

**毕 业 设 计 （论文）**

**成 果 使 用 说 明 书**



**设计(论文)题目：**基于hadoop和echarts的教育大数据可视化系统

**学生姓名：** 赵亮  **学 号：** 1513021033

**专 业：** 计算机科学与技术

**所在学院：** 计算机工程学院

**指导教师：** 谷瑞军

**职 称：** 副教授

2019 年 5 月5 日

**成果使用说明书**

# 一、运行环境

硬件环境要求

CPU:Intel I3系列以上，推荐I5系列或同等型号

内存：4GB以上内存，推荐8GB内存

硬盘：250G以上，推荐500GB硬盘或256SSD硬盘

显示器：VGA显示器或更高

软件环境要求

操作系统：Window XP/各Windows 64位系统 /Linux

开发工具：Intelli Idea/Eclipse/Mysql数据库

系统访问要求

访问工具：谷歌浏览器、火狐浏览器、360浏览器等 或IE11及以上版本的IE浏览器。

# 二、运行本系统的软件安装

## 2.1 安装RedHat Linux 7.4

（1）创建新的虚拟机，选择典型安装



图 2.1 创建新的虚拟机

（2）选择“稍后安装操作系统”



图 2.2 选择“稍后安装操作系统”

（3）选择操作系统的类型

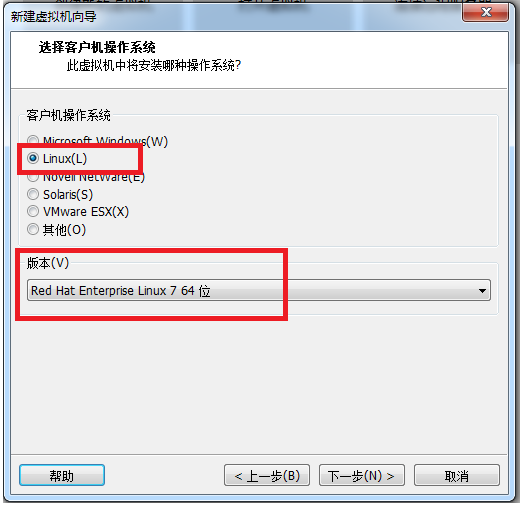


图 2.3 选择操作系统类型

（4）设置虚拟机名称和保存路径

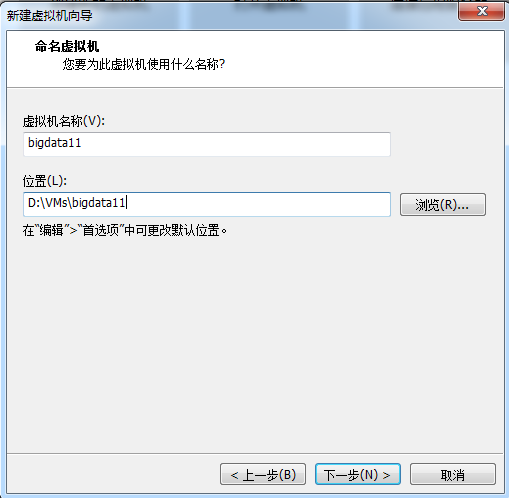


图 2.4 设置虚拟机名称和保存路径

（5）一直默认下一步即可

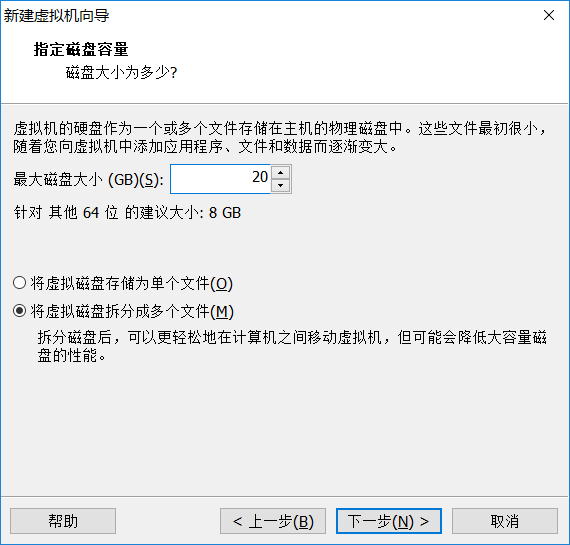


图 2.5 指定磁盘容量

（6）点击“完成”

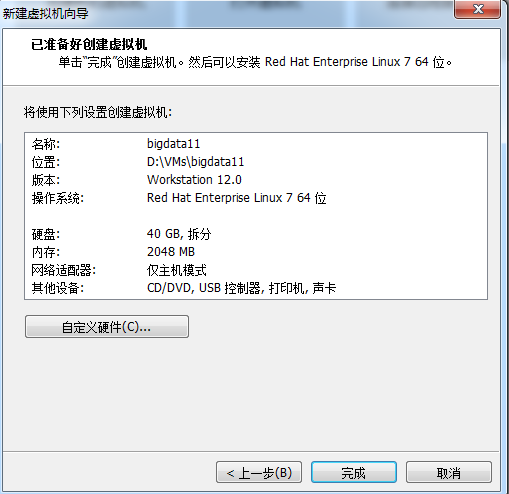


图 2.6 点击“完成”

