

# 面向金融行业数据的离线分析系统

## 使用说明

当前系统访问地址

<http://172.23.27.199:10000/>

测试用账号

	账号	密码
管理员	<a href="mailto:hufeng@cqupt.edu.cn">hufeng@cqupt.edu.cn</a>	123456789
管理员测试用户	<a href="mailto:guest@cqupt.edu.cn">guest@cqupt.edu.cn</a>	123456guest
普通用户 1	<a href="mailto:563951243@qq.com">563951243@qq.com</a>	987654321
普通用户 2	<a href="mailto:guest@163.com">guest@163.com</a>	123456guest
普通用户 3	<a href="mailto:guest@qq.com">guest@qq.com</a>	123456guest
普通用户 4	<a href="mailto:guest@fina.com">guest@fina.com</a>	123456guest

## 1 系统用户组

### 1.1 普通用户

普通用户组可以使用大数据平台看板、金融业务管理、金融数据采集与管理、金融分析模型管理、客户风险偏好数据分析与推荐、客户信用分数据分析。

### 1.2 管理员

管理员用户组除了可以使用普通用户组所有功能以外，还可以使用系统用户管理、大数据集群管理与监控功能。

## 2 系统功能说明

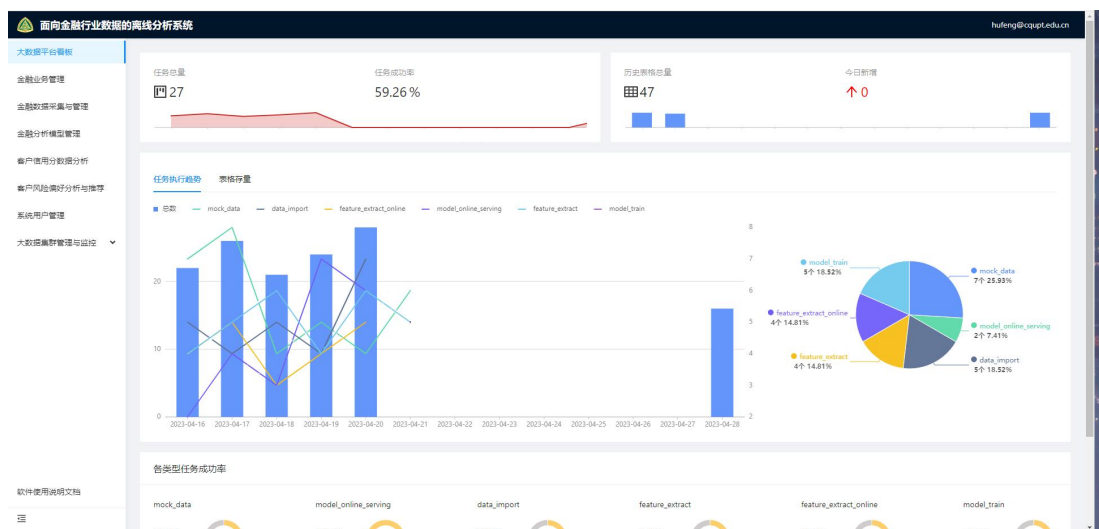
### 2.1 用户注册

游客用户如果进入系统需要注册为普通用户，进入系统注册页面后按提示输入个人信息后即可完成注册，待提示注册成功后即可（注！注册成功后账号并未激活，需要管理员进行账号激活后才能正常进入系统）。



