

课程说明

- 1、课程的意义
 - (1) 大数据时代意义；(2) 商务智能意义。
- 2、课程要求
 - (1) 上课认真听讲，做笔记；(2) 认真完成课程编程作业。
- 3、课程安排考核
 - (1) 数据采集：Requests和Scrapy，MySQL
 - (2) 数据处理：Numpy，Pandas，Matplotlib，Scikit-learn
 - (3) 最后一次课，实验考试。

课程说明

4，课程总体内容安排

章节序号	章节名称	理论学时	实验学时
1.3~1.4	安装 Python 的 <u>anaconda</u> 发行版和 <u>Jupyter</u> 使用	1	
补充材料	网络数据爬取 1-Requests 库入门	2	
补充材料	网络数据爬取 2-Requests 库网络爬取实战	2	
补充材料	网络数据爬取 3-Scrapy 爬虫框架	2	
补充材料	网络数据爬取 4-Scrapy 爬虫基本使用	2	
补充材料	网络数据爬取 5-实验 1-股票数据定向 <u>Scrapy</u> 爬虫	2	4
第 1 章 1.1~1.2	Python 数据分析概述	1	
第 2 章	<u>NumPy</u> 数值计算基础	4	
第 3 章	<u>Matplotlib</u> 数据可视化基础	4	4
第 4 章 4.1~4.3	<u>pandas</u> 统计分析基础(1)	2	
第 4 章 4.4~4.5	<u>pandas</u> 统计分析基础(2)	2	
第 5 章	使用 <u>pandas</u> 进行数据预处理	2	4
第 6 章	使用 <u>scikit-learn</u> 构建模型	3	4
	实验编程考试	3	
合计		32	16

