

Logging Everything (And Staying Sane)

DrupalCon Portland 2013
Speaker: Brian Altenhofel

IRC: VeggieMeat
@BrianAltenhofel
brian.altenhofel@vmdoh.com

A little about me...

- Own two Drupal businesses
 - VMdoh – Drupal development – <http://www.vmdoh.com>
 - Anrhizan – Drupal hosting – <http://www.anrhizan.com>
- Worked with Drupal since 2008
- Enjoy automation and data

```
127.0.0.1 - - [23/May/2013:18:07:57 +0000] "POST /editpost.php?do=updatepost&postid=2201085 HTTP/1.1" 200 11357 "http://www.veggie.com/showthread.php?175626-...-7" "Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.0; Trident/4.0; SLCC1; .NET CLR 2.0.50727; Media Center PC 5.0; InfoPath.2; .NET CLR 3.5.30729; .NET CLR 3.0.30618; .NET4.0C)"
root@web1:~# tail /var/log/nginx/access.log
97.65.100.100 - - [23/May/2013:18:08:32 +0000] "GET /image.php?u=15231&dateline=1353214186 HTTP/1.1" 200 2943 "https://www.veggie.com/showthread.php?170138-...-.../pagell&highlight=..." "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.4 (KHTML, like Gecko) Chrome/22.0.1229.94 Safari/537.4"
127.0.0.1 - - [23/May/2013:18:08:32 +0000] "GET /image.php?u=15231&dateline=1353214186 HTTP/1.0" 200 2943 "https://www.veggie.com/showthread.php?170138-...-.../pagell&highlight=..." "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.4 (KHTML, like Gecko) Chrome/22.0.1229.94 Safari/537.4"
127.0.0.1 - - [23/May/2013:18:08:32 +0000] "GET /image.php?u=28108&dateline=1365123446 HTTP/1.0" 200 2688 "https://www.veggie.com/showthread.php?170138-...-.../pagell&highlight=..." "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.4 (KHTML, like Gecko) Chrome/22.0.1229.94 Safari/537.4"
97.65.100.100 - - [23/May/2013:18:08:32 +0000] "GET /image.php?u=28108&dateline=1365123446 HTTP/1.1" 200 2688 "https://www.veggie.com/showthread.php?170138-...-.../pagell&highlight=..." "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.4 (KHTML, like Gecko) Chrome/22.0.1229.94 Safari/537.4"
97.65.100.100 - - [23/May/2013:18:08:32 +0000] "GET /image.php?u=9374&type=sigpic&dateline=1260758656 HTTP/1.1" 200 12149 "https://www.veggie.com/showthread.php?170138-...-.../pagell&highlight=..." "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.4 (KHTML, like Gecko) Chrome/22.0.1229.94 Safari/537.4"
127.0.0.1 - - [23/May/2013:18:08:32 +0000] "GET /image.php?u=9374&type=sigpic&dateline=1260758656 HTTP/1.0" 200 12149 "https://www.veggie.com/showthread.php?170138-...-.../pagell&highlight=..." "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.4 (KHTML, like Gecko) Chrome/22.0.1229.94 Safari/537.4"
97.65.100.100 - - [23/May/2013:18:08:32 +0000] "GET /image.php?u=24588&dateline=1337873971 HTTP/1.1" 200 9456 "https://www.veggie.com/showthread.php?170138-...-.../pagell&highlight=..." "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.4 (KHTML, like Gecko) Chrome/22.0.1229.94 Safari/537.4"
127.0.0.1 - - [23/May/2013:18:08:32 +0000] "GET /image.php?u=24588&dateline=1337873971 HTTP/1.0" 200 9456 "https://www.veggie.com/showthread.php?170138-...-.../pagell&highlight=..." "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.4 (KHTML, like Gecko) Chrome/22.0.1229.94 Safari/537.4"
127.0.0.1 - - [23/May/2013:18:08:33 +0000] "GET /register.php HTTP/1.1" 200 14480 "http://www.veggie.com/entry.php?..." "Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.0; Trident/4.0; Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1) ; .NET CLR 3.5.30729)"
127.0.0.1 - - [23/May/2013:18:08:33 +0000] "POST /forumrunner/request.php HTTP/1.1" 200 1032 "-" "ForumRunner 1.6.2 rv:20130201 (iPhone; iPhone OS 6.1.3; en_US)"
root@web1:~#
```

