ElasticSearch

Colin Bradford
cjbradford@gmail.com
What is ElasticSearch

- Full text search engine
- Document based, schemaless storage
- Uses Lucene Java libraries
- REST/HTTP interface
- Supports clustering out of the box
- Perl interface: ElasticSearch.pm
Installing

tar xvfz elasticsearch-0.17.6.tar.gz
cd elasticsearch-0.17.6
./bin/elasticsearch -f
Configuring ES cluster

- Configuration is in the config/ directory
- If the servers can see each other using multicast, no configuration required
ElasticSearch concepts

• Documents contain fields
• Documents have a type
• Documents are stored in indexes
• Mapping defines how documents are analysed and stored
  • Mappings can be generated automatically
Using ElasticSearch from Perl

- Can list more than one server
- Module will load list of servers in the cluster, and use any of them

```perl
use ElasticSearch;
my $es = ElasticSearch->new(
    servers => 'localhost:9200',
);
```
Inserting data

$es->index(
    index => 'mythtv',
    type => 'program',
    id => 'fp.bbc.co.uk/5a72ov',
    data => {
        callsign => "BBC ONE",
        category => "News",
        channum => 1,
        description => "Detailed weather forecast",
        endtime => "2011-07-29 00:40:00",
        programid => "fp.bbc.co.uk/5a72ov",
        starttime => "2011-07-29 00:35:00",
        title => "Holiday Weatherview",
    }
);
Mappings

- ElasticSearch generates this mapping:

```json
{
    "program" : {
        "properties" : {
            "endtime" : { "type" : "string" },
            "starttime" : { "type" : "string" },
            "programid" : { "type" : "string" },
            "callsign" : { "type" : "string" },
            "category" : { "type" : "string" },
            "channum" : { "type" : "string" },
            "title" : { "type" : "string" },
            "description" : { "type" : "string" }
        }
    }
}
```
Override the mapping

"program": {
    "properties": {
        "starttime": {
            "type": "date",
            "format": "YYYY-MM-dd HH:mm:ss"
        },
        "callsign": {
            "type": "string",
            "analyzer": "keyword"
        },
        "category": {
            "type": "multi_field",
            "fields": {
                "category": {
                    "type": "string"
                },
                "category_term": {
                    "type": "string",
                    "index": "not_analyzed"
                }
            }
        }
    }
}
Fetching an individual document

my $data = $es->get(
    index => 'mythtv',
    id    => $get_id,
);

returns

{
    _id      => "fp.bbc.co.uk/5a72ov",
    _index   => "mythtv",
    _source  => {
        callsign    => "BBC 1 Oxford",
        category    => "News",
        description => "Detailed weather forecast",
        endtime     => "2011-07-29 00:40:00",
        programid   => "fp.bbc.co.uk/5a72ov",
        starttime   => "2011-07-29 00:35:00",
        title       => "Holiday Weatherview",
    },
    _type    => "program",
    _version => 5,
    exists   => bless(do{\(my $o = 1\)}, "JSON::XS::Boolean"),
}
Searching for documents

```perl
my $data = $es->search(
    index   => 'mythtv',
    size    => 5,
    query   => {
        query_string => {
            query => "Doctor Who",
        },
    },
);
Results of a search

{  
  _shards => { failed => 0, successful => 5, total => 5 },
  hits => {
    hits => [
      {
        _id     => "fp.bbc.co.uk/sq6pm",
        _index => "mythtv",
        _score => "1.3216872",
        _source => {
          callsign    => "CBeebies",
          category    => "Kids",
          description => "Featuring Opera Oliver - ...<snip>",
          endtime     => "2011-08-13 07:35:00",
          programid   => "fp.bbc.co.uk/sq6pm",
          starttime   => "2011-08-13 07:20:00",
          title       => "Gigglebiz",
        }
      },
      {
        _type   => "program",
      },
      ( ... snip 4 more hits ... )
    ],
    max_score => "1.3216872",
    total => 1052,
  },
  timed_out => bless(do{\(my $o = 0)}, "JSON::XS::Boolean"),
  took => 11,
}
Got 1052 hits
1 (score 1.32): CBeebies: Gigglebiz :Featuring Opera Oliver ...
2 (score 1.09): BBC THREE: Doctor Who: Sci-fi drama. The Doctor ...
3 (score 1.09): BBC THREE: :Doctor Who: Science fiction drama. The...
4 (score 1.02): BBC THREE: Doctor Who: Science fiction drama. When ...
5 (score 1.02): BBC HD: The Indian Doctor: Drama about an Indian ...
Searching with highlights

