

Architect

SACC

2022 中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2022

· 激发架构性能 点亮业务活力

云上会议 网络直播 | 2022年10月27-29日

IT168.com

ChinaUnix.net

ITPUB

大型Python项目开发中的代码自动化标记、分析、调整与测试： *Bright*包的功能设计与基本原理

汪建雄

中央财经大学中国经济与管理研究院（CEMA）
北京第二外国语学院经济学院金融系

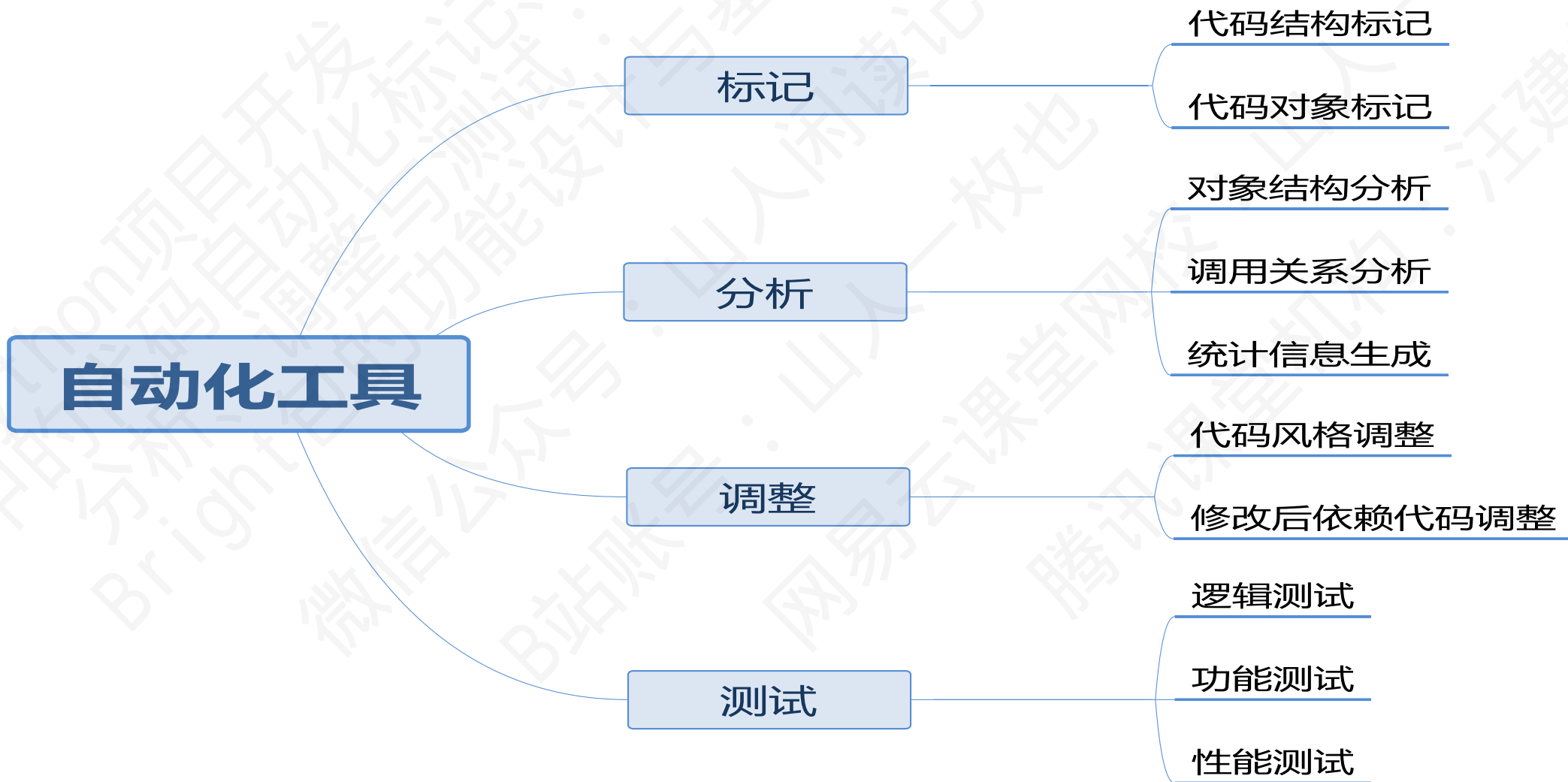
金融学博士
副教授

为啥需要自动化标记、分析、调整、测试工具？

PART-0

- 提高开发效率，加快代码迭代
 - 全面、快速、风格化的代码分析与评审
 - 统一代码风格
 - 方便代码重构
 - 提高测试效率

整体设计原理



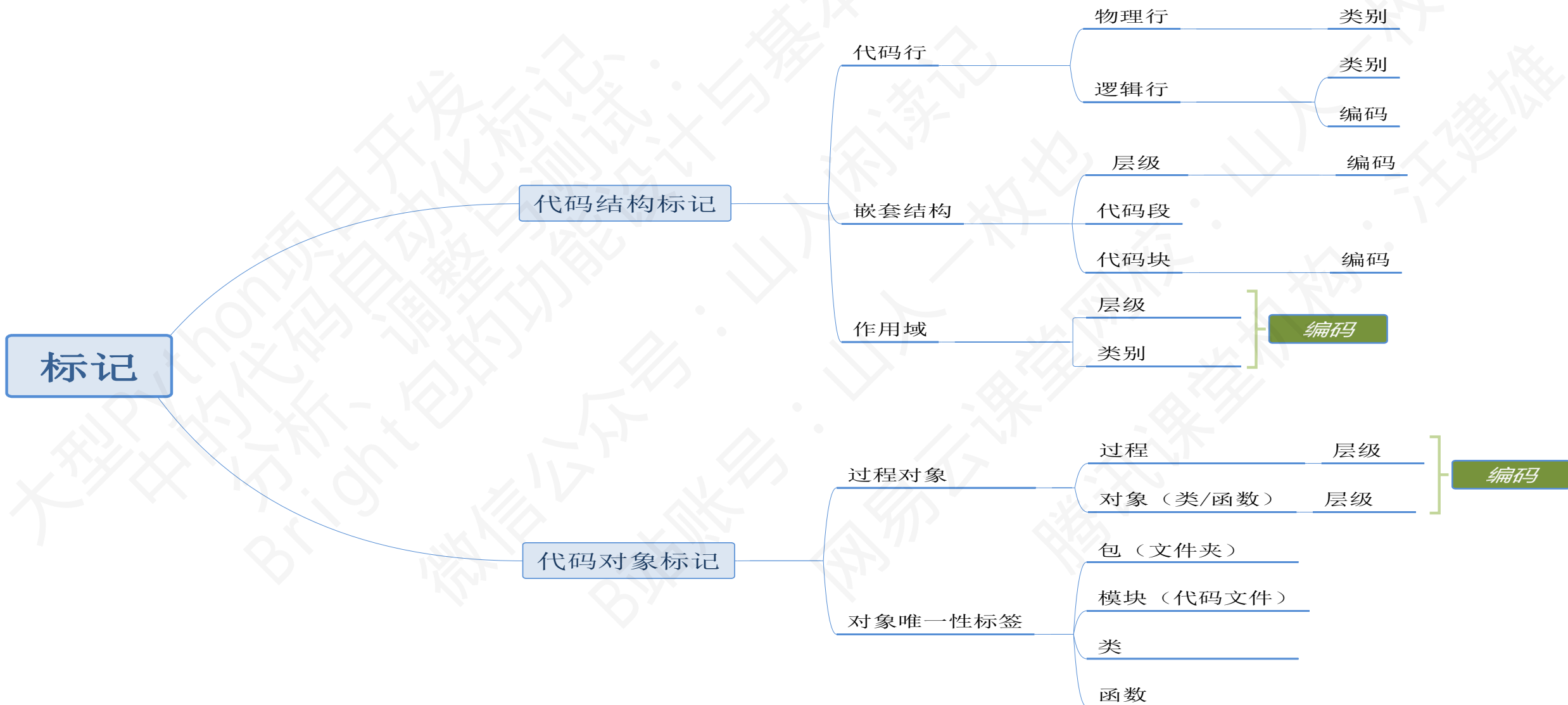
PART-0

整体设计原理

- 标记：代码结构标记；代码对象标记
