

ELK

Elasticsearch, Logstash and Kibana at
Wikimedia

Not this kind of Elk



https://commons.wikimedia.org/wiki/File:Male_Elk_%288005280487%29.jpg

E is for  **elasticsearch.**

Document oriented full text search engine built on top of Apache Lucene.

Licensed under [Apache 2 Open Source License](#) by [Elasticsearch BV](#)

L is for **logstash**

Pipeline processing system that connects "inputs" to "outputs" with optional "filters" in between.

```
$ tail -f foo.log | grep bar |  
  awk '{print $7 $9}' >> foo-pretty.log
```

Licensed under [Apache 2 Open Source License](#) by [Elasticsearch BV](#)

K is for  **kibana**

Browser based analytics and search dashboard
for Elasticsearch.

Licensed under [Apache 2 Open Source License](#) by [Elasticsearch BV](#)

How Wikimedia uses ELK

- Log events sent to Logstash by various applications.
- Logstash processes events to clean them up and normalize things where we can.
- Logstash stores events in Elasticsearch.
- Logstash sends some metrics to statsd for trend reporting.
- Use Kibana to search log events from Elasticsearch.
- Icinga alerts to watch for bad trends.

Logstash inputs

- Apache2 & HHVM via rsyslog
- MediaWiki via Monolog and Syslog handler
- scap via log2udp forwarding
- Various nodejs services via Bunyan and gelf-stream
- Cassandra via Logback
- (Beta cluster) Puppet via custom Report plugin
- (Beta cluster) Syslog
- (Labs) IRC
- Many others possible...

Logstash filters

- Strip ansi color escape sequences
- Join sequential lines into a single event for some types
- Normalizations to populate common attributes:
 - **type**: Origin (eg mediawiki, apache2, hhvm)
 - **channel**: Event type (eg memcached, proxy_fcgi)
 - **level**: Severity (eg INFO, WARNING, ERROR)
- Normalize event level naming to PSR-3 standard
- Discard junk messages
- Many others possible...