课程说明

5、意见及反馈方式

➤ 552532999@qq.com

课程软件下载位置

链接: <https://pan.baidu.com/s/1419Cwe0HPAA4ZZOD3lgTeQ>
提取码: 6w87

资料网站:https://github.com/lilin201501/data_acquisition_analysis

目录

1

安装 Python 的 Anaconda 发行版

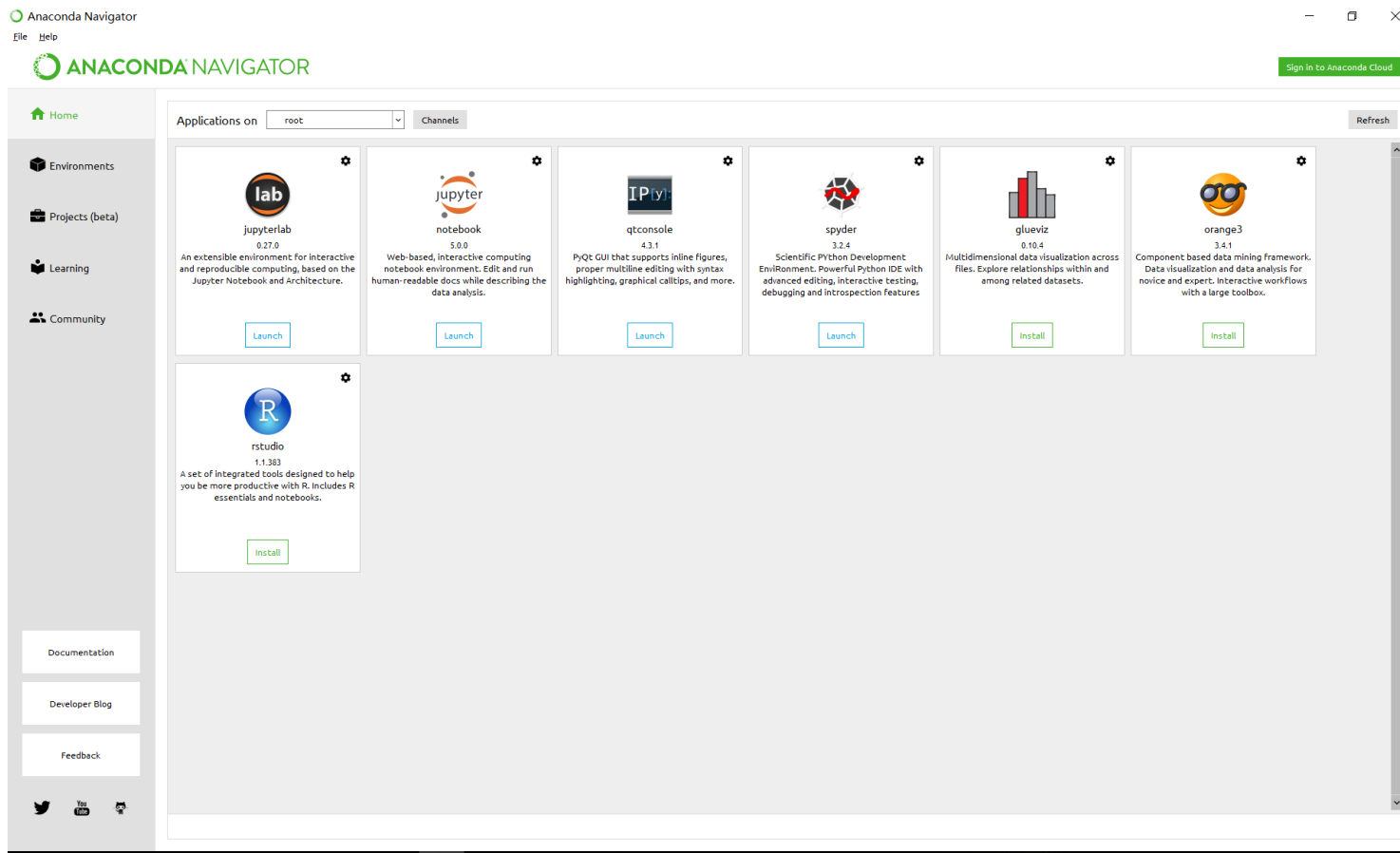
2

掌握 Jupyter Notebook 常用功能

了解 Python 的 Anaconda 发行版

Anaconda

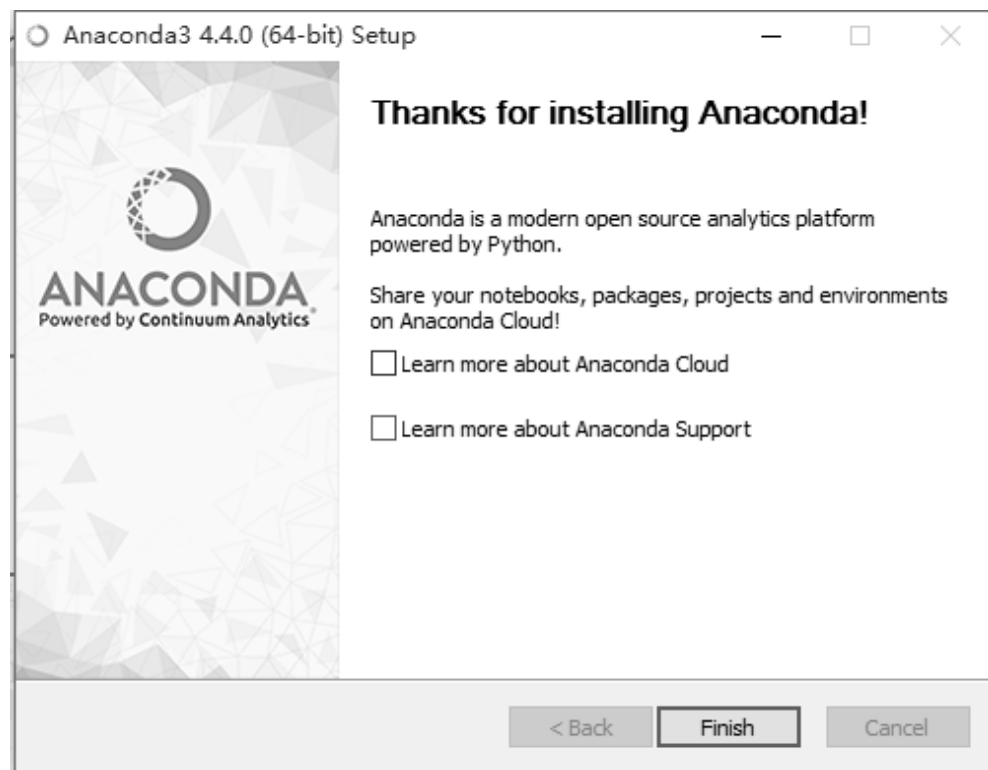
- 预装了大量常用 Packages。
- 完全开源和免费。
- 额外的加速和优化是收费的，但对于学术用途，可以申请免费的 License。
- 对全平台和几乎所有Python版本支持。



在 Windows 系统上安装 Anaconda

安装流程

安装包——“next”——“I agree”——“All Users(requires admin privileges)”——选择安装路径——“Install”——“finish”。