SYSTEM INFO

Host:	glenlivet
Uptime:	0h 0m 38s
RAM:	641MB/11.7GB
Swap usage:	0B /18.6GB
Disk usage:	26.0GB/223GB
CPU usage:	14%

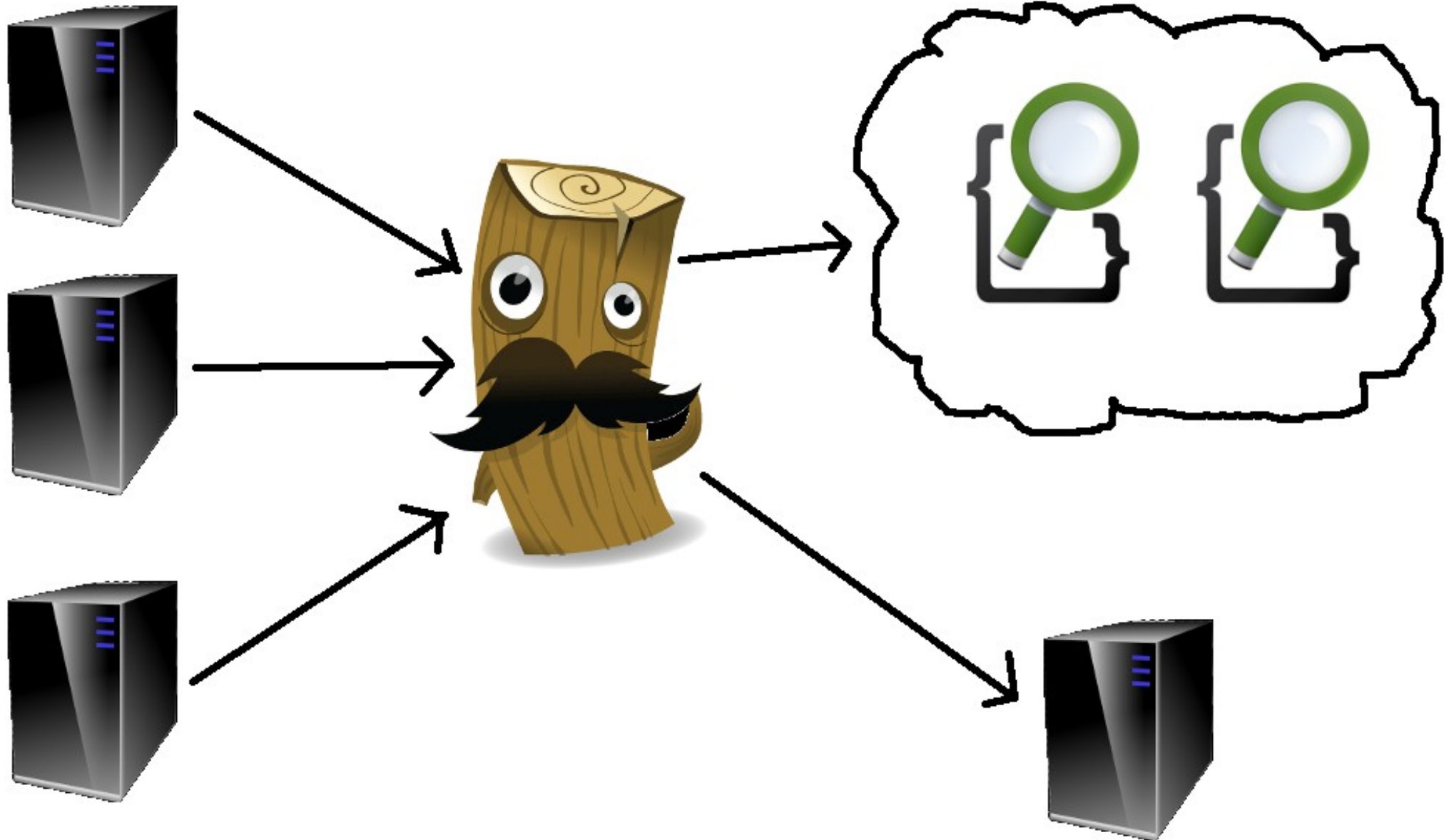
Alt+F2 Run Dialog
Alt+F3 Alt Manager
Command Terminal
Super+M File Manager
Super+H Home
Super+V Volume Control
Super+L Lock Screen

```
IP=127.0.0.1; cat /var/log/nginx/access.log | awk  
-F"[ ?]" -v ip=$IP ' $4 ~ /2013:03:/ && $1 ~ ip  
{freq[$7]++} END {for (x in freq) {print freq[x], x}}'  
| sort -rn | head -20
```