（7）加载ISO光盘，并启动虚拟机

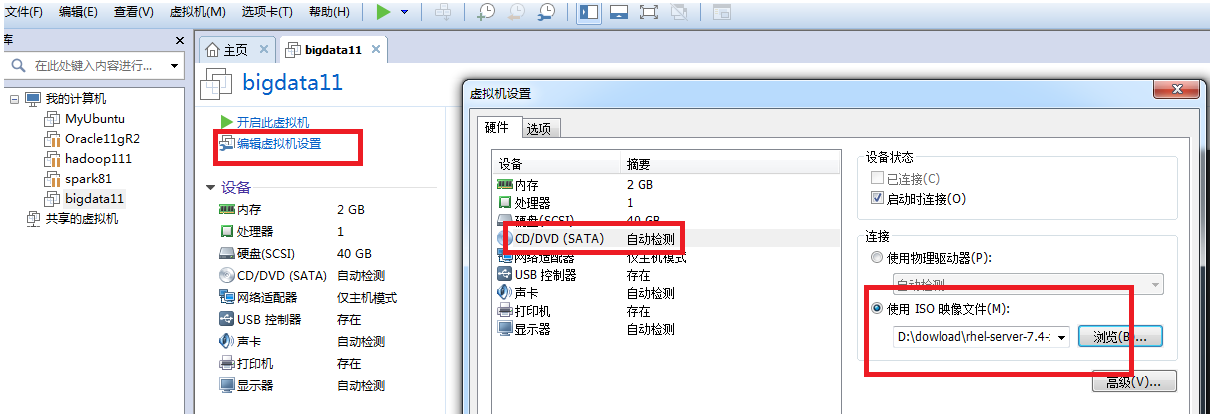


图 2.7 加载ISO光盘，启动虚拟机

（8）选择第一个选项，并回车

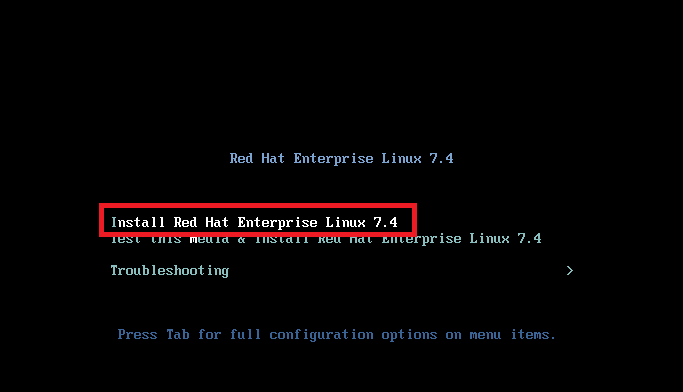


图 2.8 选择第一个选项

（9）配置安装选项

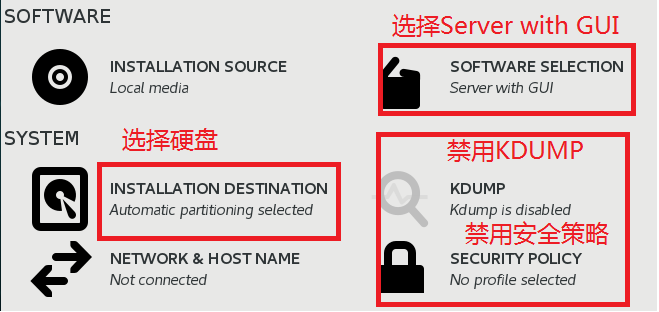


图 2.9 配置安装选项

（10）配置网络和主机名（非常重要）

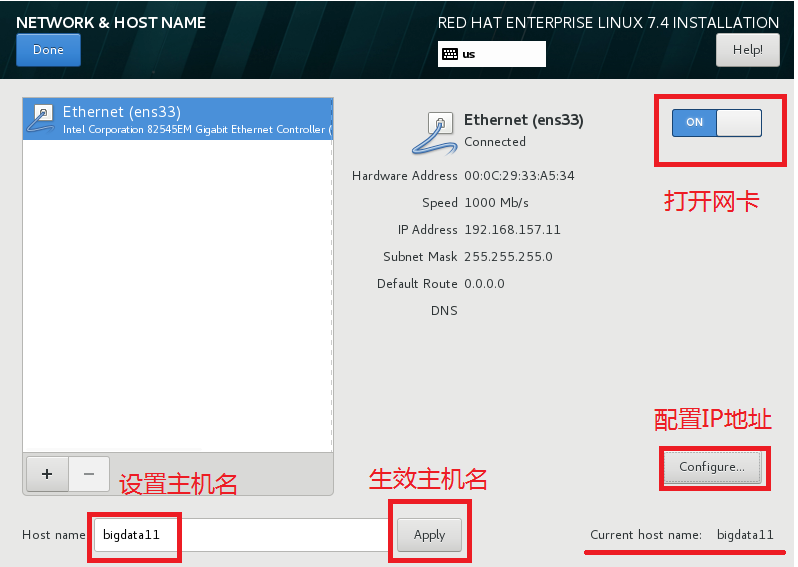


图 2.10 配置网络和主机名

（11）开始安装

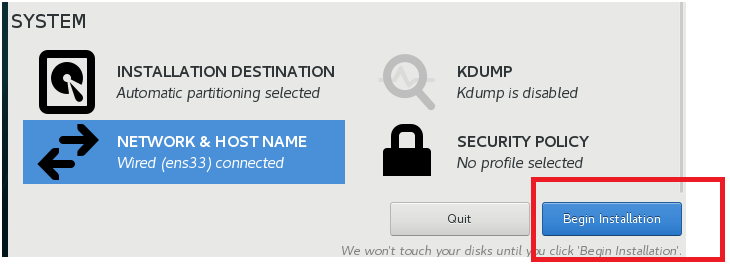


图 2.11 开始安装

（12）设置root用户的密码

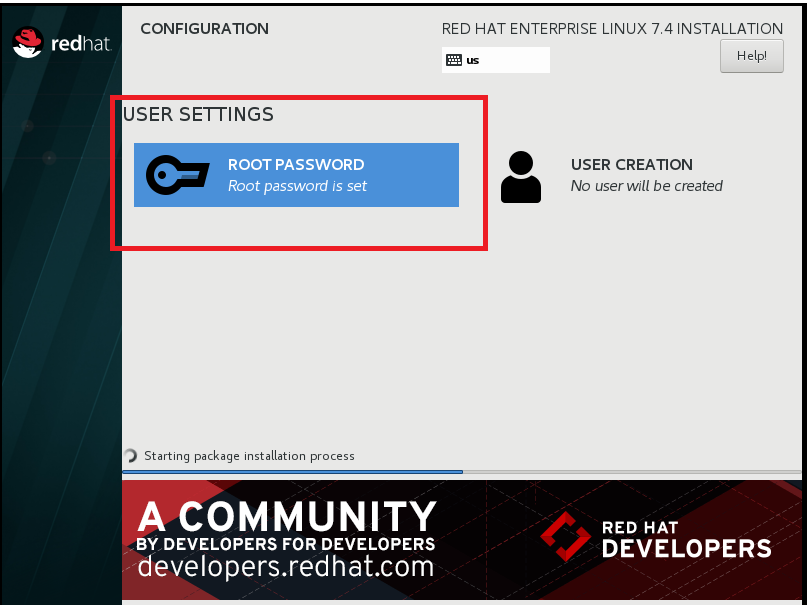


图 2.12 设置root用户的密码

（13）重启虚拟机完成配置

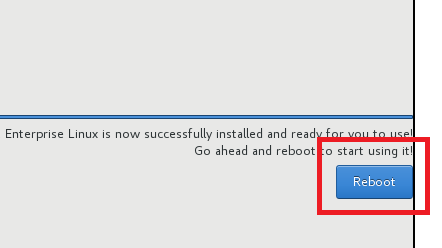


图 2.13 重启虚拟机

## 2.2 搭建伪分布式Hadoop

1、准备工作

安装Linux、JDK、关闭防火墙、配置主机名

解压：tar -zxvf hadoop-2.7.3.tar.gz -C ~/training/

设置Hadoop的环境变量: vi ~/.bash\_profile

HADOOP\_HOME=/root/training/hadoop-2.7.3

export HADOOP\_HOME

PATH=$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin:$PATH

export PATH

生效环境变量 source ~/.bash\_profile

2、伪分布式配置

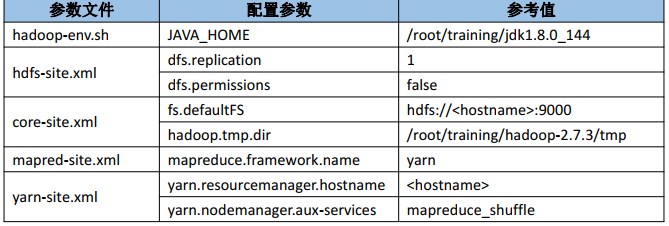


图 2.14 伪分布式配置

hdfs-site.xml

原则：一般数据块的冗余度跟数据节点（DataNode）的个数一致；最大不超过3

<!--表示数据块的冗余度，默认：3-->

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

先不设置

<!--是否开启HDFS的权限检查，默认true-->

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

core-site.xml

<!--配置NameNode地址,9000是RPC通信端口-->

<property>

<name>fs.defaultFS</name>

<value>hdfs://bigdata111:9000</value>

</property>

<!--HDFS数据保存在Linux的哪个目录，默认值是Linux的tmp目录-->

<property>

<name>hadoop.tmp.dir</name>

<value>/root/training/hadoop-2.7.3/tmp</value>

</property>

mapred-site.xml 默认没有 cp mapred-site.xml.template mapred-site.xml

<!--MR运行的框架-->

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

yarn-site.xml

<!--Yarn的主节点RM的位置-->

<property>

<name>yarn.resourcemanager.hostname</name>

<value>bigdata111</value>

</property>

<!--MapReduce运行方式：shuffle洗牌-->

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

格式化：HDFS（NameNode）

hdfs namenode -format

日志：Storage directory /root/training/hadoop-2.7.3/tmp/dfs/name has been successfully formatted.