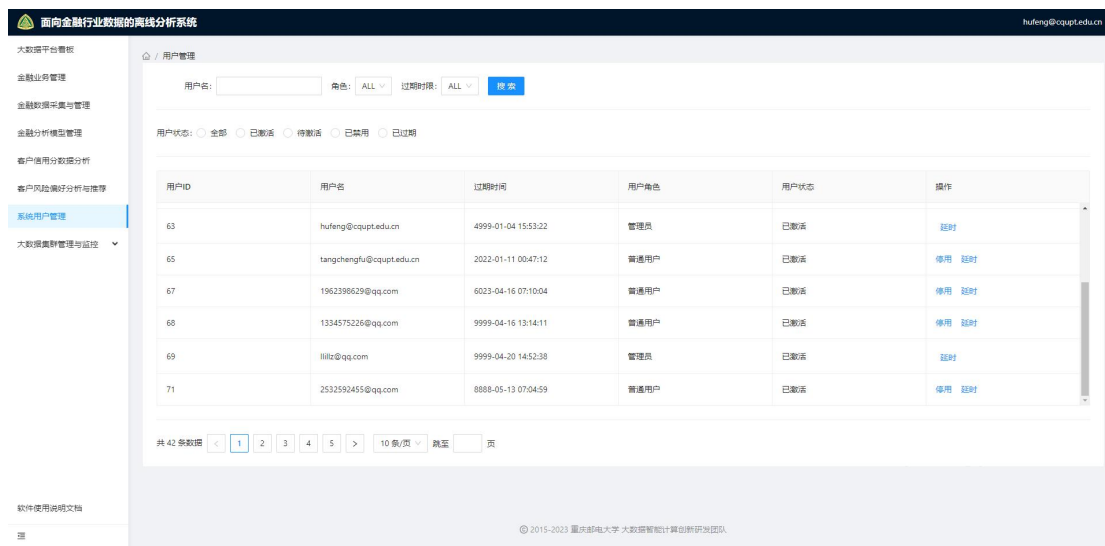
## 2.2 大数据平台看板

数据大屏，在数据大屏中可以展示当前系统中任务情况和表格情况，包括且不仅限于任务总量、成功率、趋势图等。



## 2.3 系统用户管理

用于系统用户账户管理，其中包含用户激活、用户启用\停用、用户延时功能。



## 2.3.1 用户激活

由游客用户自行填写资料注册后，需要管理员进行用户激活操作该账号才能正常使用。

tangchengfu@cqupt.edu.cn	2022-01-11 08:47:12	普通用户	待激活	激活 延时
--------------------------	---------------------	------	-----	-------

## 2.3.2 用户停用

若希望禁止某些账户登陆系统可使用用户停用功能，点击停用后会将该用户踢出系统。

60	tangcfmail@163.com	5026-07-22 22:36:14	普通用户	已激活	停用 延时
----	--------------------	---------------------	------	-----	-------

## 2.3.3 用户启用

若用户被停用，可使用用户启用功能使用户账户可以正常使用。

563951243@qq.com	2022-01-05 23:49:37	普通用户	已禁用	启用 延时
------------------	---------------------	------	-----	-------

## 2.3.4 用户延时

由于系统为计时使用，若用户授权时间到期后可以为该用户续时。

用户延时

✕

延时单位

☐ 秒

☐ 分

☐ 时

☒ 天

☐ 月

☐ 年

延长时间

1

取消

确定

## 2.4 金融数据采集与管理

支持系统中所有表格的元信息，并且可以实现随机数据表的创建、特征工程还可以完成数据导入功能。

面向金融行业数据的离线分析系统

hufeng@cqupt.edu.cn

大数据平台看板

金融业务管理

金融数据应用与管理

金融分析模型管理

客户信用分数数据分析

客户风险偏好分析与推荐

系统用户管理

大数据集群管理与监控

数据管理

搜索

创建随机数据表

创建特征提取表

导入文件数据表

刷新

表ID	表名	表创建时间	表格类型	表格状态	操作
330	t5	2023-07-07 21:04:04	导入数据表	<span>Success</span>	<a href="#">查看</a>
329	T7	2023-07-07 20:53:31	导入数据表	<span>Success</span>	<a href="#">查看</a>
328	t5	2023-07-07 16:21:37	导入数据表	<span>Success</span>	<a href="#">查看</a>
327	t3	2023-07-07 15:47:39	导入数据表	<span>Success</span>	<a href="#">查看</a>
326	t2	2023-07-07 15:32:39	导入数据表	<span>Success</span>	<a href="#">查看</a>
325	t1	2023-07-07 15:31:41	导入数据表	<span>Success</span>	<a href="#">查看</a>

共 85 条数据

1

2

3

4

5

...

