



中国科学技术大学
University of Science and Technology of China

程序语言设计与程序分析

课程简介

张昱

yuzhang@ustc.edu.cn

中国科学技术大学
计算机科学与技术学院



计算机科学的理论

□ 数理逻辑

- 一阶逻辑、高阶逻辑、公理集合论、递归论、模型论和证明论等
——也是现代数学的基础

□ 计算理论

- 可计算性和计算复杂性、算法、形式语言理论、自动机理论等

□ 程序理论

- 形式语义、类型论、 λ 演算、程序验证等等

□ 数值计算



关于程序设计语言的研究

□ 语法

形式语言和自动机理论、语法分析的实现技术

□ 语义

公理语义、**操作语义**、指称语义、属性文法

形式描述技术还有：代数规范、范畴论

□ **程序设计的范型**

命令式语言、函数式语言、逻辑程序设计语言、

面向对象程序设计语言、**并行程序设计语言**、.....



关于程序设计语言的研究

□ 类型论与类型系统

多态类型、子类型、存在类型、依赖类型等

□ 程序验证

程序正确性证明

□ 程序分析技术

■ 数据流分析、控制流分析、模型检查、抽象解释

□ 程序的自动生成技术

■ 程序变换



本课程的主要内容

□ 学习掌握和程序设计语言有关的理论和技术

■ 形式语义（**操作语义**、.....）

类型论与类型系统

以数学为工具，利用符号和公式，精确地定义和解释计算机程序设计语言的语义，使语义形式化的学科.

■ 与程序行为和**程序分析**有关的推理技术

■ 一些语言及特征的案例研究

目标：寻找精确、抽象地描述程序行为的方法

■ 精确(precise): 用数学工具形式化和证明感兴趣的性质

■ 抽象(abstract): 清晰地讨论性质，又不陷入底层细节



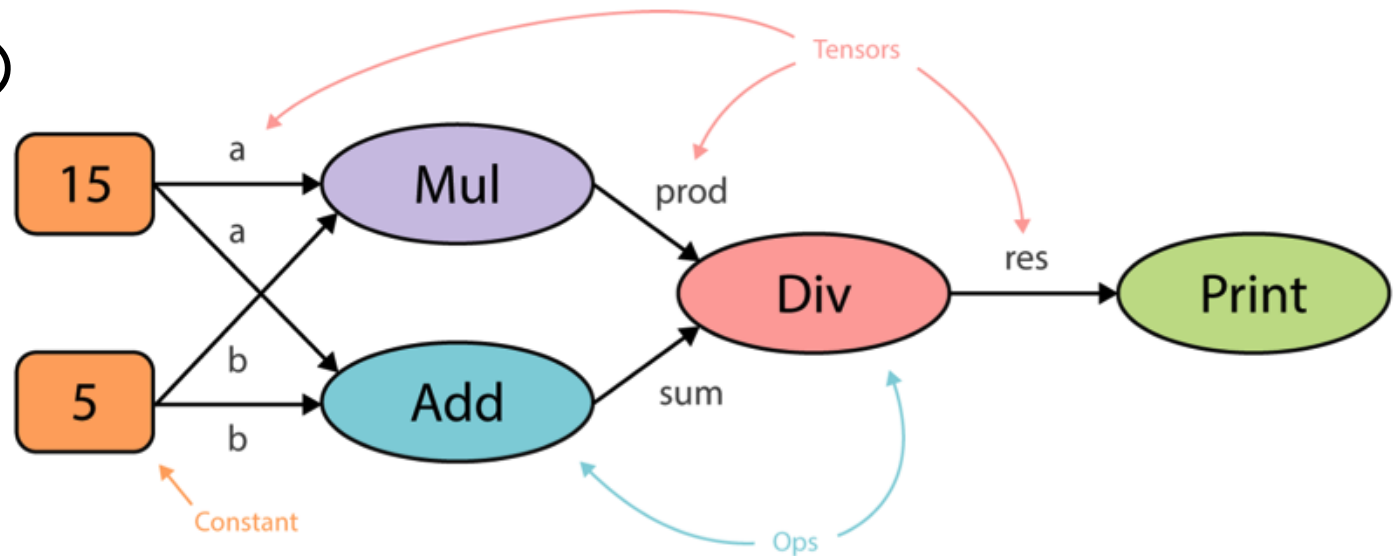
为什么学习本课程

□ 软件系统：规模日趋庞大+多语言混合+异构

例：深度学习(DL)框架(数据流图+张量)：

Python + C++ ([PyTorch](#), [Tensorflow](#), [Mindspore](#),...)

