



链滴

ES 与 MongoDB 的搭配

作者: [cactus0509](#)

原文链接: <https://ld246.com/article/1470215490067>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

<p>装之前, 请参考 https://github.com/richardwilly98/elasticsearch-river-mongodb 根据你的 MongoDB 版本号决需要的 elasticsearch 版本号和插件号。 </p>

<p>1)安装 ES</p>

<p>下载 ElasticSearch_版本号.tar.gz,官网上有, 下载好之后。 </p>

<p>tar -zvx elasticsearch-1.1.0.tar.gz

cd elasticsearch-1.1.0</p>

<p>安装一下插件, 也可以不安装, 这个插件用来监控用的</p>

<p>./bin/plugin -i elasticsearch/marvel/latest</p>

<p>想了解这个插件可以参考官方文档</p>

<p></p><div class="vditor-linkcard vditor-tooltipped vditor-tooltipped__n" aria-label="http://www.elasticsearch.org/guide/en/marvel/current/index.html" >

Free and Open Search: The Creators of Elasticsearch, ELK & Kibana | Elastic

We're the creators of the Elastic (ELK) Stack -- Elasticsearch, Kibana, Beats, and Logstash. Securely and reliably search, analyze, and visualize our data in the cloud or on-prem.

www.elasticsearch.org

</div><p></p>

<p>2) 执行程序</p>

<p>./elasticsearch</p>

<p>看到以下的就表示成功了</p>

<p>[2014-04-09 10:12:41,414][INFO][node] [Lorna Dane] version[1.1.0], pid[839] build[2181e11/2014-03-25T15:59:51Z]

[2014-04-09 10:12:41,415][INFO][node] [Lorna Dane] initializing ...

[2014-04-09 10:12:41,431][INFO][plugins] [Lorna Dane] loaded [], sites []

[2014-04-09 10:12:44,383][INFO][node] [Lorna Dane] initialized

[2014-04-09 10:12:44,384][INFO][node] [Lorna Dane] starting ...

[2014-04-09 10:12:44,495][INFO][transport] [Lorna Dane] bound_address {inet[/0:0:0:0:0:0:9300]}, publish_address {inet[/XXXXXX:9300]}

[2014-04-09 10:12:47,522][INFO][cluster.service] [Lorna Dane] new_master [Lorna Dane [MI-gTu_ZTniHR2mkpbMQ_A][XXXXXX][inet[/XXXXXX:9300]], reason: zen-disco-join (elected_a_master)

[2014-04-09 10:12:47,545][INFO][discovery] [Lorna Dane] elasticsearch/MI-gTu_ZTniHR2mkpbMQ_A

[2014-04-09 10:12:47,572][INFO][http] [Lorna Dane] bound_address {inet[/0:0:0:0:0:0:9200]}, publish_address {inet[/XXXXXX:9200]}

[2014-04-09 10:12:47,607][INFO][gateway] [Lorna Dane] recovered [0] indices into cluster_state

