

Clustering and Caching

eZ Camp 2005

Jun 3rd, 2005. Gaustablikk, Norway

Derick Rethans <dr@ez.no>

<http://derickrethans.nl/talks.php>

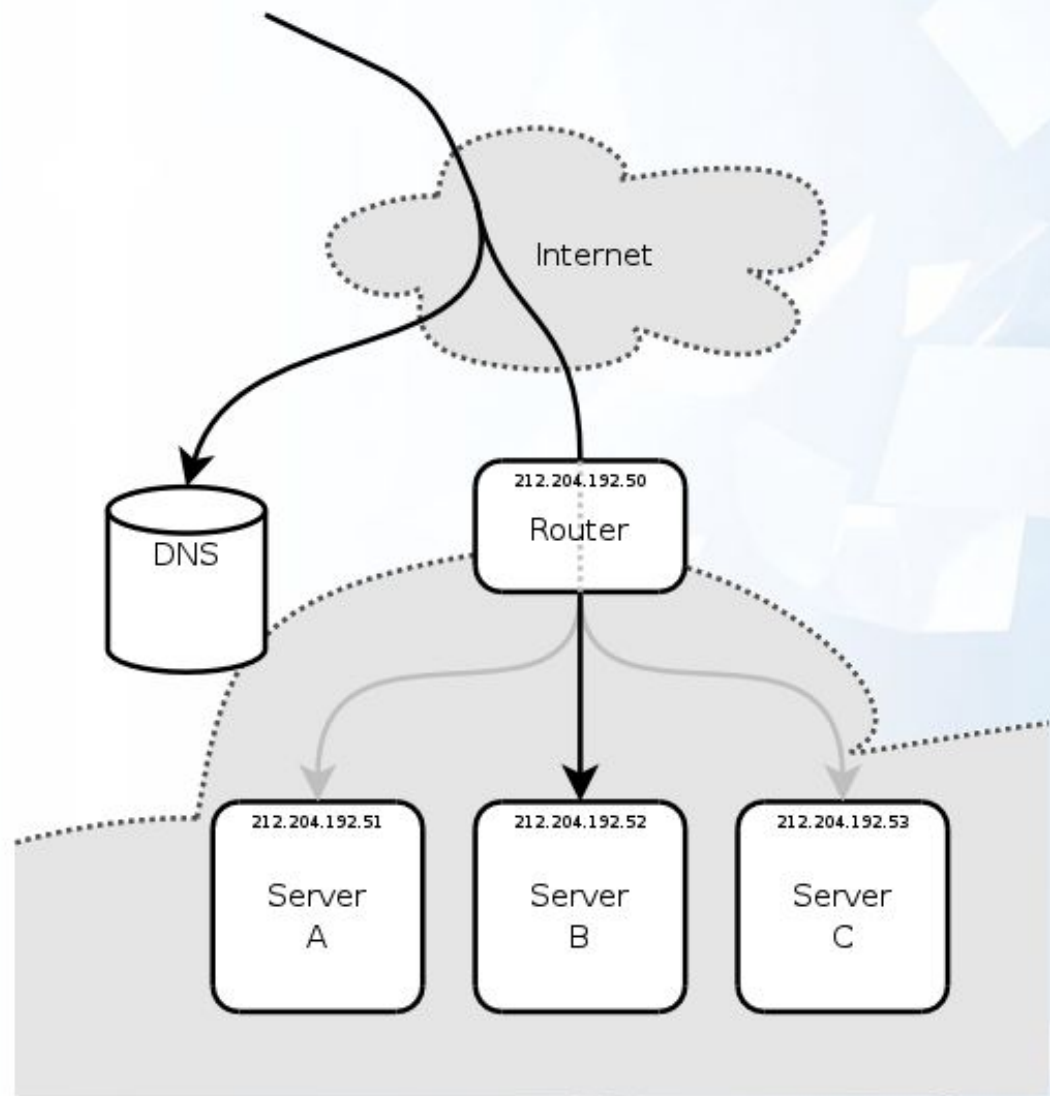
Load Balancing:

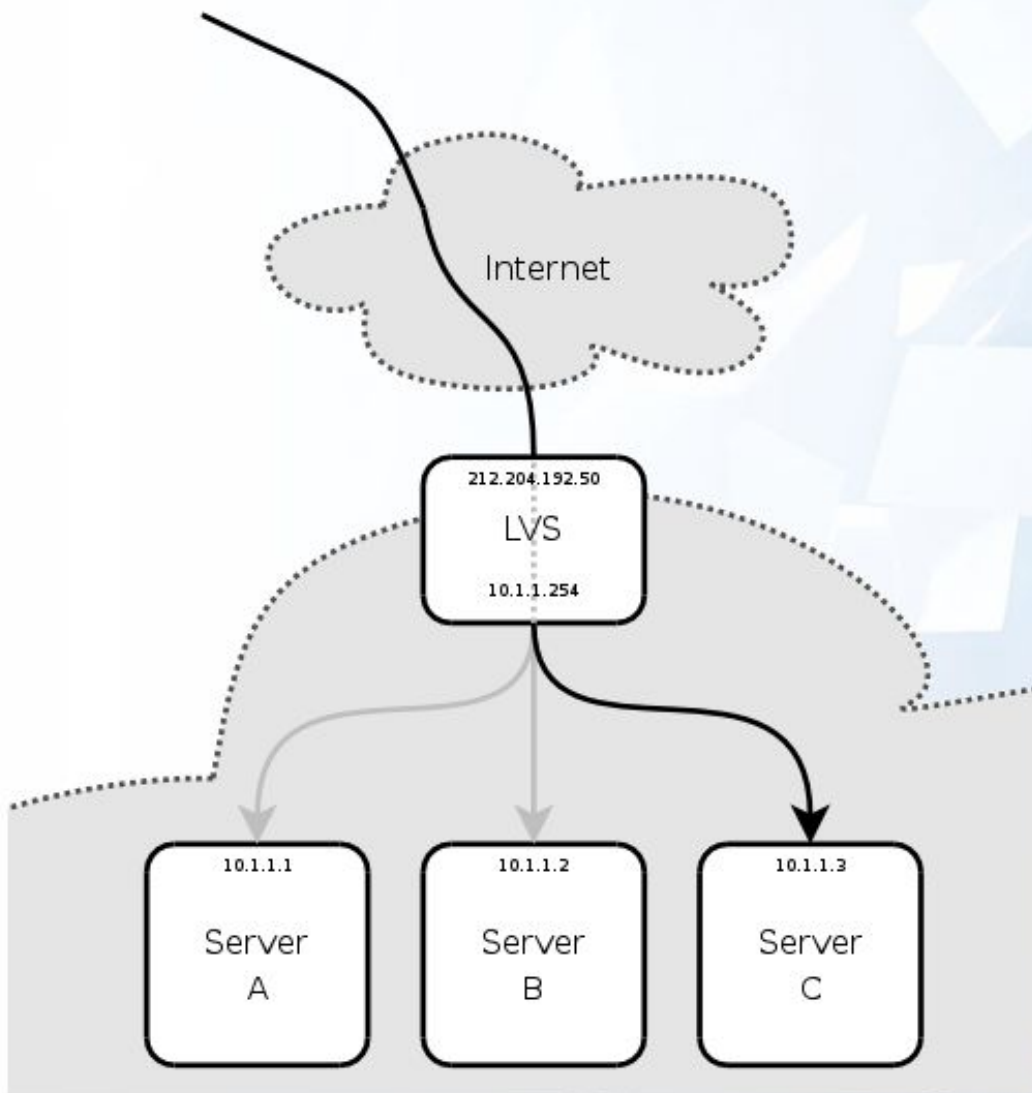
- A single server can not handle the load
- Faster machine does not always work
- Different bottlenecks: Disk IO, CPU or Network
- SMP has inherent problems

High Availability:

- Failover Capabilities
- High Availability (the 9s)

