

## NoSQL 演示具体步骤说明

鉴于大部分同学们多使用 Windows 系统，下面以 Windows 10 系统为例演示 NoSQL 相关操作，简化大家的学习负担。若使用 MacOS 系统，可利用 parallel 工具安装 Win 10 虚拟机。

### 一、安装 Java

由于 Elasticsearch 需要 Java 环境，所以需要安装 Java。建议采用 JDK 1.8 或以上版本，可参考 <https://www.cnblogs.com/xmnote/p/11306986.html> 或其他教程进行 JDK 下载、安装与环境变量设置，注意下载 JDK 时根据自己机器是 64 位还是 32 位，选择下载相应安装文件。重新打开 Windows cmd 命令行工具，在 Windows cmd 中输入 `java -version`，返回如下信息，表示 Java 安装成功。

```
C:\Users\linhao>java -version
java version "1.8.0_131"
Java(TM) SE Runtime Environment (build 1.8.0_131-b11)
Java HotSpot(TM) 64-Bit Server VM (build 25.131-b11, mixed mode)
```

### 二、安装 Python

课程演示的数据写入和读取过程由 Python 完成，因此需要安装 Python。课程演示代码在 Python 3.7 环境下测试通过，因此建议采用 Python 3.7，并使用 Anaconda 版 Python。可在官方网站进行下载 <https://www.anaconda.com/distribution/#download-section>，注意根据自己机器是 64 位还是 32 位，选择下载相应安装文件。下载完成之后双击 exe 文件进行安装，一直选 next 即可完成安装。安装完成后，将 Anaconda 实际安装路径 `C:\Users\linhao\anaconda3` 以及脚本路径 `C:\Users\linhao\anaconda3\Scripts` 加入系统环境变量 `PATH` 中。重新打开 Windows cmd 命令行工具，在 Windows cmd 中输入 `python` 以及 `pip`，如能正常显示，则表示 Python 安装成功。

```
C:\Users\linhao>python
Python 3.7.6 (default, Jan 8 2020, 20:23:39) [MSC v.1916 64 bit (AMD64)] :: Anaconda, Inc. on win32
Warning:
This Python interpreter is in a conda environment, but the environment has
not been activated. Libraries may fail to load. To activate this environment
please see https://conda.io/activation
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

出现上述 warning，需要在 cmd 中再执行 `activate base`，以加载必要的 Python Library，否则后面使用 Python 可能会有 DLL load failed 的错误。

```
C:\Users\linhao>pip
Usage:
  pip <command> [options]

Commands:
  install           Install packages.
  download          Download packages.
  uninstall         Uninstall packages.
  freeze            Output installed packages in requirements format.
  list              List installed packages.
  show              Show information about installed packages.
  check             Verify installed packages have compatible dependencies.
  config            Manage local and global configuration.
  search            Search PyPI for packages.
  wheel             Build wheels from your requirements.
  hash              Compute hashes of package archives.
  completion        A helper command used for command completion.
  debug             Show information useful for debugging.
  help              Show help for commands.
```

### 三、安装必要的 Python 第三方软件包

在 Windows 命令行 cmd 中通过 pip 命令安装用于连接 Elasticsearch 和 MySQL 数据库的 Python Client，命令分别是 `pip install elasticsearch==5.4.0 -i http://pypi.douban.com/simple --trusted-host pypi.douban.com`，`pip install PyMySQL -i http://pypi.douban.com/simple --trusted-host pypi.douban.com`。执行完成之后，在 Python 中能成功 import 各自的 package 即表示安装成功。注意针对 ES 数据库我们将使用 2.3.3 版本（较旧），因此对应的 ES python client 版本我们选择安装兼容的 5.4.0 版本。PyMySQL 官方文档见 <https://pymysql.readthedocs.io/en/latest/>。elasticsearch Python client 官方文档见 <https://elasticsearch-py.readthedocs.io/en/master/>。

