

# 06\_Windows下Solr集群搭建

## 一.Solr简介

### 1.目前最新版的Solr是7.1.0

#### Welcome to Apache Lucene

The Apache Lucene™ project develops open-source search software, including:

- **Lucene Core**, our flagship sub-project, provides Java-based indexing and search technology, as well as spellchecking, hit highlighting and advanced analysis/tokenization capabilities.
- **Solr™** is a high performance search server built using Lucene Core, with XML/HTTP and JSON/Python/Ruby APIs, hit highlighting, faceted search, caching, replication, and a web admin interface.
- **PyLucene** is a Python port of the Core project.

DOWNLOAD

Apache Lucene 7.1.0

DOWNLOAD

Apache Solr 7.1.0

Projects

### 2.系统要求

JDK1.8+

## 二.集群配置实现

### 1.解压Solr

```
~$ unzip -q solr-7.1.0.zip
~$ cd solr-7.1.0/
```

### 2.以SolrCloud模式启动

Unix或MacOS: `bin/solr start -e cloud`

Windows: `bin\solr.cmd start -e cloud`

进入目录-启动【Linux仅仅命令不同】

```
D:\>cd solr-7.1.0
```

```
D:\solr-7.1.0>bin\solr.cmd start -e cloud
```

Welcome to the SolrCloud example!

This interactive session will **help** you launch a SolrCloud **cluster** on your **local** workstation.

To begin, how many Solr nodes would you like to run in your local cluster? (specify 1-4 nodes) [2]:

**这个时候直接回车Enter:**

```
Ok, let's start up 2 Solr nodes for your example SolrCloud cluster.  
Please enter the port for node1 [8983]:
```

**再Enter:**

```
Please enter the port for node2 [7574]:
```

**Enter,后面直接出现这么一大段:**

```
Creating Solr home directory D:\solr-7.1.0\example\cloud\node1\solr  
Cloning D:\solr-7.1.0\example\cloud\node1 into  
D:\solr-7.1.0\example\cloud\node2
```

```
Starting up Solr on port 8983 using command:  
"D:\solr-7.1.0\bin\solr.cmd" start -cloud -p 8983 -s "D:\solr-7.1.0\example\cloud\node1\solr"
```

```
Waiting up to 30 to see Solr running on port 8983  
Started Solr server on port 8983. Happy searching!
```

```
Starting up Solr on port 7574 using command:  
"D:\solr-7.1.0\bin\solr.cmd" start -cloud -p 7574 -s "D:\solr-7.1.0\example\cloud\node2\solr" -z localhost:9983
```

```
Waiting up to 30 to see Solr running on port 7574  
INFO - 2017-11-03 09:37:54.245; org.apache.solr.client.solrj.impl.ZkClientClusterStateProvider; Cluster at localhost:9983 ready
```

Now let's create a new collection for indexing documents in your 2-node cluster.

Please provide a name for your new collection: `[gettingstarted]`  
Started Solr server on port 7574. Happy searching!

## 最后一段话：

系统提示你创建一个用于索引数据的集合。因为Solr提供一些样本数据，我们不妨与教程一致，在控制台输入techproducts 然后回车。

```
techproducts
```

```
How many shards would you like to split techproducts into? [2]
```

然后会让你输入想分割到两个节点上的碎片数量。这里默认为2，就是均匀分割到两个节点上。【我们目前有两个节点，所以2就是平均分。如果我们建立三个节点，那么想要平均分就要输入3】。直接回车。

```
How many replicas per shard would you like to create? [2]
```

你想为每个碎片创建多少副本？【副本是用来故障切换的】这里默认2，回车

Please choose a configuration for the techproducts collection, available options are:

`_default` or `sample_techproducts_configs` [`_default`]

这里让我们选择一个配置。在本地我也找到了这两个目录

这台电脑 > DATA (D:) > solr-7.1.0 > server > solr > configsets				
名称	修改日期	类型	大小	
 <code>_default</code>	2017/11/3 8:33	文件夹		
 <code>sample_techproducts_configs</code>	2017/11/3 8:33	文件夹		

