



2023 CCF CHINASOFT
中国软件大会

开源软件对系统实践 能力培养的重要性

王肇国

上海交通大学



计算机课程体系的三个阶段

百家争鸣阶段

大量引进西方教材和教学体系
老师什么都教，学生什么都不会



独尊儒术阶段

学习国外先进经验，着手改革教学内容
理论基础深、实践强度大



互联网、人工智能时代

课程融合
强调分布式、人工智能



我国的体系脱胎于第一个阶段
目前的计算机专业本科教学体系也是在此时构建



部分985高校按第二阶段方案进行改革



如何将**开源软件**融入到
教学课程以及科研训练？

目 录

01

本科生培养

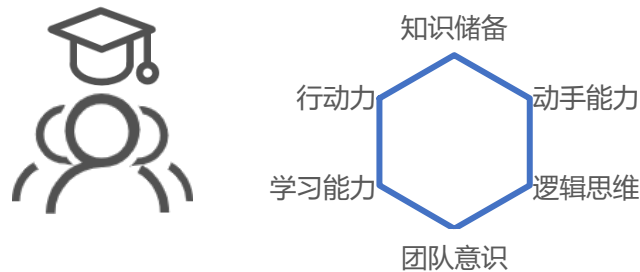
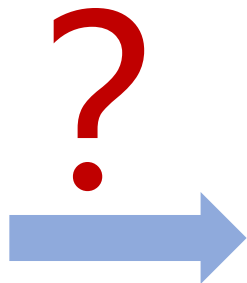
02

