概述

使用旧接口批量导入数据的过程概括来说包括以下几个部分：

1. 保证能够连接到生产环境的后台。
2. 将需要导入的数据放入生产环境的指定目录下。
3. 使用隐藏页面的功能，启动批量操作
4. 等待操作完成

然后重复2～4步，逐个商家导入。

1. 连接到生产环境后台

1.1 防火墙设置

百图使用的UCLOUD云服务器，设立了外网防火墙，防止未授权的访问。 为了能够登录后台进行控制，需要首先在UCLOUD的防火墙上，对操作者的IP进行授权。

第一步：获得自己的外网IP地址。

使用百度搜索，搜索关键字 “IP"，就可以得到本机在公网上的出口地址：

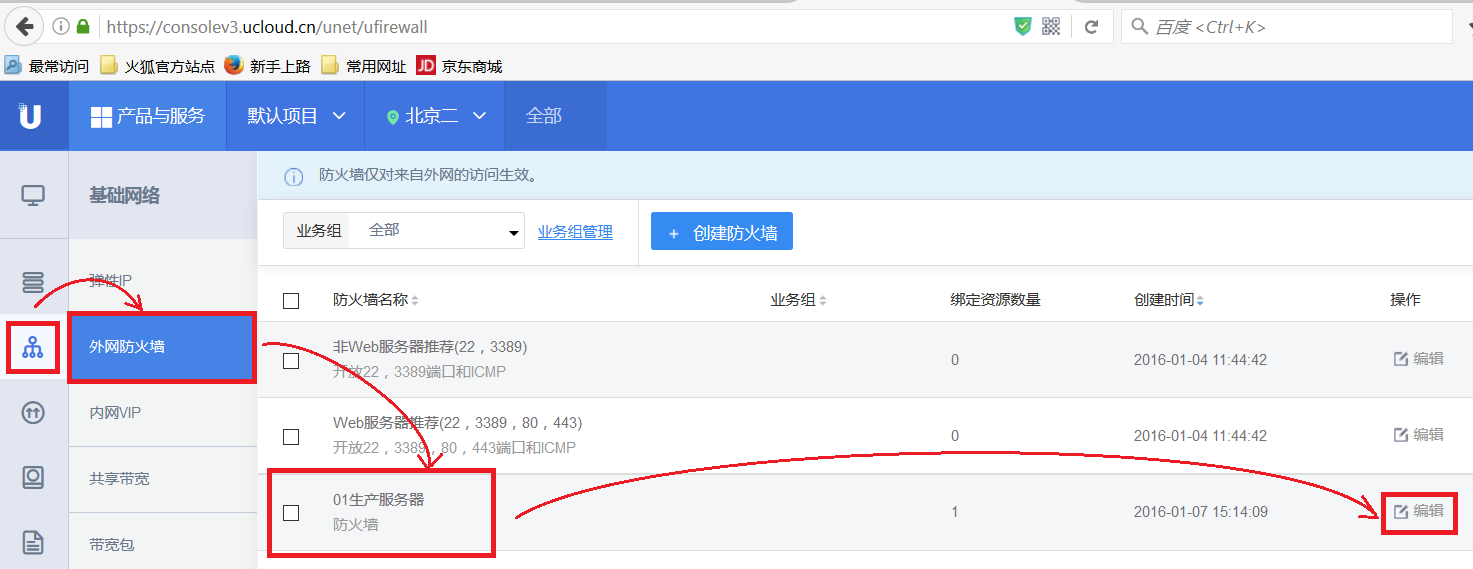


例如，写此文档时，本机的公网IP是 171.215.38.37

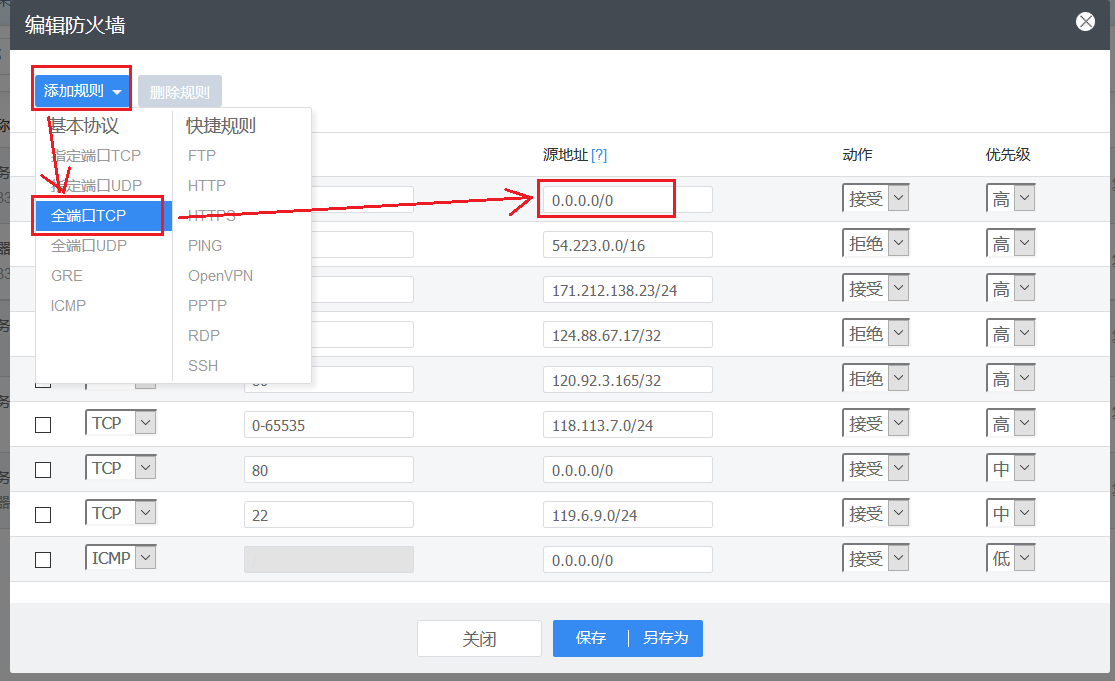
第二步，登录Ucloud的管理界面

<https://consolev3.ucloud.cn>