目录

1

安装 Python 的 Anaconda 发行版

2

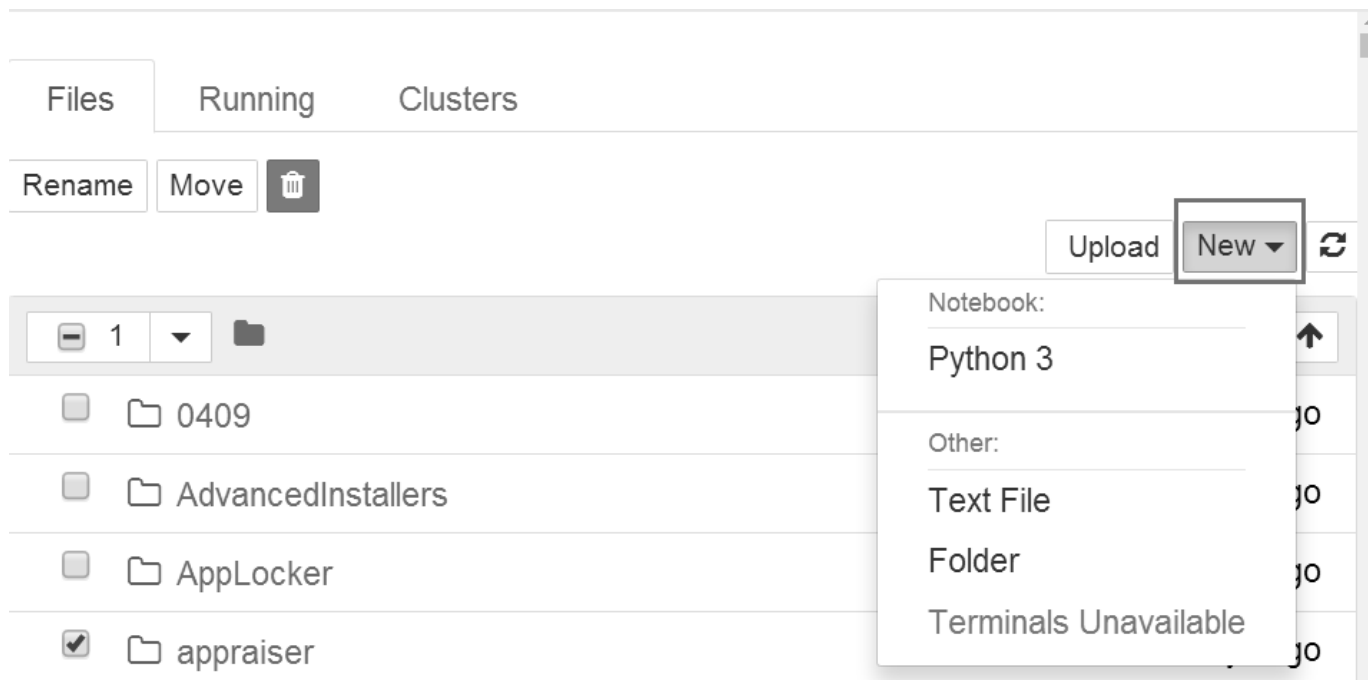
掌握 Jupyter Notebook 常用功能

掌握 Jupyter Notebook 的基本功能

打开并新建一个Notebook

打开 Jupyter Notebook

- “Text File” 为纯文本型
- “Folder” 为文件夹
- “Python 3” 表示 Python 运行脚本

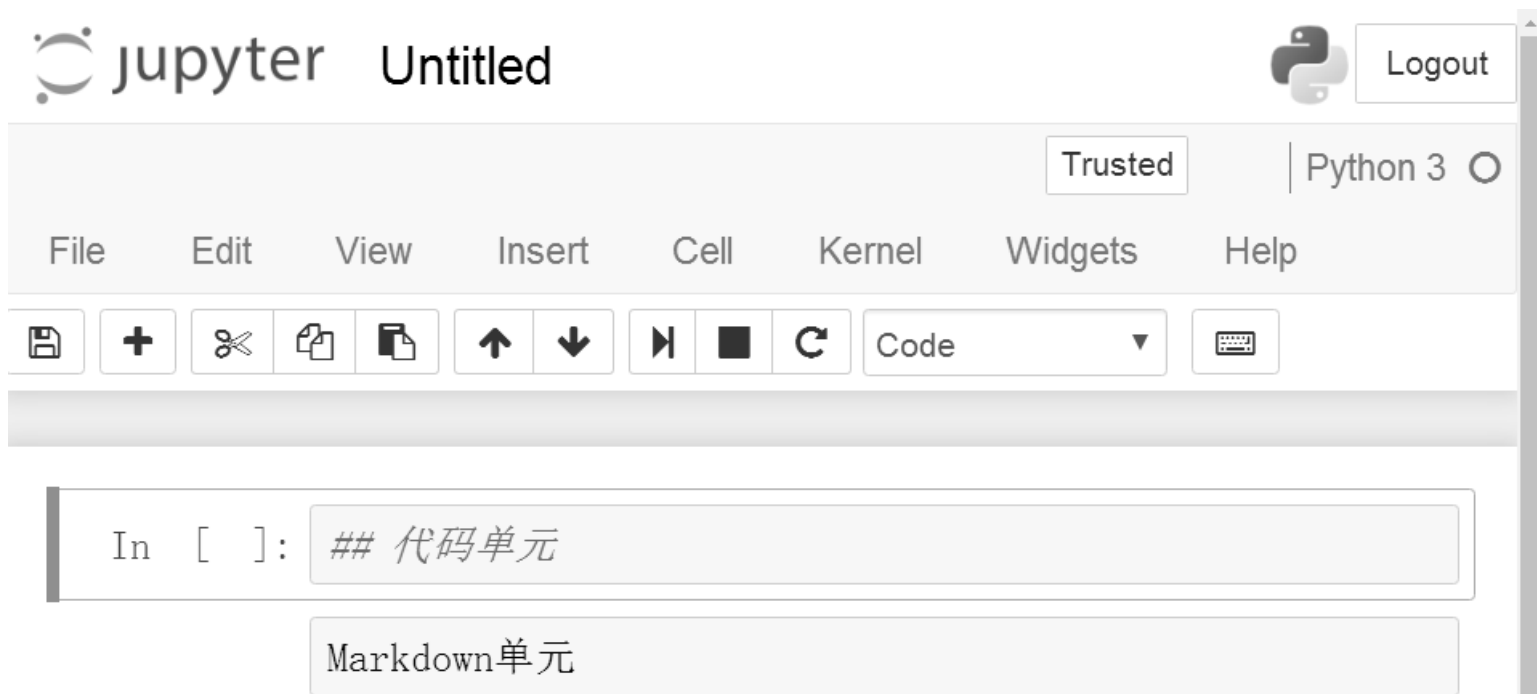


掌握 Jupyter Notebook 的基本功能

Jupyter Notebook 的界面及其构成

选择“Python 3”选项，进入 Python 脚本编辑界面，Notebook 文档由一系列单元（Cell）构成，主要有两种形式的单元。

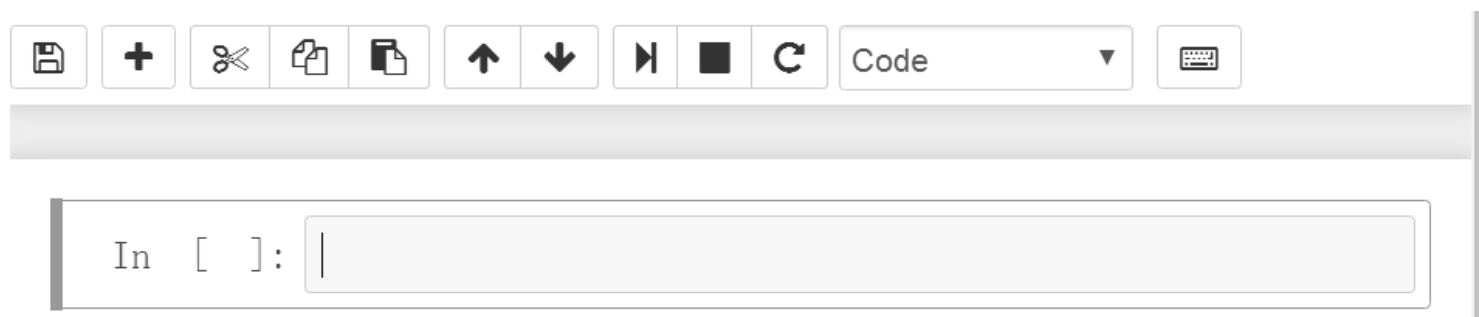
- 代码单元。这里是读者编写代码的地方。
- Markdown 单元。在这里对文本进行编辑。



掌握 Jupyter Notebook 的基本功能

编辑界面

- 编辑模式：用于编辑文本和代码。



- 命令模式：用于执行键盘输入的快捷命令。



掌握 Jupyter Notebook 的基本功能

快捷键

- “Esc” 键：进入命令模式
- “Y” 键：切换到代码单元
- “M” 键：切换到 Markdown 单元
- “B” 键：在本单元的下方增加一单元
- “H” 键：查看所有快捷命令
- “Shift + Enter” 组合键：运行代码

掌握 Jupyter Notebook 的高级功能

1 . Markdown

Markdown 是一种可以使用普通文本编辑器编写的标记语言，通过简单的标记语法，它可以使普通文本内容具有一定的格式。

➤ 标题：标题是标明文章和作品等内容的简短语句。一个 “#” 字符代表一级标题，以此类推。