Problems with Conventional Logging

- Probably need to access multiple machines to find the root of the problem
- Must be a sysadmin or ops guy to navigate logs and diagnose issues
- Can take a long time to fix problems

Solution: Centralized Logging



What is Logstash?



- Ship logs anywhere
- Manipulate log data
- Free and open source
- Unix pipe for events

Logstash Plugins

Inputs

Currently 34 inputs including syslog, file, tcp, udp, lumberjack, elasticsearch, etc.

Filters

Currently 28 filters including grok, mutate, multiline, csv, anonymize, etc.

Outputs

Currently 47 outputs including Elasticsearch, Pagerduty, Nagios, Loggly, Redis, Graphite, etc.


```
input {
  lumberjack {
    type => "nginx_access"
    port => 5001
    ssl_certificate => "/etc/ssl/logstash.pub"
    ssl_key => "/etc/ssl/logstash.key"
  }
}

filter {
  # Lumberjack sends custom fields. We're going to use those for multi-user
  # Kibana access control.
  mutate {
    add_tag => [ "%{customer}" ]
  }
  mutate {
    remove => [ "customer" ]
  }
}

grok {
  type => "nginx_access"
  pattern => "%{COMBINEDAPACHELOG}"
}

output {
  stdout { debug => true debug_format => "json" }

  elasticsearch {
    cluster => "my-logstash-cluster"
  }
}
```

SYSTEM INFO

Host:	veggie
Uptime:	0h 23m 55s
RAM:	678MiB/11.7GiB
Swap usage:	0B / 18.6GiB
Disk usage:	26.0GiB/223GiB
CPU usage:	1%

SHORTCUT KEYS

Alt+F2	Run Dialog
Alt+F3	Alt Menu
Super+Space	Main Menu
Super+Tab	Client Menu
Super+T	Terminal
Super+F	File Manager
Super+E	Editor
Super+M	Media Player
Super+W	Web Browser
Super+H	Task Manager
Super+L	Lock Screen
Super+V	Volume Control
Super+X	Logout
PrnSc	Screenshot

What is Elasticsearch?

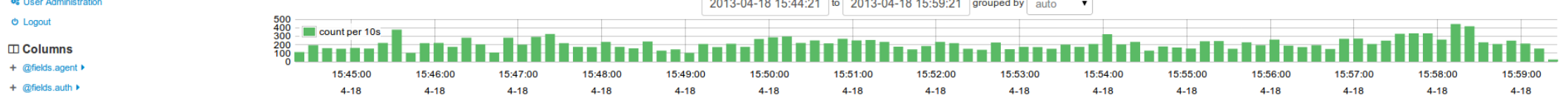


- Search engine built on Apache Lucene
- Distributed and easy high-availability
- Schema-free

What is Kibana?



Really awesome front-end
for Elasticsearch



Older 0 TO 50

Table with columns Time and @message. It lists network events including HTTP GET requests to okshooters.com and soonerstatepaw.com, and kernel messages regarding network interface statistics.