启动停止Hadoop的环境：start-all.sh、stop-all.sh

访问：通过Web界面，HDFS: <http://192.168.157.111:50070>

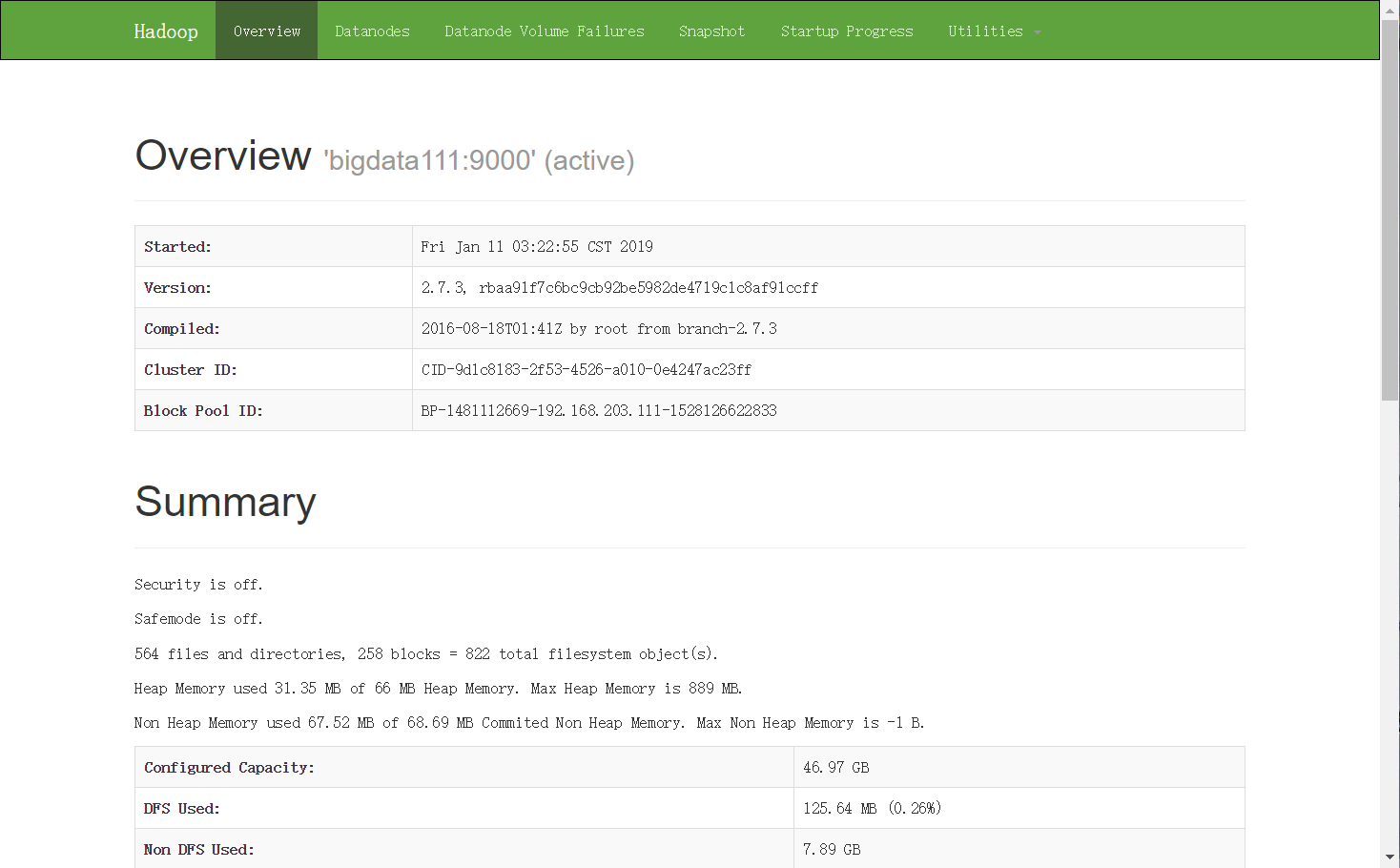


图 2.15 Hadoop欢迎页面

## 2.3 安装HBase

1、准备工作

tar -zxvf hbase-1.3.1-bin.tar.gz -C ~/training/

设置环境变量 vi ~/.bash\_profile

HBASE\_HOME=/root/training/hbase-1.3.1

export HBASE\_HOME

PATH=$HBASE\_HOME/bin:$PATH

export PATH

生效环境变量 source ~/.bash\_profile

2、伪分布式配置

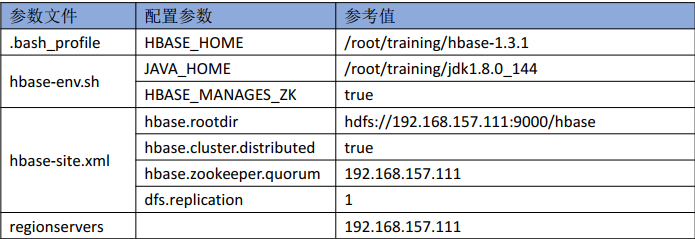


图 2.16 伪分布式配置

修改文件：hbase-env.sh

HBASE\_MANAGES\_ZK true ---> 使用HBase自带的ZK

核心配置文件: conf/hbase-site.xml

<!--HBase的数据保存在HDFS对应目录-->

<property>

<name>hbase.rootdir</name>

<value>hdfs://192.168.157.111:9000/hbase</value>

</property>

<!--是否是分布式环境-->

<property>

<name>hbase.cluster.distributed</name>

<value>true</value>

</property>

<!--配置ZK的地址-->

<property>

<name>hbase.zookeeper.quorum</name>

<value>192.168.157.111</value>

</property>

<!--冗余度-->

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

文件regionservers：配置从节点地址192.168.157.111

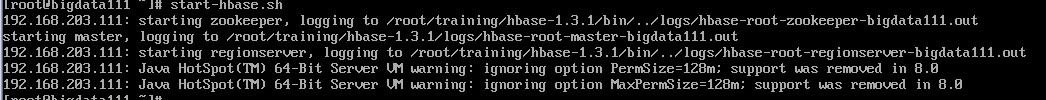


图 2.17 HBase启动成功

# 三、程序运行

1、打开项目并运行

（1）打开IDEA。



图 3.1 IDEA图标

（2）打开后，点击菜单栏上的File->New->Module from exising sources...，选择项目文件夹下的POM.xml文件，工具会自动加载项目配置。

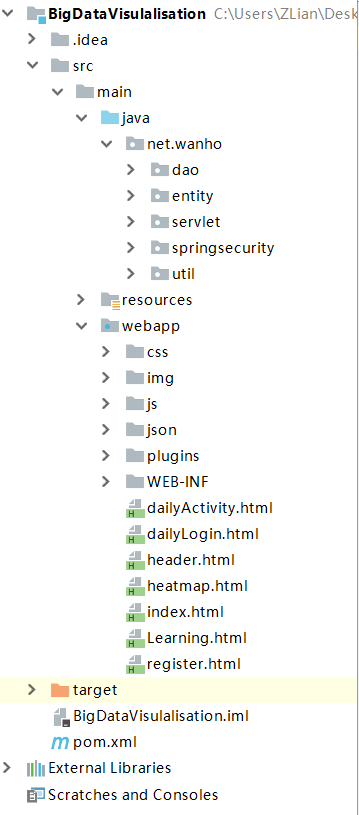


图 3.2 项目文件界面

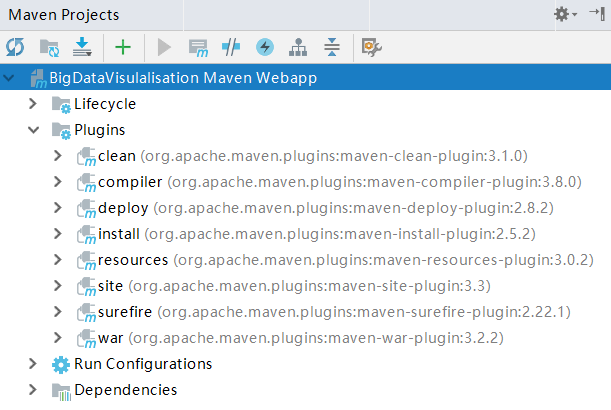
（3）打开项目后，点击右侧Maven Project标签下INSTALL的操作。