```
(base) C:\Users\linhao>python
Python 3.7.6 (default, Jan 8 2020, 20:23:39) [MSC v.1916 64 bit (AMD64)] :: Anaconda, Inc. on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> from elasticsearch import Elasticsearch
>>> import pymysql
>>>
```

### 四、安装、配置、启动 Elasticsearch 数据库服务

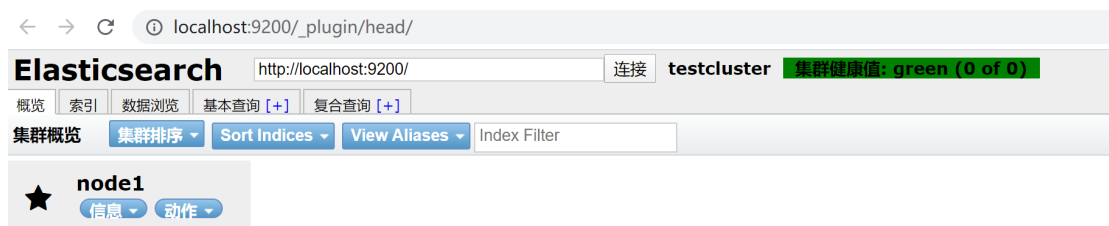
首先，推荐以源码方式安装和启动 ES 数据库服务（这种方式较为简单，且平台移植性较好，也较容易在单机上部署多个数据库实例）。具体而言，在 <https://www.elastic.co/downloads/past-releases/elasticsearch-2-3-3> 中下载 2.3.3 版本的 Elasticsearch 数据库源码 ZIP 文件，解压后置于桌面上。ES 2.3.3 文档见 <https://www.elastic.co/guide/en/elasticsearch/reference/2.3/index.html>。

随后，需要安装 Elasticsearch 数据库图形化管理界面插件 head。具体而言，打开 Windows 命令行工具 cmd，cd 进入 ES 数据库源码目录 `cd C:\Users\linhao\Desktop\elasticsearch-2.3.3`，执行 `bin\plugin install mobz/elasticsearch-head`，如下图所示即表示 head 安装成功。



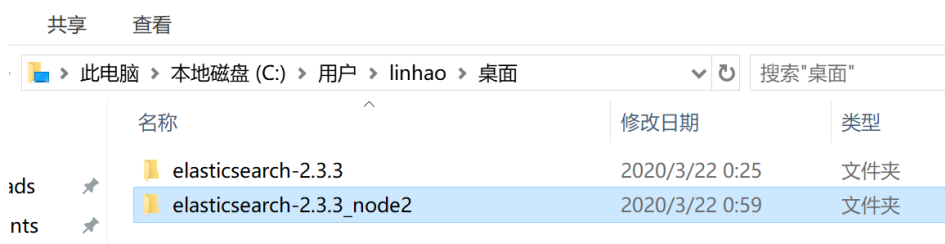
```
C:\Users\linhao>cd C:\Users\linhao\Desktop\elasticsearch-2.3.3
C:\Users\linhao\Desktop\elasticsearch-2.3.3>bin\elasticsearch.bat
[2020-03-22 00:25:27,240][INFO ][node                ] [node1] version[2.3.3], pid[5344], build[218bdf1/2016-05-17T15:40:04Z]
[2020-03-22 00:25:27,240][INFO ][node                ] [node1] initializing ...
[2020-03-22 00:25:27,771][INFO ][plugins            ] [node1] modules [reindex, lang-expression, lang-groovy], plu
gins [head], sites [head]
[2020-03-22 00:25:27,787][INFO ][env                ] [node1] using [1] data paths, mounts [[(C:)], net usable_sp
ace [220.5gb], net total_space [255.4gb], spins? [unknown], types [NTFS]
[2020-03-22 00:25:27,787][INFO ][env                ] [node1] heap size [910.5mb], compressed ordinary object poin
ters [true]
[2020-03-22 00:25:29,709][INFO ][node                ] [node1] initialized
[2020-03-22 00:25:29,709][INFO ][node                ] [node1] starting ...
[2020-03-22 00:25:29,802][INFO ][transport          ] [node1] publish_address {127.0.0.1:9300}, bound_addresses {1
27.0.0.1:9300}
[2020-03-22 00:25:29,802][INFO ][discovery          ] [node1] testcluster/li0yGjo-Qye-9e5tXQ4Ylw
[2020-03-22 00:25:33,896][INFO ][cluster.service    ] [node1] new_master {node1}{li0yGjo-Qye-9e5tXQ4Ylw}{127.0.0.1
}{127.0.0.1:9300}, reason: zen-disco-join(elected_as_master, [0] joins received)
[2020-03-22 00:25:33,912][INFO ][http               ] [node1] publish_address {127.0.0.1:9200}, bound_addresses {1
27.0.0.1:9200}
[2020-03-22 00:25:33,912][INFO ][node                ] [node1] started
[2020-03-22 00:25:33,927][INFO ][gateway            ] [node1] recovered [0] indices into cluster_state
```

在浏览器中输入网址 [http://localhost:9200/\\_plugin/head](http://localhost:9200/_plugin/head) 即可使用之前安装的 head 插件查看 ES 集群（包含一个 ES 节点 node1）状态。



ES 的一个特点是分布式，因此可以尝试构造包含多个节点的集群，本演示模拟在单机上部署两个节点。简单而言，就是找到刚才桌面上 ES 数据库源码，将其复制一份，并将其文件夹命名为 `elasticsearch-2.3.3_node2`。随后将其中的 `elasticsearch-2.3.3_node2/config/elasticsearch.yml` 文件中 `node.name` 对应值由 `node1` 修改为一个新节点名称如 `node2`，`http.port` 对应值由 `9200` 修改为一个新端口号如 `9201`，保存修改。同时删除 `elasticsearch-2.3.3_node2` 下的 `data` 文件夹。打开一个新的 cmd，启动 `node2` 的 ES 服务即可。

桌面