[2014-04-09 10:12:47,607][INFO][node] [Lorna Dane] started</p>

<p>如果想后台运行, 则执行</p>

<p>./elasticsearch -d</p>

<p>想确认程序是否运行，则运行</p>

<p>lsof -i:9200

lsof -i:9300

一个是节点对外服务端口，一个是节点间交互端口（如果有集群的话）。</p>

<p>3) 建立集群</p>

<p>配置文件路径是：</p>

<p>..... (你的实际路径)/config/elasticsearch.yml</p>

<p>默认是全部配置项都屏蔽的，</p>

<p>我修改后配置项如下：</p>

<p>cluster.name: ctoes ---配置集群的名字

node.name: "QiangZiGeGe"---配置节点的名字，注意有双引号</p>

<p>bootstrap.mlockall: true</p>

<p>没有提到的配置项都采用默认值，具体参数如何设置，还需要具体情况具体分析。</p>

<p>修改好后，启动 es,可以看到打印的消息里有别的节点名字，就表示建立集群成功。</p>

<p>注意：es 是自动探测局域网内的同名集群节点的。</p>

<p>查看集群的状态，可以通过：</p>

<p>curl 'http://localhost:9200/_cluster/health?pretty'</p>

<p>响应如下：</p>

<p>{

"cluster_name" : "ctoes",

"status" : "green",

"timed_out" : false,

"number_of_nodes" : 2,

"number_of_data_nodes" : 2,

"active_primary_shards" : 5,

"active_shards" : 10,

"relocating_shards" : 0,

"initializing_shards" : 0,

"unassigned_shards" : 0

}</p>

<p>接下来来使用一下来得到直观感受</p>

<p>4) 使用数据库感受一下</p>

<p>创建索引(相当于创建数据库)</p>

<p>示例如下：</p>

<p>[deployer@XXXXXXXX0013 ~]\$ curl -XPUT 'http://localhost:9200/test1?pretty' -d'</p>

<blockquote>

<p>{

"settings":{

"number_of_shards":2,

"number_of_replicas":1

}

}

'

{

"acknowledged" : true

}</p>

</blockquote>

<p>注意，这里的 number_of_shards 参数是一次性设置，设置之后永远不可以再修改的，但是 number_of_replicas 是可以随后可以修改的。</p>

<p>上面的 url 里的 test1 其实就是建立的索引(数据库)的名字，根据需要自己修改即可。</p>

<p>创建文档</p>

<p>curl -XPUT 'http://localhost:9200/test1/table1/1' -d'

{ "first": "dewmobile",

"last": "technology",


```
"age":3000,<br>
"about":"hello,world",<br>
"interest":["basketball","music"]<br>
}<br>
'<br>
```

响应如下:


```
{"_index":"test1","_type":"table1","_id":"1","_version":1,"created":true}</p>
```

<p>表明创建文档成功</p>

<p>test1:建立的数据库名字</p>

<p>table1:建立的 type 名字, type 与关系数据库的 table 对应</p>

<p>1:自己制定的文档的主键,也可以不指定主键由数据库自己分配。</p>

<p>5) 安装数据库同步插件</p>

<p>由于我们的数据源是放在 MongoDB 中的,所以这里只讲 MongoDB 数据源的数据同步。</p>

<p>插件源码: https://github.com/richardwilly98/elasticsearch-river-mongodb/</p>

<p>MongoDB River Plugin (作者 Richard Louapre)</p>

<p>简介: mongodb 同步插件, mongodb 必须搭成副本集的模式,因为这个插件的原理是通过定读取 mongodb 中的 oplog 来同步数据。</p>

<p>如何安装使用呢? 需要安装 2 个插件</p>

<p>1) 插件 1</p>

<p>./plugin -install elasticsearch/elasticsearch-mapper-attachments/2.0.0</p>

<p>2) 插件 2</p>

<p>./bin/plugin --install com.github.richardwilly98.elasticsearch/elasticsearch-river-mongodb 2.0.0</p>