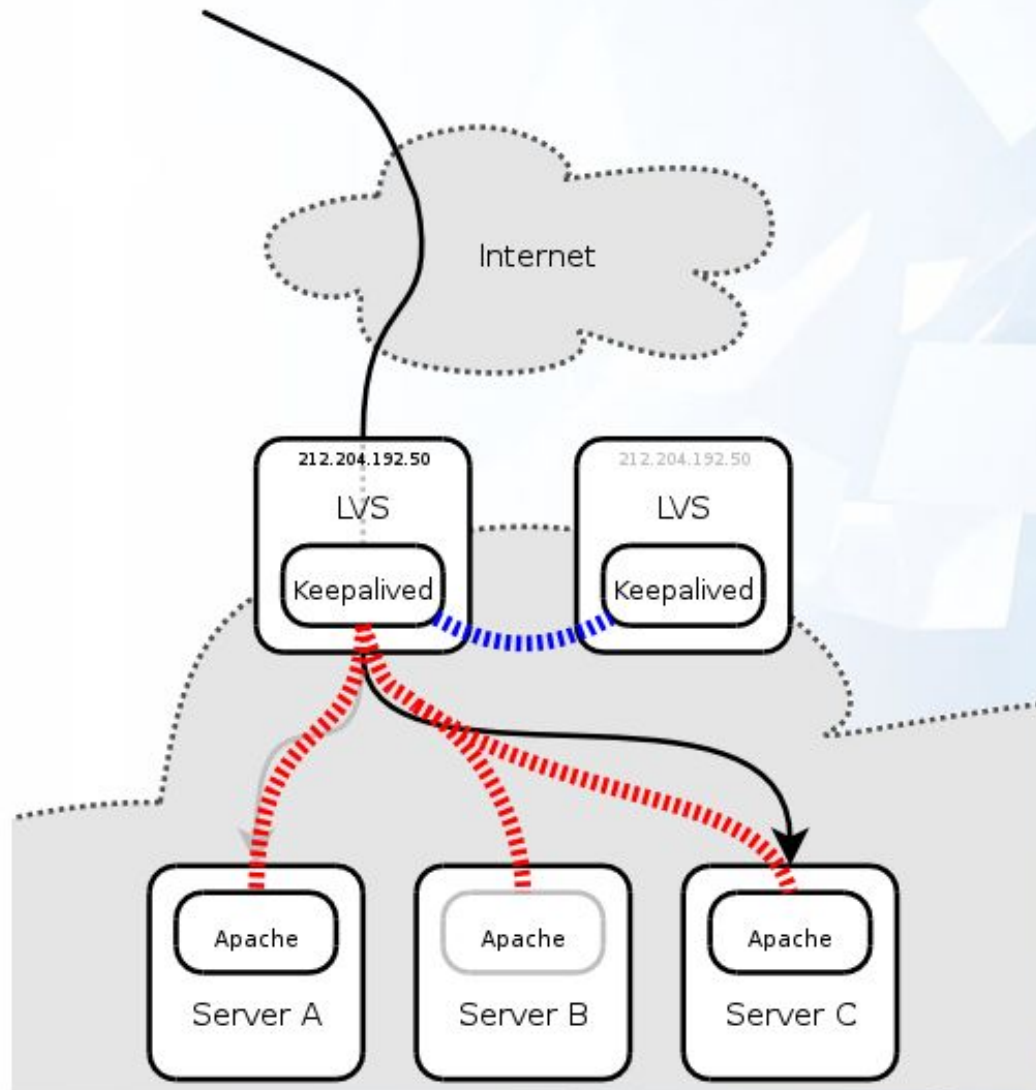
Share your information





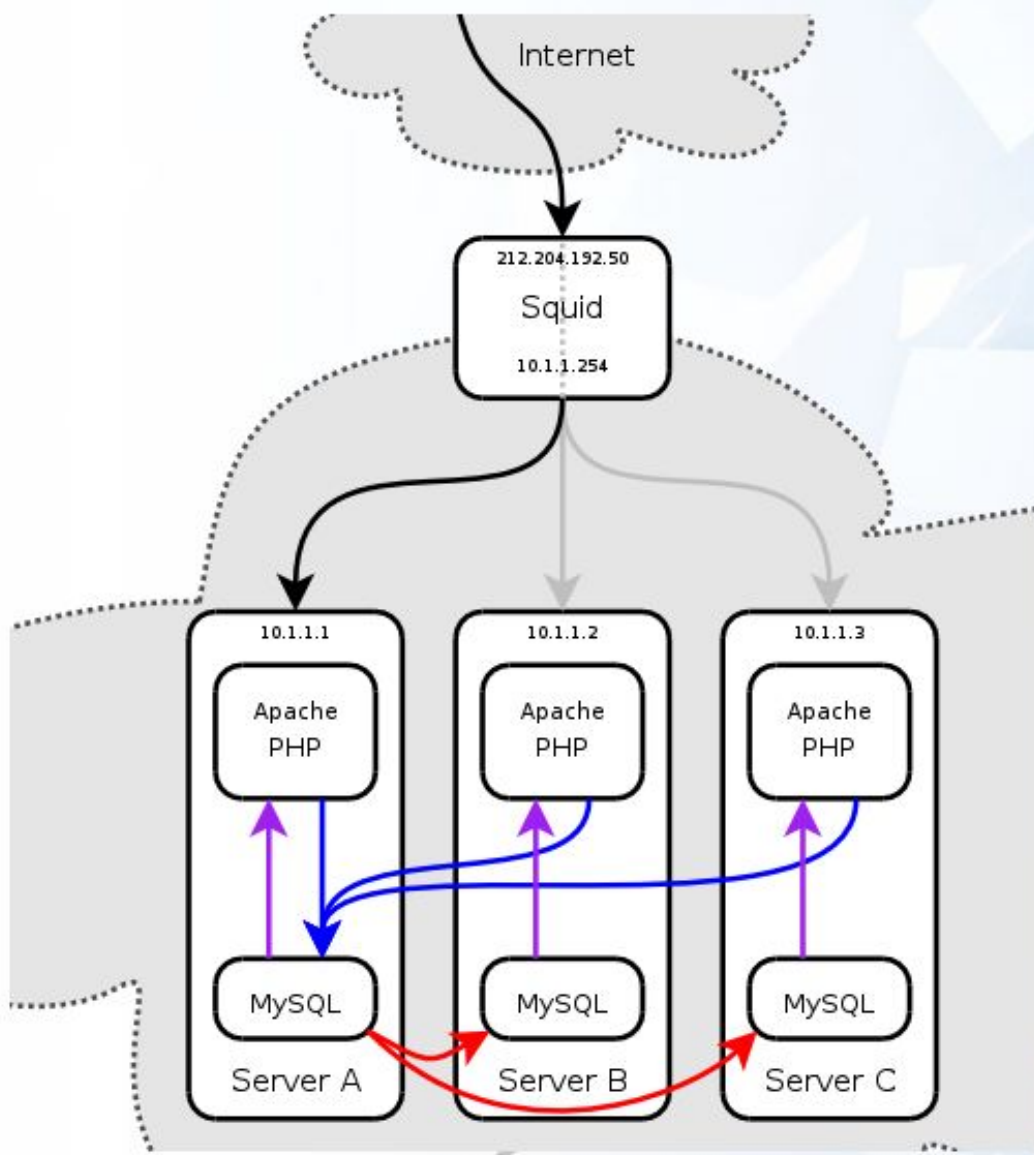
- Front End Server: Redirector (Load Balancer)
- Can use the load of specific worker servers to determine how to forward
- Uses a kernel module that implements transport-layer load balancing inside the Linux kernel, so called Layer-4 switching
- Can use different scheduling algorithms
- Does **not** provide fail-safe functionality

Share your information



- Hooks in into LVS
- Provides fail-over functionality
- Checks at different level: TCP, HTTP Get and Script

- Keepalived also implements failover with VRPP
- This allows you to have a Back-Up redirector to take over when the primary goes down