⇒ Julia ([Flux](#))? (性能)



下一代？

[Jax](#): [grad](#)(自动微分)、[jit](#)(即时编译)、[vmap](#)(向量化)、[pmap](#)(SPMD)

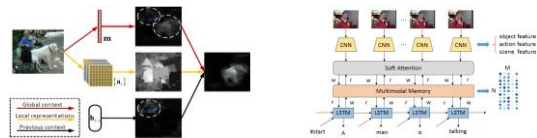
[ModularAI](#): Mojo&MAX、[重构AI](#)、[AIServing](#) 让任何人随处可用AI

深度学习软件栈

应用层



模型层



框架层



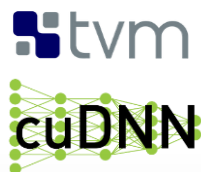
高层
中间表示

NNVM
Relay



编译

算子
实现层



优化

硬件层



Nvidia Drive
PX2



Nvidia Jetson
Nano



地平线“征途2.0”

□ TF、PyTorch、TVM等
谷歌: TF=>XLA=>JAX

□ 国内开源DL平台

- 百度 PaddlePaddle
- 华为 MindSpore
- 阿里 OneFlow
- 腾讯 TNN (推理)
- 腾讯 TNN (推理)

提升性能、适配更多芯片

=> 程序分析与优化

=> (中间)语言设计



为什么学习本课程

- 软件系统：规模日趋庞大+多语言混合+异构
- 软件无处不在 => 可信软件？
例：兼具性能优势的可信虚拟机



软件可靠性和安全性问题

□ 阿丽亚娜5号运载火箭1996年6月4日首次发射失败

- 发射点火后40秒内引爆自毁
- 控制火箭飞行的**软件故障**：数值溢出
64位浮点数变量被转换成16位整数
- 损失：1987-1996 研制费用高达80亿美元



□ 2003年8月14日美加大面积停电事故

- 电控系统软件的**并发错误**
- 损失：累计损失负荷6180万千瓦，
5000万人用电受影响，美国估计
每天损失300亿美元，加拿大安大略省损失50亿加元





软件可靠性和安全性问题

中国：2007年建立国家信息安全漏洞库

信息安全漏洞周报

(2021 年第 10 期 总第 564 期)

信息安全测评中心

2020 年 3 月 7 日

根据国家信息安全漏洞库 (CNNVD) 统计, 本周 (2021 年 3 月 1 日至 2021 年 3 月 7 日) 安全漏洞情况如下:

公开漏洞情况

本周 CNNVD 采集安全漏洞 367 个, 与上周 (243 个) 相比增加了 51.03%。

接报漏洞情况

本周 CNNVD 接报漏洞 3974 个, 其中信息技术产品漏洞 (通用型漏洞) 100 个, 网络信息系统漏洞 (事件型漏洞) 3874 个。

重大漏洞预警

Microsoft Exchange 多个安全漏洞: 包括 Microsoft Exchange 安全漏洞 (CNNVD-202103-192)、Microsoft Exchange 安全漏洞 (CNNVD-202103-191) 等。成功利用漏洞的攻击者可以在目标系统中获取敏感信息、写入任意文件、执行恶意代码等。Microsoft Exchange 2010、2013、2016、Microsoft Exchange 2019 等版本均受漏洞影响。目前, 微软官方已发布补丁修复了上述漏洞, 建议用户及时确认是否受到漏洞影响, 尽快采取修补措施。

黑客可利用漏洞攻击 D-Link VPN 路由器

发布时间 2021-01-26 来源 hackernews

香料公司Symrise感染Clop

香精香料公司Symrise感染勒索软件Clop, 500GB未加密的文件被盗, 近1000台设备被加密。Symrise是全球超过3万种产品 (包括雀巢和可口可乐) 中使用的香精香料的主要开发商, 于上周遭到了网络攻击, 并关闭了所有必要的系统以防止攻击蔓延。Clop团伙声称其通过钓鱼攻击入侵了Symrise的网络并窃取了500 GB的未加密文件, 其在数据泄露网站上发布的截图显示被盗数据包括护照、会计凭证、审计报告、化妆品成分和电子邮件等。