<p>安装过程如下: </p>

```
<p>./bin/plugin --install com.github.richardwilly98.elasticsearch/elasticsearch-river-mongodb 2.0.0<br>
```

```
-&gt; Installing com.github.richardwilly98.elasticsearch/elasticsearch-river-mongodb/2.0.0...<br>
```

```
Trying <a href="https://ld246.com/forward?goto=http%3A%2F%2Fdownload.elasticsearch.org%2Fcom.github.richardwilly98.elasticsearch%2Felasticsearch-river-mongodb%2Felasticsearch-river-mongodb-2.0.0.zip.." target="_blank" rel="nofollow ugc">http://download.elasticsearch.org/com.github.richardwilly98.elasticsearch/elasticsearch-river-mongodb/elasticsearch-river-mongodb-2.0.0.zip..</a>.<br>
```

```
Trying <a href="https://ld246.com/forward?goto=http%3A%2F%2Fsearch.maven.org%2Fremotecontent%3Ffilepath%3Dcom%2Fgithub%2Frichardwilly98%2Felasticsearch%2Felasticsearch-river-mongodb%2F2.0.0%2Felasticsearch-river-mongodb-2.0.0.zip.." target="_blank" rel="nofollow ugc">http://search.maven.org/remotecontent?filepath=com/github/richardwilly98/elasticsearch/elasticsearch-river-mongodb/2.0.0/elasticsearch-river-mongodb-2.0.0.zip..</a>.<br>
```

```
Trying <a href="https://ld246.com/forward?goto=https%3A%2F%2Foss.sonatype.org%2Fservice%2Flocal%2Frepositories%2Freleases%2Fcontent%2Fcom%2Fgithub%2Frichardwilly98%2Felasticsearch%2Felasticsearch-river-mongodb%2F2.0.0%2Felasticsearch-river-mongodb-2.0.0.zip.." target="_blank" rel="nofollow ugc">https://oss.sonatype.org/service/local/repositories/releases/content/com/github/richardwilly98/elasticsearch/elasticsearch-river-mongodb/2.0.0/elasticsearch-river-mongodb-2.0.0.zip..</a>.<br>
```

```
Downloading .....DONE<br>
```

```
Installed com.github.richardwilly98.elasticsearch/elasticsearch-river-mongodb/2.0.0 into /usr/local/elasticsearch_1.1.0/elasticsearch/elasticsearch-1.1.0/plugins/river-mongodb</p>
```

<p>3) 安装 elasticsearch-MySql 插件</p>

<p>具体请参考: </p>

<p>https://github.com/jprante/elasticsearch-river-jdbc 可以直接下载二进制 jar 包。</p>

<p>https://github.com/jprante/elasticsearch-river-jdbc</p>

<p>4) 安装 mysql 驱动 jar 包 (必须!) </p>

<p>这样, 插件就装好了。</p>

<p>6) 使用插件告知 ES 添加监听数据库任务</p>

<p>模板如下: </p>

<p>curl -XPUT localhost:9200/_river/mongo_resource/_meta -d '

{

"type":"mongodb",

"mongodb":{

"servers":

[{"host":"10.XX.XX.XX","port":"60004"}

],

"db":"zapyia_api",

"collection":"resources"

},

"index":{

"name":"mongotest",

"type":"resources"

}}</p>

<p>如果看到下面的内容表示创建成功</p>

<p>{"_index":"_river","_type":"mongodb","_id":"_meta","_version":1,"created":true}</p>