研究生培养

本科生培养目标

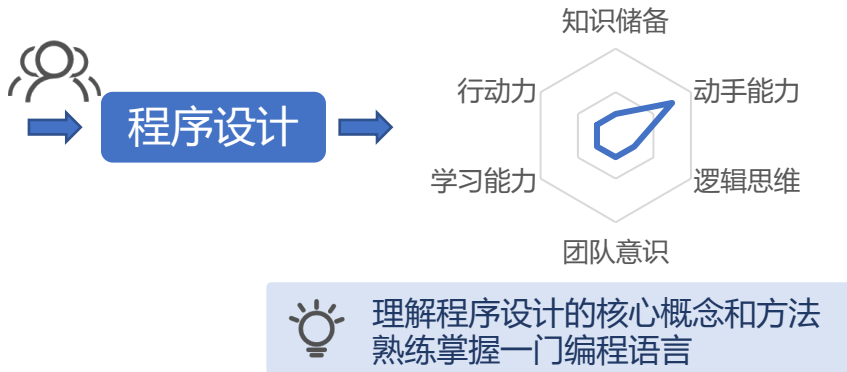


本科新生



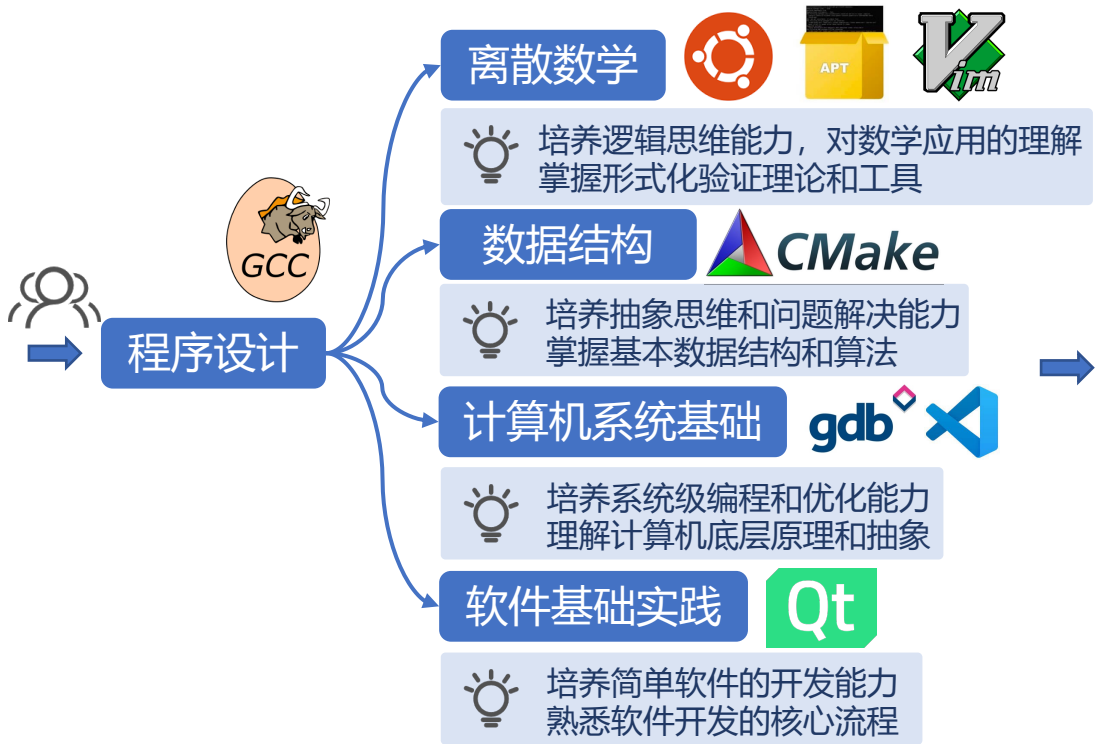
本科毕业生

具有扎实的**基础理论**
具有较强的工程**实践能力**
能够解决相关技术领域**复杂系统工程问题**



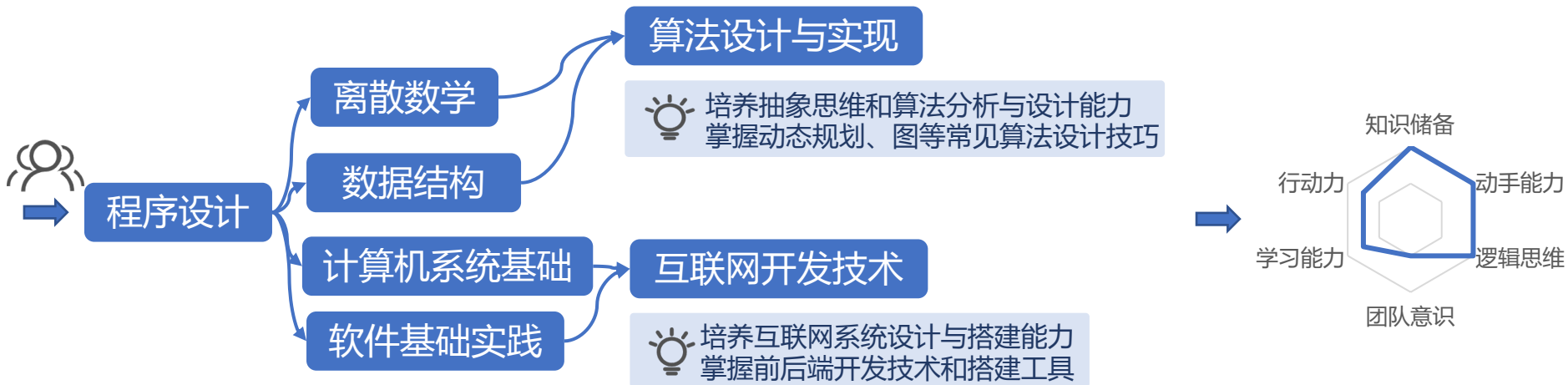




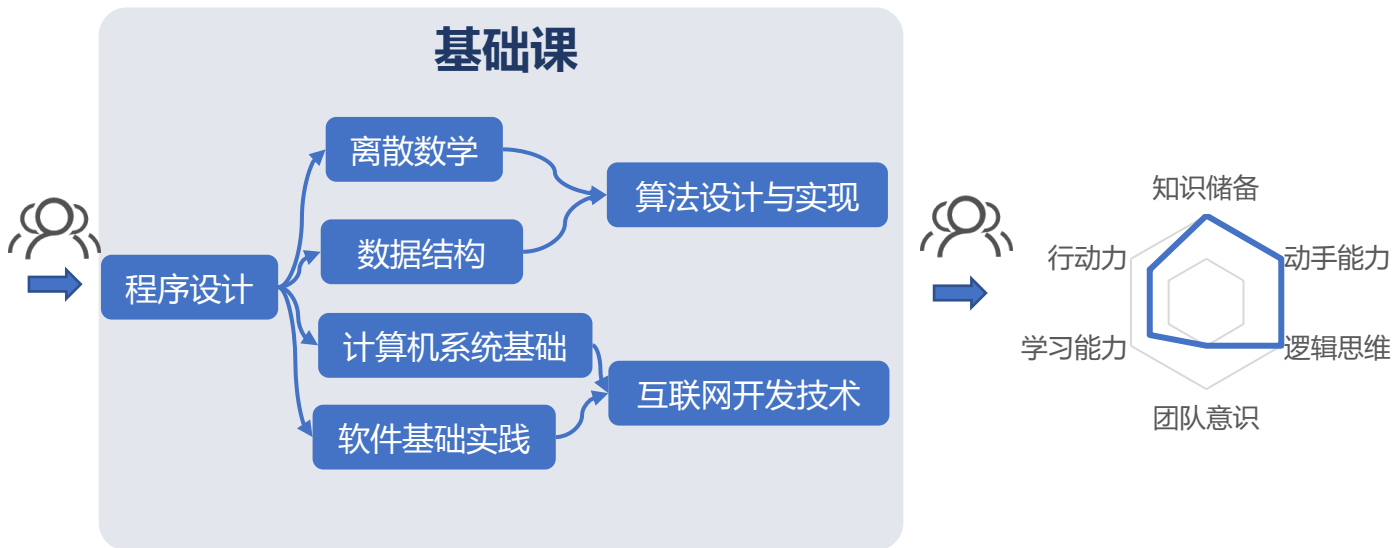


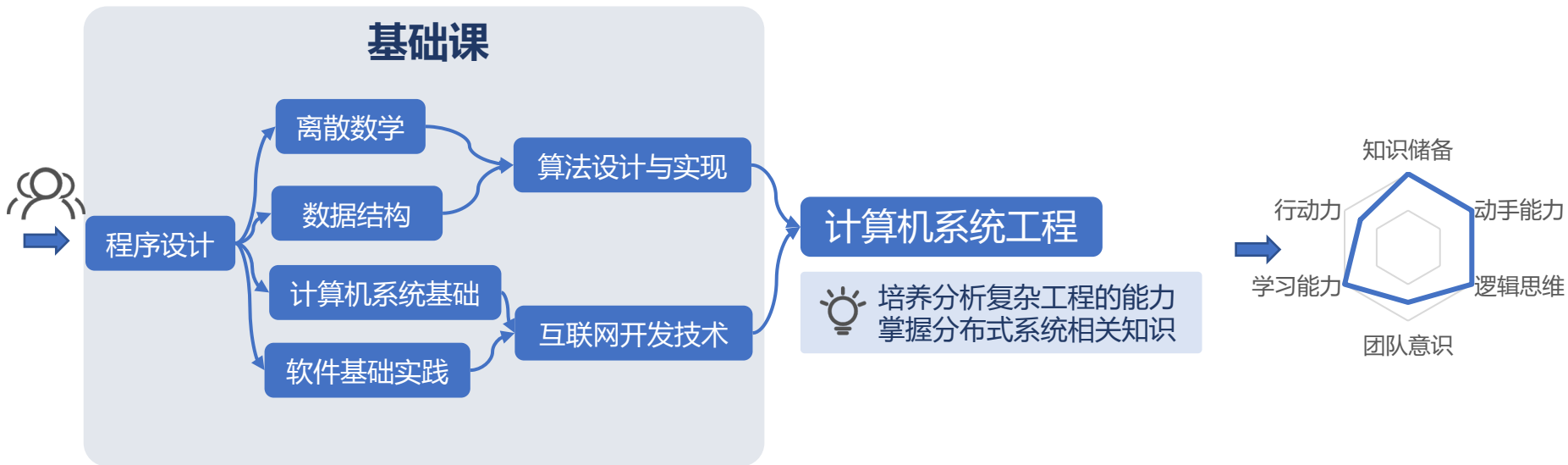
阶段一：使用**开源软件**辅助编程开发