选择“基础网络”-->”外网防火墙”--> “生产环境”， 如下图



就进入了生产环境的外网防火墙设置界面



选择 “添加规则”->“全端口TCP”就会出现一个地址全0的新条目。

第三步：将本机的IP地址填入，然后保存



然后就可以从本机访问生产环境的后台了。

1.2 使用putty工具连接到生产环境后台

Putty是一个简单小巧的远程连接工具，可以连接到Linux 后台进行命令行操作。

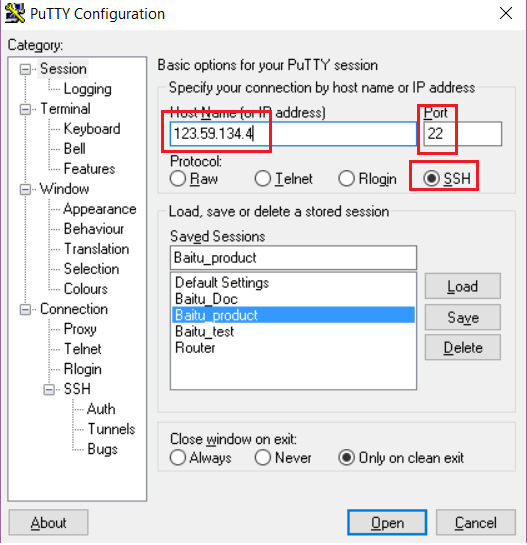
附件中有一个Putty软件包，直接使用即可。

打开Putty，设置连接生产环境的连接参数如下：

**IP：123.59.134.4**

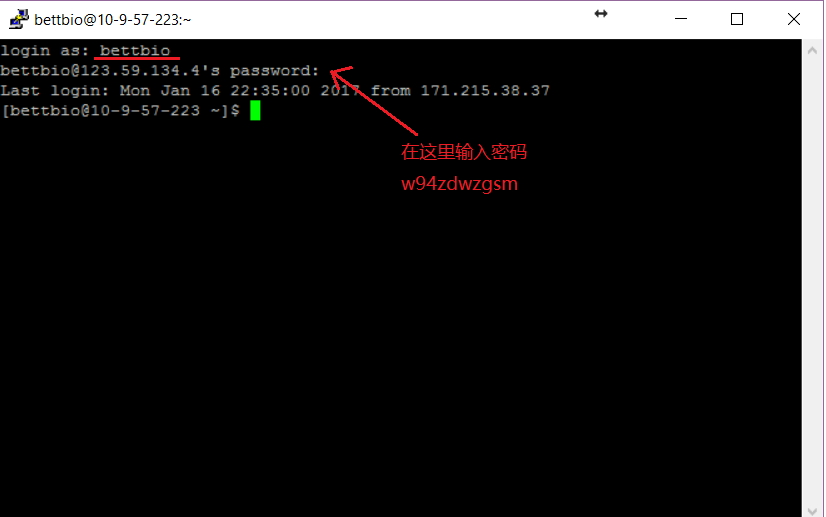
**端口：22**

**协议：SSH**



就可以连接到生产环境了。

登录生产环境的时候，使用用户名为；bettbio, 密码是 w94zdwzgsm (我就是知道我在干什么)