5m 15m 1h 6h 12h 24h 2d 5d

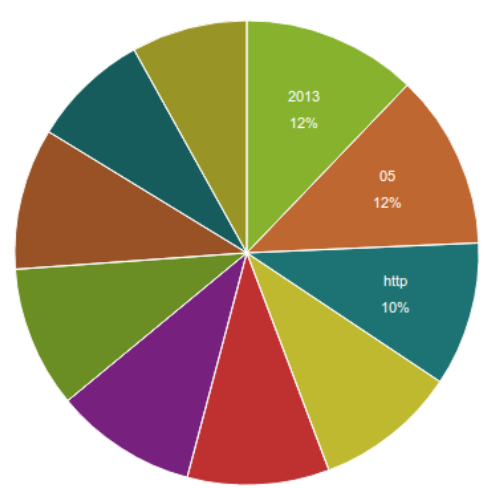
Relative | Absolute | Since | Auto-refresh

Dashboard Control

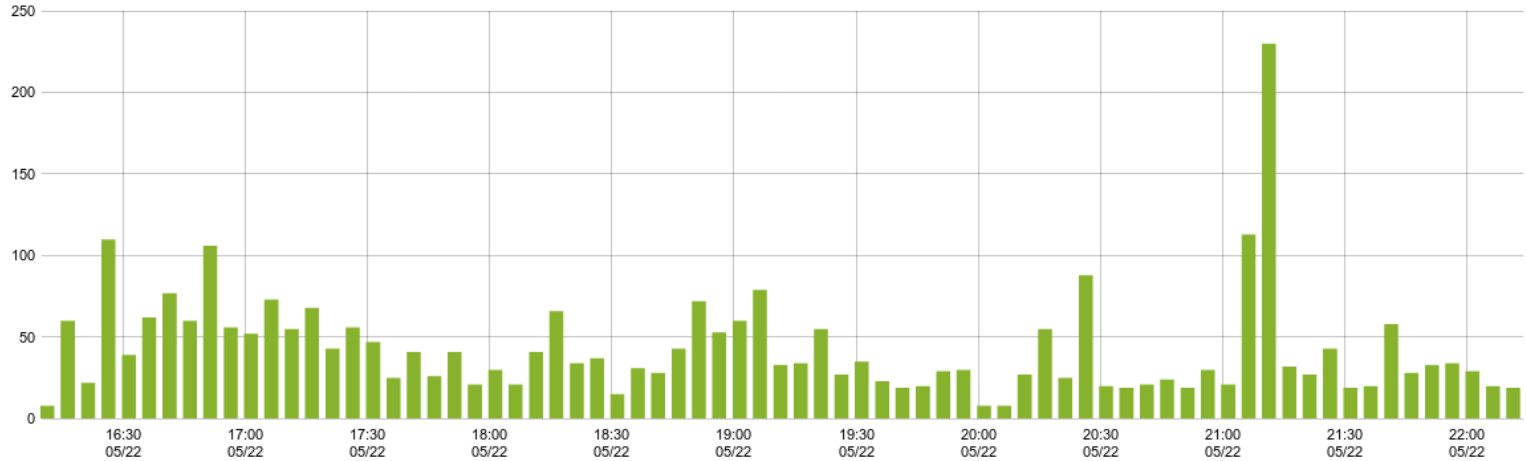


Search

Search query input field



Zoom In Zoom Out (3060) per 5m (3060 total)



- @message
- @source
- @source_host
- @source_path
- @tags
- @timestamp

@timestamp @message

```
2013-05-23T03:11:42.076Z 123.151.148.164 - - [22/May/2013:23:13:43 -0400] "GET /blog/tags/scraping HTTP/1.1" 200 10186 "http://www.semicomplete.com/" "Mozilla/5.0 (compatible; Sosospider/2.0; +http://help.soso.com/web spider.htm)"
```

What is Lumberjack?

- Ultralight shipper for Logstash
- Ships logs via SSL
- Why run 100MB+ Rsyslog or Logstash processes when you can run a few tiny Lumberjack processes with higher throughput?

```
top - 06:15:59 up 125 days, 4:33, 1 user, load average: 0.22, 0.21, 0.18
Tasks:  4 total,  0 running,  4 sleeping,  0 stopped,  0 zombie
Cpu(s): 19.2%us,  1.2%sy,  0.0%ni, 79.2%id,  0.0%wa,  0.0%hi,  0.2%si,  0.2%st
Mem:   1045340k total,  996492k used,  48848k free,  15824k buffers
Swap:  999992k total,  136776k used,  863216k free,  270460k cached
```

