



# Drupal content analysis: Elastic is here to help

Joshua Li & Dr Joshua Rich

# About Us

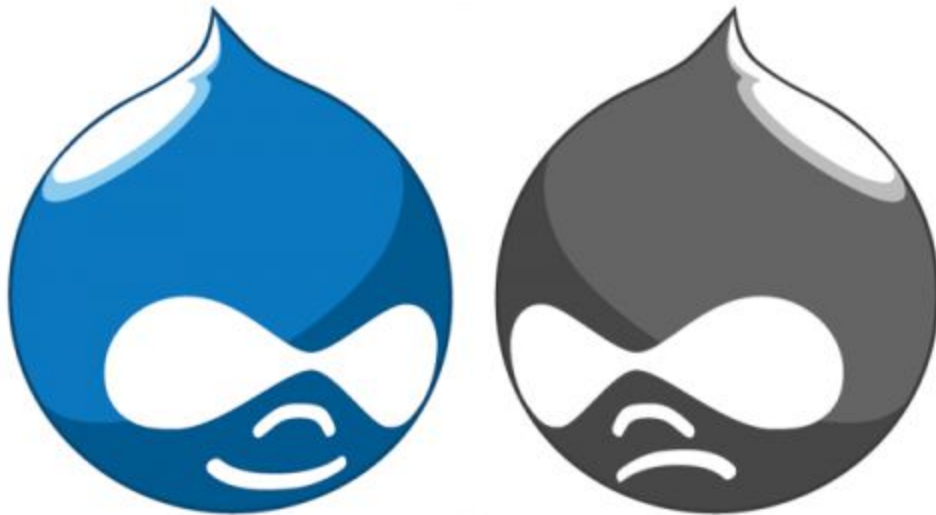
- Joshua Li:
  - Drupal developer in Canberra
  - Public sector
- Joshua Rich:
  - Technical Support Engineer with Elastic
  - PhD in Astrophysics
  - Scientific and HPC background

# The Drupal we know

- Adopted widely especially by **large** organizations
- Big data in Drupal
  - Research data
  - News
  - E-commerce data
  - Recipes
  - Everything we can put in Drupal

# The Drupal we know

- Drupal keeps data organized
  - Views
  - Taxonomy
  - Search API
  - Facet API
- Drupal makes things complicated too
  - Entity references (+ reference)
  - + Field collection + paragraphs
  - + flags



# Can I get a report of our content in our websites?

What does it mean?

