Application Development using Elasticsearch

—

By Jurgens du Toit / jrgns / eagerelk.com

Terror View

- Time series data
- Application database
- Search, tagging and categorization
- www.terrorview.com

S3Browser

- Application data
- Enhancement
- Search
- github.com/jrgns/s3browser

Common Development Tasks

- Interacting with ES
- Fields and Mappings
- Setup and Maintenance
- Taking it into Production
- Other Tools

Interacting With Elasticsearch

- Manually
  Curl
  Browser
  Kibana
  Sense
- Programatically
  elasticsearch-ruby/rails/php/js/py/dsl-py/net/groovy/hadoop/hdfs

Examples

Ruby

```
document = client.get index: ‘myapp’, type: ‘client’, id: 12345
client.index index: ‘myapp’, type: ‘client, id: 12345, body: document
client.delete index: ‘myapp’, type: ‘client’, id: 12345
client.update index: ‘myapp’, type: ‘client’, id: 12345, version: 3, body:
doc
```

PHP

```
$doc = $client->get([‘index’=>‘myapp’,‘type’=>‘client’,‘id’=>12345])
$client->index([‘index’=>‘myapp’,`type’=>‘client`,‘id’=>12345,‘body’=>$doc])
$client->delete([‘index’ => ‘myapp’, ‘type’ => ‘client’, ‘id’ => 12345])
$client->update([‘index’=>‘myapp’,‘type’=>‘client’,‘id’=>12345,
`body`=>$doc])
```

Querying

- Store the query framework in json file
- Read and convert to an internal data structure
- Replace the necessary values
- Send it to the SDK

```
$query = json_decode(file_get_contents('search_product.json'));
$query['query']['bool']['should'][0]['term']['email'] = ' joe@bloggs.com';
$client->search(['index'=>'myapp', 'type' => 'users', 'body' => $query]);
```

- Use a Repository

```
$repo = new ES\Repo;
$repo->search_users($term, $filters);
```

Fields & Mappings

Field Types

Core

- String
- Long
- Integer
- Short
- Byte
- Double
- Float
- Date
- Boolean
- Binary

Complex

- Array
- Object
- Nested
- Geo Point
- Geo Shape
- IPv4
- Completion
- Token Count
- Mapper-murmur3
- Multi Fields

Mappings

- analyzed vs not_analyzed
- Beware Dynamic Mapping
- Beware Cross Type mappings
  `/myapp/client/id = "CLI001"` vs `/myapp/user/id = 5`
- `_timestamp` is deprecated
- `_ttl`, `_parent` and `_routing`
- `_version`

Multi Fields

Definition

```
key: {
  type: :string,
  index: :analyzed,
  analyzer: :filename,
  fields: {
      raw: {
      type: :string,
      index: :not_analyzed
    }
  }
}
```

Search

```
query: {
  bool: {
    must: {
      simple_query_string: {
        fields:[ 'key','key.raw'],
        default_operator: 'OR',
        query: 'search term'
      }
    }
  }
}
```

Custom Analyzers

```
analysis: {
  analyzer: {
    filename: {
      type: ‘custom’,
      char_filter: [  ],
      tokenizer: ‘standard’,
      filter: [
        ‘word_delimiter’,’standard’,’lowercase’,’stop‘
      ]
} } }
```

Setup & Maintenance

- Quick Start
- Vagrant
- Docker?
- Hosted
- Backup & Restore
- Maintenance

Quick Start

1. `wget https://download.elasticsearch.org/elasticsearch/release/org/elasticsearch/distribution/zip/elasticsearch/2.1.1/elasticsearch-2.3.3.zip`
2. `unzip elasticsearch-2.3.3.zip`
3. `cd elasticsearch-2.3.3`
4. `bin/elasticsearch`

Setup using Vagrant

```
Vagrant.configure(2) do |config|
  config.vm.box = "ubuntu/trusty64"

  config.vm.network "forwarded_port", guest: 9200, host: 9200

  config.vm.provider "virtualbox" do |vb|
     vb.memory = "1024"
  end

  config.vm.provision "shell", privileged: false, inline: <<-SHELL
     # Repositories
     wget -qO - https://packages.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -
     echo "deb http://packages.elastic.co/elasticsearch/2.x/debian stable main" | sudo tee -a
/etc/apt/sources.list.d/elasticsearch-2.x.list
     sudo add-apt-repository ppa:webupd8team/java
     sudo apt-get update
     sudo apt-get upgrade -y
```

