

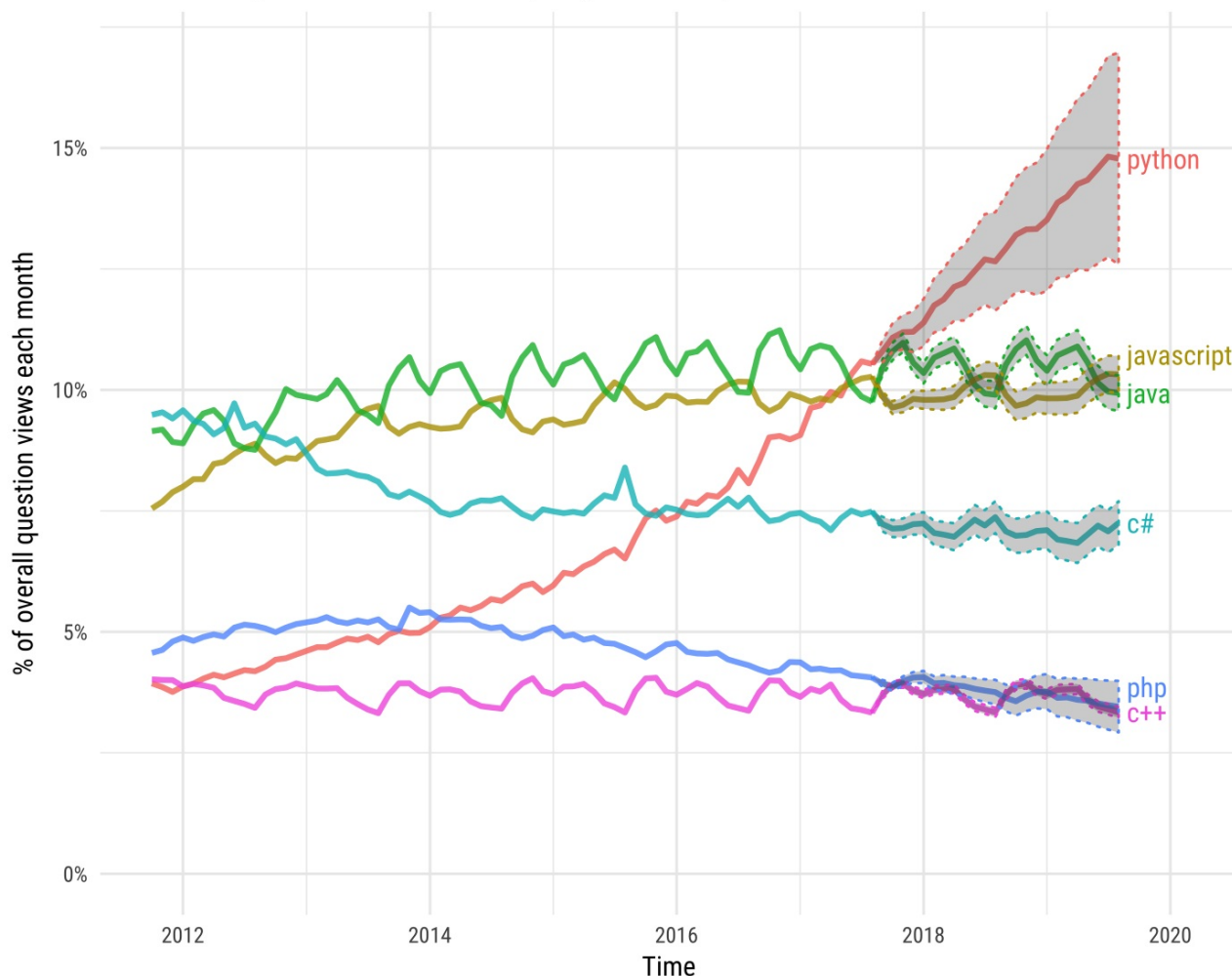
02-JupyterNotebook为什么是现代Python的必学技术？

你好，我是景霄。

Stack Overflow 曾在2017年底，发布了在该站上各种语言的提问流量。其中，Python已经超过了JavaScript成为了流量最高的语言，预测在2020年前会远远甩开JavaScript。

Projections of future traffic for major programming languages

Future traffic is predicted with an STL model, along with an 80% prediction interval.



