



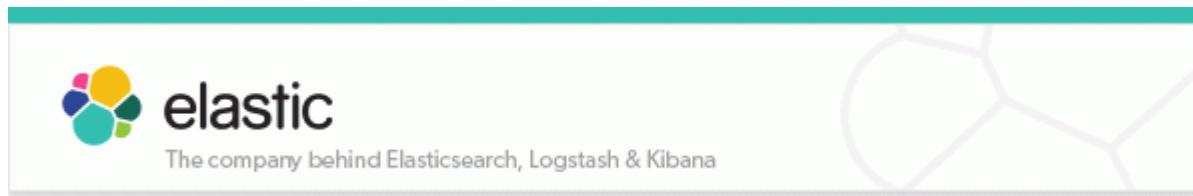
ALL ABOUT ELASTICSEARCH

Jelle Vink
Software Architect
SugarCRM
@skymeyer

2015
SUGARCON
CONNECTING i2i

What is Elasticsearch ?

- What is Elasticsearch ?
 - Open source full text search engine based on Apache Lucene
 - Fast and highly scalable
 - Low maintenance
 - REST API



What is Elasticsearch ?

Elasticsearch is much more than just Lucene and much more than “just” full-text search:

- A distributed real-time document store where every field is indexed and searchable
- A distributed search engine with real-time analytics
- Capable of scaling to hundreds of servers and petabytes of structured and unstructured data

SugarCRM integration

Fields

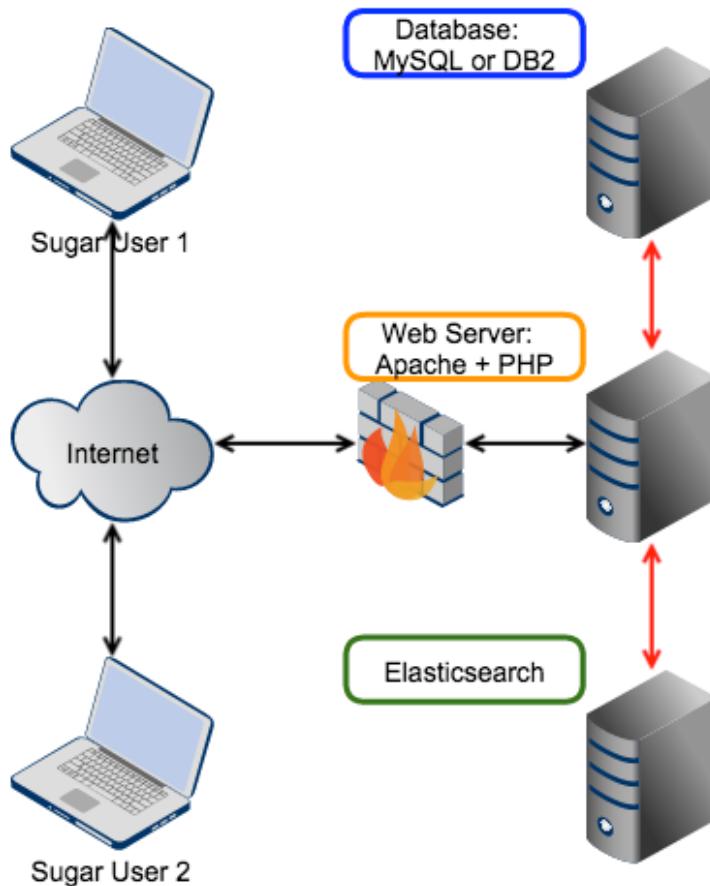
Index

Search

- Modules and fields are configured for full text search in vardefs.
- Studio has basic controls.
- SugarBean::save() triggers the indexing of a record
- Records are pushed inline or using a queue to Elasticsearch
- The global search page leverages the indexed records for search.
- Customizations and future extensions will make more use of the index.

Deploying Elasticsearch

Topology



- Elasticsearch needs to be placed in a secured network.
- No direct access should be allowed.
- All communication with ES are flowing through SugarCRM.