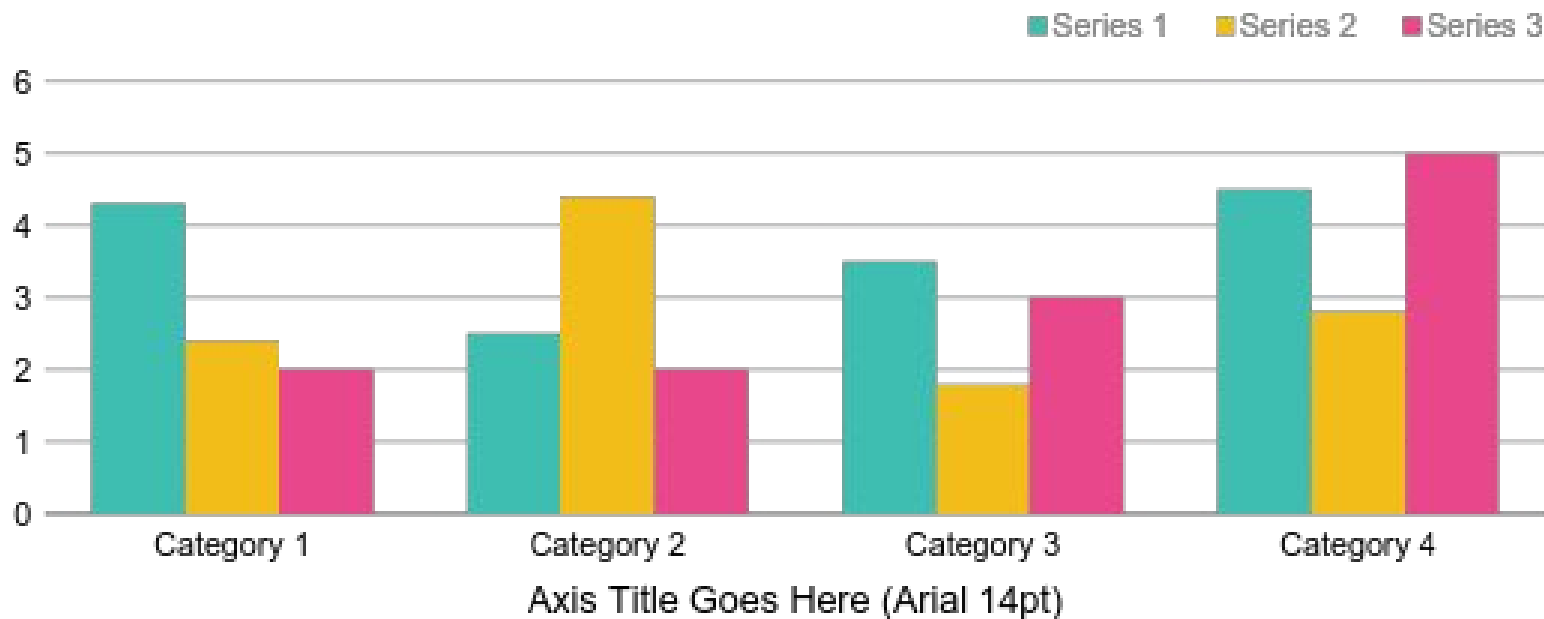


# It does not mean:

- Give him the total number of each content type
- Give him the total number of each content type + per term
- or grouped by user/organic group/etc

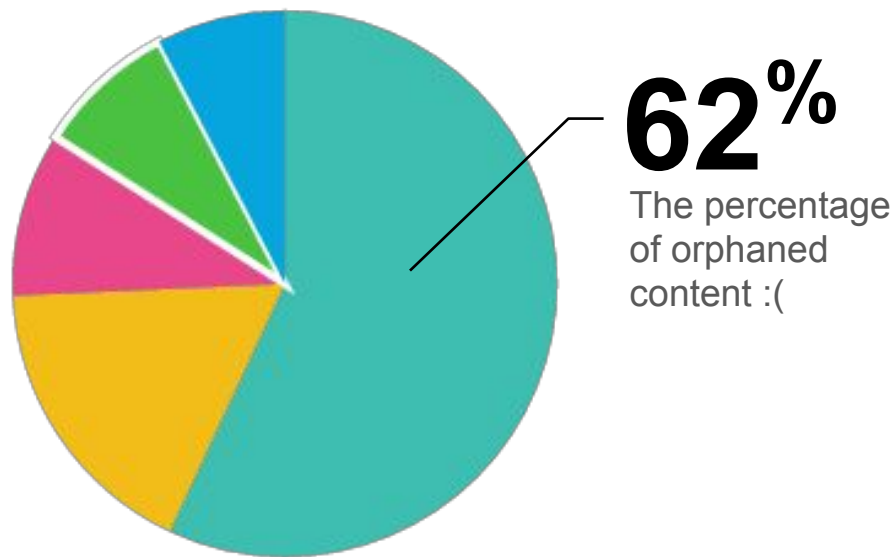


# The report my boss likes



# The report my boss likes

1st Quarter 2nd Quarter 3rd Quarter 4th Quarter 5th Quarter





# Visualisation dashboard

- Client also requires some dashboard ...
- Normally they don't know what they want
- But it should look nice ...
- And it should be public accessible





**And can you export it and print it out ?**

Anonymous boss

# What I can do after the swearing?

- Diving into mySQL start to join the tables
- Starting to create the new views
- Luckily I have facet API ...
- I think I may have to create a module for it
  - Hmm, D3 module, heard of it
  - Print friendly ?



**Until one day I  
had a drink with  
Dr Rich ...**



The background is a solid teal color with several overlapping, semi-transparent, wavy lines in a lighter shade of teal. These lines create a sense of movement and depth, resembling stylized waves or abstract patterns.

**Let's talk about Elastic...**

# Elastic Overview



**70,000+**

Community  
Members



**70M+**

Product  
Downloads



**2,400+**

Subscription  
Customers

**We aim to help you make your data usable in real time to power mission critical applications that solve today's real problems**