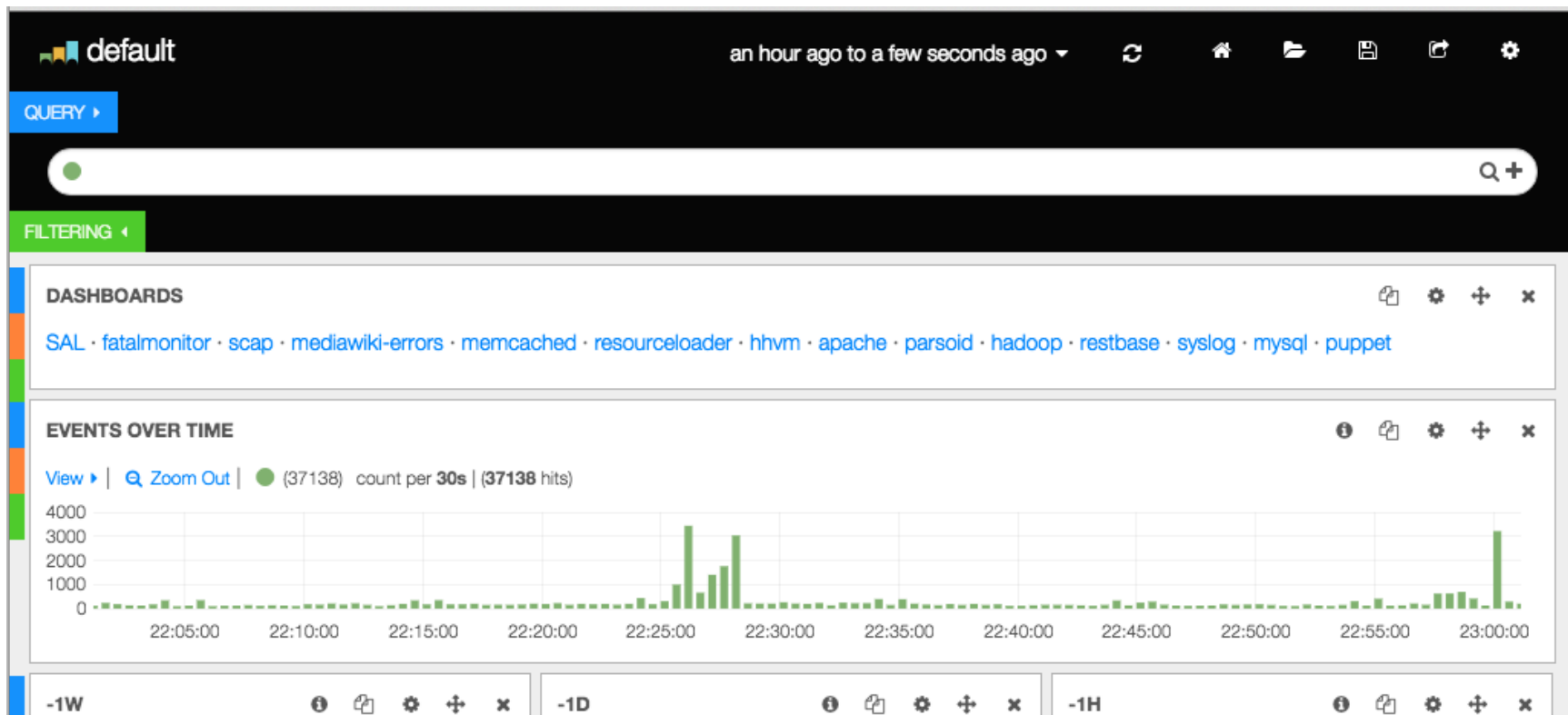
Logstash outputs

- Local Elasticsearch cluster
- Remote Elasticsearch cluster
- Statsd metrics for Graphite storage
- Many others possible...

Using Kibana

- WMF production: <https://logstash.wikimedia.org/>
 - Requires a [signed NDA](#) because of access to potentially sensitive data.
- WMF Beta cluster: <https://logstash-beta.wmflabs.org/>

Using Kibana



Using Kibana

The screenshot displays the Kibana interface with a dark theme. At the top left, the text "default" is visible. In the top right corner, a dropdown menu is open, showing the selected time range "an hour ago to a few seconds ago", which is circled in yellow. Below the search bar, a white box contains the text "Select time range to search". The interface also shows a "DASHBOARDS" section with a list of dashboards and an "EVENTS OVER TIME" section with a bar chart showing event counts over time.

default

an hour ago to a few seconds ago

QUERY

Select time range to search

FILTERING

DASHBOARDS

SAL · fatalmonitor · scap · mediawiki-errors · memcached · resourceloader · hvm · apache · parsoid · hadoop · restbase · syslog · mysql · puppet

EVENTS OVER TIME

View | [Q Zoom Out](#) | ● (37138) count per 30s | (37138 hits)

4000
3000
2000
1000
0

22:05:00 22:10:00 22:15:00 22:20:00 22:25:00 22:30:00 22:35:00 22:40:00 22:45:00 22:50:00 22:55:00 23:00:00

-1W -1D -1H

Using Kibana

The screenshot displays the Kibana dashboard interface. At the top left, the text "default" is visible. In the top right, there is a time range selector set to "an hour ago to a few seconds ago" and a refresh button (a circular arrow icon) which is circled in yellow. Below the search bar, a white callout box contains the text "Refresh the current dashboard". The main content area is divided into two sections: "DASHBOARDS" and "EVENTS OVER TIME". The "DASHBOARDS" section lists various dashboards such as "SAL", "fatalmonitor", "scap", "mediawiki-errors", "memcached", "resourceloader", "hvm", "apache", "parsoid", "hadoop", "restbase", "syslog", "mysql", and "puppet". The "EVENTS OVER TIME" section shows a bar chart with the title "View | Q Zoom Out | (37138) count per 30s | (37138 hits)". The chart's x-axis represents time from 22:05:00 to 23:00:00, and the y-axis represents the count of events, ranging from 0 to 4000. The chart shows a significant spike in events around 22:25:00 and another spike at 23:00:00. At the bottom of the dashboard, there are three panels labeled "-1W", "-1D", and "-1H", each with its own set of control icons.

Using Kibana

The screenshot displays the Kibana user interface. At the top left, the word "default" is shown next to a small icon. To the right, there is a navigation bar with several icons: a refresh icon, a home icon (circled in yellow), a folder icon, a document icon, a share icon, and a settings icon. Below the navigation bar is a search bar with the text "Load the default dashboard" and a search icon. The main content area is divided into two sections: "DASHBOARDS" and "EVENTS OVER TIME". The "DASHBOARDS" section lists various dashboards: SAL, fatalmonitor, scap, mediawiki-errors, memcached, resourceloader, hvm, apache, parsoid, hadoop, restbase, syslog, mysql, and puppet. The "EVENTS OVER TIME" section shows a bar chart with a y-axis ranging from 0 to 4000 and an x-axis showing time from 22:05:00 to 23:00:00. The chart displays a series of green bars representing event counts over time, with a notable peak around 22:25:00. The bottom of the interface shows three panels for different time ranges: "-1W", "-1D", and "-1H", each with its own set of control icons.