Solr有两个开箱即用的配置文件样本集，称为“configSet”。而索引集合必须有一个“ConfigSet”。它包括两个主要配置文件：

电脑 > DATA (D:) > solr-7.1.0 > server > solr > configsets > _default > conf				▼	🔄	搜索
名称	修改日期	类型	大小			
lang	2017/11/3 8:33	文件夹				
managed-schema	2017/10/13 14:59	文件	50 KB			
params.json	2017/10/13 16:16	JSON 文件	1 KB			
protwords.txt	2017/10/13 16:16	文本文档	1 KB			
solrconfig.xml	2017/10/13 16:16	XML 文件	55 KB			
stopwords.txt	2017/10/13 16:16	文本文档	1 KB			
synonyms.txt	2017/10/13 16:16	文本文档	2 KB			

大家注意 `sampletechproductsconfigs`，它包括了我們定义的索引集合名称：`techproducts`。这个config包含了我們需要的示例数据。所以在控制台上输入 `sample_techproducts_configs` 并回车【可能会慢一点，耐心】：

```
sample_techproducts_configs
Created collection 'techproducts' with 2 shard(s), 2 replica(s) with
config-set
'techproducts'

Enabling auto soft-commits with maxTime 3 secs using the Config API

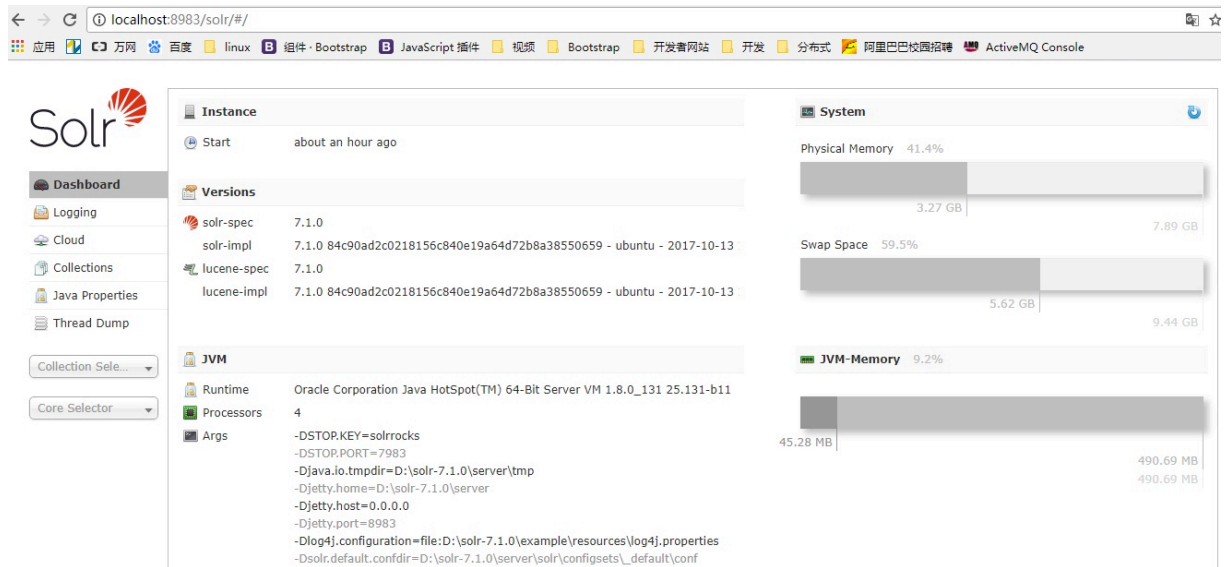
POSTing request to Config API: http://localhost:8983/solr/techproduct
s/config
{"set-property":{"updateHandler.autoSoftCommit.maxTime":"3000"}}
Successfully set-property updateHandler.autoSoftCommit.maxTime to 300
0

SolrCloud example running, please visit: http://localhost:8983/solr

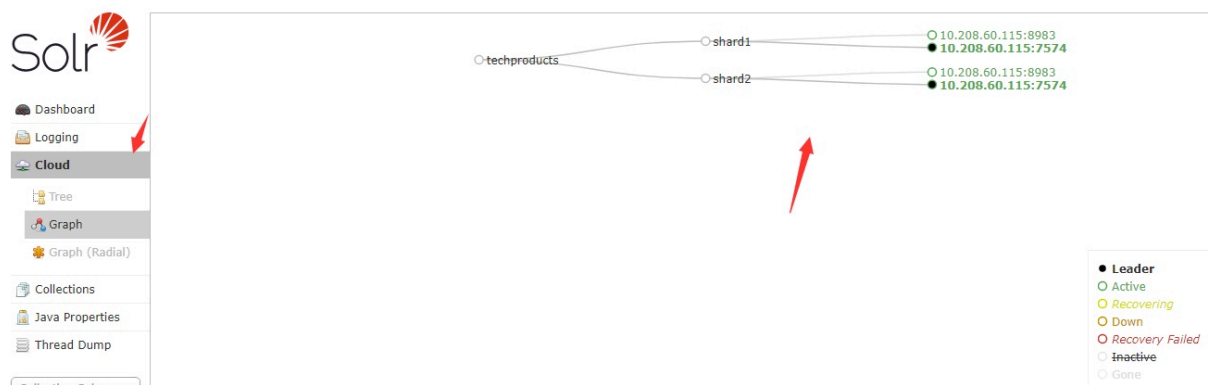
D:\solr-7.1.0>
```

