# 一、系统开发环境

Linux系统配置项

|  |  |
| --- | --- |
| 配置项 | 配置内容 |
| Hadoop | 2.7.3 |
| Flume | 1.8 |
| Kafka | 2.4 |
| Scala | 2.11 |
| JDK | 1.8 |
| Python | 3.5.1 |
| ZooKeeper | 3.6.1 |

windows系统配置项

|  |  |
| --- | --- |
| 配置项 | 配置内容 |
| Hadoop | 2.7.3 |
| Spark | 2.3.3 |
| maven | 3.6.1 |
| Scala | 2.11 |
| Redis | 5.1 |
| JDK | 1.8 |
| Python | 3.7.2 |

开发工具

IDEA 2021

PyCharm 2021

Python第三方库(pip 安装)

|  |  |
| --- | --- |
| 配置项 | 版本 |
| django | 3.1.7 |
| jieba | 0.42.1 |
| pandas | 1.3.5 |
| genism | 4.0.1 |
| pyLDAvis | 3.2.2 |
| redis | 1.8 |
| pyyaml | 3.13 |
| pymysql | 0.10.1 |
| platform | 1.0.8 |
| psutil | 5.8.0 |

# 二、相关组件安装步骤

## 2.1 Hadoop安装

1、安装好虚拟机后安装Linux系统，以ubuntu16.04为例

2、关闭防火墙

3、安装JDK1.8

4、安装ssh-server

5、解压hadoop安装包，配置环境

## 2.2 Spark安装

windows下安装

1、安装jdk 1.8

2、安装python 3.7.2

3、安装scala 2.11

4、解压配置Hadoop

5、解压配置spark

## 2.3 Zookeeper安装

Linux下单节点安装

1、解压

tar -zxvf apache-zookeeper-3.6.1-bin.tar.gz

2、备份

cd conf

cp zoo\_sample.cfg zoo.cfg

3、修改配置文件

vim zoo.cfg

#修改默认存储路径和日志路径。

dataDir=/data2/zk/data

dataLogDir=/data2/zk/logs

## 2.4 Flume安装

1、解压

sudo tar xzvf apache-flume-1.8.0-bin.tar.gz

2、备份

cd /opt/flume-1.8.0/conf/

mv flume-env.sh.template flume-env.sh

3、配置环境变量

export FLUME\_HOME=/opt/flume-1.8.0

export PATH=$PATH:$FLUME\_HOME/bin

## 2.5 Kafka安装

Linux下单节点安装

1. 解压

tar -zxvf kafka\_2.11-2.4.0.tgz

## 2.6 Redis安装

Windows 安装

1. 下载
2. 解压