



Search: A Journey of Delivery on a Budget

Sam McLeod [Operations]
Ricky Cook [Web Development]

smcleod.net | infoxchange.org

WHO IS @INFOXCHANGE?

Infoxchange is a not-for-profit community organisation that delivers technology for social justice.

We work to strengthen communities and organisations, using information technology as the primary tool to create positive social change.

Web Applications
IT Efficiency Consulting
Community Empowerment

OUR PRODUCTS



20 Web Applications (Django, Perl & PHP)
300+ hosted CMS websites
Application RPM = 4000 - 8000

Budget: *Tight*



ServiceSeeker

Community directors
An Infoxchange enterprise

WEB APPLICATIONS

Electronic Referral

Electronic Waitlist

Service Coordination

Health Project Management

Patient Management

Content Management



DigitalInclusion

Essential to social inclusion
An Infoxchange enterprise

SPECIALISING IN

Homelessness

Mental Health

Eldercare

Welfare



JobSeeker

Work in the social sector
An Infoxchange enterprise



VolunteerMatch
An Infoxchange enterprise



Lantern
Interpreting services
An Infoxchange enterprise



InfoWatch
An Infoxchange enterprise



Graffiti Visual
An Infoxchange enterprise

WebMultimedia

Information management and digital design
An Infoxchange enterprise

earch for: Tips

Suburb/Town or Postcode:

[Search]

RICH SEARCH HISTORY

Infoxchange established the first searchable electronic directory of community services in Australia in 1990

Since the 90's Service Seeker has included:

Name search

Keyword search

Regional maps

Vacancy search

1990 - ARCHIE

Webpage Screenshot

Welcome to archie.icm.edu.pl

Archie Query Form 

Search for:

Database: Worldwide Anonymous FTP Polish Web Index
Search Type: Sub String Exact Regular Expression
Case: Insensitive Sensitive

Do you want to look up strings only (no sites returned):
 NO YES

Output Format For Web Index Search: Keywords Only Excerpts Only Links Only

[Wersja polska](#) | [Help](#) | [Advanced Search](#)

Last changed on 01/24/1999 14:25:29 by [Wojtek Sylwestrzak](#)

1991 - SERVICESEEKER

Netscape - [Criteria]

File Edit View Go Bookmarks Options Directory Window Help

The info Xchange Criteria Search

The criteria search form is under development - we'd like [feedback](#) on this to help us get it working as you'd like it!

Please enter criteria words:

psychiatric

Related Terms

ringwood Suburb

Local Government Area

Barwon South Western Region
Gippsland Region
Grampians Region
Hume Region
Loddon Mallee Region
Metropolitan East Region
Metropolitan North Region
Metropolitan South Region
Metropolitan West Region

Human Services Region

DSS Region

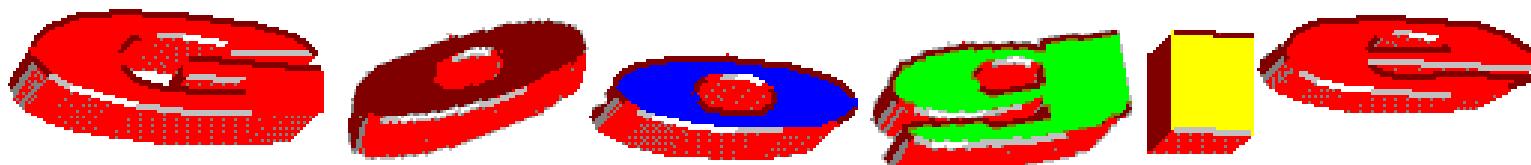
North
South
East

Select a service specifically for:

Children
 Youth
 Aged
 Family
 Female
 Male

Document: Done

1995 - GOOGLE



Search Stanford

10 results ▾ clustering on ▾

Search The Web

10 results ▾ clustering on ▾

1996 - SERVICESEEKER 2

 **ServiceSeeker**
Community directories
An Infoxchange enterprise

 **Infoxchange** Australia
Technology for social justice

[Home](#) | [About](#) | [Add Your Service](#) | [Usage Policy](#) | [Copyright](#) | [Bookmark](#) | [Contact Us](#)

Search

Australia's most extensive directory of community support services providing access to **331,979** health, welfare and community service records.

TEXT SEARCH [ACT](#) | [NSW](#) | [NT](#) | [QLD](#) | [SA](#) | [TAS](#) | **VIC** | [WA](#)

Victoria

[Quick Search](#) [Keyword](#) [Focus](#) [Provider](#) [Health Practitioner Name](#) [Health Focus](#) [Vacancy](#)

[Help](#)

Search for: [Suburb/Town or Postcode:](#) [Show search options](#) [Search](#)



© Infoxchange Australia 2012. This work is copyright. Apart from any use permitted under the Copyright Act 1968, all other rights are reserved.
Created and hosted by 

2014 - SERVICESEEKER 3 (API)

Services or keywords and suburb or postcode in NSW

Find services

Advanced Search

Displaying 100 results

Rosebank Child Sexual Abuse Service Liverpool

(02) 9601 3790 A counselling and support service for people who have been sexually abused.

[More information >](#)

Pregnancy Help Sydney Liverpool

(02) 9602 6543 Provides counselling, support and practical assistance for pregnant women and new mothers as well as pre and post termination counselling.

[More information >](#)

Pregnancy Counselling Service Liverpool

Pregnancy Help Sydney (02) 9602 6543 Confidential counselling and emotional support for those involved in a worrying pregnancy or those distressed after an abortion. Free pregnancy testing and a telephone support service are available.

[More information >](#)

The map displays the Liverpool town center with major streets like Bathurst St, Elizabeth Drive, and Macquarie St. Key landmarks include Woolworths Liverpool, Myer Liverpool, Liquorland, Westfield Liverpool, JB Hi-Fi Liverpool, Nando's, Liverpool Plaza, Roma Function Centre, and Liverpool City Council Libraries. Several blue pins indicate the locations of the services listed on the left: Rosebank Child Sexual Abuse Service, Pregnancy Help Sydney, and Pregnancy Counselling Service. Other pins are scattered across the map, likely representing other services in the area.

THE OPS JOURNEY



NGINX

kibana

XenServer



PostgreSQL
the world's most advanced open source database



GITLAB



ENVIRONMENTAL CONSTRAINTS

Small Operations Team

Slow SAN Performance

Limited Hardware Resources

No Search experience & limited Java experience

~~LOTS of Technical Debt~~

~~Elderly Network~~

ENVIRONMENTAL STRENGTHS

Modern Tech: BTRFS, Docker, ZRAM & Kernel 3.14

Strong DevOps and Agile Practises

No Bureaucracy == Agility

Open Source Architecture

Newly Deployed Network

Hardworking, *Passionate* People

HIGHLY DIVERSE ENVIRONMENT

What the hell have you built.

- Did you just pick things at random?
- Why is Redis talking to MongoDB?
- Why do you even *use* MongoDB?

