**Cài đặt Redis-Cluster + Load-blancing (High Availability)**

Follow this procedure to setup a High Available Redis Caching Layer.

**Environment**

1. - haproxy
2. - redis
3. - redis-sentinel

Haproxy 1.5+ do have a TCP health check feature for redis. Haproxy will be configured in such a way that proxy connections will be forwarded to master instance only. Sentinel will constantly monitor the redis master instance and will promote slave node with lowest priority as next redis master in case of failure. Redis will be configured in such a way that one node will be master and other two nodes will be configured as slaveof the master instance.

Sentinel will be running as a separate service. we'll need minimum 3 sentinel instances to monitor redis master instance. Our cluster will have quorum of 2 ie; two sentinel instances should agree / vote for a slave to be promoted as master incase of master failure.

**Architecture Diagram**

**Server setup**

**Redis, Redis-sentinel instances - redis-server-001,002,003**

Install system updates

yum update –y

Install remi repository

centos 6

rpm -Uvh http://download.fedoraproject.org/pub/epel/6/x86\_64/epel-release-6-8.noarch.rpm

rpm -Uvh http://rpms.famillecollet.com/enterprise/remi-release-6.rpm

centos 7

rpm -Uvh http://rpms.famillecollet.com/enterprise/remi-release-7.rpm

Install redis

yum --enablerepo=remi install redis

Configure redis and redis-sentinel

Create redis-sentinel working directory

mkdir /var/lib/redis-sentinel

Download and install redis and redis-sentinel configuration from redis-server-001,002 and 003 directories and Update file ownership, redis user should have write privileges.

chown redis /etc/redis.conf

chown redis /etc/redis-sentinel.conf

Bootstrap Redis HA enviorment - Start Services and add services into system startup

chkconfig redis on ; chkconfig redis-sentinel on

service redis start ; service redis-sentinel start

**Haproxy Instance - haproxy-server-001**

Install system updates

yum update -y

Install epel repository

yum -y install epel-release

Install haproxy

yum -y install haproxy

Configure haproxy

cp /etc/haproxy/haproxy.cfg /etc/haproxy/haproxy.cfg\_bak

Download and install haproxy configuration from haproxy-server-001 directory

File cấu hình HAProxy:/etc/haproxy/haproxy.cfg

defaults REDIS

mode tcp

timeout connect 4s

timeout server 30s

timeout client 30s

frontend ft\_redis

bind HAPROXY\_IPADDR:6379 name redis

default\_backend bk\_redis

backend bk\_redis

option tcp-check

tcp-check send PING\r\n

tcp-check expect string +PONG

tcp-check send info\ replication\r\n

tcp-check expect string role:master

tcp-check send QUIT\r\n

tcp-check expect string +OK

server R1 REDIS\_001\_IPADDR:6379 check inter 1s

server R2 REDIS\_002\_IPADDR:6379 check inter 1s

server R2 REDIS\_003\_IPADDR:6379 check inter 1s

Hoặc có thể cấu hình theo file cấu hình dưới, cài đặt haproxy lên trực tiếp 3 nodes: /etc/haproxy/haproxy.cfg

global

log 127.0.0.1 local0

log 127.0.0.1 local1 debug

maxconn 45000 # Total Max Connections. This is dependent on ulimit

daemon

nbproc 8 # Number of processing cores. Dual Dual-core Opteron is 4 cores for example.

