

# Python 普及

[lida03@meituan.com](mailto:lida03@meituan.com)

# 入门教程

Python 简明教程 前10章

[http://woodpecker.org.cn/abyteofpython\\_cn/chinese/](http://woodpecker.org.cn/abyteofpython_cn/chinese/)

# 我们做一个练习

- 抓取豆瓣正在上映的电影名称
- 分析需求

# 拆解任务

- 环境
- 抓
- 分析
- 存

# 轮子

在别人做好的工具的基础上去做事半功倍

# 环境

- 云
- 安装 Python 的virtualenv环境
- 安装 Package

# 私有云

- <http://cloud.sankuai.com/users/login/>
- `ssh-keygen -t rsa`
- `ssh sankuai@[IP]`

# 虚拟环境

- `sudo apt-get install python-pip`
- `sudo pip install virtualenv`
- `mkdir [PROJECT]`
- `cd [PROJECT]`
- `virtualenv .env --distribute`
- `source .env/bin/activate`



# 0.可以干活了

- 初步了解

# I. 抓取

- 抓取网页需要的工具
- 不要使用 urllib or urllib2
- 使用 pip install requests
- <http://docs.python-requests.org/en/latest/>

# 2/3.分析

- re (正则表达式)
- XPATH
- pip install BeautifulSoup

# 4.保存

- 文本?
- SQL?
- ORM?
- `pip install SQLAlchemy`
- <http://www.sqlalchemy.org/>
- [http://www.desertsandsoftware.com/?page\\_id=99](http://www.desertsandsoftware.com/?page_id=99)

# 谢谢, 善用搜索和提问

<http://google.com/>

<http://stackoverflow.com/>

<https://github.com/yoyicue/tutorial>