

## 2.安装Pyspark虚拟环境连接pycharm

### 下载pyspark环境并连接pycharm

#### 1、切换到pyspark虚拟环境，安装清华镜像源的pyspark

安装pyspark时候一定要进入到pyspark的虚拟环境中下载，

进入pyspark命令：conda activate pyspark

```
Safe mode is OFF in hadoop008/192.168.10.101:8020
[hadoop@hadoop007 sunyaxu]$ conda activate pyspark
(pyspark) [hadoop@hadoop007 sunyaxu]$
```

##### 1.1、执行安装pyspark

pip install -i <https://pypi.tuna.tsinghua.edu.cn/simple> pyspark==3.2.4

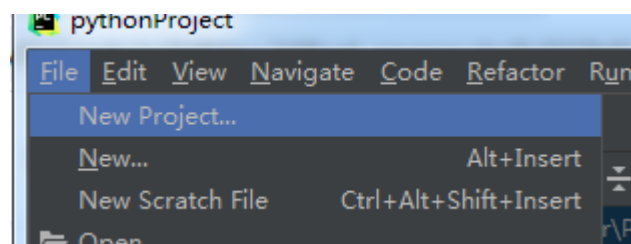
安装完成执行进入Python编译器，import一下如果没报错说明安装成功

```
Successfully built pyspark
Installing collected packages: py4j, pyspark
Successfully installed py4j-0.10.9.5 pyspark-3.2.4
(pyspark) [hadoop@hadoop007 sunyaxu]$ python
Python 3.8.16 (default, Mar 2 2023, 03:21:46)
[GCC 11.2.0] :: Anaconda, Inc. on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import pyspark
>>>
```

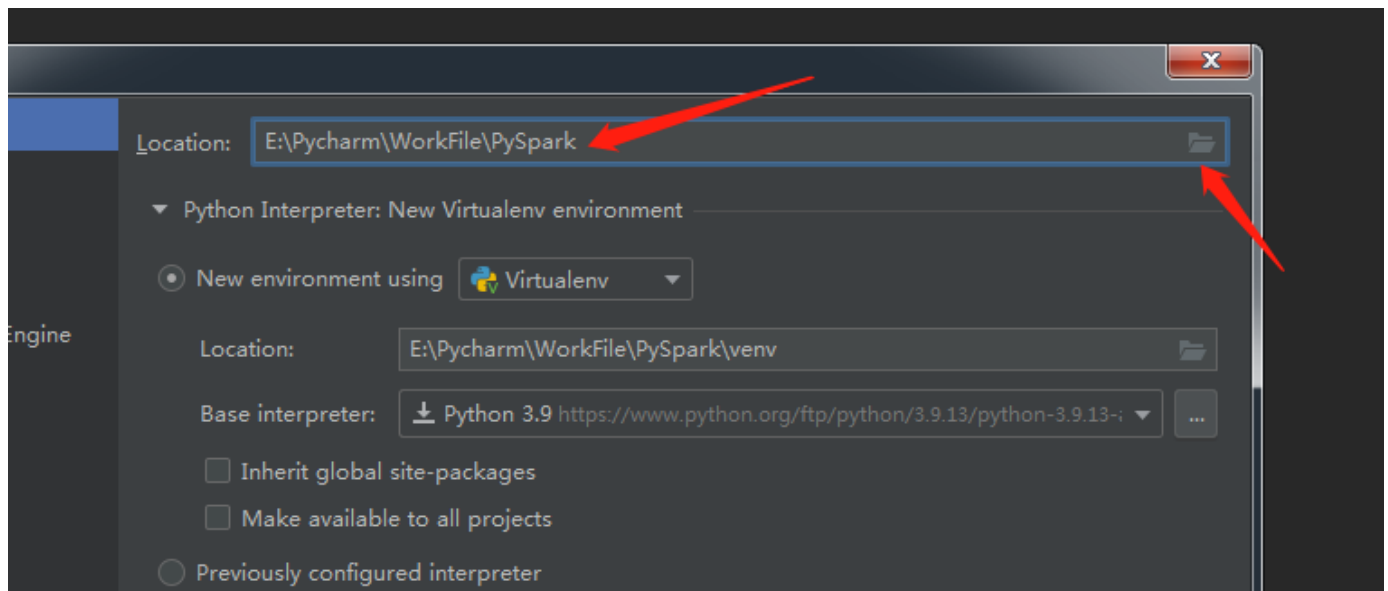
1.2、主节点安装完成，其他两个节点同样方式安装（注意都需要先切换到pyspark环境），配置完成记得验证。