Goddamnit

Nevermind FREE



ES HOSTS

3 Node Clusters

OS = *Debian 7 (Wheezy)*

CPU = 8x *E5-2660 @ 2.20GHz (24 Core Cluster)*

RAM = *10GB (30GB Cluster)*

Virt = *XenServer 6.2*

Kernel = *linux-3.14-amd64*

ES PERFORMANCE & TWEAKS

Heap size = half the available RAM:

```
ES_HEAP_SIZE = 5g
```

Faster shard recovery:

```
routing.allocation.node_initial_primaries_recoveries = 4,  
routing.allocation.node_concurrent_recoveries      = 15  
recovery.max_bytes_per_sec                         = 100mb,  
recovery.concurrent_stream                        = 5
```

Attempt to lock the process address space so it won't be swapped:

```
mlockall          = true,
```

Stop people from deleting all your nice things:

```
disable_delete_all_indices = true,
```

Ensure kernel keep-alive timeouts < routers timeouts!

TAKING OUT THE GARBAGE

"Stop-the-world garbage collection and blocking for disk IO can cause runtime latencies on the order of seconds to minutes."

Garbage collecting more often
=
less 'hit' to performance when it occurs

s/CMSInitiatingOccupancyFraction=[0-9]* /CMSInitiatingOccupancyFraction=35/g

<http://aphyr.com/posts/288-the-network-is-reliable>

MONITORING

Royrusso/ElasticHQ

Nodes: 3 Indices: 36 Shards: 297 Data: 6.65 GB CPU: 69% Memory: 3.53 GB / 11.35 GB JVM: 7.0.1 Memory: 1.20

MONITORING

Elasticsearch/Marvel

Powerful Cluster Insights

Easily Visualise Cluster Performance

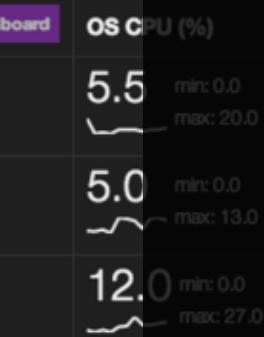
Puts Metrics In Context

Creates Very Large Indexes

Doesn't Clean Up After Itself

No Longer Free

Nodes / 0 selected / Last 10m



Load (1m)

0.7 min: 0.2 max: 1.9

0.5 min: 0.0 max: 1.2

0.9 min: 0.2 max: 1.2

JVM Mem (%)

26.0 min: 20.0

33.0 min: 21.0

27.5 min: 21.0 max: 38.0

Disk Free Space

17.3 GB min: 17.3 GB max: 17.3 GB

17.3 GB min: 17.3 GB max: 17.3 GB

17.0 GB min: 17.0 GB max: 17.0 GB

Indices / 0 selected / Last 10m

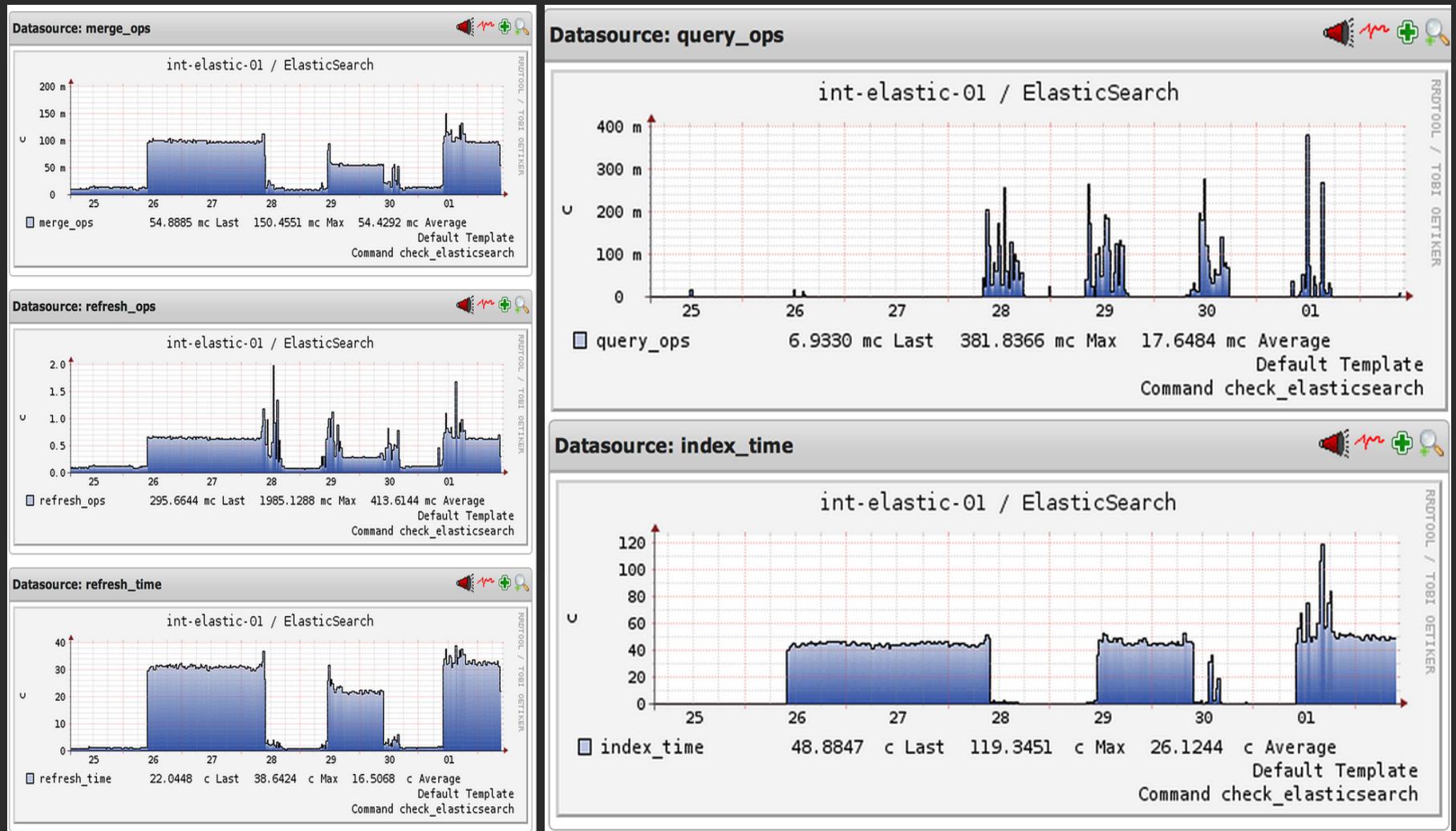
page refresh rate to 2m or greater for more options.