掌握 Jupyter Notebook 的高级功能

1 . Markdown

- 列表：列表是一种由数据项构成的有限序列，即按照一定的线性顺序排列而成的数据项的集合。
- 对于无序列表，使用星号、加号或者减号作为列表标记
- 对于有序列表，则是使用数字 “，” “（一个空格）”。

```
* Python  
+ Python2  
- Python3  
  
1. Python  
2. Python2  
3. Python3
```

```
• Python  
• Python2  
• Python3  
  
1. Python  
2. Python2  
3. Python3
```

掌握 Jupyter Notebook 的高级功能

1 . Markdown

- 加粗 / 斜体：前后有两个星号或下划线表示加粗，前后有 3 个星号或下划线表示斜体。

Python数据分析

****Python数据分析****

******Python数据分析******

__Python数据分析__

__python数据分析__

Python数据分析

Python数据分析

Python数据分析

Python数据分析

python数据分析

掌握 Jupyter Notebook 的高级功能

1 . Markdown

- 表格：代码的第一行表示表头，第二行分隔表头和主体部分，从第三行开始，每一行代表一个表格行；列与列之间用符号 “|” 隔开，表格每一行的两边也要有符号 “|” 。

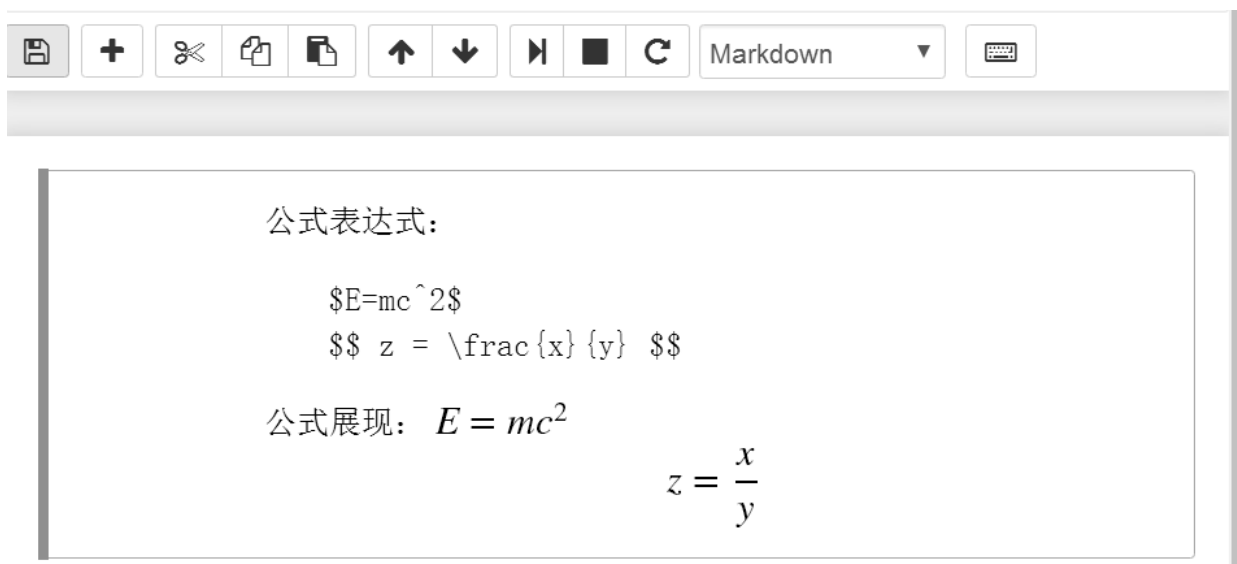
```
Python | R| MATLAB |  
-----|-----|----|  
接口统一，学习曲线平缓 | 接口众多，学习曲线陡峭 | 自由度大，学习曲线较为平缓 |  
开源免费 | 开源免费 | 商业收费 |
```