Setup using Vagrant

```
    # Java
    echo debconf shared/accepted-oracle-license-v1-1 select true | sudo debconf-set-selections
    echo debconf shared/accepted-oracle-license-v1-1 seen true | sudo debconf-set-selections
    sudo apt-get install -y oracle-java8-installer

    # Elasticsearch
    sudo apt-get install -y elasticsearch
    sudo service elasticsearch stop
    echo "ES_HEAP_SIZE=512m" | sudo tee -a /etc/default/elasticsearch

    # Services
    sudo service elasticsearch restart
  SHELL
end
```

Hosted Options

- Found / Elastic Cloud
- AWS Elasticsearch
- Google Cloud
- Bonsai
- QBox
- SearchBox
- FacetFlow
- Logz.io
- Papertrail
- Others?

Backup & Restore

Repository

```
PUT /_snapshot/backups {
“type”: “fs”,
“settings”: { .. }
}
```

Backup

```
PUT /_snapshot/backups/b-160525
```

Validate

```
GET /_snapshot/backups/b-160525
```

Restore

```
POST /_snapshot/backups/b-
160525/_restore
{
  "indices": "index_1,index_2"
}
```

Maintenance

```
#!/bin/bash
echo 'Creating a backup'
curl -X PUT localhost:9200/_snapshot/s3_backups/`date +%Y.%m.%d`?wait_for_completion=true
echo 'Remove old backup'
/usr/local/bin/curator --host localhost delete snapshots --older-than 30 --time-unit days
--timestring "%Y.%m.%d" --repository s3_backups

echo 'Remove old marvel indices'
/usr/local/bin/curator --host localhost delete indices --older-than 14 --time-unit days --
timestring "%Y.%m.%d" --prefix .marvel

echo 'Cleaning out old indices'
/usr/local/bin/curator --host localhost delete indices --older-than 30 --time-unit days --
timestring "%Y.%m.%d"
```

Taking it into Production

- Sizing
- Monitoring
- Health and Recovery

Sizing

RAM / ES_HEAP_SIZE

ES_HEAP_SIZE = RAM / 2 (Mostly)