输入密码的时候，没有回显，只能盲打，所以要小心。输错了有3次机会。全错了就要重新连接了。

登录成功，就会出现”$”提示符

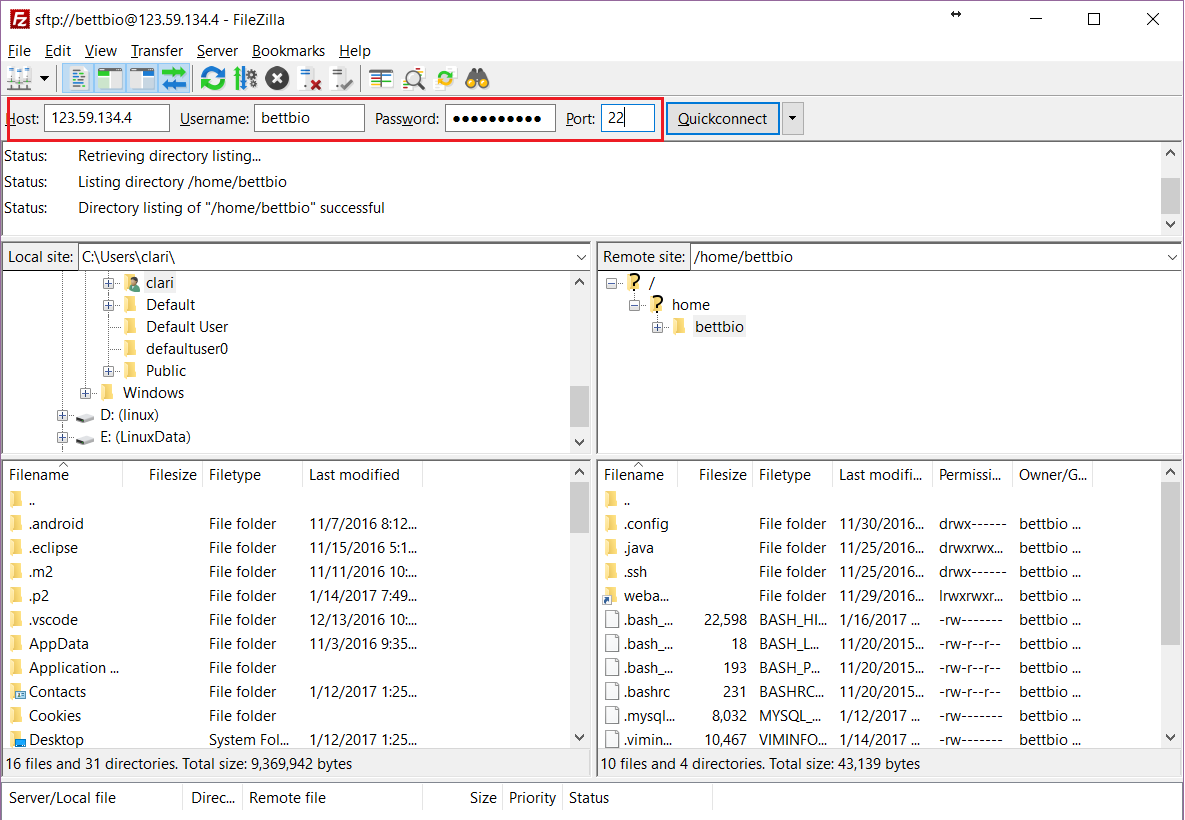
1.3 使用Filezilla 连接生产环境

Filezilla 是一个开源的FTP客户端软件，在这里我们用它来上传文件到生产环境的服务器上。

这个软件需要安装，下载地址是

<https://filezilla-project.org/download.php?type=client>

安装完成后就可以连接了。连接服务器的账号密码，地址端口都和putty的一样，协议选择 SFTP，也就是端口22。