<p>然后, 数据就导入到了 es 中了, 索引建立成功。</p>

```
<pre><code class="highlight-chroma"><span class="highlight-line"><span class="highlight-cl">
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">如果是导入mysql, 模板如下:
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">[deployer@XXX004 ~]$ curl -XPUT 'localhost:9200/_river/my_jdbc_river/_meta' -d '{
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "type":"jdbc",
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "jdbc":{
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "url":"jdbc:mysql://localhost:3306/fastooth",
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "user":"XXX",
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "password":"
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "sql":"select * base62Decode(display_name) as name from users"
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; }</span></span><span class="highlight-line"><span class="highlight-cl">&gt; }</span></span><span class="highlight-line"><span class="highlight-cl">&gt; ' </span></span><span class="highlight-line"><span class="highlight-cl">&gt; 更详细的是:
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "strategy": "simple",
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "url": null,
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "user": null,
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "password": null,
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "url": null,
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "user": null,
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "password": null,
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "url": null,
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "user": null,
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "password": null,
```

```
</span></span><span class="highlight-line"><span class="highlight-cl">&gt; "url": null,
```

```

</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">
ull,
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">
ull,
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">
: false,
</span></span><span class="highlight-line"><span class="highlight-cl">
0, /* Integer.MIN for MySQL */
</span></span><span class="highlight-line"><span class="highlight-cl">
0,
</span></span><span class="highlight-line"><span class="highlight-cl">
3,
</span></span><span class="highlight-line"><span class="highlight-cl">
wait" : "30s",
</span></span><span class="highlight-line"><span class="highlight-cl">
le.getDefault().toLanguageTag(),
</span></span><span class="highlight-line"><span class="highlight-cl">
",
</span></span><span class="highlight-line"><span class="highlight-cl">
"type" : "jdbc
',
</span></span><span class="highlight-line"><span class="highlight-cl">
"bulk_size" :
00,
</span></span><span class="highlight-line"><span class="highlight-cl">
"max_bulk_r
quests" : 30,
</span></span><span class="highlight-line"><span class="highlight-cl">
"bulk_flush_i
terval" : "5s",
</span></span><span class="highlight-line"><span class="highlight-cl">
"index_settin
s" : null,
</span></span><span class="highlight-line"><span class="highlight-cl">
"type_mappi
g" : null
</span></span><span class="highlight-line"><span class="highlight-cl"> }
</span></span><span class="highlight-line"><span class="highlight-cl">>}
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">对于schedule参数
设置调度时刻的
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">格式参考: http://
www.quartz-scheduler.org/documentation/quartz-1.x/tutorials/crontrigger
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">http://elasticsearc
-users.115913.n3.nabble.com/Ann-JDBC-River-Plugin-for-ElasticSearch-td4019418.html
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">http://www.quartz
scheduler.org/documentation/quartz-1.x/tutorials/crontrigger
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">https://github.co
/jprante/elasticsearch-river-jdbc/issues/186
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">官方文档:
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">http://elasticsearc

```

-users.115913.n3.nabble.com/Ann-JDBC-River-Plugin-for-ElasticSearch-td4019418.html

<https://github.com/jprante/elasticsearch-river-jdbc/wiki/JDBC-River-parameters>

<https://github.com/jprante/elasticsearch-river-jdbc/wiki/Quickstart> (包含如何删除任务)

附录: http://my.o.china.net/wenhaowu/blog/215219#OSC_h2_7

测试过程中, 会出错误:

```
[7]: index [yyyy], type [rrrr], id [1964986], message [RemoteTransportException[[inet[/xxxxxxxx:9300]] [ulk/shard]]; nested: EsRejectedExecutionException[rejected execution (queue capacity 50) on rg.elasticsearch.action.support.replication.TransportShardReplicationOperationAction$AsyncSardOperationAction$1@3e82ee89]; ]
```

修改配置文件, 在后增加:

```
threadpool:
  bulk:
    type: fixed
    size: 60
    queue_size: 000
```

至于这几个参数是什么意思, 还请读者自己去弄明白。

参考:

<http://stackoverflow.com/questions/20683440/elasticsearch-gives-error-about-queue-size>

<http://www.elasticsearch.org/guide/en/elasticsearch/reference/current/modules-threadpool.html>

~~~~~  
~~~