	Dashboard	Documents	Index Rate	Search Rate	Merge Rate
drupal_node		51	0.0	0.0	0.0 B
changeapps_net_au_drupal_node		37	0.0	0.0	0.0 B
_dev_drupal_node		37	0.0	0.0	0.0 B
dev_drupal_node		195	0.0	0.0	0.0 B

MONITORING

Nagios - PNP4Nagios

Aggregates Nagios Performance Information



MONITORING

Nagios + check_mk multisite

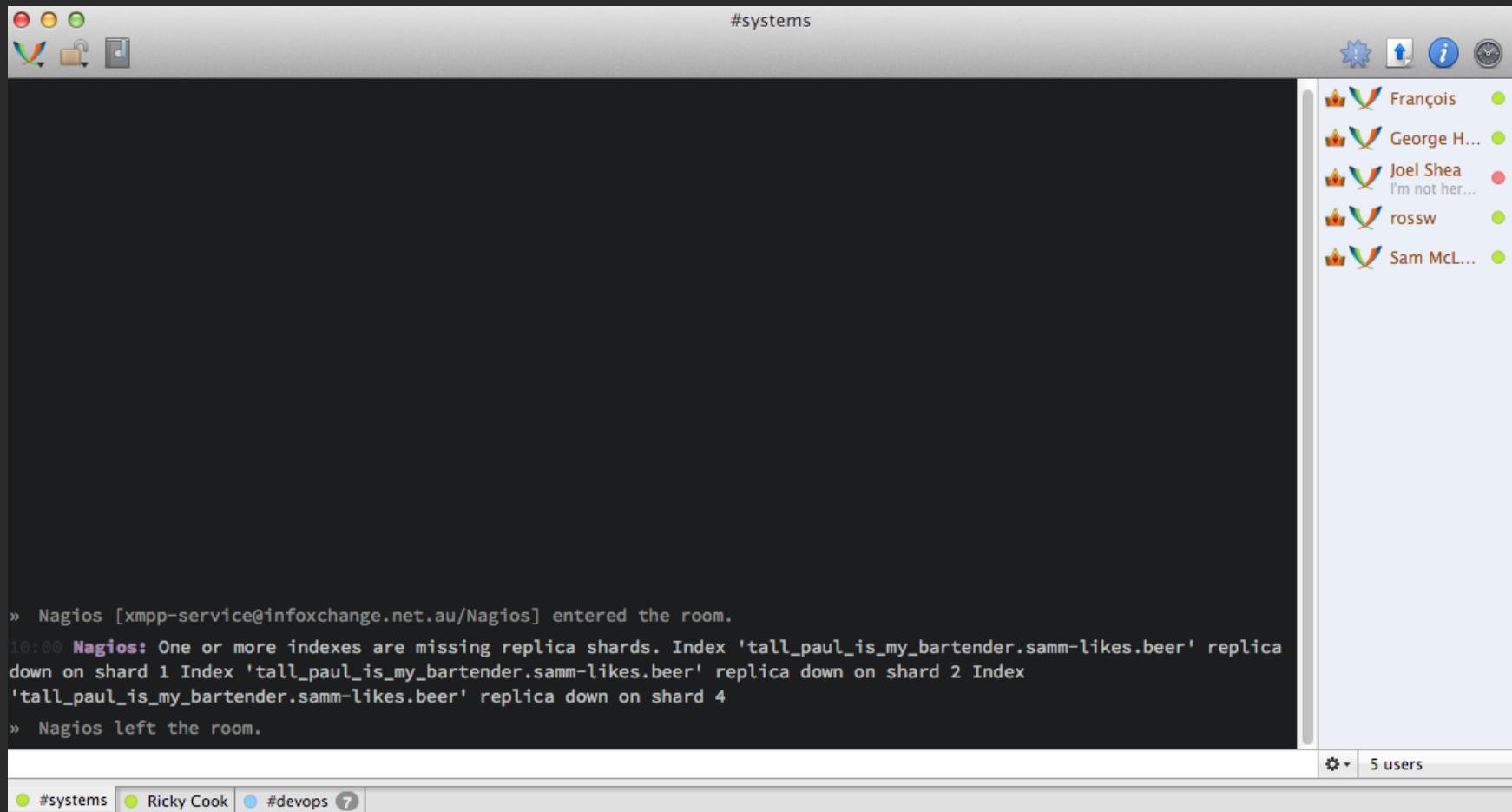
Cross-Site Elasticsearch Cluster Information

Events of recent 4 hours					
	Time	Host	Service	Check output	
	10 sec	int-elastic-03	ElasticSearch	Monitoring cluster 'int-elastic-cluster'	
	10 sec	int-elastic-01	ElasticSearch	Monitoring cluster 'int-elastic-cluster'	
	11 sec	int-elastic-02	ElasticSearch	Monitoring cluster 'int-elastic-cluster'	
	85 sec	int-elastic-01	ElasticSearch	One or more indexes are missing replica shards. Index 'test_iss3_small.rperez-django.dev-20140502-094121' replica down on shard 1 Index 'test_iss3_small.rperez-django.dev-20140502-094121' replica down on shard 2 Index 'test_iss3_small.rperez-django.dev-20140502-094121' replica down on shard 4	
	85 sec	int-elastic-03	ElasticSearch	One or more indexes are missing replica shards. Index 'test_iss3_small.rperez-django.dev-20140502-094121' replica down on shard 1 Index 'test_iss3_small.rperez-django.dev-20140502-094121' replica down on shard 2 Index 'test_iss3_small.rperez-django.dev-20140502-094121' replica down on shard 4	
	85 sec	int-elastic-02	ElasticSearch	One or more indexes are missing replica shards. Index 'test_iss3_small.rperez-django.dev-20140502-094121' replica down on shard 1 Index 'test_iss3_small.rperez-django.dev-20140502-094121' replica down on shard 2 Index 'test_iss3_small.rperez-django.dev-20140502-094121' replica down on shard 4	

MONITORING

Nagios + XMPP (Jabber)

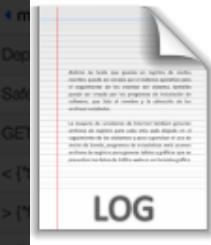
Realtime Cluster Notifications



_type:"docker" (371695) count per 5m | (371695 hits)

ES + LOGSTASH = SPLUNK4FREE

(Almost)



GET http://int-elastic-01.serv.abb.ixa.net.au:9200/_search/test...