Deploying Elasticsearch

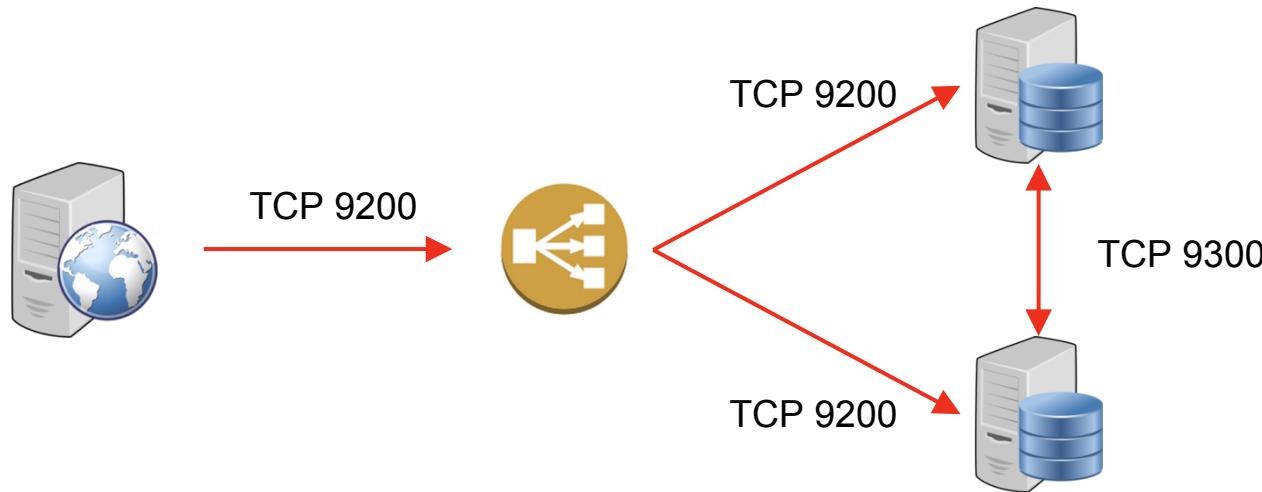
Sizing

- Disk space: minimum 2x database size
 - Mostly depends on the amount of fields being indexed
- Memory:
 - 2 - 64 GB
 - Keep 50% reserved for system/file cache
 - Keep Java heap size below 32 GB
- Elasticsearch scales easily by adding nodes, starting with one or two node is possible.

Deploying Elasticsearch

Communication

- TCP 9200 for HTTP REST API
- TCP 9300 is used for inter node communication
- SugarCRM allows one connection setting, use additional load balancer for multi node setup.



Deploying Elasticsearch

Installation prerequisites

- Linux, OSX and Windows are supported.
- Requires latest stable Java, see <https://www.elastic.co/subscriptions/matrix>

	Oracle JRE 1.6	Oracle JVM 1.7u55+	Oracle JVM 1.8u20+	IcedTea OpenJDK 1.6	IcedTea OpenJDK 1.7.0.55+	IBM JVM 6	IBM JVM 7	Azul Zing
ES.1.0.x	✗	✓	✗	✗	✓	✗	✗	✗
ES.1.1.x	✗	✓	✗	✗	✓	✗	✗	✗
ES.1.2.x	✗	✓	✗	✗	✓	✗	✗	✗
ES.1.3.x	✗	✓	✓	✗	✓	✗	✗	✗
ES.1.4.x	✗	✓	✓	✗	✓	✗	✗	✗
ES.1.5.x	✗	✓	✓	✗	✓	✗	✗	✗

Deploying Elasticsearch

Verify Installation

- Test connectivity after install:

```
–curl -XGET http://localhost:9200/
```

```
{  
    status: 200,  
    name: "tauriel",  
    cluster_name: "skymeyer",  
    version: {  
        number: "1.5.0",  
        build_hash: "544816042d40151d3ce4ba4f95399d7860dc2e92",  
        build_timestamp: "2015-03-23T14:30:58Z",  
        build_snapshot: false,  
        lucene_version: "4.10.4"  
    },  
    tagline: "You Know, for Search"  
}
```

Deploying Elasticsearch

SugarCRM Connection settings

- Basic connection settings through UI

The screenshot shows a configuration dialog titled "Full Text Search Settings". It has three input fields: "System Type" set to "ElasticSearch", "Host" set to "tauriel.sugar", and "Port" set to "9200". A "Test" button is located to the right of the host field. At the bottom are three buttons: "Save" (blue), "Schedule System Index" (blue), and "Cancel" (gray).

Full Text Search Settings	
System Type:	ElasticSearch
Host:	tauriel.sugar
Port:	9200

Save Schedule System Index Cancel

- Advanced configuration available using config_override.php supporting:
 - Transport mode
 - Authentication
 - Any cURL parameters as needed

Deploying Elasticsearch

SugarCRM advanced connection settings

- Use `config_override.php` for

```
$sugar_config['full_text_engine']['Elastic']
```

- `['transport'] = 'Http' or 'Https'`

- `['curl']` to pass in any cURL parameter

- Example configuration using TLS and basic HTTP authentication:

```
$sugar_config['full_text_engine']['Elastic']['host'] = 'lb01.edge.local';
$sugar_config['full_text_engine']['Elastic']['port'] = '8888';
$sugar_config['full_text_engine']['Elastic']['transport'] = 'Https';
$sugar_config['full_text_engine']['Elastic']['curl'][CURLOPT_USERPWD] = 'user:password';
$sugar_config['full_text_engine']['Elastic']['curl'][CURLOPT_SSL_VERIFYPEER] = true;
$sugar_config['full_text_engine']['Elastic']['curl'][CURLOPT_SSL_VERIFYHOST] = 2;
$sugar_config['full_text_engine']['Elastic']['curl'][CURLOPT_CAPATH] = '/etc/ssl/trusted';
```

