Modernizing HPC for Cloud

Steven Carter
Chief Cloud Architect, Public Sector
2 May 2017
Wall Clock VS Results
Science is Multidisciplinary
But, not just a Supercomputer in the cloud...
Posix Filesystem vs Object Storage
Monolithic VS MicroServices
Circuit Breaker using 3Scale

Dealing with Microservice failures

Tom Corcoran, Sr. Solution Architect, Red Hat
Veer Muchandi, Principal Architect, Red Hat
05/02/2017
Problem
Client calling service
Service Failure
Service failure effect multiplied
Circuit Breaker Implementation using 3Scale
Normal Operation - 3 Scale forwards requests to the main service

http://myappgateway/currencyconverter

API Gateway

main service

currencyconvertor.subdomain1.com

alternate service

currencyconvertor.subdomain2.com
Fail over - Circuit broken to the main service and requests forwarded to the alternate service.
Graceful handling on Alternative service failure with cached response

API Gateway

- Acts as circuit breaker object
- Cached Response

Main service
- currencyconverter.subdomain1.com

Alternate service
- currencyconverter.subdomain2.com
Graceful handling of Alternative service failure with Error response

- **API Gateway**
  - Acts as circuit breaker object
  - Error Response

- **Main service**
  - currencyconverror.subdomain1.com

- **Alternate service**
  - currencyconverror.subdomain2.com
Advantages of using 3Scale

Circuit breaker becomes an infrastructure service

- Works with simple configuration changes in the nginx.conf
- No code intrusion, No annotations in business logic

Underlying technology does not matter

Language Neutral
SUPPORT – SLIDE - Circuit Breaker – on Nginx

Nginx uses

- Upstreams – for API Back Ends

```nginx
upstream backend_sms_API {
    ip_hash;
    server subdomain1.sms-backend.com;
    server subdomain2.sms-backend.com;
}
```

- Locations – to capture incoming API requests and proxy pass them

```nginx
location /sms {
    proxy_pass http://backend_sms_API;
}
```
SUPPORT – SLIDE - Circuit Breaker – on Nginx

Back end failover – fine grained control

Override default selection of Next Upstream.

```bash
location /sms {
    proxy_next_upstream  error timeout invalid_header http_504 http_503;
    proxy_connect_timeout 2;
    proxy_pass http://backend_sms_API;
}
```
SUPPORT – SLIDE - Circuit Breaker - on Nginx

Failure Mitigation

Slow Start

```
upstream backend_sms_API {
    slow_start=30s;
    server subdomain1.sms-backend.com;
    server subdomain2.sms-backend.com;
}
```

Rate Limiting

```
http {
    limit_req_zone $server_addr zone=moderateReqs:1m rate=100r/s;
    ... 
    server {
        ... 
        limit_req zone=moderateReqs burst=150;
        limit_req_status 503;
        ... 
    }
}
```
SUPPORT – SLIDE - Circuit Breaker – on Nginx

Last Resort Options – no back end available

Response caching.

```bash
proxy_cache_path /app/cache levels=1:2 keys_zone=oauth_cache:10m max_size=10m inactive=15s

location /v1/users {
    ....
    proxy_cache_valid 200 30s;
    proxy_cache_use_stale error timeout invalid_header updating
    http_500 http_502 http_503 http_504;
}
```

Error Page.

```bash
location / {
    error_page 503 = @fallback;
}

location @fallback {
    proxy_pass http://alternative-backend;
}
```
Fast as lightning: NVMe and NVDIMM's Impact on your data center

Tom Coughlan
Sr. Engineering Manager
Filesystem and Storage Team, Red Hat
May 2, 2017
I/O Latency

Little change over 40+ years...
An order of magnitude improvement

Flash and NVMe arrive.

I/O Latency

10 ns

I/O Ctrlr.

PCI

1 µs

HBA

SAS/SATA

1 ms

Flash SSD

FC, Ethernet

1 s

Local HDD

Shared Array

CPU

NVMe Flash

SSD
Another order of magnitude improvement
Lower latency for shared storage.
I/O Latency

Even lower-latency for shared storage...
RHEL Status

- Persistent Memory
  - Block-mode access supported in 7.3
  - Byte-mode direct access (DAX) is Tech. Preview in 7.3 (and 7.4)
    - Uses mmap on XFS or ext4
  - HPE results, using NVDIMMs for...
    - Oracle OLTP redologs

- PostgreSQL Write-Ahead Logging

- NVMeF - try it in RHEL 7.4 Beta
Additional sessions of interest...

- “On the path to persistent memory”
  BOF B113761
  Tuesday, 10:15 to 11 am (oops...) Room 158
  Linda Knippers (HPE) and Jeff Moyer (Red Hat)

- “Bring your performance and scale problems to the experts”
  Session 101558
  Tuesday 5:30 PM - 7:00 PM Room 205C

- “Utilizing Persistent Memory to Improve DB Performance and Reduce Costs”
  Session S111008
  Wednesday at 4:30 pm Room 105
  Karen Dorhamer (HPE)
THANK YOU

plus.google.com/+RedHat
linkedin.com/company/red-hat
youtube.com/user/RedHatVideos
facebook.com/redhatinc
twitter.com/RedHatNews

#redhat #rhsummit
LEARN. NETWORK. EXPERIENCE OPEN SOURCE.