

(Relatório final)

Linux Cluster with 3 or more Applications

GitHub:

<https://github.com/RubenABM/Sistemas-operativos>

Ana Bernardino (202000330), João Pires(20200459),
Marta Sousa (20200278) e Rúben Martins (20200453)

Licenciatura de Engenharia Informática

IADE – Faculdade de Design Tecnologias e Comunicação

Sistemas Operativos

Professor Pedro Rosa

14 de Maio 2023

Introdução

Cluster (ou clustering) é o nome dado a um sistema que relaciona dois ou mais computadores para que estes trabalhem de maneira conjunta no intuito de processar uma tarefa. Estas máquinas dividem entre si as atividades de processamento e executam este trabalho de maneira simultânea. Caso um node falhe, os serviços presentes serão transportados para outro node. Com este trabalho tivemos como objetivo implementar uma solução de clustering, em ambiente linux, com pelo menos 3 aplicações.

Relativamente à proposta do trabalho, alteramos duas aplicações, ou seja em vez de termos o server dns passamos a ter um servidor nodejs e uma base de dados replicada.

Solução a implementar

Como referido em relatórios anteriores, a solução de clustering escolhida foi o SUSE Linux Enterprise Server com recurso à extensão SUSE High Availability Extension.

Descrição

O SUSE Linux Enterprise Server foi a distro selecionada para dar host ao cluster deste trabalho pelos diversos aspetos positivos encontrados no decorrer da avaliação intermédia dos softwares de clustering disponíveis. Foi registada uma conta com a SUSE, pois é necessário para obter as keys de utilização do trial de 60 dias de tanto a distro como a extension utilizada para configurar o cluster.

As máquinas em si foram colocadas em VirtualBox, tendo os nodes 2cpus, 15gb de disco e 2gb de RAM cada, há exceção da máquina sbd que dispões de um disco adicional de 4g utilizado como disco partilhado para node fencing.

Em termos de serviços foram implementados um server nodejs, um mail server e uma base de dados replicada.