#### 2、配置pycharm连接Linux的Python环境

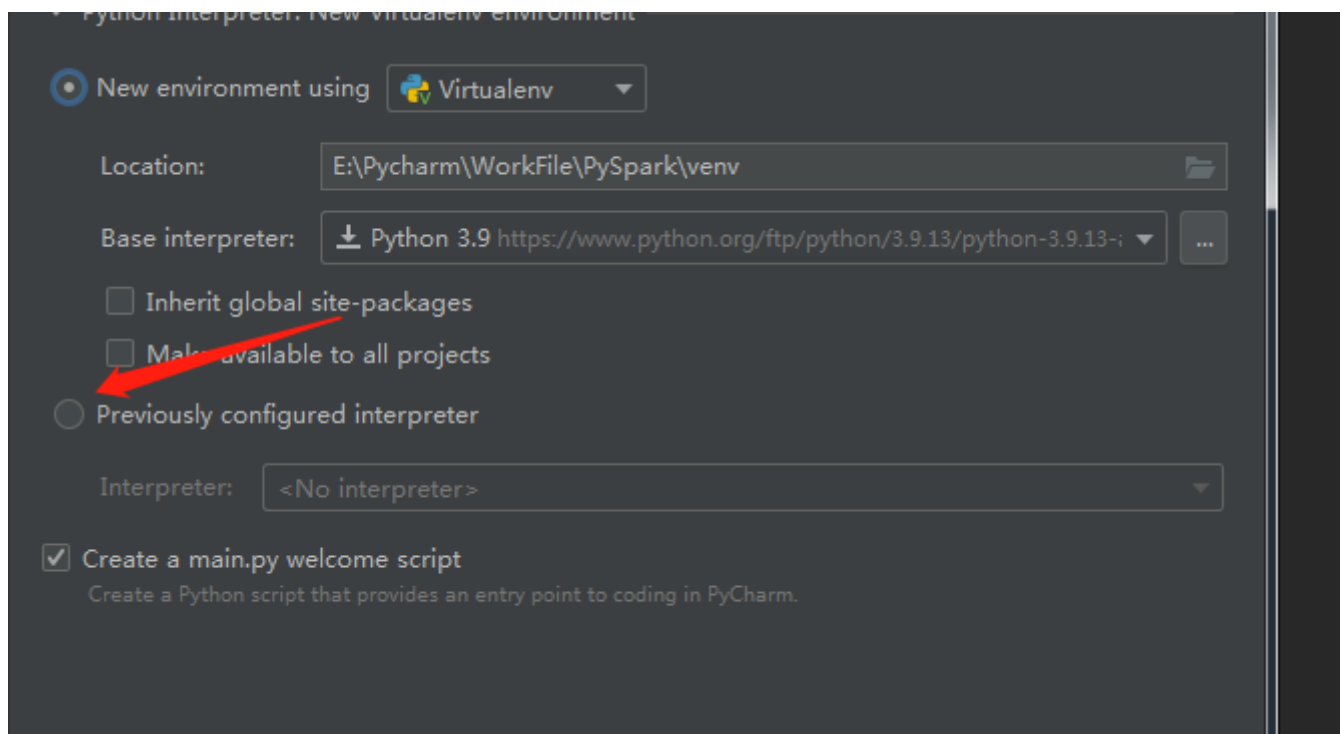
##### 2.1、打开pycharm之后创建新的项目



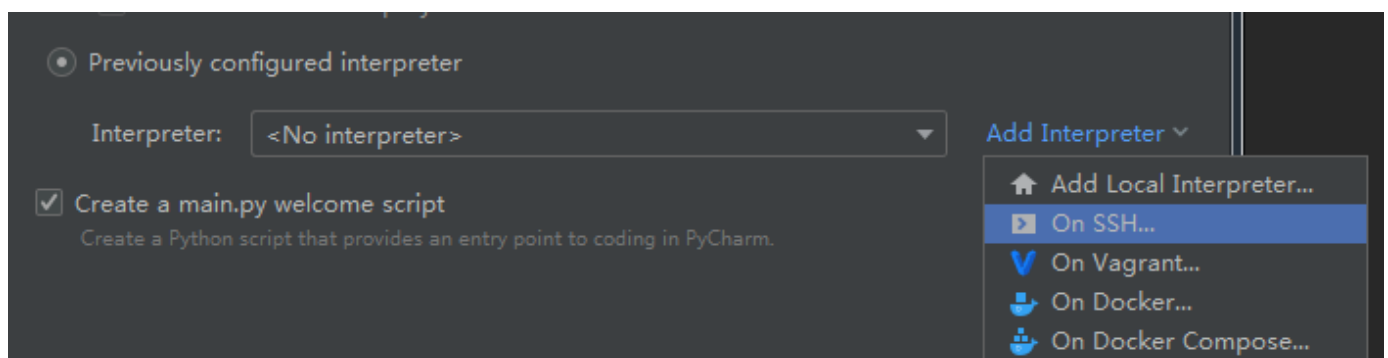
##### 2.2、选择你项目存放的路径和名称



### 2.3、勾选Python环境配置



### 2.4、选择ssh连接我们的虚拟机



### 2.5、填ip端口号，类似于xshell的连接，填写主机名和登陆账号。

SSH connection: ☒ New ☐ Existing

