GMOペパボ

Consul in Production

consul の用途いろいろ

柴田博志 / GMO PEPABO inc.

2016.08.01 Consul Casual Talks

GMOペパボ



柴田博志　@hsbt

チーフエンジニア

技術部技術基盤チーム

https://www.hsbt.org

self.introduce

```
=>
{
  github: “@hsbt”,
  commit_bits: [
    “ruby”, “rake”, “rubygems”, “rdoc”,
    “psych”, “syck”, “ruby-build”, “jenkins.rb”…
  ],
  sites: [
    “hsbt.org”, “ruby-lang.org”, “rubyci.org”,
    “railsgirls.com”, “railsgirls.jp”
  ],
}
```



consul でやってることを紹介

- minne (minne.com)でやってること in 2016/08
- minne は OpenStack 9割、AWS 1割の multi DC、VM は 100 くらい、全部に consul 入れてます
- consul の dc は使ってません、OpenStack と AWS で別のクラスタ
- consul + dnsmasq でやる内部 DNS は省略
- stretcher を使ったデプロイは省略
- consul で tls session ticket の更新は省略



おまけ: systemd

```
[Unit]
Description=Consul is a tool for service discovery and configuration. Consul is distributed, highly available, and extreme
Documentation=http://www.consul.io
After=network-online.target
Wants=network-online.target

[Service]
User=consul
Group=consul
EnvironmentFile=-/etc/sysconfig/consul
ExecStart=/usr/bin/consul $CMD_OPTS
ExecReload=/bin/kill -HUP $MAINPID
KillSignal=SIGINT
Restart=on-failure
LimitNOFILE=8192

[Install]
WantedBy=multi-user.target
```



NewRelic のホスト選定

こういうことをやりたい

- minne ではアプリケーションのパフォーマンスモニタリングに NewRelic (https://newrelic.com) を使っている
- NewRelic は高い($200/host)

  → バルクディスカウントを交渉すると $99/host くらいまで下がる
- それでもサービスのロールが 5 つあれば最低 5 host に入れる必要がある
- ロールの中からいい感じに 1 台だけ選んで NewRelic を有効にしたい



こうやった

- NewRelic の設定ファイル(Rails 用)に consul からホストを取得して 1 ホストだけ有効とするようなコードを埋め込む
- newrelic_rpm の gem がロードされた時に呼び出されて欲しいので Ruby のコードは `Rails.root.join('lib')` に置く
- Consul へのアクセスは Diplomat (https://github.com/WeAreFarmGeek/diplomat)を使う
- オリジナル: https://makandracards.com/makandra/19145-enable-newrelic-monitoring-for-rails-on-specific-hosts-only



こんな感じ

```
common: &default_settings

(snip)

test:
  <<: *default_settings
  monitor_mode: false

production:
  <<: *default_settings
  monitor_mode: <%= CustomNewRelicConfiguration.enable_on_current_host? rescue false %>
```



こんな感じ(2)

```
require 'socket'

class CustomNewRelicConfiguration
  def self.enable_on_current_host?
    current_hostname = Socket.gethostname.strip
    Rails.logger.info ">>> Checking whether to enable NewRelic on host '#{current_hostname}'"

    if newrelic_hosts == current_hostname
      Rails.logger.info ">>> NewRelic should be enabled on this host."
      true
    else
      Rails.logger.info ">>> NewRelic is disabled on this host! Edit config/newrelic.yml to enable."
      false
    end
  end

  def self.newrelic_hosts
    require 'diplomat'
    Diplomat.configuration.url = "http://127.0.0.1:8500"

    node = Diplomat::Node.get_all.
           select{|n| n.Node =~ /^www-/ }.
           sort_by{|n| n.Node }.
           first.Node
    "#{node}.minne.com"
  end
end
```