Python	R	MATLAB
接口统一，学习曲线平缓	接口众多，学习曲线陡峭	自由度大，学习曲线较为平缓
开源免费	开源免费	商业收费

掌握 Jupyter Notebook 的高级功能

1 . Markdown

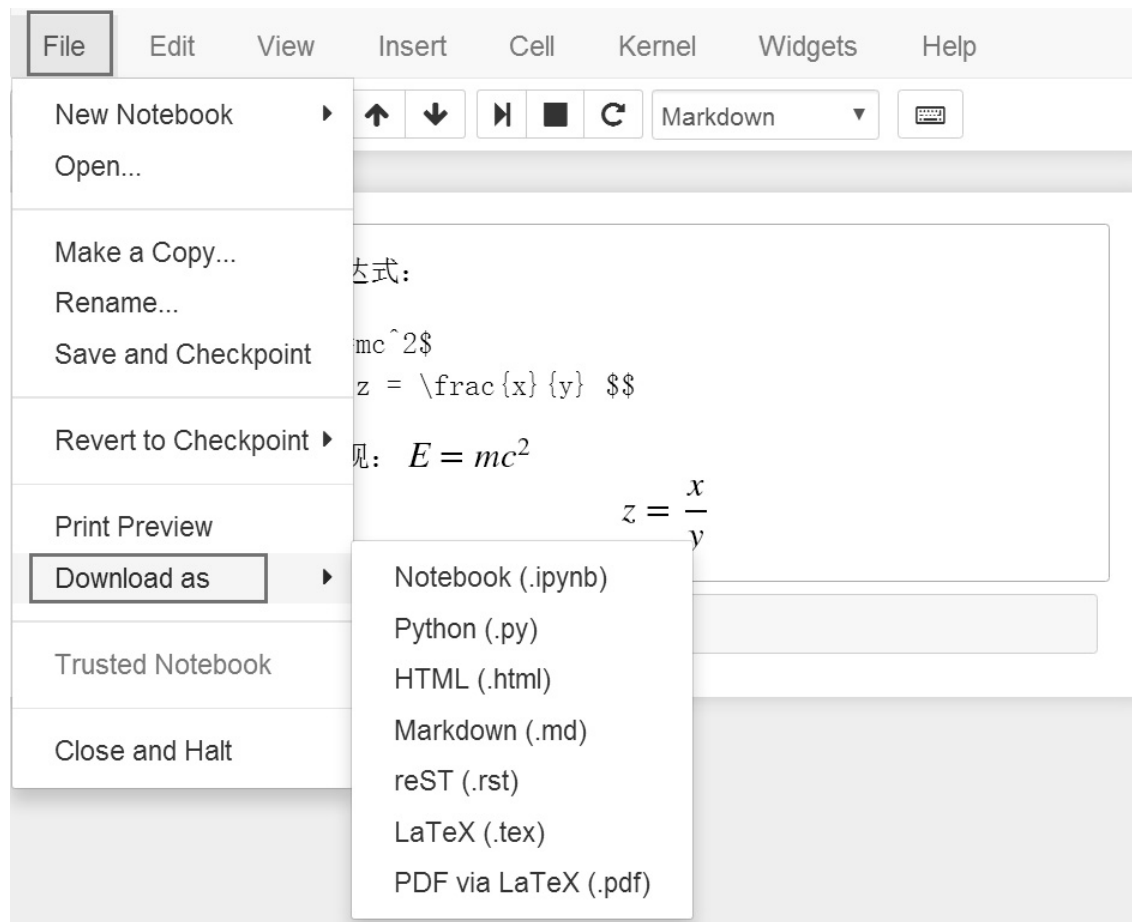
- 数学公式编辑：LaTeX 是写科研论文的必备工具，Markdown 单元中也可以使用 LaTeX 来插入数学公式。
- 在文本行中插入数学公式，应在公式前后分别加上一个 “\$” 符号
- 如果要插入一个数学区块，则在公式前后分别加上两个 “\$\$” 符号。