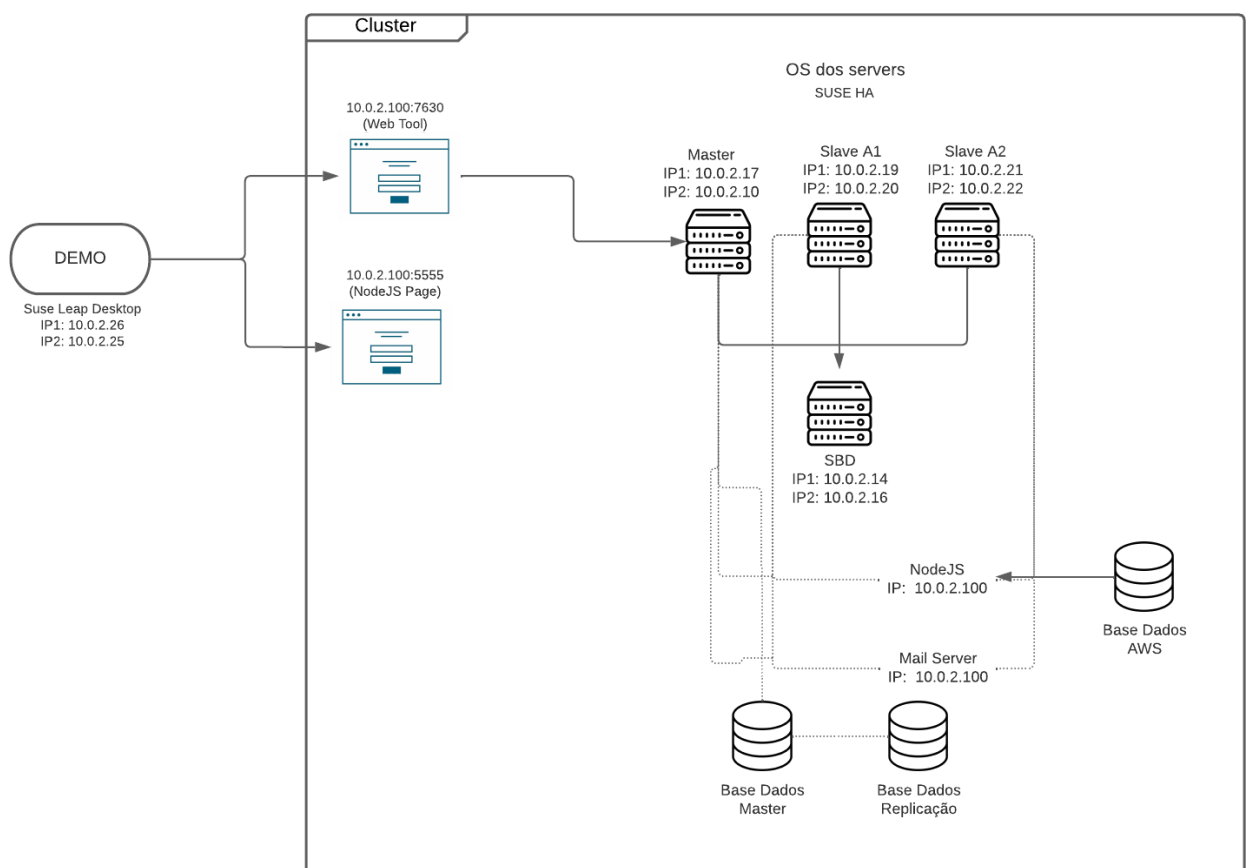
Enquadramento na Unidade Curricular

Relativamente ao enquadramento na unidade curricular, este projeto teve como objetivo estudar e representar a fiabilidade que os sistemas operativos podem ter. Ou seja, ao utilizar um cluster, é possível indisponibilizar um dos elementos e ainda ter outros que o substituam. Ao mesmo tempo melhora o desempenho do sistema operativo, fazendo com que as tarefas sejam executadas mais rapidamente.

Requisitos Técnicos

- Sistema de Tolerância a falhas e ambiente de alta disponibilidade com replicação
- Escalabilidade
- Virtualização

