主机信息

hostname	ip	username	password	sklearn	tensorflow 1.10.0
hadoop210	192.168.0.210	hadoop	123	√	√
hadoop211	192.168.0.211	hadoop	123	×	×
hadoop212	192.168.0.212	hadoop	123	×	×

CDH管理页面地址

http://hadoop210:7180

http://192.168.0.210:7180

账号：admin

密码：admin

CDH版本6.3.0

Spark版本2.4.0（华为的环境是1.5.0，开发时注意使用的api不要超过1.5.0）

Pycharm远程调试

1.添加hosts

windows : "C:\Windows\System32\drivers\etc\hosts"
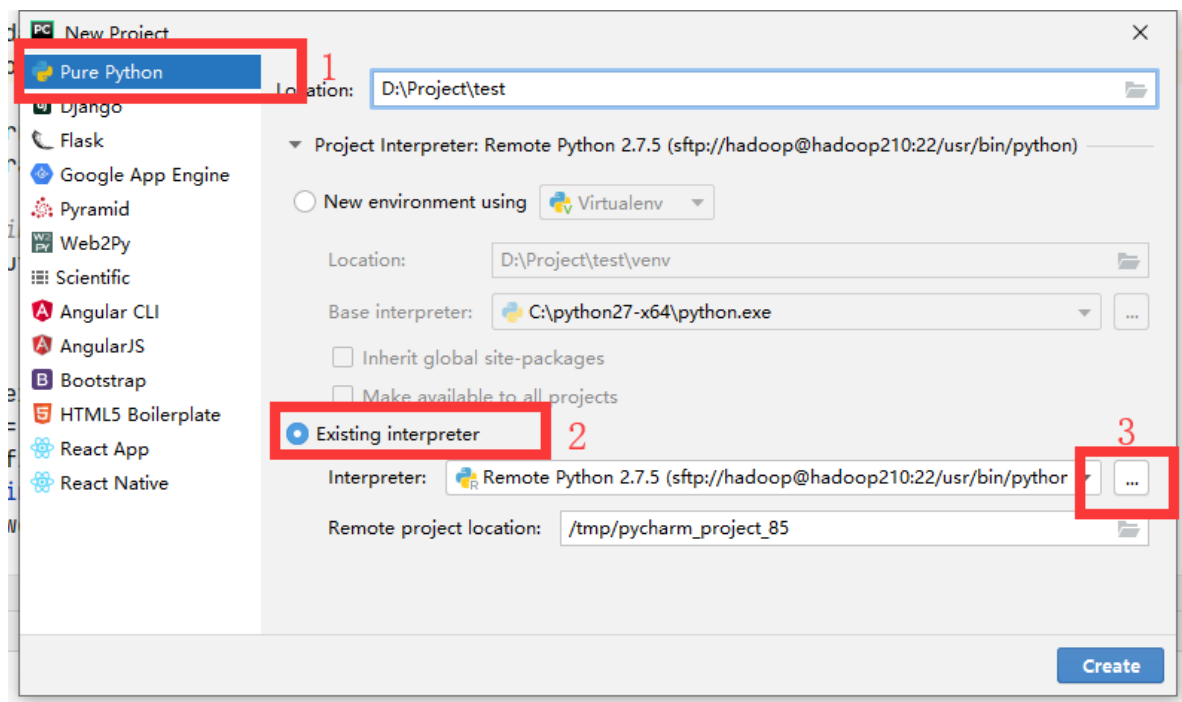
linux : "/etc/hosts"

```
192.168.0.210 hadoop210
192.168.0.211 hadoop211
192.168.0.212 hadoop212
```
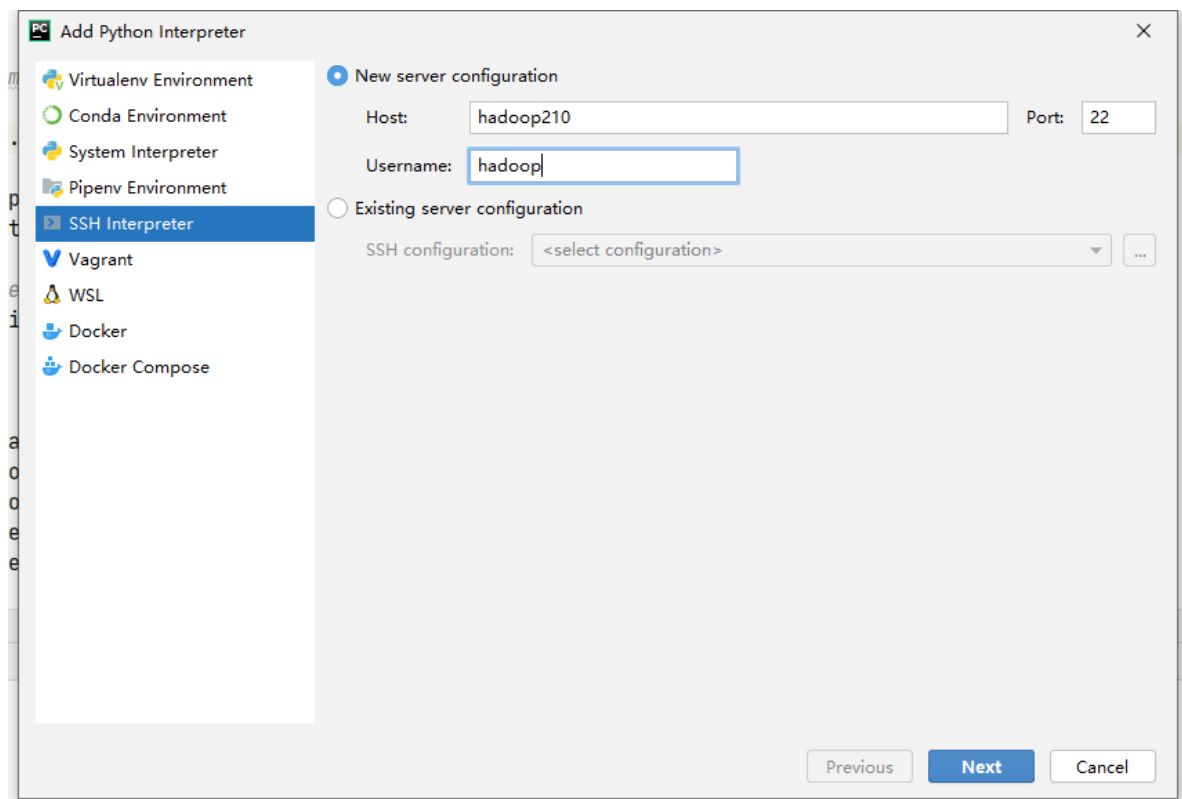
2.pycharm新建项目

