编码：项目级-软件配置库-Version\_2.0.0-release-20200813

文档等级：□发布级 ■项目级 □团队级

采纳标准：团队文档质量标准

网络文学大数据分析系统

**Big Data Analysis System of Network Literature**

软件配置库

**Version 2.0.0-release**

编写人：张婷睿

西工大-华迪实训第四小组

**All Rights Reserved**

**目 录**

**[1引言 3](#_Toc19203)**

[1.1 标识 3](#_Toc24279)

[1.2 系统概述 3](#_Toc21041)

[1.3 文档概述 3](#_Toc22673)

[1.4 基线 4](#_Toc7513)

**[2 运行环境 4](#_Toc7091)**

[2.1 硬件环境 4](#_Toc32067)

[2.2 软件环境 4](#_Toc15019)

**[3 部署方案 5](#_Toc13403)**

[3.1 总体部署说明 5](#_Toc26886)

[3.2 环境配置流程 5](#_Toc32645)

[3.2.1 JDK配置 5](#_Toc4470)

[3.2.2 安装SSH、配置SSH无密码登录 7](#_Toc17027)

[3.2.3 安装Hadoop-2.8.1 7](#_Toc1326)

[3.2.4 Spark-2.2.0环境搭建 8](#_Toc23660)

[3.2.5 MySQL环境搭建 9](#_Toc22245)

[3.2.6 PyCharm安装 10](#_Toc3341)

# 1引言

## 标识

适用系统：网络文学大数据分析系统

中文名称：软件配置库（SCL）

英文名称：Software Configuration Library (SCL)

文档版本：2.0.0

文档编号：项目级-软件配置库-Version\_2.0.0-release-20200813

## 系统概述

本文档适用于“网络文学大数据分析系统”项目的开发过程。网络文学大数据分析系统项目由西北工业大学软件学院提出，由西工大-华迪实训第四小组成员负责实施，该项目编号为项目级-软件配置库-Version\_2.00-release-20200813，其软件产品版本号为“1.0”。

该项目基于B/S（Brower/Server）浏览器/服务器结构，采用传统的三层结构方式进行解耦开发。对当下最热门的网络文学阅读机构——“晋江文学”提供的各项数据进行可视化展示和分析。数据分析采用Hadoop和Spark。

## 文档概述

本文档依据《GB/T 8567-2006 计算机软件文档编制规范》制定，属于技术文档，仅限于项目相关人员阅读。

本文档记录与配置项相关的所有信息，是配置管理的有力工具，查询该文档利用库中的信息可回答许多配置管理的问题。本文档处于项目计划过程的需求分析阶段，使用配置库管理文档可以帮助配置管理员把信息系统开发过程的各种工作产品，包括半成品或阶段产品和最终产品管理得井井有条，使其不致管乱、管混、管丢。

本文档的保密范围如下：

①开发人员

②测试阶段人员

③对本文档进行评审的人员或机构

④项目组以及其他有权需要调用本文档的人员

## 基线

基线：GB/T 8567-2006

# 运行环境

## 硬件环境

(1)CPU：Intel CoreI5 1.8GHz 及以上

(2)内存：8G 及以上

(3)硬盘：60G 及以上

## 软件环境

(1)服务器： Tomcat 8.0

(2)操作系统：Ubuntu 16.04 LTS 或 Ubuntu 18.04 LTS

(3)数据库：MySQL 6.5

(4)基本配置：Spark 2.2.0（Built for Hadoop 2.8.1）伪分布式环境、JDK 1.8、Scala 2.12.2

(5)开发工具：Intellij IDEA 2017.3.5 及以上

(6)PC 端：Chrome21.0及以上等的浏览器

# 部署方案

## 总体部署说明

系统的运行环境为ubuntu。系统整体分为三部分，分别是：数据预处理、数据挖掘分析和数据可视化。数据预处理部分使用scala语言编写，将原始数据和静态数据进行清洗和处理，存入HBase数据库中。数据挖掘分析部分使用Scala语言和python语言进行编写，将收集到的数据进行挖掘分析并把结果存储在MySQL数据库中。数据可视化部分则通过前端轮询ajax数据库，在不同页面上对数据进行可视化。

## 环境配置流程

### JDK配置

（1）访问官网，下载JDK，下载1.8版本。

（2）在Linux命令行界面中，执行如下Shell命令（注意：当前登录用户名是hadoop）：

1. cd /usr/lib
2. sudo mkdir jvm *#创建/usr/lib/jvm目录用来存放JDK文件*
3. cd ~ *#进入hadoop用户的主目录*
4. cd Downloads *#注意区分大小写字母，刚才已经通过FTP软件把JDK安装包jdk-8u162-linux-x64.tar.gz上传到该目录下*
5. sudo tar -zxvf ./jdk-8u162-linux-x64.tar.gz -C /usr/lib/jvm *#把JDK文件解压到/usr/lib/jvm目录下*

