

Pyflink 的安装和 windows 开发环境配置

作者: [Sakura6868](#)

原文链接: <https://ld246.com/article/1631936123228>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)



在分布式大数据计算引擎这个领域，现在最常用的Apache Spark 早已支持 python语言编写，而且对 ML（机器学习）和DM（数据挖掘）也都有api的支持，而作为第三代计算引擎的flink，从 1.9.0 版开始增加了对 Python 的支持（PyFlink），在Flink 1.10 中，PyFlink 增加了对 Python UDFs（自定义函数）的支持，可以在 Table API/SQL 中注册并使用自定义函数，而从Flink 1.11开始还支持在 Windows 上本地运行 PyFlink 作业，所以可以在 Windows 上开发和调试 PyFlink 作业

Pyflink的安装

Pyflink的安装十分之简单

- 首先先看看自己系统的python版本（PyFlink 需要 Python 版本（3.6、3.7 或 3.8））

```
$ python3 --version  
# the version printed here must be 3.6, 3.7 or 3.8
```

- 环境配置

由于系统可能包含多个 Python 版本，因此也包含多个 Python 二进制可执行文件。运行以下 `ls` 命令找出系统中可用的 Python 二进制可执行文件：

```
$ ls /usr/bin/python*
```

- 选择软链接 `python` 指向您的 `python3` 解释器

```
ln -s /usr/bin/python3 python
```

- 安装 Pyflink

由于 `pyflink` 还在持续火热的更新之中每个版本的变化较大，所以无脑安装最新版就行（更新此文章时新版为 `apache-flink1.13.2`）

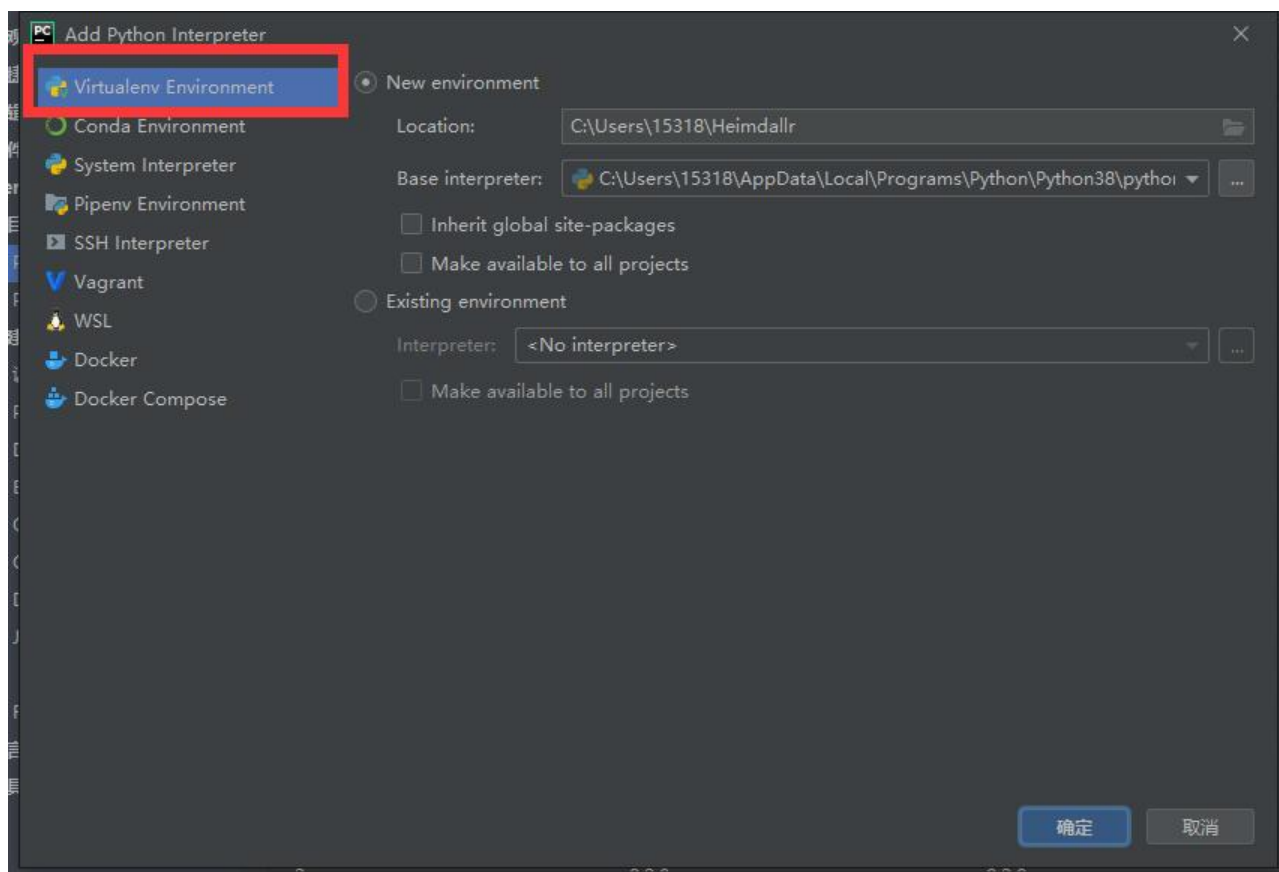
```
$ python3 -m pip install apache-flink
```