```
C:\Users\linhao\Desktop>cd C:\Users\linhao\Desktop\elasticsearch-2.3.3_node2
C:\Users\linhao\Desktop\elasticsearch-2.3.3_node2>bin\elasticsearch.bat
[2020-03-22 01:07:23,111][INFO ][node                ] [node2] version[2.3.3], pid[9664], build[218bdf1/2016-05-17T15:40:04Z]
[2020-03-22 01:07:23,111][INFO ][node                ] [node2] initializing ...
[2020-03-22 01:07:23,501][INFO ][plugins            ] [node2] modules [reindex, lang-expression, lang-groovy], plu
gins [head], sites [head]
[2020-03-22 01:07:23,517][INFO ][env                ] [node2] using [1] data paths, mounts [[(C:)], net usable_sp
ace [220.4gb], net total_space [255.4gb], spins? [unknown], types [NTFS]
[2020-03-22 01:07:23,517][INFO ][env                ] [node2] heap size [910.5mb], compressed ordinary object poin
ters [true]
[2020-03-22 01:07:25,095][INFO ][node                ] [node2] initialized
[2020-03-22 01:07:25,095][INFO ][node                ] [node2] starting ...
[2020-03-22 01:07:25,189][INFO ][transport          ] [node2] publish_address {127.0.0.1:9301}, bound_addresses {1
27.0.0.1:9301}
[2020-03-22 01:07:25,189][INFO ][discovery          ] [node2] testcluster/x1KhzMd6Q4CZdpEhk4PpBg
```

在浏览器中刷新网址 [http://localhost:9200/\\_plugin/head](http://localhost:9200/_plugin/head)，即可看到集群有两个节点都正常提供服务了（原先的 `node1` 和新配置的 `node2` 都启动成功并自动连成一个集群了，实际是通过 `config.yml` 文件中 `cluster.name` 关联起来的）。至此实现单机双节点的 ES 集群配置。类似

地，如果想增加更多节点，可以以类似方式进行配置。



注：可能存在刷新上述网址仍只有一个节点 node1 的情况，这种情况下可先选中启动 node1 ES 服务的命令行窗口敲击回车，再刷新网址即可看到两个节点。这是 Windows 命令行的一个 bug，即命令偶然出现卡顿的情况，需回车才能继续执行。

## 五、安装、配置并启动 MySQL 数据库服务

可搜索网上教程，进行 MySQL Server 的安装与服务启动。

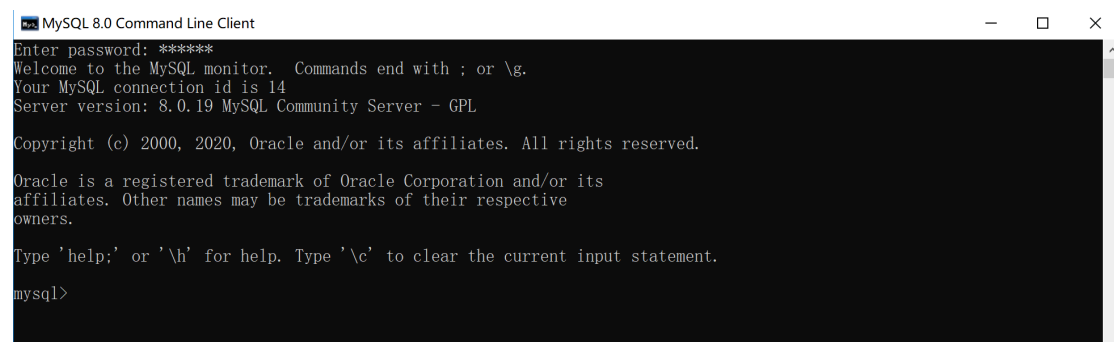
## 六、创建表结构

ES：在 Windows cmd 中进入到 code 目录中运行 `python es_create.py`

MySQL 创建表的过程稍微复杂一些，需要结合 MySQL Command Line Client 进行执行，可参考 <https://blog.csdn.net/chenriyang0306/article/details/54587034>，具体步骤如下：

首先，在 Windows 搜索栏中搜索 mysql 8.0 Command Line Client（我的 mysql 是 8.0 版本），打开该客户端工具。

随后输入安装时候给 root 用户设置的密码，我的是 123456，回车。



创建 test 数据库：