defaults

log global

option forwardfor except 127.0.0.1

option abortonclose

option httpclose

option http-server-close

mode http

maxconn 45000

timeout server 6s

timeout connect 3s

timeout client 6s

timeout queue 60s

timeout http-keep-alive 20s

timeout http-request 20s

# Redis

listen cluster\_redis \*:6378

mode tcp

option forceclose

option tcp-check

#balance roundrobin

tcp-check connect

tcp-check send PING\r\n

tcp-check expect string +PONG

tcp-check send info\ replication\r\n

tcp-check expect string role:master

tcp-check send QUIT\r\n

tcp-check expect string +OK

server redis\_node1 REDIS\_001\_IPADDR:6379 check inter 1s

server redis\_node2 REDIS\_002\_IPADDR:6379 check inter 1s

server redis\_node3 REDIS\_003\_IPADDR:6379 check inter 1s

Start Haproxy service and service into system startup

service haproxy start

chkconfig haproxy on

**Các điểm lưu ý:**

1. Redis master running on redis-server-001 instance
2. Redis slaves running on redis-server-002 and redis-server-003 instances
3. Redis Sentinel services running on all 3 redis nodes
4. Haproxy Server listerning on port 6379 and forwarding connections to master redis instance.
5. Upon master failure, redis-sentinel service will promote a slave into the master and haproxy's tcp health-check will detect next master and forward connections to the new master instance.

We setup multiple HAProxy instances and manage them using cluster management services. That's how we ensure redundancy and HA for haproxy instances.

**File cấu hình:**

**Lưu ý:**

Replace : REDIS\_001\_IPADDR with Node1's IP Address

Replace : REDIS\_002\_IPADDR with Node2's IP Address

Replace : REDIS\_003\_IPADDR with Node3's IP Address

Replace : CLUSTER\_NAME with your identity

**Node 1: Chạy master**

**File /etc/redis.conf:**

################################ GENERAL CONFIGURATION ###############################

bind REDIS\_001\_IPADDR

port 6379

dir /var/lib/redis

protected-mode yes

tcp-backlog 511

timeout 0

tcp-keepalive 300

daemonize no

supervised no

pidfile /var/run/redis\_6379.pid

loglevel notice

logfile /var/log/redis/redis.log

databases 16

################################ SNAPSHOTTING ################################

# Snapshot settings

save 900 1

save 300 10

save 60 10000

stop-writes-on-bgsave-error yes

rdbcompression yes

rdbchecksum yes

dbfilename dump.rdb

dir /var/lib/redis/

################################# REPLICATION #################################

slave-serve-stale-data yes

slave-read-only yes

repl-diskless-sync no

repl-diskless-sync-delay 5

repl-disable-tcp-nodelay no

# repl-backlog-size 1mb

# repl-backlog-ttl 3600

slave-priority 100

############################## APPEND ONLY MODE ###############################

appendonly no

appendfilename "appendonly.aof"

# appendfsync always

appendfsync everysec

# appendfsync no

no-appendfsync-on-rewrite no

auto-aof-rewrite-percentage 100

auto-aof-rewrite-min-size 64mb

aof-load-truncated yes

################################ LUA SCRIPTING ###############################

lua-time-limit 5000

################################## SLOW LOG ###################################

slowlog-log-slower-than 10000

slowlog-max-len 128

################################ LATENCY MONITOR ##############################

latency-monitor-threshold 0

############################# EVENT NOTIFICATION ##############################

notify-keyspace-events ""