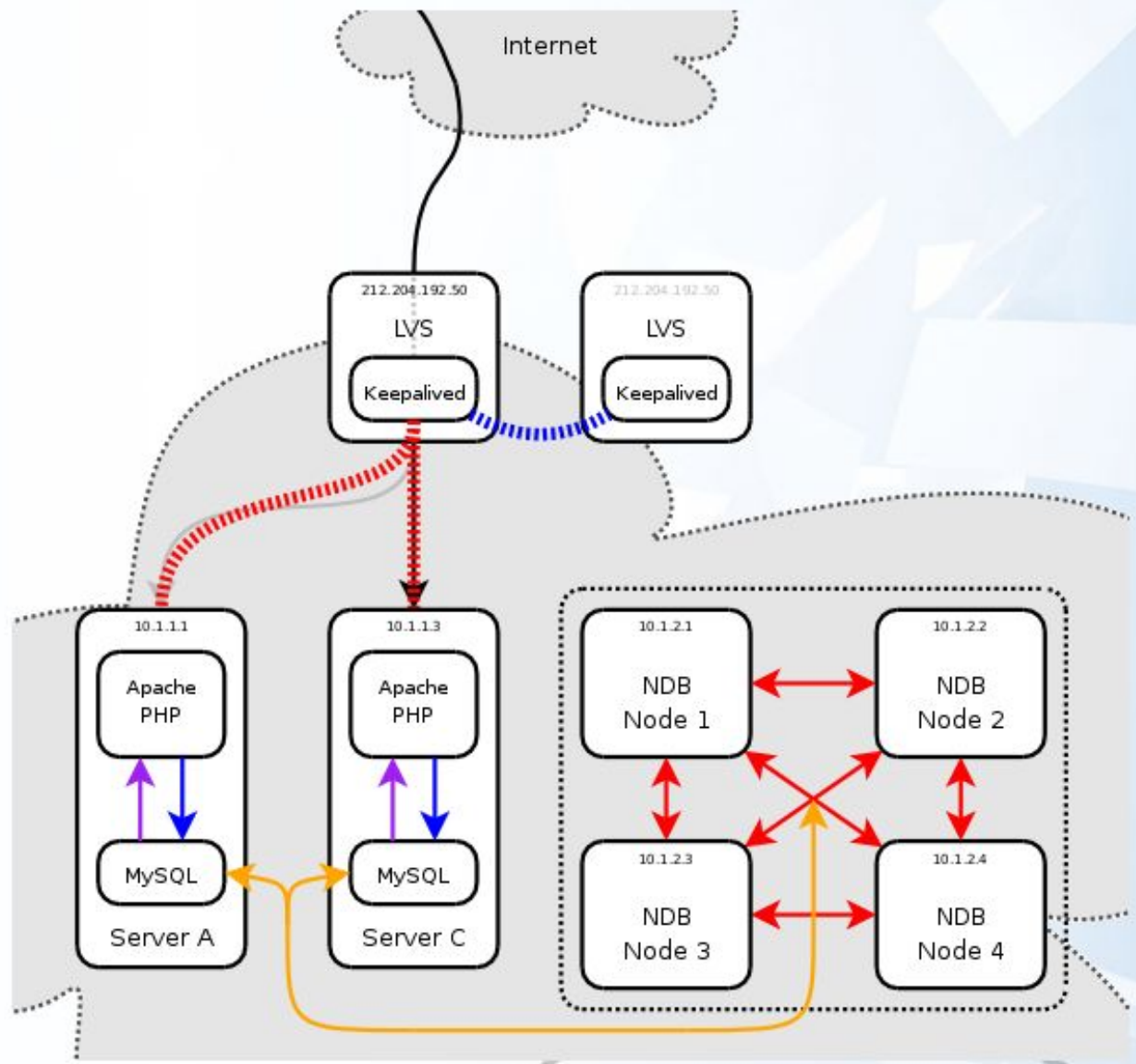


- Master with multiple slaves
- All writes go to the master
- All reads go to one of the slaves

Problems:

- No real fail-safe
- Transactions are not atomic over all nodes

Share your information



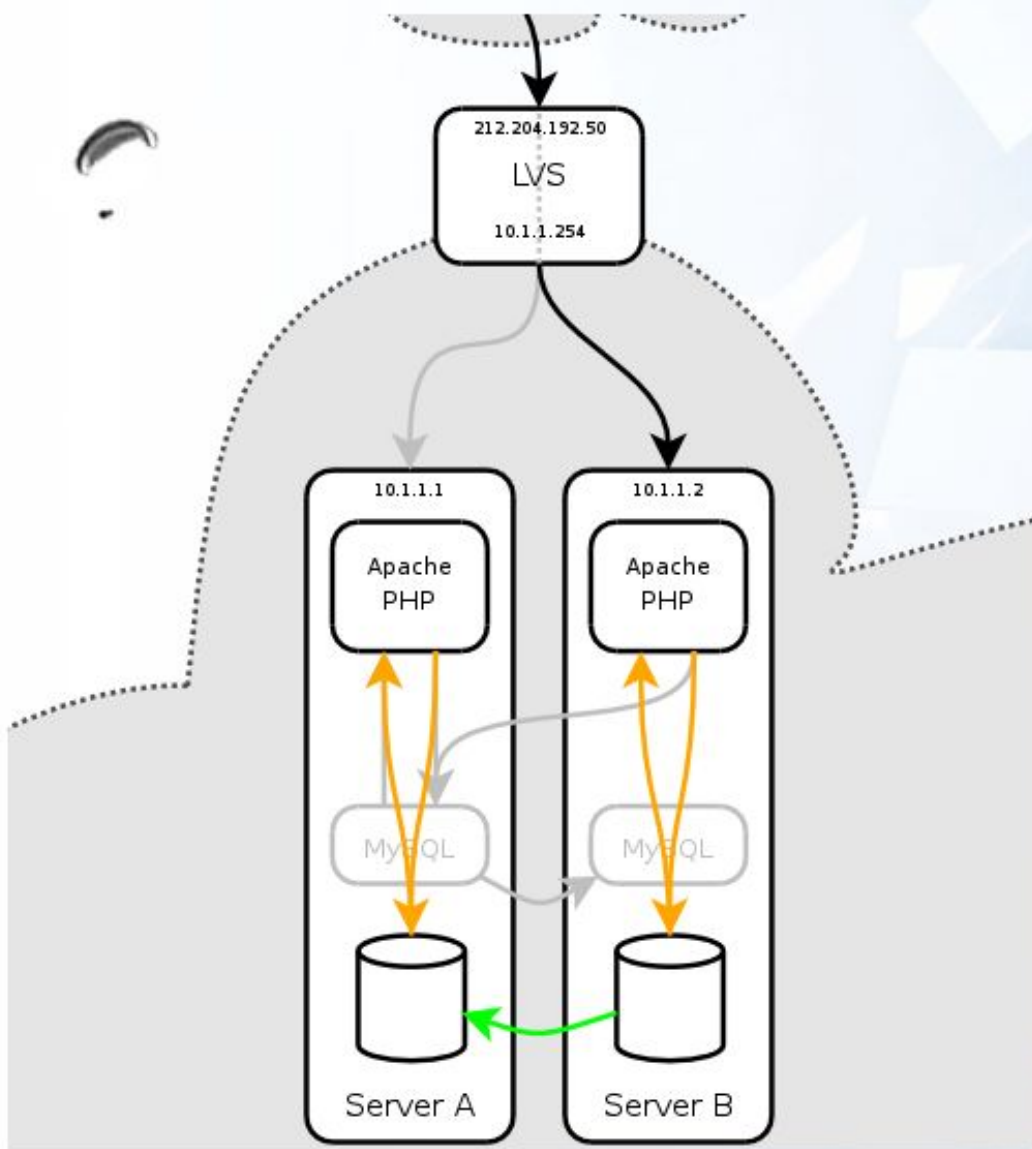
- NDB Support
- Multiple replica's for fail-safe
- Transactions are atomic cluster wise

Different nodes:

- Management Node
- SQL Nodes
- Storage Nodes

Problems:

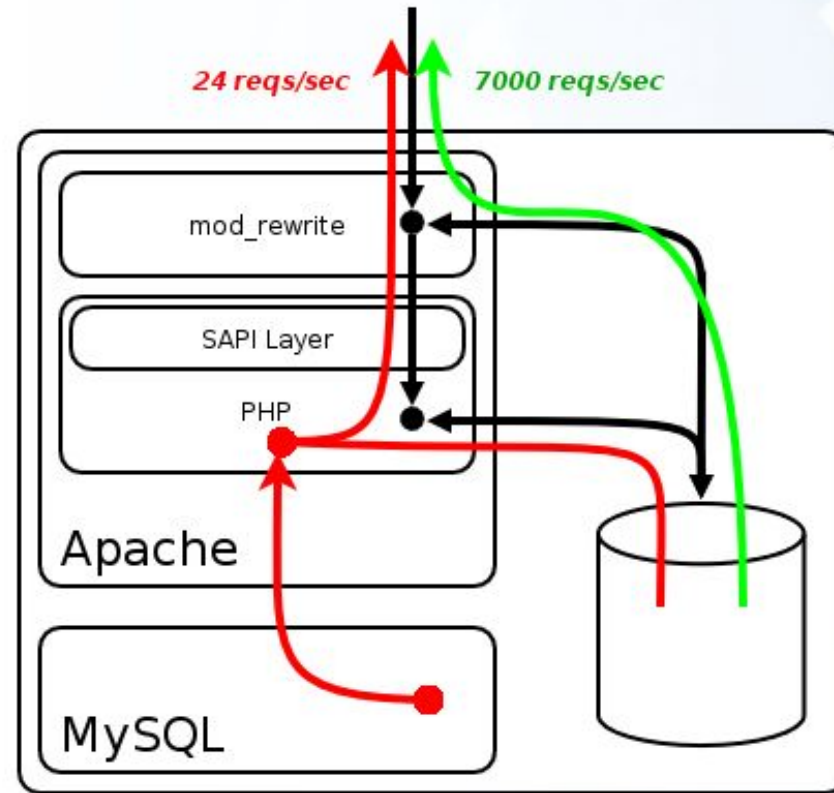
- Everything is in-memory
- It likes using a lot of memory



- All file actions in PHP are logged with a patch
- Uses a patch to eZ publish for additional file actions
- rsync is used to sync a file lists created from the file action log
- Problems with image variants and uploaded files

- Network File Systems: NFS, CODA
- Dedicated Hardware: NAS/SAN (Network Attached Storage)

Share your information



- Uses Apache `mod_rewrite` to check if files exist
- New cache files are written on publish
- Configuration can determine which nodes should always be updated

```
RewriteEngine On
```

```
RewriteCond /dat/ez.no/static/index.html -f  
RewriteRule ^/$ /static/index.html [L]  
RewriteCond /dat/ez.no/static/index.html -f  
RewriteRule ^/$ /static/index.html [L]
```

```
RewriteCond %{REQUEST_METHOD} !^POST$  
RewriteCond /dat/ez.no/static$1/index.html -f  
RewriteRule ^(.*)$ /static$1/index.html [L]
```

```
RewriteRule !\.(gif|css|jpg|png|jar|ico|js)$ /index.php
```

staticcache.ini:

```
[CacheSettings]
HostName=localhost
MaxCacheDepth=4

CachedURLArray[]
CachedURLArray[]=news*
CachedURLArray[]=weblog*

CachedSiteAccesses[]
CachedSiteAccesses[]=news_en
CachedSiteAccesses[]=news_fr

AlwaysUpdateArray[]
AlwaysUpdateArray[]=/
```

- *CachedURLArray*: Which URLs should be statically cached
- *CachedSiteAccesses*: For which site accesses should we cache
- *AlwaysUpdateArray*: Which nodes should always have it's static cache regenerated when publishing content

Static Cache

Updating

On publish:

- The published node's cache
- The parent node's cache of the published node
- Everything from the *AlwaysUpdateArray* array

With a script:

```
php bin/php/makestaticcache.php -s ezno -f
```

Nodes are cached as:

- The 'fancy' url name (news/my_article)
- content/view/full/*nodeid*

Resources

These Slides: <http://derickrethans.nl/talks.php>

PHP: <http://www.php.net>

Squid redirecting: <http://www.squid-cache.org/Doc/FAQ/FAQ-15.html>

Linux Virtual Server: <http://linuxvirtualserver.org>

Keepalived: <http://keepalived.org/>

MySQL Cluster: <http://dev.mysql.com/doc/mysql/en/ndbcluster.html>

Static Caching: http://ez.no/ez36_staticcache

Questions?: <mailto:dr@ez.no>