Host: 192.168.10.100

Username: hadoop

2.6、输入hadoop用户的密码，点击next

New Target: SSH

Authentication options for connecting to hadoop@hadoop007:22

☒ Password: ..... ☒ Save password

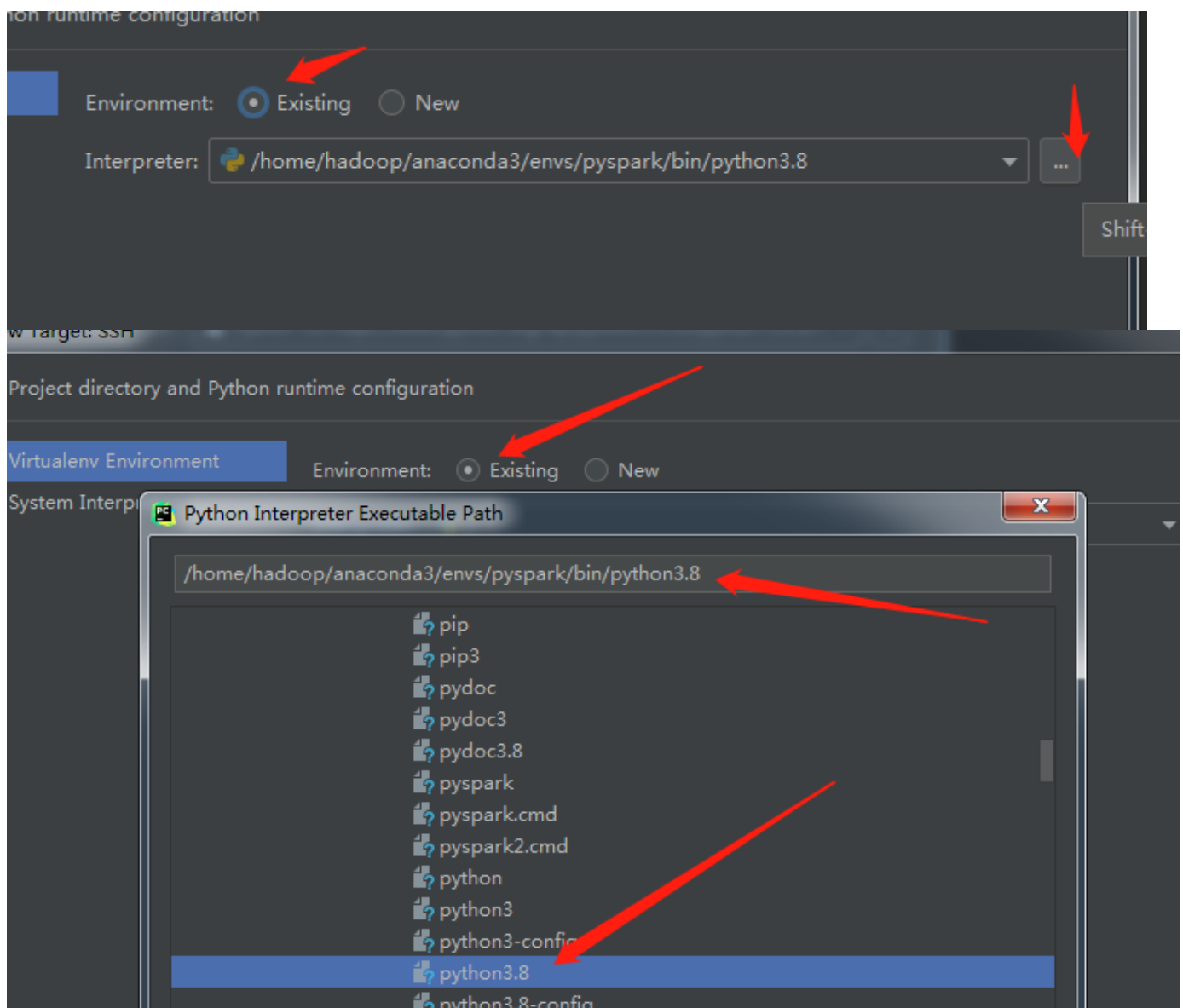
☐ Key pair: OpenSSH or PuTTY

Private key:

Passphrase:  ☒ Save passphrase

? Previous Next Cancel

2.7、指定虚拟环境pyspark的路径, /home/hadoop/anaconda3/envs/pyspark/bin/python3.8

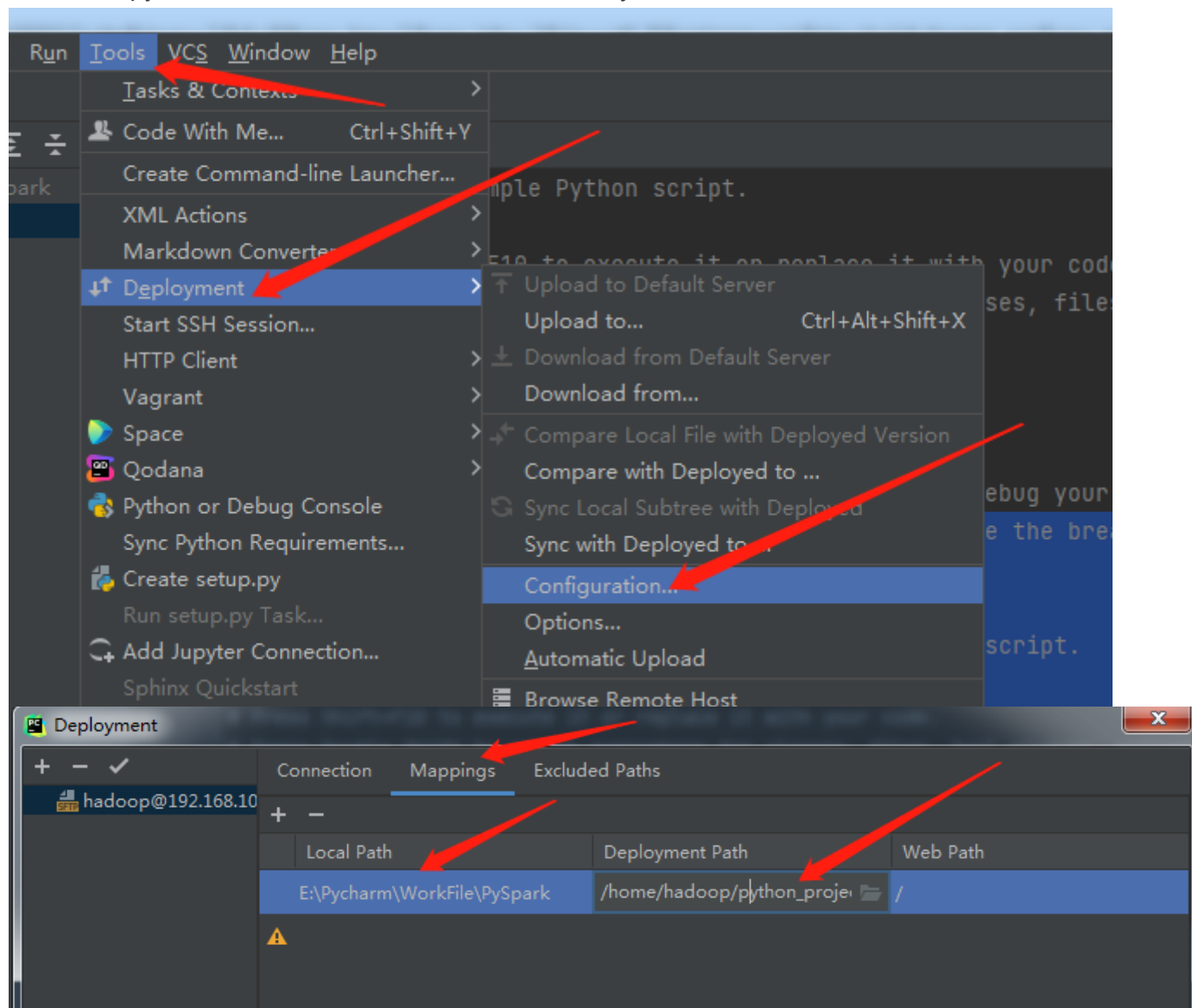


2.8、由于配置了连接远程Linux服务器，执行Python代码必定需要将代码同步到Linux环境执行，也就是说我们本地会有一份，Linux也会有一份。pycharm上带有自动同步功能，我们只需要进行一些配置即可。