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
27080	root	39	19	66200	4136	1140	S	0.3	0.4	126:39.11	lumberjack
27370	root	39	19	66200	4264	1140	S	0.3	0.4	126:33.76	lumberjack
27096	root	39	19	66200	2984	1096	S	0.0	0.3	8:06.93	lumberjack
27143	root	39	19	82592	4344	1140	S	0.0	0.4	40:41.21	lumberjack

publisher	transport	consumer	publish rate	consume rate
logstash	N/A	N/A	33500	
yes netcat	tcp	logstash		32000
yes	stdin	logstash		31000
lumberjack	lumberjack	logstash	25000*	25000
logstash	zmq pushpull	logstash	22000	15000
logstash	zmq pushpull	logstash	15800	15800
logstash	lumberjack	logstash	8000	8000
logstash	redis	logstash	7500	5900
logstash	redis >batch=true	logstash	22000	6000
logstash	redis >batch=true < batch_count=50	logstash	24000	13000
logstash	redis >batch=true < batch_count=50 < threads=2	logstash	23500	23500
logstash	rabbitmq	logstash	6000	3000

Why not Drupal's DBLog and Statistics core modules?

- Statistics module is S-L-O-W.
- Writing log messages to the database can be really slow, especially if your site has a bug that's throwing warnings or errors.

But....

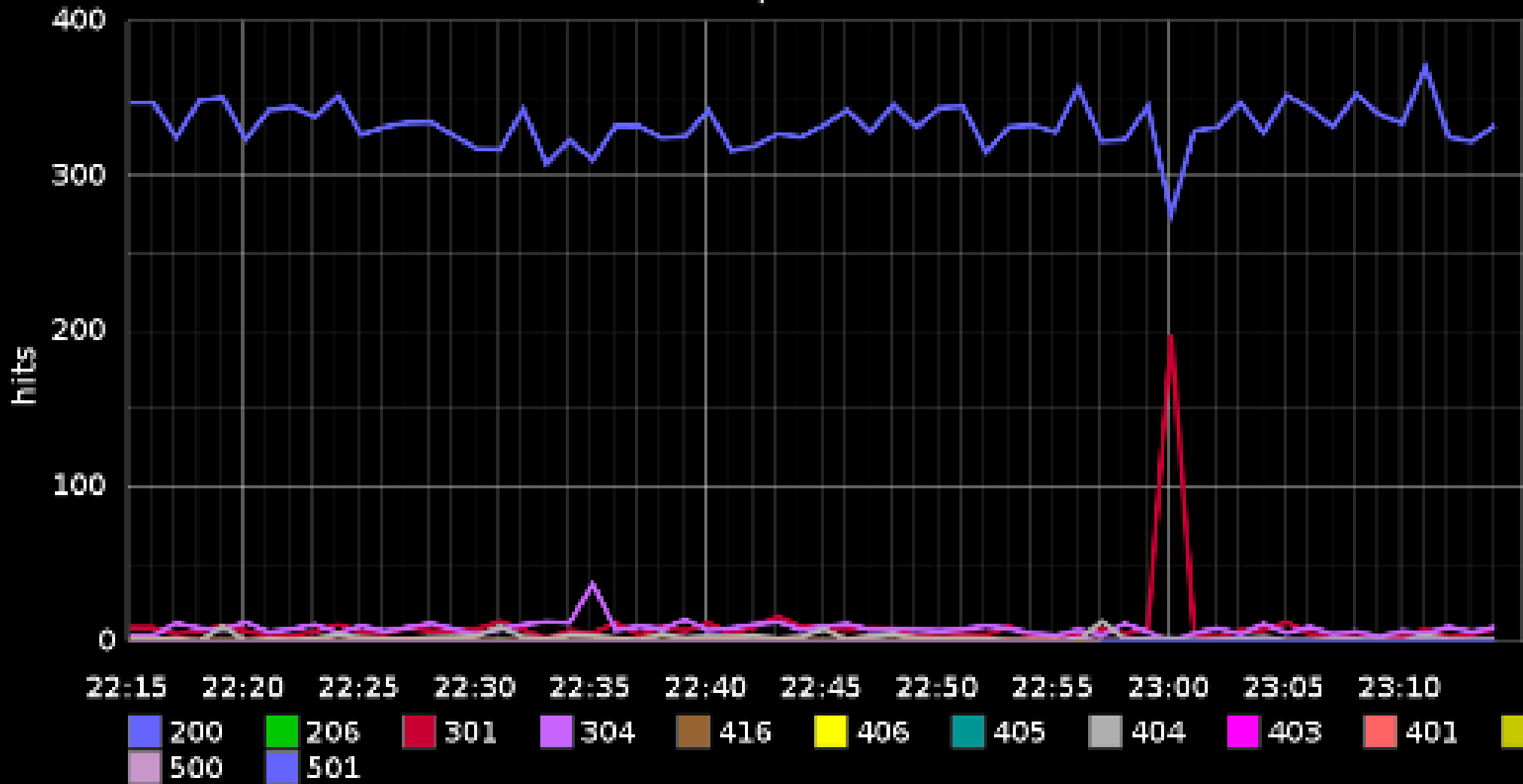
There is a Logstash input for DBLog.

