



Kako napraviti Google od zgrade s računalima?

Dobrica Pavlinušić

<http://www.rot13.org/~dpavlin/>

HULK, Knjižnica Filozofskog fakulteta u Zagrebu

DORS/CLUC, Zagreb, 7. svibnja 2010.

Zašto?

**Imate neka računala koja nisu
sasvim iskorištena,
a troše struju**

Uvijek ste htjeli imati svoj cluster!

Sadržaj

- Računala
 - Web kiosci za pretraživanje kataloga
- Network boot i monitoring
 - PXElator
- Distribute data
 - Sack
- Distribute computing
 - Gearman, Narada
- Dokumentacija
 - Sysadmin Cookbook, bak-git

Nadite računala

- Web kiosci za pretraživanje kataloga knjižnice
- Dell Optiplex755
 - 2Gb RAM
 - Core2 Duo E7200 @ 2.53GHz
 - 80Gb Hitachi disk
- 5 katova+prizemlje
 - 24 Gb RAM
 - 30.36 Ghz CPU
 - 960 Gb disk



Pretraživanje kataloga

- Koha – knjižnični software s web sučeljem
- Webconverger
 - Debian Live distribucija
 - Firefox, skrivanje URL-a
 - Boot s mreže bez korištenja diska
- Svaki ormar ima 100Mb/s switch (sigh!) zbog manjska mrežnih utičnica
- <http://webconverger.com/>

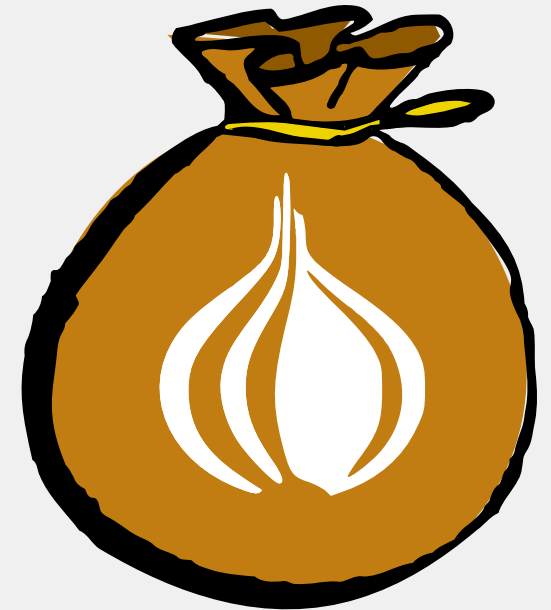


PXElator

- Replace system administration with perl script
- Protocol soup:
 - DHCP, TFTP, HTTP, DNS, syslog, WOL, AMT, DRAC5
- Store audit data in MongoDB
- Configuration in files/perl code
- Provision netboot, KVM, nfsroot
- Additional toos: nmap, wireshark
- <http://blog.rot13.org/pxelator/>

Sack

- Distribuirani **memorija** sa **ad-hoc upitima** u obliku **perl** funkcija
- Svi ostali pristupi su prespori ili prekomplikirani za korištenje
 - CouchDB
 - MongoDB
 - HyperTable
- Shard-nothing pristup
 - Cluster, multi-core
- <http://blog.rot13.org/sack/>