############################### ADVANCED CONFIG ###############################

hash-max-ziplist-entries 512

hash-max-ziplist-value 64

list-max-ziplist-size -2

list-compress-depth 0

set-max-intset-entries 512

zset-max-ziplist-entries 128

zset-max-ziplist-value 64

hll-sparse-max-bytes 3000

activerehashing yes

client-output-buffer-limit normal 0 0 0

client-output-buffer-limit slave 256mb 64mb 60

client-output-buffer-limit pubsub 32mb 8mb 60

hz 10

aof-rewrite-incremental-fsync yes

**File /etc/redis-sentinel.conf:**

bind REDIS\_001\_IPADDR

port 16379

dir /var/lib/redis-sentinel

logfile /var/log/redis/redis-sentinel.log

sentinel monitor CLUSTER\_NAME REDIS\_001\_IPADDR 6379 2

sentinel down-after-milliseconds CLUSTER\_NAME 10000

sentinel parallel-syncs CLUSTER\_NAME 1

sentinel failover-timeout CLUSTER\_NAME 20000

**Node 2: Chạy slave**

**File /etc/redis.conf:**

################################ GENERAL CONFIGURATION ###############################

bind REDIS\_002\_IPADDR

port 6379

dir /var/lib/redis

protected-mode yes

tcp-backlog 511

timeout 0

tcp-keepalive 300

daemonize no

supervised no

pidfile /var/run/redis\_6379.pid

loglevel notice

logfile /var/log/redis/redis.log

databases 16

################################ SNAPSHOTTING ################################

# Snapshot settings

save 900 1

save 300 10

save 60 10000

stop-writes-on-bgsave-error yes

rdbcompression yes

rdbchecksum yes

dbfilename dump.rdb

dir /var/lib/redis/

################################# REPLICATION #################################

slave-serve-stale-data yes

slave-read-only yes

repl-diskless-sync no

repl-diskless-sync-delay 5

repl-disable-tcp-nodelay no

# repl-backlog-size 1mb

# repl-backlog-ttl 3600

slaveof REDIS\_001\_IPADDR 6379

slave-priority 150

############################## APPEND ONLY MODE ###############################

appendonly no

appendfilename "appendonly.aof"

# appendfsync always

appendfsync everysec

# appendfsync no

no-appendfsync-on-rewrite no

auto-aof-rewrite-percentage 100

auto-aof-rewrite-min-size 64mb

aof-load-truncated yes

################################ LUA SCRIPTING ###############################

lua-time-limit 5000

################################## SLOW LOG ###################################

slowlog-log-slower-than 10000

slowlog-max-len 128

################################ LATENCY MONITOR ##############################

latency-monitor-threshold 0

############################# EVENT NOTIFICATION ##############################

notify-keyspace-events ""

############################### ADVANCED CONFIG ###############################

hash-max-ziplist-entries 512

hash-max-ziplist-value 64

list-max-ziplist-size -2

list-compress-depth 0

set-max-intset-entries 512

zset-max-ziplist-entries 128

zset-max-ziplist-value 64

hll-sparse-max-bytes 3000

activerehashing yes

client-output-buffer-limit normal 0 0 0

client-output-buffer-limit slave 256mb 64mb 60

client-output-buffer-limit pubsub 32mb 8mb 60

hz 10

aof-rewrite-incremental-fsync yes

**File /etc/redis-sentinel.conf:**

bind REDIS\_002\_IPADDR

port 16379

dir /var/lib/redis-sentinel

logfile /var/log/redis/redis-sentinel.log

sentinel monitor CLUSTER\_NAME REDIS\_001\_IPADDR 6379 2 # Vì node 1 là Master

sentinel down-after-milliseconds CLUSTER\_NAME 10000

sentinel parallel-syncs CLUSTER\_NAME 1

sentinel failover-timeout CLUSTER\_NAME 20000

**Node 3: Chạy slave File cấu hình** tương tự Node 2, chỉ thay REDIS\_002\_IPADDR bằng REDIS\_003\_IPADDR

**File /etc/redis.conf:**

################################ GENERAL CONFIGURATION ###############################

bind REDIS\_003\_IPADDR

port 6379

dir /var/lib/redis

protected-mode yes

tcp-backlog 511

timeout 0

tcp-keepalive 300

daemonize no

supervised no

pidfile /var/run/redis\_6379.pid

loglevel notice

logfile /var/log/redis/redis.log

databases 16

################################ SNAPSHOTTING ################################

# Snapshot settings

save 900 1

save 300 10

save 60 10000

stop-writes-on-bgsave-error yes

rdbcompression yes

rdbchecksum yes

dbfilename dump.rdb

dir /var/lib/redis/

################################# REPLICATION #################################

slave-serve-stale-data yes

slave-read-only yes

repl-diskless-sync no

repl-diskless-sync-delay 5

repl-disable-tcp-nodelay no

# repl-backlog-size 1mb

# repl-backlog-ttl 3600

slaveof REDIS\_001\_IPADDR 6379

slave-priority 200

############################## APPEND ONLY MODE ###############################

appendonly no

appendfilename "appendonly.aof"