Filezilla的操作都是图形化的，非常容易理解，支持拖动。后面具体操作的时候再举例说明。

1. 将需要导入的数据放入生产环境的指定目录下

2.1 批量数据的存放路径

生产环境的应用程序是用Tomcat运行的，Tomcat的安装路径为

**/data/apache-tomcat-8.0.36**

批量试剂数据是放在tomcat下的

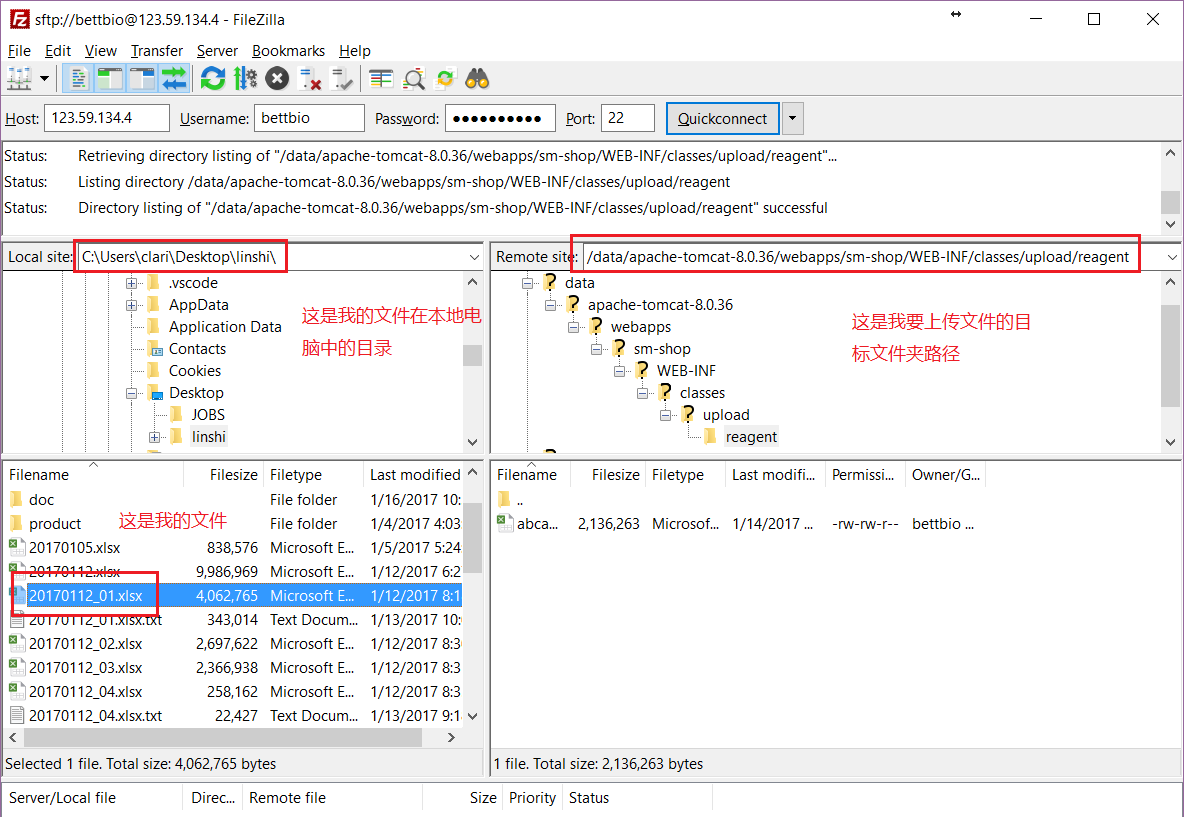
**webapps/sm-shop/WEB-INF/classes/upload/reagent** 目录