图 3.3 maven install

（4）接着在Edit Configuration 处 添加MAVEM，将项目添加，运行即可。

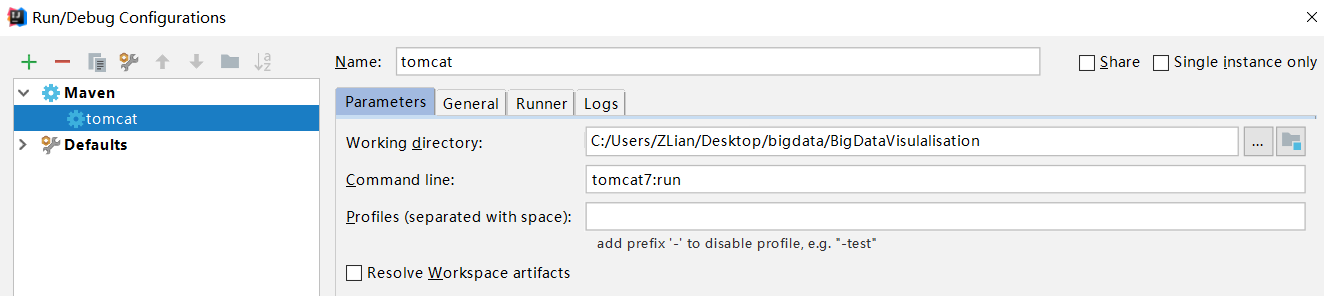


图 3.4 tomcat 添加

2、运行CAS

（1）打开apache-tomcat-cas文件中bin目录中startup.bat，运行CAS。



图 3.5 运行CAS

（2）运行成功

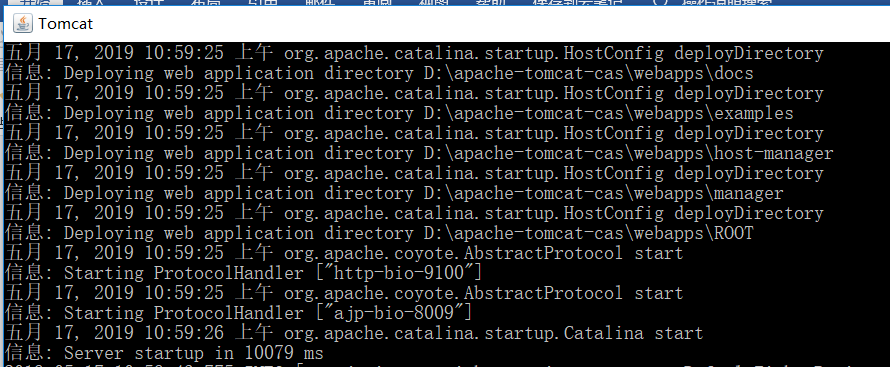


图 3.6 CAS启动成功

# 四、运行界面

本系统提供给用户查看平台中数据，首先要求用户简单注册以登录系统查看页面数据，注册过程中需要提供手机号进行短信验证。用户成功登录后通过点击导航栏中不同的按钮，可以查看不同数据维度的图表展示。



图 4.1 系统界面

# 五、系统各模块的实现

## 5.1 注册界面



图 5.1 注册界面

当新用户没有账号登录时徐注册账号，注册时输入用户名昵称和密码，获取手机短信验证码验证功能。



图 5.2 注册验证

功能上进行输入检查和验证，密码二次输入一致性检查，短信验证码是否正确，密码以md5加密的密文形式存储数据库中。

## 5.2 登录界面

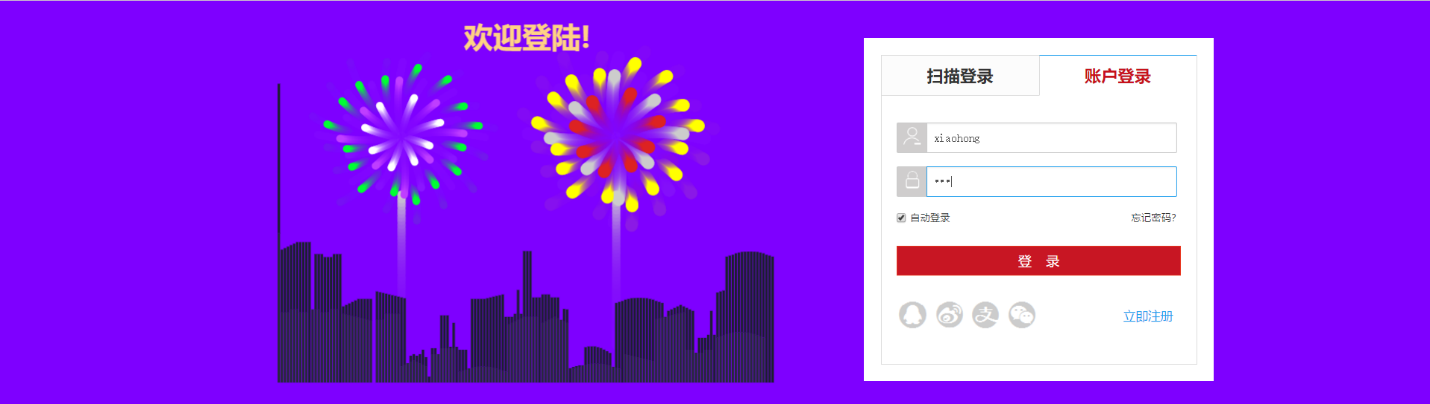


图5.3 登录界面

在浏览器中输入系统地址<http://localhost:9004>时，系统会验证cas中是否有用户已登录，如果没有则跳转登录页面<http://localhost:9100/cas/login?service=http%3A%2F%2Flocalhost%3A9004%2Flogin%2Fcas>，有则直接跳转首页。