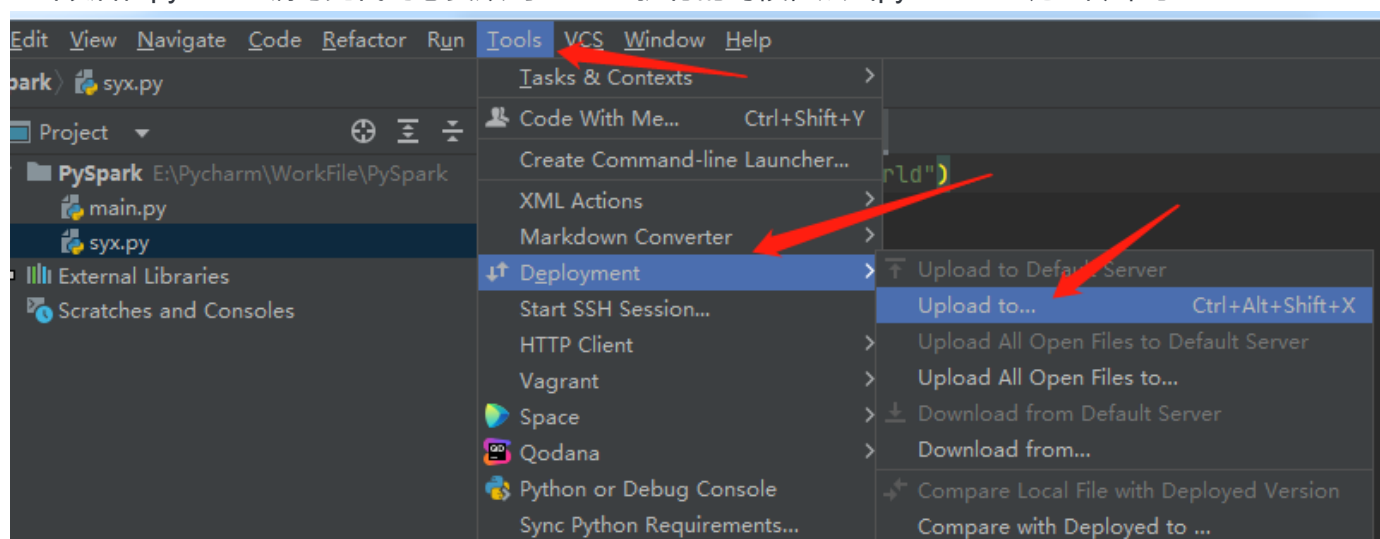
\* 首先在Linux创建一个用于你存放Python代码的文件夹，我这里创建了一个python\_project文件夹存放在了/home/hadoop/python\_project路径上

```
total 0
[hadoop@hadoop007 python_project]$ pwd
/home/hadoop/python_project
[hadoop@hadoop007 python_project]$
```