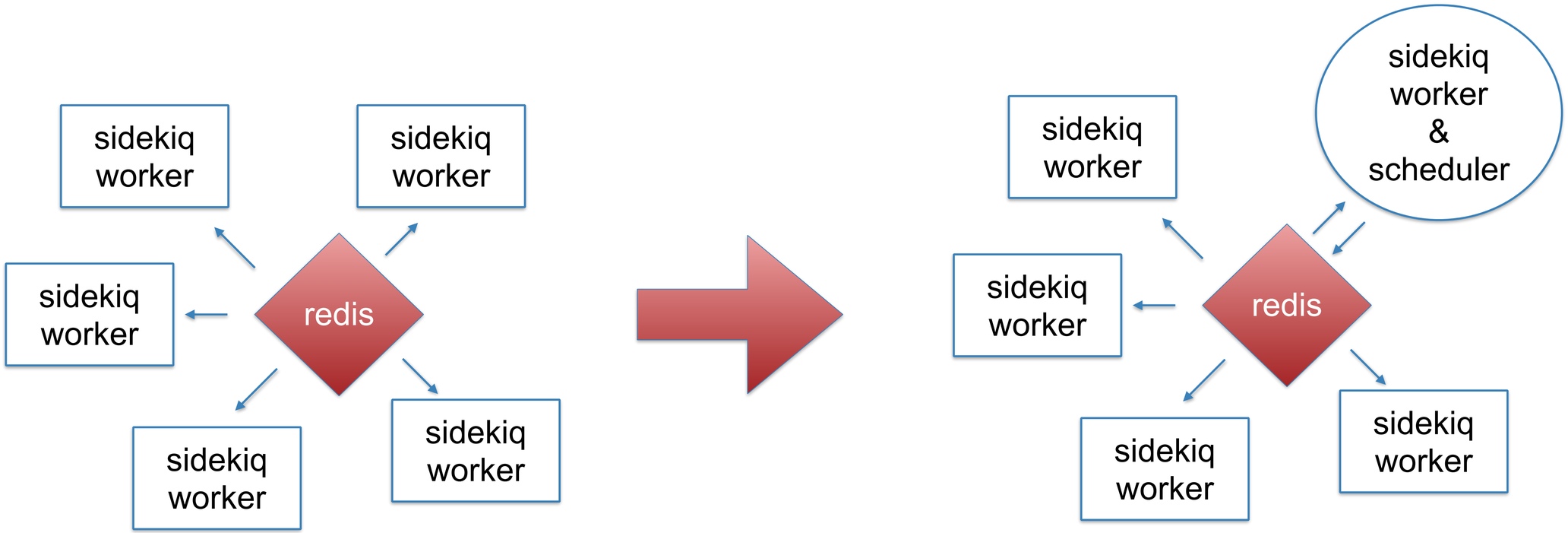
sidekiq-scheduler

こういうことをやりたい

- cron は 1vm が SPOF となりがちなのでやめたい
- https://github.com/moove-it/sidekiq-scheduler を使うと sidekiq に periodical job が使えるようになる(sidekiq pro でもできる)
- sidekiq-scheduler は単純に使うと worker 全台で有効になって、ジョブが重複して困る
- consul で常に一台だけ sidekiq scheduler を有効にしたい



概念図


sidekiq
worker
sidekiq
worker
sidekiq
worker
redis
sidekiq
worker
sidekiq
worker
sidekiq
worker
sidekiq
worker
&
scheduler
sidekiq
worker
sidekiq
worker
redis
sidekiq
worker
sidekiq
worker

こうやった

- consul を使って、一定のルールで一台だけ scheduler を有効とするようなホストネームを取り出す
- scheduler がいなくなると困るので、scheduler ホストがプラットフォームのトラブルで落ちた場合は自動で再選出する(まだできてない…)



こんな感じ

```
Sidekiq.configure_server do |config|
  if Rails.env.production? # for sidekiq-scheduler
    config.on(:startup) do
      if Sidekiq::Scheduler.master_server?
        Sidekiq::Scheduler.register_server
        Sidekiq::Scheduler.reload_schedule!
      else
        Sidekiq::Scheduler.enabled = false
      end
    end
    %i(quiet shutdown).each do |state|
      config.on(state) do
        Sidekiq::Scheduler.reset_master_server!
      end
    end
  end
end
```

```
module Sidekiq
  class Scheduler
    def self.master_server
      Diplomat::Node.get_all.
        select{|n| n.Node =~ /^worker-/ }.
        sort_by{|n| n.Node }.
        first.Node
    end

    def self.master_server?
      Socket.gethostname.chomp == self.master_server
    end

    def self.register_server
      Sidekiq.redis {|c| c.set(‘sidekiq:schedules:master’, Socket.gethostname.chomp) }
    end

    def self.reset_master_server!
      if Socket.gethostname.chomp == Sidekiq.redis {|c| c.get('sidekiq:schedules:master')}
        Sidekiq.redis {|conn| conn.del('sidekiq:schedules:master')}
      end
    end
  end
end
```