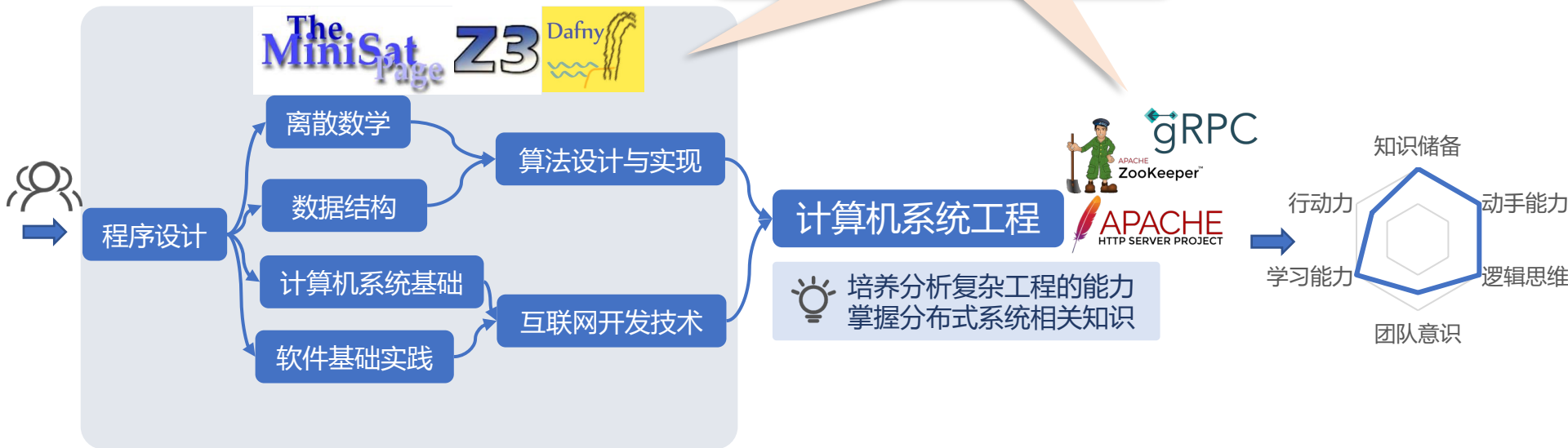


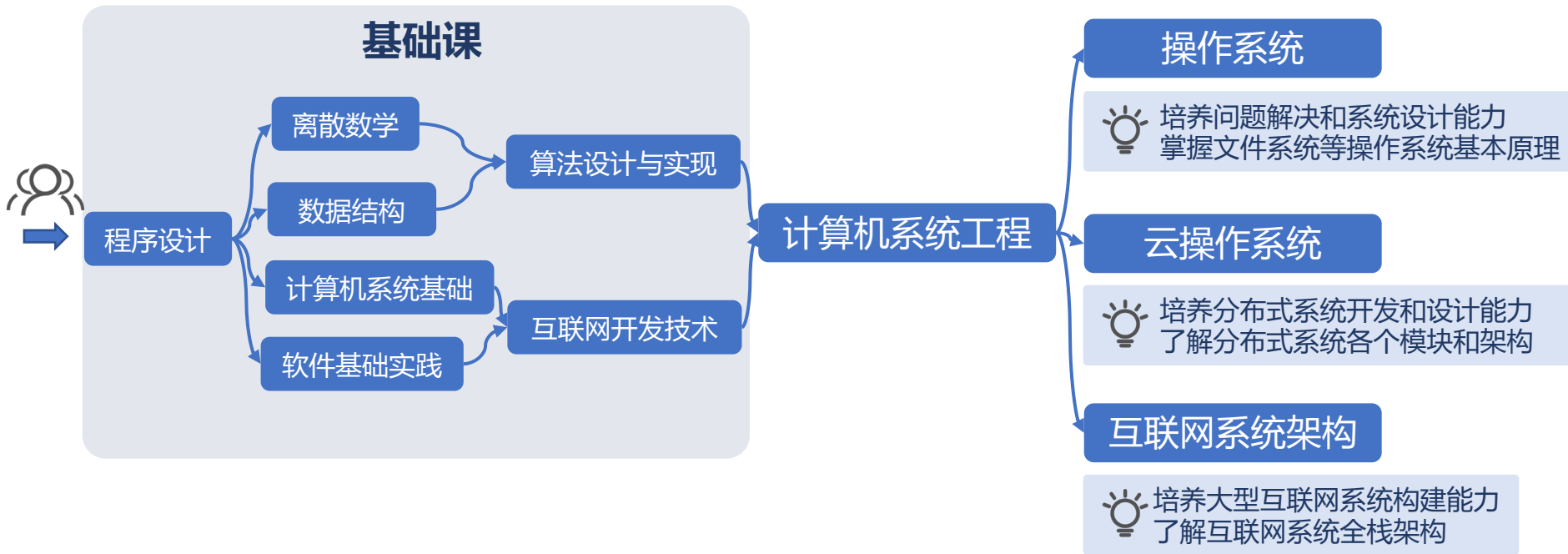




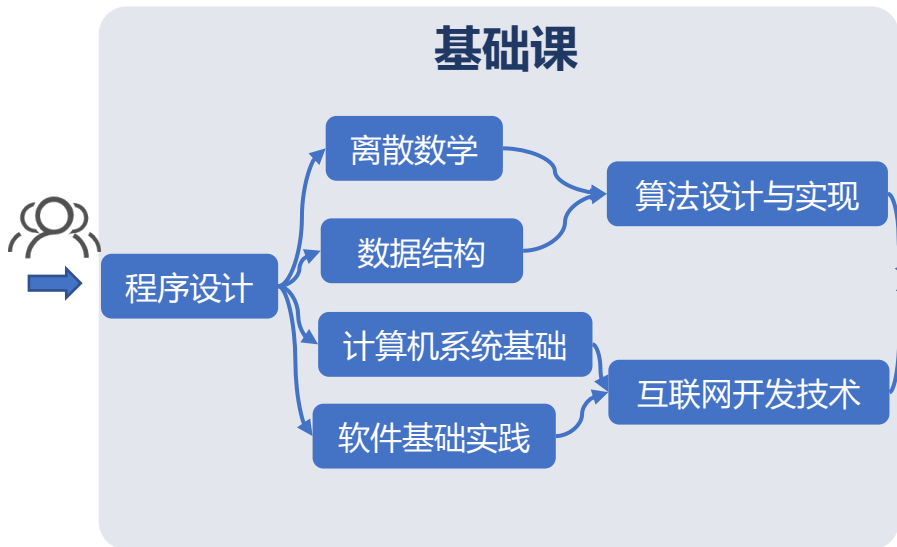


阶段三: 了解**开源软件**底层原理





阶段四: 基于**开源系统软件**
设计并构建复杂系统



操作系统



培养问题解决和系统设计能力
掌握文件系统等操作系统基本原理

云操作系统

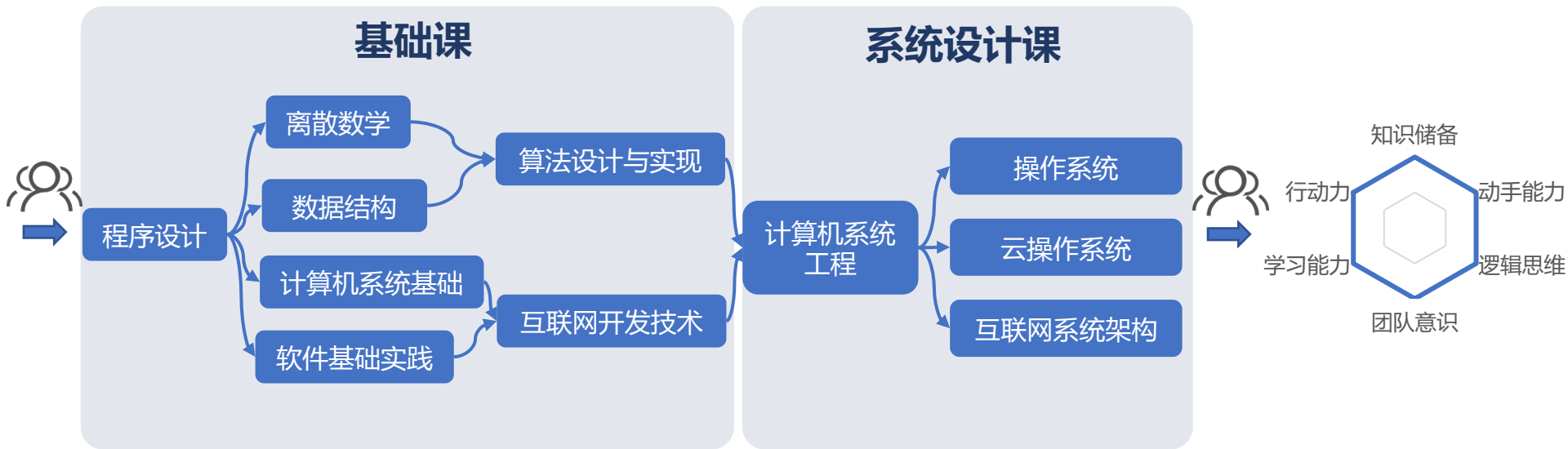


培养分布式系统开发和设计能力
了解分布式系统各个模块和架构

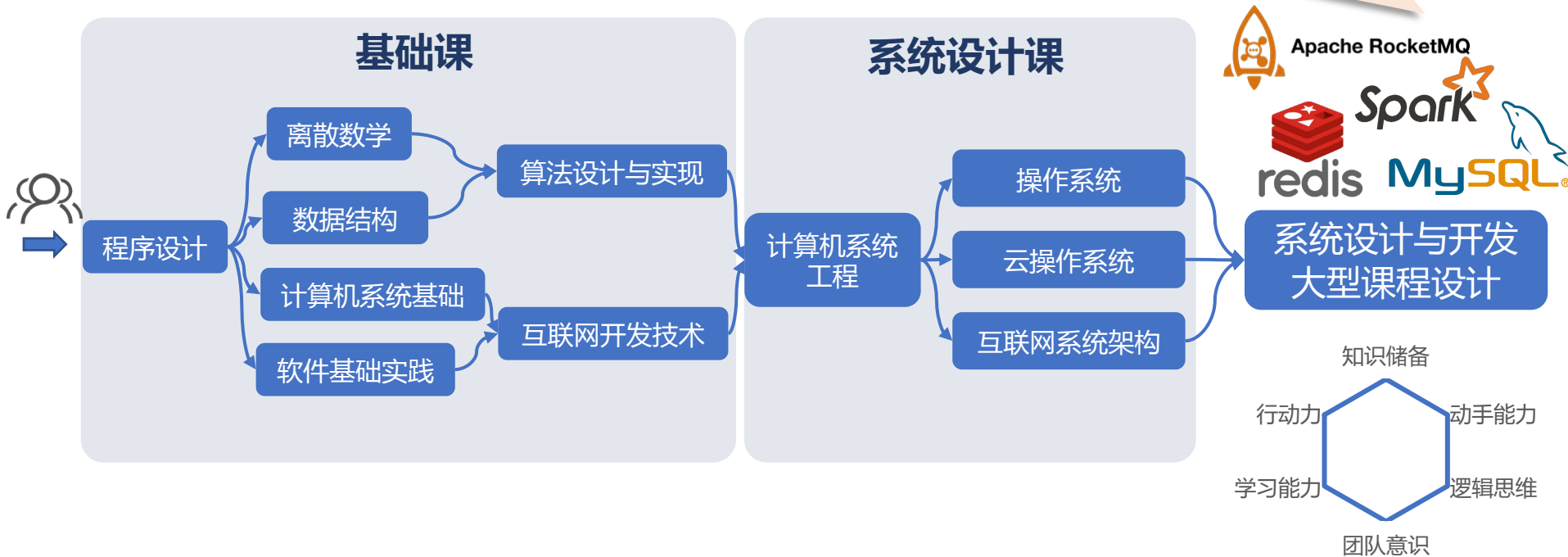
互联网系统架构