For logging

Hot Capacity = CHS * 8

Application

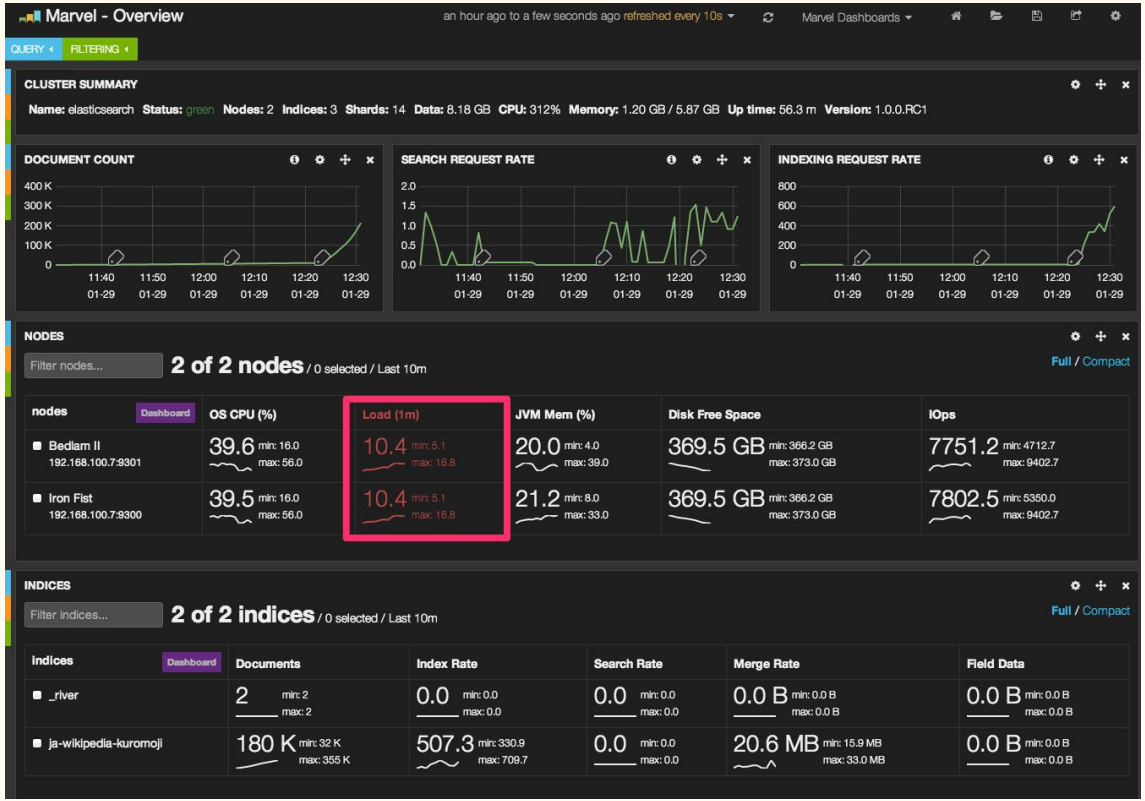
Hot Capacity = CHS * 2

CPU and Storage

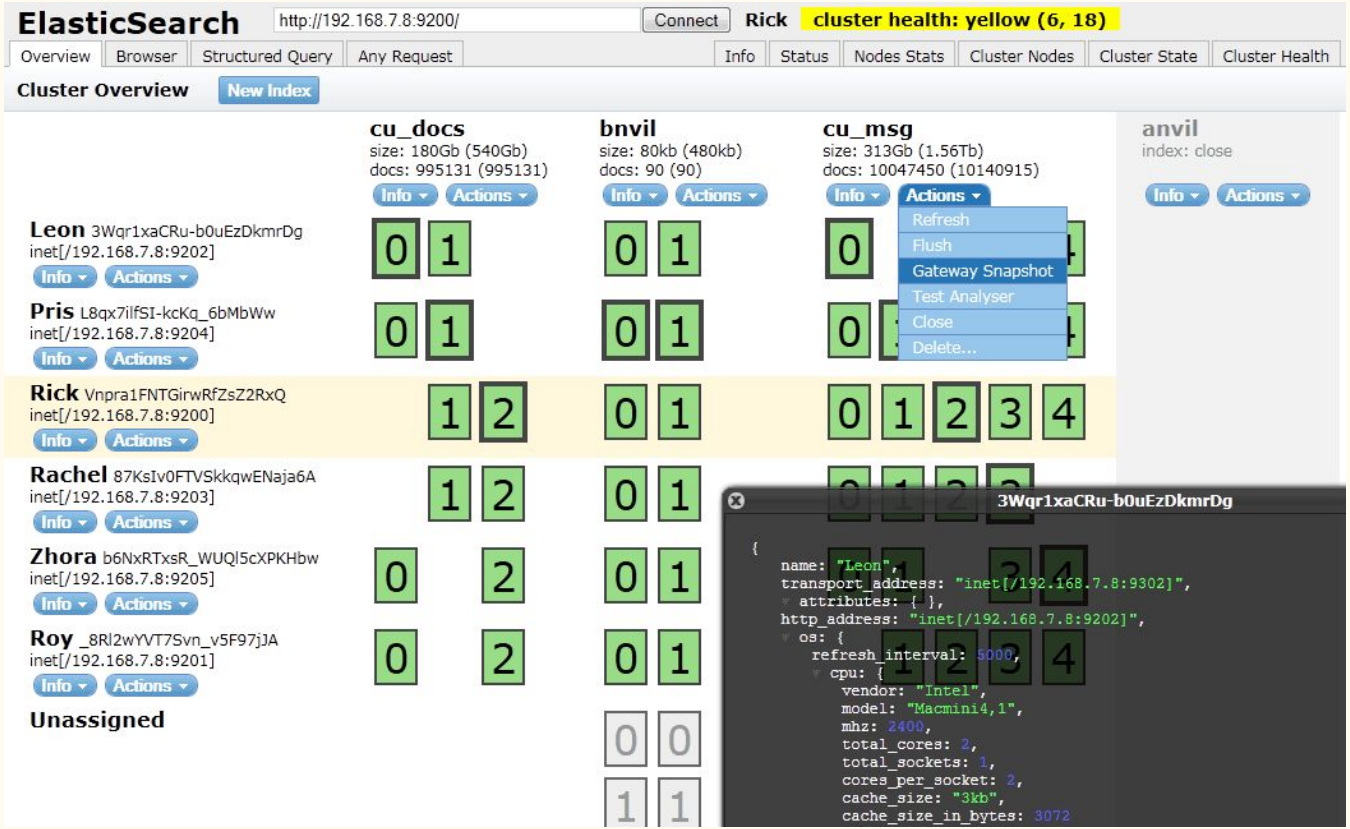
Index vs Search

Long term storage vs Replication

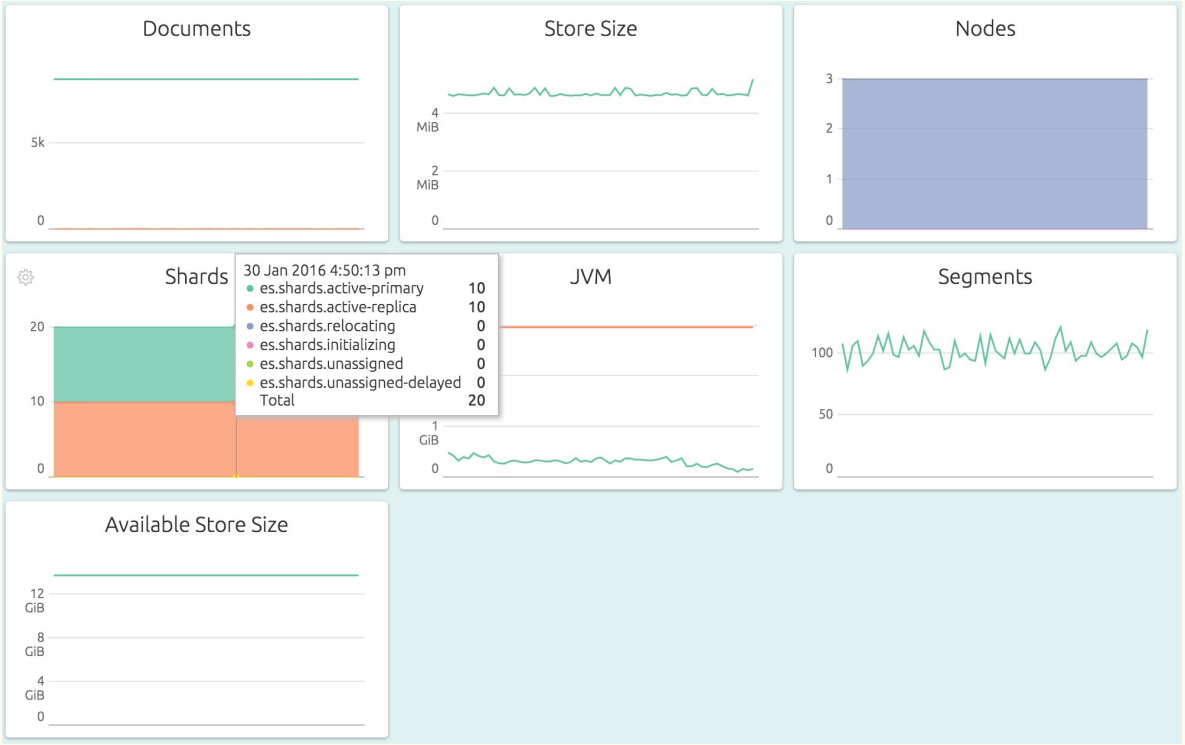
Monitoring - Marvel



Monitoring - Head



Monitoring - OpsDash



Health and Recovery

Argh!

```
GET /_cluster/health?pretty
{
  "cluster_name" : "myescluster",
  "status" : "red",
  "number_of_nodes" : 20,
  "number_of_data_nodes" : 16,
  "active_primary_shards" : 2558,
  "active_shards" : 5628,
  "relocating_shards" : 0,
  "initializing_shards" : 4,
  "unassigned_shards" : 22
}
```

Fix it!

https://t37.net/how-to-fix-your-elasticsearch-cluster-stuck-in-initializing-shards-mode.html

- Identify the stuck shards
- Reassign them to a healthy node
- Restart the node with initializing
- Tread carefully and hope for the best

Other Tools

- curator
- Beats
- Logstash
- Kibana
- geronime/es-reindex

Questions?

@jrgns

jrgns@jrgns.net

eagerelk.com