_ss3-test.docker.dev

program	severity
wsxtest-deploy	Informational
wsxtest-deploy	Informational
elasticsearch	Informational
elasticsearch	Debug
elasticsearch	Debug
elasticsearch	Debug
elasticsearch	Informational

Action	Value
Q, ⚙	2014-06-04T04:55:02.000Z
Q, ⚙	1
Q, ⚙	QSYr8qq3Qe2rxsvUj44DTA
Q, ⚙	logstash-2014.06.04
Q, ⚙	docker
Q, ⚙	1
Q, ⚙	user-level
Q, ⚙	10.51.10.220
Q, ⚙	_ss3-test.docker.dev
Q, ⚙	GET http://int-elastic-01.serv.abb.ixa.net.au:9200/_ss3-test.docker.dev/_search/_search [status:200 request:0.013s]
Q, ⚙	56
Q, ⚙	14
Q, ⚙	elasticsearch
Q, ⚙	6
Q, ⚙	Informational
Q, ⚙	Jun 04 14:55:02
Q, ⚙	docker

LOGSTASH

Puppet Managed Logstash Config

```
ixalogstash::input { 'syslog-docker':
  input_type      => 'syslog',
  type_tag       => 'docker',
  input_port     => 5550,
}
ixalogstash::input { 'syslog-nginx':
  input_type      => 'syslog',
  type_tag       => 'nginx',
  input_port     => 5551,
}
```

```
ixalogstash::filter { 'syslog-nginx':
  type_tag      => 'nginx',
  filter_match  => '%{COMBINEDAPACHELOG} %{QS:vhost}' ,
}
```

```
ixalogstash::output { "${elastic_host}":
  output_type    => 'elasticsearch',
}
```

<https://github.com/elasticsearch/puppet-logstash>

LESSONS LEARNT

php (ext) (659) ● png (ext) (317) ● gif (ext) (229) ● css (ext) (121) count per 30m | (2214 hits)

Small Budget = Creative Solutions

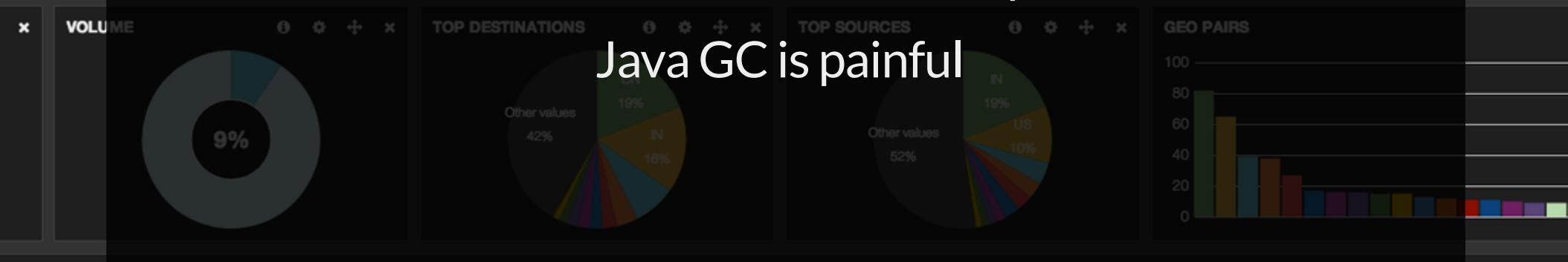
Puppet is great for managing Elasticsearch & Logstash

Bulk indexing can be slow (more from ricky soon...)

Test for split-brain (before it happens)

Modern OS & kernel helps

Java GC is painful



0 to 100 of 500 available for paging

	geo.srccdst	extension	clientip	bytes	id	phpmemory
s,security	MY:VN	html	167.12.22.189	8540	1066	
s,info	IT:MM	png	164.87.170.73	2045	1903	
s,info	AR:ES	html	222.23.102.238	1801	1133	
s,info	IN:DZ	html	138.226.66.81	7029	1801	

LESSONS LEARNT

Puppet is great for managing Elasticsearch

Small budget = creative solutions

Bulk indexing can be slow

Test for split-brain (before it happens)

Modern OS & kernel helps

Java GC is painful

Kernel keep-alive timeouts < routers timeouts

RAM is cheap, buy lots of it

LINKS & REFERENCES

[Elasticsearch Pre-Flight Checklist](#)

[Elasticsearch JVM Settings Explained](#)

[On Elasticsearch Performance](#)

[Managing Elasticsearch Cluster Restart Time](#)

[Slow Cluster Restarts](#)

[Nagios Elasticsearch Plugin](#)

[Puppet Elasticsearch Module](#)

[Puppet Logstash Module](#)

[Grok Debugger](#)

[Linux TCP/IP Tuning](#)

[The Network Is Reliable](#)

THE DEV JOURNEY

It "worked" in dev :/



ISS2



Home | About | Add Your Service | Usage Policy | Copyright | Bookmark | Contact

Custom indexing (not just full text search!)

Stemming (sort of)

Keywords/Synonyms

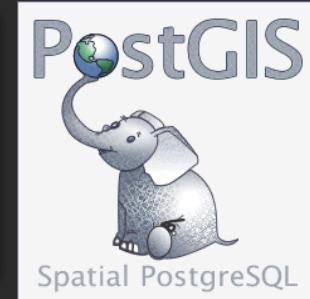
Area search (not quite GIS)

© Infoxchange Australia 2012. This work is copyright. Apart from any use permitted under the Copyright Act 1968, all other rights are reserved.

Created and hosted by



ISS3



mozilla

(ElasticUtils)

... and Tasty Pie (which has no logo)

THE HARD PARTS

GIS searches

Private data

Query strings

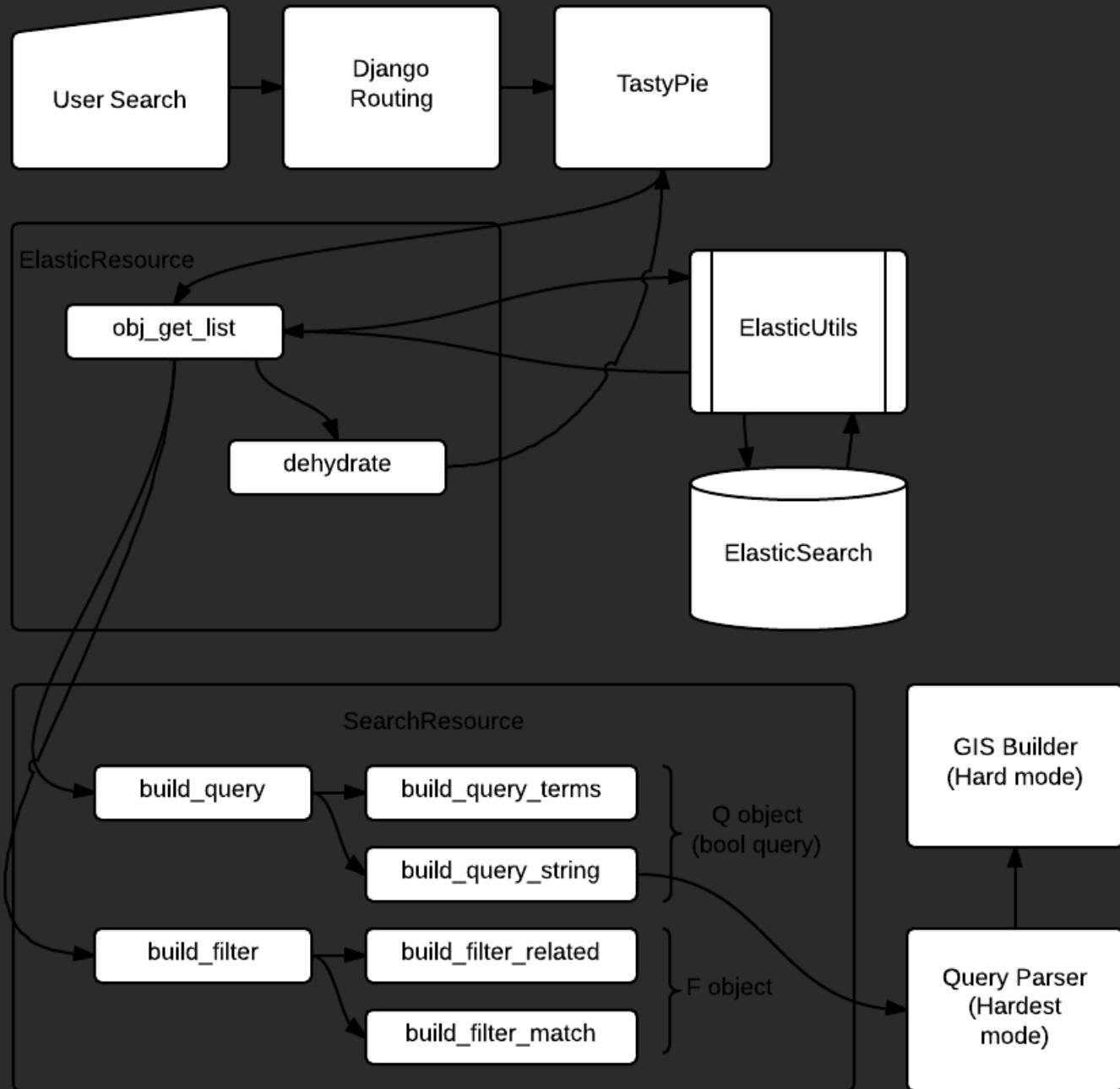
Keywords (I KNOW RIGHT?!)