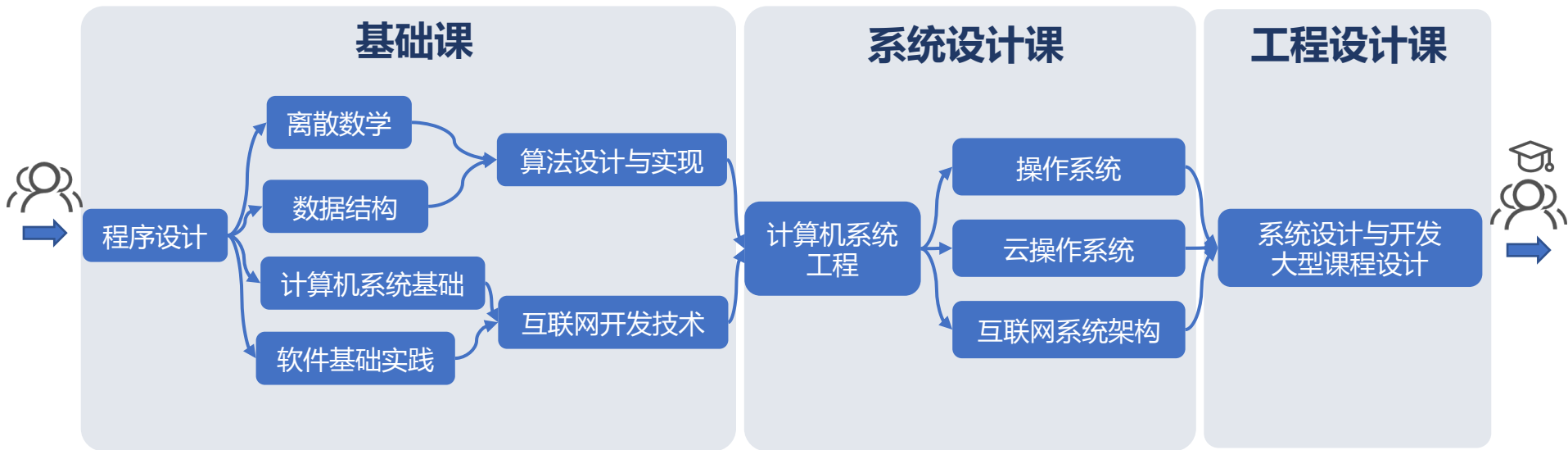


培养大型互联网系统构建能力
了解互联网系统全栈架构



阶段五: 深度优化**开源系统软件**,
提升复杂系统性能和可用性






基础课



```
bool is_power2(unsigned int v)
{
    unsigned int i;
    for(i = 0; i <= 31; ++i)
    {
        if(v == (1 << i))
            return true;
    }
    return false;
}
```



一个含有Bug的简单程序

这个程序正确吗？



回答 “如何验证程序的正确性？”