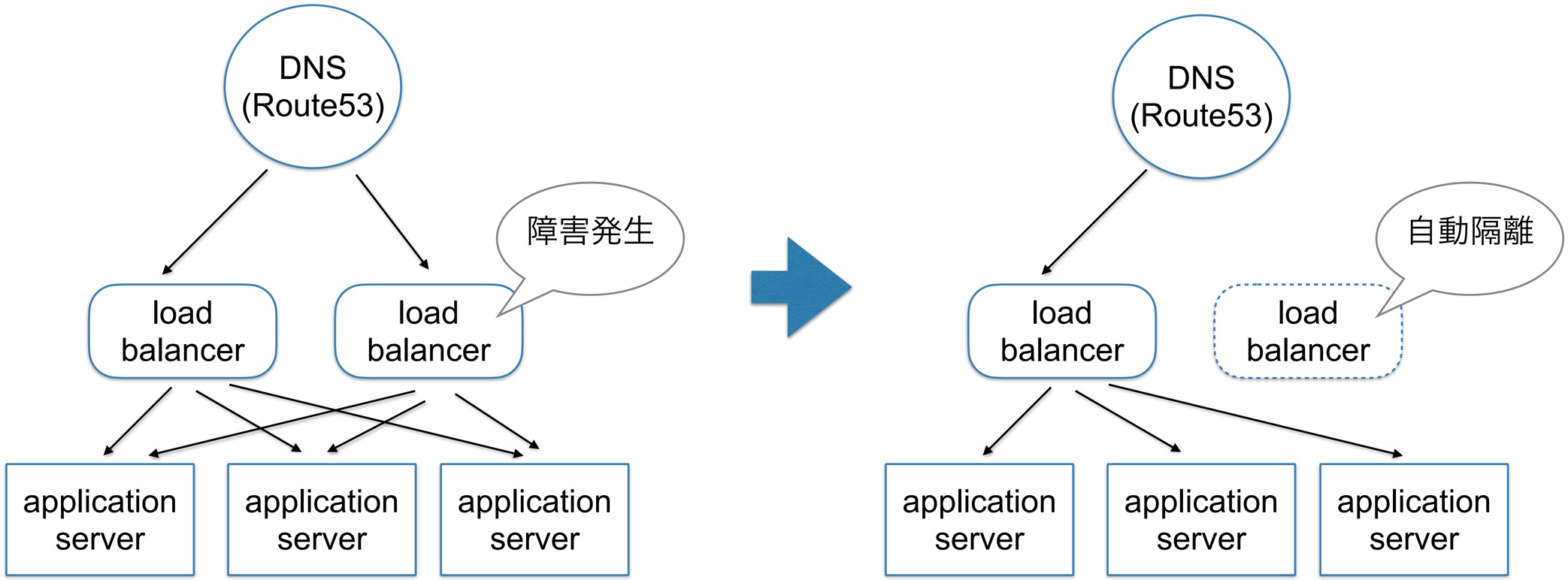
ロードバランサの DNS 更新

こういうことをやりたい

- load balancer として nginx を n 台利用している
- nginx の upstream として m 台配置し、クロスメッシュで繋いでいる
- load balancer を DNS ラウンドロビンで分散アクセスしている
- load balancer がピヨった時に DNS ラウンドロビンの対象から自動で外したい



