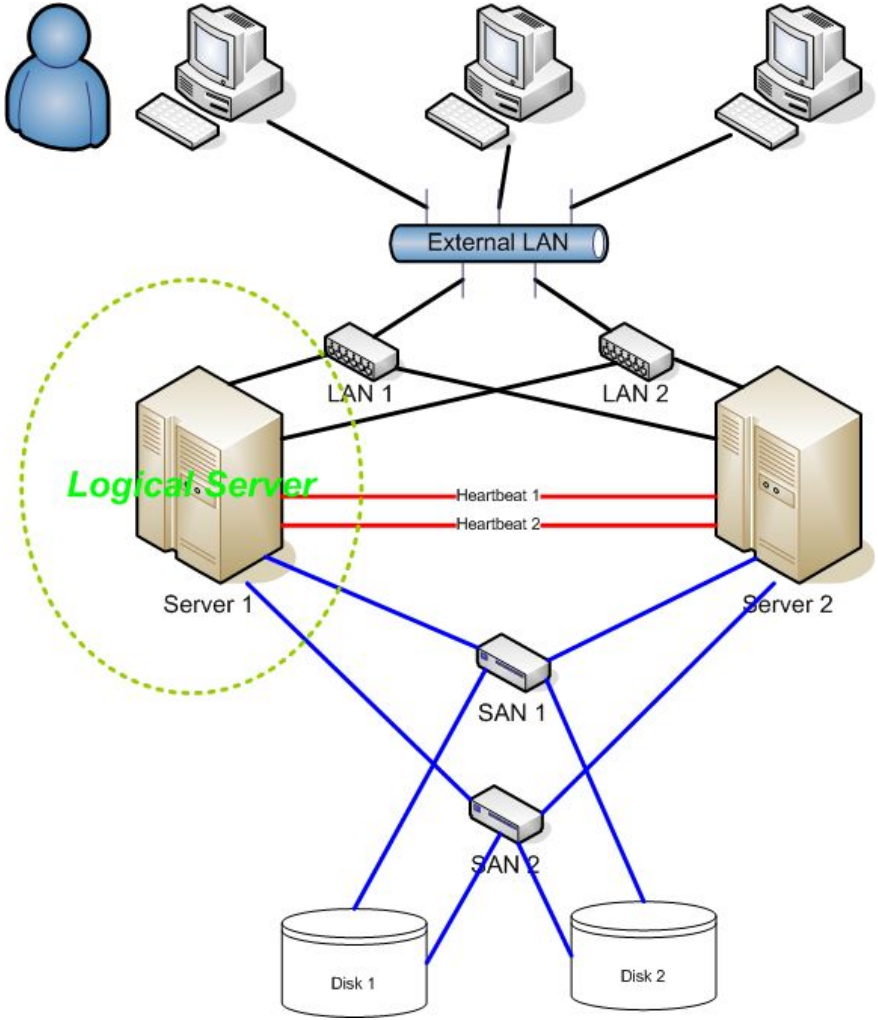


Image from https://en.wikipedia.org/wiki/High-availability_cluster

Linux HA Cluster komponentai

- Shared storage
- Different networks
- Bonded network devices
- Multipathing
- Fencing/STONITH

Programinė įranga

Cluster membership layer

corosync

cman

heartbeat

openais

Resource manager

pacemaker

Pacemaker vidiniai komponentai

CIB - Cluster Information Base

crmd - Cluster resource manager daemon

pengine - Policy engine

lrmd - Local resource manager daemon

stonithd/fenced

“Cluster” veikimo režimai

Active / Active

Active / Passive

N+1

N+M

N-to-1

N-to-N

Kokie yra “resource agent”?

Filesystem

IPaddr2

ProFTPd / proftpd

SphinxSearchDaemon

apache

mysql

mysql-proxy

nfsserver

nginx

pgsql

postfix

rabbitmq-cluster

redis

Jų yra žymiai daugiau.

STONITH (Shoot The Other Node In The Head)

Neveikiančio serverio izoliavimas, siekiant apsaugoti “Cluster” nuo jo neigiamo poveikio.

Dažniausiai tai išjungimas arba perkrovimas.

Įrankiai: IPMI, APC, DRAC, Virtualization Management, SSH, Meatware

Tai yra pagrindiniai Linux High Availability Cluster dalis

Diegimo principas

Į visus narius diegiama ir konfigūruojama reikalinga programinė įranga.

Į visus narius diegiamas ir konfigūruojamas “membership layer”.

Į visus narius narius diegiamas “resource manager”.

Viename iš narių konfigūruojami resursai per “resource manager”.

Konfigūracinių failų valdymas

Kiekviename serveryje atskirai.

Sunku suvaldyti, lengva suklysti. Automatizuoti sinchronizaciją?

“Shared storage”.

Svarbu, kad netaptų SPOF.

Tad grįžkime prie mano eksperimento.