Arquitetura Final da Solução

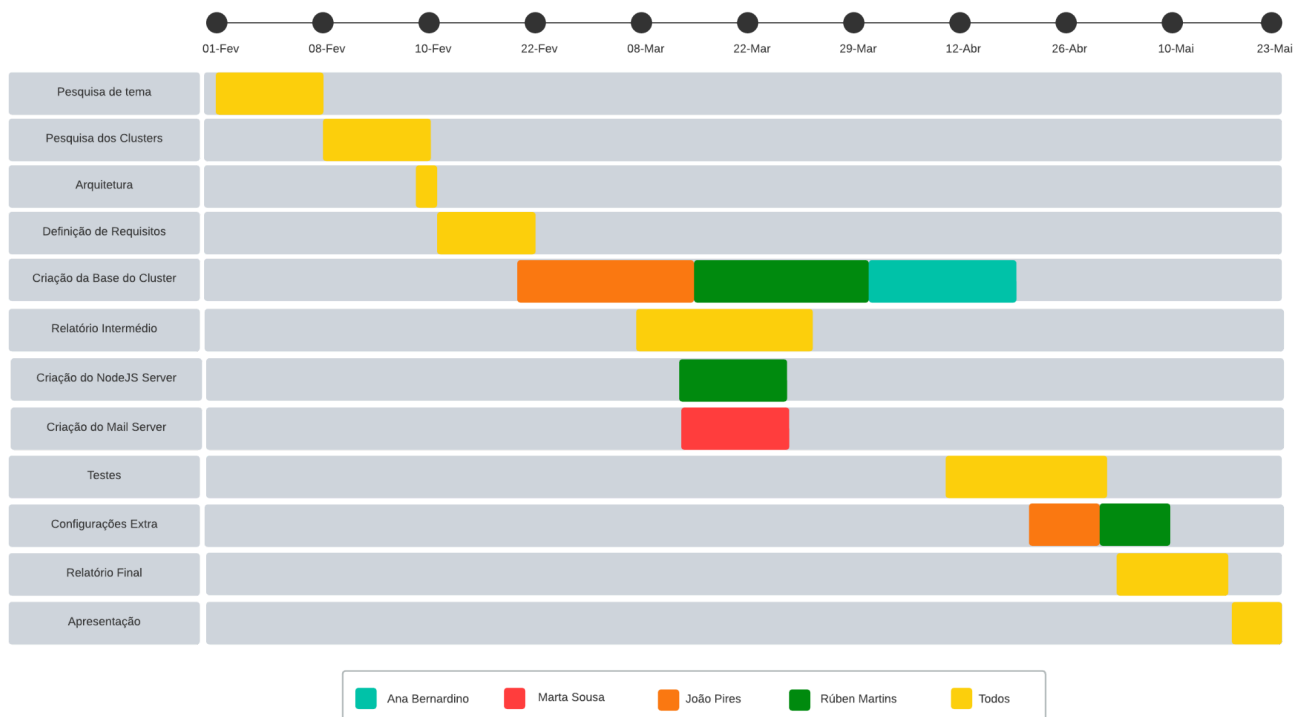


Tecnologias a Utilizar

As tecnologias que foram utilizadas foram o sistema operativo OpenSuse Leap sp15, máquina virtual da Oracle e por fim foi usada a extensão “SUSE Linux Enterprise High Availability Extension” . Para as 3 aplicações foram utilizadas as seguintes extensões: para o mail server foi usado o Postfix, base de dados em MariaDB/MySQL e PM2 para o web server NodeJS.

Planeamento e Calendarização

A calendarização inicial não foi totalmente respeitada, tendo em conta que foram feitas várias alterações e devido à quantidade de problemas que foram aparecendo à medida que o projeto foi avançando e pela maneira como esses problemas foram resolvidos.



Resultados

Concluimos que foram atingidos os objetivos principais da construção de um cluster mas que houveram etapas propostas que não foram conseguidas.

O cluster foi desenvolvido em cima de uma plataforma SUSE linux, assim como foi proposto no relatório intermédio, onde concluimos ser a melhor escolha para o nosso caso estudo.

Inicialmente construído em cloud (aws) mas devido a dificuldades encontradas foi passado para localhost, deixando apenas em cloud a base de dados do serviço nodejs. É composto por 3 nodes, 1 master e 2 slaves, 1 serviço de sbd para node fencing, também baseado em SUSE linux e por fim uma máquina SUSE Desktop externa ao cluster para visualização da web tool e Express page do serviço nodejs. Todos os nodes estão sincronizados ao mesmo server NTP para obtenção da melhor performance aquando fencing.

Foram implementados 3 serviços, server nodejs, mail server e base de dados replicada. Tanto o cluster como os serviços cumprem o objetivo de passarem para outro node disponível após falha no que lhes faz host atual.

Bibliografia

<https://doc.opensuse.org/documentation/leap/reference/html/book-reference/cha-dhcp.html>

<https://documentation.suse.com/sle-ha/15-SP1/html/SLE-HA-all/cha-ha-fencing.html>

<https://documentation.suse.com/sles/15-SP2/html/SLES-all/cha-iscsi.html>

<https://learn.microsoft.com/en-us/azure/sap/large-instances/ha-setup-with-fencing-device>

Anexo A

```
sbdhost:~ # date
Mon May 8 19:02:08 WEST 2023
sbdhost:~ # _

slaveA2:~ # _

sbdhost:~ # chronyc sources
=====
* ntp1.tecnico.ulisboa.pt 2 6 77 55 +285us[+1114us] +/- 43ms
sbdhost:~ # S

slaveA2:~ # _

slaveA1:~ # _

masterhost:~ # _
```