# appendfsync always

appendfsync everysec

# appendfsync no

no-appendfsync-on-rewrite no

auto-aof-rewrite-percentage 100

auto-aof-rewrite-min-size 64mb

aof-load-truncated yes

################################ LUA SCRIPTING ###############################

lua-time-limit 5000

################################## SLOW LOG ###################################

slowlog-log-slower-than 10000

slowlog-max-len 128

################################ LATENCY MONITOR ##############################

latency-monitor-threshold 0

############################# EVENT NOTIFICATION ##############################

notify-keyspace-events ""

############################### ADVANCED CONFIG ###############################

hash-max-ziplist-entries 512

hash-max-ziplist-value 64

list-max-ziplist-size -2

list-compress-depth 0

set-max-intset-entries 512

zset-max-ziplist-entries 128

zset-max-ziplist-value 64

hll-sparse-max-bytes 3000

activerehashing yes

client-output-buffer-limit normal 0 0 0

client-output-buffer-limit slave 256mb 64mb 60

client-output-buffer-limit pubsub 32mb 8mb 60

hz 10

aof-rewrite-incremental-fsync yes

**File /etc/redis-sentinel.conf:**

bind REDIS\_003\_IPADDR

port 16379

dir /var/lib/redis-sentinel

logfile /var/log/redis/redis-sentinel.log

sentinel monitor CLUSTER\_NAME REDIS\_001\_IPADDR 6379 2 # Vì node 1 là Master

sentinel down-after-milliseconds CLUSTER\_NAME 10000

sentinel parallel-syncs CLUSTER\_NAME 1

sentinel failover-timeout CLUSTER\_NAME 20000

**Ví dụ với 2 node: Ta cấu hình các file cấu hình reidis và haproxy như sau:**

**Ta sử dụng cấu hình file redis-sentinel.conf như sau:**

Node 1, Master:

bind 10.84.70.145

port 16378

dir "/var/lib/redis-sentinel"

logfile "/var/log/redis/redis-sentinel.log"

sentinel myid b4500470f622364e03689001a2d6ad0434f630c7

sentinel monitor redis-cluster 10.84.70.146 6378 2

sentinel down-after-milliseconds redis-cluster 10000

sentinel failover-timeout redis-cluster 20000

# Generated by CONFIG REWRITE

supervised systemd

sentinel config-epoch redis-cluster 44

sentinel leader-epoch redis-cluster 44

sentinel known-slave redis-cluster 10.84.70.145 6378

sentinel known-sentinel redis-cluster 10.84.70.146 16378 73c3633c769a8bb99d06f6185edea0613888f105

sentinel current-epoch 44

Node 2, Slave:

bind 10.84.70.146

port 16378

dir "/var/lib/redis-sentinel"

logfile "/var/log/redis/redis-sentinel.log"

sentinel myid 73c3633c769a8bb99d06f6185edea0613888f105

#sentinel deny-scripts-reconfig yes

sentinel monitor redis-cluster 10.84.70.146 6378 2

sentinel down-after-milliseconds redis-cluster 10000

# Generated by CONFIG REWRITE

supervised systemd

sentinel failover-timeout redis-cluster 20000

sentinel config-epoch redis-cluster 44

sentinel leader-epoch redis-cluster 44

sentinel known-slave redis-cluster 10.84.70.145 6378

sentinel known-sentinel redis-cluster 10.84.70.145 16378 b4500470f622364e03689001a2d6ad0434f630c7

sentinel current-epoch 44

**Ta sử dụng cấu hình file redis.conf như sau:**

Node 1, Master:

bind 10.84.70.145

port 6378

dir "/var/lib/redis"

protected-mode yes

tcp-backlog 511

timeout 0

tcp-keepalive 300

daemonize no

supervised systemd

pidfile "/var/run/redis\_6378.pid"

loglevel notice

logfile "/var/log/redis/redis.log"

databases 16

################################ SNAPSHOTTING ################################

# Snapshot settings

#save 900 1

#save 300 10

#save 60 10000

stop-writes-on-bgsave-error yes

rdbcompression yes

rdbchecksum yes

dbfilename "dump.rdb"