windows 开发环境配置

这里我们选择 Pycharm IDE 来进行 windows Pyflink 开发

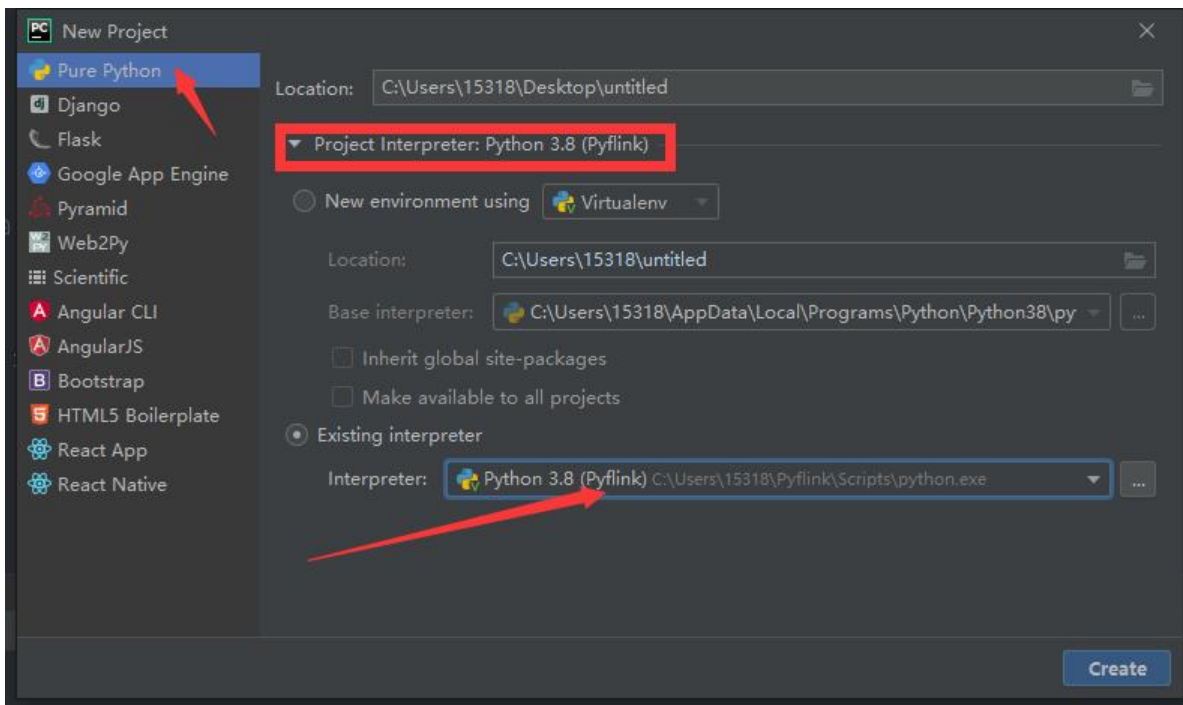
- 首先配置 Python 虚拟环境

配置路径 `PyCharm -> Preferences -> Project Interpreter`



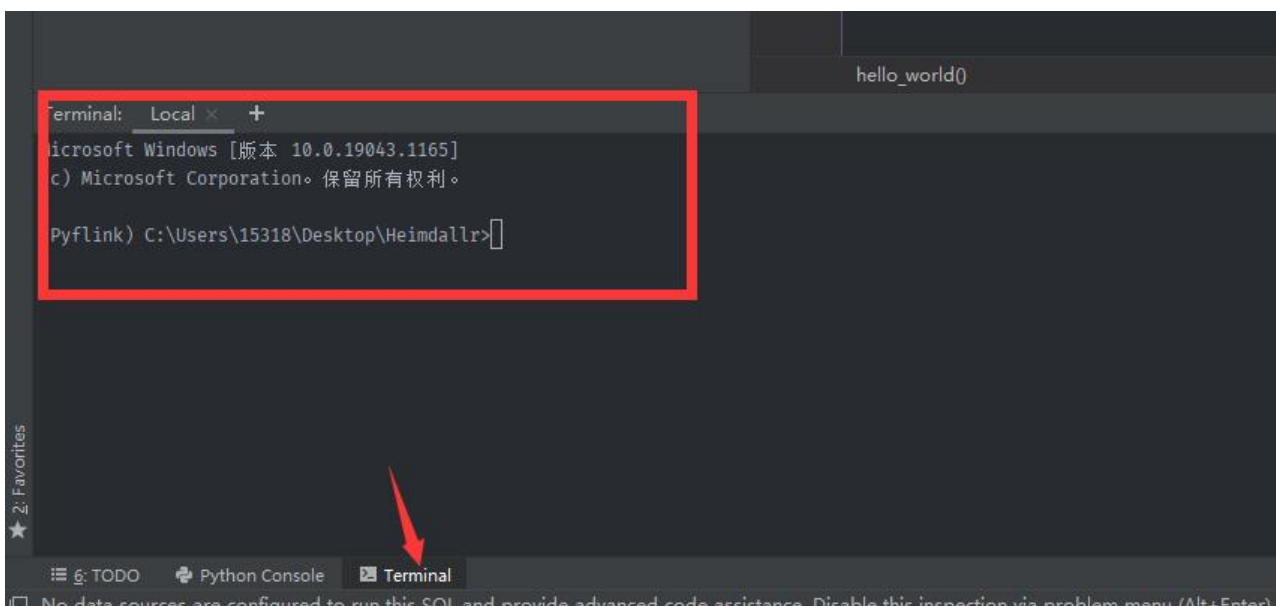
记住选择的 python 版本一定要是 3.6、3.7 或 3.8

- 新建一个项目，环境选择我们刚刚配置的 python 虚拟环境



• 安装Pyflink

进入到终端界面