Deploying Elasticsearch

Elasticsearch recommended configuration

- Default out of the box is not production ready. The following config options are recommended for elasticsearch.yml:
 - index.mapper.dynamic: false
 - action.destructive_requires_name : true
 - action.auto_create_index : false
 - script.disable_dynamic: true
 - rest.action.multi.allow_explicit_index: false
 - http.jsonp.enable: false
 - discovery.zen.ping.multicast.enabled: false

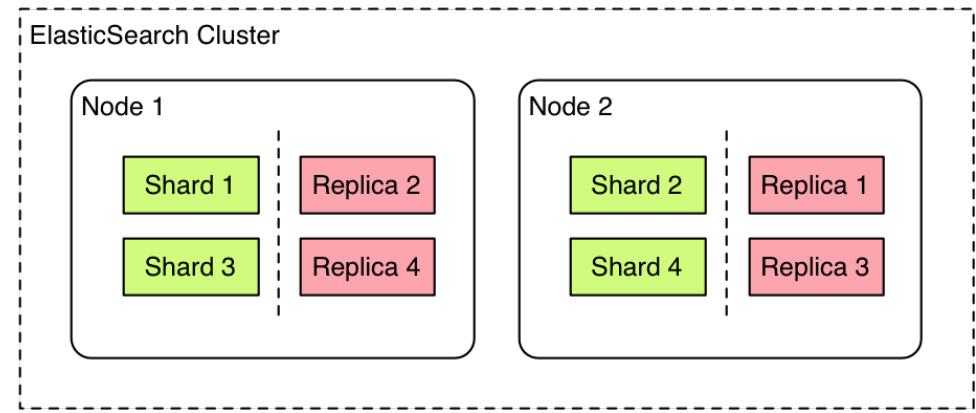
Deploying Elasticsearch

Tuning

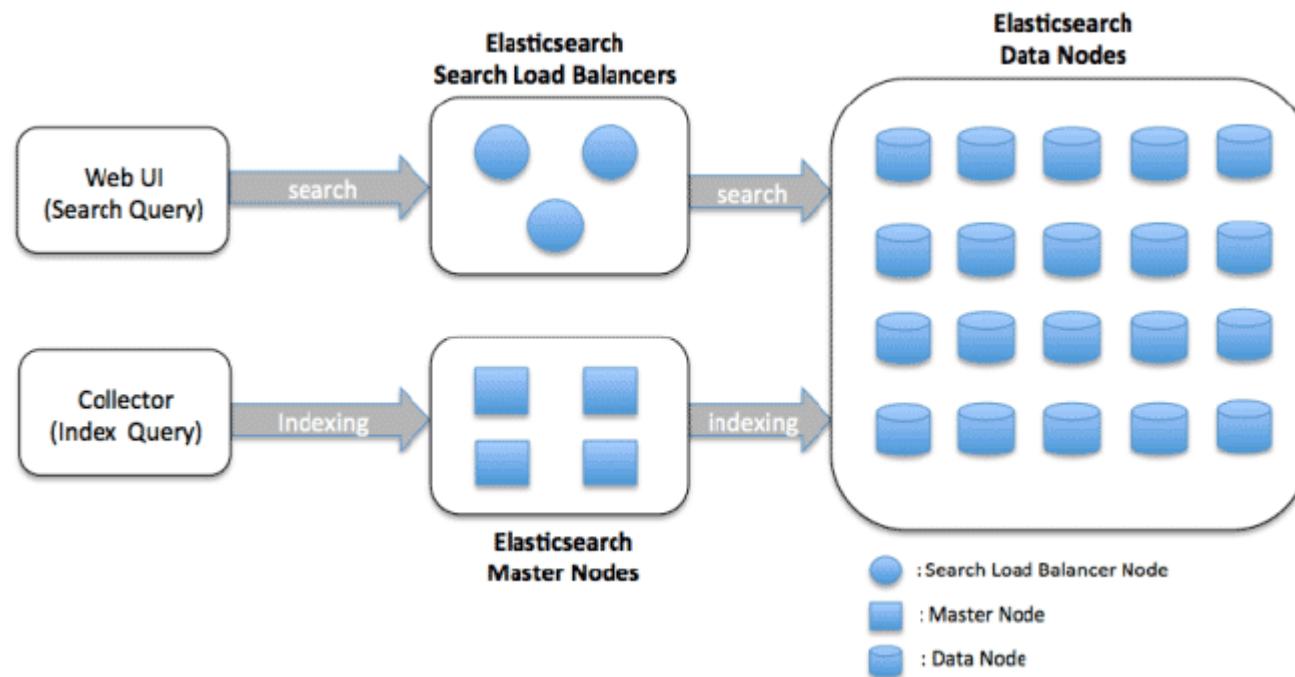
- Elasticsearch start script:
 - Set ES_HEAP_SIZE to half of system RAM
 - MAX_OPEN_FILES=65535
 - MAX_LOCKED_MEMORY=unlimited
 - MAX_MAP_COUNT=262144
- elasticsearch.yml
 - bootstrap.mlockall: true