发布时间 2020-12-30 来源 启明星辰



兴盛的自动驾驶及隐患

2月召回62.12万辆 特斯拉召回数量超去年

2021.3.1

美国要求特斯拉召回15.8万辆车 路透华盛顿

2021.1.13 因媒体控制单元(MCU)故障可能导致触屏无效而构成安全风险

特斯拉“自动驾驶技术”屡发事故遭质疑

2021.1.7



2020年7月29日，北京南四环特斯拉Model S 自动驾驶交通事故，速度80km/h，前方的道路养护货车正在作业，特斯拉发现前方车辆后车辆自动向右并线避让，然而在避让到另外一个车道时，特斯拉却突然向左加速撞到了货车右轮处

任泽平：中国自动驾驶发展报告2020（上）
2021.1.15

2017年12月，工信部发布的《国家车联网产业标准体系建设指南（智能网联汽车）》

安徽省于2020年提出了《2020年安徽省汽车和新能源汽车发展工作要点》

2020年9月3日全长4.4公里的安徽首条自动驾驶汽车5G示范线在包河区正式开通

42辆东风自动驾驶出租车“上路”运行，武汉市民可免费预约试乘 2021.2.28

智能驾驶新一轮芯片争夺战开启 2021.2.26

全球自动驾驶芯片龙头Mobileye分别在2008~2020年发布EyeQ1~EyeQ5五代芯片，其**计算能力**从2008年的4.4GOPS（十亿次）提升到24000GOPS，**超五千多倍**。

软硬件协同的性能优化 + 安全可靠



为什么学习本课程

□ 软件系统：规模日趋庞大+**多语言混合**+**异构**

□ 软件无处不在 => **可信**软件？

例：兼具性能优势的可信虚拟机

□ 软件无处不在 => **快速**运维？

例：基础架构即代码（IaC, infrastructure as code）

<https://octoverse.github.com/> 2022 HCL增长最快

Hashicorp Configuration Language ([HCL](https://github.com/hashicorp/hcl)) +Shell+Go

<https://github.com/LarryDpk/terraform-101>

2023：AI-powered developer platform TypeScript



十四五数字经济发展规划

□ 2022.1发布

□ “东数西算”工程

“数”指数据，“算”是算力，即对数据的处理能力，“东数西算”是通过构建数据中心、云计算、大数据一体化的新型算力网络体系，将东部算力需求有序引导到西部，优化数据中心建设布局，促进东西部协同联动。



中华人民共和国中央人民政府

www.gov.cn

http://www.gov.cn/zhengce/content/2022-01/12/content_5667817.htm



<https://www.ndrc.gov.cn/xwdt/ztzl/dsxs/?code=&state=123>



“东数西算”工程



全国设立8个算力枢纽节点，10个国家数据中心集群——

京津冀枢纽：张家口集群

长三角枢纽：长三角生态绿色一体化发展示范区集群、芜湖集群

粤港澳枢纽：韶关集群

成渝枢纽：天府集群、重庆集群

贵州枢纽：贵安集群

内蒙古枢纽：和林格尔集群

甘肃枢纽：庆阳集群

宁夏枢纽：中卫集群



“东数西算”工程取得积极进展

数据来源：国家数据局

截至2024年3月底

10个国家数据中心集群算力总规模超 **146万** 标准机架

整体上架率 **62.72%**

较2022年提升**4个百分点**



已有**226**个各级地方建立了公共数据开放平台，开放的有效数据集超过了**34**万个

- **西部**：处理后台加工、离线分析、存储备份等对网络要求不高的业务
- **东部**：处理工业互联网、金融证券、灾害预警、远程医疗、视频通话、人工智能推理等对网络要求较高的业务

- **加快实现云网协同**
- **降低网络、电力成本**
- **提升资源使用效率**
- **降低上云用数成本**



“东数西算” 工程

- (云核心网)网络处理
- 异构的云边端计算

性能极致
低功耗
低时延
高可靠

- API 编程/云原生架构
- 人工智能编程框架等

易适配 异构算力
易调度 云边端异构核
易编程 无/低代码编程

- 加快实现云网协同
- 降低网络、电力成本
- 提升资源使用效率
- 降低上云用数成本

云原生技术栈



- **西部：**处理后台加工、离线分析、存储备份等对网络要求不高的业务
- **东部：**处理工业互联网、金融证券、灾害预警、远超医疗、视频通话、人工智能推理等对网络要求较高的业务