关于客户端, 我们用了Play框架, 正如数据库都需要驱动包一样, 我们从官方网站上看到了这个

<https://github.com>

/cleverage/play2-elasticsearch

```
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">关于中文分词，可
尝试使用Ansj.
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span></code></pre>
<p>关于创建索引: </p>
<p>curl -i -XPUT 'XXX:9200/fasth' -d '<br>
{<br>
"settings" :<br>
{<br>
"number_of_shards" : 3 ,<br>
"number_of_replicas" : 1<br>
}</p>
<p>}<br>
'</p>
<pre><code class="highlight-chroma"><span class="highlight-line"><span class="highlight
cl">
</span></span><span class="highlight-line"><span class="highlight-cl">创建映射
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">curl -i -XPUT 'http
//localhost:9200/fa/users/_mapping' -d '
</span></span><span class="highlight-line"><span class="highlight-cl">{
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">"properties":
</span></span><span class="highlight-line"><span class="highlight-cl"> {
</span></span><span class="highlight-line"><span class="highlight-cl"> "_id":
</span></span><span class="highlight-line"><span class="highlight-cl"> {
</span></span><span class="highlight-line"><span class="highlight-cl"> "type":"string",
</span></span><span class="highlight-line"><span class="highlight-cl"> "index":"not_ana
yzed"
</span></span><span class="highlight-line"><span class="highlight-cl"> },
</span></span><span class="highlight-line"><span class="highlight-cl"> "name":
</span></span><span class="highlight-line"><span class="highlight-cl"> {
</span></span><span class="highlight-line"><span class="highlight-cl"> "type":"string"
</span></span><span class="highlight-line"><span class="highlight-cl"> },
</span></span><span class="highlight-line"><span class="highlight-cl"> "gender":
</span></span><span class="highlight-line"><span class="highlight-cl"> {
</span></span><span class="highlight-line"><span class="highlight-cl"> "type":"string",
</span></span><span class="highlight-line"><span class="highlight-cl"> "index":"not_ana
yzed"
</span></span><span class="highlight-line"><span class="highlight-cl"> },
</span></span><span class="highlight-line"><span class="highlight-cl"> "primary_avatar":
</span></span><span class="highlight-line"><span class="highlight-cl"> {
</span></span><span class="highlight-line"><span class="highlight-cl"> "type":"string",
</span></span><span class="highlight-line"><span class="highlight-cl"> "index":"not_ana
yzed"
</span></span><span class="highlight-line"><span class="highlight-cl"> },
</span></span><span class="highlight-line"><span class="highlight-cl"> "signature":
</span></span><span class="highlight-line"><span class="highlight-cl"> {
</span></span><span class="highlight-line"><span class="highlight-cl"> "type":"string",
</span></span><span class="highlight-line"><span class="highlight-cl"> "index":"not_ana
```


yzed"

```
</span></span><span class="highlight-line"><span class="highlight-cl"> }
</span></span><span class="highlight-line"><span class="highlight-cl"> }
</span></span><span class="highlight-line"><span class="highlight-cl"> }
</span></span><span class="highlight-line"><span class="highlight-cl">}
</span></span><span class="highlight-line"><span class="highlight-cl">'
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl">全量任务:
</span></span><span class="highlight-line"><span class="highlight-cl">curl -XPUT 'xxx:9
00/_river/mysql_users/_meta' -d '
</span></span><span class="highlight-line"><span class="highlight-cl">{
</span></span><span class="highlight-line"><span class="highlight-cl"> "type":"jdbc",
</span></span><span class="highlight-line"><span class="highlight-cl"> "jdbc":
</span></span><span class="highlight-line"><span class="highlight-cl"> {
</span></span><span class="highlight-line"><span class="highlight-cl"> "url":"jdbc:mysql:
/XXX:3306/fastooth",
</span></span><span class="highlight-line"><span class="highlight-cl"> "user":"XXX",
</span></span><span class="highlight-line"><span class="highlight-cl"> "password":"XXX"

</span></span><span class="highlight-line"><span class="highlight-cl"> "sql":"select disti
ct _id,base62Decode(display_name) as name,gender,primary_avatar,signature from users",
</span></span><span class="highlight-line"><span class="highlight-cl"> "index":"XXX",
</span></span><span class="highlight-line"><span class="highlight-cl"> "type":"XXX"
</span></span><span class="highlight-line"><span class="highlight-cl"> }
</span></span><span class="highlight-line"><span class="highlight-cl">}
</span></span><span class="highlight-line"><span class="highlight-cl">'
</span></span><span class="highlight-line"><span class="highlight-cl">
</span></span><span class="highlight-line"><span class="highlight-cl"> http://www.nosql
b.cn/1368777378160.html
</span></span></code></pre>
```