# Introducing the Elastic Stack, X-Pack, and Cloud

## Elastic Stack

User Interface



Kibana

Store, Index,  
& Analyze



Elasticsearch

Ingest



Logstash



Beats



X-Pack

Security

Alerting

Monitoring

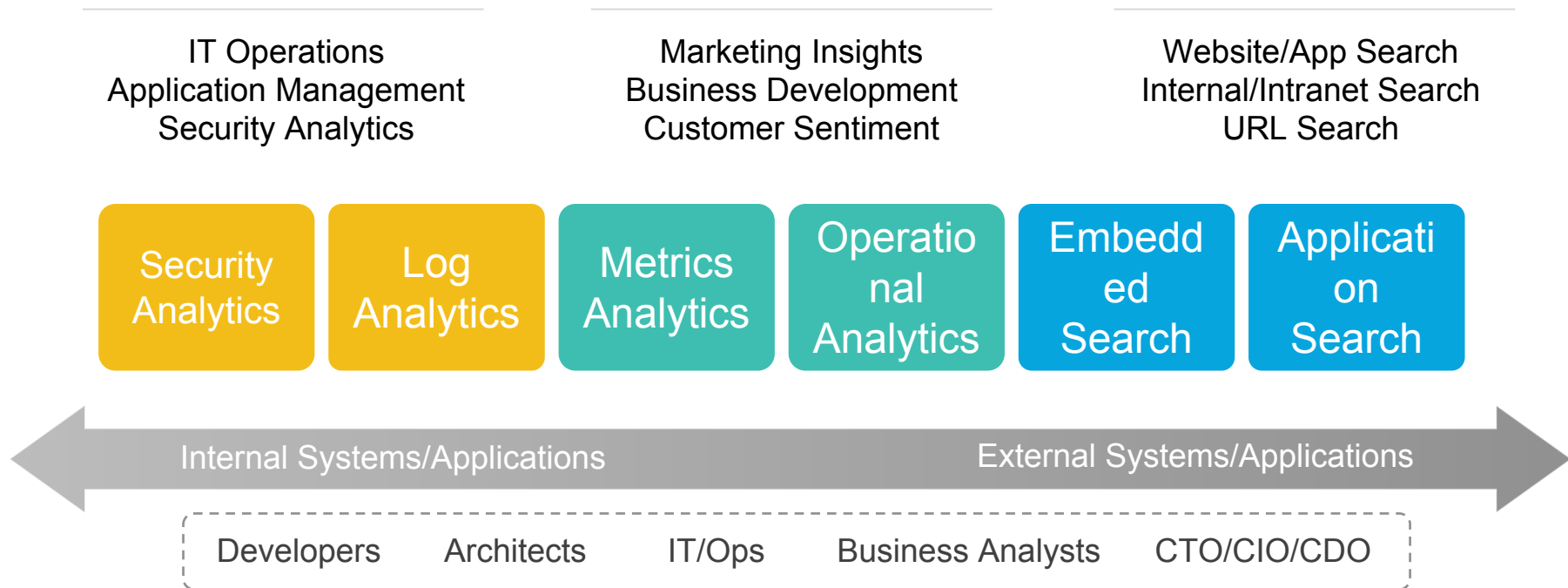
Reporting

Graph



Elastic Cloud

# A Solution for *Every Use Case* and *Everyone*





# What does it mean?

How can I leverage the Elastic stack for my project?



# Know your audience

## Users



## Managers



We are also the audience

---

**You know, for search!**

---

### Distributed & Scalable

---

- Resilient; designed for scale-out
- High availability; multitenancy
- Structured & unstructured data

### Developer Friendly

---

- Schemaless
- Native JSON
- Client libraries
- Apache Lucene

### Search & Analytics

---

- Real-time
- Full-text search
- Aggregations
- Geospatial
- Multilingual

# For full text search (with auto-complete):

WIKIPEDIA



**English**  
*The Free Encyclopedia*  
5 010 000+ articles

**Español**  
*La enciclopedia libre*  
1 212 000+ artículos

**Русский**  
*Свободная энциклопедия*  
1 267 000+ статей

**Deutsch**  
*Die freie Enzyklopädie*  
1 875 000+ Artikel

**日本語**  
*フリー百科事典*  
991 000+ 記事

**Français**  
*L'encyclopédie libre*  
1 696 000+ articles

**Italiano**  
*L'enciclopedia libera*  
1 235 000+ voci

**中文**  
*自由的百科全书*  
847 000+ 條目

**Português**  
*A enciclopédia livre*  
894 000+ artigos

**Polski**  
*Wolna encyklopedia*  
1 143 000+ haseł

English

Find a language:

- Elastic
- Elastica
- Elasticity (physics)
- Elasticity
- Elastic modulus
- Elastic p-wave
- Elastic energy
- Elastic Heart
- Elastic fiber
- Elastic Love

• Dansk • Eesti • Ελληνικά • Esperanto • E  
lynporsk) • Нохчийн • O'zbekcha / Ўзбекча • Português • Қазақша / Qazaqşa / قازاقشا • Ror

# Faceted/Structured search:

Work type (up to 4)

- All Work Types
- Full Time
- Part Time
- Contract/Temp
- Casual/Vacation

Show results from

Any

Call Centre & Customer Service

CEO & General Management

Community Services & Development

Sub-Classification (up to 10)

- Any Sub-Classification
- Account & Relationship Management
- Accounts Officers/Clerks
- Accounts Payable
- Accounts Receivable/Credit Control
- Actuarial
- Administration
- Administrative Assistants

Fewer options

SEEK

Your SEEK Profile never stops searching for your next job.