AI/云原生 编程语言



□ Python: 1991发布

- 多范型：函数式、指令式、反射式、结构化、面向对象
- 解释执行

□ UCBerkeley CS61A

- <https://cs61a.org/>
Structure and Interpretation of Computer Programs

[Python学习路线图](#)

[2020 Python 开发者调查结果](#)





AI/云原生 编程语言

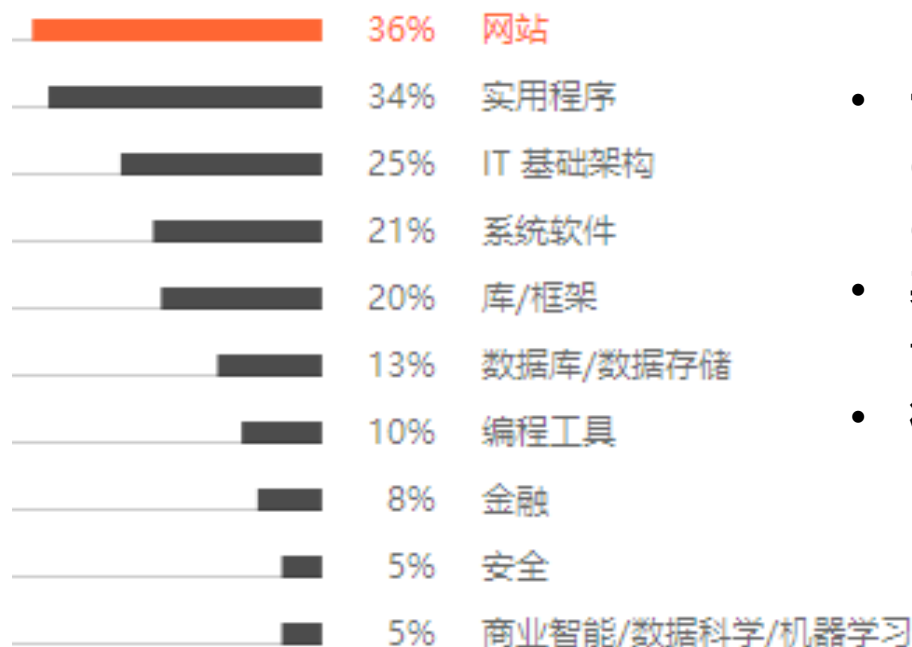
□ Go: 2007设计、2009发布

- 静态强类型,并发型,垃圾收集
- 110 万专业的开发者 选择 Go 作为其主要开发语言
亚洲最多 57万



[\[2021年2月\]Go语言现状](#)

[Go 开发者路线图](#)



- **云计算基础设施**领域代表项目：
docker、kubernetes、etcd、consul、cloudflare CDN、七牛云存储等
- **基础软件**代表项目：
tidb、influxdb、cockroachdb 等
- **微服务**代表项目：go-kit、micro、monzo bank 的 typhon、bilibili 等
 - **互联网基础设施**代表项目：以太坊、hyperledger 等



为什么学习本课程

- 软件系统：规模日趋庞大+多语言混合+异构
- 软件无处不在 => 可信软件？
- 软件无处不在 => 快速运维？
- 异构硬件和新兴应用 => 期待更多语言解决方案

- 降低编程难度、发挥异构硬件优势

芯片(2020.8国家芯片五年计划；2025:70%芯片自给)