http://logstash.net/docs/1.1.12/inputs/drupal_dblog

Advantages to Shipping Drupal's Logs to Logstash (and elsewhere)

- Easy to search through log messages
- Easily graph things like 404 errors or failed login attempts
- Less database bloat
- Fewer database writes, meaning happier users on high-traffic sites

Frontend Response Codes

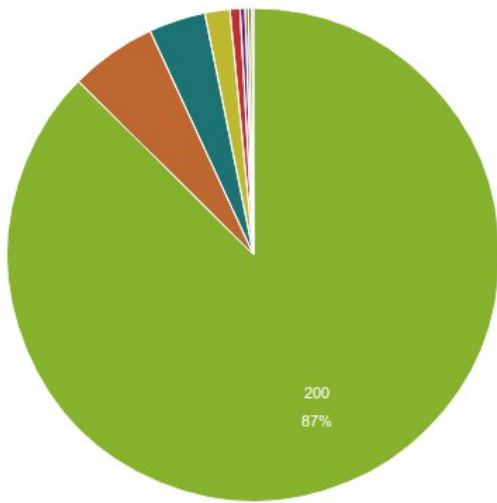


So really...
How does this help me?

Scenario #1 – Status Codes

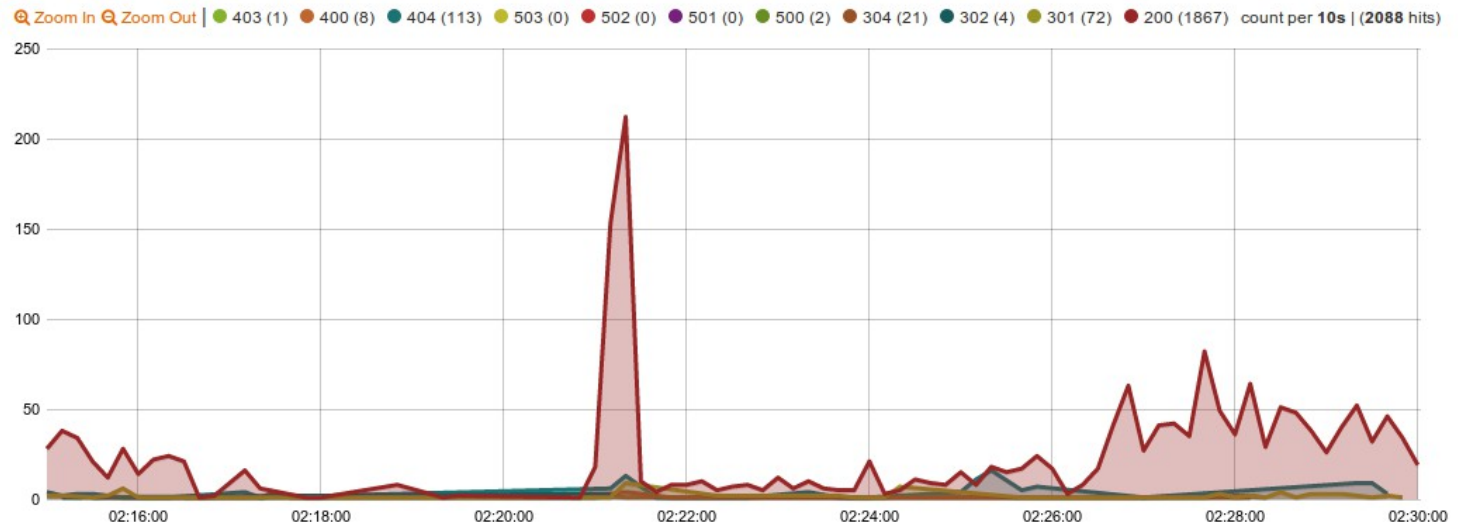
As a website administrator, I need to know if there is a trend of increasing errors.