fig.1 - configuração NTP dos nodes, master e sbd

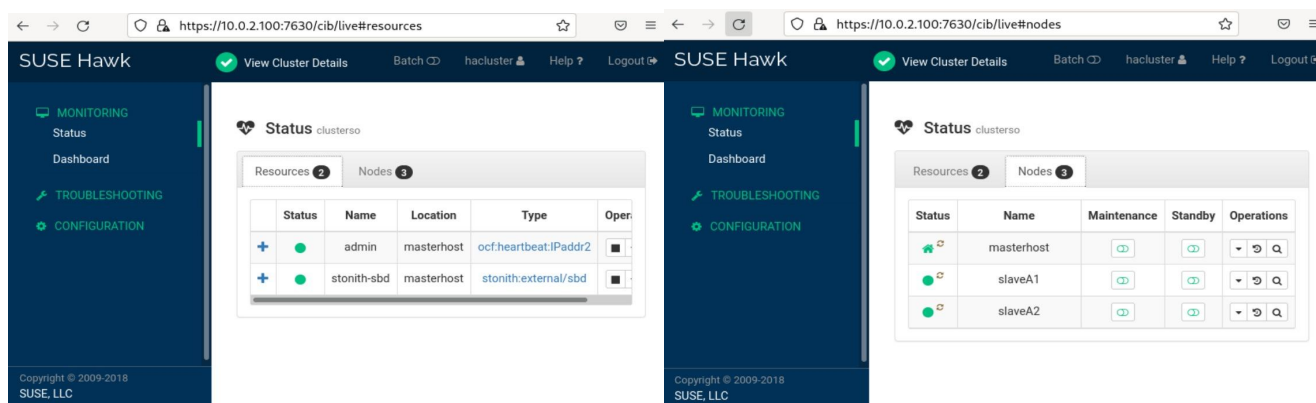


fig.2 - web tool do cluster Hawk2 (ip público 10.0.2.100:7630)

```
Cluster Summary:
* Stack: corosync
* Current DC: masterhost (version 2.1.2+20211124.ada5c3b36-150400.4.9.2-2.1.2+20211124.ada5c3b36)
partition with quorum
* Last updated: Mon May 8 19:20:25 2023
* Last change: Mon May 8 18:17:56 2023 by hacluster via cibadmin on masterhost
* 3 nodes configured
* 2 resource instances configured

Node List:
* Online: [ masterhost slaveA1 slaveA2 ]

Active Resources:
* stonith-sbd (stonith:external/sbd): Started masterhost
* admin (ocf::heartbeat:IPaddr2): Started masterhost
```

fig.2.1 - informações base do cluster, incluindo resources disponíveis.

```
masterhost:~/Sistemas-operativos # lsblk
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINTS
sda 8:0 0 20G 0 disk
├─sda1 8:1 0 8M 0 part
├─sda2 8:2 0 12.6G 0 part /usr/local
│ /var
│ /opt
│ /srv
│ /snapshots
│ /boot/grub2/i386-pc
│ /root
│ /tmp
│ /boot/grub2/x86_64-efi
└─sda3 8:3 0 5.4G 0 part /home
sda4 8:4 0 2G 0 part [SWAP]
sdb 8:16 0 4G 0 disk
sr0 11:0 1 12.1G 0 rom
masterhost:~/Sistemas-operativos #
```