Elasticsearch components

RDBMS	Elasticsearch
schema	mapping
database	index
table	type
row/record	document
column	field



Elasticsearch components advanced setup



Elasticsearch

Shards and replicas

- SugarCRM uses one index by default.
 - Based on "unique_id"
- Shards and replicas can be configured in `elasticsearch.yml` directly, but its better to configure this through SugarCRM.
- Amount of shards cannot be changed per index once set.
- Replicas can be added/removed on the fly.

Elasticsearch

Index settings

- \$sugar_config['full_text_search']['Elastic']
 - ['index_strategy']
 - ['Contacts']
 - ['index'] => 'contactsindex'
 - ['strategy'] => 'static'

Elasticsearch

Index settings

- \$sugar_config['full_text_search']['Elastic']
 - ['index_settings']
 - ['default']
 - ['index.number_of_shards'] => 3
 - ['index.number_of_replicas'] => 2
 - ['contactsindex']
 - ['index.number_of_shards'] => 2

Elasticsearch

Mapping

- Running full reindex with "drop data" option recreates the mapping.

```
{  
    "state": "open",  
    "settings": {  
        "index": {  
            "creation_date": "1429517813062",  
            "mapping": {  
                "coerce": "true",  
                "ignore_malformed": "true"  
            },  
            "uuid": "_cucmdVWTPeqckwigAMK7A",  
            "number_of_replicas": "0",  
            "analysis": {  
                "char_filter": {  
                    "gs_char_num_pattern": {  
                        "type": "pattern_replace",  
                        "pattern": "[\\d\\s]+",  
                        "replacement": ""  
                    }  
                },  
                "analyzer": {  
                    "gs_analyzer_phone_ngram": {  
                        "type": "custom",  
                        "char_filter": [  
                            "gs_char_num_pattern"  
                        ],  
                        "filter": [  
                            "gs_filter_ngram_3_15"  
                        ],  
                        "tokenizer": "whitespace"  
                    }  
                }  
            }  
        }  
    }  
}
```

```
    "mappings": {  
        "Accounts": {  
            "dynamic": "false",  
            "include_in_all": false,  
            "_source": {  
                "excludes": [  
                    "Accounts_email_search"  
                ]  
            },  
            "properties": {  
                "date_entered": {  
                    "include_in_all": false,  
                    "index": "not_analyzed",  
                    "type": "string",  
                    "fields": {  
                        "gs_datetime": {  
                            "format": "YYYY-MM-dd HH:mm:ss",  
                            "type": "date"  
                        }  
                    }  
                },  
                "team_set_id": {  
                    "include_in_all": false,  
                    "index": "not_analyzed",  
                    "type": "string"  
                },  
                "sic_code": {  
                    "include_in_all": false,  
                    "index": "not_analyzed",  
                    "type": "string",  
                    "copy_to": [  
                        "Accounts_sic_code"  
                    ]  
                }  
            }  
        }  
    }  
}
```

Elasticsearch

Sending documents

- After reindex `fts_queue` is populated with beans to be indexed.
- Running `cron.php` processes the queued records
- `After_save` logic hooks send beans inline to ES.
- Fallback to `fts_queue` when ES is unavailable.
- Cron is always required to process records from the `fts_queue`.

Elasticsearch

Debugging tools

- Elasticsearch "head" plugin
- Using cURL directly
- Using intermediate reverse proxy with logging
- In SugarCRM 7.7 new logging functionality is integrated through sugarcrm.log

Vardef settings

- Enable module
 - 'full_text_search' => true
- Enable fields
 - 'full_text_search' =>
 - 'enabled' => true
 - 'type' => '...' (optional type override)
 - 'searchable' => true/false (*)

(*) New in 7.7

Mappings

- Index settings are used to register (custom) analyzers (tokenizers, token/char filters)
- Mapping (schema) is required for proper search functionality.
- Sugar types are mapping to ES types
 - SugarSearchEngineElasticMapping
 - Using "providers" (new in 7.7)



Q&A

@skymeyer

www.sugarcrm.com

2015
SUGARCON
CONNECTING i2i