第一步：将程序转换为计算机可验证的语言

理论教学



命题逻辑
谓词逻辑
霍尔逻辑



开源软件

The
MiniSat
Page

SAT求解器

Z3

SMT求解器

回答 “如何验证程序的正确性？”

第一步：将程序转换为计算机可验证的语言

第二步：自动证明

理论教学



命题逻辑
谓词逻辑
霍尔逻辑



程序验证
符号执行



程序正确性

开源软件

The MiniSat Page

SAT求解器

Z3

SMT求解器



程序验证器

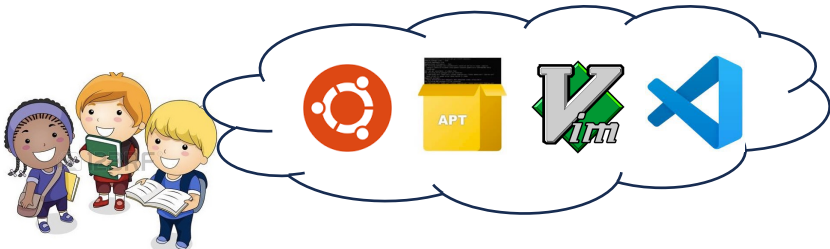


符号执行引擎

离散数学：与开源软件结合

阶段一

通过开源软件了解应用场景

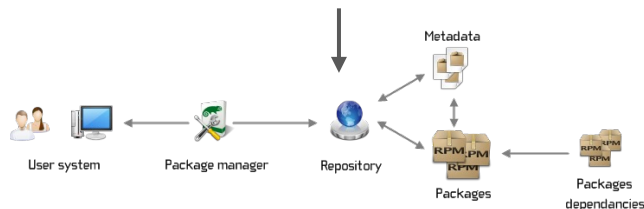


首次接触Linux系统

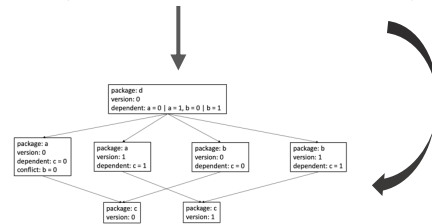
The following information may help to resolve the situation:

```
The following packages have unmet dependencies:  
libc6-dev : Depends: libc6 (= 2.35-0ubuntu3.4) but 2.35-0ubuntu3 is to be installed  
E: Unable to correct problems, you have held broken packages.
```