Update now

Browse jobs by:

**Classifications** | Accounting | Education & Training | Government & Defence | Healthcare & Medical | Sales | [view all classifications](#)

**Major Cities** | Sydney | Melbourne | Brisbane | Gold Coast | Perth | Adelaide | Hobart | Darwin | Canberra

**Salary** | Jobs paying more than \$150k

**Advertisers** | Companies | Recruiters

# Geo search:

Domain

BuyRentNew HomesSoldCommercialNewsAdviceAgentsMore

Gold Coast, QLD

Filters

List

Map

Featured

There's too many pins to display on the map, please zoom in to refine search

1000+ Found

FEATURED PROPERTY

Offers Above \$695,000

58 Greg Norman Crescent,  
Parkwood, QLD 4214

4

2

2

Inspection Sat 15 Oct

Telly Karadimos

Barry Plant

Give Feedback?

23

elastic

---

**You know, for analytics!**

---





### Discover Insights

---

- Explore and analyze patterns in data; drill down to any level
- Leverage powerful analytical capabilities in Elasticsearch

### Customize & Share

---

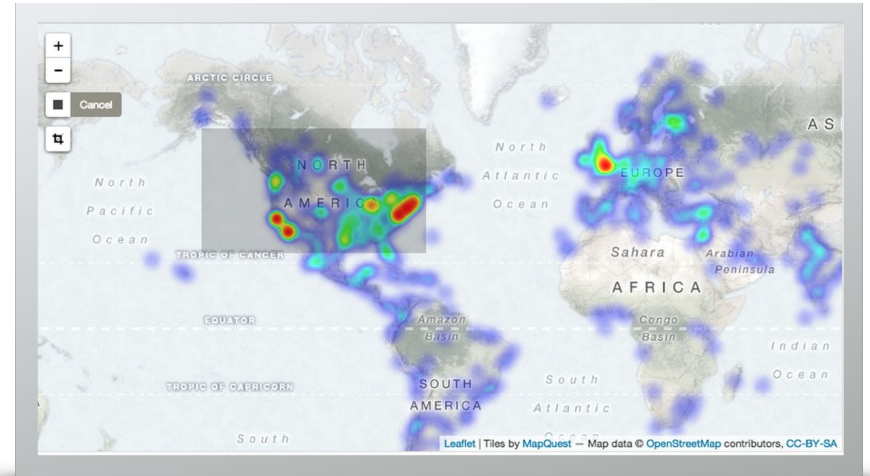
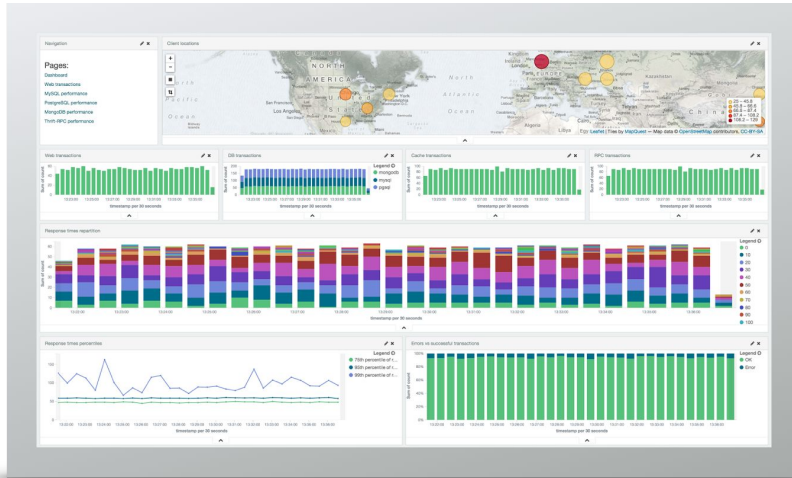
- Create bar charts, line and scatter plots, maps and histograms
- Share and embed dashboards into operational workflows

### Window into Elastic Stack

---

- Unified user interface for data visualization
- Administration and management for the Elastic Stack
- Pluggable architecture to create custom visualizations and applications

## Visualize and Explore



Easily create bar charts, line and scatter plots, histograms, pie charts, and maps

---

**You know, for ingest!**

---



- Data collection and enrichment; 200+ plugins
- Next generation data pipeline; micro-batches, process groups of events



- Platform to build lightweight, data shippers
- Forward host-based metrics and any data to Elasticsearch