HTTP Status Codes



Http Status Codes

histogram



Scenario #2 - MySQL

MySQL slow query logs suck



- @fields.duration
- @fields.lock_wait
- @fields.results
- @fields.scanned
- @message
- @source
- @source_host
- @source_path
- @tags
- @timestamp
- @type

0 to 100 of 404 available for paging



table

@timestamp	@message	@fields.scanned	@fields.results
2013-05-23T07:29:43.000Z	# User@Host: [redacted]@ [172.16.1.1] # Thread_id: 11237533 Schema: vb Last_errno: 0 Killed: 0 # Query_time: 0.000519 Lock_time: 0.000076 Rows_sent: 50 Rows_examined: 100 Rows_affected: 0 Rows_read: 50 # Bytes_sent: 2634 Tmp_tables: 0 Tmp_disk_tables: 0 Tmp_table_sizes: 0 # InnoDB_trx_id: 7F640EC SET timestamp=1369294183; SELECT moderator.*, user.username, IF(user.displaygroupid = 0, user.usergroupid, user.displaygroupid) AS displaygroupid, infractiongroupid FROM moderator AS moderator INNER JOIN user AS user USING(userid);	100	50
2013-05-23T07:29:42.000Z	# User@Host: [redacted]@ [172.16.1.6] # Thread_id: 11237514 Schema: vb Last_errno: 0 Killed: 0 # Query_time: 0.000311 Lock_time: 0.000073 Rows_sent: 0 Rows_examined: 36 Rows_affected: 0 Rows_read: 0 # Bytes_sent: 56 Tmp_tables: 0 Tmp_disk_tables: 0 Tmp_table_sizes: 0 # InnoDB_trx_id: 7F640B1 SET timestamp=1369294182; SELECT (primaryid + 2500000000) AS kill_id FROM digitalpoint_sphinx_delta AS digitalpoint_sphinx_delta WHERE contenttypeid = (SELECT contenttypeid FROM contenttype AS contenttype WHERE class = 'BlogEntry') AND	36	0

Scenario #3 – Twitter Trends

I may want to monitor what is being said about my brand.

Or I could just want to easily search through DrupalCon tweets.

Query

Search



Graph

Events

- @fields.client
- @fields.retweeted
- @fields.urls
- @fields.user
- @message
- @source
- @source_host
- @source_path
- @tags
- @timestamp
- @type

0 to 17 of 17 available for paging

table

@timestamp ▶	◀ @message ▶	◀ @fields.user ▶	◀ @fields.urls ▶
2013-05-23T07:18:08.334Z	RT @SebCorbin: Drupal community built a website to help victims of Oklahoma tornado during #DrupalCon http://t.co/8SJ71S1kbT	AnaelBoulier	http://t.co/8SJ71S1kbT
2013-05-23T07:22:58.127Z	Know what you need after a night of #drupalcon parties? Warm DoubleTree chocolate chip cookies. MMMmmmmzzzzzz.....	nategasser	
2013-05-23T07:23:23.209Z	Good @SPEAKINGinTECH podcast on Drupal, @ramkump1 worth a listen	MartinLeggatt	
2013-05-23T07:29:49.079Z	Trio is on 9th & Broadway #drupalcon folks	nonsie	
2013-05-23T07:26:02.020Z	RT @SebCorbin: Drupal community built a website to help victims of Oklahoma tornado during #DrupalCon http://t.co/8SJ71S1kbT	learningdrupal	http://t.co/8SJ71S1kbT
2013-05-23T07:15:24.381Z	Ya hemos publicado el último capítulo de nuestro blog. Integrar Twitter de una manera sencilla en nuestro sistema. http://t.co/BaNDjp6fOC	Abretutienda	http://t.co/BaNDjp6fOC

Scenario #4 – Unauthorized Access

As a site administrator, I want to see failed login attempts to my site.