1.选择Pure Python

2.选择Existing interpreter

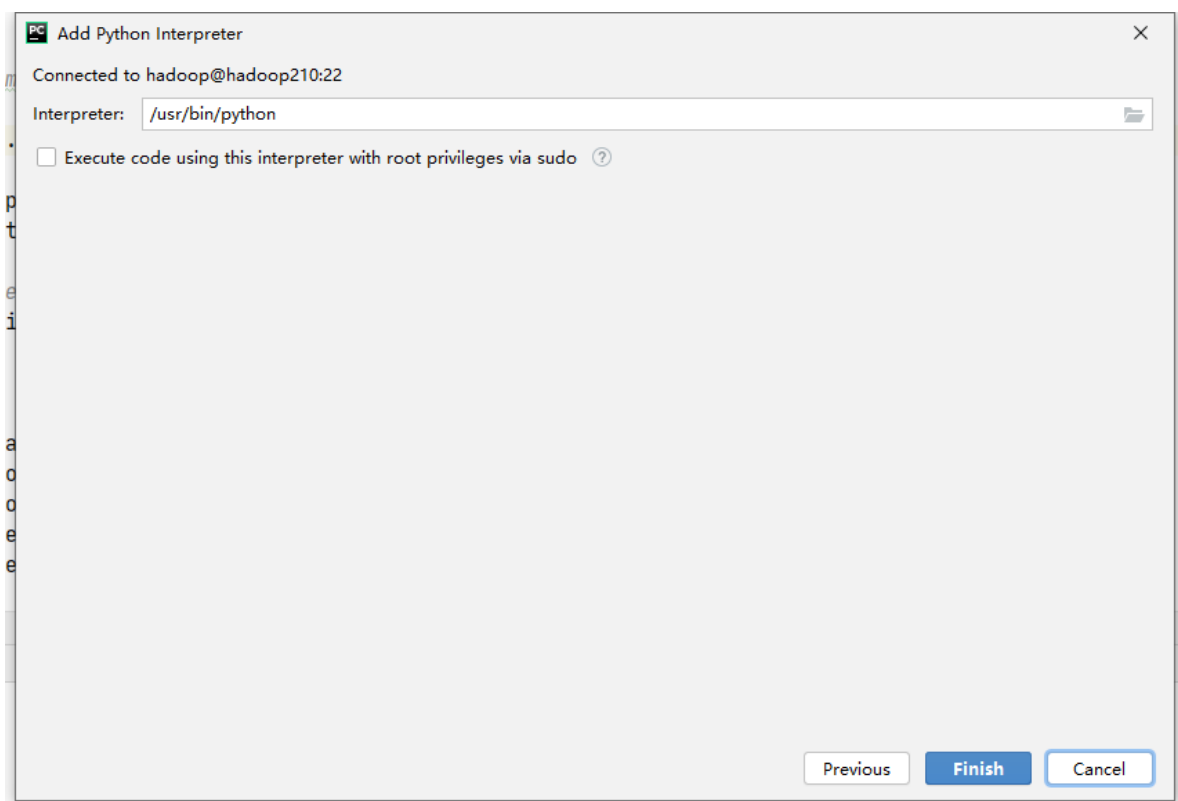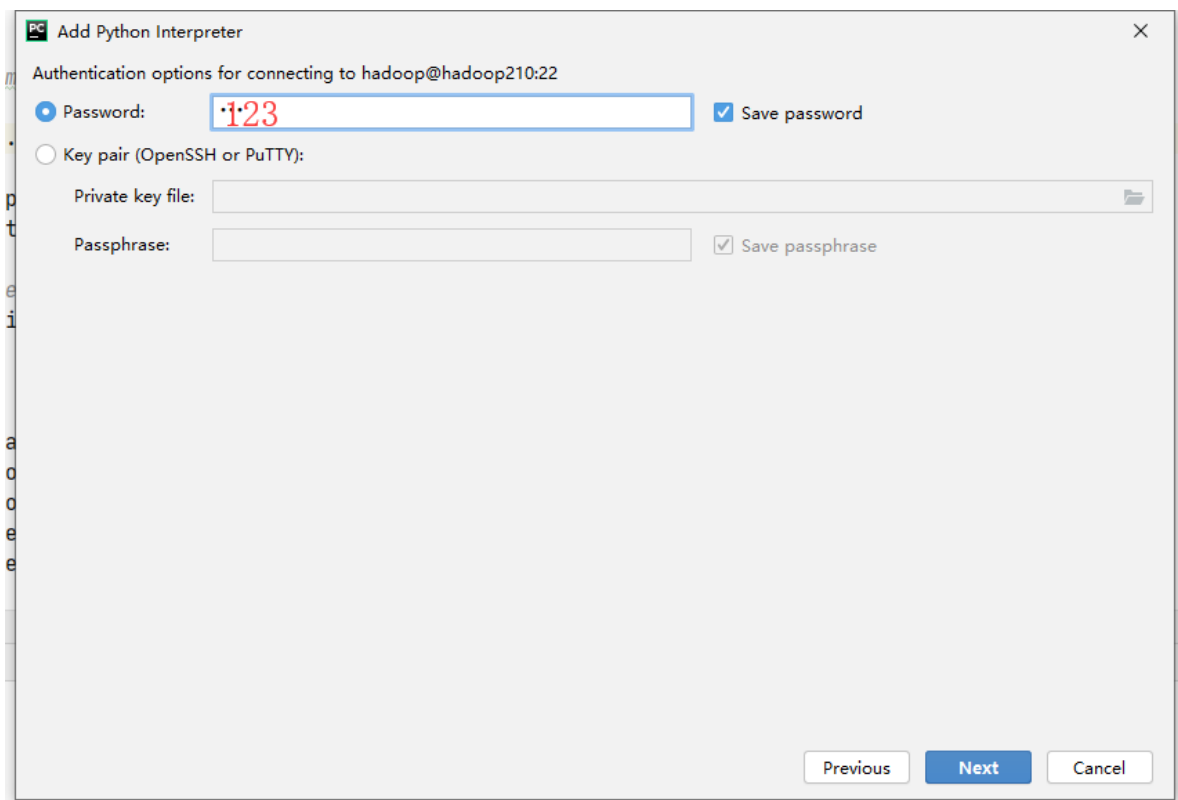3.点击...（Remote project location是自动生产的，无需和图里保持一致）

4.选择SSH Interpreter

5.输入Host：hadoop210，Port：22，Username：hadoop

6.输入Password：123

3.添加系统环境变量

代码中需要添加如下代码段，否则无法远程运行

```python
import os

SPARK_HOME = "/opt/cloudera/parcels/CDH-6.3.0-1.cdh6.3.0.p0.1279813/lib/spark/"
PYSPARK_PYTHON = "/usr/bin/python"
PYSPARK_DRIVER_PYTHON = "/usr/bin/python"
os.environ['SPARK_HOME'] = SPARK_HOME
os.environ['PYSPARK_PYTHON'] = PYSPARK_PYTHON
os.environ['PYSPARK_DRIVER_PYTHON'] = PYSPARK_DRIVER_PYTHON
```