- 分析：对象结构分析；调用关系分析；统计信息生成
- 调整：代码风格调整；修改后依赖代码调整
- 测试：逻辑；功能；性能

PART-1

自动化标记：整体设计



PART-1

自动化标记：代码行

• 概念与原理

- 概念：物理行，逻辑行
- 原理：基于人为换行标记

• 类别

- 物理行：（多行注释_首行、多行注释、多行注释_尾行）、单行注释、空行、纯代码、代码+注释
- 逻辑行：空行、单行注释、多行注释、导入语句、类定义_首行、函数定义_首行、函数定义_返回值、装饰器定义、控制语句_分支、控制语句_循环、控制语句_环境管理、控制语句_异常处理、普通语句_无显性赋值、普通语句_普通赋值、普通语句_增强赋值

• 编码

- 原理：仅针对“逻辑行”

自动化标记：嵌套结构

- 基础概念

- 嵌套
- 嵌套层级
- 代码段
- 代码块

- 编码

- 原理：基于“嵌套层级”和“代码块”

PART-1

自动化标记：作用域

- 类别

- 执行代码 (X)
- 类的定义 (C)
- 函数定义 (F)

- 层级

- 编码

- 原理：基于“类别”和“层级”

PART-1

自动化标记：过程对象

- 类别

- 过程 (P)
- 对象：函数 (F)，类 (C)

- 层级

- 编码

- 原理：基于“类别”和“层级”

自动化标记：对象唯一性标签 1

- 标记目标

- 包（文件夹）
- 模块（代码文件）
- 类
- 函数

- 原理与设计

- 原理：生成唯一性标签，确保互不重复
- 方案：

PART-1

自动化标记：代码演示

- 代码行
- 嵌套结构
- 作用域
- 过程对象
- 对象唯一性标签

自动化分析：整体设计

PART-2



自动化分析：对象结构分析

- 模块级对象结构

- 仅含模块（包）信息
 - 不拆分模块名
 - 拆分模块名
- 包含模块内对象（类、函数）信息

- 模块内对象结构

- 仅含模块内对象（类、函数）信息

PART-2

自动化分析：调用关系分析

- 项目外调用
- 项目内模块间调用
 - 导入
 - 包
 - 模块
 - 调用
 - 模块中的类
 - 模块中的函数
 - 模块中的变量
- 模块内调用
 - 类调用
 - 函数调用