Using Kibana

The screenshot shows the Kibana dashboard interface. At the top, the word "default" is on the left, and a navigation bar on the right includes a time range selector set to "an hour ago to a few seconds ago", a refresh icon, a home icon, a folder icon (circled in yellow), a save icon, a share icon, and a settings icon. Below the navigation bar is a "QUERY" section with a search input field. A callout box points to the folder icon with the text "Load a saved dashboard". Below the query section is a "FILTERING" section. The main content area is divided into two panels. The top panel, titled "DASHBOARDS", shows a list of saved dashboards: SAL, fatalmonitor, scap, mediawiki-errors, memcached, resourceloader, hvm, apache, parsoid, and hadoop. A callout box points to this list with the text "Try typing a few letters in the text field to find saved dashboards." The bottom panel, titled "EVENTS OVER TIME", shows a bar chart of event counts over time. The chart has a y-axis from 0 to 4000 and an x-axis from 22:05:00 to 23:00:00. The data shows a significant spike in event counts between 22:25:00 and 22:30:00, and another spike at 23:00:00. The chart includes a legend for "count per 30s" with 37138 hits. At the bottom of the dashboard, there are three panels for different time ranges: "-1W", "-1D", and "-1H", each with its own set of control icons.

default

an hour ago to a few seconds ago

QUERY

Load a saved dashboard

FILTERING

DASHBOARDS

SAL · fatalmonitor · scap · mediawiki-errors · memcached · resourceloader · hvm · apache · parsoid · hadoop

EVENTS OVER TIME

View | Zoom Out | (37138) count per 30s | (37138 hits)

4000
3000
2000
1000
0

22:05:00 22:10:00 22:15:00 22:20:00 22:25:00 22:30:00 22:35:00 22:40:00 22:45:00 22:50:00 22:55:00 23:00:00

-1W -1D -1H

Try typing a few letters in the text field to find saved dashboards.

Using Kibana

The screenshot shows the Kibana dashboard interface. At the top left, the word "default" is displayed next to a small icon. To the right, there is a time range selector set to "an hour ago to a few seconds ago", a refresh button, a home button, a folder icon, a save icon (circled in yellow), a share icon, and a settings icon. Below the top bar, there is a "QUERY" button and a search input field. A "FILTERING" button is also visible. The main content area is divided into two sections: "DASHBOARDS" and "EVENTS OVER TIME". The "DASHBOARDS" section lists several dashboards: SAL, fatalmonitor, scap, mediawiki-errors, memcached, resourceloader, hvm, apache, parsoid, and hadoop. The "EVENTS OVER TIME" section shows a bar chart with a y-axis ranging from 0 to 4000 and an x-axis showing time from 22:05:00 to 23:00:00. The chart displays a series of green bars representing event counts over time. At the bottom of the dashboard, there are three panels, each with a time range selector (-1W, -1D, -1H) and a set of control icons (info, copy, settings, zoom, close).

Save the current dashboard

Make sure to change the name if you are making a new dashboard.

Dashboards are NOT versioned so if you save over an existing dashboard the old version is lost.

Using Kibana

The screenshot shows the Kibana dashboard interface. At the top, the word "default" is on the left, and a time range "an hour ago to a few seconds ago" is in the center. On the right, there are icons for refresh, home, folders, save, share, and settings. The share icon is circled in yellow. Below the top bar, there is a "QUERY" section with a search bar and a "FILTERING" section. The main content area is divided into two panels: "DASHBOARDS" and "EVENTS OVER TIME". The "DASHBOARDS" panel lists various dashboards like "SAL", "fatalmonitor", "scap", etc. The "EVENTS OVER TIME" panel shows a bar chart of event counts over time, with a legend indicating "(37138) count per 30s | (37138 hits)". At the bottom, there are three time range selectors: "-1W", "-1D", and "-1H".

default an hour ago to a few seconds ago

QUERY

Share the current dashboard

FILTERING

DASHBOARDS

SAL · fatalmonitor · scap · mediawiki-errors · memcached · resourceloader · hvm · apache · parsoid · hadoop

EVENTS OVER TIME

View | [Q Zoom Out](#) | ● (37138) count per 30s | (37138 hits)

4000
3000
2000
1000
0

22:05:00 22:10:00 22:15:00 22:20:00 22:25:00 22:30:00 22:35:00 22:40:00 22:45:00 22:50:00 22:55:00 23:00:00

-1W -1D -1H

Note: A shared dashboard is saved for 30 days.

Using Kibana

The screenshot shows the Kibana dashboard interface. At the top left, the text "default" is displayed next to a small icon. To the right, there is a time range selector set to "an hour ago to a few seconds ago" and several utility icons: refresh, home, search, save, share, and a settings gear icon which is circled in yellow. Below the top bar, there are three main sections: a "QUERY" section with a search bar, a "FILTERING" section, and a "DASHBOARDS" section. The "DASHBOARDS" section lists various dashboards: SAL, fatalmonitor, scap, mediawiki-errors, memcached, resourceloader, hvm, apache, parsoid, hadoop, restbase, syslog, mysql, and puppet. Below this is the "EVENTS OVER TIME" section, which features a bar chart showing event counts over time. The chart has a y-axis from 0 to 4000 and an x-axis from 22:05:00 to 23:00:00. A callout box points to the settings gear icon in the top right, stating "Configure dashboard settings". Another callout box points to the "DASHBOARDS" section, stating "Mostly used to add a new row." At the bottom, there are three panels for different time ranges: "-1W", "-1D", and "-1H", each with its own set of utility icons.