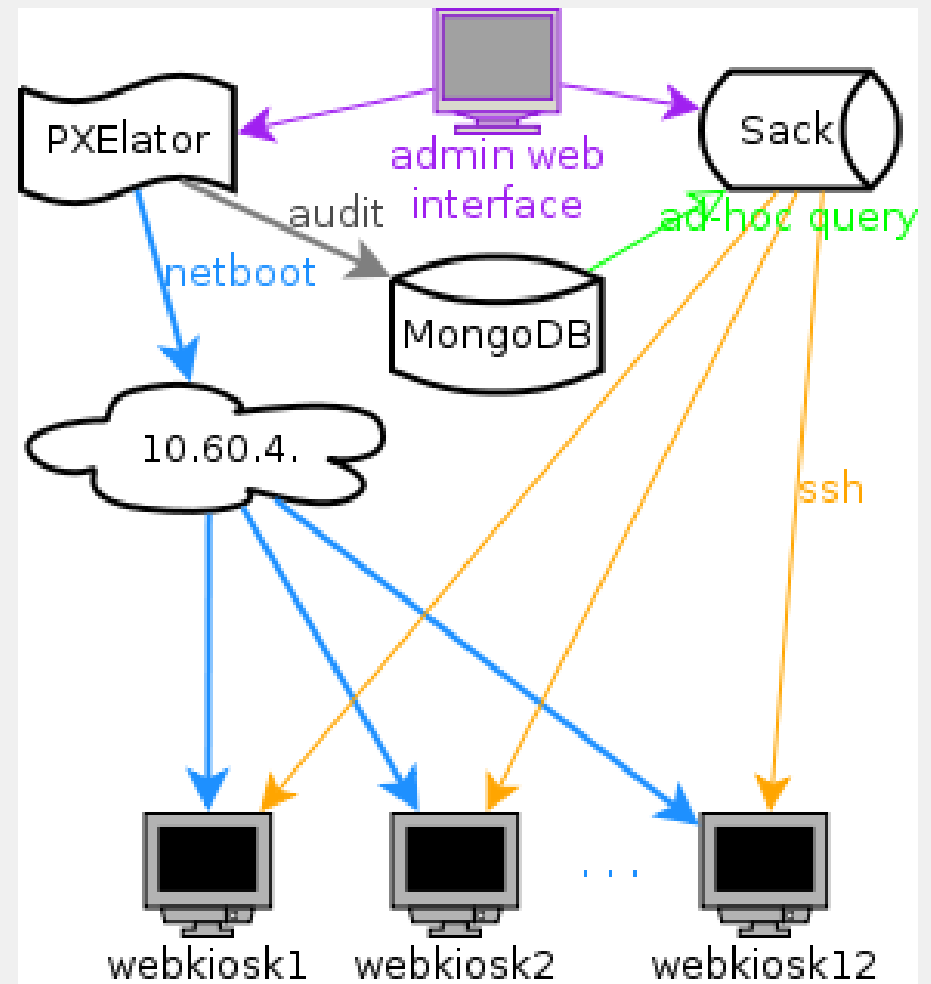


Kako napuniti vreću?

- pure-perl client
- configure (hosts+ssh config/keys)
- deployment over ssh/cpio on startup
 - ssh compression for network speedup
- source data for shards from CouchDB, MongoDB, WebPAC, text files...
- Edit perl ad-hoc query in vi
- Web interface
 - query execution and overview
 - with gnuplot graphs for time data

Cijeli sustav

- PXElator
 - boot nodes
 - audit data
- Sack
 - shard data
 - distribute
 - ad-hoc query
- Fun and profit!

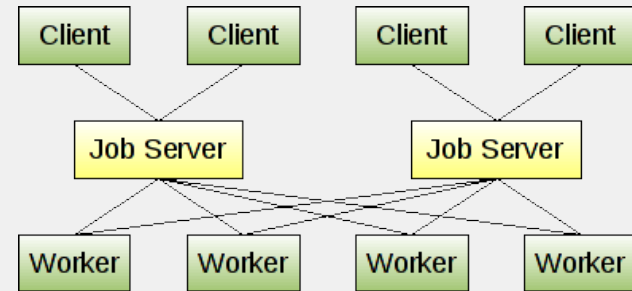
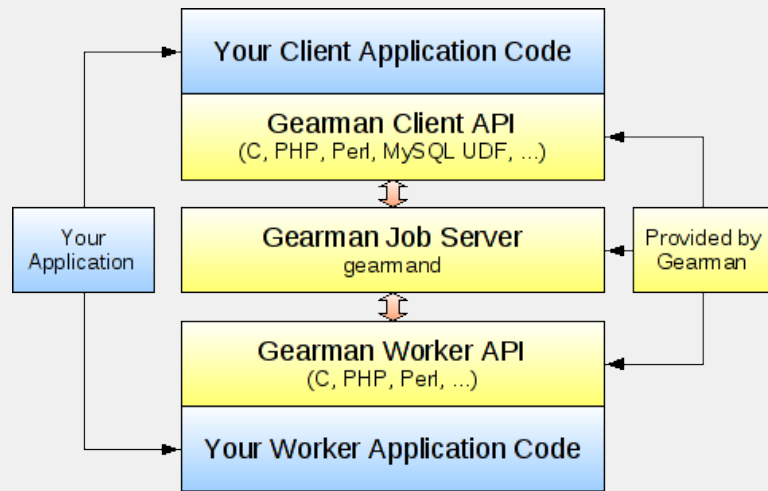


