# 新浪微博数据抓取用户使用文档

目录

[1 运行环境清单 1](#_Toc404197000)

[2 JDK配置 1](#_Toc404197001)

[3 Mongodb配置 2](#_Toc404197002)

[3.1 安装Mongo数据库： 2](#_Toc404197003)

[3.2 配置Mongo服务端： 2](#_Toc404197004)

[3.3 Windows Service方式 3](#_Toc404197005)

[3.4 使用时候发生的异常及解决方法： 5](#_Toc404197006)

[4 启动BDG.jar 5](#_Toc404197007)

图表目录

[图 1 MongoDB配置成功后的画面 3](#_Toc404196302)

[图 2 MongoDB服务启动画面 4](#_Toc404196303)

[图 3 MongoDB服务开启查看 4](#_Toc404196304)

[图 4 MongoDB异常画面 5](#_Toc404196305)

## 运行环境清单

系统:windos7旗舰版

系统处理器:Intel(R) Core™ i5-2450M CPU @ 2.50GHZ 2.5GHZ

系统内存：8.00GB

系统类型：64位操作系统

数据库使用：MongoDB 版本1.6.5(64位)

JDK版本：1.7.0\_51-b13（工作版本）

开发工具：Eclipse 版本Kepler Service Release 2 (64位 英文版)

平台使用：新浪微博API平台版本：2.0

软件开发包：新浪微博Java SDK （包含新版接口及OAuth2.0）

## JDK配置

进行java开发，首先要安装jdk，安装了jdk后还要进行环境变量配置：

* 1. 下载jdk（<http://java.sun.com/javase/downloads/index.jsp>）
  2. 安装jdk-6u14-windows-i586.exe
  3. 配置环境变量：右击“我的电脑”-->"高级"-->"环境变量"
  4. 在系统变量里新建JAVA\_HOME变量，变量值为：C:\Program Files\Java\jdk1.6.0\_14（根据自己的安装路径填写）
  5. 新建classpath变量，变量值为：.;%JAVA\_HOME%\lib;%JAVA\_HOME%\lib\tools.jar
  6. 在path变量（已存在不用新建）添加变量值：%JAVA\_HOME%\bin;%JAVA\_HOME%\jre\bin（注意变量值之间用“;”隔开）
  7. “开始”-->“运行”-->输入“javac”-->"Enter"，如果能正常打印用法说明配置成功！
  8. 补充环境变量的解析:

JAVA\_HOME:jdk的安装路径

classpath:java加载类路径，只有类在classpath中java命令才能识别，在路径前加了个"."表示当前路径。

path：系统在任何路径下都可以识别java,javac命令。

## Mongodb配置

### 安装Mongo数据库：

第一步：下载安装包：如果是win系统，注意是64位还是32位版本的，请选择正确的版本。

第二步：新建目录“D:\MongoDB”，解压下载到的安装包，找到bin目录下面全部.exe文件，拷贝到刚创建的目录下。

第三步：在“D:\MongoDB”目录下新建“data”文件夹，它将会作为数据存放的根文件夹。

注：官方下载地址：<http://www.mongodb.org/downloads>

### 配置Mongo服务端：

打开CMD窗口，按照如下方式输入命令：//以管理员方式运行

> d:> cd D:\MongoDB> mongod --dbpath D:\MongoDB\data

配置成功后会看到如下画面：

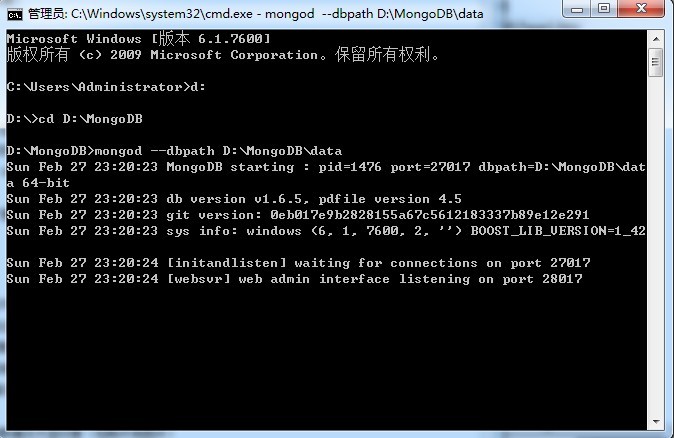
[](http://images.51cto.com/files/uploadimg/20110308/1049310.jpg)

图 1 MongoDB配置成功后的画面

在浏览器输入：http://localhost:27017/，可以看到如下提示： You are trying to access MongoDB on the native driver port. For http diagnostic access, add 1000 to the port number

如此，MongoDB数据库服务已经成功启动了。

### Windows Service方式

运行cmd.exe

> cd C:\Program Files\mongodb\bin

> C:\Program Files\mongodb\bin>mongod --dbpath "C:\Program Files\mongodb\data\db" --logpath "C:\Program Files\mongodb\data\log\MongoDB.log" --install --serviceName "MongoDB"

这里MongoDB.log就是开始建立的日志文件，--serviceName "MongoDB" 服务名为MongoDB

运行命令成功为如下图：

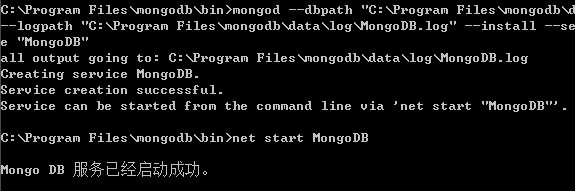


图 2 MongoDB服务启动画面

引时服务已经安装成功，运行

>NET START MongoDB (开启服务）

>NET stop MongoDB (关闭服务)

>

> C:\Program Files\mongodb\bin

>mongod --dbpath "C:\Program Files\mongodb\data\db" --logpath "C:\Program Files\mongodb\data\log\MongoDB.log" --remove --serviceName "MongoDB" (删除，注意不是--install了）

其它命令可查阅help命令或官网说明。

查看服务方式如下图：



图 3 MongoDB服务开启查看

运行bin文件夹下mongo.exe 客户端测试一下吧。测试同3.1相同 。

### 使用时候发生的异常及解决方法：

**异常：**

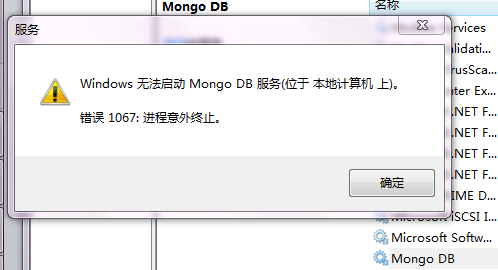


图 4 MongoDB异常画面

**解决方法：**

删掉mongod.lock文件就重启就可以了，但是在我的机子上依然出现上述问题。如果之前一直好好的，仅仅是因为插入了数据导致服务无法使用，处理问题的方法就很简单了，把你刚刚复制的数据和mongod.lock文件删掉（都在data文件夹下）。有可能是因为mongodb-win32-i386-2.2.1这个版本的某个文件夹下有你刚刚复制进去的数据，数据的重复导致只删掉mongod.lock无法启动。

## 启动BDG.jar

* 开启MongoDB
* 进入BDG.jar包内的\grab\sina\config内配置两个文档date\_config.json和access\_token\_config.json。access\_token\_config.json存放有access\_token，在access\_token仍有权限获取数据的情况下，不需要配置。date\_config.json存放的是抓取时限的信息，需要配置从哪一天开始，持续抓多少天的信息。
* 用cmd在JVM上运行BDG.jar