- Two-way connector to integrate with HDFS, Spark, MapReduce, etc.
- Enable real-time search queries on Hadoop data

# Demo Data

ABC Local Online Photo Stories

# Info

- Metadata and links for around 8,500 stories created around Australia by ABC Local Online reporters.
- All the stories are currently published live on the [abc.net.au](http://abc.net.au) site.
- Data consists of:
  - Photo Caption
  - Photo URL
  - Story URL
  - Reporting station
  - State
  - Geo coordinates of story location
- <https://data.gov.au/dataset/abc-local-online-photo-stories-2009-2014>

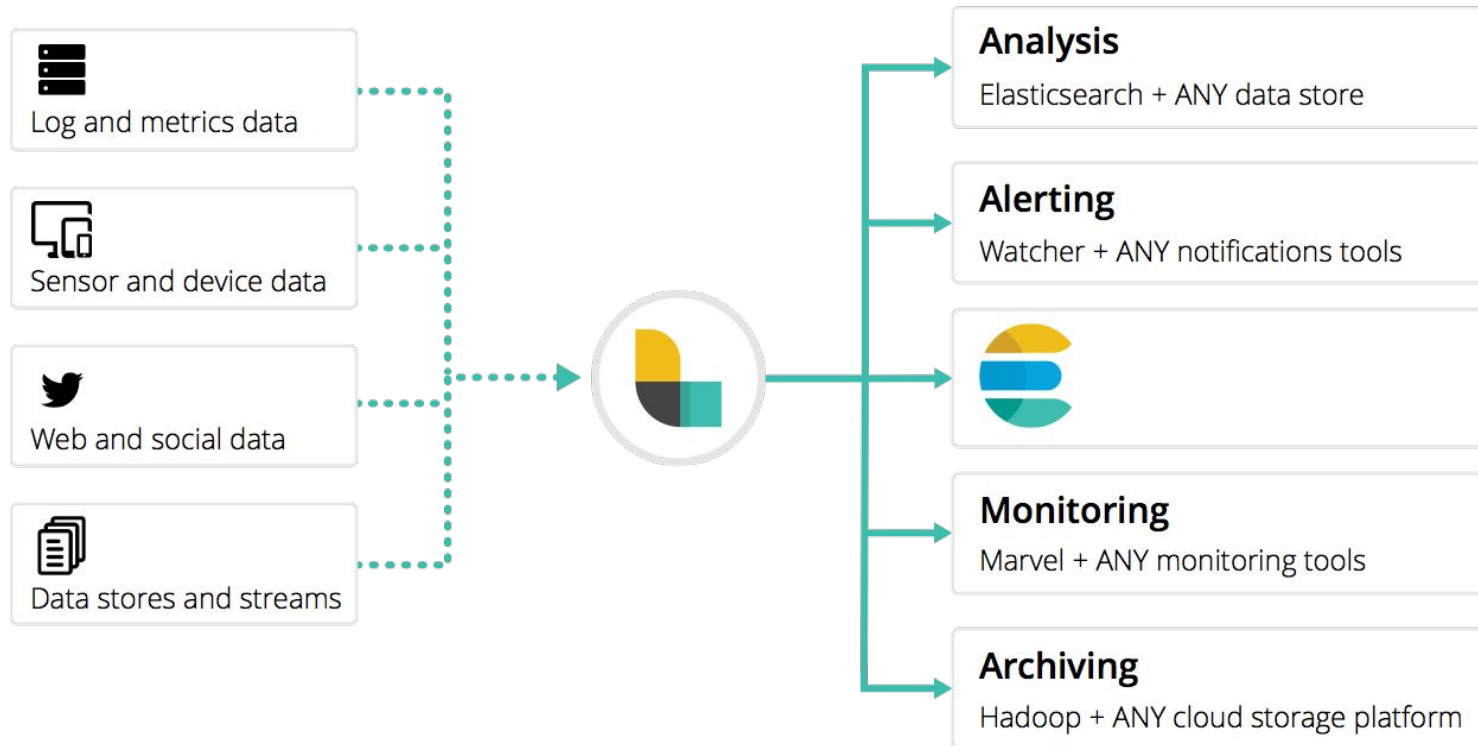
# How easy is it to use the Elastic Stack?