概念図

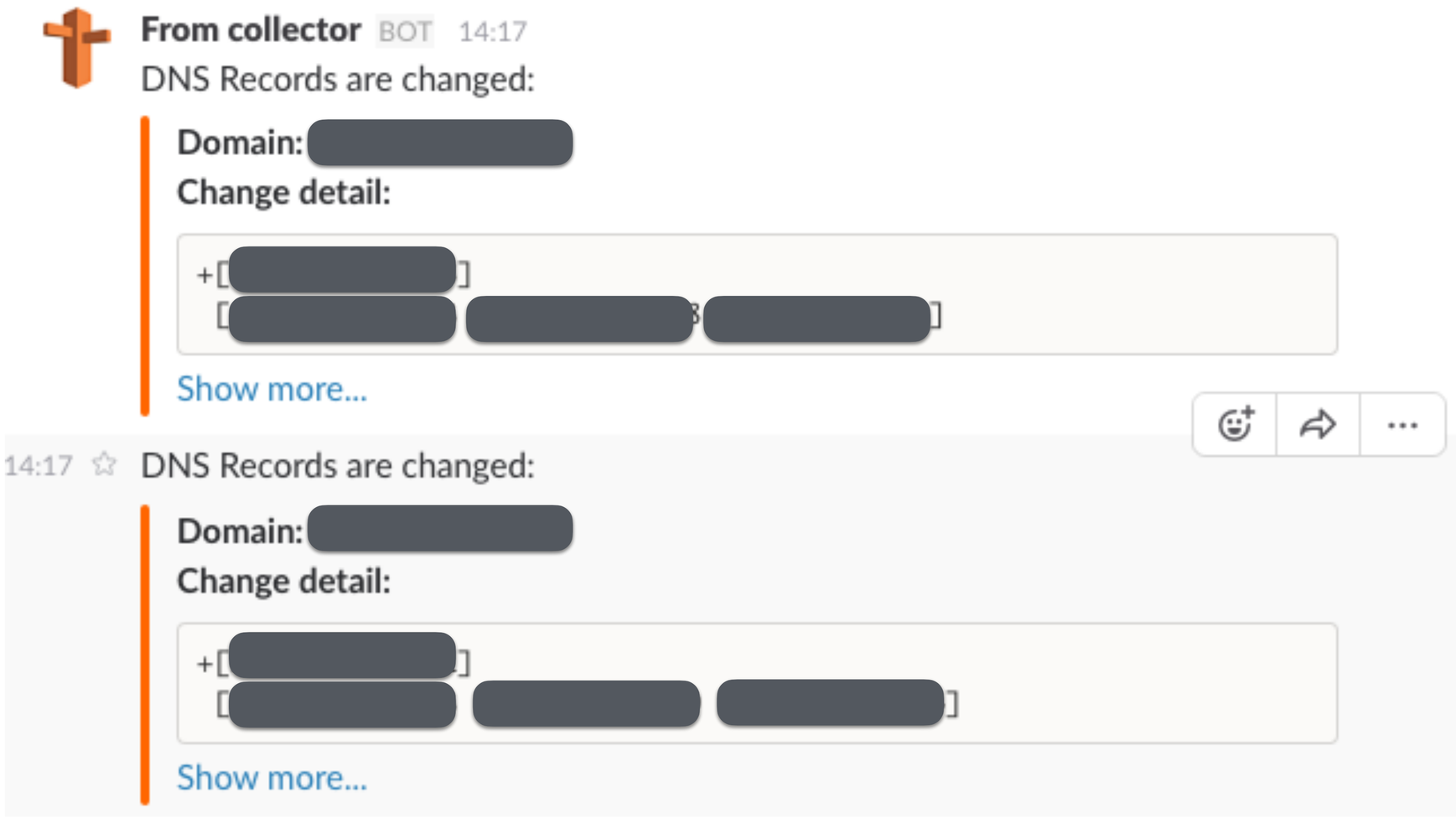

DNS
(Route53)
障害発生
load
balancer
load
balancer
application
server
application
server
application
server
DNS
(Route53)
自動隔離
load
balancer
load
balancer
application
server
application
server
application
server

こうやった

- https://github.com/udzura/collector を利用
- consul の service 状況を監視(nagios-plugin-all 使ってましたが mackerel に移行中)
- consul の service(load balancer の nginx)がダウンしている場合は該当の consul node を consul watch と collector を用いて dns から切りはずし
- 詳細は設定ファイルなどは README.md をみてください(省略)



こんな感じ


From collector BOT 14:17
DNS Records are changed:
Domain: [redacted]
Change detail:
+[redacted]
[redacted]
Show more...
14:17
DNS Records are changed:
Domain: [redacted]
Change detail:
+[redacted]
[redacted]
Show more...