最后来访问：

`http://localhost:8983/solr`



Solr现在将运行两个“节点”，一个在端口7574上，一个在端口8983上。一个集合techproducts，两个分片集合，每个都有两个副本。【点击Cloud可见】



到此Solr已经启动啦，可是还没有数据。下面来完成。  
加入数据:

```
Linux: solr-7.1.0:$ bin/post -c techproducts example/exampledocs/*
Windows: D:\solr-7.1.0> java -jar -Dc=techproducts -Dauto example\exampledocs\post.jar example\exampledocs\*
```

结果:

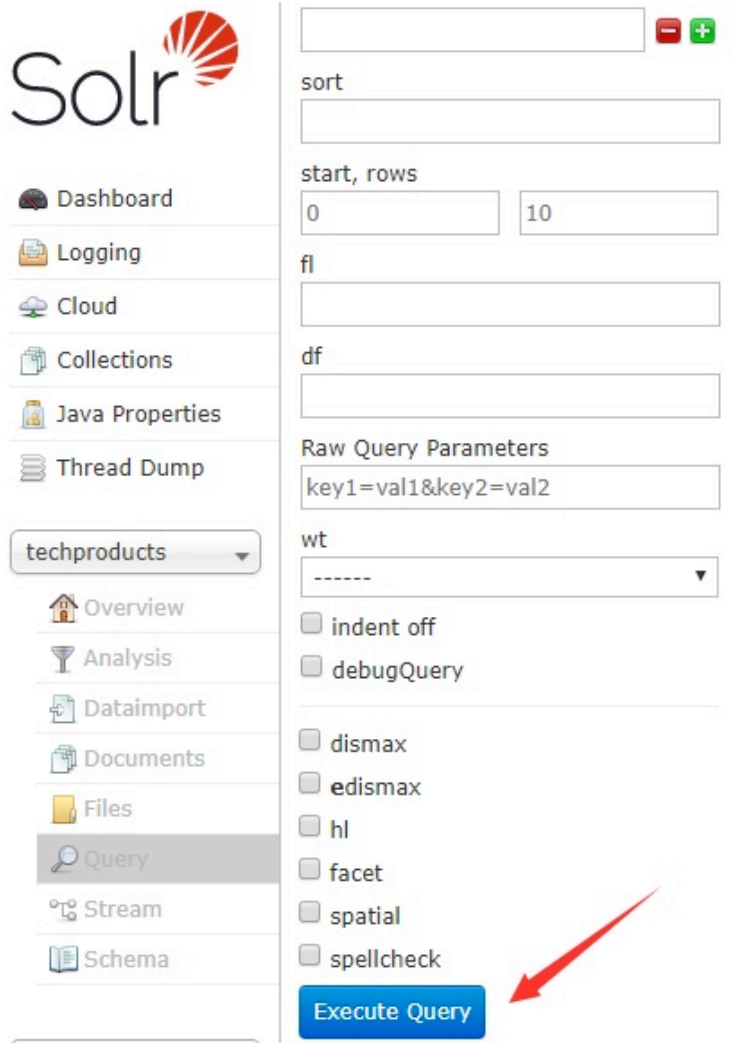
```
D:\solr-7.1.0>java -jar -Dc=techproducts -Dauto example\exampledocs\post.jar example\exampledocs\*
SimplePostTool version 5.0.0
Posting files to [base] url http://localhost:8983/solr/techproducts/update...
Entering auto mode. File endings considered are xml,json,jsonl,csv,pdf,doc,docx,ppt,pptx,xls,xlsx,odt,odp,ods,ott,otp,ots,rtf,htm,html,txt,log
```

```
POSTing file books.csv (text/csv) to [base]
POSTing file books.json (application/json) to [base]/json/docs
POSTing file gb18030-example.xml (application/xml) to [base]
POSTing file hd.xml (application/xml) to [base]
POSTing file ipod_other.xml (application/xml) to [base]
POSTing file ipod_video.xml (application/xml) to [base]
POSTing file manufacturers.xml (application/xml) to [base]
POSTing file mem.xml (application/xml) to [base]
POSTing file money.xml (application/xml) to [base]
POSTing file monitor.xml (application/xml) to [base]
POSTing file monitor2.xml (application/xml) to [base]
POSTing file more_books.jsonl (application/json) to [base]/json/docs
POSTing file mp500.xml (application/xml) to [base]
POSTing file post.jar (application/octet-stream) to [base]/extract
POSTing file sample.html (text/html) to [base]/extract
POSTing file sd500.xml (application/xml) to [base]
POSTing file solr-word.pdf (application/pdf) to [base]/extract
POSTing file solr.xml (application/xml) to [base]
POSTing file test_utf8.sh (application/octet-stream) to [base]/extract
t
POSTing file utf8-example.xml (application/xml) to [base]
POSTing file vidcard.xml (application/xml) to [base]
21 files indexed.
COMMITting Solr index changes to http://localhost:8983/solr/techproducts/update.
Time spent: 0:00:06.768
D:\solr-7.1.0>
```