Suggestions

Cascaded data (slow indexing)

ELASTICPIE

... for consuming, not eating





words, Suburb or Postcode in NSW

100 results

GEOSPATIAL SEARCH

ne Pharmacy, St
5339

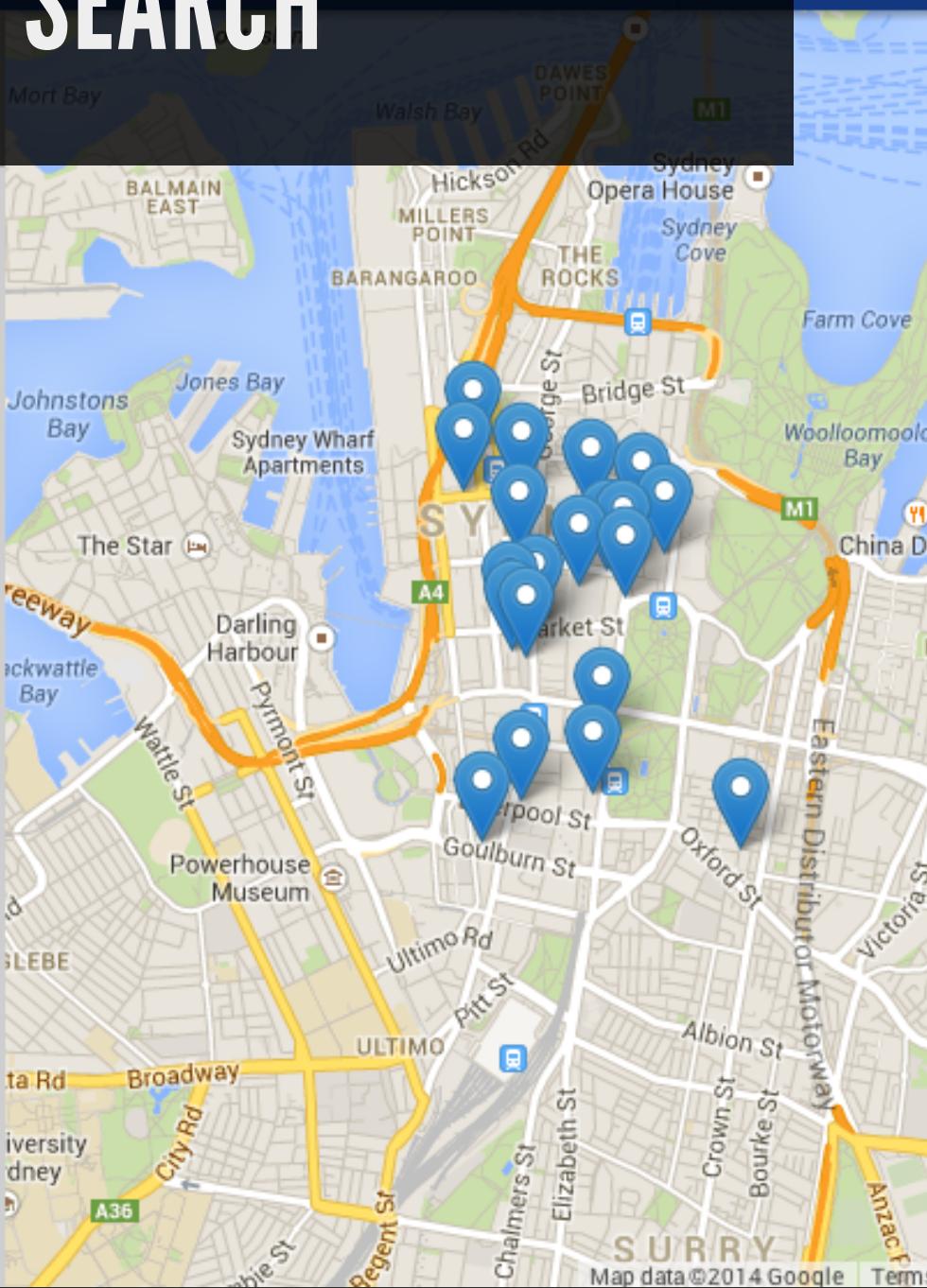
Pharmaceutical services include provision of prescription medicines and over-the-counter health products.

[More information >](#)ne Pharmacy, King
8650

Pharmaceutical services include provision of prescription medicines and over-the-counter health products. Specific services include National Diabetes Services Scheme (NDSS).