As easy as one-two-three:

1. Download the CSV from [data.gov.au](https://data.gov.au)
2. Index the data into an Elasticsearch instance on [cloud.elastic.co](https://cloud.elastic.co)\*
3. Make purty visualisations in Kibana :)



## Collect, Enrich, and Transport





# Comparison to other search engines/data stores

# Elasticsearch and Solr

## Some quick points:

- Both are based on Apache Lucene.
- Elasticsearch is accessed by a JSON-based RESTful API.
- Elasticsearch has near real time search and instant direct document retrieval capabilities.
- Elasticsearch is designed to scale horizontally.
  - All features are implemented to be distributed.
  - It has clustering baked in (inc. HA and resiliency).
- Elasticsearch has powerful data analytics built-in using aggregations.
  - This ranges from simple statistics (e.g., sums, averages) through to derivatives and moving averages.

# Elasticsearch, MySQL, Postgres & other RDBMS

We can't compare these vastly different tools!

- If you have well-structured data and don't need search, use an RDBMS.
- If your data isn't well structured, or you anticipate its structure will change, try Elasticsearch.
- If you need search, use Elasticsearch.
- Completely different interface languages, SQL to JSON.
- Elasticsearch has clustering built-in, multi-node replication.

---

**Bottom line:**  
**Use the right tool for the job.**

---

*In this case, the right tool is Elasticsearch :)*

# Putting/getting the content from Drupal to Elasticsearch



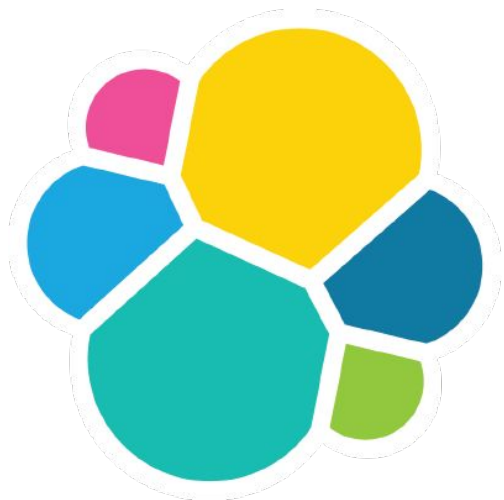
---

**There is a module for that ...**

---

# The Elasticsearch Drupal modules

- Elasticsearch connector (Porting to D8 now)
- Search API elasticsearch (No D8 branch yet)
- Elasticsearch helper (Working in D8, but needs more work)



## Write your own modules

---

# Elasticsearch-PHP

---

PHP API: 2.0 (current) ⬆⬇⬇⬆



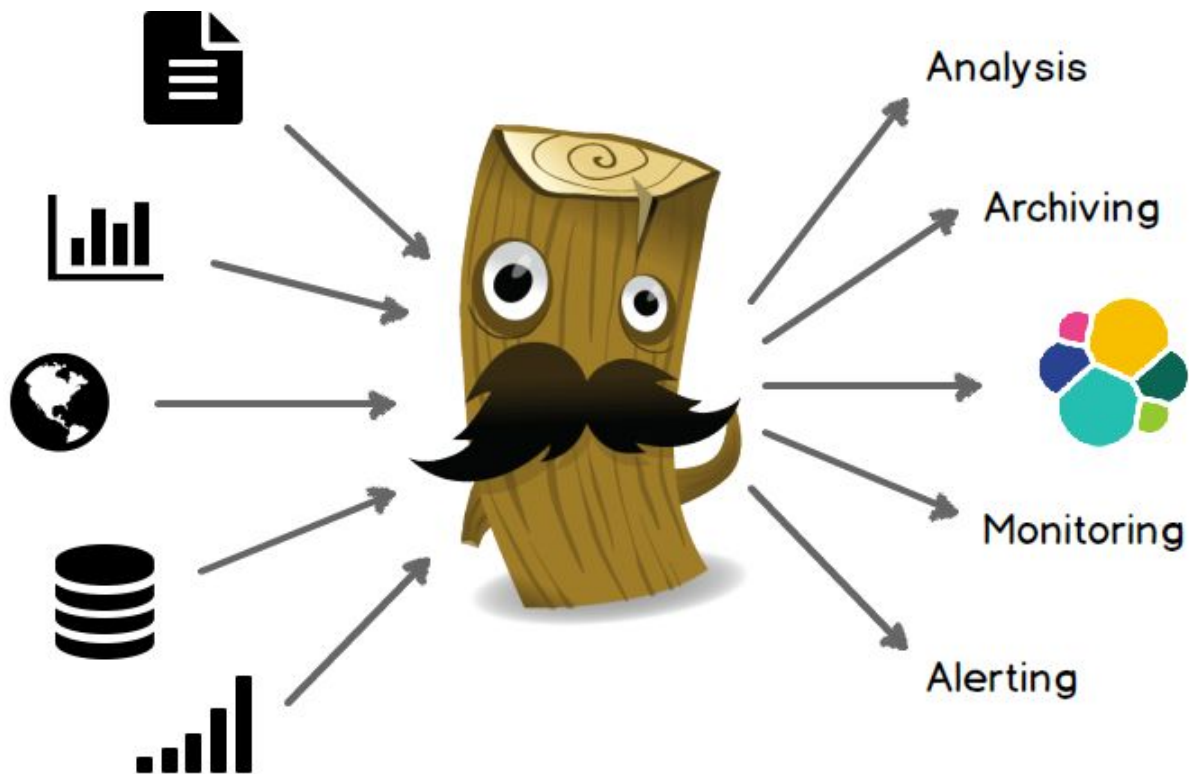
# And how ?

```
/**
 * Implements hook_entity_insert().
 */
function elasticsearch_helper_entity_insert(EntityInterface $entity) {
    \Drupal::service('plugin.manager.elasticsearch_index.processor')->indexEntity($entity);
}

/**
 * Implements hook_entity_update().
 */
function elasticsearch_helper_entity_update(EntityInterface $entity) {
    elasticsearch_helper_entity_insert($entity);
}

/**
 * Implements hook_entity_delete().
 */
function elasticsearch_helper_entity_delete(EntityInterface $entity) {
    /** @var ElasticsearchIndexManager $elasticsearchPluginManager */
    \Drupal::service('plugin.manager.elasticsearch_index.processor')->deleteEntity($entity);
}
```