在生产环境就是

“/data/apache-tomcat-8.0.36/webapps/sm-shop/WEB-INF/classes/upload/reagent”

这个就是它的绝对路径。

假设我有一个文件，名字叫做 20170112\_1.xlsx, 在桌面的linshi 目录下，想放到这个目录下，那么用Filezilla，就可以实现。如下图所示：



批量服务类产品的数据是放在

**webapps/sm-shop/WEB-INF/classes/upload/server** 目录

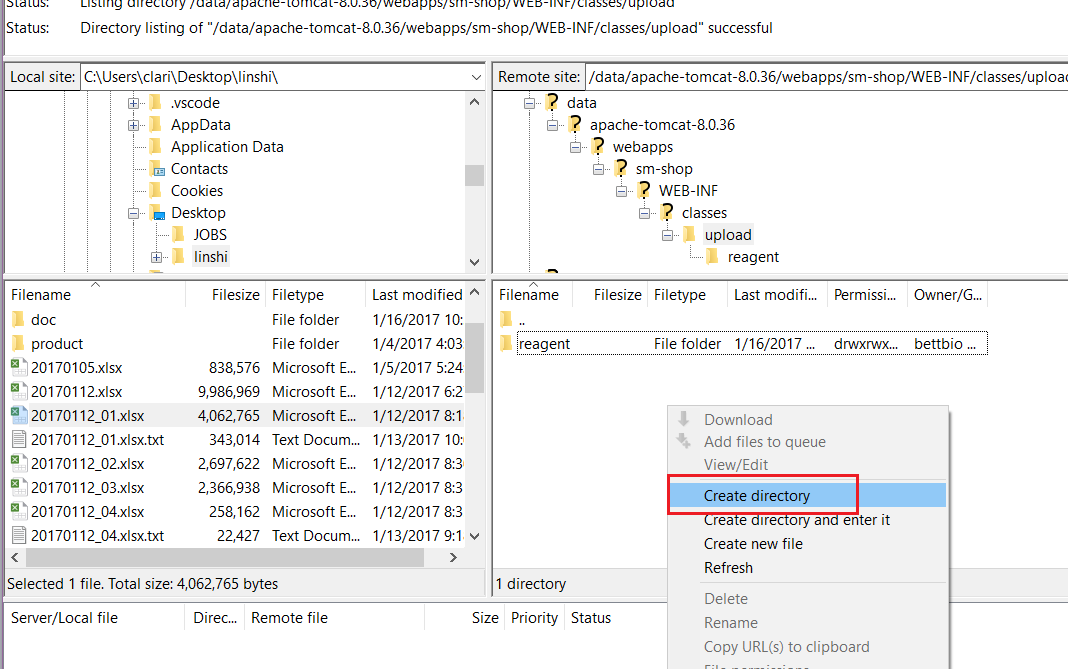
批量耗材数据是放在tomcat下的

**webapps/sm-shop/WEB-INF/classes/upload/other** 目录

如果目标文件夹不存在，还可以用FileZilla创建，操作很简单，在需要创建目录的地方，右键单击，选择Create Directory/新建目录 即可。 例如现在只有 “reagent”，需要上传服务类产品，那么就可以在

