



《分布式系统》 课程建设与教学实践

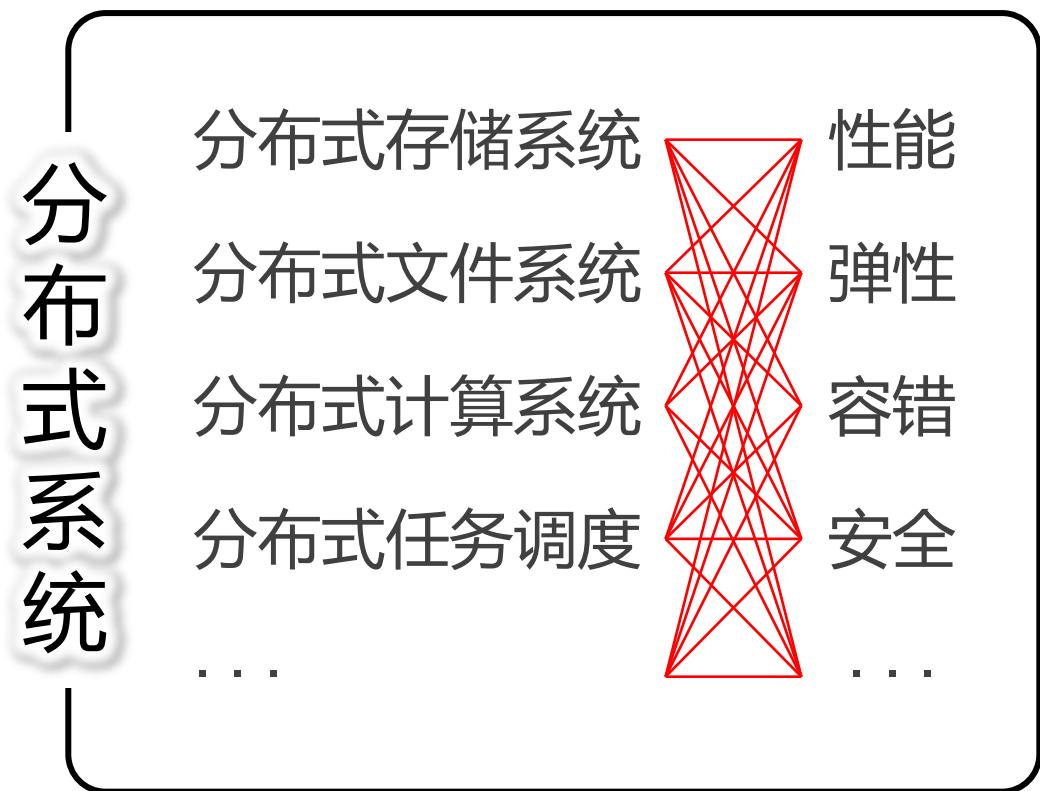
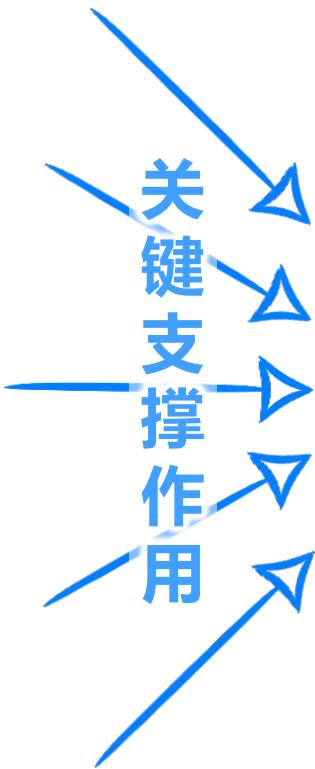
陈榕

