Coderwall   Tips   **Sign In or Up**

Last Updated: October 04, 2017   ·   jwwest

Setting up ElasticSearch with MongoDB

MONGODB   ELASTICSEARCH

I've recently set up an instance of ElasticSearch indexing data from MongoDB. I thought it would be helpful to others to put my notes online since I spent hours collating instructions from various sites.

Generally speaking, the ElasticSearch documentation sucks. It assumes a level of familiarity with search indexing out of the box that the average dev doesn't have. Setting up ES or Solr or Sphinx is very much a system administration chore, and configuring it is a discipline on its own.

The Environment

I used EC2 running an Unbuntu instance, so my notes are pretty specific to that environment. Wherever possible, I used the built-in package manager (apt/aptitude) and service manager (Upstart).

MongoDB

If you're not starting from scratch, you can skip this section as it pertains to installing Mongo.

To get Mongo up and running on Ubuntu, you will need to follow the excellent directions on 10gen's site.

You're going to be adding another upstream for apt, then pulling down and installing the mongod service into Upstart. Don't worry, apt takes care of doing all of this for you.

Don't try to skip this step. While the default Ubuntu apt package library contains Mongo, it's a really old version. I've had this bite me before.

Configure Mongo as a replica set

Even if you're running a standalone Mongo instance, you're going to need to convert it into a replica set. The reason is that the ElasticSearch plugin depends on the operation log (or 'oplog', a log of all changes used by Mongo to replicate itself) to push new updates into ElasticSearch. Mongo lacks built-in support for triggers, so this is the next best thing.

Again, follow the instructions here to convert your instance. However, ignore the line about starting it with the `-replSet` command line. Since we're running as a service, edit `/etc/mongod.conf` to include the `replSet` parameter. It doesn't matter what you name your replication set, just that you remember it and are consistent with it.

You can find more information about the Mongo configuration file here.

Installing ElasticSearch

I found this gist to be invaluable in getting the lastest version of ES up and running on my system. It'll first install a Java runtime, followed by ES and then a simple service wrapper.

After you've installed ES, and confirmed it's working by sending an HTTP request to `localhost:9200`, you have to install two plugins to enable support for MongoDB.

The first is a dependency called Mapper Attachments. You can install via the ES plugin script:

```
$ES_HOME/bin/plugin -install elasticsearch/elasticsearch-mapper-
attachments/1.6.0
```

$ES_HOME is wherever ES is installed. If you followed the instructions above, it'll be in `/usr/local/share/elasticsearch`.

The second plugin is the ES 'river' for Mongo. The syntax to install it is slightly different as it's a third-party plugin:

```
plugin -url
https://github.com/downloads/richardwilly98/elasticsearch-river-
mongodb/elasticsearch-river-mongodb-1.6.1.zip -install river-
mongodb
```

Once those two plugins are successfully installed, you should be ready to go. Restart the ES service before proceeding to configuration just in case.

Configuring ElasticSearch

I don't want to go too deep into setting up your search index here. It's a very rich topic with lots of stuff specific to your data. The default analyzers that ES uses are pretty good anyway, so let's look at how you drive this thing.

You create and manage indexes in ES via a RESTful interface. The easiest way I've found is using `curl` inside of bash scripts to send JSON payloads which will instruct ES on how to look at our data.

Example:

```sh
#!/bin/sh
curl -XPUT "localhost:9200/_river/ffxi/_meta" -d '
{
  "type": "mongodb",
  "mongodb": {
    "servers": [
      { "host": "127.0.0.1", "port": 27017 }
    ],
    "options": { "secondary_read_preference": true },
    "db": "ffxi",
    "collection": "pages"
  },
  "index": {
    "name": "pagesidx",
    "type": "page"
  }
}'
```

◄   ►

Notice that the url contains `_river`. This is a small difference you'll notice when using 'river' plugins, that is non-native support for databases. This format is only for setting up and configuring an index however. When you query, it'll look more like:

```
curl -XGET 'localhost:9200/pagesidx/_search?q=param1
```

Written by James West

♡ **Recommend**

🐦 **Say Thanks**

🔕 **Update Notifications Off**

💬 **Respond**

9 Responses

ocnvn
Hi, thanks for this tutorial, just one question which version of every software are you using?

over 1 year ago · ♡

maziyar_sh
Hi, thanks for the tutorial. Since you are using Elasticsearch, how is that different from its built-in full-text search MongoDB has?

over 1 year ago · ♡

richardwilly98
Hi,

MongoDB is a datastore not initially designed to do full-text search. This feature is just been released recently (still in beta). I don't believe you should expect this feature matching Elasticsearch.

over 1 year ago · ♡

richardwilly98

Hi,

I would suggest to check the version matrix available here [1] to make sure you are using the correct version of the river.

[1] - https://github.com/richardwilly98/elasticsearch-river-mongodb#mongodb-river-plugin-for-elasticsearch

over 1 year ago · ♡

mjlevitt

There are no date stamps on this article nor the comments. Can someone please tell me, when was this posted? ES info seems to be a moving, emerging topic, and I want to know how current this is.

over 1 year ago · ♡

mattjanssen

When was this article published?

over 1 year ago · ♡

rkroondotnet

It is now the 7th of July, 2014. The article's version numbers are now somewhat out of date. The working versions I have found are as follows:

MongoDB v2.4.9 (current installation of mongodb, this was the driving force in choosing the others)
ElasticSearch v1.1.1
ElasticSearchMapperAttachments v2.0.0
ElasticSearchRiverMongoDb/2.0.0

over 1 year ago · ♡

ansup

I am trying this tutorial , but in my data folder in Elastic search, No indices are created but a _river folder is created. When i am querying the indices there is no output.

Any help ?

over 1 year ago · ♡

Mohamed Ramzan

Hi this works perfectly for elasticsearch version 1.X. Why don't you update your plugin to support elastic version 5.X.
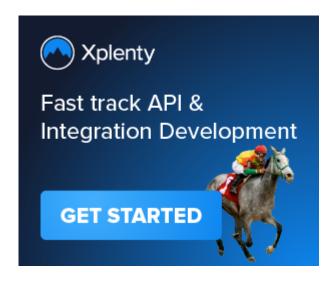
12 months ago · ♡

👍 **Sponsors**

💎 **Awesome Job**   **See All Jobs →**

Web and Multimedia Developer
EUMETSAT · Darmstadt · Full Time

Post a job for only $299

@coderwall 🐦   🐙 **Contact Privacy Terms** Copyright 2018