/data/apache-tomcat-8.0.36/webapps/sm-shop/WEB-INF/classes/upload

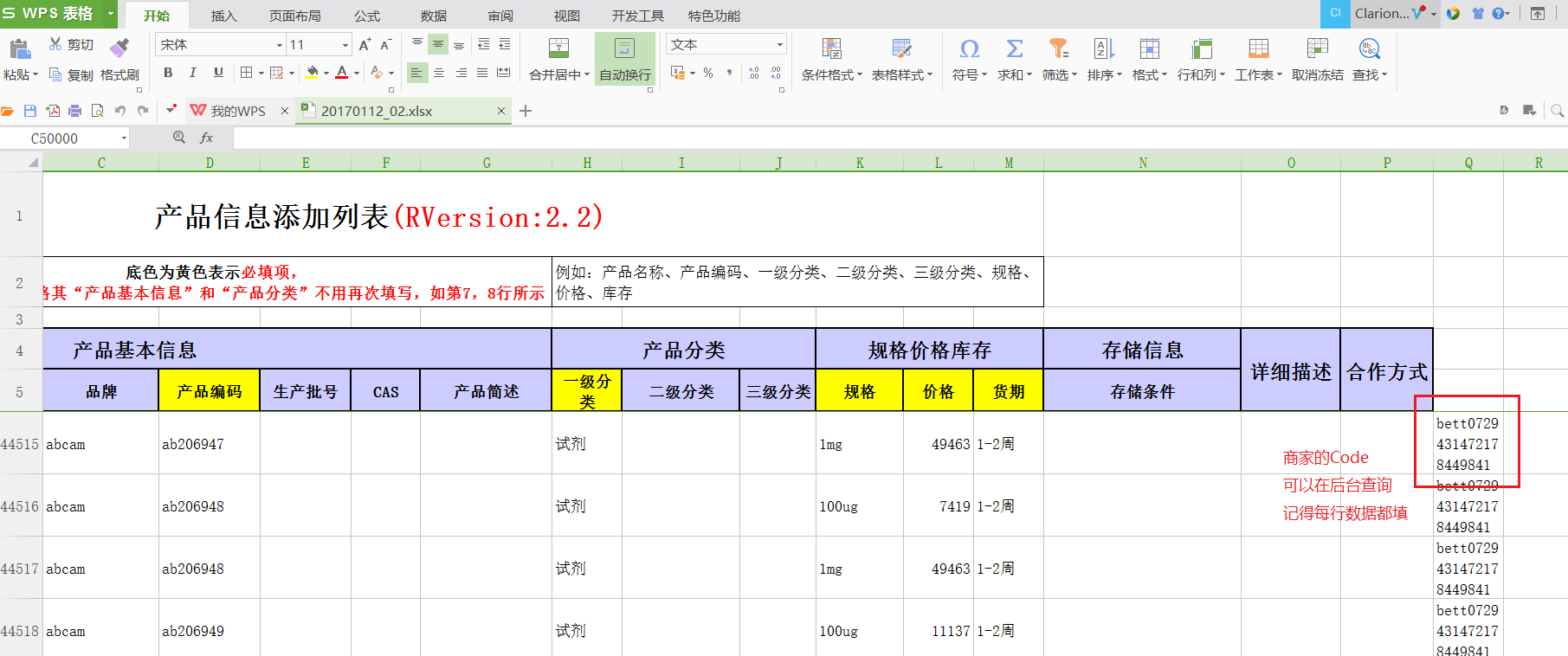
目录下右键单击，然后创建一个 server 目录



2.2 准备好数据

通常，都会把数据保存在一个统一的目录下，然后确认内容正确：

1. 不要出现 #N/A 这样的数据
2. 要记得填写商家的code：



1. 文件名最好全是英文，这样在linux后台中可以看到正确的文件名。 如果是中文，由于生产环境的语言包暂未支持中文，会看到乱码
2. 文件内容最好不要超过5万条记录

