

4、部署Python节点集群

- 1、系统环境准备
 - 1.1 防火墙配置
- 2、部署Python计算机点集群
 - 1. 安装前配置
 - 2. 安装Python主程序
 - 3. 离线安装Python插件包
 - 4. 部署数据挖掘引擎包
 - 5. 创建Python执行用户
 - 6. 启动Python执行代理
 - 7. 运维操作

Python节点主要用于机器学习的DBSACN算法和自定义模块的python脚本扩展。如果没有用到这两个功能模块，可以不用部署python节点。

数据挖掘服务引擎如果没有用到这两个功能模块，也可以不用部署python节点

文档环境

集群部署数据挖掘组件环境如下：

服务器IP	主机名	组件实例	部署目录
10.10.35.64	10-10-35-64	数据挖掘-1, Zookeeper-1, Python-1	/data
10.10.35.65	10-10-35-65	数据挖掘-2, Spark-1, Hadoop-1	/data
10.10.35.66	10-10-35-66	Spark-2, Zookeeper-2, Hadoop-2	/data
10.10.35.67	10-10-35-67	Spark-3, Zookeeper-3, Hadoop-3, Python-2	/data
10.10.204.250	10-10-204-250	Smartbi-Proxy	/data

如果Python计算任务较多，建议Python节点单独部署

1、系统环境准备

1.1 防火墙配置

为了便于安装，建议在安装前关闭防火墙。使用过程中，为了系统安全可以选择启用防火墙，但必须启用服务相关端口。

1. 关闭防火墙

临时关闭防火墙（立即生效）

```
systemctl stop firewalld
```

永久关闭防火墙（重启后生效）

```
systemctl disable firewalld
```

查看防火墙状态

```
systemctl status firewalld
```

2. 开启防火墙

相关服务及端口对照表：

服务名	需要开放端口
Python	8980

如果确实需要打开防火墙安装，需要给防火墙放开以下需要使用到的端口
开启端口：8980

```
firewall-cmd --zone=public --add-port=8980/tcp --permanent
```

配置完以后重新加载firewalld，使配置生效

```
firewall-cmd --reload
```

查看防火墙的配置信息

```
firewall-cmd --list-all
```

3. 关闭selinux

临时关闭selinux，立即生效，不需要重启服务器。

```
setenforce 0
```

永久关闭selinux，修改完配置后需要重启服务器才能生效

```
sed -i 's/=enforcing/=disabled/g' /etc/selinux/config
```

2、部署Python计算机点集群



Python集群说明

Python计算节点集群没有特别设置，只需要在每个服务器中将Python计算节点部署启动完成即可，数据挖掘服务会自动将Python计算任务分发到Python计算节点中。

多个Python节点时，参考以下步骤部署即可。

1. 安装前配置

① 配置本地yum源，参考文档：<https://www.cnblogs.com/diantong/p/10105805.html>



注意事项

Centos7.4 以上，可不设置本地yum源。

② 配置主机名映射

将数据挖掘组件中的服务器主机名映射到hosts文件中

```
vi /etc/hosts
```

文件末尾添(根据实际环境信息设置，如果添加过则无需重复添加)：

```
10.10.35.64 10-10-35-64
10.10.35.65 10-10-35-65
10.10.35.66 10-10-35-66
10.10.35.67 10-10-35-67
```

③安装javva环境

解压jdk到指定目录：

```
tar -zxvf jdk-8u181-linux-x64.tar.gz -C /data
```

添加环境变量

```
vi /etc/profile
```

在文件末尾添加以下内容:

```
export JAVA_HOME=/data/jdk1.8.0_181
export JAVA_BIN=$JAVA_HOME/bin
export CLASSPATH=:$JAVA_HOME/lib/dt.jar:$JAVA_HOME/lib/tools.jar
export PATH=$PATH:$JAVA_BIN
```

让配置生效

```
source /etc/profile
```

验证安装

```
java -version
```

2. 安装Python主程序

上传Python安装包到服务器的/data目录

①安装依赖包

```
rpm -ivh bzip2-1.0.6-13.el7.x86_64.rpm
yum install gcc
```



注意事项

Centos7.4 以上, 无需安装依赖包

②安装Anaconda3并指定安装目录为/data/anaconda3

```
cd /data
chmod +x Anaconda3-2019.10-Linux-x86_64.sh
bash Anaconda3-2019.10-Linux-x86_64.sh -b -p /data/anaconda3
```

③配置环境变量

```
vi /etc/profile
```

在末尾添加以下内容:

```
export PATH=/data/anaconda3/bin:$PATH
```

使配置生效, 查看python版本

```
source /etc/profile
python --version
```

显示Python 3.7.4 表示安装成功。

3. 离线安装Python插件包

①上传离线压缩包到服务器/data目录, 并解压

```
cd /data
unzip plugin.zip
```

②批量安装插件

```
cd /data
pip install --no-index --find-links=./plugin/ -r requirements.txt
```

4. 部署数据挖掘引擎包

上传安装包到服务器，并解压到指定目录

```
tar -zxvf SmartbiMiningEngine-V10.0.64186.21183.tar.gz -C /data
```



注意事项

数据挖掘引擎包版本需要与Smartbi版本一致。

Python计算节点和数据挖掘引擎部署在相同服务器，无需重复解压安装包，使用相同安装包即可。

5. 创建Python执行用户

创建用户组、用户并设置密码

```
groupadd mining                #mining
useradd -g mining mining-ag     #(mining-ag)mining
passwd mining-ag                #mining-ag
```

给引擎安装目录附权限(为了使用mining-ag用户启动执行代理程序时候，有权创建agent-data跟agent-logs目录)

```
chgrp mining -R /data/smartbi-mining-engine-bin
chmod 775 -R /data/smartbi-mining-engine-bin
```

6. 启动Python执行代理

①浏览器访问Smartbi，打开[系统运维 - 数据挖掘配置 - 引擎设置](#)，复制Python代理器启动命令

引擎设置

引擎地址:	<input type="text" value="http://10-10-204-248:8899"/>	示例(http://localhost:8899)	<input type="button" value="测试"/>
服务地址:	<input type="text" value="https://10-10-204-248:8900"/>	示例(https://localhost:8900)	<input type="button" value="测试"/>
代理地址:	<input type="text"/>	示例(https://ip:port)	
Python代理器启动命令:	<input type="text" value="/agent-daemon.sh start --master http://10-10-204-248:8900"/>		<input type="button" value="复制"/>
全局词典文件地址:	<input type="text"/>	?	<input type="button" value="复制"/>



注意事项

复制Python代理器启动命令前，请确认数据挖掘引擎能正常测试连接成功

②登录到部署Python节点机器，并切换到mining-ag用户



注意事项

为了避免出现安全问题，一定要切换到mining-ag用户去启动执行代理服务，不要使用root用户安装或带有sudo权限的用户来启动执行代理服务

```
su - mining-ag
```

进入引擎启动目录

```
cd /data/smartbi-mining-engine-bin/engine/sbin
```

把拷贝命令粘贴，并执行，例如：

```
./agent-daemon.sh start --master http://10.10.35.64:8899 --env python
```

7. 运维操作

1、更新Python数据挖掘引擎包

Smartbi更新war包版本时，Python执行节点需要同步更新对应版本的数据挖掘引擎。