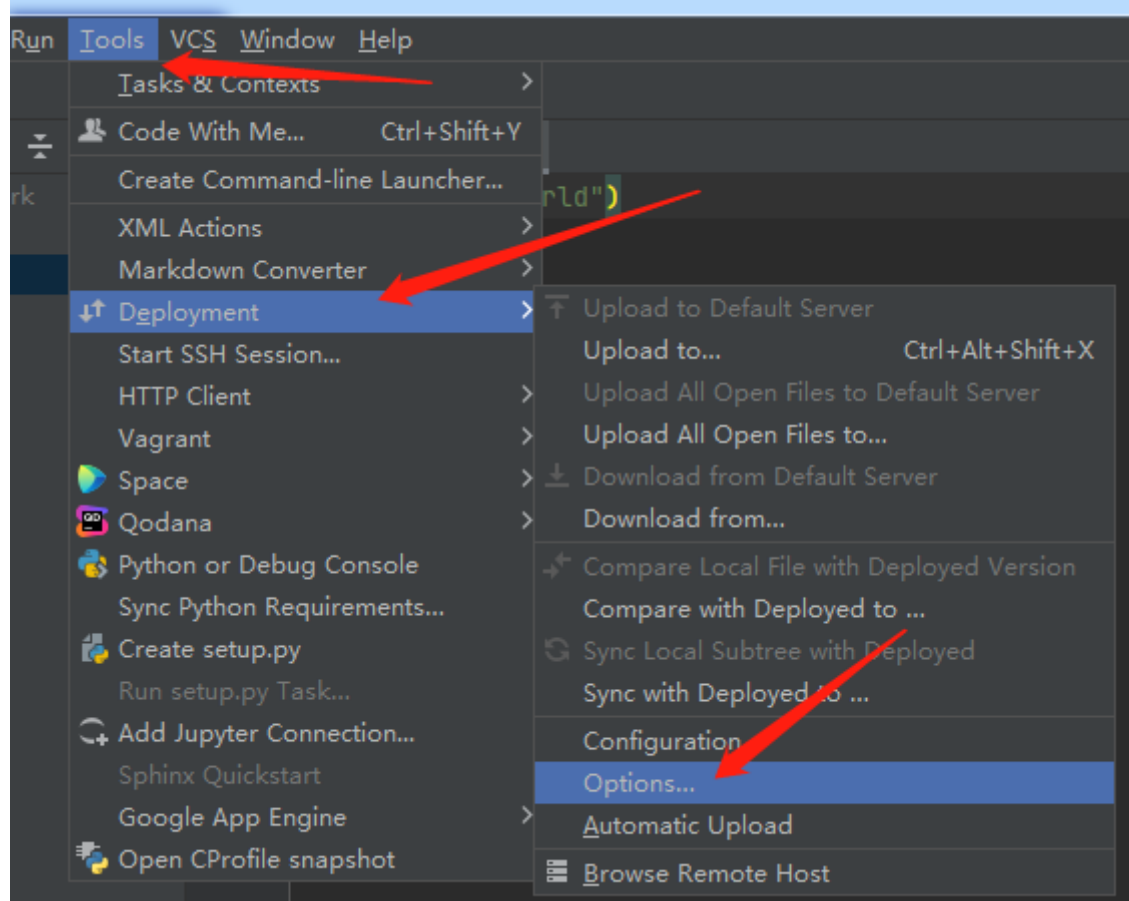
\* 然后点击pycharm的工具配置中，配置Linux中存放Python代码的目录。



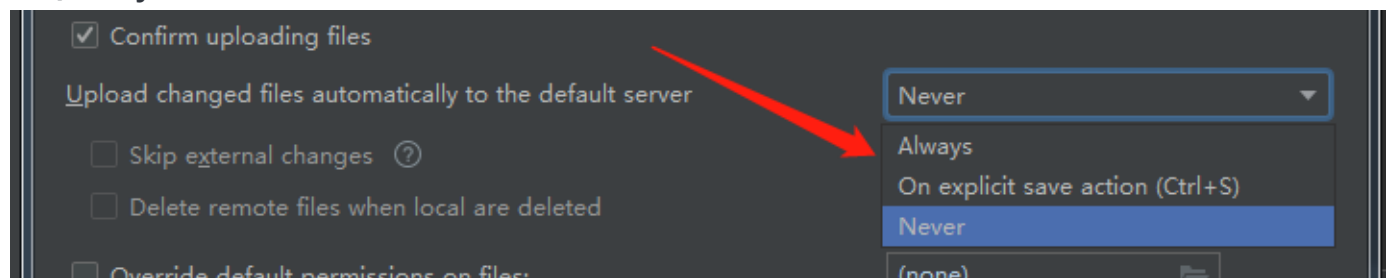
\* 当以后在pycharm编写完代码想要放到Linux上执行的时候，点击pycharm一键上传即可