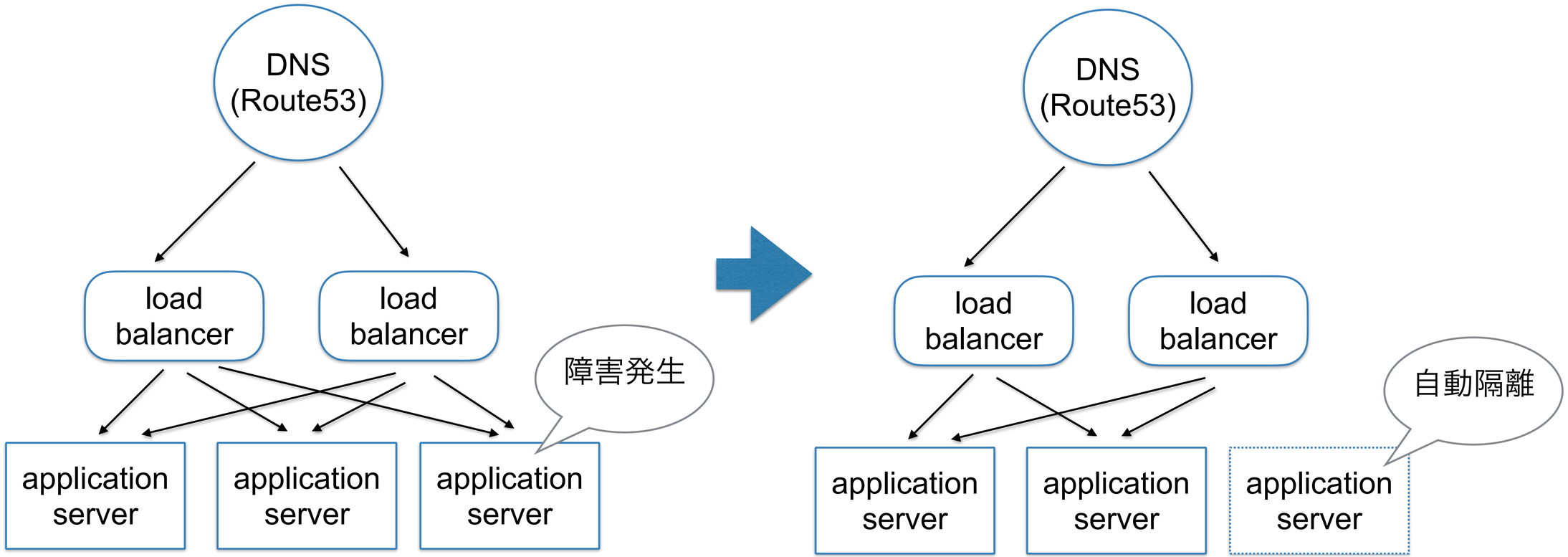
nginx の upstream の自動更新

こういうことをやりたい

- load balancer で使っている nginx の upstream を自動更新したい
- https://github.com/hashicorp/consul-template で nginx の configuration を自動生成
- 異常がある vm は upstream から外すし、新しく立ち上げた vm は自動で load balancer に join する
- 地雷: http://tomohisaoda.com/posts/2015/use_systemd_with_consul-template.html



概念図


DNS
(Route53)
load
balancer
load
balancer
application
server
application
server
application
server
障害発生
DNS
(Route53)
load
balancer
load
balancer
application
server
application
server
application
server
自動隔離

こうやった

- load balancer に consul-template を入れて素朴に実行
- configuration の command に nginx reload と slack 通知を挿入
- 余談:
  - consul cluster にルーティング間違えた node を入れて、クラスタ全体の調子が悪くなった際に 1min に 3-4 回 nginx を reload する症状が発生
  - ngx_mruby に reload するとメモリリークするバグがあり load balancer が崩壊…(hiboma が valgrind で発生箇所を特定して報告した)



同僚が直してくれた 😇

https://github.com/matsumoto-r/ngx_mruby/pull/183

まとめ

- consul を使うことで vm を束ねて一つのシステムとして扱うことが可能となる
- consul maint を活用できてないのでいい感じに vm を退避するみたいのを作りたい
- 全自動なめらかサーバーリソース制御くん、全自動なめらかジョブスケジューラを作っているところ
- こういうのをやりたい人は今すぐ僕とペパランチョン → https://pepabo.com/recruit/pepaluncheon/