```perl
my $data = $es->search(
    index => 'mythtv',
    size => 5,
    query => {
        query_string => {
            query => "Doctor Who",
        },
    },
    highlight => {
        fields => {
            title => {},
            description => {},
        },
    },
);
Highlighted results

```json
{
  _id   => "fp.bbc.co.uk/sq6pm",
  _index => "mythtv",
  _score => "1.0836935",
  _source => {
    callsign => "CBeebies",
    category => "Kids",
    description => "Featuring Opera Oliver <snip>",
    endtime => "2011-08-13 07:35:00",
    programid => "fp.bbc.co.uk/sq6pm",
    starttime => "2011-08-13 07:20:00",
    title => "Gigglebiz",
  },
  _type => "program",
  highlight => {
    description => [
      "Featuring Opera Oliver - the restaurant owner <em>who</em> just loves to sing about his food, and <em>Doctor</em> <em>Doctor</em>, <em>who</em> prescribes fun and giggles."
    ],
  },
}
```
1 (score 1.08): CBeebies: Gigglebiz: Featuring Opera Oliver - the restaurant owner who just loves to sing about his food, and Doctor, who prescribes fun and giggles.

2 (score 1.02): BBC THREE: Doctor Who: Science fiction drama. When he follows a Time Lord distress signal, the Doctor puts Amy, Rory and his beloved TARDIS in grave danger.

3 (score 1.02): BBC HD: The Indian Doctor: Drama about an Indian doctor who moves to a Welsh mining village in 1963. The doctor's friendship

4 (score 1.02): BBC THREE: Doctor Who: Science fiction drama. The Doctor, Amy and Rory must solve the mystery of the disappearance of a pirate crew at the hands of a beautiful Siren.

5 (score 1.00): BBC THREE: Doctor Who: Sci-fi drama. The Doctor and his friends are reunited in Utah, and are enlisted by President Nixon to save a terrified girl from a mysterious spaceman.
Asking for an explanation

```json
index => 'mythtv',
size => 5,
explain => 1,
query => {
    query_string => {
        query => "Doctor Who",
    },
},
highlight => {
    fields => {
        title => {},
        description => {},
    },
},
```
Explained results

```json

{  
  _explanation => {  
    description => "sum of:",  
    details => [  
      {  
        description => "weight(_all:doctor in 262), product of:",  
        details => [  
          {  
            description => "queryWeight(_all:doctor), product of:",  
            details => [  
              {  
                description => "idf(_all: doctor=15)", value => "0.0089674" },  
              {  
                description => "queryNorm", value => "0.3441825" }  
            ]  
          }  
        ]  
      },  
      {  
        description => "fieldWeight(_all:doctor in 262), product of:",  
        details => [  
          {  
            description => "btq, product of:",  
            details => [  
              {  
                description => "tf(phraseFreq=1.0)", value => 1 },  
              {  
                description => "allPayload(...)", value => 1 },  
              {  
                description => "idf(_all: doctor=15)", value => "0.0089674" },  
              {  
                description => "fieldNorm(field=_all, doc=262)", value => "0.15625" }  
            ]  
          }  
        ]  
      }  
    ]  
  }  

  _id          => "fp.bbc.co.uk/sq6pm",  
  _index       => "mythtv",  
  _node        => "UIX1sqZ5RTOhB4y2erH45g",  
  _score       => "1.0836935",  
  _shard       => 1,  
  _source      => {  
    callsign    => "CBeebies",  
    category    => "Kids",  
    description => "Featuring Opera Oliver - the restaurant owner who just loves to sing about his food, and Doctor Doctor, who prescribes fun and giggles.",  
    endtime     => "2011-08-13 07:35:00",  
    programid   => "fp.bbc.co.uk/sq6pm",  
    starttime   => "2011-08-13 07:20:00",  
    title       => "Gigglebiz",  
    }  

  _type        => "program",  
  highlight    => [  
    {  
      description => "Featuring Opera Oliver - the restaurant owner <em>who</em> just loves to sing about his food, and <em>Doctor</em> <em>Doctor</em>, <em>who</em> prescribes fun and giggles.",  
    }  
  ]  
}
```
my $data = $es->search(
    index   => 'mythtv',
    size    => 5,
    query   => {
        query_string => {
            query   => "Doctor Who",
            fields  => [ 'title^3', 'description' ],
            use_dis_max => 1,
        },
    },
);
Better results