default

an hour ago to a few seconds ago

QUERY

Configure dashboard settings

FILTERING

DASHBOARDS

SAL · fatalmonitor · scap · mediawiki-errors · memcached · resourceloader · hvm · apache · parsoid · hadoop · restbase · syslog · mysql · puppet

EVENTS OVER TIME

View | [Q Zoom Out](#) | ● (37138) count per 30s | (37138 hits)

4000
3000
2000
1000
0

22:05:00 22:10:00 22:15:00 22:20:00 22:25:00 22:30:00 22:35:00 22:40:00 22:45:00 22:50:00 22:55:00 23:00:00

-1W -1D -1H

Mostly used to add a new row.

Using Kibana

The screenshot displays the Kibana interface with a search query and a corresponding chart. The search bar at the top contains the query `type:mediawiki AND channel:fatal`, which is highlighted with a yellow oval. Below the search bar, a text box contains the instruction: "Enter an Elasticsearch ["query string query"](#)".

The interface shows a dashboard with the following components:

- QUERY**: A blue button with a right-pointing arrow.
- SEARCH BAR**: Contains the query `type:mediawiki AND channel:fatal` and a search icon.
- FILTERING**: A green button with a left-pointing arrow.
- DASHBOARDS**: A section with a title and a list of dashboard names: `SAL · fatalmonitor · scap · mediawiki-errors · memcached · resourceloader · hvm · apache · parsoid · hadoop · restbase · syslog · mysql · puppet`.
- EVENTS OVER TIME**: A section with a title and a chart. The chart shows the count of events over time, with a peak around 22:25:00. The chart is titled "View | [Zoom Out](#) | ● (37138) count per 30s | (37138 hits)".

The bottom of the interface shows three tabs: `-1W`, `-1D`, and `-1H`, each with its own set of control icons.

Using Kibana

The screenshot displays the Kibana dashboard interface. At the top, there is a navigation bar with the text "default" and a time range selector set to "an hour ago to a few seconds ago". Below this, a "QUERY" button is visible. A search bar is present, and the "FILTERING" section is highlighted with a yellow circle. A callout box points to this section with the text: "Click to expand filters section. Filters are AND'ed with the query string." Below the filtering section, the "DASHBOARDS" section is visible, showing a list of dashboards including "elasticsearch", "syslog", "mysql", and "puppet". The main content area features an "EVENTS OVER TIME" chart, which is a histogram showing event counts over a period from 22:05:00 to 23:00:00. The chart has a y-axis ranging from 0 to 4000 and a x-axis with 5-minute intervals. The data shows a significant spike in events between 22:25:00 and 22:30:00, reaching nearly 4000 hits. The chart is titled "EVENTS OVER TIME" and includes a legend indicating "(37138) count per 30s | (37138 hits)".

default an hour ago to a few seconds ago

QUERY

FILTERING

DASHBOARDS

Click to expand filters section. Filters are AND'ed with the query string.