## 三.基本搜索界面介绍

---

访问: <http://localhost:8983/solr/#/techproducts/query>



The image shows the Solr Admin web interface. On the left is a sidebar with navigation links: Dashboard, Logging, Cloud, Collections, Java Properties, Thread Dump, a dropdown menu showing 'techproducts', and a list of options including Overview, Analysis, Dataimport, Documents, Files, Query (which is highlighted), Stream, and Schema. The main area on the right contains various configuration fields: an empty search box with minus and plus icons, a 'sort' label with an empty input field, a 'start, rows' section with input fields for '0' and '10', an 'fl' (fields list) input field, a 'df' (default field) input field, a 'Raw Query Parameters' section with the text 'key1=val1&key2=val2', a 'wt' (output format) dropdown menu currently set to '-----', and several checkboxes for query features: 'indent off', 'debugQuery', 'dismax', 'edismax', 'hl', 'facet', 'spatial', and 'spellcheck'. At the bottom of the main area is a blue 'Execute Query' button, which is pointed to by a red arrow.

点击Execute Query.开始查询,你会发现右边有好几条JSON数据。准确来说:是十条.

fl

df

Raw Query Parameters

wt

☐ indent off  
☐ debugQuery

☐ dismax  
☐ edismax  
☐ hl  
☐ facet  
☐ spatial  
☐ spellcheck


[Execute Query](#)

```

    "price_c": "0.00, USD",
    "inStock": false,
    "author": "Glen Cook",
    "author_s": "Glen Cook",
    "series_t": "The Chronicles of The Black Company",
    "sequence_i": 1,
    "genre_s": "fantasy",
    "_version_": 1583011027217285120,
    "price_c___l_ns": 699},
  {
    "id": "0441385532",
    "cat": ["book"],
    "name": "Jhereg",
    "price": 7.95,
    "price_c": "7.95, USD",
    "inStock": false,
    "author": "Steven Brust",
    "author_s": "Steven Brust",
    "series_t": "Vlad Taltos",
    "sequence_i": 1,
    "genre_s": "fantasy",
    "_version_": 1583011027301171200,
    "price_c___l_ns": 795},
  {
    "id": "0380014300",
    "cat": ["book"],
    "name": "Nine Princes In Amber",
    "price": 6.99,
    "price_c": "6.99, USD",
    "inStock": true,
    "author": "Roger Zelazny",
    "author_s": "Roger Zelazny",