- @fields.logsource
- @fields.message
- @fields.program
- @fields.timestamp
- @message
- @source
- @source_host
- @source_path
- @tags
- @timestamp
- @type

0 to 17 of 17 available for paging

table

@timestamp ▶	◀ @message
2013-05-23T07:05:32.536Z	May 23 07:05:31 prod1 drupal: http://www.vmdoh.com 1369292731 access denied 178.32.86.157 http://www.vmdoh.com/user/register http://www.vmdoh.com/blog/may-curtain-rise 0 user/register
2013-05-23T07:00:47.601Z	May 23 06:20:59 prod1 drupal: http://www.vmdoh.com 1369290059 access denied 216.59.18.39 http://www.vmdoh.com/user/register http://www.vmdoh.com/blog/centralizing-logs-lumberjack-logstash-and-elasticsearch 0 user/register
2013-05-23T07:01:19.846Z	May 23 06:46:09 prod1 drupal: http://www.vmdoh.com 1369291569 access denied 199.204.45.137 http://www.vmdoh.com/user/register 0 user/register
2013-05-23T06:59:22.504Z	May 23 05:14:06 prod1 drupal: http://www.vmdoh.com 1369286046 access denied 94.23.255.24 http://www.vmdoh.com/user/register http://www.vmdoh.com/?page=21 0 user/register
2013-05-23T07:05:32.533Z	May 23 07:05:31 prod1 drupal: http://www.vmdoh.com 1369292731 access denied 178.32.86.157 http://www.vmdoh.com/user/register http://www.vmdoh.com/blog/may-curtain-rise 0 user/register
2013-05-	May 23 06:20:59 prod1 drupal: http://www.vmdoh.com 1369290059 access

Scenario #5 - IRC

There is a ton of “documentation” in IRC that never makes it to Drupal.org.

Query

Search

@type:irc AND @message:"~views~"



Graph

Events

- @fields.channel
- @fields.nick
- @message
- @source
- @source_host
- @source_path
- @tags
- @timestamp
- @type

0 to 1 of 1 available for paging

table

@timestamp	@message	@fields.channel	@fields.nick
2013-05-23T12:18:53.190Z	http://drupal.org/node/1421656 => Extend the "modify entity values" action to add a delete mode for multiple-value properties and fields [#1421656] => Views Bulk Operations (VBO), Core, normal, needs review, 19 comments, 1 IRC mention	#drupal	Druplicon

0 to 1 of 1 available for paging

Scenario #6 – Smarter Notifications

As a system administrator, I might need certain messages sent directly to me or my staff.

```
output {
  #stdout { debug => true debug_format => "json" }

  elasticsearch {
    bind_host => "172.16.0.20"
    cluster => "XXXXXXXXXXXXXXXXXXXX"
  }

  pagerduty {
    type => "ouch"
    description => "Superbad event for #{@source_host}"
    event_type => "trigger"
    service_key => "XXXXXXXXXXXXXXXXXXXX"
  }

  statsd {
    type => "nginx_access"
    host => '172.16.0.76'
    increment => "nginx.response.#{response}"
  }
}
```

More resources...

- <http://www.logstash.net>
- <http://www.elasticsearch.org/download/>
- <https://github.com/jordansissel/lumberjack>
- <http://kibana.org/>

Feedback

<http://portland2013.drupal.org/node/1323>

Logging Everything (And Staying Sane)

DrupalCon Portland 2013
Speaker: Brian Altenhofel

IRC: VeggieMeat
@BrianAltenhofel
brian.altenhofel@vmdoh.com