（3）JDK文件解压缩以后，可以执行如下命令到/usr/lib/jvm目录查看一下：

1. cd /usr/lib/jvm
2. ls

可以看到，在/usr/lib/jvm目录下有个jdk1.8.0\_162目录。  
（4）下面继续执行如下命令，设置环境变量：

1. cd ~
2. vim ~/.bashrc

上面命令使用vim打开了hadoop这个用户的环境变量配置文件，请在这个文件的末尾位置，添加如下几行内容：

export JAVA\_HOME=/usr/lib/jvm/jdk1.8.0\_112

export JRE\_HOME=${JAVA\_HOME}/jre

export CLASSPATH=.:${JAVA\_HOME}/lib:${JRE\_HOME}/lib

export PATH=${JAVA\_HOME}/bin:$PATH

保存.bashrc文件并退出vim编辑器。然后，继续执行如下命令让.bashrc文件的配置立即生效：

source ~/.bashrc

这时，可以使用如下命令查看是否安装成功：

java -version

如果能够在屏幕上返回如下信息，则说明安装成功：

hadoop@ubuntu:~$ java -version

java version "1.8.0\_112"

Java(TM) SE Runtime Environment (build 1.8.0\_112-b12)

Java HotSpot(TM) 64-Bit Server VM (build 25.112-b12, mixed mode)

### 安装SSH、配置SSH无密码登录

（1）集群、单节点模式都需要用到 SSH 登陆，Ubuntu 默认已安装了 SSH client，此外还需要安装 SSH server：

sudo apt-get install openssh-server

（2）安装后，可以使用如下命令登陆本机：

ssh localhost

（3）此时会有如下提示(SSH首次登陆提示)，输入 yes 。然后按提示输入密码 hadoop，登陆到本机。

（4）配置SSH无密码登录。

利用 ssh-keygen 生成密钥，并将密钥加入到授权中：

1. exit *# 退出刚才的 ssh localhost*
2. cd ~/.ssh/ *# 若没有该目录，请先执行一次ssh localhost*
3. ssh-keygen -t rsa *# 会有提示，都按回车就可以*
4. cat ./id\_rsa.pub >> ./authorized\_keys *# 加入授权*

此时再用 ssh localhost 命令，无需输入密码就可以直接登录。

### 安装Hadoop-2.8.1

Hadoop 2可以通过 [http://mirror.bit.edu.cn/apache/hadoop/common/](http://mirror.bit.edu.cn/apache/hadoop/common/" \t "_blank) 或者 [http://mirrors.cnnic.cn/apache/hadoop/common/](http://mirrors.cnnic.cn/apache/hadoop/common/" \t "_blank) 下载，一般选择下载最新的稳定版本，即下载 “stable” 下的 **hadoop-2.x.y.tar.gz** 这个格式的文件，另一个包含 src 的则是 Hadoop 源代码，需要进行编译才可使用。

将 Hadoop 安装至 /usr/local/ 中：

1. sudo tar -zxf ~/下载/hadoop-2.8.1.tar.gz -C /usr/local *# 解压到/usr/local中*
2. cd /usr/local/
3. sudo mv ./hadoop-2.6.0/ ./hadoop *# 将文件夹名改为hadoop*
4. sudo chown -R hadoop ./hadoop *# 修改文件权限*

Hadoop 解压后即可使用。输入如下命令来检查 Hadoop 是否可用，成功则会显示 Hadoop 版本信息：

1. cd /usr/local/hadoop
2. ./bin/hadoop version

### Spark-2.2.0环境搭建

（1）访问官网，获得Spark2.2.0的安装包.

（2）安装Spark（Local模式）

1. sudo tar -zxf ~/下载/spark-2.2.0-bin-without-hadoop.tgz -C /usr/local/
2. cd /usr/local
3. sudo mv ./spark-2.2.0-bin-without-hadoop/ ./spark
4. sudo chown -R hadoop:hadoop ./spark *# 此处的 hadoop 为你的用户名*

（3）安装后，还需要修改Spark的配置文件spark-env.sh

1. cd /usr/local/spark
2. cp ./conf/spark-env.sh.template ./conf/spark-env.sh

编辑spark-env.sh文件(vim ./conf/spark-env.sh)，在第一行添加以下配置信息:

export SPARK\_DIST\_CLASSPATH=$(/usr/local/hadoop/bin/hadoop classpath)

（4）添加spark环境变量

vi ~/.bashrc

添加：

export SPARK\_HOME=/usr/local/spark

export PYTHONPATH=$SPARK\_HOME/python:$SPARK\_HOME/python/lib

/py4j-0.10.4-src.zip:$PYTHONPATH

export PYSPARK\_PYTHON=python3

export PATH=$HADOOP\_HOME/bin:$SPARK\_HOME/bin:$PATH

（5）通过运行Spark自带的示例，验证Spark是否安装成功。

### MySQL环境搭建

（1）使用以下命令即可进行mysql安装，注意安装前先更新一下软件源以获得最新版本：

1. sudo apt-get update *#更新软件源*
2. sudo apt-get install mysql-server *#安装mysql*

上述命令会安装以下包：  
 apparmor  
 mysql-client-5.7  
 mysql-common  
 mysql-server  
 mysql-server-5.7  
 mysql-server-core-5.7  
（2）安装过程会提示设置mysql root用户的密码，设置完成后等待自动安装即可。默认安装完成就启动了mysql。