掌握 Jupyter Notebook 的高级功能

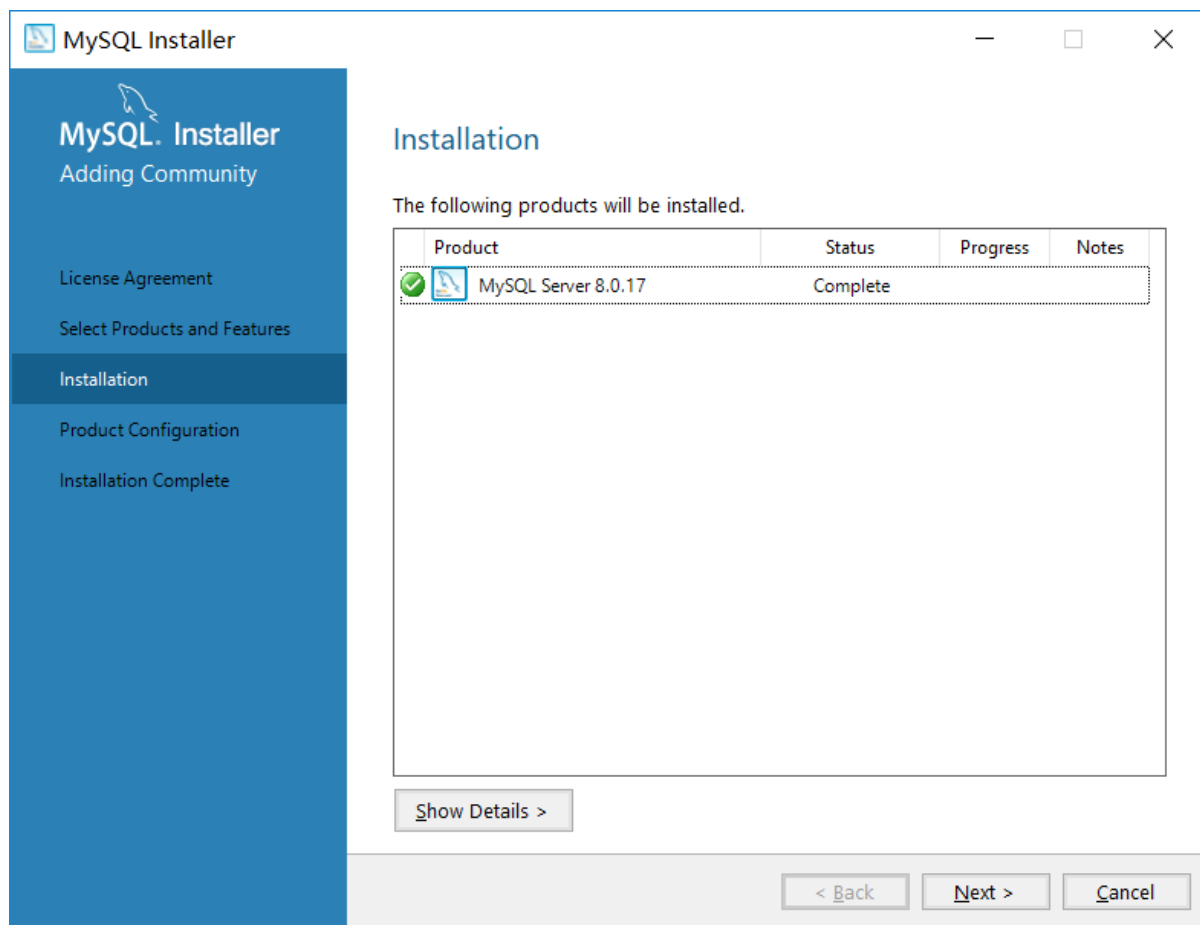
2. 导出功能

- Notebook 还有一个强大的特性，就是导出功能。可以将 Notebook 导出为多种格式，如HTML、Markdown、reST、PDF（通过 LaTeX）等格式。
- 导出功能可通过选择“File” → “Download as”级联菜单中的命令实现。