问题：无法安装软件包



原因：包管理器发现软件包之间存在依赖冲突



开源代码

SAT问题（离散数学的应用）

阶段一

通过**开源软件**了解应用场景

阶段二

使用**开源软件**求解简单的逻辑命题



具备使用**开源软件**解决
简单问题的能力

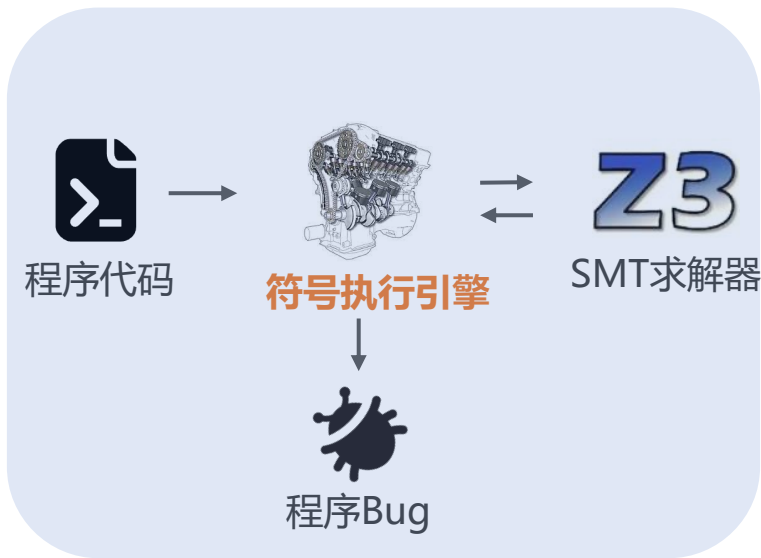


课程实验: 基于**开源SAT求解器**求解简单逻辑问题

- 阶段一
通过**开源软件**了解应用场景
- 阶段二
使用**开源软件**求解简单的逻辑命题
- 阶段三
熟悉**开源软件**底层原理



对**开源软件**原理产生好奇



Lab2: 基于**开源SMT求解器**
实现符号执行引擎

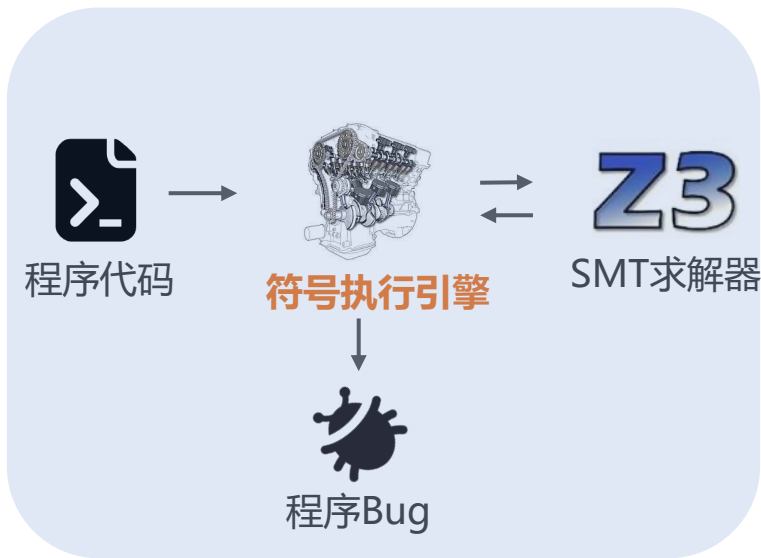
离散数学：与开源软件结合

- 阶段一
通过开源软件了解应用场景
- 阶段二
使用开源软件求解简单的逻辑命题
- 阶段三
熟悉开源软件底层原理

阶段四
使 **验证了程序的正确性**



对开源软件系统和原理有较深理解



Lab2: 基于开源SMT求解器实现符号执行引擎



Lab3: 基于开源程序验证器验证程序正确性

离散数学：与开源软件结合

2023 CCF CHINASOFT
中国软件大会



阶段一

通过开源软件了解应用场景

阶段二

使用开源软件求解简单的逻辑命题

阶段三

熟悉开源软件底层原理

阶段四

使用开源软件构建模块，搭建复杂系统

阶段五

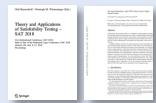
优化开源系统



希望更加深入理解
开源软件系统和原理



SAT求解器
开源代码



前沿论文



SAT Competition

修复现有代码Bug
优化求解算法性能

Bonus Lab: 优化开源SAT求解器的性能

Submissions

Submission ID	Submit Time	Status	Message	Score	Solved	Score SAT	Solved SAT
41	2022/10/27 21:09:28	Accepted	Accepted	0.018510384	4	0.0014977237	1
109	2022/11/7 17:57:37	Running					
28	2022/10/26 14:59:54	Accepted	Accepted	0.012236633	4	0.0009900981	1
47	2022/11/1 20:00:29	Accepted	Accepted	0.018510226	4	0.0014977109	1
42	2022/10/29 16:50:52	Accepted	Accepted	0.012442511	3	0.0014977151	1

学生打榜

目 录

01

本科生培养

02

研究生培养

研究生培养目标



具有扎实的**基础理论**
具有较强的**工程实践能力**
能够解决相关技术领域**复杂系统工程问题**

本科生



培养能够解决**复杂工程问题**的工程师
具备**开发、调试、优化真实系统**的能力

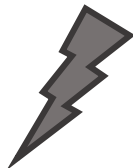
研究生

挑
战

学校内缺乏工业界**开源系统**的实践案例
学生缺乏开发调试**开源系统**的实践经验



学生对工业界**开源系统**的了解不深
缺少接触工业界**开源系统**的机会



希望学生了解工业界**开源系统**
希望学生参与工业界**开源系统**开发



学校



企业

解决
方案

上海交通大学



华为



将**开源软件**的原理知识融入课程
在**课程项目**中使用**开源软件**

校企

联合培养



丰富的**工业界开源系统**
完善的**开源系统支持**



学校



企业

- 离散的培养体系
 - 五门合作课侧重点不同
- 校企配合授课
 - 校企围绕一个关键技术设计教学内容
 - 学校铺垫基础知识
 - 华为讲解最新技术
 - 校：企 \approx 4：1



五门课程合作	用于教学的开源系统软件
计算机系统设计与实现	• openHarmony
并行与分布式数据库	• openGauss
高级操作系统	• openEuler
高级云操作系统	• openEuler
暑期实践类课程 (分组完成)	• openEuler 开源创新实践 • openGauss AI特性课程实验 • MindSpore 目标检测算法实战 • 仓颉编程语言课程实践 • 毕昇编译器

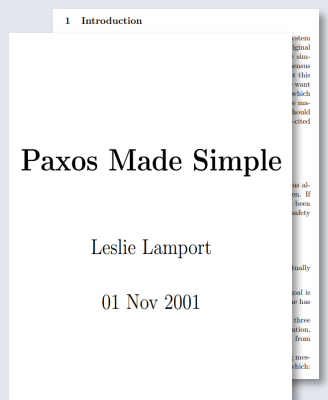
目标

- 在掌握数据库系统**核心原理**的基础上
- 通过对**典型案例**的分析
- 指导学习**现代数据库系统**在**设计和实现**中的**关键问题**

课程内容

- 围绕典型并行与分布式数据库系统和相关**前沿论文**与**工业界开源系统**展开
- 运用分析、比较和讨论等形式
- 深入学习数据库系统的**可靠性**、**可伸缩性**、**系统容错**和**安全**等方面