登录页面的动画效果是手写的echarts图标效果，功能上同样有用户名和密码的验证并提示信息。

## 5.3 首页界面



图5.4 首页界面

用户登录成功后，系统将用户名并存储在用户浏览器的COOKIE中，在导航栏的右上角显示：欢迎，xxx!，当鼠标悬停在用户名上，下方会显示退出登录，点击退出登录，系统重新跳转到登录界面。

首页页面读取HBase中处理好的数据，展示平台中课程数、学生数，以及课程上课受欢迎度降序排序，下方采用仪表盘开发**错误!未找到引用源。**，反映平台每周的活跃人数和周活跃平台健康度。在仪表盘的右上方有工具配置项，可供下载和装换数据视图。

## 5.4 每日登陆人数分析界面



图5.5 每日登陆人数分析界面

此页面左侧饼图显示每月登录人数比，数据库中只有四月和五月相关数据。右侧默认显示四月每日登陆人数信息。当点击下拉按钮后，系统展示所选月份相关数据。

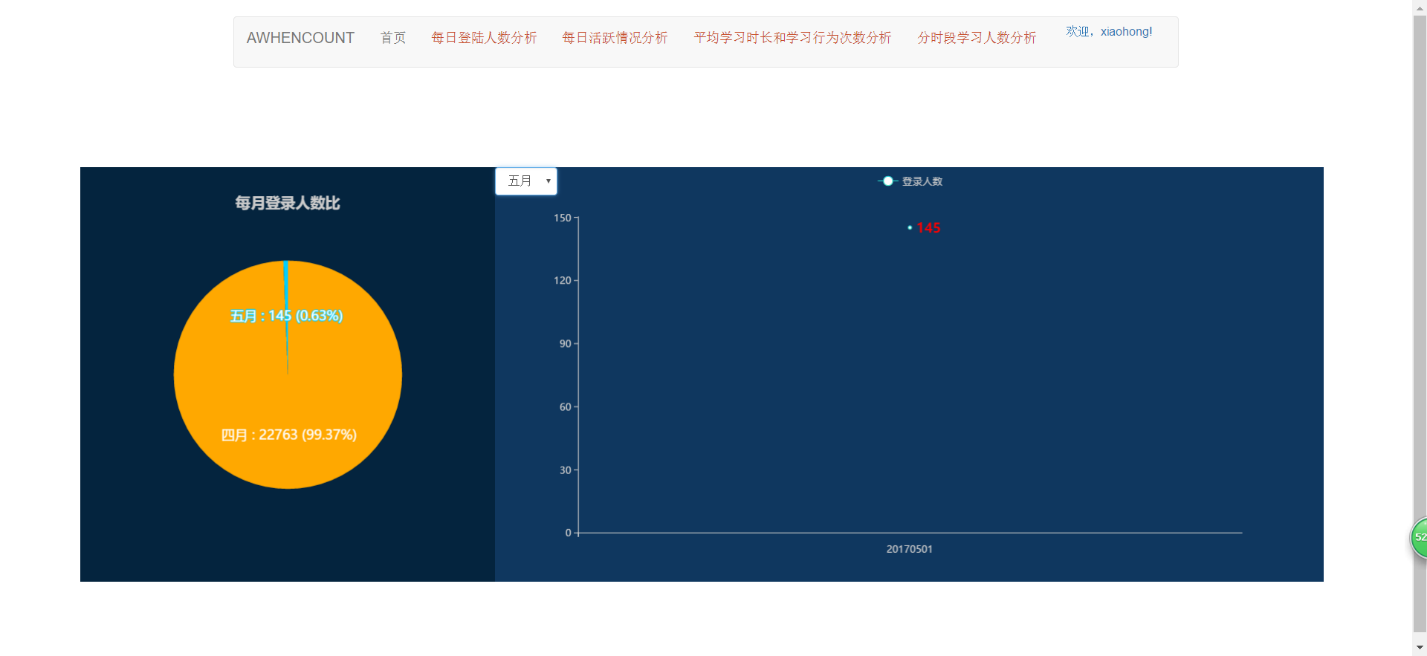


图5.6 选择按钮切换数据界面

## 5.5 每日活跃情况分析界面



图 5.7 每日活跃清空分析页面

这里设定每日至少进行3次学习行为的用户为活跃用户。

该页面通过柱状图显示每天的活跃学生情况，第一层以月份形式展示。当点击201705所对应的柱状图时，进行数据下砖，第二层以当月的每天形式展示，点击放回按钮，又返回第一层数据。