Gearman

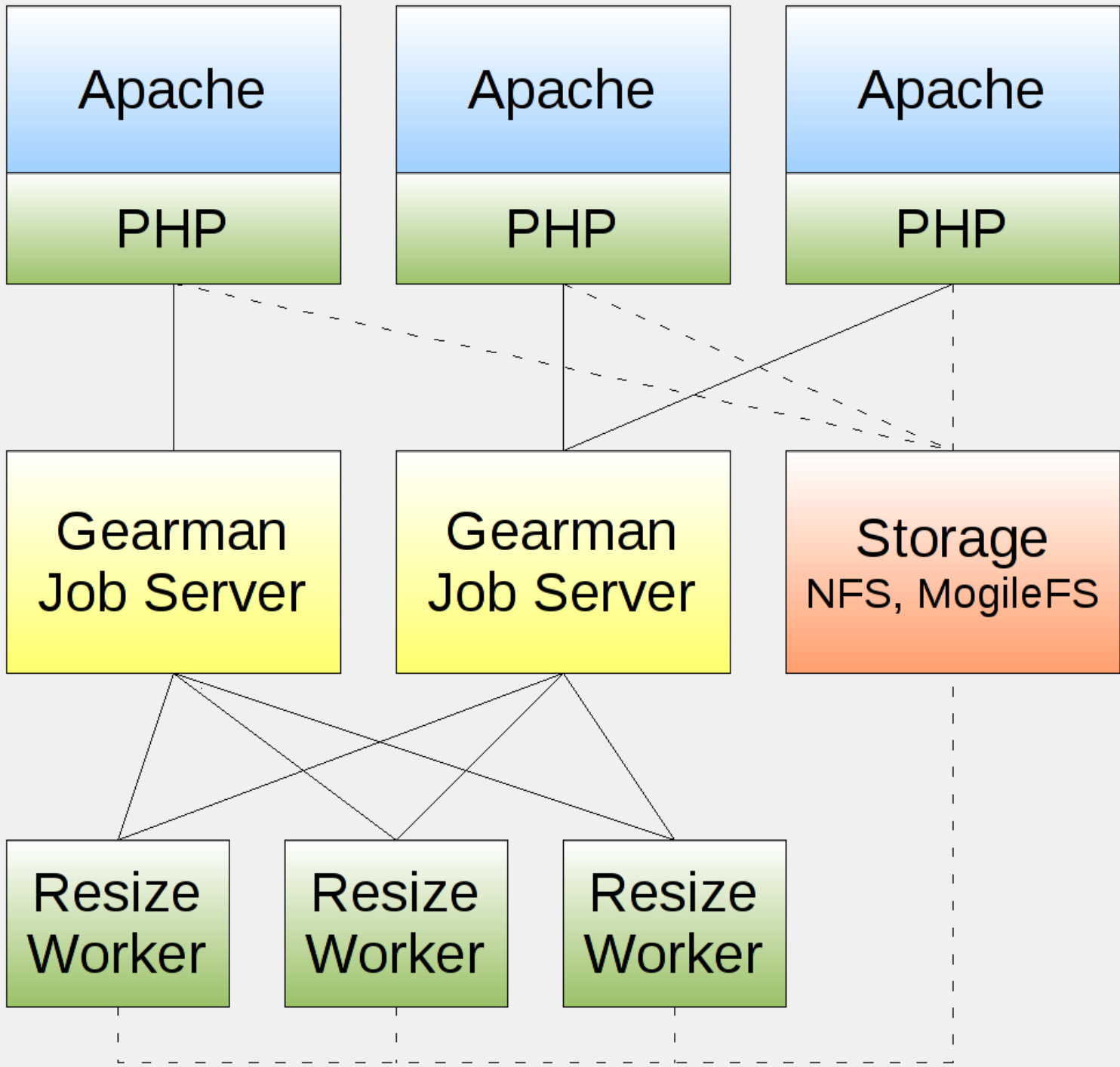
“The way I like to think of Gearman is as a massively distributed, massively fault tolerant fork mechanism.”