前沿论文



开源系统



共建课程：并行与分布式数据库系统

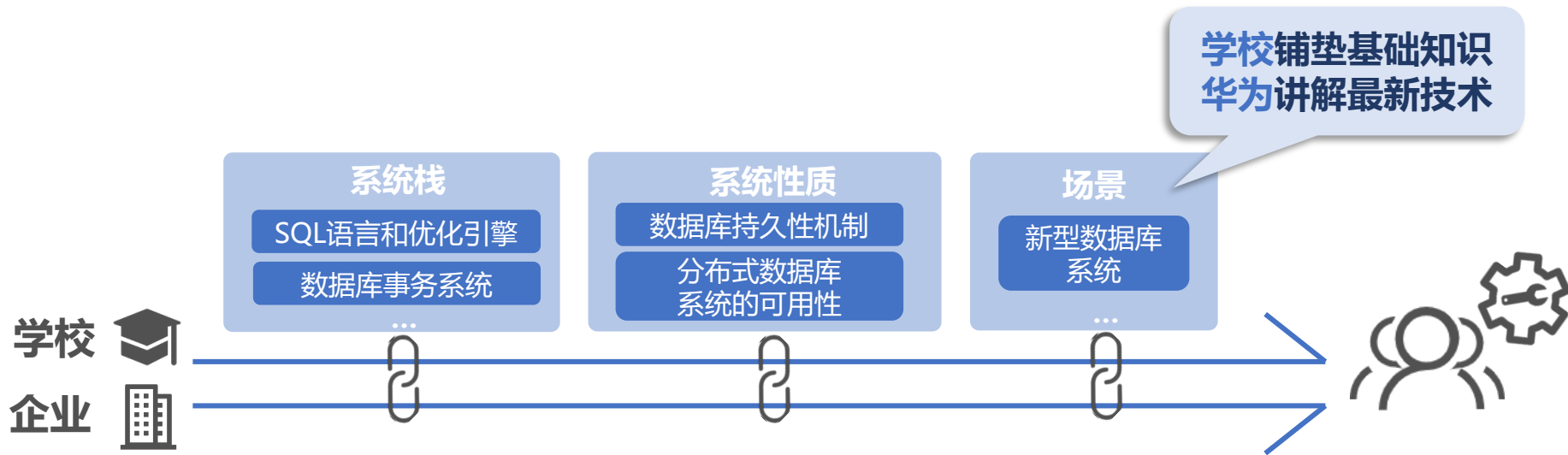
2023 CCF CHINASOFT
中国软件大会



	章节	主要内容	课时
概览	数据库系统设计原则	探讨数据库设计中的典型问题、解决方案和设计原则	3
系统栈	SQL语言和优化引擎	介绍数据库中SQL语言的核心理论以及SQL优化引擎基本原理和设计原则	3
	数据库事务系统	介绍一致性在多核、分布式系统中的作用；探讨不同场景下不同的一致性需求；分析强一致性的经典并发控制	3
	数据库数据索引结构	探讨高性能、高可扩展的数据索引结构的设计原则以及实现方法	6
系统性质	数据库持久性机制	介绍持久性在系统崩溃时的作用；探讨如何使用日志保证持久性；案例分析：MySQL数据库和PostgreSQL数据库的日志系统	6
	分布式数据库系统的可用性	介绍可用性在分布式系统中的作用；探讨如何使用数据备份保证高可用性；分析不同数据备份方式的优劣	9
场景	新型数据库系统	探讨前沿数据库系统（键值存储、云原生）在设计实现中遇到的问题、机遇与挑战；分析不同种类的存储系统在设计实现上的权衡	6

共建课程：并行与分布式数据库系统

2023 CCF CHINASOFT 中国软件大会



课程项目：结合工业界**开源系统软件**实现**应用与优化**

开放型课题

- ① 当前**大语言模型 (GPT)** 研究火热，将**GPT** 应用于数据库有什么想法？并尝试将其应用在 **openGauss** 上，给出原型演示。
- ② 调研业界商用数据库**性能Profiling**方法，并设计**openGauss**的细粒度全链路的耗时分布，以便于性能分析、定位以及优化
- ③ 从CPU利用率、IO利用率以及鲲鹏软硬协同方面，寻找**openGauss性能优化**点，给出分析结论、优化思路以及优化原型。

具象型课题

- ① **openGauss** 支持行存、支持列存。设计一种**数据生命周期管理方法**，将行存冷数据按照设定周期，转为列存并压缩。同时，如果数据访问热度变热以后，将其转为行存。
- ② 当前**编译执行技术**在数据库应用越来越广泛，尝试在**openGauss**数据库执行器中找出一个编译执行的应用点，实现demo并给出性能效果。
- ③ 总结SQL引擎**查询重写规则**，尽可能给出较多的规则，找出其中未在**openGauss**上实现的规则，尝试将该规则在**openGauss**上demo实现。

THANKS

恳请各位专家批评指正