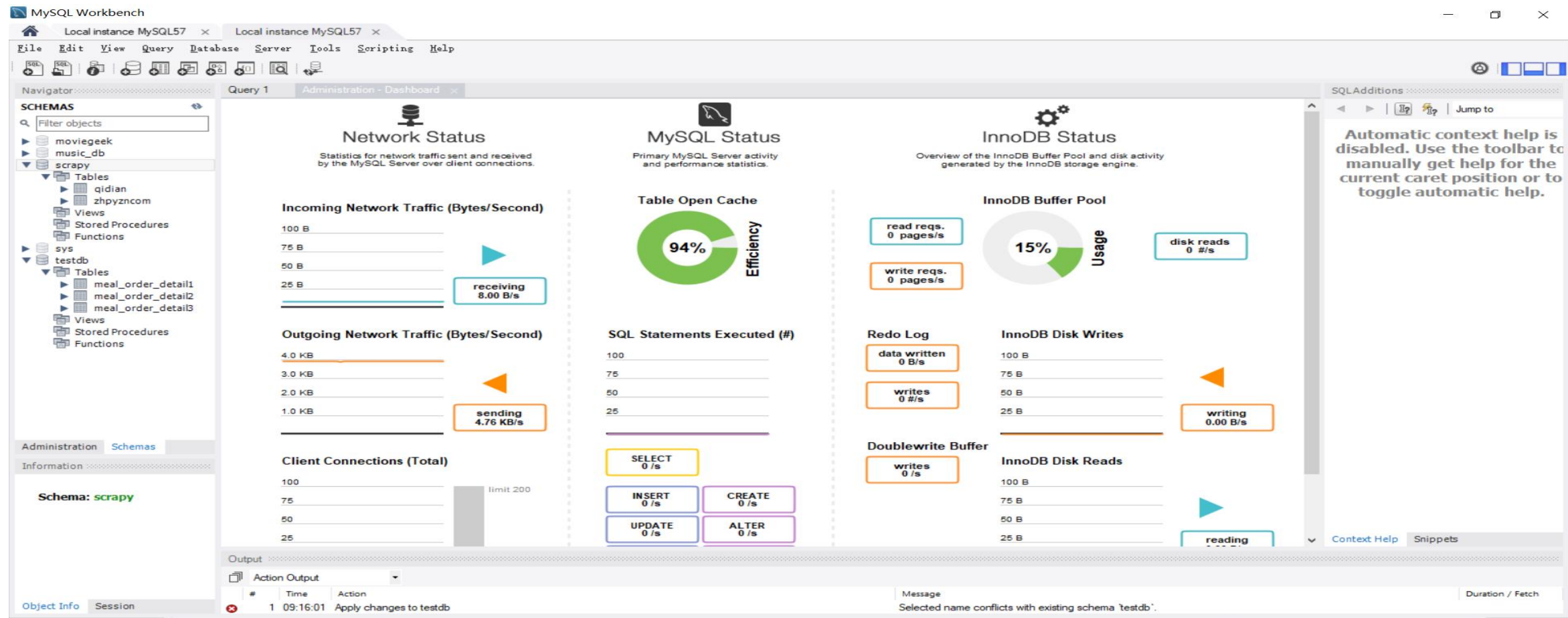
MySQL数据库安装

服务器安装

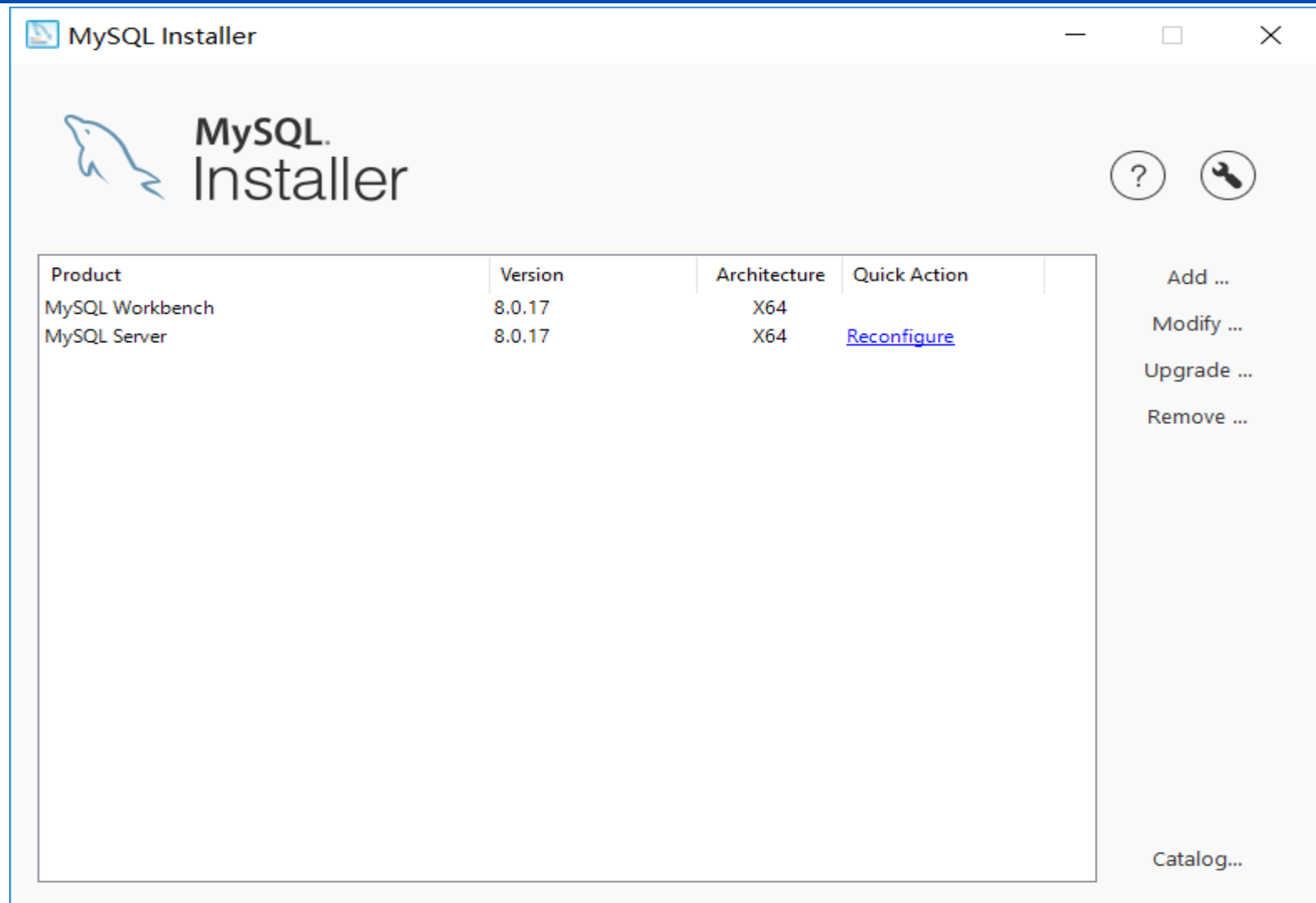


MySQL数据库安装

客户端连接



MySQL数据库安装



MySQL数据库管理

MySQL Workbench

Local instance MySQL57 x Local instance MySQL57 x

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

Filter objects

moviegeek

music_db

scrapy

Tables

qidian

zhpyzncom

Views

Stored Procedures

Functions

sys

testdb

Tables

meal_order_detail1

meal_order_detail2

meal_order_detail3

Views

Stored Procedures

Functions

Administration Schemas

Information

Table: meal_order_detail1

Columns: detail_id varci order_id varci dishes_id varci logicprn_name varci parent_class_name varci

Object Info Session

qidian x meal_order_detail1

Limit to 1000 rows

1 • SELECT * FROM scrapy.qidian;