```

右上角是发送的URL，你可以点击它，点击试试看



[Dashboard](#)  
[Logging](#)  
[Cloud](#)  
[Collections](#)  
[Java Properties](#)  
[Thread Dump](#)

Request-Handler (qt)

/select

common

q

\*:\*

fq

sort

start, rows

0, 10

[http://localhost:8983/solr/techproducts/select?q=\\*:\\*](http://localhost:8983/solr/techproducts/select?q=*:*)

```

{
  "responseHeader": {
    "zkConnected": true,
    "status": 0,
    "QTime": 75,
    "params": {
      "q": "/*:*",
      "_: "1509677094712"}},
  "response": [{"numFound": 52, "start": 0, "maxScore": 1.0, "docs": [
    {
      "id": "0812521390",
      "cat": ["book"],
      "name": "The Black Company",

```

查询参数解释：



Request-Handler (qt)

/select

— common —

q

fq

sort

start, rows

0 10

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt

----- ▾

☐ indent off

☐ debugQuery

select请求处理器

主查询参数：搜索与其中参数相关的文档

过滤查询：通过过滤器筛选结果集文档

指定排序字段和排序方式


搜索结果的起始页【第一页为0】；页面大小：每一页的数量

为搜索结果中每个文档指定需要返回的字段

默认搜索字段

响应输出类型：好多种

修改q为foundation，结果显示有四个命中。



- Dashboard
- Logging
- Cloud
- Collections
- Java Properties
- Thread Dump
- techproducts ▾

Request-Handler (qt)

/select

— common —

q  
foundation

fq

sort

start, rows  
0 10

fl

http://localhost:8983/solr/techproducts/select?q=foundation

```
{
  "responseHeader": {
    "zkConnected": true,
    "status": 0,
    "QTime": 38,
    "params": {
      "q": "foundation",
      "_: "1509677662382"
    }
  },
  "response": {
    "numFound": 4, "start": 0, "maxScore": 2.671101, "docs": [
      {
        "id": "0553293354",
        "cat": ["book"],
        "name": "Foundation",
        "price": 7.99,
        "price_c": "7.99, USD",

```

将id放入fl中，再次查询，你会发现显示的字段只有id

The Solr Admin UI shows the 'Request-Handler (qt)' section with the path '/select'. The 'q' field contains 'foundation'. The 'fl' field contains 'id', highlighted with a red arrow. The 'start' is 0 and 'rows' is 10. The 'Raw Query Parameters' section shows 'key1=val1&key2=val2'. The URL bar shows 'http://localhost:8983/solr/techproducts/select?fl=id&q=foundation'. The JSON response shows 4 results, with the first result having an 'id' of '0553293354'.

```
Request-Handler (qt)
/select

common
q
foundation

fq

sort

start, rows
0 10

fl
id

df

Raw Query Parameters
key1=val1&key2=val2

wt
-----

[indent off]
[dehighlight]
```

```
http://localhost:8983/solr/techproducts/select?fl=id&q=foundation

{
  "responseHeader": {
    "zkConnected": true,
    "status": 0,
    "QTime": 11,
    "params": {
      "q": "foundation",
      "fl": "id",
      "start": "0",
      "rows": "10"
    }
  },
  "response": {
    "numFound": 4,
    "start": 0,
    "maxScore": 2.671101,
    "docs": [
      {
        "id": "0553293354",
        "name": "VTF8TEST",
        "id": "SOLR1000",
        "id": "D:\\solr-7.1.0\\example\\exampledocs\\test_utf8.sh"
      }
    ]
  }
}
```