上海交通大学

2021. 12

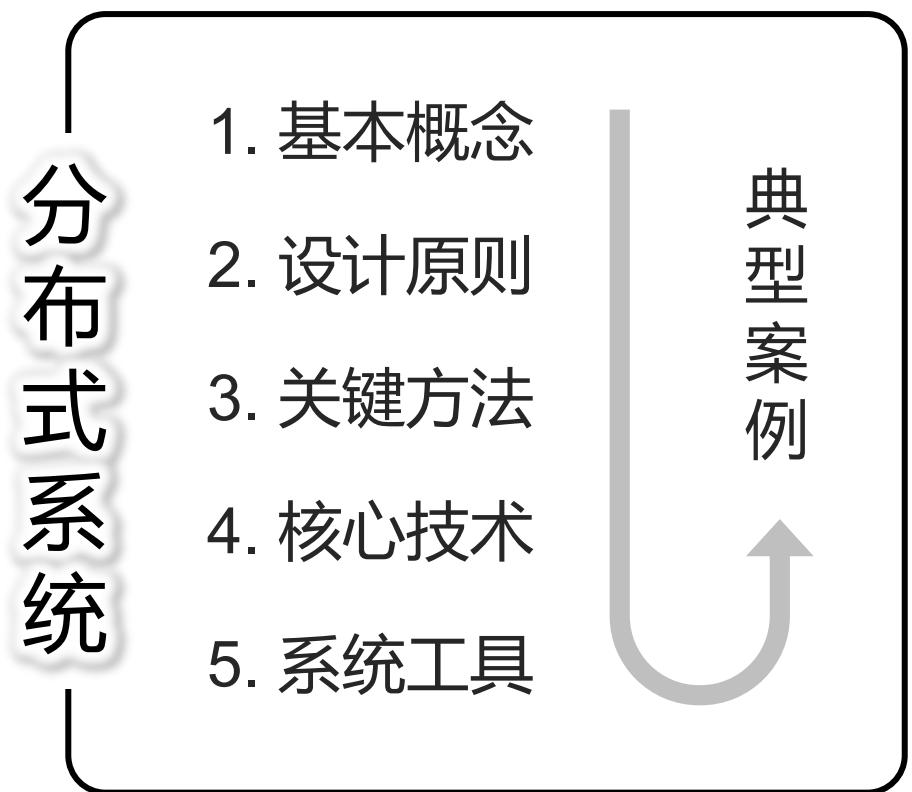
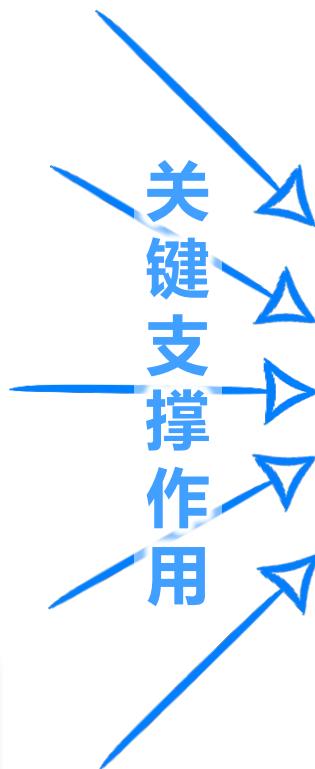
《分布式系统》课程背景

2



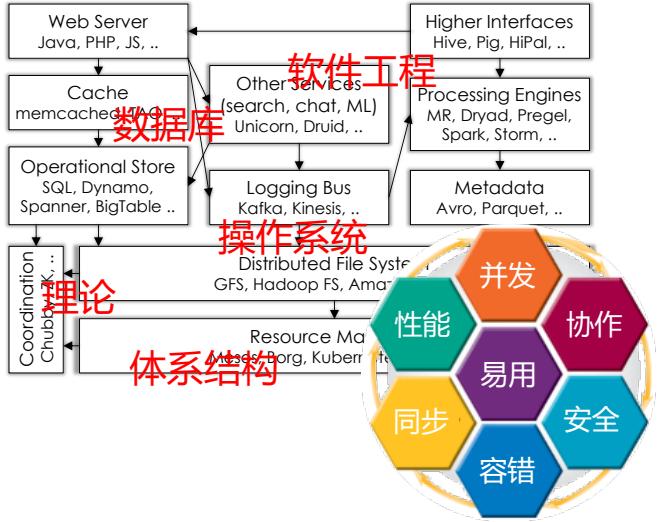
《分布式系统》课程背景

3



《分布式系统》课程挑战

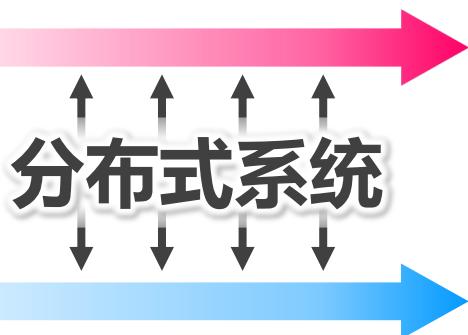
4



知识体系：多、散、变、活

应用

硬件

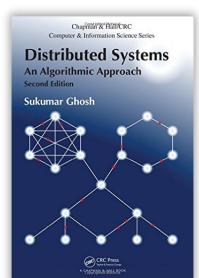
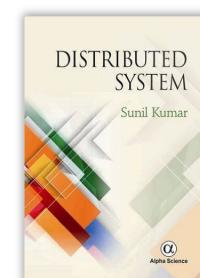
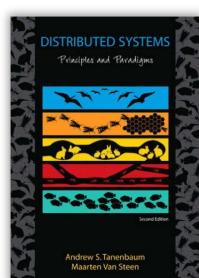