[More information >](#)rmacy
1805

Pharmaceutical services include provision of prescription



GIS AND IDF LIKE OIL AND WATER

$$idf(t) = 1 + \log(\frac{numDocs}{docFreq + 1})$$

Use function score! Ignores IDF and query normalization completely

INTELLIGENT QUERY STRINGS

doctor in richmond

doctor near richmond

doctor richmond

doctor near richmond nsw

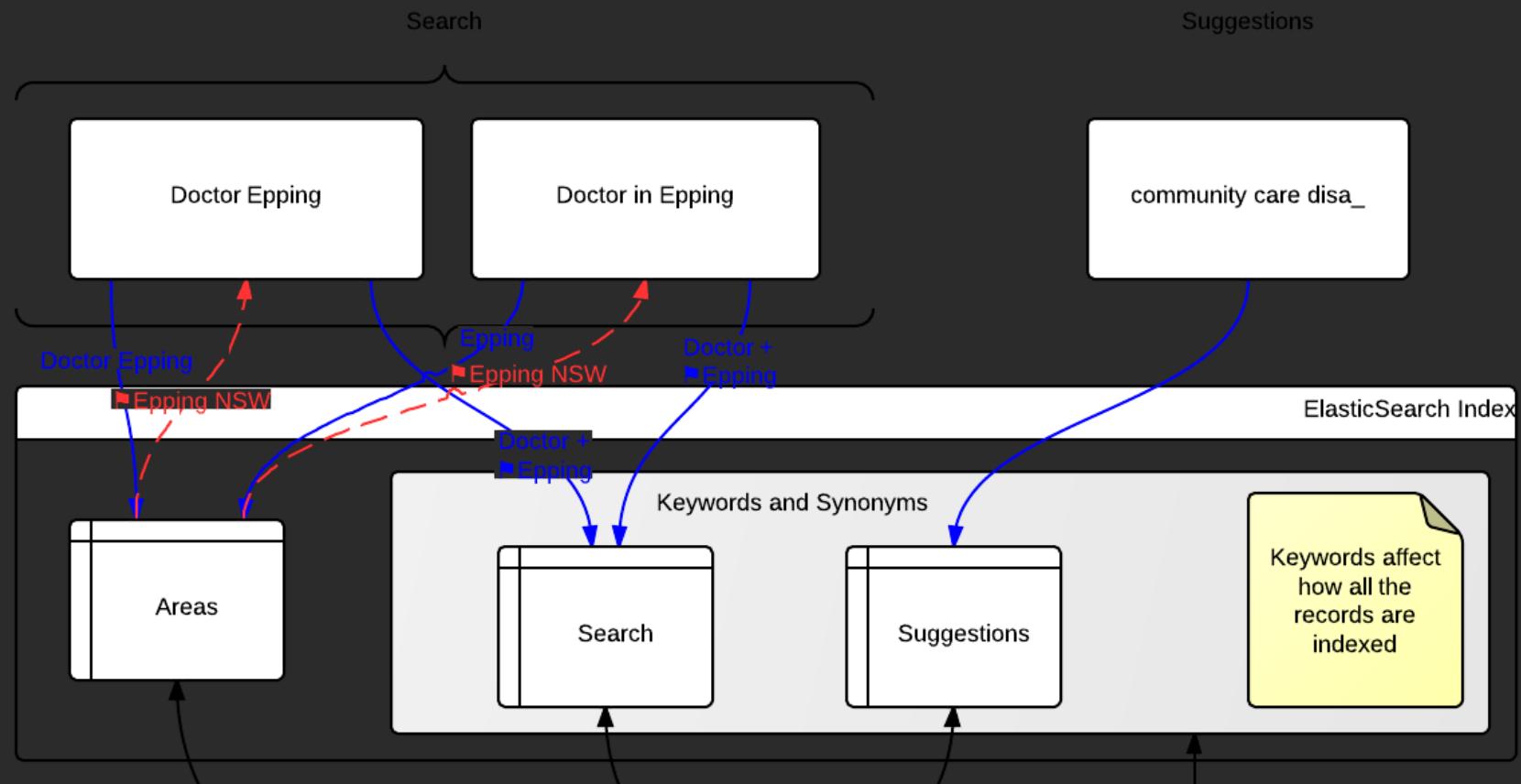
doctor near richmond new south wales

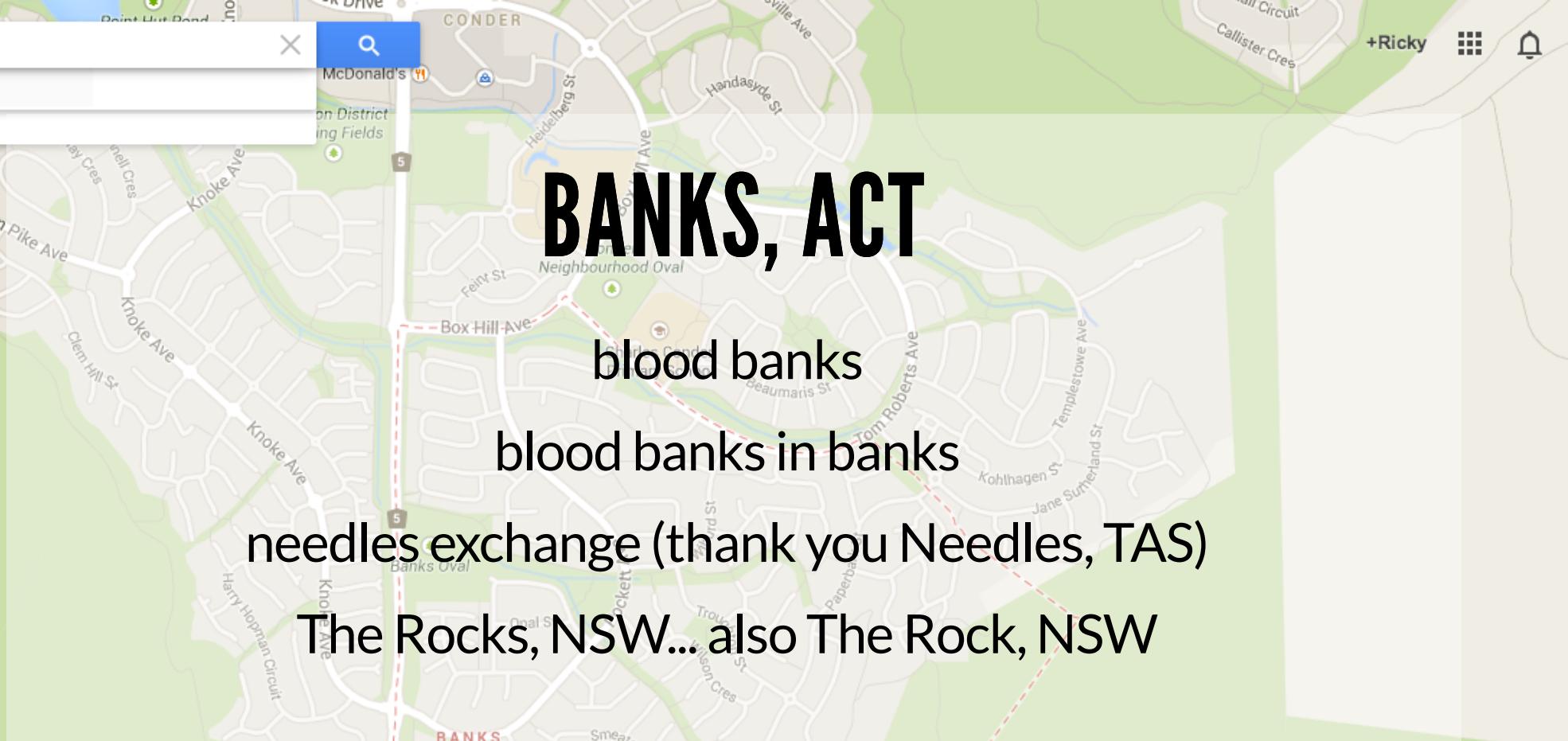
doctor near richmond town hall

doctor around richmond town hall that speaks chinese

female doctor around richmond qld that speaks chinese

ISS 4 DUMMIES

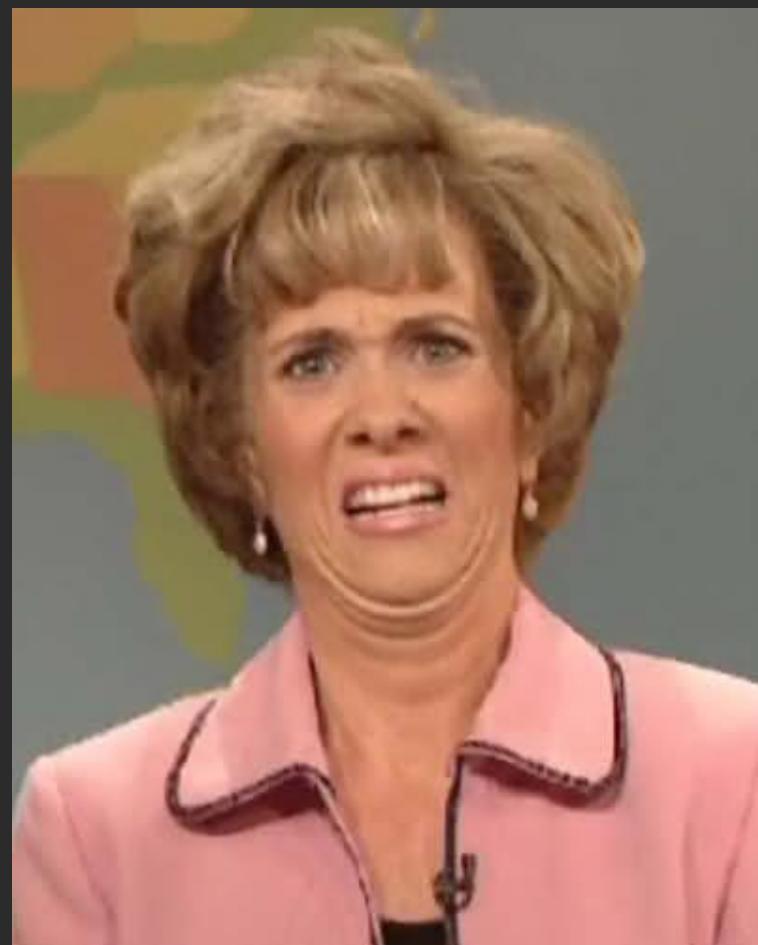




Google

KEYWORDS

How is this even a problem?



KEYWORDS

New South Wales

Allied Health Services

Primary school

Alternative therapy

(Ab)use Synonym filters

```
"type": "synonym",
"synonyms": [
    "new south wales => new__south__wales,nsw",
    "primari school => primari school,primary__school",
    "altern therapi => altern therapi,alternative__therapy"
]
```

SUGGESTIONS

s my ip
not to wear
s my ip address
are these strawberries doing on my nipples i need them for the fruit salad
to expect when you're expecting
to do when your bored
s love
time is it
s twitter
hurts the most lyrics

Completion suggester