先看看 python 版本



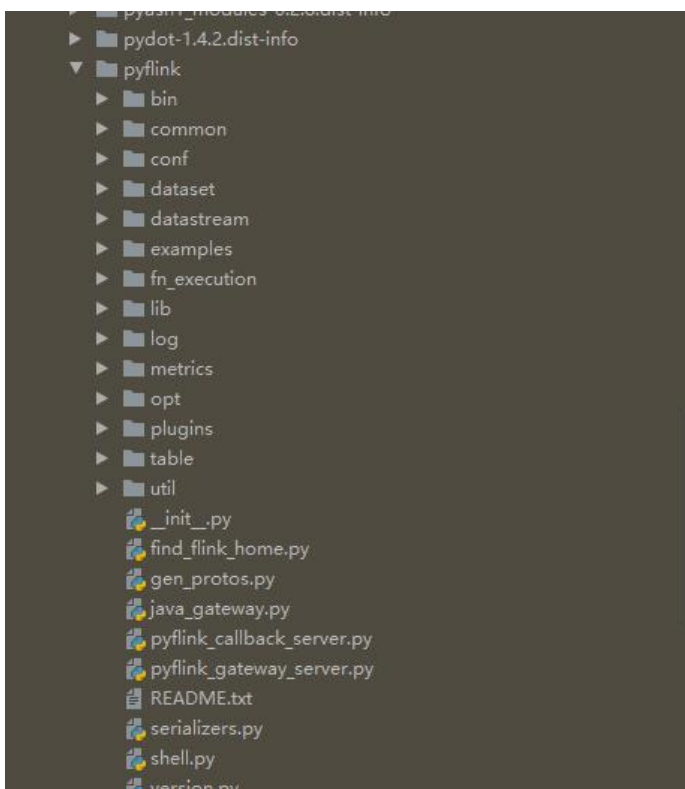
然后安装 Pyflink

```
Terminal: Local x +
Microsoft Windows [版本 10.0.19043.1165]
(c) Microsoft Corporation。保留所有权利。

(Pyflink) C:\Users\15318\Desktop\Heimdallr>python --version
Python 3.8.2

(Pyflink) C:\Users\15318\Desktop\Heimdallr>python3 -m pip install apache-flink
```