15

>

6 条/页

跳至

页

软件使用说明文档

返回

© 2015-2023 重庆邮电大学 大数据智能计算创新研发团队

### 2.4.1 创建随机数据表

支持创建一张包含固定数据量的随机数据表格。支持的随机字段类型包括字符枚举、整型数字、浮点型数字和日期。

基础信息

表名

table1

表简介

简介

配置信息

生成数据量

10000

+ 添加列

\* 列1

\* 数据类型

\* 数据生成器

\* 最小值

\* 最大值

\* 列注释

1

整型

范围

1

111

1

## 2.4.2 特征工程

支持特征工程表的创建，支持 log、加、除、onehot、label encoding 算子

基础信息

表名

fetest1

表简介

d

配置信息

原始表

cfmodksuccess

+ 添加列

\* 列1

\* 特征提取算子

\* 计算列

\* 列注释

列名

Log

列注释

Log

加

除

OneHot编码

Label编码

保持

提交

## 2.4.3 导入数据表

当前支持上传 csv 文件作为数据表，注意需要在下面列字段中将 csv 中所有的字段均按照实际的数据类型配置好。

基础信息

表名

表名

表简介

表简介

数据文件

选择文件

未选择文件

上传

数据列

+ 添加列

\* 列1

\* 数据类型

\* 列注释

col1

整型

列注释

## 2.5 模型管理

模型管理

搜索

添加模型

刷新

模型ID	模型状态	模型名称	模型训练表	模型Jar包	模型训练类名	模型推理类名	模型在线服务类名	模型保存路径	模型标签列
86	standby	stroke_prediction	stroke_data	hdfs://172.23.27.199...	cn.edu.cqupt.ModelTr...	cn.edu.cqupt.Offline...	cn.edu.cqupt.OnlineS...	hdfs://172.23.27.199...	2 stroke
82	standby	高血压预测	excel_classifier	hdfs://172.23.27.199...	cn.edu.cqupt.ModelTr...	cn.edu.cqupt.Inferen...	cn.edu.cqupt.OnlineS...	hdfs://172.23.27.199...	2 hypertensionlabel
80	standby	协同过滤推荐模型	cf	hdfs://172.23.27.199...	cn.edu.cqupt.ModelTr...	cn.edu.cqupt.Offline...	cn.edu.cqupt.OnlineS...	hdfs://172.23.27.199...	2 rate
78	standby	客户信用分数分析模型	user_certificate_fe	hdfs://172.23.27.199...	cn.edu.cqupt.ModelTr...	cn.edu.cqupt.Offline...	cn.edu.cqupt.OnlineS...	hdfs://172.23.27.199...	2 label

共 4 条数据

< 1 >

10 条/页

### 2.5.1 模型新增与训练

在模型新增中需要明确模型中的三个类，训练类（用于模型训练和独立部署后的模型迭代）、推理类名（当前无用，后续可作为模型分析使用）、实时流类名（实时数据流下的推理类名）

需要明确模型使用的特征列和标签列。

当前金融系统总共提供两个模型包（线性回归、协同过滤）



业务 uri /recommend

添加业务

\* 业务名称: 客户风险偏好数据分析与推荐系统

模型: 协同过滤推荐模型:standby

\* 业务Uri: /recommend

\* 业务简介: 客户风险偏好数据分析与推荐系统

取消 确定

### 2.6.1.2 新增客户信用分数数据分析

业务模型需要为“线性回归模型”