繁荣时期



课程教材

- 覆盖面差异大
- 缺乏公认体系
- 内容相对陈旧



大部分国外课程
并未指定教材，
课程内容体系也
有较大差异

《分布式系统》课程借鉴

5

国内外大学《分布式系统》课程调研

- 围绕“主题”安排课程内容
- 无指定教材(仅参考读物)
- 配套论文研读、实验项目等
- 分布式系统领域专家主讲

MIT: 6.824

Monday	Tuesday	Wednesday	Thursday	Friday
feb 15 feb 16 LEC 1: Introduction, video Preparation: Read Ch 1 of the book (Introduction) (2004) Assigned: Lab 1: MapReduce First day of classes	feb 17 feb 18 LEC 2: RPC and Threads, crawler.go, kv.go, vote examples, video Preparation: Do Online Go tutorial (FAQ) (Question)	feb 19		
feb 22 feb 23 LEC 3: GFS, video Preparation: Read GFS (2003) (FAQ) (Question) Assigned: Lab 2: Raft	feb 24 feb 25 LEC 4: Primary-Backup Replication, video Preparation: Read Raft, Tornet, Virtual Machines (2010) (FAQ) (Question)	feb 26 feb 27 LEC 5: Fault Tolerance, Raft (1), video Preparation: Read Raft (extended) (2014), to end of Section 3 (FAQ) (Question)	feb 28 feb 29 LEC 6: Q&A Lab 1, video Preparation: (Question)	feb 29 DUE: Lab 1
mar 1 No Class	mar 2 Monday schedule	mar 3 mar 4 LEC 7: Fault Tolerance, Raft (2), video Preparation: Read Raft (extended) (2014), Section 7 to end (but not Section 8) (FAQ) (Question)	mar 5 mar 6 LEC 8: Paxos, video Preparation: (Question)	mar 5 DUE: Lab 2A
mar 5 No Class	mar 6 Monday schedule	mar 10 mar 11 LEC 9: GAA Lab 2, video Preparation: (Question)	mar 12 mar 13 LEC 10: Fault Tolerance, Raft (2), video Preparation: Read Raft (extended) (2014), Section 7 to end (but not Section 8) (FAQ) (Question)	mar 12 DUE: Lab 2B ADD DATE
mar 15 No Class	mar 16 LEC 11: GAA Lab 3, video Preparation: (Question)	mar 17 mar 18 LEC 12: Zookeeper, video Preparation: Read Zookeeper (2010) (FAQ) (Question)	mar 19 mar 20 LEC 13: Guest lecturer on Go (Russo Cos, Google/Go), video Preparation: (FAQ) (Question)	mar 19 DUE: Lab 2C ADD DATE
mar 22 No Class	mar 23 No Class Lab 3: KV Raft	mar 24 mar 25 LEC 14: Guest lecturer on Go (Russo Cos, Google/Go), video Preparation: (FAQ) (Question)	mar 26 mar 27 LEC 15: Guest lecturer on Go (Russo Cos, Google/Go), video Preparation: (FAQ) (Question)	mar 26 DUE: Lab 2D
mar 29 No Class	mar 30 LEC 16: Chain Replication, video Preparation: Read CR (2004) (Question)	mar 31 apr 1 Remote Midterm Exam Materials: Open books, notes, laptop Schedule: 1pm through 10pm, Lab 1 and 2 Our exams	apr 2 apr 3 DUE: Project proposals (If you are doing a project)	apr 2 DUE: Lab 3A
apr 5 No Class	apr 6 Lab 4: Consistency, Fingers, video Preparation: Read Ergocons (FAQ) (Question)	apr 7 apr 8 LEC 17: Distributed Transactions, video Preparation: Read E033 Chapter 3, lab 9.1.5, 9.1.6, 9.5.2, 9.5.3, 9.5.4 (FAQ) (Question)	apr 9 apr 10 DUE: Lab 3B	apr 9 DUE: Lab 3A