################################# REPLICATION #################################

slave-serve-stale-data yes

slave-read-only yes

repl-diskless-sync no

repl-diskless-sync-delay 5

repl-disable-tcp-nodelay no

# repl-backlog-size 1mb

# repl-backlog-ttl 3600

slave-priority 100

############################## APPEND ONLY MODE ###############################

#cluster-enabled yes

appendonly no

appendfilename "appendonly.aof"

# appendfsync always

appendfsync everysec

# appendfsync no

no-appendfsync-on-rewrite no

auto-aof-rewrite-percentage 100

auto-aof-rewrite-min-size 64mb

aof-load-truncated yes

################################ LUA SCRIPTING ###############################

lua-time-limit 5000

################################## SLOW LOG ###################################

slowlog-log-slower-than 10000

slowlog-max-len 128

################################ LATENCY MONITOR ##############################

latency-monitor-threshold 0

############################# EVENT NOTIFICATION ##############################

notify-keyspace-events ""

############################### ADVANCED CONFIG ###############################

hash-max-ziplist-entries 512

hash-max-ziplist-value 64

list-max-ziplist-size -2

list-compress-depth 0

set-max-intset-entries 512

zset-max-ziplist-entries 128

zset-max-ziplist-value 64

hll-sparse-max-bytes 3000

activerehashing yes

client-output-buffer-limit normal 0 0 0

client-output-buffer-limit slave 256mb 64mb 60

client-output-buffer-limit pubsub 32mb 8mb 60

hz 10

aof-rewrite-incremental-fsync yes

# Generated by CONFIG REWRITE

slaveof 10.84.70.146 6378

Node 2, Slave:

################################ GENERAL CONFIGURATION ###############################

bind 10.84.70.146

port 6378

dir "/var/lib/redis"

protected-mode yes

tcp-backlog 511

timeout 0

tcp-keepalive 300

daemonize no

supervised systemd

pidfile "/var/run/redis\_6378.pid"

loglevel notice

logfile "/var/log/redis/redis.log"

databases 16

################################ SNAPSHOTTING ################################

# Snapshot settings

#save 900 1

#save 300 10

#save 60 10000

stop-writes-on-bgsave-error yes

rdbcompression yes

rdbchecksum yes

dbfilename "dump.rdb"

################################# REPLICATION #################################

slave-serve-stale-data yes

slave-read-only yes

repl-diskless-sync no

repl-diskless-sync-delay 5

repl-disable-tcp-nodelay no

# repl-backlog-size 1mb

# repl-backlog-ttl 3600

slave-priority 100

############################## APPEND ONLY MODE ###############################

#cluster-enabled yes

appendonly no

appendfilename "appendonly.aof"

# appendfsync always

appendfsync everysec

# appendfsync no

no-appendfsync-on-rewrite no

auto-aof-rewrite-percentage 100

auto-aof-rewrite-min-size 64mb

aof-load-truncated yes

################################ LUA SCRIPTING ###############################

lua-time-limit 5000

################################## SLOW LOG ###################################

slowlog-log-slower-than 10000

slowlog-max-len 128

################################ LATENCY MONITOR ##############################

latency-monitor-threshold 0

############################# EVENT NOTIFICATION ##############################

notify-keyspace-events ""