elasticsearch · syslog · mysql · puppet

EVENTS OVER TIME

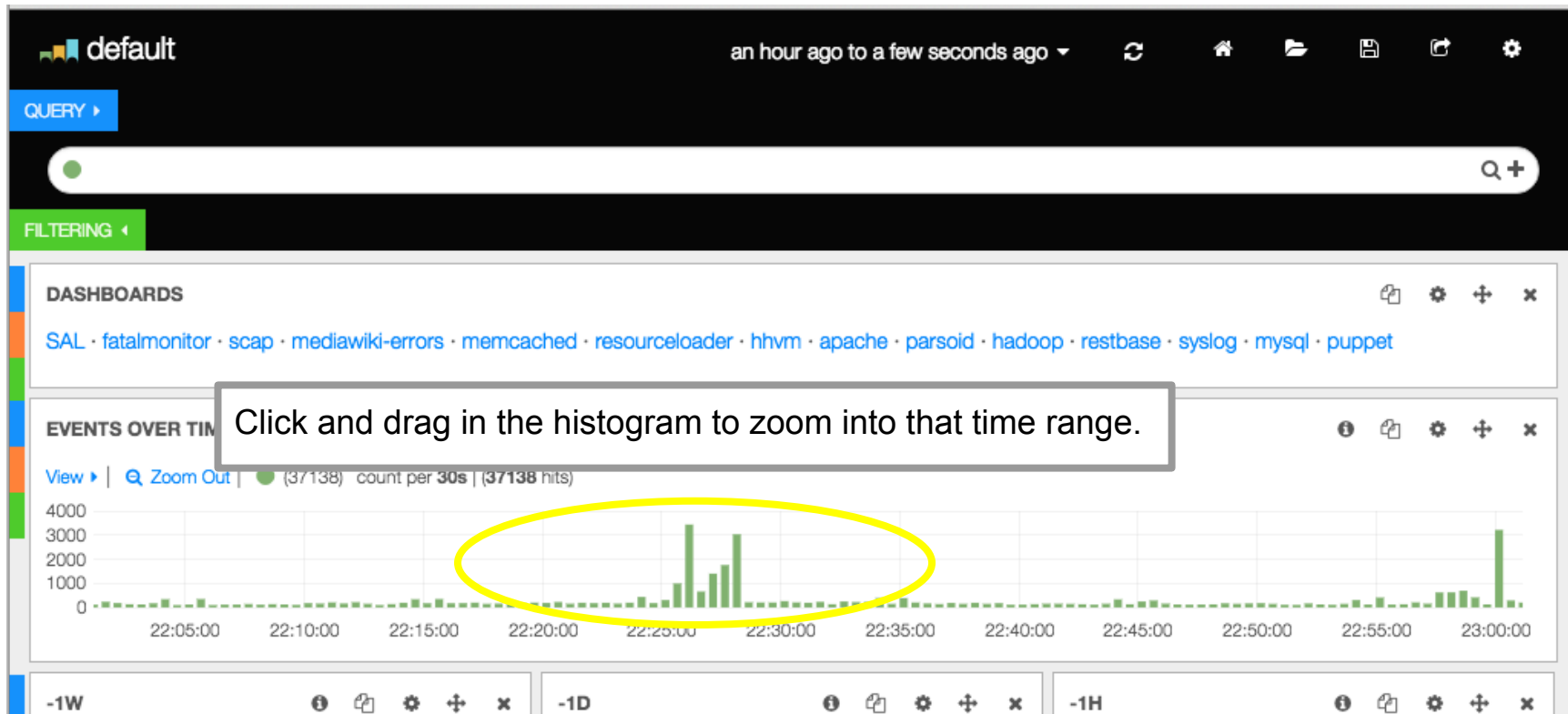
View | Zoom Out | (37138) count per 30s | (37138 hits)

4000
3000
2000
1000
0

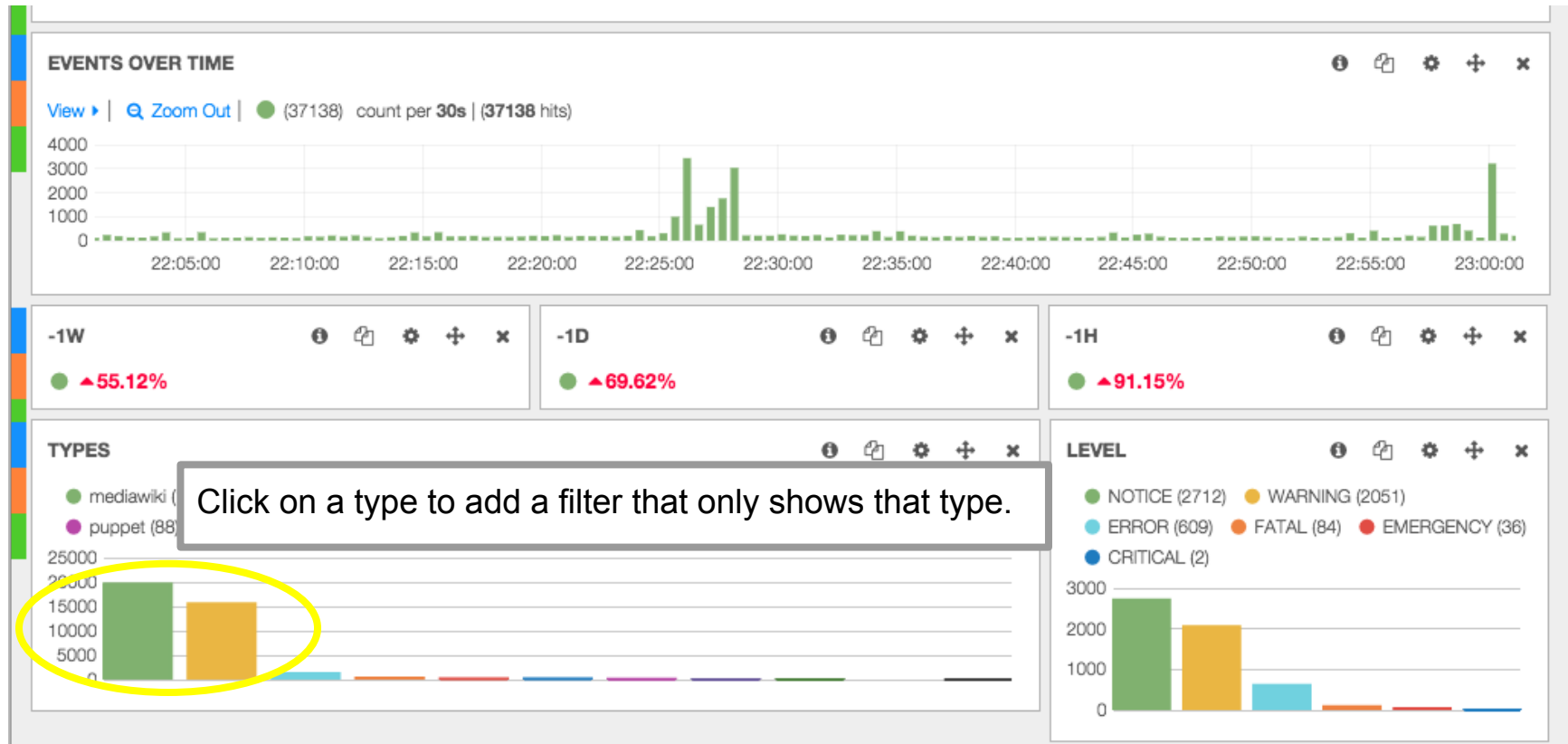
22:05:00 22:10:00 22:15:00 22:20:00 22:25:00 22:30:00 22:35:00 22:40:00 22:45:00 22:50:00 22:55:00 23:00:00