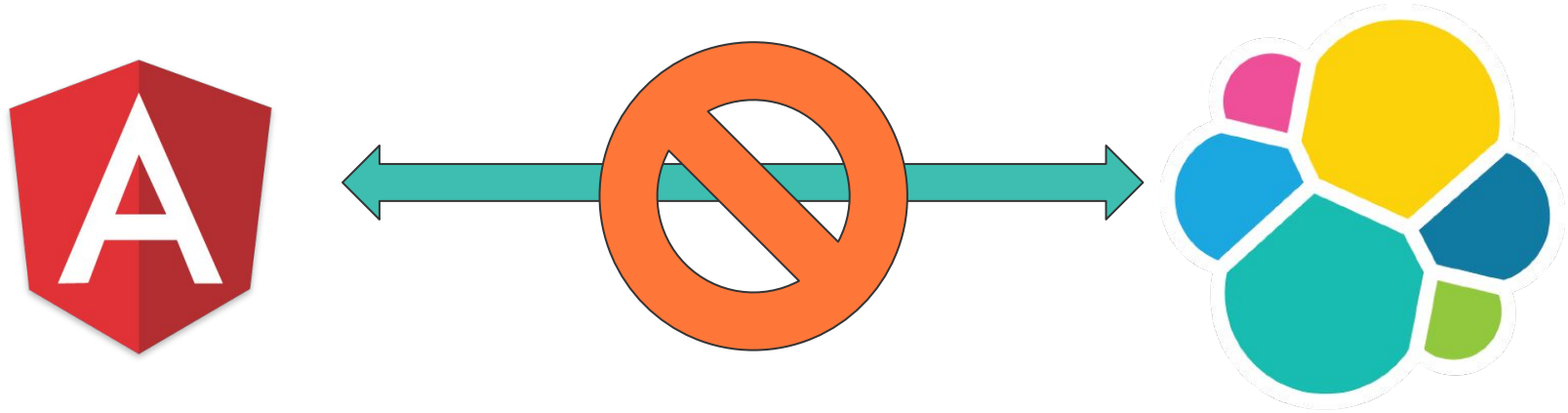
# Logstash



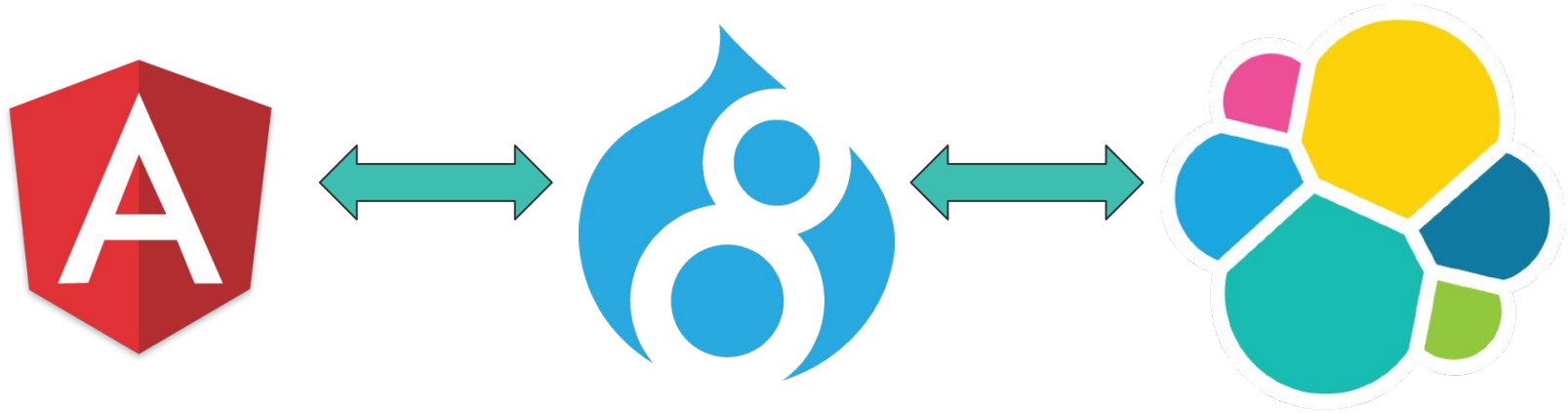
**BEYOND**



# Web app and Elasticsearch



# Web app and Elasticsearch



# Questions ?





# Talk to us:

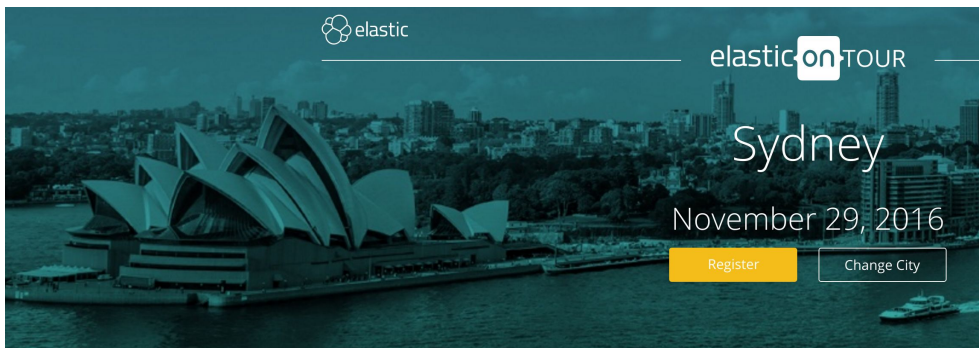
Joshua Li  
joshua@lirujia.com | @rujiali

Joshua Rich  
joshua.rich@elastic.co | @devopswannabe

WED 26TH - FRI 28TH  
OCTOBER 2016



Q1, GOLD COAST  
QUEENSLAND, AUSTRALIA

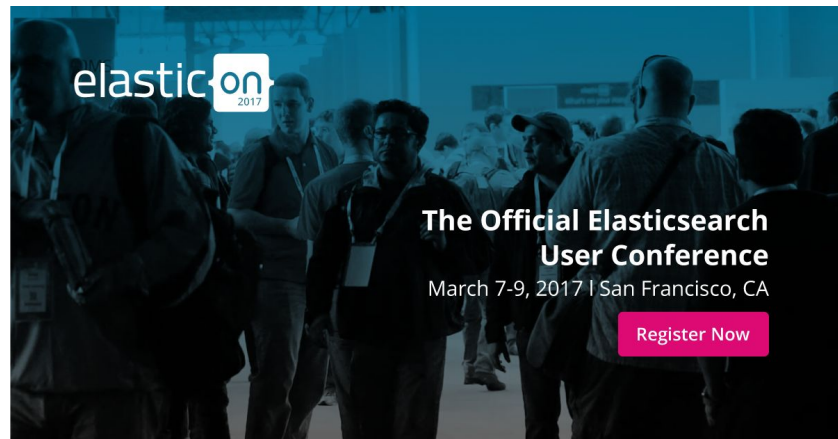


**This year: we're on tour!**

<https://goo.gl/czDqfU>

**Next year: join us at our user  
conference in San Francisco!**

<http://go.es.io/2e38u90>





# References

- [https://www.drupal.org/project/elastic\\_search\\_connector](https://www.drupal.org/project/elastic_search_connector)
- [https://www.drupal.org/project/search\\_api\\_elasticsearch](https://www.drupal.org/project/search_api_elasticsearch)
- [https://www.drupal.org/project/elastic\\_search\\_helper](https://www.drupal.org/project/elastic_search_helper)
- <https://www.elastic.co/guide/en/elasticsearch/client/php-api/current/index.html>
- <https://www.elastic.co/guide/en/elasticsearch/client/index.html>
- <https://www.elastic.co/guide/en/logstash/current/index.html>