Result Grid

ID	title	author	information	introduction	word_num	clicks_num	recommen
1	全球高武	老鹰吃小鸡	重生过去、畅想未来、梦幻现实，再塑传奇...	地窟入侵，武道崛起。	762.04万字	872.95万总会员点击	10.79万总
2	玄月三声泣	【2019“玄幻王者”征文】参赛作品	《大荒第一嗜帝》沙雕1号：普通群58583...	《大荒第一嗜帝》沙雕1号：普通群58583...	40.69万字	6778万总会员点击	3096万总
3	我有一座冒险屋	我会修空调	考古探险、鉴宝收藏，侦探推理、诡秘分析...	【本作品简体实体书由次元书馆出版】	171.81万字	380.84万总会员点击	3.34万总
4	黎明之剑	远瞳	星海漫游，时空穿梭，机械科技，目标是未...	高文穿越了，但穿越的时候稍微出了点问题。	225.46万字	487.95万总会员点击	4.98万总
5	伏天氏	净无痕	心潮澎湃，无限幻想，迎风挥击千层浪，少...	东方神州，有人皇立道统，有圣贤宗门传道...	407.25万字	518.01万总会员点击	1.45万总
6	轮回乐园	那一只蚊子	热血的少年，为打破次元壁一往无前！	苏晓与轮回乐园签订契约，穿梭在各个动漫...	530.6万字	252.53万总会员点击	3.09万总
7	我真没想出名啊	巫马行	重生过去、畅想未来、梦幻现实，再塑传奇...	那一年，穷疯了的陆远拿着五个小时写好的...	165.65万字	75.04万总会员点击	1.89万总
8	全职高手	蝴蝶蓝	拥有十年游戏经验的他，带着对往昔的回忆...	网游荣耀中被誉为教科书级别的顶尖高手，...	534.97万字	563.46万总会员点击	7183万总
9	大唐第一狠人	山下出水	醒掌天下权，醉卧美人膝，五千年风华烟雨...	民间故老相传，将不过李，王不过霸，但是...	100.59万字	101.29万总会员点击	658万总
10	学霸的黑科技系统	晨星LL	星海漫游，时空穿梭，机械科技，目标是未...	“系统，积分能兑钱吗？”	299.61万字	436.93万总会员点击	3.56万总
11	修真聊天群	圣骑士的...	宋书航意外加入一个仙侠中二病资深患者的...	某天，宋书航意外加入了一个仙侠中二病资深...	780.71万字	1461.47万总会员点击	5.5万总
12	超神机械师	齐佩甲	肆意挥洒激情的游戏人生，打破现实框架的...	韩萧，《星海》骨灰级代练，被来自东（zu...	332.52万字	433.17万总会员点击	4.03万总
13	明朝败家子	上山打老...	醒掌天下权，醉卧美人膝，五千年风华烟雨...	重要通知，明朝败家子，又名《明颂》。	477.59万字	346.22万总会员点击	2.46万总
14	穹顶之上	人间武库	重生过去、畅想未来、梦幻现实，再塑传奇...	韩青禹生在一个铸就史诗的年代，平静地生...	125.94万字	175.84万总会员点击	2.01万总
15	拜见教主大人	封七月	借三尺明月，衔两袖青龙，轻剑快马恣意，...	意外身亡，楚休发现自己竟然穿越到了游戏...	404.35万字	166.07万总会员点击	8348万总
16	退后让为师来	隐语者	星海漫游，时空穿梭，机械科技，目标是未...	穿越唐僧，取经功成。	193.56万字	41.06万总会员点击	7874万总
17	还看今朝	瑞根	重生过去、畅想未来、梦幻现实，再塑传奇...	当扑面而来的时代巨澜把懵懵懂懂的沙正阳...	379.02万字	280.99万总会员点击	1.47万总
18	漫威里的德鲁伊	骑行拐杖	星海漫游，时空穿梭，机械科技，目标是未...	暗黑2召唤系德鲁伊上身，穿越成为漫威世...	270.9万字	56.81万总会员点击	6041万总
19	边个	七条	醒掌天下权，醉卧美人膝，五千年风华烟雨...	我们接受了祖先的遗言，让让中化超越了数...	381.50万字	626.08万总会员点击	1705万总

qidian 1 x

Output

Action Output

#	Time	Action	Message	Duration / Fetch
2	09:21:00	SELECT * FROM scrapy.qidian LIMIT 0, 1000	376 row(s) returned	0.000 sec / 0.000 sec

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

VS CODE 安装

安装程序 - Microsoft Visual Studio Code (User)	安装程序 - Microsoft Visual Studio Code (User)
<p>许可协议</p> <p>请在继续操作前阅读以下重要信息。</p>	<p>选择其他任务</p> <p>应执行哪些其他任务?</p>
<p>请阅读以下许可协议。必须接受此协议条款才可继续安装。</p> <div><p>此许可适用于 Visual Studio Code 产品。Visual Studio Code 的源代码可根据 MIT 许可协议 (https://github.com/Microsoft/vscode/blob/master/LICENSE.txt) 在以下网址获取: https://github.com/Microsoft/vscode。有关其他许可信息, 请查看我们的常见问题解答, 网址为 https://code.visualstudio.com/docs/supporting/faq。</p></div> <p><input checked="" type="radio"/> 我接受协议(A)</p> <p><input type="radio"/> 我不接受协议(D)</p>	<p>选择安装 Visual Studio Code 时希望安装程序来执行的其他任务, 然后单击“下一步”。</p> <p>其他快捷方式:</p> <p><input type="checkbox"/> 创建桌面快捷方式(D)</p> <p>其他:</p> <p><input checked="" type="checkbox"/> 将“通过 Code 打开”操作添加到 Windows 资源管理器文件上下文菜单</p> <p><input checked="" type="checkbox"/> 将“通过 Code 打开”操作添加到 Windows 资源管理器目录上下文菜单</p> <p><input checked="" type="checkbox"/> 将 Code 注册为受支持的文件类型的编辑器</p> <p><input checked="" type="checkbox"/> 添加到 PATH (重启后生效)</p>
<p>下一步(N) > 取消</p>	<p>< 上一步(B) 下一步(N) > 取消</p>