-1W -1D -1H

Using Kibana



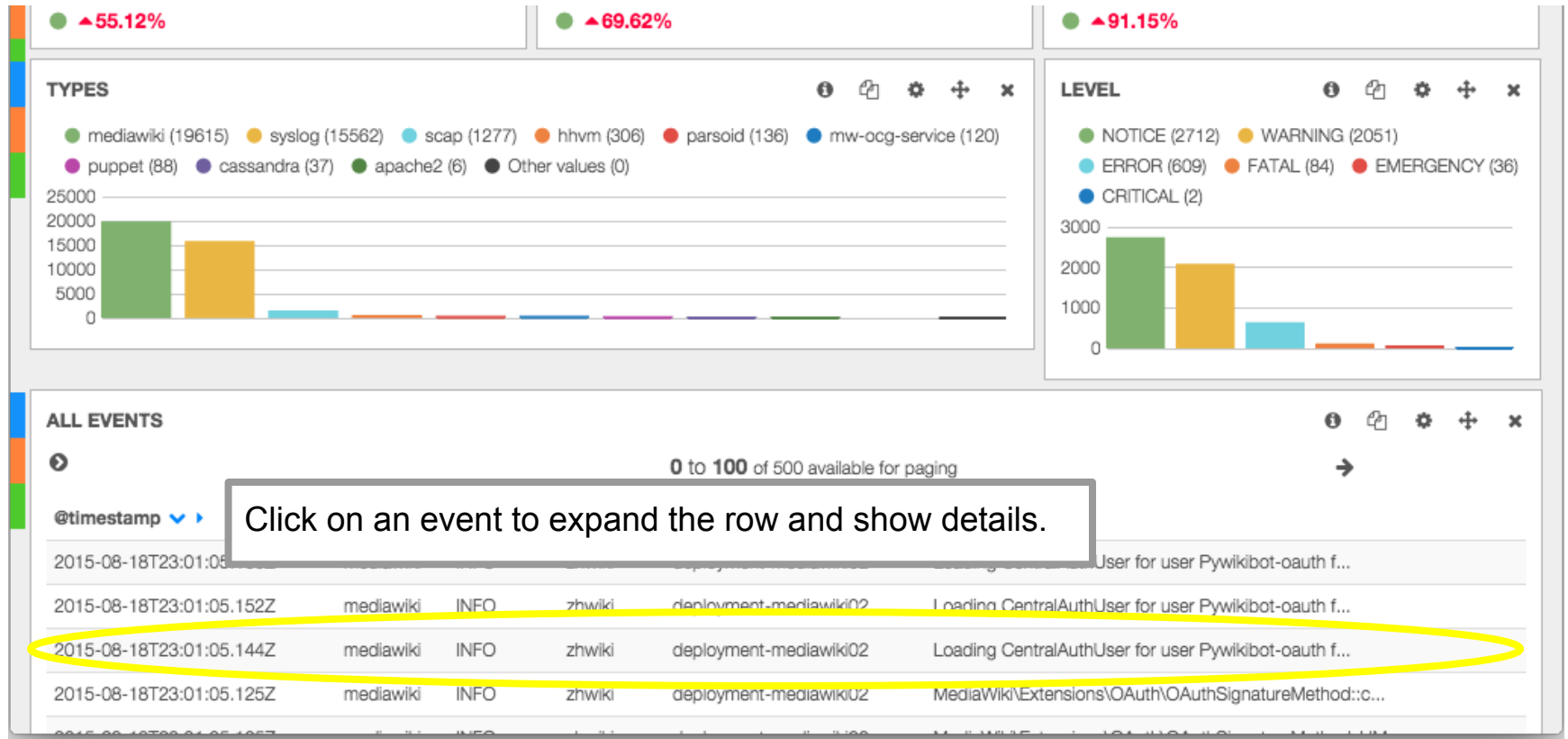
Using Kibana



Using Kibana



Using Kibana



Using Kibana

2015-08-19T00:52:45.997Z

mediawiki




















































INFO

enwiki

deployment-mediawiki01

Loading CentralAuthUser for user Test3544535452 fr...