十四五数字经济发展规划、东数西算系统工程

端到端DL编译器、Go等的运行时、

量子计算编程系统、图数据及图查询语言



量子计算

全球量子计算布局详解 2020.8

美、中、英、德、俄 国家量子计划

重磅！国务院政府工作报告首次提及量子信息

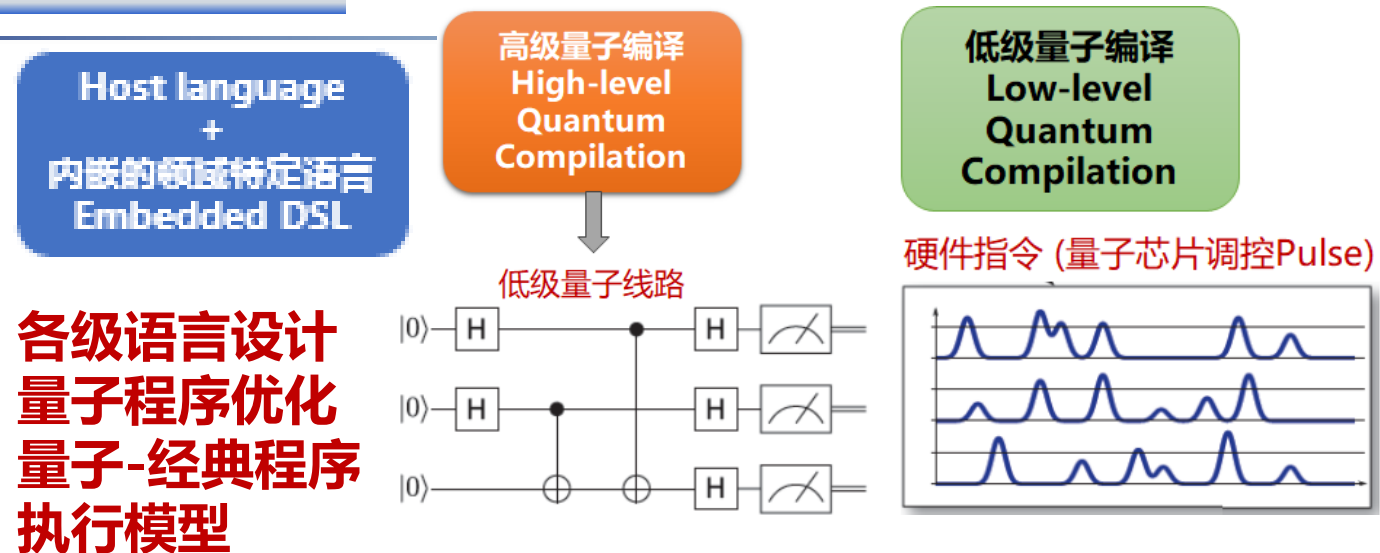
2021年3月5日，国务院总理李克强在政府工作报告中介绍：

...创新型国家建设成果丰硕，在载人航天、探月工程、深海工程、超级计算、量子信息等领域取得一批重大科技成果。...

这是“量子信息”首次出现在国务院政府工作报告中，此前，只有2016年和2018年的政府工作报告两次提及“量子通信”，但是量子信息不仅包括量子通信，还包括量子计算、量子精密测量等，这意味着中国将从优先发展量子通信转变为量子信息科学的全面发展。