2.3 清理批量数据目录下的旧文件。

目前的代码，是遍历批量数据目录下的所有文件，然后全部导入的。如果已经导入的文件没有删除，下次会再导入一遍，就会有多个产品重复了。

所以每次开始一个批量的数据导入时，需要将所有已经操作完成的数据文件清除掉。在FileZilla中，简单的右键删除，或者Delete键删除均可。

2.4 使用FileZilla工具将文件上传。

这个非常简单，从左边拖动到右边，就把本地文件上传到服务器上了。

1. 使用隐藏页面的功能，启动批量操作

原有接口经过测试，在多次使用后会经常报 OutofMemory 异常，为了争取更高的批量操作成功率，建议没批量操作过10万以上的数据后，就重启一下Tomcat

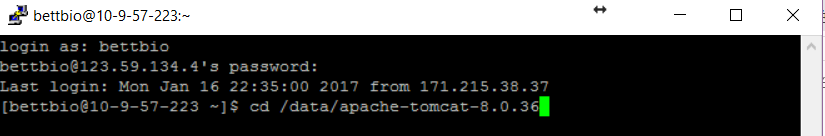
3.1 停止 tomcat

前面介绍了如何使用putty连接到生产环境后台，出现$提示符后，进入tomcat的安装目录,

输入命令

|  |
| --- |
| cd /data/apache-tomcat-8.0.36 |

就进入tomcat目录了



然后运行命令

|  |
| --- |
| bin/shutdown.sh |

就发送一个终止消息给Tomcat。(请确定通过前面的命令，当前目录是/data/apache-tomcat-8.0.36）



通常情况下，这个命令会正常停止tomcat，但是在批量数据操作后，很大肯定内存已经不足，无法正常的停止tomcat, 需要使用更底层的kill命令来终止tomcat