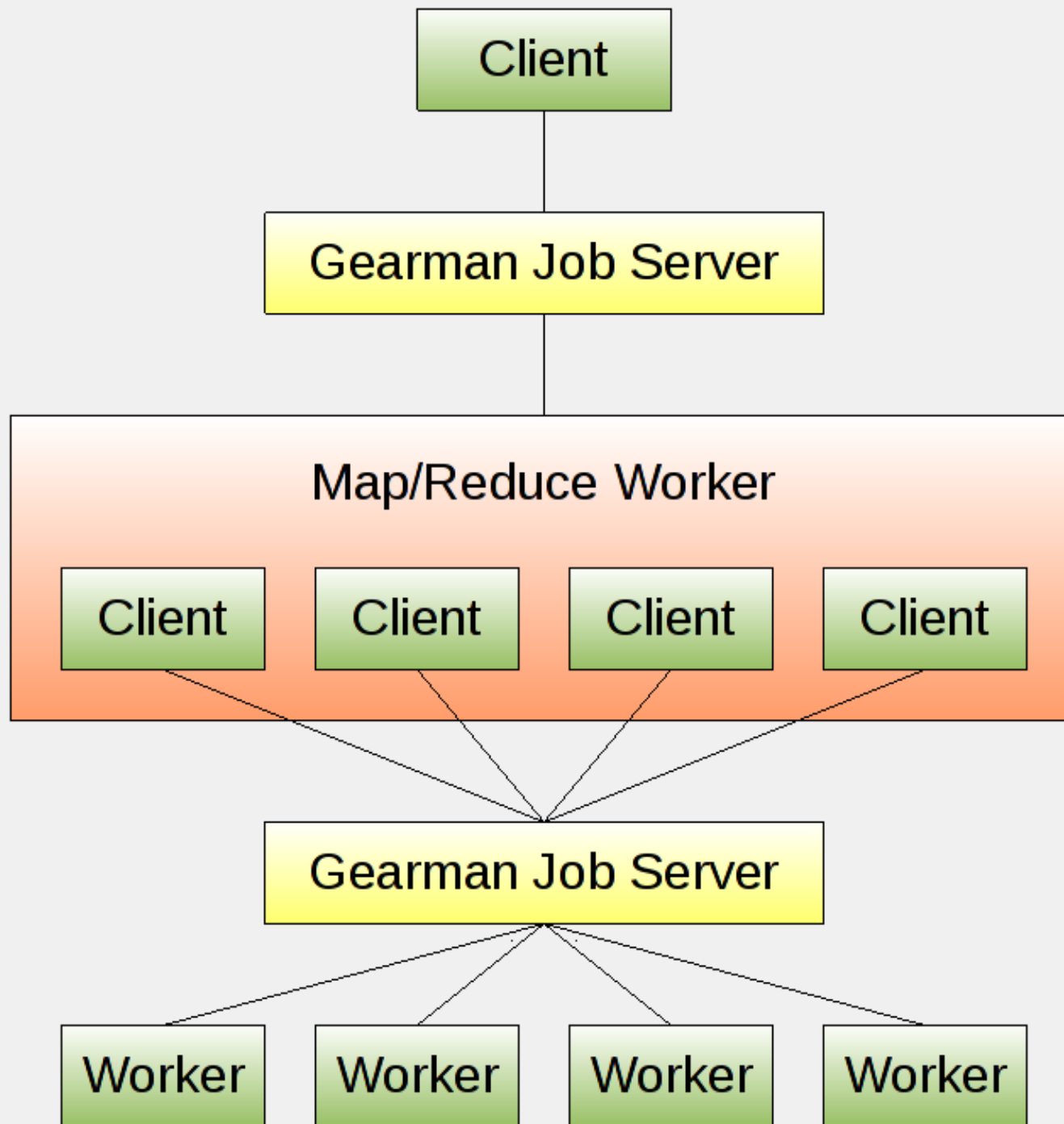
- Joe Stump, Digg

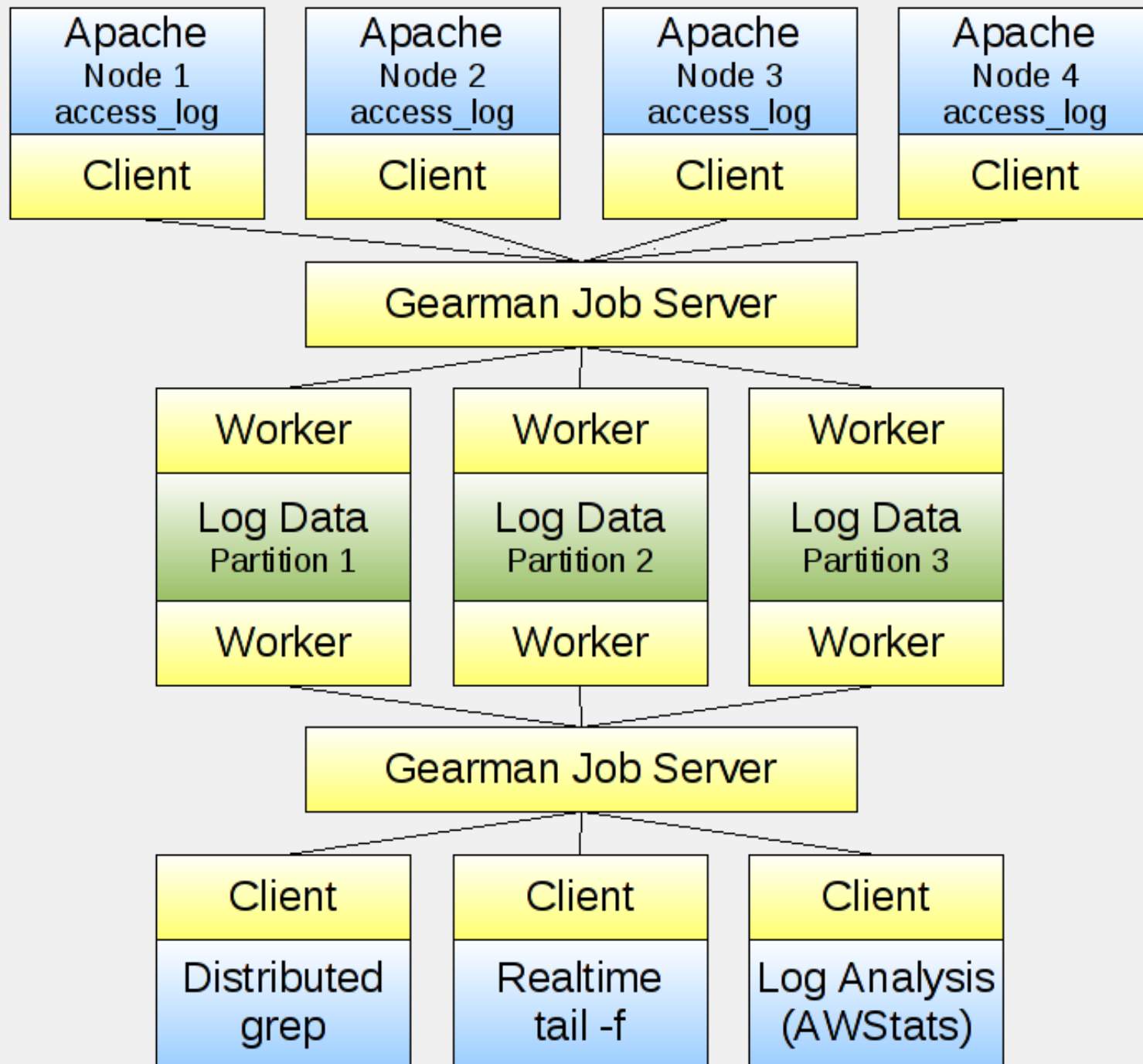
Gearman dijelovi



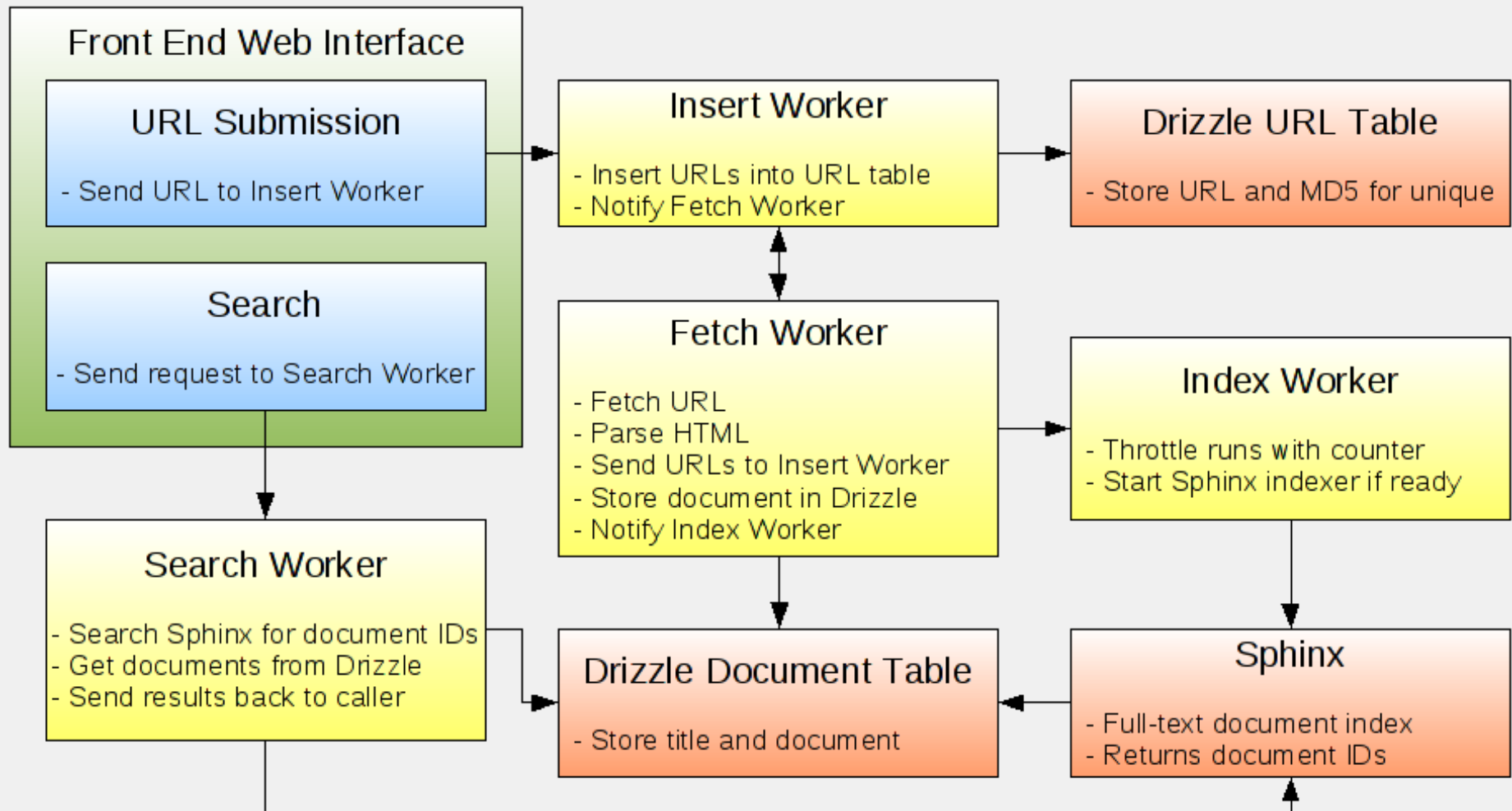
- fault-tolerant
- perl, php, C, Drizzle, MySQL, PostgreSQL, ...
- Command-line tool
- Pluggable protocols: HTTP
- <http://sysadmin-cookbook.rot13.org/#gearman>







Narada



Dokumentacija

Virtualizacija i clusteri su krasne stvari, sve dok ne počnete pisati dokumentaciju!

**Standard Operating Procedure
Configuration Management**

Sysadmin Cookbook

- Dokumentation for me, two years later
- Fast and easy deployment
- Convention over documentation
 - number.name
 - magic names
 - sh, append, replace
 - symlinks to keep original locations
 - commit messages for documentation
- <http://sysadmin-cookbook.rot13.org/>

bak-git centralni SCM

- System configuration management
- Centralni git repozitorij
 - Poznate git komande (s/git/bak/)
- Jednostavan deployment
 - ssh, rsync, netcat, shell (bak)
- Audit log svih promjena (security!)
- Svako računalo ima svoj direktorij
 - Uspoređivanje konfiguracija
- Web sučelje: gitweb
- <http://blog.rot13.org/bak-git/>

O čemu smo pričali?

- Nađite neiskorištenja računala
- Pojednostavite upravljanje
 - PXElator, netboot
- Podjelite podatke po njima
 - Sack, ad-hoc query
- Paralelizirajte obrade podataka
 - Gearman
- Dokumentirajte učinjeno
 - Ne postoji bez dokumentacije

Pitanja?
42