Got 1052 hits

1 (score 6.68): BBC THREE: Doctor Who: Science fiction drama. When he follows a Time Lord distress signal, the Doctor puts Amy, Rory and his beloved TARDIS in grave danger.

2 (score 6.50): BBC THREE: Doctor Who: Sci-fi drama. The Doctor and his friends are reunited in Utah, and are enlisted by President Nixon to save a terrified girl from a mysterious spaceman.

3 (score 6.50): BBC THREE: Doctor Who: Science fiction drama. The Doctor mounts a rebellion against an alien invasion dating back to the very beginnings of human civilisation.

4 (score 6.08): BBC THREE: Doctor Who: Science fiction drama. The last hope for all of reality is a little girl who still believes in stars.

5 (score 6.08): BBC THREE: Doctor Who: Science fiction drama. The Doctor, Amy and Rory must solve the mystery of the disappearance of a pirate crew at the hands of a beautiful Siren.
my $data = $es->search(
    index   => 'mythtv',
    size    => 5,
    query   => {
        query_string => {
            query   => "Doctor Who",
            fields  => [ 'title^3', 'description' ],
            use_dis_max => 1,
        },
    },
    facets  => {
        callsign_facet => {
            terms   => { field => 'callsign', size => 5, }
        },
        title_facet => {
            terms   => { field => 'title_term', size => 5, }
        },
        category_facet => {
            terms   => { field => 'category_term', size => 5, }
        },
    },
);
Faceted results

```json
{
  facets => {
    callsign_facet => {
      _type => "terms",
      missing => 0,
      other => 736,
      terms => [
        { count => 113, term => "CBeebies" },
        { count => 88, term => "CITV" },
        { count => 58, term => "Channel 5" },
        { count => 34, term => "Channel 4 HD" },
        { count => 23, term => "BBC 2 England" },
      ],
      total => 1052,
    },
    ( ... snip two more facets ... )
  },
  hits => { ( ... snip hits ... ) },
}
```
my $data = $es->search(
    index => 'mythtv',
    size  => 5,
    query => {
        bool => {
            must => [
                {
                    query_string => {
                        query   => "Doctor Who",
                        fields  => [ 'title^3', 'description' ],
                        use_dis_max => 1,
                    },
                },
                { term => { callsign => "BBC THREE", } },
            ],
        },
    },
    facets => {
        title_facet => { terms => { field => 'title_term', size => 5, } }
    },
);
Results

Got 22 hits

1 (score 7.36): BBC THREE: Doctor Who: Science fiction drama. When he follows a Time Lord distress signal, ...

2 (score 7.23): BBC THREE: Doctor Who: Sci-fi drama. The Doctor and his friends are reunited in Utah, and are enlisted by President Nixon ...

3 (score 7.23): BBC THREE: Doctor Who: Science fiction drama. The Doctor mounts a rebellion against an alien invasion dating back to the ...

4 (score 7.19): BBC THREE: Doctor Who: Science fiction drama. The last hope for all of reality is a little girl who still believes in stars.

5 (score 7.19): BBC THREE: Doctor Who: Science fiction drama. The Doctor, Amy and Rory must solve the mystery of the disappearance of a pirate ...

Title facet:
Doctor Who (5)
American Dad! (4)
Russell Howard's Good News (3)
Family Guy (3)
Young, Rich and House Hunting (2)
Further information

- Active mailing list
- Regular releases
- http://www.elasticsearch.org/