✗ NOPE

200,000,000 results

43,900,000 results

51,800,000 results

,000,000 for what. (0.12 seconds)

13,400 results

11,000,000 results

16,600,000 results

391,000,000 results

601,000,000 results

620,000,000 results

553,000 results

close

Term suggester

✗ NOPE

Phrase suggester

✓ YES

Just kidding, ✗ NOPE

[unes/emotion.htm](#) - [Cached](#) - [Similar](#)

- always on **WHAT.com**

ot, top, now, popular and **what** on YouTube, iTunes, Google

ore - always on **WHAT.com**.

at?

- 3 Apr 2008 -

1.09 : I will do a new updated version once the show has finished

0. THE ULTIMATE **WHAT**. Hold me to it. Update 01.29.09 ...

[e.com/watch?v=GcatQSyRK6c](#) - [Related videos](#)

FUN WITH MARKOV CHAINS

You can read it later...

meals on whee

meals on wheels

meals on wheelchairs

meals on Wheeling

meals on Wherrett

meals on Wheelabout

meals on Wheatley

meals on White's

meals on Wyee

meals on Weethalle

meals on Wreck

Community Hall, Chatsworth

 (02) 6646 4338

A community hall.

[More information >](#)

Fernvale Hall

 Fernvale (02) 6677 9304

A community hall available for hire.

[More information >](#)

Norah Head Hall

 Ourimbah (02) 4350 5538

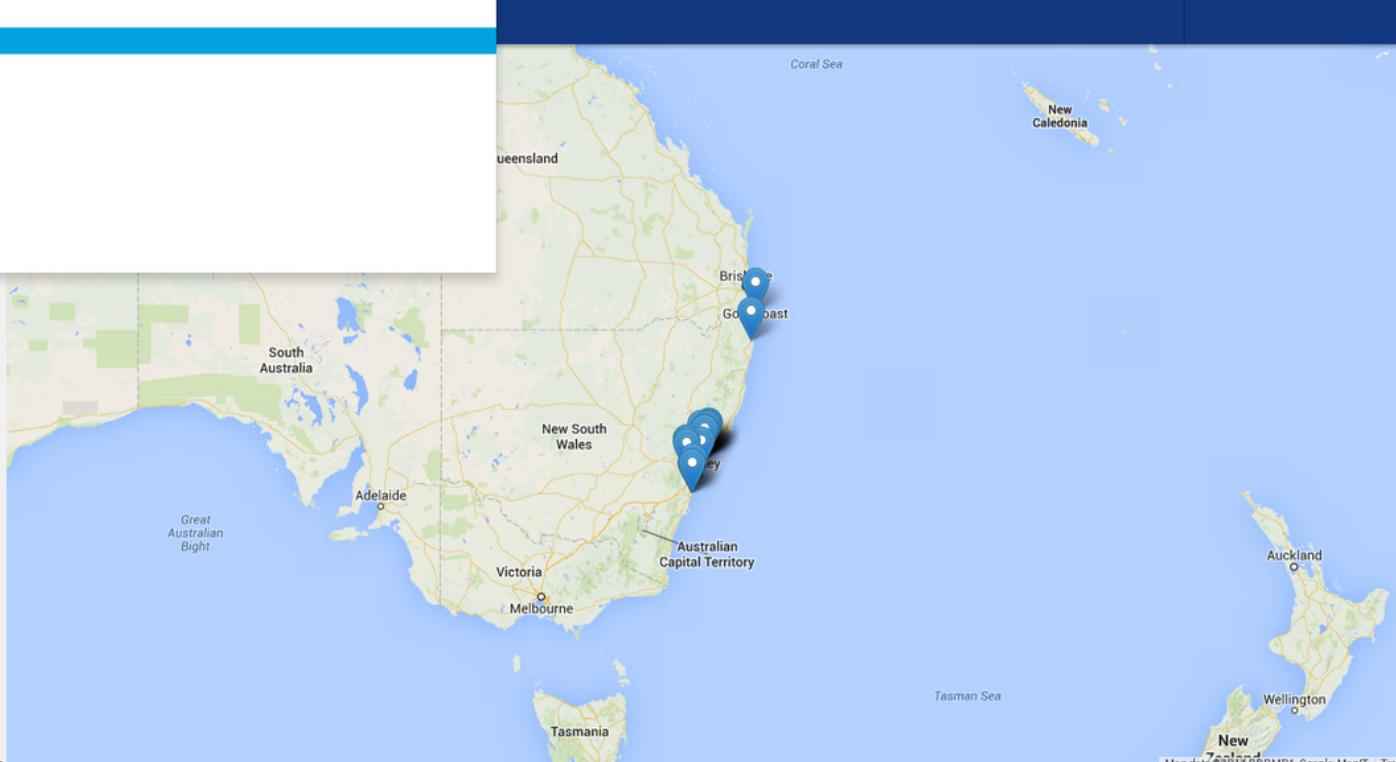
A public hall available for hire to private and community groups.