############################### ADVANCED CONFIG ###############################

hash-max-ziplist-entries 512

hash-max-ziplist-value 64

list-max-ziplist-size -2

list-compress-depth 0

set-max-intset-entries 512

zset-max-ziplist-entries 128

zset-max-ziplist-value 64

hll-sparse-max-bytes 3000

activerehashing yes

client-output-buffer-limit normal 0 0 0

client-output-buffer-limit slave 256mb 64mb 60

client-output-buffer-limit pubsub 32mb 8mb 60

hz 10

aof-rewrite-incremental-fsync yes

# Generated by CONFIG REWRITE

**Ta sử dụng cấu hình file haproxy.cfg như sau:**

global

log 127.0.0.1 local0

log 127.0.0.1 local1 debug

maxconn 45000 # Total Max Connections. This is dependent on ulimit

daemon

nbproc 8 # Number of processing cores. Dual Dual-core Opteron is 4 cores for example.

defaults

log global

option forwardfor except 127.0.0.1

option abortonclose

option httpclose

option http-server-close

mode http

maxconn 45000

timeout server 6s

timeout connect 3s

timeout client 6s

timeout queue 60s

timeout http-keep-alive 20s

timeout http-request 20s

# Redis

listen cluster\_redis \*:6379

mode tcp

option forceclose

option tcp-check

#balance roundrobin

tcp-check connect

tcp-check send PING\r\n

tcp-check expect string +PONG

tcp-check send info\ replication\r\n

tcp-check expect string role:master

tcp-check send QUIT\r\n

tcp-check expect string +OK

server redis\_node1 10.84.70.145:6378 check inter 1s

server redis\_node2 10.84.70.146:6378 check inter 1s

listen stats 10.84.70.146:8123

mode http

stats enable

stats hide-version

stats refresh 30s

stats show-node

stats uri /haproxy\_stats

stats realm HAProxy\ Statistics

stats auth admin:smsmkt@123!

stats admin if TRUE

**Ta sử dụng cấu hình file /etc/sysctl.conf như sau:**

# sysctl settings are defined through files in

# /usr/lib/sysctl.d/, /run/sysctl.d/, and /etc/sysctl.d/.

#

# Vendors settings live in /usr/lib/sysctl.d/.

# To override a whole file, create a new file with the same in

# /etc/sysctl.d/ and put new settings there. To override

# only specific settings, add a file with a lexically later

# name in /etc/sysctl.d/ and put new settings there.

#

# For more information, see sysctl.conf(5) and sysctl.d(5).

#Disable slow start

net.ipv4.tcp\_slow\_start\_after\_idle=0

net.core.wmem\_max=12582912

net.core.wmem\_default = 2048000

net.core.rmem\_max=12582912

net.core.optmem\_max=25165824

net.ipv4.tcp\_window\_scaling=1

net.ipv4.tcp\_rmem=10240 87380 12582912

net.ipv4.udp\_mem=10240 87380 12582912

net.ipv4.tcp\_wmem=102400 2048000 12582912

net.ipv4.tcp\_mem=196608 262144 393216

net.ipv4.tcp\_congestion\_control=scalable

#no metrics -> avoid confuse for user behind NAT

net.ipv4.tcp\_no\_metrics\_save=1

net.ipv4.tcp\_moderate\_rcvbuf = 1

#enable fast open

#net.ipv4.tcp\_fastopen=1

net.ipv4.tcp\_frto=1

net.ipv4.tcp\_low\_latency=1

net.ipv4.tcp\_dsack = 1

net.ipv4.tcp\_sack = 1

net.ipv4.tcp\_timestamps = 0

#flush cache of ssth

net.ipv4.route.flush=1

net.ipv4.tcp\_adv\_win\_scale=1

#Add more local port

net.ipv4.ip\_local\_port\_range = 1024 61000

net.ipv4.tcp\_fin\_timeout = 10

#Change keep alive in Linux

net.ipv4.tcp\_keepalive\_time = 60

net.ipv4.tcp\_keepalive\_intvl = 20

net.ipv4.tcp\_keepalive\_probes = 10

#Increase network buffer

net.core.netdev\_max\_backlog = 300000

net.ipv4.tcp\_congestion\_control = cubic

net.core.somaxconn = 50000

fs.file-max = 9400000

#Fixing CVE-2016-5696 vulnerability

net.ipv4.tcp\_challenge\_ack\_limit = 999999999