法国启动量子技术国家战略 2021年1月

计划5年内在量子领域投资18亿欧元，其中4.3亿欧元用于未来成熟量子计算机研究



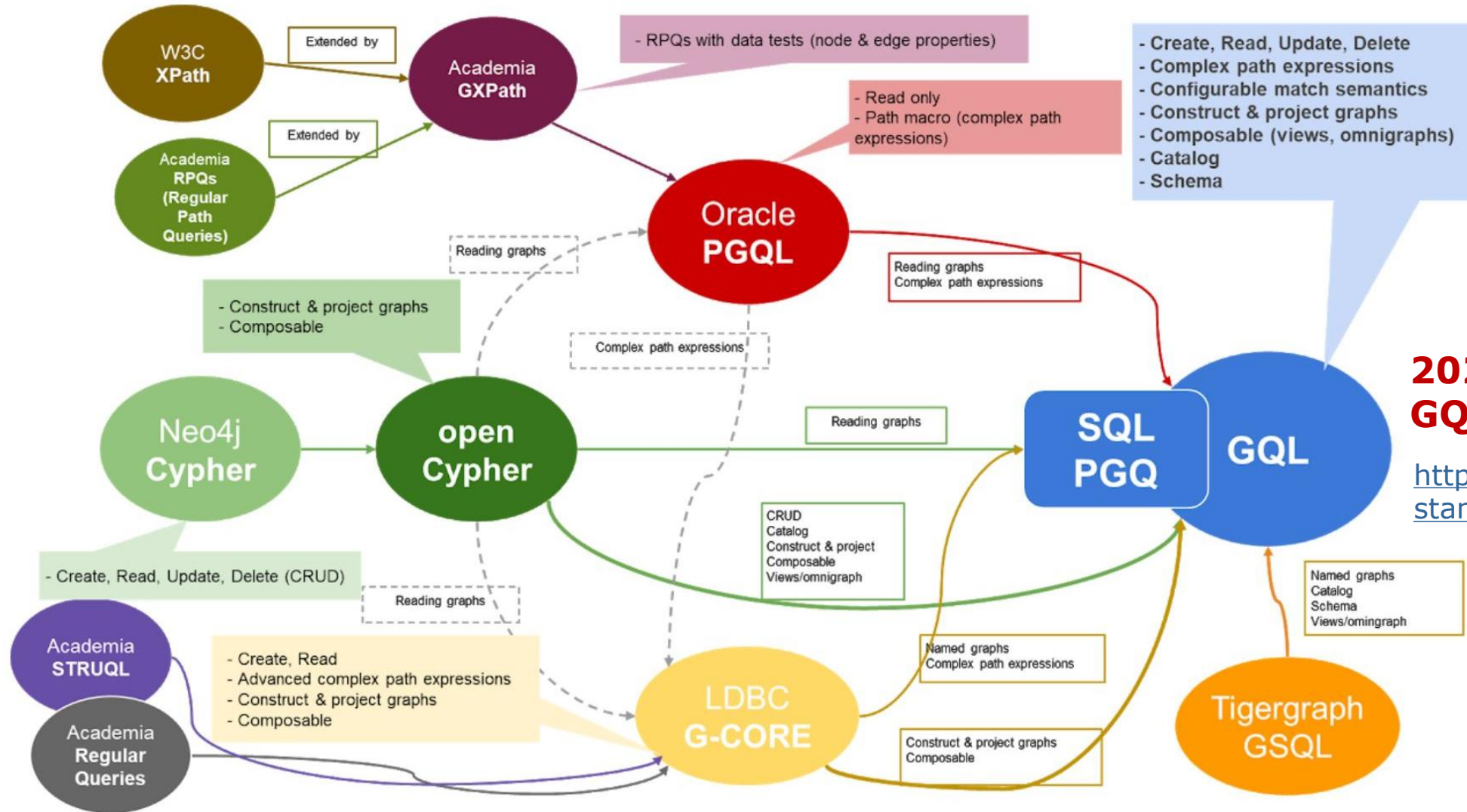
2024.4 中国电信19亿元控股国盾量子 加速“掘金”量子科技

2024.5 中电信量子集团发布“天衍”量子计算云平台

2024.9下达的安徽省科技创新攻坚计划项目

多体系量子算力与超智算力融合的量子计算云平台研发与应用示范	本源量子计算科技(合肥)股份有限公司	4056.2 (811)
低温芯片型离子阱量子计算系统	合肥么正量子科技有限公司	3845 (750)

图查询语言



2024.4.17
GQL标准发布

<https://www.gqlstandards.org/>

图查询语言的历史回顾短文



未来的重要发展方向

基于高层抽象的编程系统 (程序语言+编译运行时系统)

高性能计算+大数据处理+AI



课程主要目标

- 介绍现有的语言特征
 - 含义、作用、对比

- 定义程序行为和语义的方法
 - 操作语义+类型论

- 分析和推理程序性质的方法
 - 程序分析和验证



PL Concepts

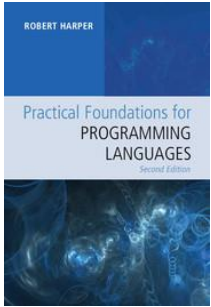
- Programming models (编程模型)
 - A precise, composable specification of things
- Programming paradigms (编程范型)
 - Common properties of models
- Programming languages
 - Syntax for expressing a programming model
- Programs
 - Instance of a model (abstractly) or a language (concretely)



PL Concepts to be Discussed

- ☐ **Scopes**
- ☐ **Object System**
- ☐ **Embedded languages**
- ☐ **Lambda calculus and Type System**
- ☐ **Control flow (coroutine, exception)**
- ☐ **Types and Propositions, Logic Prog., Hoare Logic**
- ☐ **Memory management, Memory Model**
- Concurrency**

Lua



- Practical Foundations for Programming Languages (2nd Edition) by *Robert Harper*. ([pdf](#))
Cambridge University Press, 2016.

2022.4 出版



- Lecture notes on Static Program Analysis
by Anders Møller and Michael I. Schwartzbach,
- Principles of Program Analysis (2nd Edition) by *Flemming Nielson et al.* Springer, 2005.
- 程序设计语言理论(第二版), 陈意云、张昱,
高等教育出版社, 2010.

张昱: 《程序语言设计与程序分析》课程简介