可能你已经知道，Python在14年后的“崛起”，得益于机器学习和数学统计应用的兴起。那为什么Python如此适合数学统计和机器学习呢？作为“老司机”的我可以肯定地告诉你，Jupyter Notebook (<https://jupyter.org/>) 功不可没。

毫不夸张地说，根据我对Facebook等硅谷一线大厂的了解，一个Python工程师如果现在还不会使用Jupyter Notebook的话，可能就真的太落伍了。

磨刀不误砍柴工，高效的工具让我们的编程事半功倍。这一节课，我就来带你学习一下Jupyter Notebook，为后面的Python学习打下必备基础。

什么是Jupyter Notebook？

说了这么多，到底什么是Jupyter Notebook？按照Jupyter 创始人 Fernando Pérez的说法，他最初的梦想是做一个综合 Ju（Julia）、Py（Python）和 R 三种科学运算语言的计算工具平台，所以将其命名为Ju-Py-te-R。发展到现在，Jupyter 已经成为一个几乎支持所有语言，能够把软件代码、计算输出、解释文档、多媒体资源整合在一起的多功能科学运算平台。

英文里说一图胜千言 (A picture is worth a thousand words)。看下面这个图片，你就明白什么是Jupyter Notebook了。

▼ Rich, interactive outputs

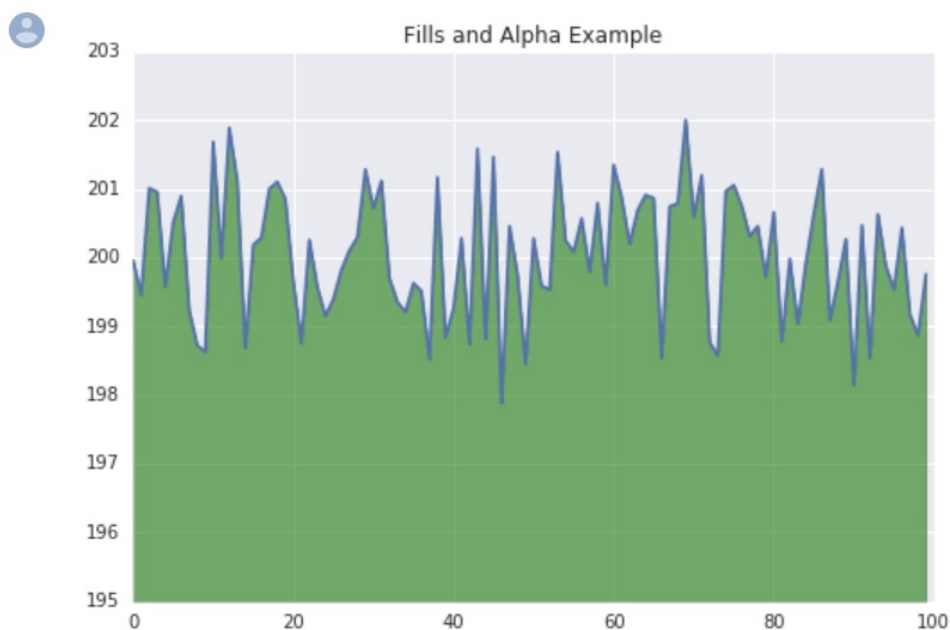
Until now all of the generated outputs have been text, but they can be more interesting, like the chart below.

```
[ ] import numpy as np
    from matplotlib import pyplot as plt

    ys = 200 + np.random.randn(100)
    x = [x for x in range(len(ys))]

    plt.plot(x, ys, '-')
    plt.fill_between(x, ys, 195, where=(ys > 195), facecolor='g', alpha=0.6)

    plt.title("Fills and Alpha Example")
    plt.show()
```



你在一个框框中直接输入代码，运行，它立马就在下面给你输出。怎么样，是不是很酷？你可能会纳闷儿，这样一个看起来“华而不实”的玩意儿，真的就成了Python社区的颠覆者吗？说实话放在几年前我也是不信的。所以 Jupyter Notebook 的影响究竟有多大呢？

Jupyter Notebook 的影响力

我们衡量一个技术的影响力，或者说要用自己的技术去影响世界时，必定绕不开这个技术对教育界的影响力。

就拿微软的Word文本处理系统来说吧。从纯技术角度来讲，Word的单机设计理念早已落后时代20年。但以 Google Doc为代表的在线文档系统，却并没有像想象中那样，实现对Word的降维打击。

直观的原因是用户习惯，使用Word修改文档，那就来回发几十遍呗，用着也还可以。但更深刻来想，之所以养成这样的用户习惯，是因为我们的教育根源。教育系统从娃娃抓起，用小学中学大学十几年的时间，训练了用户Word的使用习惯。到工作中，老员工又会带着新员工继续使用Word，如此行程技术影响力生生不息的正向反馈。

回到我们今天的主题，我们来看Jupyter Notebook。从2017年开始，已有大量的北美顶尖计算机课程，开始完全使用Jupyter Notebook作为工具。比如李飞飞的CS231N《计算机视觉与神经网络》课程，在16年时作业还是命令行Python的形式，但是17年的作业就全部在Jupyter Notebook上完成了。再如UC Berkeley的《数据科学基础》课程，从17年起，所有作业也全部用Jupyter Notebook完成。

而Jupyter Notebook在工业界的影响力更甚。在Facebook，虽然大规模的后台开发仍然借助于功能齐全的IDE，但是几乎所有的中小型程序，比如内部的一些线下分析软件，机器学习模块的训练都是借助于Jupyter Notebook完成的。据我了解，在别的硅谷一线大厂，例如Google的AI Research部门Google Brain，也是清一色地全部使用Jupyter Notebook，虽然用的是他们自己的改进定制版，叫 Google Colab。

看到这里，相信你已经认可了Jupyter Notebook如今的江湖地位。不过，说到技术的选择，有些人会说，这个技术流行，我们应该用；有些人认为，阿里已经在用这个技术了，这就是未来，我们也要用等等。不得不说，这些都是片面的认知。不管是阿里还是Facebook用的技术，其实不一定适用你的应用场景。

我经常会鼓励技术同行，对于技术选择要有独立的思考，不要人云亦云。最起码你要去思考，Facebook为什么选择这个技术？这个技术解决了哪些问题？Facebook为什么不选择别的技术？有哪些局限？单从选择结果而言，Facebook选择的技术很可能是因为它有几百个产品线，几万个工程师。而同样的技术，在一个十人的团队里，反而成了拖累。

在这里，我不想忽悠你任何技术，我想教会你的是辩证分析技术的思考方法。接下来，我们就来看看，Jupyter究竟解决了哪些别人没有解决的问题。

Jupyter的优点

整合所有的资源

在真正的软件开发中，上下文切换占用了大量的时间。什么意思呢？举个例子你就很好理解了，比如你需要切换窗口去看一些文档，再切换窗口去用另一个工具画图等等。这些都是影响生产效率的因素。

正如我前面提到的，Jupyter通过把所有和软件编写有关的资源全部放在一个地方，解决了这个问题。当你打开一个Jupyter Notebook时，就已经可以看到相应的文档、图表、视频和相应的代码。这样，你就不需要切换窗口去找资料，只要看一个文件，就可以获得项目的所有信息。

交互性编程体验

在机器学习和数学统计领域，Python编程的实验性特别强，经常出现的情况是，一小块代码需要重写100遍，比如为了尝试100种不同的方法，但别的代码都不想动。这一点和传统的Python开发有很大不同。如果是在传统的Python开发流程中，每一次实验都要把所有代码重新跑一遍，会花费开发者很多时间。特别是在像Facebook这样千万行级别的代码库里，即使整个公司的底层架构已经足够优化，真要重新跑一遍，也需要几分钟的时间。

而Jupyter Notebook引进了Cell的概念，每次实验可以只跑一小个Cell里的代码；并且，所见即所得，在代码下面立刻就可以看到结果。这样强的互动性，让Python研究员可以专注于问题本身，不被繁杂的工具链所累，不用在命令行直接切换，所有科研工作都能在Jupyter上完成。

零成本重现结果

同样在机器学习和数学统计领域，Python的使用是非常短平快的。常见的场景是，我在论文里看到别人的

方法效果很好，可是当我去重现时，却发现需要pip重新安装一堆依赖软件。这些准备工作可能会消耗你80%的时间，却并不是真正的生产力。

Jupyter Notebook如何解决这个问题呢？

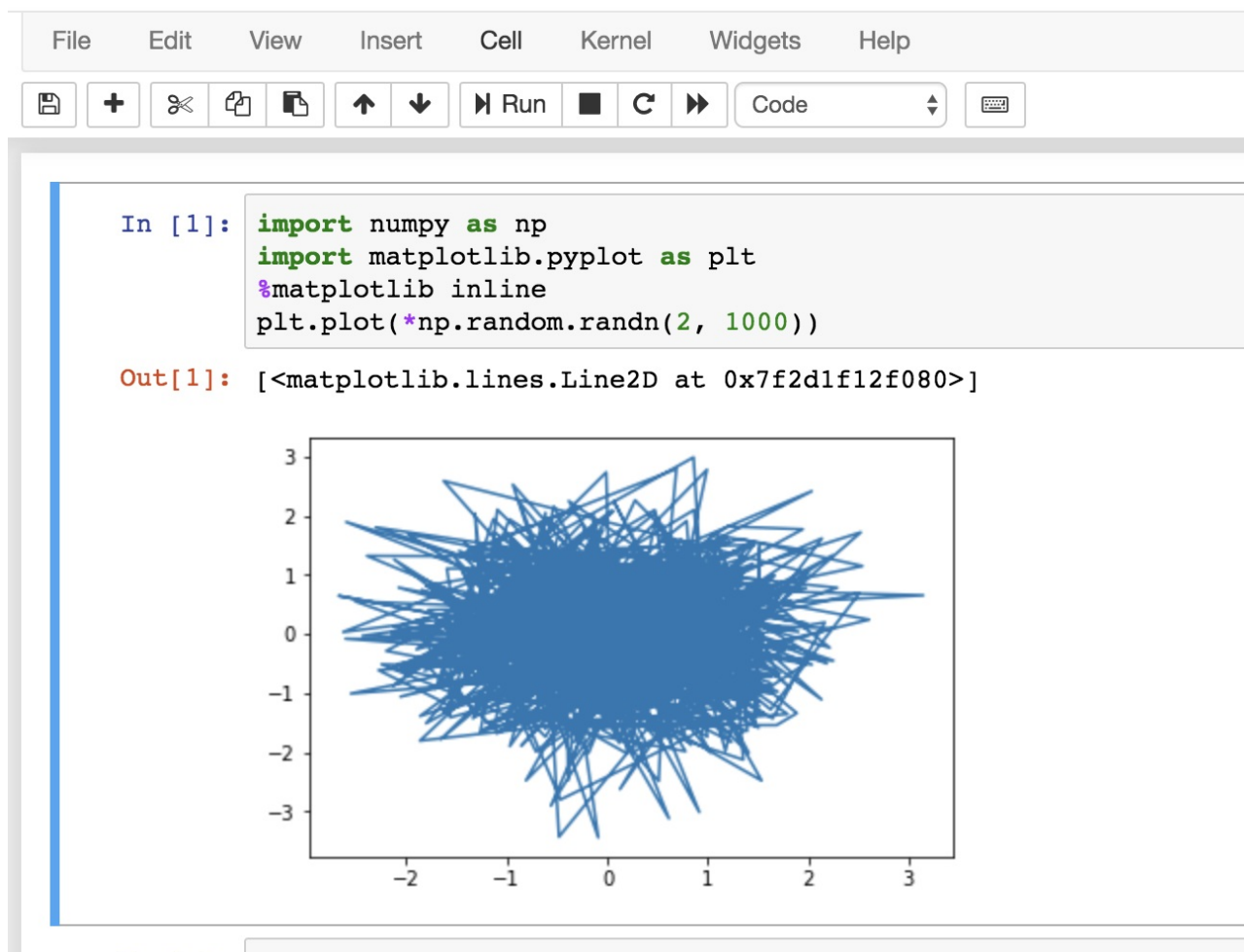
其实最初的Jupyter Notebook也是挺麻烦的，需要你先在本机上安装IPython引擎及其各种依赖软件。不过现在的技术趋势，则是彻底云端化了，例如Jupyter官方的Binder平台（介绍文档：<https://mybinder.readthedocs.io/en/latest/index.html>）和Google提供的 Google Colab环境（介绍：<https://colab.research.google.com/notebooks/welcome.ipynb>）。它们让Jupyter Notebook变得和石墨文档、Google Doc在线文档一样，在浏览器点开链接就能运行。

所以，现在当你用Binder打开一份GitHub上的Jupyter Notebook时，你不需要安装任何软件，直接在浏览器打开一份代码，就能在云端运行。

Jupyter Notebook 初体验

学习技术的最好方法就是用技术。不过，在今天的篇幅里，我不可能带你完全学会Jupyter Notebook的所有技巧。我想先带你直接感受一下，使用Jupyter Notebook的工作体验。

比如这样一个[GitHub文件](#)。在[Binder](#)中，你只要输入其对应的GitHub Repository的名字或者URL，就能在云端打开整个Repository，选择你需要的[notebook](#)，你就能看到下图这个界面。



每一个Jupyter的运行单元都包含了In、Out的Cell。如图所示，你可以使用Run按钮，运行单独的一个Cell。当然，你也可以在此基础上加以修改，或者新建一个notebook，写成自己想要的程序。赶紧打开链接试一试吧！

另外，我还推荐下面这些Jupyter Notebook，作为你实践的第一站。

- 第一个是Jupyter官方：https://mybinder.org/v2/gh/binder-examples/matplotlib-versions/mpl-v2.0?filepath=matplotlib_versions_demo.ipynb
- 第二个是Google Research提供的Colab环境，尤其适合机器学习的实践应用：https://colab.research.google.com/notebooks/basic_features_overview.ipynb

如果你想在本地或者远程的机器上安装Jupyter Notebook，可以参考下面的两个文档。

安装：<https://jupyter.org/install.html>

运行：<https://jupyter.readthedocs.io/en/latest/running.html#running>

总结

这节课，我为你介绍了Jupyter Notebook，并告诉你它为什么日趋成为Python社区的必学技术。这主要是因为它的三大特点：**整合所有的资源、交互性编程体验和零成本重现结果**。但还是那句话，学习技术必须动手实操。这节课后，希望你能自己动手试一试Jupyter Notebook，后面我们的一些课程代码，我也会用Jupyter Notebook的形式分享给你。

思考题

你尝试Jupyter Notebook了吗？欢迎在留言区和我分享你的使用体验。



Python 核心技术与实战

系统提升你的 Python 能力

景霄
Facebook 资深工程师



新版升级：点击「 请朋友读」，20位好友免费读，邀请订阅更有**现金**奖励。

精选留言：

- kyle 2019-05-13 07:48:52
个人认为，工具类的介绍，一个工具没必要花一篇的时间去讲，文中提一下，读者自己去弄就行了~~ [41赞]

作者回复2019-05-13 12:16:46

你的自学能力很强很好，但是我们的专栏设计要兼顾各种背景阶段的同学。

- PyGeek 2019-05-13 08:24:30

Jupyter本地的两种安装办法：

- 1.安装Anaconda。conda包含科学计算的几乎所有包，包含jupyter。
- 2.仅安装了Python。可以pip install ipython, jupyter。安装即可。

命令行启动jupyter：jupyter notebook [9赞]

- 老贼 2019-05-13 00:44:20

jupyter lab呢？ [6赞]

作者回复2019-05-13 12:18:43

Jupyter lab和binder慢慢整合了，就是云端环境

- 江湖海 2019-05-13 07:46:57

工于善其事，必先利其器。把更多的精力和资源用到去解决实际最关键的问题，很好，谢谢老师的推荐。 [5赞]

作者回复2019-05-13 12:16:51

嗯嗯

- Fullmetal Alchemist 2019-05-13 07:28:51

一直使用Pycharm的咋办？表示从来没有用过Jupyter [5赞]

作者回复2019-05-13 12:17:10

如果你现在的问题都能解决也可以。

- 契卡 2019-05-13 11:16:06

个人是一个生物信息学的研三学生，科研工作中大量涉及数据分析。我把Jupyter Notebook安装在课题组服务器上，用个人电脑远程连接使用。这样就可以在任何地方用一个浏览器就使用服务器的强大性能来处理课题组数据。ipynb也很方便和同学分享协作，他们可以很清晰准确地看到我处理数据的方法、过程和结果。向低年级新生传授知识时，Jupyter的交互性也使得演示变得快速简单直接。 [3赞]

- 坎中满 2019-05-13 07:50:25

老师，我用的pharcm可以吗

[2赞]

作者回复2019-05-13 12:15:54

可以，另一个留言也回复到了。工具要看应用场景。

- Tango 2019-05-13 07:44:45

第三天打卡。除了丘比特笔记本，还喜欢用vscode。 [2赞]

作者回复2019-05-13 12:16:56

哈哈

- Geek_f4ef24 2019-05-13 07:19:34

安装Anaconda就有了这个工具，可能用的少，我并没有发觉到它带来的革命性。 [2赞]

作者回复2019-05-13 12:17:17

可以尝试一下

- Geek_59f23e 2019-05-13 21:32:50
pycharm2019.1已经支持jupyter notebooks了。老师好，我现在在一家培训机构学python，8月份学完出来就要找工作了，那时候正好我们的课也结束了，期待老师后面多分享点干货，对初入职场的小白比较有用^_^ [1赞]
- cronusqiu 2019-05-13 20:36:34
特地增加一篇来讲工具，觉得不是很值得，不同人使用工具习惯场景不同，很像市面上绝大部分书开头都是介绍环境，语法，hello world，我觉得这个平台还是应该注重语言自身的核心，已经哪些大厂如何很好地利用它来实现功能以及一些好的思想框架 [1赞]
- Danpier 2019-05-13 18:40:54
个人使用Anaconda+VSCode。Anaconda包含了conda和jupyter notebook，conda方便管理切换python环境，jupyter notebook非常适合用来学习数据分析。至于VSC既可以运行调试代码，还可以直接新建一个终端在交互模式下跑代码，支持powershell和ipython。相比PyCharm，内存占用也比较低，对我那台老笔记非常友好。但是VSC刚上手需要学习配置一下工具环境。如果是小白，这点不介意的话，个人推荐Anaconda+VSCode，不然就直接Anaconda+PyCharm。
最后附个清华大学开源软件镜像站方便下载Anaconda（官网下载网速可能比较慢，要下载就赶紧。由于目前没有取得Anaconda官方授权，5月16号清华大学将彻底关闭下载入口）
<https://mirrors.tuna.tsinghua.edu.cn/anaconda/archive/?C=M&O=A> [1赞]
- 郭杰 2019-05-13 10:28:28
谢谢啦，我已经开始重视这个工具了
[1赞]
- Vincent 2019-05-14 06:06:42
Jupyter确实是一个不错的工具。
但是我一般用python写一些工程类的应用。
jupyter在组织代码的表现不够idea这类IDE差。比如代码跳转，方法补全，包组织方面不够好。
也许是我见识少，希望老师和同学们不吝赐教。
- Geek_6cf7a7 2019-05-14 00:11:58
景老师，您好，我的系统环境是windows 看网上说window s不适合python 是不是要在linux上搞，输求解惑
- vivien_zh 2019-05-14 00:02:36
第三天打卡。明天一早起来折腾环境！搞起搞起
- 仙鱼 lggl.cn 2019-05-13 22:19:08
第一次听说丘比特，来射我吧!python!☺☺☺
- 蜉蝣 2019-05-13 20:00:55
jupyter 的交互性很爽，通常用它写些小 demo，或者测试一些语法。平时还是用 pycharm 后台开发，并且暂时感觉它无可替代。但还是期待 vsc 突发奇力，让我把 pycharm 这个笨重家伙换掉 @_@
- GLADIATOR 2019-05-13 19:59:57
Mac应该可以吧，回去装一下，一直用的pycharm

- Geek_e918c5 2019-05-13 19:40:14

想问问老师，jupyter 的配置文件jupyter_notebook_config.py不存在，手动配置：jupyter notebook --generate-config也不成功怎么办啊，网上很多方法都失败了