View: [Table](#) / [JSON](#) / [Raw](#)

Field	Action	Value
@timestamp	  	2015-08-19T00:52:45.997Z
@version	  	1
_id	  	AU9Dcad_L84K6YCA-e4F
_index	  	logstash-2015.08.19
_type	  	mediawiki
channel	  	CentralAuthVerbose
host	  	deployment-mediawiki01
normalized_message	  	Loading CentralAuthUser for user Test3544535452 from cache object
private	  	false
process_id	  	29383
referrer	  	http://en.wikipedia.beta.wmflabs.org/wiki/Talk:Flow
server	  	en.wikipedia.beta.wmflabs.org
tags	  	syslog_es_es_normalized_message_untrimmed
type	  	mediawiki
uid	  	6171597
url	  	/w/api.php?action=flow&format=json&submodule=view-topiclist&page=Talk%3AFlow&vtloffset=0&vtllimit=50&vtlsortby=updated&vtlsavesortby=1
wiki	  	enwiki

Click magnifying glass to add a filter to only show events with fields that exactly match this field.

Using Kibana

2015-08-19T00:52:45.997Z

mediawiki




















































INFO

enwiki

deployment-mediawiki01

Loading CentralAuthUser for user Test3544535452 fr...

View: [Table](#) / [JSON](#) / [Raw](#)

Field	Action	Value
@timestamp	  	2015-08-19T00:52:45.997Z
@version	  	1
_id	  	AU9Dcad_L84K6YCA-e4F
_index	  	logstash-2015.08.19
_type	  	mediawiki
channel	  	CentralAuthVerbose
host	  	deployment-mediawiki01
normalized_message	  	Loading CentralAuthUser for user Test3544535452 from cache object
private	  	false
process_id	  	29383
referrer	  	http://en.wikipedia.beta.wmflabs.org/wiki/Talk:Flow
server	  	en.wikipedia.beta.wmflabs.org
tags	  	syslog_es_es_normalized_message_untrimmed
type	  	mediawiki
uid	  	6171597
url	  	/w/api.php?action=flow&format=json&submodule=view-topiclist&page=Talk%3AFlow&vtloffset=0&vtllimit=50&vtlsortby=updated&vtlsavesortby=1
wiki	  	enwiki

Click ban circle to add a filter to exclude events with fields that exactly match this field.

Using Kibana

2015-08-19T00:52:45.997Z

mediawiki





























































INFO

enwiki

deployment-mediawiki01

Loading CentralAuthUser for user Test3544535452 fr...

View: [Table](#) / [JSON](#) / [Raw](#)

Field	Action	Value
@timestamp	  	2015-08-19T00:52:45.997Z
@version	  	1
_id	  	AU9Dcad_L84K6YCA-e4F
_index	  	logstash-2015.08.19
_type	  	mediawiki
channel	  	CentralAuthVerbose
host	  	deployment-mediawiki01
http_	  	
ip	  	
level	  	INFO
message	  	Loading CentralAuthUser for user Test3544535452 from cache object
normalized_message	  	Loading CentralAuthUser for user Test3544535452 from cache object
private	  	false
process_id	  	29383
referrer	  	http://en.wikipedia.beta.wmflabs.org/wiki/Talk:Flow
server	  	en.wikipedia.beta.wmflabs.org
tags	  	syslog_es_es_normalized_message_untrimmed
type	  	mediawiki
uid	  	6171597
url	  	/w/api.php?action=flow&format=json&submodule=view-topiclist&page=Talk%3AFlow&vtloffset=0&vtllimit=50&vtlsortby=updated&vtlsavesortby=1
wiki	  	enwiki

Click grid to add this field as a table column.