16主题 | 16篇论文 | 4个实验
主讲 : Frans Kaashoek

Stanford: CS244b

Week of	Monday	Wednesday
Apr 6 - Apr 10	Introduction Handout: FLP Zookeeper	
Apr 13 - Apr 17	Two-phase Commit Paxos Mode Simple	
Apr 20 - Apr 24	Raft HARP	
Apr 27 - May 1	Practical Byzantine Fault Tolerance Honey Badger	
May 4 - May 8	HotStuff Coral	
May 11 - May 15	Dynamo COPS	
May 18 - May 22	GFS Aurora	

15主题 | 15篇论文 | 自选项目
主讲 : Jinyang Li

NYU: Distributed Systems

Date	Lecture	Lecture Preparation
9/4	Introduction, RPC and threads ds-intro.pdf ds-rpcthreads.pdf	[Online Go Tutorial]
9/11	primary/backup replication ds-pb.pdf ds-gfs.pdf	[Google File System] [Viewstamp Replication]
9/18	linearizability linearizability.pdf	[Linearizability]
9/25	consensus: paxos paxos.pdf	[Paxos]
10/2	consensus: raft ds-raft.pdf	[Raft]
10/9	causal consistency ds-causal.pdf	[Lamport Clocks] [Bayou]

11主题 | 21篇论文 | 5个实验
主讲 : David Mazières



国外大学《分布式系统》课程分析

优点

- ▶ 适合于分布式系统特点(技术点分散、不系统、变化快)
- ▶ 强调问题解决能力、系统设计能力，能紧跟技术发展前沿

问题

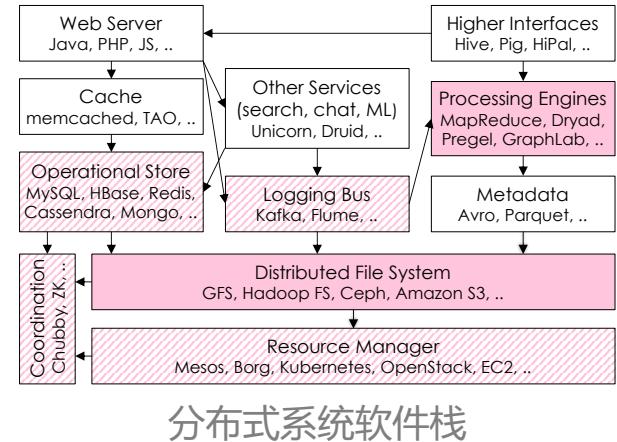
- ▶ 缺少展示技术/理论的发展历程、对于不同设计权衡的思辨
- ▶ 欠缺基础知识/背景介绍，对系统软件开发能力有较高要求

课程对象

- ▶ 软件工程专业研究生(各方向: SE | Theory | AI | System ...)

课程目标

- ▶ 掌握分布式系统的**核心问题、共性技术、(S|H需求/特征)设计权衡**
- ▶ 了解领域研究与产业前沿、历史变迁、未来趋势
- ▶ 重点培养学生(全体研究生)
 1. 解决问题的**系统化手段和方法**
 2. **能力**: 选型、组合、扩展、优化等





课程负责人

- ▶ 陈榕（软件学院）
- ▶ 研究方向：计算机系统、**分布式系统**
- ▶ 主要职责：课程建设、课程主讲等

在OSDI/SOSP/NSDI上发表多项**分布式系统研究**工作领导研发分布式数据库、图计算、键值存储等系统



课程信息

- ▶ 开设时间：2017～2021（春）
- ▶ 课程信息：**专业基础课**、48学时（3学分）
- ▶ 选课人数：49、74、75、69、70（**全体研究生**）

建设思路

- ▶ “经典”与“前沿”相结合 + “指定”与“自选”相结合

建设特色

- ▶ 以“主题”为中心，“新(2010s)旧(1990s)”对比
 - 先“旧”：核心问题、基础概念
 - 再“新”：(软硬件需求/环境)变化 >> 技术变迁和演化(权衡)
- ▶ 半自选团队SHOW：前沿论文(同主题、5年内)的讲解+演示
 - 团队任务、现学现用(变化>>权衡)、发挥自身特色和优势
 - 演示：围绕论文自选内容(如、应用/技术/实验 | 复现/开发/分析 ...)