fig.3 - configurações do sbd, target iscsi e confirmação de reach nodes-sbd

```
slaveA1:~/Sistemas-operativos # iscsiadm -m node
10.0.2.14:3260,1 iqn.2023-05.sbdstorage.:101c57a2944ff62f9ce1

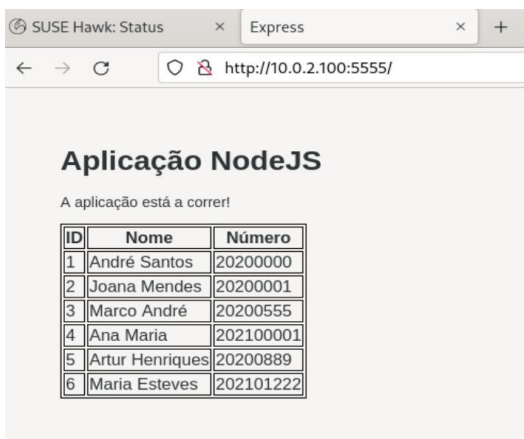
masterhost:~/Sistemas-operativos # sbd -d /dev/sdb list
0 masterhost clear
1 slaveA1 clear
```

```

masterhost:/Sistemas-operativos # ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP group default qlen 1000
    link/ether 00:00:27:e3:d1:3d brd ff:ff:ff:ff:ff:ff
    altname enp8s3
    inet 10.0.2.17/24 brd 10.0.2.255 scope global eth0
        valid_lft forever preferred_lft forever
    inet6 fe80::a00:27ff:fe3d:d13d/64 scope link
        valid_lft forever preferred_lft forever
3: eth1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP group default qlen 1000
    link/ether 00:00:27:36:9c:f3 brd ff:ff:ff:ff:ff:ff
    altname enp8s8
    inet 10.0.2.10/24 brd 10.0.2.255 scope global eth1
        valid_lft forever preferred_lft forever
    inet 10.0.2.100/24 brd 10.0.2.255 scope global secondary eth1
        valid_lft forever preferred_lft forever
    inet6 fe80::a00:27ff:fe36:9cf3/64 scope link
        valid_lft forever preferred_lft forever

```

fig.4 - eth0 e eth1 no master node e confirmação da atribuição do ip virtual do cluster 10.0.2.100



Aplicação NodeJS

A aplicação está a correr!

ID	Nome	Número
1	André Santos	20200000
2	Joana Mendes	20200001
3	Marco André	20200555
4	Ana Maria	202100001
5	Artur Henriques	20200889
6	Maria Esteves	202101222

fig.5 - serviço nodejs a correr no cluster, exibindo tabela de estudantes armazenada em cloud (aws)


```

slaveA1:/etc/postfix # mailx -r "sistemasoperativos@zohomail.eu" "rubenbatistamartins@gmail.com"
Subject: Email teste

Cluster sistemas operativos
EOT

```

fig.6 - serviço mail, (mensagem de email teste enviada com sucesso)



Email teste

sistemasoperativos@zohomail.eu

para mim

Cluster sistemas operativos

Responder Encaminhar

fig.6.1 - confirmação de recepção de mensagem email teste

```

slaveA1:/etc/postfix # mailx -r "sistemasoperativos@zohomail.eu" "rubenbatistamartins@gmail.com"
Subject: Teste
EOT
slaveA1:/etc/postfix # sendmail: fatal: parameter inet_interfaces: no local interface found for 10.0.2.100

```

fig.6.2 - tentativa de envio de mensagem email de server em "maintenance", ou seja, do cluster por momentos. Falha o envio pois o serviço passou para outro node do cluster.

Anexo B

Boot order:

boot: first storage vm then node (because storage vm is sbd target)

shutdown: first nodes then storage vm

Requisitos iniciais:

NTP configured to "ntp1.tecnico.ulisboa.pt"

All machines configured to the same NAT Network (static IP addresses)

List all cluster nodes in the /etc/hosts file with their fully qualified host name and short host name.

Configure iSCSI Storage

Server:

zypper install yast2-iscsi-lio-server

yast2 iscsi-lio-server (configure)

Client:

zypper install yast2-iscsi-client

systemctl start iscsid

yast2 iscsi-client (configure)

Add the stable device name to /etc/sysconfig/sbd

(SBD_DEVICE=/dev/disk/by-id/scsi-SLIO-ORG_IBLOCK_DEVICE_ID_STRING)

Timeline dos comandos:

zypper install -t pattern ha_sles (all nodes)

Using SBD for NodeFencing:

echo softdog > /etc/modules-load.d/watchdog.conf (each node)

systemctl restart systemd-modules-load

Setting Up the First Node

ha-cluster-init --name CLUSTERNAME (configure)

Adding the Second Node

ha-cluster-join (configure)

Hawk Web Interface:

username: hacluster

password: linux