图5.8 月份下钻界面

## 5.6 平均学习时长和学习行为次数分析界面

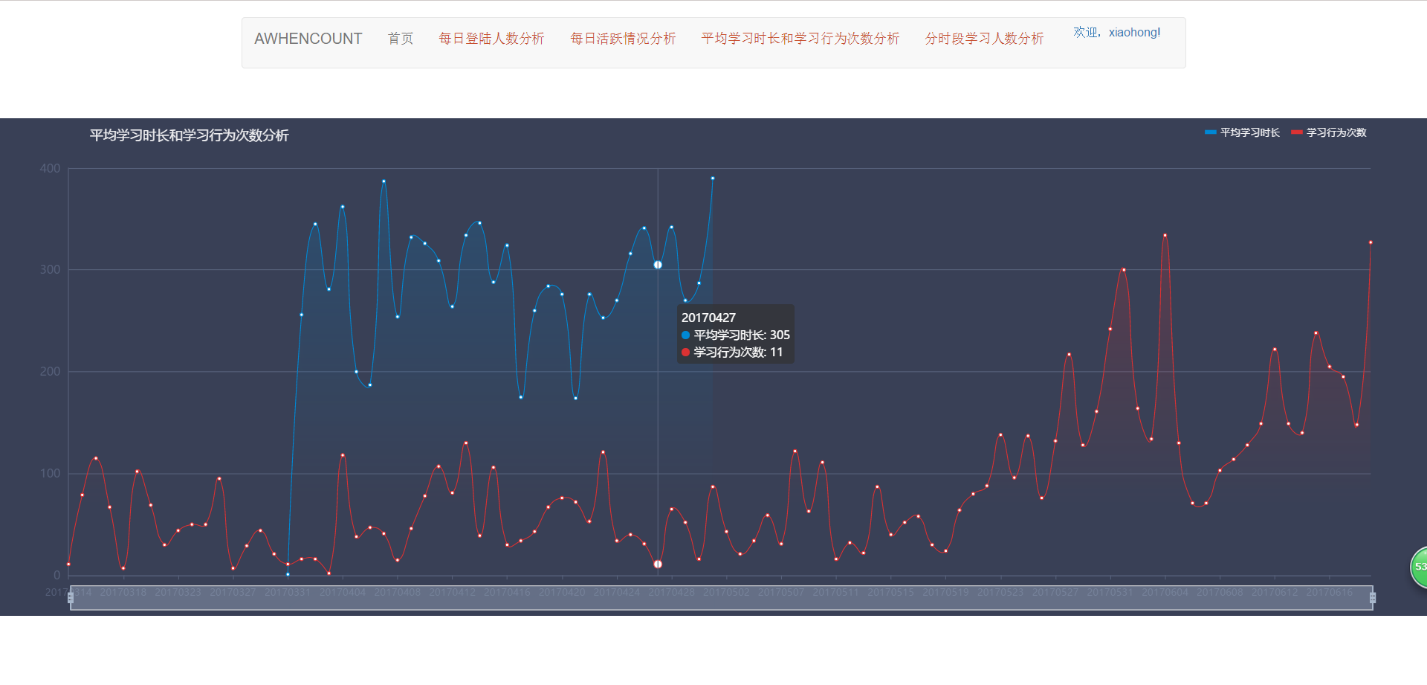


图 5.9 平均学习时长和学习行为次数分析界面

此页面展示了平台中每条学生的平均学习时长和行为次数，可以从图中看出这二者关系正相关，当学习行为次数增加时，平台中学习时长就增加。

## 5.7 分时段学习人数分析界面

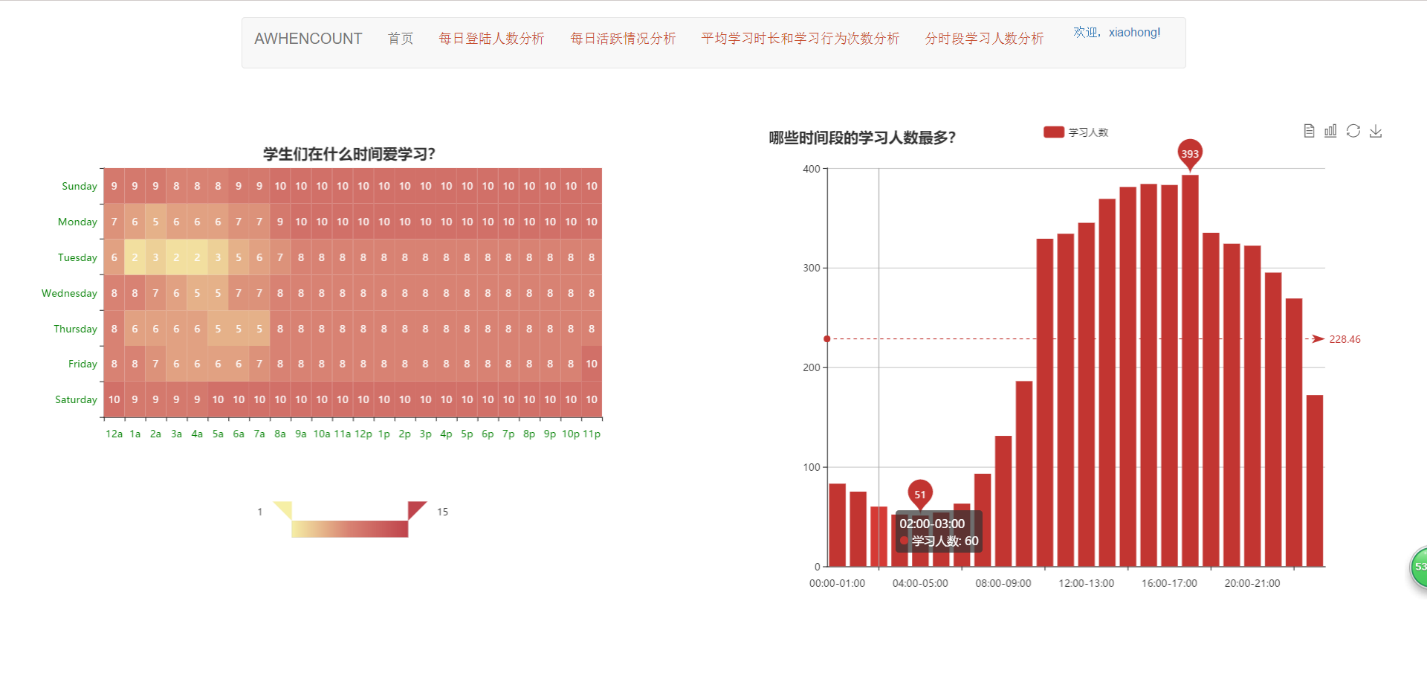


图 5.10 分时段学习人数分析页面

该页面左侧以热力图的图标展示学生们在什么时间段中学习，可以从图表中纵坐标看出在一星期中的星期一、星期六、星期日学习人数最多，从横坐标可以看出，在每天的0-8点学习人数比白天人数少很多，这符合人们的学习作息时间。

页面右侧细化分一天中每个时间段学习人数在下午的2点-6点，这个时间段中学习人数最多，晚上和早上人数次之，1点-7点人数最少。图表的右上角同样实现柱状图与折线图的转换。