这里有一个脚本已经放在tomcat的bin目录下，可以直接调用：

|  |
| --- |
| bin/forceKill.sh |

(请确定当前目录是/data/apache-tomcat-8.0.36）



3.2 启动 tomcat

执行脚本

|  |
| --- |
| bin/startup.sh |

(请确定当前目录是/data/apache-tomcat-8.0.36）

就启动tomcat了

Tomcat的启动大约需要2分钟，在启动完全完成前，暂时还不能正常的执行批量数据导入功能。 我们可以通过查看tomcat的日志来看启动是否完全完成

3.3 查看tomcat的日志

使用命令

|  |
| --- |
| tail -f logs/catalina.out |

(请确定当前目录是/data/apache-tomcat-8.0.36）

就可以看到tomcat的日志输出了。

通常当你发现日志输出不再变化时，就是启动完成，并且没有其他请求需要处理了。

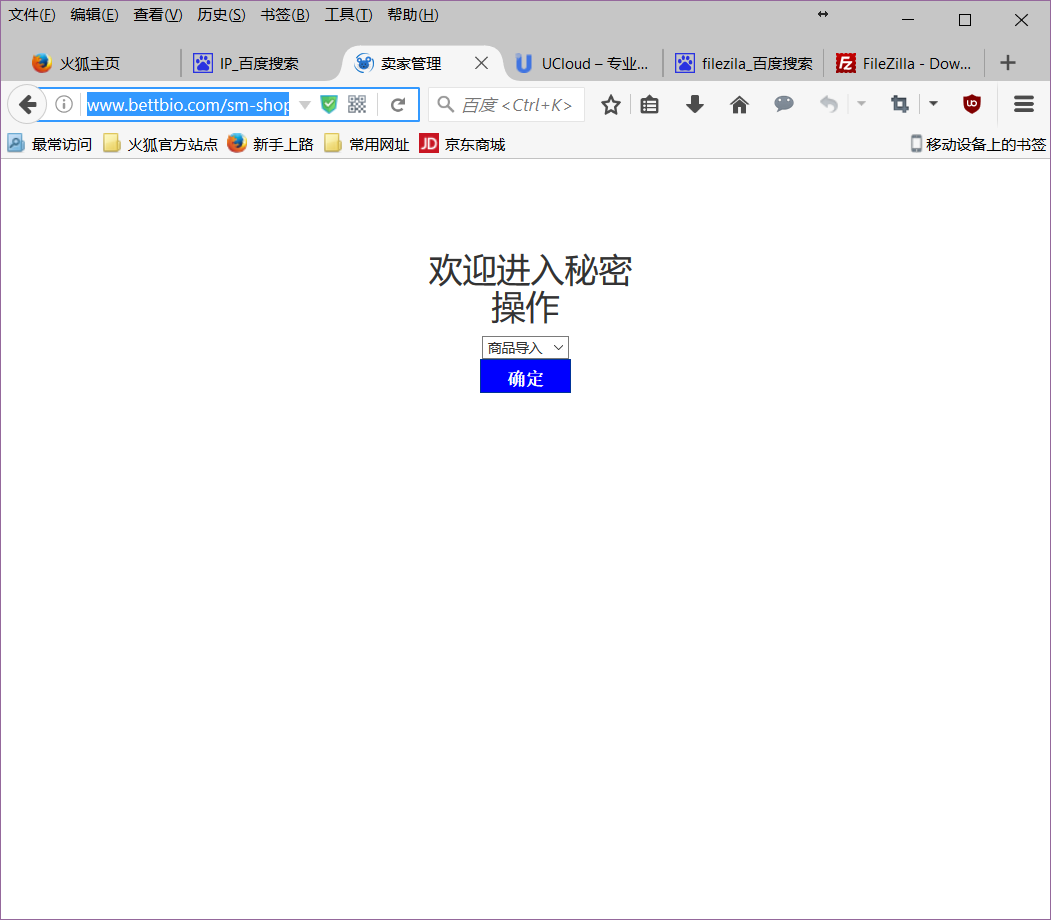
3.4 启动批量数据导入

首先，请确认几件事情：

1. 应用程序，即tomcat的启动已经顺利完成
2. 需要导入的数据已经正确的放置在目录里
3. 已经完成的数据文件，已经被移除了
4. 你已经开始监控查看tomcat的日志输出了

然后，在页面上登录超级管理员，在地址栏输入：

<http://www.bettbio.com/sm-shop/admin/loaddata.html>



确认页面上选择的是“商品导入”，点击确定，开始导入。

1. 等待操作完成

旧接口速度较慢，需要耐性等待。可以观察日志输出。

正常情况下会看到日志输出中有你上传的数据文件的文件名。 然后等待一段时间（视数据量大小不同，从半分钟到几十分钟不等），然后日志输出开始快速的输出大量的

Add Add index of product XXXXXX

这样的日志。

耐心等待日志不再输出此类信息，大约几十分钟到几个小时不等，数据都导入完成了。

导入完成后，用FileZIlla可以查看批量数据存放目录中，除了原来上传的文件外，还多了一个txt文件。这个是批量数据的一些记录信息。

建议将这个文件用filezilla下载到本地，如果发现其中有 Error 字样，表面数据导入出现了问题。 那么结合这个日志文件，和导入使用的批量数据文件，就可以在后面逐条的手工做修复。 如果错误太多，也可以考虑重新做数据，重启tomcat，然后重新导入。

无论何时发现需要处理错误，请记录此次数据导入的开始和结束时间，这个信息可以在数据库中做为一个过滤条件，将此次操作的产品ID找出来。

1. 重复2～4的步骤，导入各个商家

建议每次导入的数据累计超过10W即重启一次tomcat。