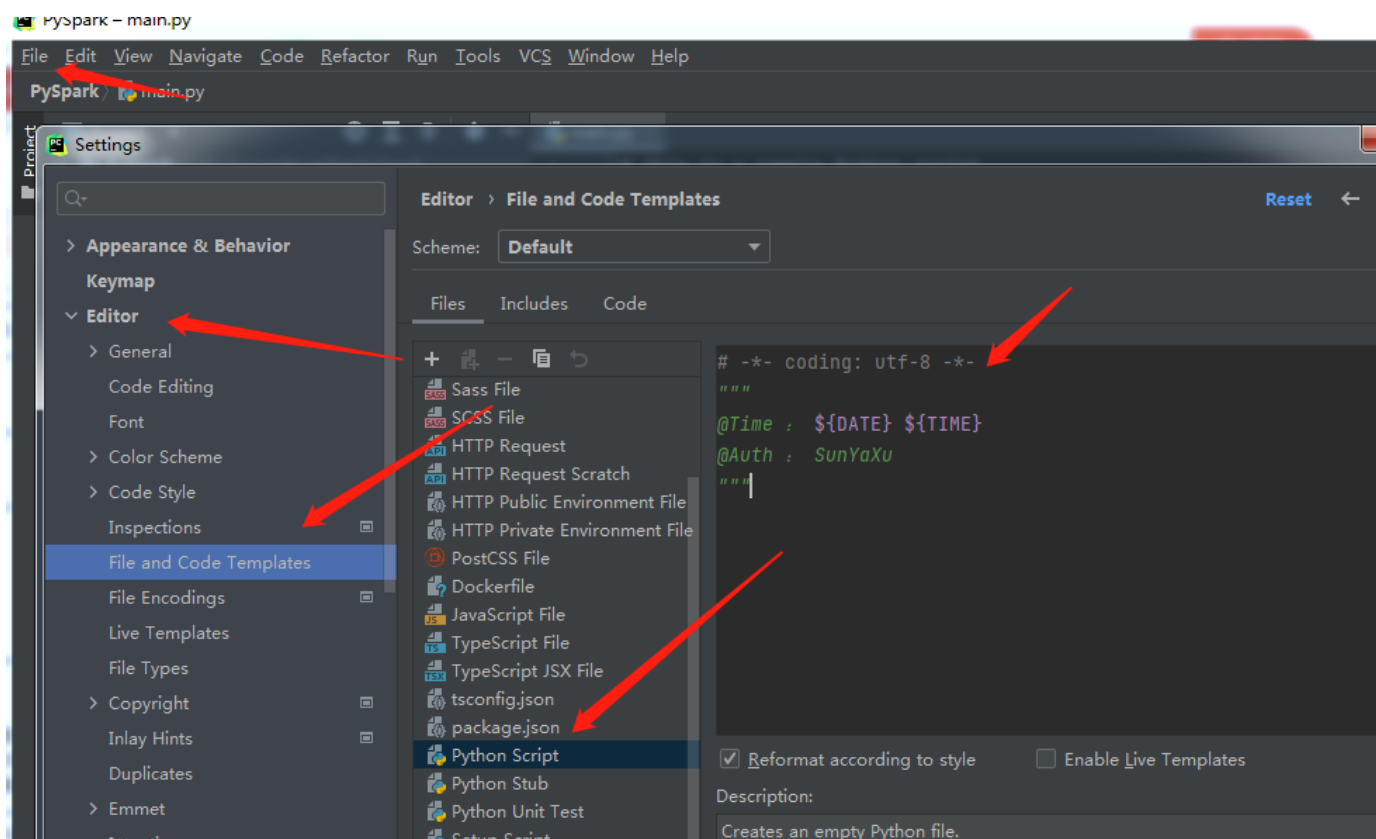
\* 或者说你想让pycharm每次只要新建代码都会自动同步到Linux的，可以勾选一下配置选项



选择always



### 3、配置项目开发者信息（可选项，提高代码可读性）



## 配置内容

```
# -*- coding: utf-8 -*-  
"""  
@Time :  ${DATE}  ${TIME}  
@Auth :  SunYaXu  
"""
```

## 效果展示:

