

## 第2章: Python介绍

数据科学实战：Python篇

讲师：Ben

# 自我介绍

- 天善商业智能和大数据社区      讲师 –Ben
- 天善社区 ID - Ben\_Chang
- <https://www.hellobi.com> – 学习过程中有任何相关的问题都可以提到技术社区数据挖掘版块。

# 1 Python简介



# Python是什么？

## Python 官方网站的描述

Python is a programming language that lets you work more quickly and integrate your systems more effectively. You can learn to use Python and see almost immediate gains in productivity and lower maintenance costs.

## 《简明Python教程》的描述

Python是一种简单易学，功能强大的编程语言。它有高效率的高层数据结构，能够简单、有效地实现面向对象编程。





















Python语法简洁，支持动态输入，是解释性语言。

在大多数平台上，对于众多领域，Python都是一个理想的开发语言，特别适合于应用程序的快速开发。

# Why Python ?

## IEEE Spectrum 的排行

在 2016 年排名第三的 Python 在今年已经成为世界上最受欢迎的语言

Language Rank	Types	Spectrum Ranking
1. Python	 	100.0
2. C	  	99.7
3. Java	  	99.5
4. C++	  	97.1
5. C#	  	87.7
6. R		87.7
7. JavaScript	 	85.6
8. PHP		81.2
9. Go	 	75.1
10. Swift	 	73.7

- **作者 : Guido Van Rossum (GvR)**
- <http://www.python.org/~guido/>
- 1989年，为了打发圣诞节假期，Guido开始写Python语言的编译/解释器，Python取名得自Guido喜欢的一部Monty Python's Flying Circus的情景喜剧。
- Python是ABC语言的后代，Guido曾在CWI工作，并参与到ABC语言的开发，Python设计理念是能够像C语言那样能够全面调用计算机的功能接口，又可以像shell那样可以轻松的编程。
- 1991年，第一个Python编译器(同时也是解释器)诞生。它是用C语言实现的，并能够调用C库(.so文件)。



Guido被称为Benevolent Dictator For Life，他持续关注 Python 的开发进程，指导支持Python开源社区的活动，并在必要的时刻做出决定，享有绝对的仲裁权。

- *Computer Programming for Everybody*
- 1999年，Guido向DARPA 阐述Python语言的特性：
  - 简单、直观、强大
  - 开源，以便任何人都可以为它做贡献
  - 代码像纯英语那样容易理解
  - 适用于短期开发的日常任务
- 其它包括可扩展性、丰富的库、面向对象、高级语言、可移植性等  
等特点

# Python缺点

- 强制缩进
- 构架选择太多
- 性能问题



## 2 Anaconda的安装与使用



# Anaconda简介

- ❑ Anaconda Python 是 Python 科学计算包的合集，包含了常用科学计算、数据分析、自然语言处理、绘图等包，所有的包几乎都是最新的，容量适中。
- ❑ 使用了conda和pip包管理工具，安装第三方包非常方便，避免了管理各个库之间依赖性的麻烦。
- ❑ 安装后就默认安装了python、IPython、集成开发环境Spyder和众多的包和模块。支持Python2和Python3，包括免费版、协作版、企业版等。

Anaconda Python 是完全免费的企业级的Python发行版，用于大规模数据处理、预测分析和科学计算。

# Anaconda下载地址

<https://www.anaconda.com/download/>



Windows



macOS



Linux

## Anaconda 5.1 For Windows Installer

Python 3.6 version \*

↓ Download

[64-Bit Graphical Installer \(537 MB\)](#) [?](#)

[32-Bit Graphical Installer \(436 MB\)](#)

Python 2.7 version \*

↓ Download

[64-Bit Graphical Installer \(523 MB\)](#) [?](#)

[32-Bit Graphical Installer \(420 MB\)](#)

# 3 Python数据分析包简介

# 常用数据分析包

## 常用数据分析库\包：

- numpy：数组、矩阵
- scipy：统计、优化、线性代数、信号、图像处理等
- pandas：数据框、序列、数据处理、绘图
- statsmodel：统计检验、统计建模
- scikit-learn：数据变换、机器学习、交叉验证
- matplotlib：绘图
  
- seaborn：高级统计制图包

## 安装包：

首选：CDM->conda install 包名

失败后尝试：CDM-> pip install 包名

—— 秦路主讲 ——  
**七周成为数据分析师**  
七周为期，Get一条数据分析师职业黄金通道！



—— Python ——  
**数据分析与挖掘**  
集Python爬虫、数据采集、数据处理、数据分析与数据挖掘于一体，打造Python全栈工程师  
主讲老师：韦玮  
VIP会员群+在线答疑+录播复习+1年反复观看



**案例为师，实战为王**  
开启Python机器学习之路  
科学规划全套课程体系，从入门到进阶，从理论到技巧，嵌入丰富课程案例讲解，逐步推进  
讲师：唐宇迪 深度学习领域多年一线实践研究专家



**独一无二的  
数据仓库**建模指南系列教程升级版  
• 从企业视角进行数据规划以及数据仓库模型的搭建  
• 高质量的数据库模型和技巧，以及丰富的例子  
• 数据仓库架构理论和实践要领  
资深讲师：BAO胖子 15年+BI从业经验  
涉足电力、快消品、医药、信息服务行业的BI老兵



**业务知识一站通**  
技术+业务，挣钱有门路！  
—— 讲师：陈文 ——



自己动手 丰衣足食  
**Python3网络爬虫实战案例**  
—— 循序渐进，案例为王，诠释全面，思路制胜 ——  
讲师：崔庆才 北航硕士，百万级热度爬文博主



讲师 丘祐玮  
**人人都爱数据科学家**  
Python数据科学精华实战课程



**数据分析报告制作**  
秘籍升级版  
讲师：陈丹奕 知乎大神，前百度资深数据分析师



**先机致胜 破冰AI**  
—— 深度学习模型/框架与实战 ——  
讲师：唐宇迪 同济大学硕士  
深度学习领域多年一线实践研究专家



BI、商业智能  
数据挖掘 大数据  
数据分析师  
R语言 Python  
机器学习  
深度学习  
人工智能  
Hive Hadoop  
Tableau  
BIEE ETL  
数据科学家  
PowerBI