Sena sistema


Round Robin DNS


NGINX1


NGINX 2

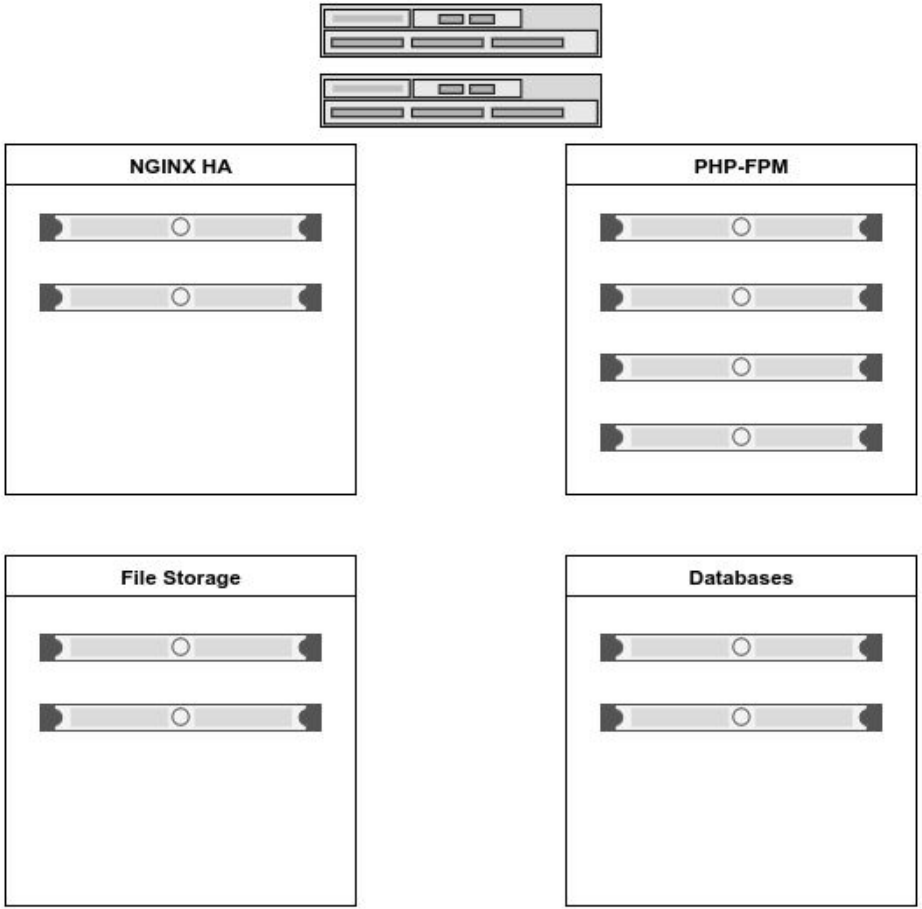

PHP-FPM


PHP-FPM


MySQL


File Storage

Tai į ką aš taikiausi


NGINX HA
PHP-FPM
File Storage
Databases

Demo One

Jei pavyks :-)

Mano pasirinktas ir sukurtas sprendimas turėtų perkelti IP į veikiantį serverį ir jame paleisti NGINX.

Kas panaudota

VirtualBox

4 VM’ai viename “host-only” tinkle

Debian Jessie + Backports

Pacemaker + corosync

NGINX

PHP-FPM

Demo Two

Jei pavyks :-)

Mano pasirinktas ir sukurtas sprendimas turėtų perkelti IP į veikiantį serverį ir jame paleisti NFS Share.

Kas panaudota

VirtualBox

2 VM’ai dvejuose “host-only” tinkluose

Debian Jessie + Backports

Pacemaker + corosync

DRBD

NFS

Kas nepavyko

Network bound - kažkaip neveikia man su VirtualBox.

NFS Cluster nevisada teisingai daro “failover”.

Testavimas

Migracijos ir resursų išjungimas

Staigus resurso išjungimas (pvz.: kill -9)

Elektros dingimas viename iš “node”

Tinklo sutrikimai

STONITH testai

Daugiau įrangos != didesnis HA

Ne, tai ne paradoksas!

Didesnis įrangos kiekis padidina reikiamas pastangas norint užtikrinti HA, nes atsiranda daugiau galimų trikčių vietų ir sudėtingėja įgyvendinimas.

Manau, kad mano eksperimentas pavyko.

Išvados

Reikia labai gerai išmanyti programinę įrangą kuri keliama į “Cluster”.

Reikia labai gerai išmanyti “Cluster” programinę įrangą ir konfigūravimą.

Labai stipriai sudėtingėja projektų diegimas ir priežiūra.

Duomenų centras turi atitikti tam tikrus reikalavimus.

Techniniai įrangai (serveriai, tinklas) turi būti keliami specifiniai reikalavimai.

Prieš paleidimą reikia atlikti žymiai daugiau “Cluster” testavimo darbų.

Galbūt vertėjo naudoti RHEL/CentOS (Red Hat Cluster Suite), o ne Debian.

Papildoma informacija

https://en.wikipedia.org/wiki/High_availability

https://en.wikipedia.org/wiki/High-availability_cluster

https://wiki.debian.org/Debian-HA/ClustersFromScratch

http://www.linux-ha.org/wiki/Main_Page

http://clusterlabs.org/

https://ourobengr.com/ha/

Knygos

“Clusters from Scratch”
http://clusterlabs.org/doc/

“Pro Linux High Availability Clustering”
https://www.apress.com/la/book/9781484200803

Diskusija

Sergej Kurakin

@paštas: sergej.kurakin@nfq.lt