Search tips

- Equivalent of /a/mw-log/fatal.log on fluorine:
 - `type:mediawiki AND channel:fatal`
- In general:
 - `type:mediawiki AND channel:$logGroup`
- If you get no results:
 - try a larger time range
 - check for filters in the collapsed filter section
- Shorter time ranges == faster searches

Better log output from MediaWiki

```
<?php
use MediaWiki\Logger\LoggerFactory;

$logger = LoggerFactory::getInstance( "channel_name" );

$logger->debug( "Something too spammy for prod logging" );
$logger->info( "Valuable state change info" );
$logger->warning( "Soft error condition" );
$logger->error( "Hard error that needs attention" );
```


Adding context to messages

```
<?php
```

```
$logger->info( "Valuable state change", array(  
    'thing' => $thing,  
    'method' => __METHOD__,  
    'line' => __LINE__,  
) );
```

```
$logger->error( "Badness {exception}", array(  
    'exception' => $exception,  
) );
```

More fun with Logstash


 SAL Prod | RelEng | Other

search production SAL 1-50 of 53207 results (21ms)
« [1](#) [2](#) [3](#) [4](#) [5](#) ... »

2015-08-20

16:38	<YuviPanda>	puppet swat done
15:44	<akosiaris>	uploaded to apt.wikimedia.org jessie-wikimedia: etherpad-lite_1.5.7-1
15:43	<krenair@tin>	Synchronized php-1.26wmf18/extensions/ContentTranslation/modules/tools/ext.cx.tools.reference.js: https://gerrit.wikimedia.org/r/#/c/232729/ (duration: 00m 12s)
15:42	<krenair@tin>	Synchronized php-1.26wmf19/extensions/ContentTranslation/modules/tools/ext.cx.tools.reference.js: https://gerrit.wikimedia.org/r/#/c/232730/ (duration: 00m 13s)
15:39	<krenair@tin>	Synchronized wmf-config/InitialiseSettings.php: https://gerrit.wikimedia.org/r/#/c/206480/ (duration: 00m 13s)
15:38	<krenair@tin>	Synchronized wmf-config/CommonSettings.php: https://gerrit.wikimedia.org/r/#/c/206480/ (duration: 00m 13s)
15:32	<krenair@tin>	Synchronized php-1.26wmf18/extensions/ContentTranslation/api/ApiContentTranslationPublish.php: https://gerrit.wikimedia.org/r/#/c/232687/ (duration: 00m 13s)
15:31	<krenair@tin>	Synchronized php-1.26wmf19/extensions/ContentTranslation/api/ApiContentTranslationPublish.php: https://gerrit.wikimedia.org/r/#/c/232688/ (duration: 00m 11s)
15:27	<greg-g>	on mw2187: rsync: failed to set times on "/srv/mediawiki/wmf-confia": Read-only file system (30)
15:25	<krenair@tin>	Synchronized wmf-config/InitialiseSettings.php: https://gerrit.wikimedia.org/r/#/c/206480/
15:08	<urandom>	restarting restbase1001 to apply temporary heap size of https://tools.wmflabs.org/sal/

More fun with Logstash

 **Quips**

random search

141-160 of 507 results (23ms)

[«](#) ... [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) ... [»](#)

<TimStarling> just saying, since I'm a java developer, I have to tell you that PHP is not a real language and you couldn't possibly do anything complex in it

[\[link\]](#)

<MarkTraceur> Then again, I've seen projects in the past that just used free licenses for convenience, so if you don't actually want people to use free software, you could probably ignore this safely.

[\[link\]](#)

Silke> I am not a coder so I did not realize I have to ask for a style guide before starting. Now I got a lot of contradictory comments from the "tabs vs. spaces universe", but nobody has actually given me feedback

[\[link\]](#)

<sumanah> I have just discovered that Roan is a GitHub repo <https://github.com/cobrateam/roan> specifically, he is a per-model url purging Django app "Roan uses only an optional setting: ROAN_PURGE_URL"

[\[link\]](#)

<kaldari> jesus! this is the article on Islam!

[\[link\]](#)

<https://tools.wmflabs.org/bash/>

Learning more

- <https://www.elastic.co/>
- <https://www.elastic.co/webinars/introduction-elk-stack>
- <https://www.elastic.co/guide/en/kibana/3.0/index.html>
- <https://wikitech.wikimedia.org/wiki/Logstash>
- <https://phabricator.wikimedia.org/tag/elk/>
- <https://phabricator.wikimedia.org/diffusion/OPUP/browse/production/files/logstash/>
- https://www.mediawiki.org/wiki/Structured_logging
- <https://phabricator.wikimedia.org/tag/psr3/>

Credits

- "[Male Elk \(8005280487\)](#)" by [Tony Hisgett](#) from Birmingham, UK. Licensed under [CC BY 2.0](#) via [Wikimedia Commons](#).
- Elasticsearch is a trademark of Elasticsearch BV, registered in the U.S. and in other countries.
- Kibana is a trademark of Elasticsearch BV, registered in the U.S. and in other countries.
- logstash is a trademark of Elasticsearch BV

Copyright © 2015, [Bryan Davis](#) and the [Wikimedia Foundation](#).

This work is licensed under a [Creative Commons Attribution-ShareAlike 3.0 Unported License](#).