## 5.8 大屏可视化界面

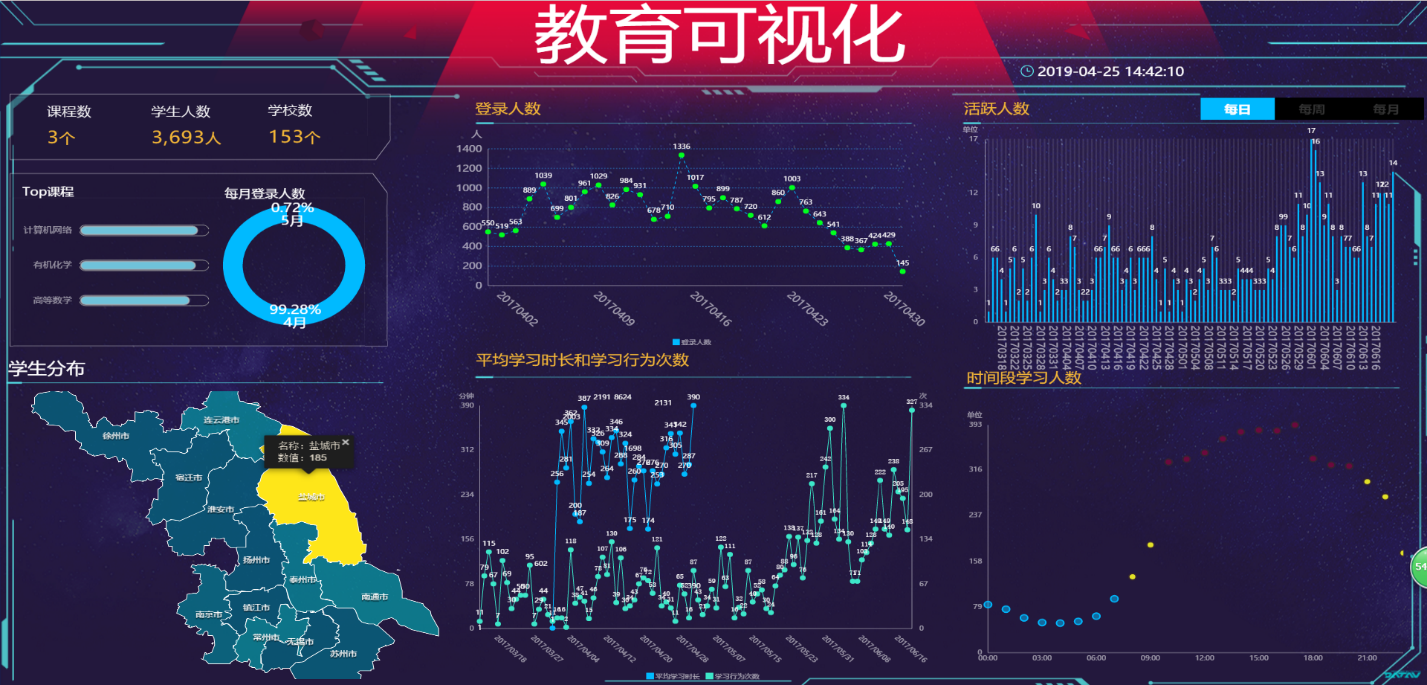


图 5.11 大屏可视化页面

该页面使用阿里云DataV数据可视化平台，同样实现相同的功能，点击每日，每周，每月按钮，对应折线图，柱状图图表会联动显示数据。学生分布处使用地图显示学生的地区分布人数，以南通市学生人数为榜首。在时间段学习人数处使用气泡图展示，图中恰好展示学习人数在不同时间段中的分布。

使用DataV数据可视化平台，可以看出数据展示效果更加紧凑和大屏数据的显示，看到数据可视化的魅力，可以帮助非专业的工程师通过图形化的界面轻松搭建专业水准的可视化应用多种业务的展示需求。

# 六、系统测试

系统测试是为了更早的发现错误并完善系统，没有任何系统及软件做到0错误。想成为一个优秀的开发者，开发和测试不分家，应把“尽早地不断地进行软件测试”立于心中，并且在开发中做好最坏的打算，写一步测一步。一个完整的系统的开发流程中，开发和测试都是一一对应，编码完成后就要及时完成单元测试，详细设计时就要写好测试用例供功能测试时测试等。测试方式常见与黑盒测试和白盒测试。

前者不考虑程序内部结构和逻辑结构，以准备好的测试用例测试模块的输入和输出，是否满足预期结果和实际结果一致。

后者主要应用在单元测试阶段，开发者自己对方法中代码的自我测试，主要是对代码级的测试，针对程序内部逻辑是否异常，增加程序的健壮性。

系统测试是在经过以上各阶段测试确认之后，把系统完整地模拟客户环境来进行的测试。使其开发的软件产品符合客户的需求。

根据预计的目的和方法，测试方案如下所示：

（1）首先要尽快进行连续系统测试。我们知道，越快发现错误，它的损失也就越低。但是因为程序的错误通常难以避免，所以系统测试并非是一个独立开放阶段，它是从始至终贯穿整个开放过程的。

（2）通常我们在开发测试方案时，遵循一个判断标准：可以先利用输入数据来获取程序运行后的测试结果，然后可以将整个结果的数据与我们之前预测的数据进行比较，看看看这两个数据是否一致。

（3）档我们在进行测试时，要兼顾两种测试输入的数据，一种是合法的输入数据，另一种是非法的即错误的数据。这个原理是什么呢？其实在平时的系统程序日常实际使用的过程中，用户可能会因为某种原因，在里面输入一些错误的非法的数据，比如错误的密钥或非法使用的命令，这个过程可能会导致系统出现bug，严重时可能后引起系统的奔溃，因此，在这个测试方面，我们要尤其要关注非法测试，加多测试次数。