对应 jar 包: linear\_regression-1.0-SNAPSHOT-jar-with-dependencies.jar

业务 uri /userCertificate

添加业务

\* 业务名称: 客户信用分数数据分析

模型: 客户信用分数数据分析模型:standby

\* 业务Uri: /userCertificate

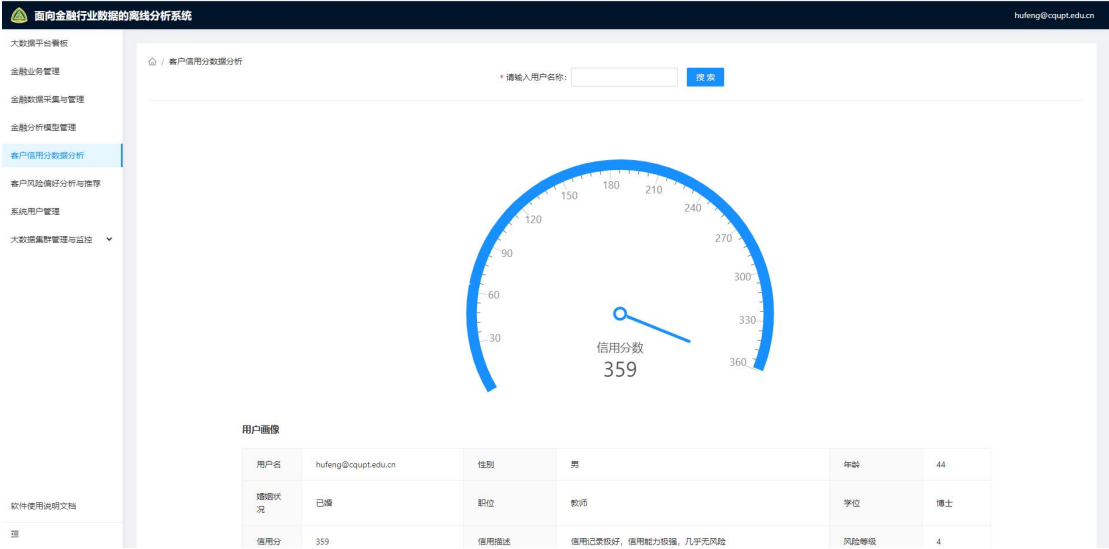
\* 业务简介: 客户信用分数数据分析

取消 确定



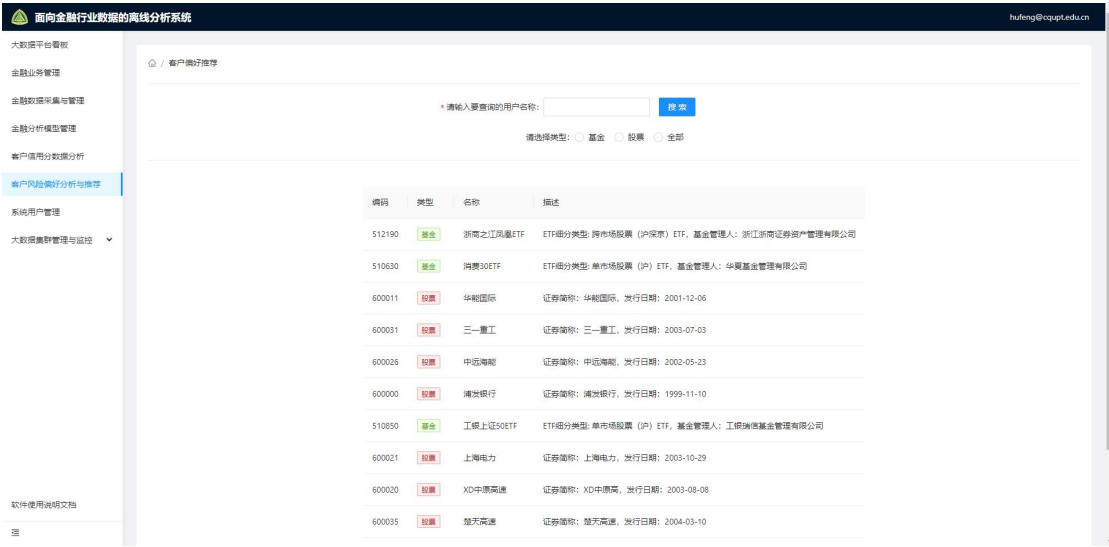
## 2.7 客户信用分数数据分析

默认展示当前登录用户的信用分信息，输入想要查询的用户名点击提交即可分析要查询客户的信用分信息。



## 2.8 客户风险偏好分析与推荐系统

默认展示当前登录用户的风险偏好分析与推荐结果，输入想要查询的用户名点击提交即可分析要查询客户的风险偏好分析与推荐信息。



## 2.9 大数据集群管理与监控

通过该功能卡可进入大数据平台管理页面

大数据集群管理与监控

HDFS

YARN

MapReduce

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Overview 'node1:8020' (active)