[More information >](#)

Berkeley Community Hall

 Berkeley (02) 4251 7777

A community hall & meeting room



alcoholics kings cross

[Find services](#)
 Advanced Search

alcoholics kings cross
 alcoholics kings crows
 alcoholics kings crossdressing
 alcoholics kings Crossley
 alcoholics kings Crossman
 alcoholics kings crossroads
 alcoholics kings Croats
 alcoholics kings Crossdale
 alcoholics kings crossways
 alcoholics kings Crosslands

[1300 222 222](#)
[More information >](#)

Support for Alcoholics

[Darlinghurst](#)

 Alcoholics Anonymous (AA),
 Darlinghurst
 (02) 9387 7788
 Now open

Alcoholics Anonymous is a fellowship of men and women who share their experiences in order to help solve their common problems and assist recovery from alcoholism.

[More information >](#)

Support for Alcoholics

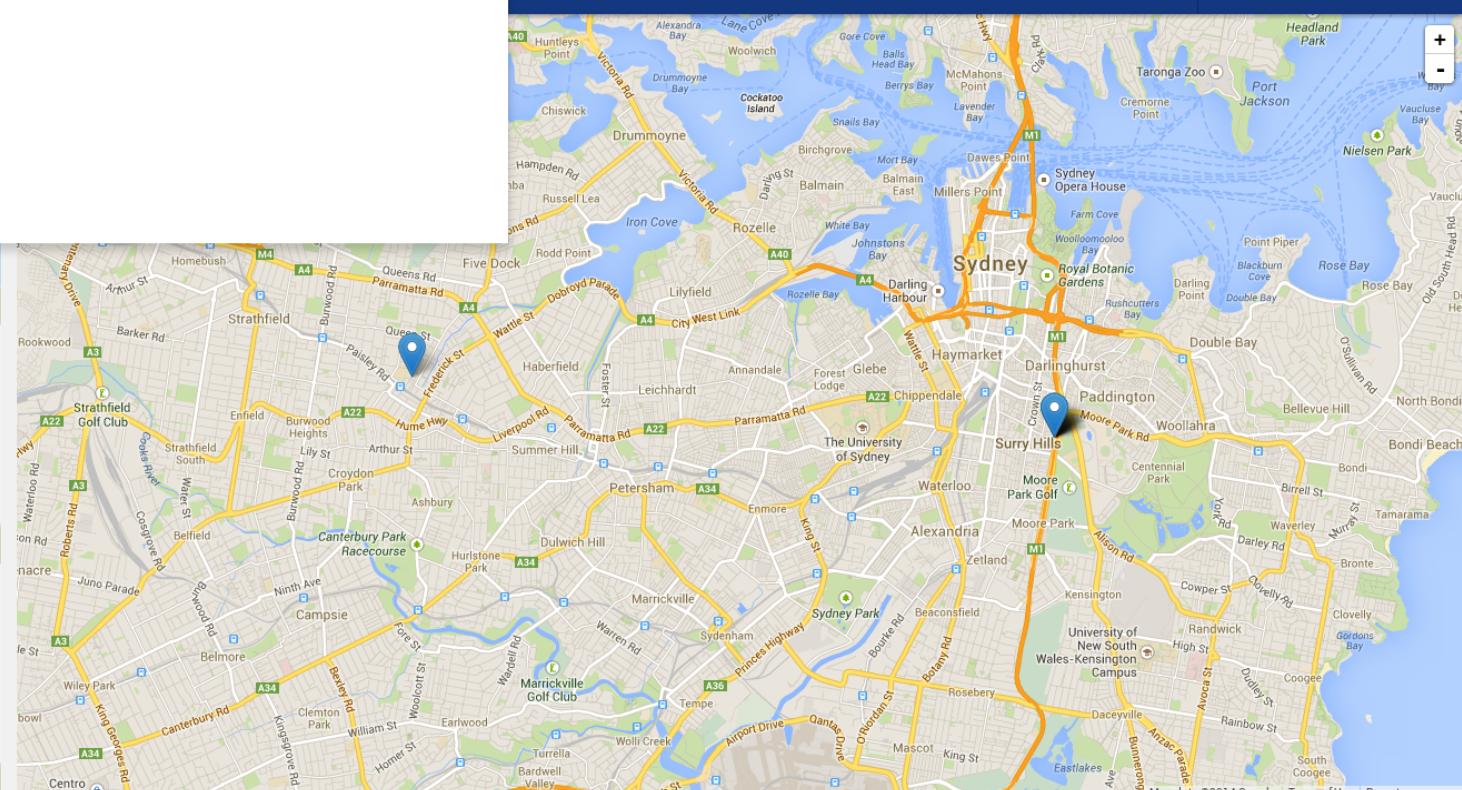
[Woolloomooloo](#)

 Alcoholics Anonymous (AA),
 Woolloomooloo
 (02) 9387 7788
 Now open

Alcoholics Anonymous is a fellowship of men and women who share their experiences in order to help solve their common problems and assist recovery from alcoholism.

[More information >](#)

Support for Alcoholics

[Potts Point](#)
[Contact us](#) | [Privacy policy](#) | [Disclaimer](#)


CELERY: DISTRIBUTED TASKS

aka how we "fixed"
our indexer

(and addressed fault-tolerance)

THANK YOU

Sam McLeod - [@s_mcleod](#)
Ricky Cook - [@ThatPandaDev](#)

smcleod.net: [smcleod.net](#)

Infoxchange website: [infoxchange.org](#)

LINKS & REFERENCES

Dev Links

something



[HTTPS://SMCLEOD.NET](https://smcLeod.net)