```
MySQL 8.0 Command Line Client
Enter password: *****
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 14
Server version: 8.0.19 MySQL Community Server - GPL

Copyright (c) 2000, 2020, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> create database test;
Query OK, 1 row affected (0.01 sec)

mysql>
```

创建 news 表结构 (SQL 语句在 mysql\_create.sql 中):

```
MySQL 8.0 Command Line Client
Enter password: *****
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 16
Server version: 8.0.19 MySQL Community Server - GPL

Copyright (c) 2000, 2020, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use test;
Database changed
mysql> CREATE TABLE news (
->   id varchar(255) NOT NULL,
->   url varchar(255) NOT NULL,
->   datestr varchar(255) NOT NULL,
->   title varchar(255) NOT NULL,
->   content varchar(255) NOT NULL,
->   PRIMARY KEY (id)
-> ) ENGINE=InnoDB DEFAULT CHARSET=utf8;
Query OK, 0 rows affected, 1 warning (0.03 sec)

mysql>
```

## 七、执行数据写入

ES: 在 Windows cmd 中进入到 code 目录中运行 `python es_write.py`

MySQL: 在 Windows cmd 中进入到 code 目录中运行 `python mysql_write.py`。如果报错: `pymysql.err.DataError: (1406, "Data too long for column 'content' at row 1")`, 可以在 MySQL client 中设置 `SET @@global.sql_mode= ''`; 取消严格模式再重新执行 python 脚本即可。

## 八、执行全文检索

ES: 在 Windows cmd 中进入到 code 目录中运行 `python es_search.py`

MySQL: 在 Windows cmd 中进入到 code 目录中运行 `python mysql_search.py`

注: 注意保持 code 文件夹下 `config.py` 中相关数据库名称、表名称、密码等配置与实际配置一置, 避免因不一致产生运行错误信息。比如 `config.py` 中 `MYSQL_PASSWORD` 代表的是 MySQL 数据库 root 用户的密码, 演示用的是 123456, 注意修改成安装 MySQL 时配置的 root 用户密码。