PART-2

自动化分析：统计信息生成

- 代码行信息

- 数量
 - 空行
 - 注释
 - 代码
- 结构

- 嵌套结构信息（仅针对代码行，空行、注释行略过）

- 过程代码嵌套结构
- 对象定义代码嵌套结构
- 全部代码嵌套结构

- 过程对象信息

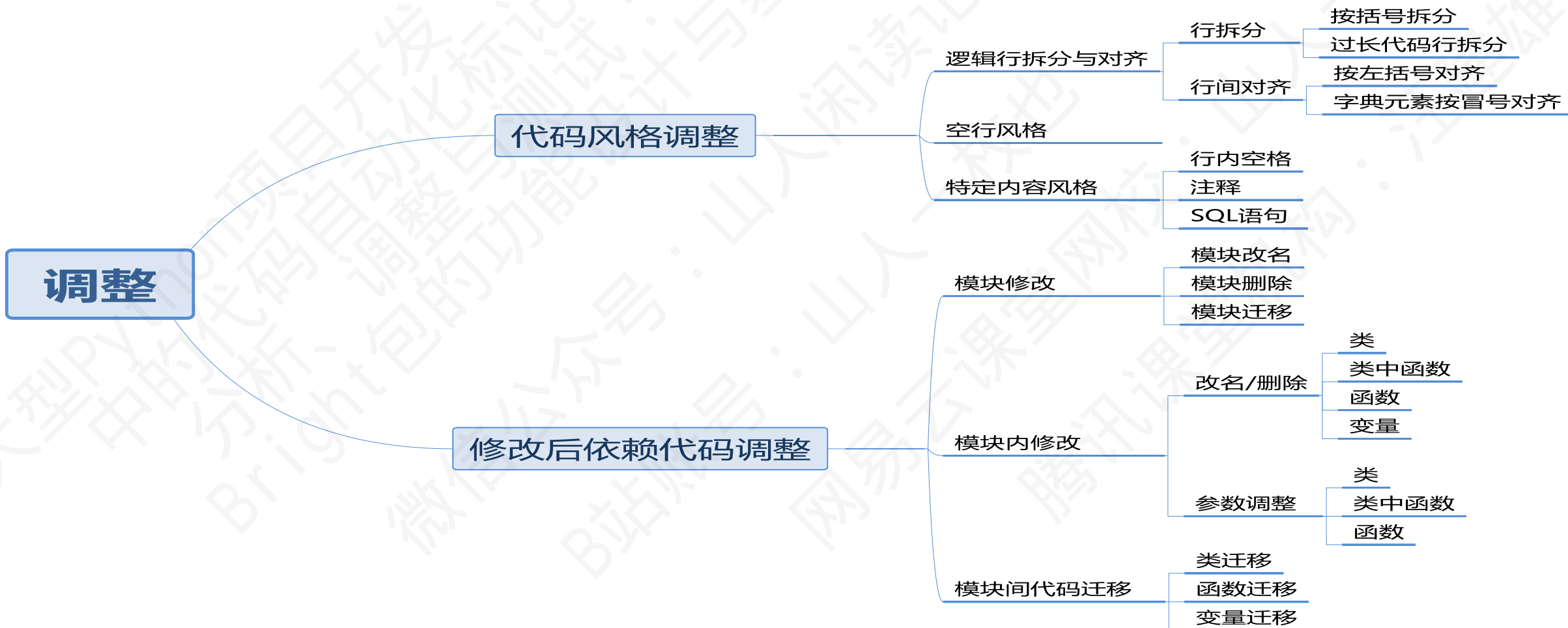
- 数量
- 结构

PART-2

自动化分析：代码演示

- 对象结构分析
- 调用关系分析
- 统计信息生成

自动化调整：整体设计



PART-3

自动化调整：代码风格调整

- 逻辑行拆分与对齐
 - 行拆分
 - 按括号拆分
 - 过长代码行拆分
 - 行间对齐
 - 行首对齐
 - 按左括号对齐
 - 按指定缩进规则对齐
 - 行中对齐
 - 字典元素按冒号对齐
- 空行风格
- 特定内容风格
 - 行内空格
 - 注释
 - SQL语句

自动化调整：修改后依赖代码调整 PART-3

- 模块（包）修改
 - 模块（包）改名
 - 模块（包）删除
 - 模块（包）迁移
- 模块内修改
 - 改名/删除
 - 类
 - 类中函数
 - 函数
 - 变量
 - 参数调整
 - 类
 - 类中函数
 - 函数
- 模块间代码迁移
 - 类迁移
 - 函数迁移
 - 变量迁移

自动化调整：代码演示

- 代码风格调整
- 修改后依赖代码调整

自动化测试：整体设计



PART-4

自动化测试：逻辑测试

- 作用域测试
 - 参数作用域
 - 类参数
 - 类中函数的参数
 - 函数参数
 - 主体代码作用域
 - 直接写于类主体代码中的变量
 - 类中函数主体代码中的变量
 - 函数主体代码中的变量
- 特定工具相关测试
 - 控制语句相关测试
 - DataFrame相关测试
- 常见逻辑错误测试

PART-4

自动化测试：功能测试

- 添加执行标记信息
 - 输出执行行号
 - 输出循环执行轮数

PART-4

自动化测试：性能测试

- 执行速度
- 内存占用
- CPU/GPU资源占用
- I/O状况
- 数据库读写负载

PART-4

自动化测试：代码演示

- 作用域测试
- 功能测试
- 性能测试

代码重要贡献者：樊雪超



Fancy
北京 朝阳



扫一扫上面的二维码图案，加我为朋友。

手机号：17862925273

邮箱：fancy678y@163.com

《并行大数据处理：基于Python、PostgreSQL及其他》

▼ 第1篇 基础：工具与理念

- ▶ CH1 Python
- ▶ CH2 PostgreSQL
- ▶ CH3 编程理念
- ▶ CH4 自动化开发：基于Bright

▼ 第2篇 主题应用：数值、文本、地理及网络数据处理

- ▶ CH5 数值数据处理：基于Pandas
- ▶ CH6 文本数据处理：基于正则表达式
- ▶ CH7 地理数据处理：基于PostGIS和QGIS
- ▶ CH8 社会网络数据处理：基于Neo4j