我们再改变p为electronics，有14个结果。

The Solr Admin UI shows the 'Request-Handler (qt)' section with the path '/select'. The 'q' field contains 'electronics', highlighted with a red arrow. The 'fl' field contains 'id'. The 'start' is 0 and 'rows' is 10. The 'Raw Query Parameters' section shows 'key1=val1&key2=val2'. The URL bar shows 'http://localhost:8983/solr/techproducts/select?q=electronics'. The JSON response shows 14 results, with the first result having an 'id' of 'SP2514N'.

```
Request-Handler (qt)
/select

common
q
electronics

fq

sort

start, rows
0 10

fl
id

df

Raw Query Parameters
key1=val1&key2=val2

wt
-----

[indent off]
[dehighlight]
```

```
http://localhost:8983/solr/techproducts/select?q=electronics

{
  "responseHeader": {
    "zkConnected": true,
    "status": 0,
    "QTime": 20,
    "params": {
      "q": "electronics",
      "start": "0",
      "rows": "10"
    }
  },
  "response": {
    "numFound": 14,
    "start": 0,
    "maxScore": 1.8609722,
    "docs": [
      {
        "id": "SP2514N",
        "name": "Samsung SpinPoint P120 SP2514N - hard drive - 250 GB - ATA-133",
        "manu": "Samsung Electronics Co. Ltd.",
        "manu_id_s": "samsung",
        "cat": ["electronics", "hard drive"],
        "features": ["7200RPM, 8MB cache, IDE Ultra ATA-133", "NoiseGuard, SilentSeek technology, Fluid Dynamic Bearing (FDB) motor"],
        "price": 92.0,
        "price_c": "92.0, USD",
        "popularity": 6,
        "inStock": true,
        "manufacturedate_dt": "2006-02-13T15:26:37Z",
        "store": "35.0752, -97.032",
        "version": "1583011027602112512",
        "price_c__l_ns": 9200
      }
    ]
  }
}
```

这种查询是在所有索引字段查询包括electronics字段的文档，如果想更精确一点，我们看到有一个cat字段，那么将p改为cat:electronics，意思为仅仅cat字段中包含electronics信息的数据。

Request-Handler (qt)  
/select  
— common —  
q  
cat:electronics  
fq  
sort  
start, rows  
0 10  
fl  
df  
Raw Query Parameters  
key1=val1&key2=val2

http://localhost:8983/solr/techproducts/select?q=cat:electronics

```
{
  "responseHeader": {
    "zkConnected": true,
    "status": 0,
    "QTime": 24,
    "params": {
      "q": "cat:electronics",
      "_: "1509677862382"
    }
  },
  "response": {
    "numFound": 12, "start": 0, "maxScore": 0.9614112, "docs": [
      {
        "id": "SP2514N",
        "name": "Samsung SpinPoint P120 SP2514N - hard drive - 250 GB - ATA-133",
        "manu": "Samsung Electronics Co. Ltd.",
        "manu_id_s": "samsung",
        "cat": [ "electronics",
          "hard drive" ],
        "features": [ "7200RPM, 8MB cache, IDE Ultra ATA-133",
          "NoiseGuard, SilentSeek technology, Fluid Dynamic Bearing (FDB) motor" ],
        "price": 92.0,
        "price_c": "92.0, USD",
        "popularity": 6,
        "inStock": true,

```

可以看到，只有12条数据，更加精确了。

## 2.下面进行短语搜索

p改为"multiple terms here"，短语搜索必需加上双引号。

Request-Handler (qt)  
/select  
— common —  
q  
"CAS latency"  
fq  
sort  
start, rows  
0 10  
fl  
df  
Raw Query Parameters  
key1=val1&key2=val2

http://localhost:8983/solr/techproducts/select?q="CAS latency"

```
{
  "responseHeader": {
    "zkConnected": true,
    "status": 0,
    "QTime": 165,
    "params": {
      "q": "\"CAS latency\"",
      "_: "1509677862382"
    }
  },
  "response": {
    "numFound": 2, "start": 0, "maxScore": 3.878659, "docs": [
      {
        "id": "VDBDB1A16",
        "name": "A-DATA V-Series 1GB 184-Pin DDR SDRAM Unbuffered DDR 400 (PC 3200) System Memory - OEM",
        "manu": "A-DATA Technology Inc.",
        "manu_id_s": "corsair",
        "cat": [ "electronics",
          "memory" ],
        "features": [ "CAS latency 3, 2.7v" ],
        "popularity": 0,
        "inStock": true,
        "store": "45.18414, -93.88141",
        "manufacturedate_dt": "2006-02-13T15:26:37Z",
        "payloads": "electronics|0.9 memory|0.1",
        "_version_": 1583011027800293376,

```

最后,如果你希望删除此集合:

```
D:\solr-7.1.0> bin\solr delete -c techproducts
```

注意: Linux的路径为/

停止启动的Solr:

```
D:\solr-7.1.0> bin\solr stop -all
```