* 启动和关闭mysql服务器：

1. service mysql start
2. service mysql stop

* 确认是否启动成功，mysql节点处于LISTEN状态表示启动成功：

sudo netstat -tap | grep mysql

* 进入mysql shell界面：

mysql -u root -p

### PyCharm安装

**（1）下载Pycharm**

PyCharm的下载地址（[Linux版本](https://www.jetbrains.com/pycharm/download/" \l "section=linux" \t "_blank)）。

**（2）安装PyCharm**

执行以下命令解压文件：

1. cd ~/下载
2. tar -xvf pycharm-professional-2018.2.4.tar.gz

在下载目录看到一个pycharm-2018.2.4文件夹。接下来我们把它放到/usr/local下，并且重命名

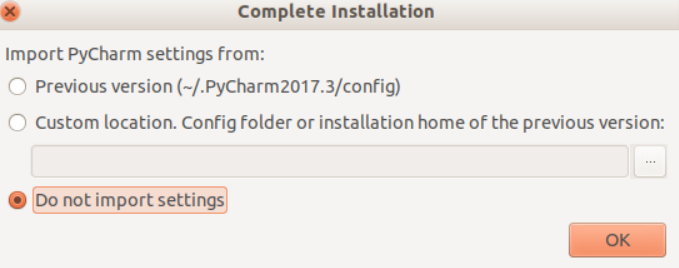
sudo mv ./pycharm-2018.2.4 /usr/local/pycharm

然后我们要执行pycharm.sh文件，完成首次安装：

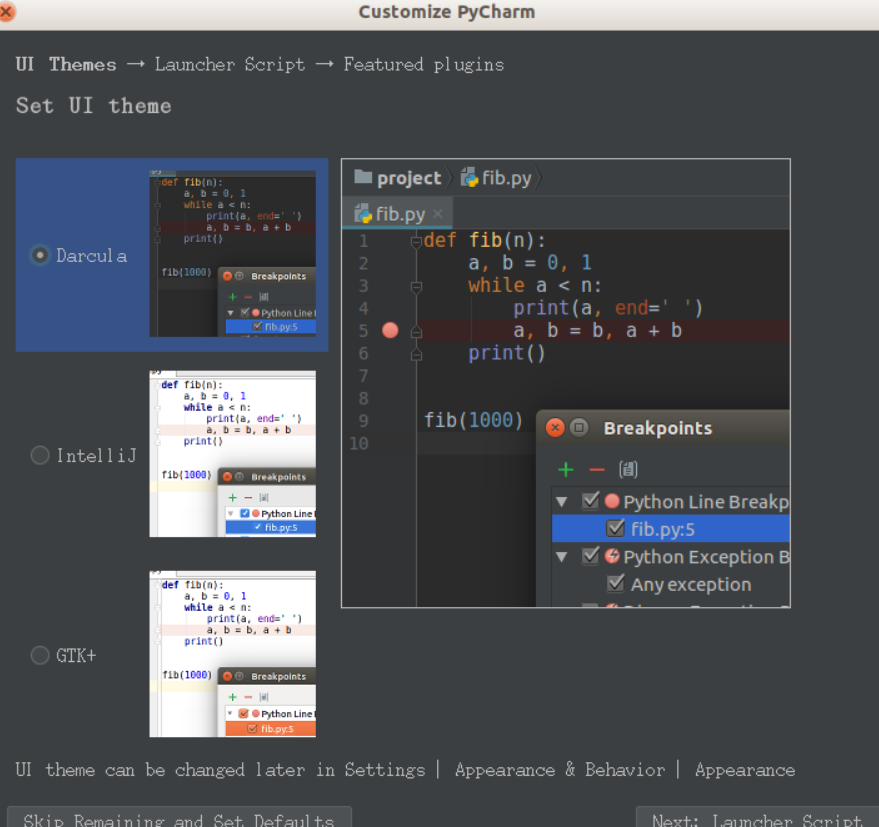
cd /usr/local/pycharm/bin

./pycharm.sh

等待之后我们可以看到如下图界面：



选择不导入设置，点击OK。然后我们会看到以下界面：



选择左下角“Skip Remaining and Set Defaults”，默认设置即可

**（3）配置环境变量**

配置环境变量的意义在于我们以后不需要每次都到pycharm文件夹下去启动程序。

sudo vim ~/.bashrc

将下面内容复制到文件的开头部分。

#pycharm

export PyCharm\_HOME=/usr/local/pycharm

export PATH=${PyCharm\_HOME}/bin:$PATH

完成以上操作后你就可以在终端直接使用：pycharm.sh命令打开程序了。