最终完成之后你可以在 `site-packages` 下面找的 `pyflink` 目录，如下



• Hello World 示例

新建一个.py文件输入以下代码

```
#!/usr/bin/env python38
#-*- coding:utf-8 -*-
from pyflink.datastream import StreamExecutionEnvironment
from pyflink.table import EnvironmentSettings, StreamTableEnvironment

def hello_world():
    """
    从随机Source读取数据，然后直接利用PrintSink输出。
    """
    settings = EnvironmentSettings.new_instance().in_streaming_mode().use_blink_planner().build()
    env = StreamExecutionEnvironment.get_execution_environment()
```

```

t_env = StreamTableEnvironment.create(stream_execution_environment=env, environment
settings=settings)
source_ddl = """
    CREATE TABLE random_source (
        f_sequence INT,
        f_random INT,
        f_random_str STRING
    ) WITH (
        'connector' = 'datagen',
        'rows-per-second'='5',
        'fields.f_sequence.kind'='sequence',
        'fields.f_sequence.start'='1',
        'fields.f_sequence.end'='1000',
        'fields.f_random.min'='1',
        'fields.f_random.max'='1000',
        'fields.f_random_str.length'='10'
    )
    """

sink_ddl = """
    CREATE TABLE print_sink (
        f_sequence INT,
        f_random INT,
        f_random_str STRING
    ) WITH (
        'connector' = 'print'
    )
    """

# 注册source和sink
t_env.execute_sql(source_ddl)
t_env.execute_sql(sink_ddl)

# 数据提取
tab = t_env.from_path("random_source")
# 这里我们暂时先使用 标注了 deprecated 的API, 因为新的异步提交测试有待改进...
tab.execute_insert("print_sink").wait()
# 执行作业
t_env.execute_sql("Flink Hello World")

if __name__ == '__main__':
    hello_world()

```

结果如下

```
Run: demo x
C:\Users\15318\Pyflink\Scripts\python.exe C:/Users/15318/Desktop/Heimdallr/demo.py
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.flink.api.java.ClosureCleaner (file:/C:/Users/15318/Pyflink/Lib/site-packages/pyflink/lib/flink-
WARNING: Please consider reporting this to the maintainers of org.apache.flink.api.java.ClosureCleaner
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
9> +I[9, 317, 92f95cc04f]
2> +I[2, 990, 74e83e550d]
7> +I[7, 609, 4efb98a28f]
1> +I[1, 568, 08f8da34c1]
4> +I[4, 817, 99795a375e]
10> +I[10, 772, 16ff7cd57a]
12> +I[12, 886, 06fedcb8b]
6> +I[6, 326, 29683a2c39]
3> +I[3, 513, 2ced3a76fd]
5> +I[5, 380, ad6d7a4a45]
11> +I[11, 969, 1afc9837da]
8> +I[8, 702, eaed78f6a3]
9> +I[9, 721, 87de36b52f]
6> +I[6, 956, b8a61e8cfa]
8> +I[8, 450, 8b0613b988]
2> +I[2, 288, 354dh07a30]
```