Started:

Thu Jul 06 21:57:38 +0800 2023

Version:

3.1.3, rba631c436b806728f8ec2f54ab1e289526c90579

Compiled:

Thu Sep 12 10:47:00 +0800 2019 by ztang from branch-3.1.3

Cluster ID:

CID-2fb35b64-9ba4-46d8-z9d4-eb553090a3e

Block Pool ID:

BP-952110410-10.16.62.146-1668532064349

Summary

Security is off.

Safemode is off.

97 files and directories, 27 blocks (27 replicated blocks, 0 erasure coded block groups) = 124 total filesystem object(s).

Heap Memory used 1.89 GB of 5.96 GB Heap Memory. Max Heap Memory is 26.67 GB.

Non Heap Memory used 82.85 MB of 84.63 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:

229.16 TB

Configured Remote Capacity:

0 B

DFS Used:

22.49 MB (0%)

Non DFS Used:

115.09 GB

DFS Remaining:

229.05 TB (99.95%)

Block Pool Used:

22.49 MB (0%)

DataNodes usages% (Min/Median/Max/stdDev):

0.00% / 0.00% / 0.00% / 0.00%

1 Data Node

3 HadoopBlock A to HadoopBlock B

hadoop

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted

0

Apps Pending

0

Apps Running

6

Apps Completed

0

Containers Running

0 B

Memory Used

12 GB

Memory Total

0 B

Memory Reserved

0

VCores Used

24

Cluster Nodes Metrics

Active Nodes

3

Decommissioning Nodes

0

Decommissioned Nodes

0

Lost Nodes

0

Unhealthy Nodes

0

Rebooted Nodes

0

Scheduler Metrics

Scheduler Type

Capacity Scheduler

Scheduling Resource Type

[memory-mb (unit=M), vcores]

Minimum Allocation

<memory:512, vCores:1>

Maximum Allocation


<memory:4096, vCores:4>

Maximum

0

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores	Allocated Memory MB	Reserved CPU Vcores	Reserved Memory MB	% of Queue
application_1688652574655_0006	root	cn.edu.cqupt.OfflineFeatureExtract	SPARK	default	0	Fri Jul 7 15:27:08 +0800 2023	Fri Jul 7 15:27:08 +0800 2023	Fri Jul 7 15:27:15 +0800 2023	FAILED	FAILED	N/A	N/A	N/A	N/A	N/A	0.0
application_1688652574655_0005	root	cn.edu.cqupt.GenerateData	SPARK	default	0	Fri Jul 7 15:25:20 +0800 2023	Fri Jul 7 15:25:20 +0800 2023	Fri Jul 7 15:25:27 +0800 2023	FAILED	FAILED	N/A	N/A	N/A	N/A	N/A	0.0
application_1688652574655_0004	root	cn.edu.cqupt.GenerateData	SPARK	default	0	Fri Jul 7 15:23:29 +0800 2023	Fri Jul 7 15:23:29 +0800 2023	Fri Jul 7 15:23:37 +0800 2023	FAILED	FAILED	N/A	N/A	N/A	N/A	N/A	0.0
application_1688652574655_0003	root	cn.edu.cqupt.GenerateData	SPARK	default	0	Fri Jul 7 14:47:14 +0800 2023	Fri Jul 7 14:47:15 +0800 2023	Fri Jul 7 14:47:22 +0800 2023	FAILED	FAILED	N/A	N/A	N/A	N/A	N/A	0.0
application_1688652574655_0002	root	cn.edu.cqupt.GenerateData	SPARK	default	0	Fri Jul 7 14:46:14 +0800 2023	Fri Jul 7 14:46:14 +0800 2023	Fri Jul 7 14:46:22 +0800 2023	FAILED	FAILED	N/A	N/A	N/A	N/A	N/A	0.0
application_1688652574655_0001	root	cn.edu.cqupt.GenerateData	SPARK	default	0	Fri Jul 7 14:28:45 +0800 2023	Fri Jul 7 14:28:46 +0800 2023	Fri Jul 7 14:28:56 +0800 2023	FAILED	FAILED	N/A	N/A	N/A	N/A	N/A	0.0



Logged in as: root

## JobHistory

Application  
About  
Jobs

Retired Jobs  
Show 20 entries

Search

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduces Total	Reduces Completed	Elapsed Time
No data available in table												

Showing 0 to 0 of 0 entries

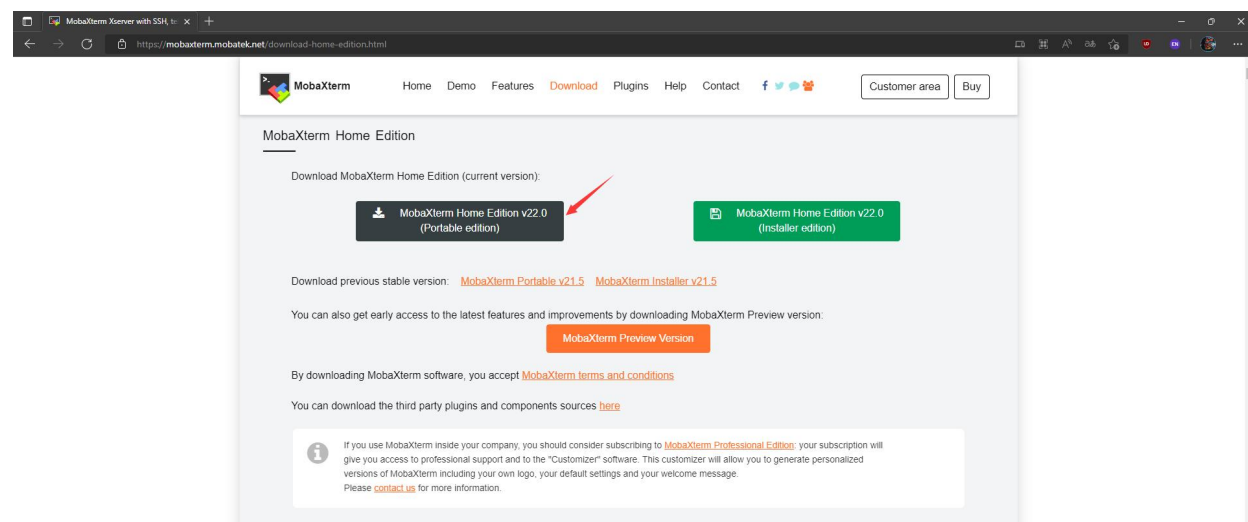
First Previous Next Last

## 3 程序启动相关

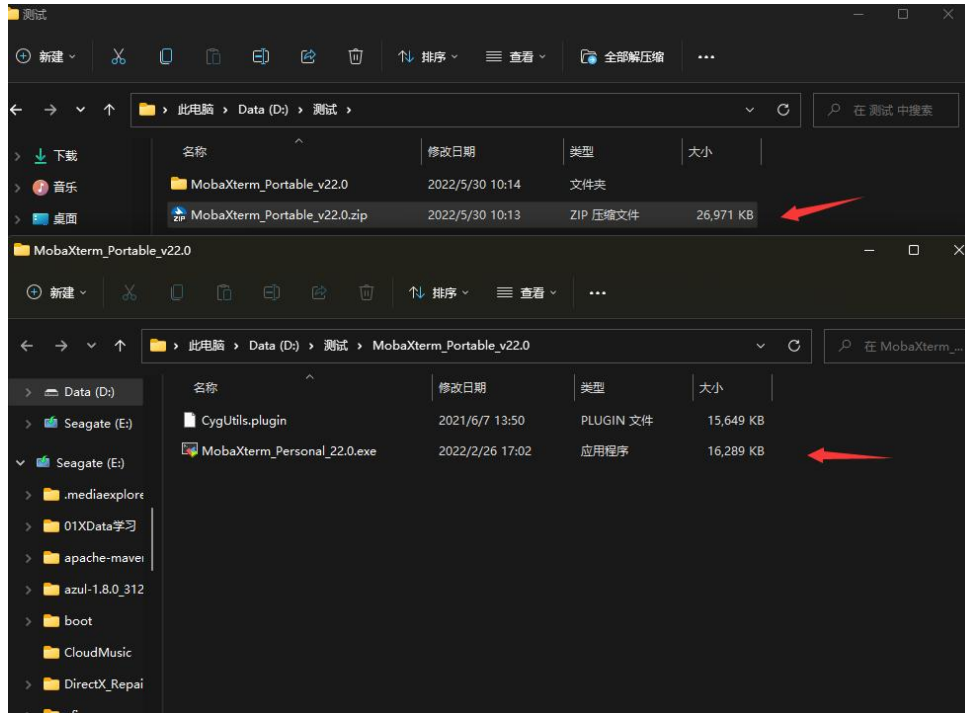
### 3.1 进入部署服务器的通用步骤

步骤 1. 首先下载一个 ssh 软件，推荐 mobaxterm

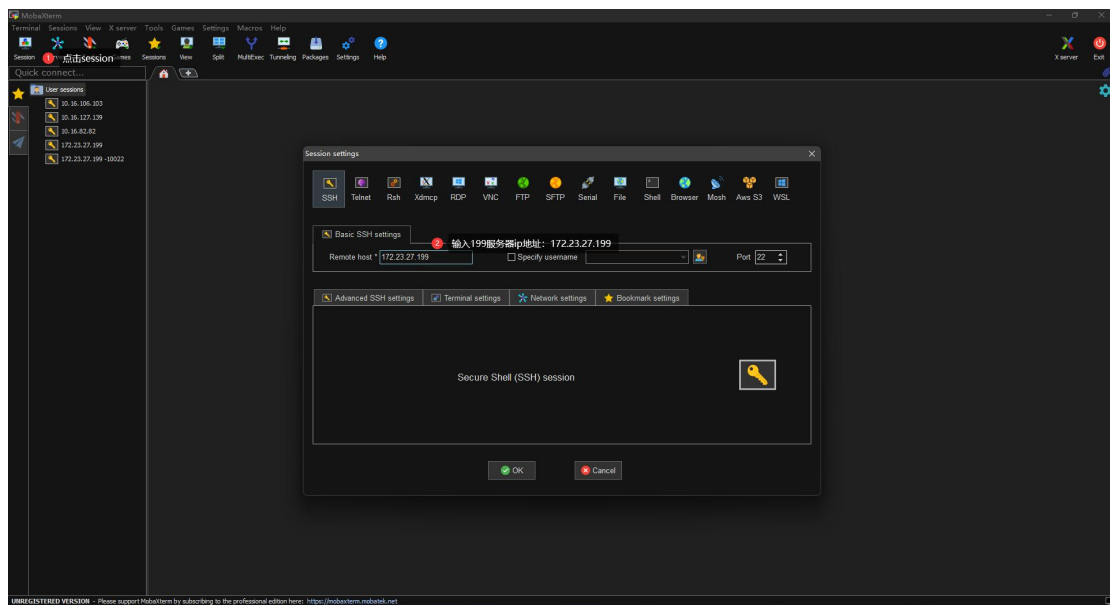
下载链接 <https://mobaxterm.mobatek.net/download-home-edition.html> [MobaXterm Xserver with SSH, telnet, RDP, VNC and X11 - Home Edition \(mobatek.net\)](https://mobaxterm.mobatek.net/download-home-edition.html)



步骤 2，下载后解压文件，得到 mobaxterm.exe 文件



步骤 3.连接 199 服务器，输入信息后点击 ok



步骤 3，输入账户和密码，账户 root，密码 1，注意输入密码的时候不会有显示，直接输入 1 后回车。

Login as: root

root@172.23.27.199's password:1

## 3.2 重启各项服务

登入进入 199 服务器后输入对应命令即可完成对应的操作。

重启后端服务命令

docker restart bigdata\_platform

重启 web 服务命令  
`docker restart bigdata_web`