▼ 第3篇 性能提升：加速、并行及容器化

- ▶ CH9 加速
- ▶ CH10 分布式并行处理
- CH11 Linux：主要基于Ubuntu与CentOS
- ▶ CH12 容器化：基于Docker
- CH13 容器编排：基于Kubernetes
- ▶ 附录A：重要概念备查
- 附录B：硬件环境配置方案
- ▶ 附录C：团队协作编程：基于Git

预计：2023年出版；前期视频课程已于2022年6月上线



山人闲读记

微信扫描二维码，关注我的公众号

B站账号：山人一枚也



网易云课堂



腾讯课堂

学术之外，世界广阔：读闲书，拍山水，交朋友



山人闲读记

微信扫描二维码，关注我的公众号

B站账号：山人一枚也



网易云课堂



腾讯课堂

“从宽广大路出发的人常常走投无路，从羊肠小道出发的却能走到天边。”——某友人（Zhao-QuanMing）朋友圈签名

周一到周五，每天晚上 7:00-8:30， B站在线咨询、答疑；

《基于Python和PostgreSQL的并行大数据处理》系列课程

- 预备课程：《经济学和金融学中的大数据处理》（无需任何基础，免费）
- 课程之一：《Python及其在数据分析中的应用》（无需任何基础）
 - ◆ 以数据分析代码为例，讲解全套Python语法，确保学过之后，可实现“以数据处理与分析为导向”的各类项目
- 课程之二：《PostgreSQL及其Python应用》（前期基础：需已掌握Python；无需数据库基础）
 - ◆ 讲解PostgreSQL数据库的原理、服务器安装、配置、图形化管理工具；并以股票数据为例，讲解如何通过Python连接和操作PostgreSQL数据库
- 课程之三：《并行大数据处理：基于Python和PG》（前期基础：需已掌握Python和PostgreSQL）
 - ◆ 提供一套利用Python和PostgreSQL进行大规模并行数据处理和分析的方案，有面向单机器和服务器集群的两个版本

《基于Python和PostgreSQL的并行大数据处理》

系列在线课程

一月时间，拿下最热门语言和最受欢迎数据库
让机器以大数据之名全力跑起来



腾讯课堂



网易云课堂

山人一枚也

学员福利：

课件代码数据全部公开
每周五天定时在线答疑
微信群组随时自由交流

Python及其在数据分析中的应用

一顿饭钱，半月时间
拿下热门语言，畅享大数据时代



腾讯课堂



网易云课堂

山人一枚也

学员福利：

课件代码数据全部公开
每周五天定时在线答疑
微信群组随时自由交流

章节1: 第0章: 系列课程及本门课程安排

试看

课时1 视频 系列课程及本门课程安排

可试看 19:52

章节2: 第1章: Python编程环境

试看

课时2 视频 Python编程环境: Anaconda、Jupyter Notebook与Spyder

可试看 45:35

章节3: 第2章: Python语法结构概述

试看

课时3 视频 Python语法结构概述

可试看 41:30

章节4: 第3章: Python中的基础概念

试看

课时4 视频 代码、注释、保留字、对象

可试看 29:10





课时5 视频 变量

47:36

课时6 视频 基本运算符、代码规范初步

30:57

章节5: 第4章: Python基本数据类型

- | | | | |
|------|--------------------|-------------------------|---|
| 课时7 | 视频 | 整数、布尔、浮点与字符串型, 多元素类型的运算 |  40:40 |
| 课时8 | 视频 | 列表、元组与集合 |  39:55 |
| 课时9 | 视频 | 字典、文件、NoneType |  53:17 |
| 课时10 | 视频 | DataFrame |  28:39 |