课程内容

48课时

= 16周

x3课时

1. 旧

2. 新

3. SHOW

课程建设案例（上海交通大学/软件学院）



11

课程内容	基础	系统
48课时 = 16周	1. 分布式系统简介 2. 顺序一致性 3. 最终一致性 4. 原子性/日志 5. 分布式事务处理 6. 分布式提交协议 7. 分布式共识算法	8. 分布式文件系统 9. 分布式键值存储 10. 数据并行计算系统 11. 分布式图计算系统 12. 分布式数据划分 13. 分布式任务调度 14. 分布式容错方法 15. 分布式系统安全(可选) 16. 新网络硬件优化(可选)
x3课时 — 1. 旧 2. 新 3. SHOW		



课程内容

48课时
= 16周

x3课时

1. 旧

2. 新

3. SHOW

基础

1. 分布式系统简介
2. 顺序一致性
3. 最终一致性
4. 原子性/日志
5. 分布式事务处理
6. 分布式提交协议
7. 分布式共识算法

系统

8. 分布式文件系统
9. 分布式键值存储
10. 数据并行计算系统
11. 分布式图计算系统
12. 分布式数据划分
13. 分布式任务调度
14. 分布式容错方法
15. 分布式系统安全(可选)
16. 新网络硬件优化(可选)

存储

课程建设案例（上海交通大学/软件学院）



13

课程内容

48课时
= 16周

x3课时

1. 旧

2. 新

3. SHOW

基础

1. 分布式系统简介
2. 顺序一致性
3. 最终一致性
4. 原子性/日志
5. 分布式事务处理
6. 分布式提交协议
7. 分布式共识算法

系统

8. 分布式文件系统
9. 分布式键值存储
10. 数据并行计算系统
11. 分布式图计算系统
12. 分布式数据划分
13. 分布式任务调度
14. 分布式容错方法
15. 分布式系统安全(可选)
16. 新网络硬件优化(可选)

计算



课程内容

48课时
= 16周

x3课时

1. 旧

2. 新

3. SHOW

基础

1. 分布式系统简介
2. 顺序一致性
3. 最终一致性
4. 原子性/日志
5. 分布式事务处理
6. 分布式提交协议
7. 分布式共识算法

系统

8. 分布式文件系统
9. 分布式键值存储
10. 数据并行计算系统
11. 分布式图计算系统
12. 分布式数据划分
13. 分布式任务调度
14. 分布式容错方法
15. 分布式系统安全(可选)
16. 新网络硬件优化(可选)

高级

课程建设案例（上海交通大学/软件学院）

15

考核方式

- ▶ **书面考核(50%)**：问题分析、方法设计、优化 | 权衡、... (半开卷)
- ▶ **团队SHOW(40%)**：论文讲解、代码/功能/实验演示 (评分：教师 + 学生)
- ▶ **平时成绩(10%)**：论文预习问答、团队SHOW提问

半开卷：一张A4



课件

新工具辅助

二维码 (问卷/投票)



课程主页

A screenshot of the course homepage titled 'ADVANCED DISTRIBUTED SYSTEM SPRING 2017'. It shows staff information for 'Advanced Distributed System' and a brief description of the course.

上交CANVAS系统

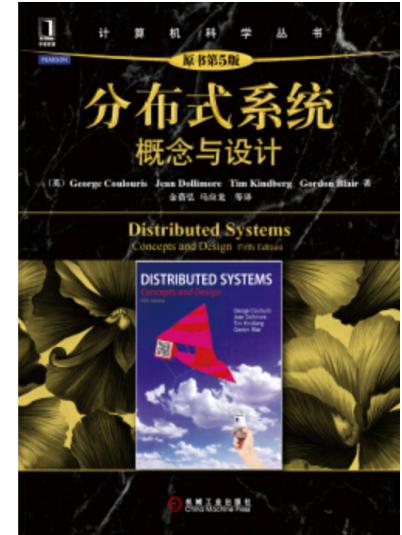
A screenshot of the course homepage titled '并行与分布式处理'. It lists course materials such as '1.1 Introduction to Parallel and Distributed Systems', '1.2 Sequential Consistency', '1.3 Sequential Consistency', '1.4 Sequential Consistency', and '1.5 Recovery & Replication'.

课程建设案例（上海交通大学/软件学院）

16

课程材料

- ▶ 课件/PPT（公开）
- ▶ 无指定教材(提供自学参考读物)
《分布式系统：概念与设计（原书第5版）》
George Coulouris, Jean Dollimore, Tim Kindberg, Gordon Blair
- ▶ 课程主页：<https://ipads.se.sjtu.edu.cn/courses/ads/index.shtml>
历年资料：课件、论文、团队SHOW



Home/News Schedule Staffs Prior Materials

ADVANCED DISTRIBUTED SYSTEM

SPRING 2021

Advanced Distributed System

Distributed systems help programmers aggregate the resource of many networked computers to construct highly available and scalable services. This course teaches the abstractions, design and implementation techniques that allow you to build fast, scalable, fault-tolerant distributed systems.

Staffs

Lecturer: Rong Chen
rongchen@sjtu.edu.cn
Teaching Assistant: Sijie Shen
ds_ssj@sjtu.edu.cn

Schedule			
Date	Lecture	Pre-course reading	Review
2.22	Lec.1 Introduction & Distributed Systems Lecture Slides		
3.01	Lec.2 Sequential Consistency Lecture Slides	Memory Coherence in Shared Virtual Memory System Lecture Slides	R01: Making Causally-Sorted Systems Fast as Possible. Consistent when Necessary [Slides & Demo]
3.08	Lec.3 Eventual Consistency Lecture Slides	Don't Settle for Eventual: Basic Causal Consistency for Wide-Area Storage with COPs Lecture Slides	R01: Making Causally-Sorted Systems Fast as Possible. Consistent when Necessary [Slides & Demo]
3.15	Lec.4 Recovery & Logging Lecture Slides	Reimplementing the Cedar File System Using Logging and Group Commit Lecture Slides	R02: After A Scalable Approach to Logging [Slides & Demo]
3.22	Lec.5 Concurrency Control: 2PL / SI Lecture Slides	A Critique of ANSI SQL Isolation Levels Q02	R03: PSI [Slides & Demo]
3.29	Lec.6 Consensus: 2PC Lecture Slides	Sinfonia: A New Paradigm for Building Scalable Distributed Systems Q03	R04: TAPIR [Slides & Demo]

Paper & Questions

Lec.2 Question (Do not need to submit)

Paper: [Memory Coherence in Shared Virtual Memory System](#)

[ivy-code.txt](#) is a version of the code in Section 3.1 with some clarifications and bug fixes. The write fault handler ends by sending a confirmation to the manager, and the "Write server" code in the manager waits for this confirmation. Suppose you eliminated this confirmation (both the send and the wait) from the system. Describe a scenario in which lack of the confirmation would cause the system to behave incorrectly. You should assume that the network delivers all messages, and that none of the computers fail.

Lec.3 Question

Paper: [Don't Settle for Eventual: Scalable Causal Consistency for Wide-Area Storage with COPs](#)

Suppose an application client at data center D1 writes object x with version 2 (x_2) and then object y with version 3 (y_3). Suppose y_3 has propagated from data center D1 to data center D2 but x_2 has not yet arrived at D2. Suppose another application client data center D2 has just read y_3 , is it possible that it might read x_1 next? (If not, why not?) Will the client be blocked waiting for x_2 to arrive from D1? (If not, why not?)



教学效果

学生评教结果¹

- ▶ 2021年春：评教档次:A1 (全校排名:130/1047、学院排名:10/164)
- ▶ 2019年春：评教档次:A1 (全校排名:312/1354、学院排名:28/175)

评教内容：

最高档次

建议在之后教学中改进

教学规范 | 教学内容 | 教学方法 | 师生互动 | 教学效果 | 教学资源

后续科研基础

如：2018春学生谢夏婷 | 基于RDMA的分布式图数据动态迁移 (ATC 2019)

2017春学生王思源 | 基于RDMA+GPU的分布式图查询优化 (ATC 2018)

1. 2018-2019学年起可查询学生评教结果，2020年因疫情未进行评教



分布式系统成为云计算、大数据、人工智能等新兴领域的关键支撑

《分布式系统》是软件工程一级学科研究生核心课程之一

课程建设应充分考虑课程特点、学生情况、培养目标

特色化道路（上海交通大学软件学院）

- ▶ “经典”与“前沿”相结合：以“主题”为中心，采用“新旧”对比
- ▶ “指定”与“自选”相结合：半自选团队SHOW(运用所学、发挥特长)

五年课程建设初步成型，内容/方法/工具等仍需保持**与时俱进**



衷心感谢 敬请不吝指正