章节6: 第5章: Python控制语句

- | | | | |
|------|--------------------|--------------|---|
| 课时11 | 视频 | 控制语句概述, 分支语句 |  20:46 |
| 课时12 | 视频 | 循环语句 |  48:48 |
| 课时13 | 视频 | 异常处理与环境管理语句 |  16:55 |

章节7: 第6章: Python代码封装结构

课时14	视频	Python的代码封装结构体系, 函数的定义与传参	44:55
课时15	视频	函数参数、命名空间、匿名函数、函数的调用与编写规范	39:26
课时16	视频	类的定义、类实例、运算符重载	40:18
课时17	视频	类的继承、钻石规则	37:18
课时18	视频	模块	47:20
课时19	视频	包	49:43

章节8: 第7章: Python编程原则

课时20	视频	Python编程原则	08:12
------	----	------------	-------

章节9: 扩展学习资料 试看

课时21	文本	扩展学习资料	可试看 
------	----	--------	---

PostgreSQL及其Python应用

一张电影票钱，加一周时间
掌握PostgreSQL数据库，轻松操纵大数据



腾讯课堂



网易云课堂

山人一枚也

学员福利：

课件代码数据全部公开
每周五天定时在线答疑
微信群组随时自由交流

章节1: PostgreSQL的发展历程、基本原理和整体架构

试看

课时1 视频 PostgreSQL的现状与优势，数据库的分类

可试看 27:30

课时2 视频 PostgreSQL的基本原理和整体架构

可试看 35:12

章节2: PostgreSQL服务器的安装、配置与管理

课时3 视频 Windows和MacOS下的服务器安装、删除、开启及关闭

54:07

课时4 视频 PostgreSQL服务器的网络配置，Postgresql.conf文件配置

49:01

课时5 视频 pg_hba.conf 文件配置、pgAdmin配置与管理、表空间管理

31:07

章节3: PostgreSQL的数据与索引类型，大规模编程理念

课时6 视频 PostgreSQL的数据与索引类型

14:16

课时7 视频 大规模编程中的编程理念









36:24

章节4: 通过Python操作PostgreSQL数据库

课时8 视频 通过Python创建PostgreSQL数据库连接，并操作数据库

38:04

章节5: PostgreSQL数据库操作

- | | | | |
|------|-----------------|-----------------------------|---|
| 课时9 | 视频 | SQL基础, 数据库操作 |  31:04 |
| 课时10 | 视频 | 数据表操作之一, 包括建表、删表、改名、清空等 |  17:28 |
| 课时11 | 视频 | 数据表操作之二, 包括创建、删除、重建索引, 创建约束 |  16:22 |
| 课时12 | 视频 | 数据表内操作之一, 包括列操作、简单查询、复杂查询 |  24:23 |
| 课时13 | 视频 | 数据表内操作之二, 行操作之插入操作 |  31:45 |
| 课时14 | 视频 | 数据表内操作之三, 行操作之文件插入、更新、删除 |  34:00 |
| 课时15 | 视频 | 数据表间操作之跨表查询 |  06:53 |
| 课时16 | 视频 | 导出数据到文件 |  07:42 |

章节6: 扩展学习资料

试看

- | | | |
|------|-----------------|--------|
| 课时17 | 文本 | 扩展学习资料 |
|------|-----------------|--------|

可试看

并行大数据处理：基于Python和PG

一份外卖钱，加三天时间
让机器以大数据之名全力跑起来，无论多少台



腾讯课堂



网易云课堂

山人一枚也

学员福利：

课件代码数据全部公开
每周五天定时在线答疑
微信群组随时自由交流

章节1: 分布式并行计算的原理 试看课时1 视频 分布式并行计算的原理可试看 09:53

章节2: 并行计算方案：适用于服务器集群

课时2 视频 服务器集群版并行方案（第一部分）26:50课时3 视频 服务器集群版并行方案（第二部分）43:42

章节3: 并行计算方案：适用于单台服务器

课时4 视频 单机器版并行方案_110:40课时5 视频 单机器版并行方案_244:53章节4: 扩展学习资料 试看课时6 文本 扩展学习资料可试看



THANKS

Architect