

华中科技大学-武汉光电国家研究中心

并行数据存储研究室（PDSL）

实习生招生简章

一、研究室介绍：

本项目组是谢长生领导的项目组之一，本项目组老师有谢长生（原国光副主任）、万继光、谭志虎。万继光负责具体的日常项目研发，有 20 多年的 Linux 内核开发经验，热爱计算机科研，在学生中有很好的口碑。

实验室项目介绍：这几年，实验室先后申请了 2 个国家重点基金，2 个 973 项目，1 个 863 项目；另外跟华为、PingCAP、商汤等企业长期合作项目，经费充足。目前非常需要核心的开发人员，特别是对 Linux 下算法编程有兴趣的同学。

有国外教授指导：项目组跟 4 个美国教授长期合作研究。实验室老师和部分博士有美国学习交流经历。

有很好的团队：项目组有一个激情的强力团队，可以互相学习，目前有 10 名博士（全部本科保送直博的），25 名硕士（基本上都是本校保送的），比如 2015 级第 2,3 名，2016 级第 1 名，2017 级的第 7 名，2018 级的第 9 名（CS+ACM+校交班），2018 级计卓第 2 名。另外，每年都有多名联创团队和点团队学生来读研。

人性化管理：计算机科研是创造性的工作，需要激发大家的主动性和创新性，采用团队气氛带动研究生科研，没有打卡的严格管理制度，老师、博士、硕士以及实习生平等交流，相互学习。实验室的制度是老师与学生一起商量制定的，在做项目过程中，鼓励大家提出不同的意见和建议，对事不对人；对我提意见最多的同学，往往就是跟我关系最好的。

项目组织形式：做项目采用自由组合方式，靠的是学习的愿望和编程的兴趣。参加项目上，充分尊重学生的意愿。我们没有惩罚，是采用教育与奖励结合，优秀的同学，我们就从多方面给与更多支持和奖励。

项目特点：系统级软件开发，创新研究与工程开发结合；注重研究成果的应用，论文基于企业碰到的真实问题，做有实际意义的创新研究，企业没有解决的问题我们去解决，创新研究难度大；不追求数量，但要求每个论文都要把原型系统做出来，工程开发工作量大。

研究生培养计划：大四参与项目，重点学习编程；硕一独立负责项目模块；硕二负责整个项目，负责项目组织，设计与实现全过程；研三找工作，写毕业论文，并做项目指导工作。每个新同学刚到实验室，会安排高年级同学指导。

丰富多彩的活动：我们的目标是**创新研究、开心生活**，有一个值得怀念的研究生生活。实验室注重课外活动，一般包括每周组织一些体育锻炼的集体活动，比如光谷体育馆打球等；每个月组织一次集体的户外活动，比如东湖绿道骑车、森林公园烧烤、KTV、轰趴和联谊等。

研究方向：

1. **分布式大数据键值存储系统研究（跟 PingCAP 合作）：**大数据背景下，传统文件系统或数据库系统已经无法满足复杂应用对存储性能的需求，而键值存储使用线性数据组织方式并具有简洁的用户接口，实现起来简单高效。键值存储保证了服务质量和用户体验，因而成为当前数据中心大规模应用的主要存储方式，目前国外 Google、微软和 Amazon 等，国内的 PingCAP、华为、阿里、腾讯和字节等都在大力研究，比如阿里的 OceanBase 和 PolarDB 底层存储都采用键值技术。项目组毕业的**姚婷**博士获得华为**天才少年 offer**。

2. **云存储技术研究（跟华为合作）：**包括云平台内部的分布式云存储技术研究、云平台上的分布式存储技术研究，云平台上的存储+AI 技术研究等。目前跟华为签了两个联合研究中心：“**云存储技术创新中心**”（我们组独立承担）和“**变革型存储技术创新中心**”（承担创新中心的三个核心项目组之一“分布式存储系统”）。

3. **智能存储技术研究（跟商汤，PingCAP 合作）：**一方面，人工智能的是因为计算能力和海量数据集的发展而火起来的，人工智能系统需要底层的存储系统支持，因此，我们研究面向海量人工智能数据集的存储系统技术；另外，也研究将人工智能技术用于存储系统，比如采用智能调度技术进行存储性能优化。

开发平台：基于 Linux 平台系统级开发，开发工具用 C，C++、Go 或 RUST 语言编程，强调操作系统级算法编程。有些同学以为存储是做硬件的，这个是一个误解，存储主要是做**系统级软件开发**，涉及的比较密切的课程有 C 语言，数据结构，操作系统。

二、实习介绍（实验室与 PingCAP 联合组织）

1. **实习内容：** **分布式大数据键值存储系统开发**；具体课程见：

<https://university.pingcap.com/talent-plan/>

<https://hustport.com/t/talent-plan>

2. 实习安排:

- 1) 第一阶段（**课外时间或者假期**）：分布式 KV 相关课程学习，学习掌握分布式一致性算法 Raft，掌握分布式存储基本原理，实现一个简单的分布式 KV 存储系统。实习期间，表现优秀的，有意向保研的，我们实验室**优先录取**。
- 2) 第二阶段（**远程参与 TiDB 开源项目实训 2~3 个月**）：在光电研究中心完成第一阶段的实习，验收合格的，可以申请 TiDB 开源项目实训（也就是 **Talent Challenge Program**），时间 2-3 个月，完成项目挑战可获得 PingCAP 赞助的项目**挑战基金 5 千~1 万元不等**。
- 3) 第三阶段（参与**实验室项目**或者在 **PingCAP 公司实习**）：顺利完成 Talent Challenge Program 的同学，将获得 Talent Plan 结业证书；表现突出的，可以申请到 PingCAP 公司实习，实习待遇为 **300 元/8 小时**（不打卡，自己计时报给公司），此外每月还会有 **1500 元餐费**补助。

3. 招收要求:

- 1) **学习态度认真，责任心强**：不管做什么事情，也不管是否是自己喜欢的工作，只要接手，就尽力做好。
- 2) **要有团队合作精神**：能够跟团队成员很好的合作开发；
- 3) **有一定的编程基础，对软件编程有兴趣**：对项目编程经验不要求，我们看重的是态度和团队精神，相信大家只要认真负责，刻苦学习，编程不是问题；
- 4) **实习自愿原则**：如果录取了，在正式实习之前，随时可以退出。

4. 实习地址：光电国家研究中心 B521 室。

5. **联系人：万继光**；有兴趣的同学请将简历发送到 jgwan@hust.edu.cn，也欢迎加 QQ：24977886 交流。

三、PingCAP 公司介绍

PingCAP 成立于 2015 年，是面向全球的开源的**新型分布式数据库**公司，其

核心产品 **TiDB** 是一款定位于在线事务处理/在线分析处理（HTAP: Hybrid Transactional/Analytical Processing）的融合型数据库产品，实现了一键水平伸缩，强一致性的多副本数据安全，分布式事务，实时 OLAP 等重要特性。同时兼容 MySQL 协议和生态，迁移便捷，运维成本极低，是**大数据**时代理想的数据库集群和云数据库解决方案。

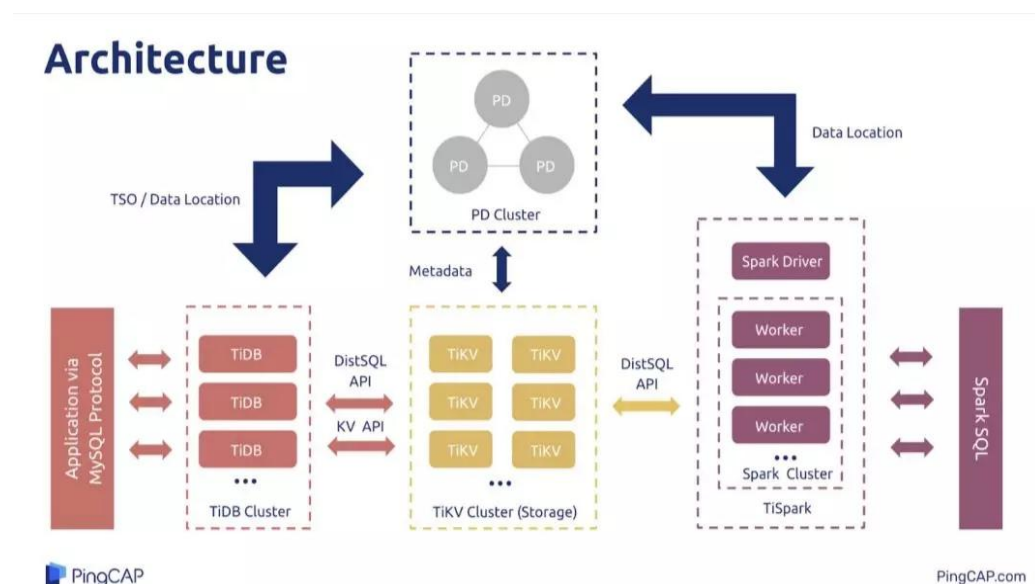


图 1 TiDB 的分布式架构图

TiDB 架构如图 1，其中，TiDB Server：接收 SQL 请求，处理 SQL 逻辑；通过 PD 找到数据所在 TiKV 地址，并获取数据；不存储数据，无限水平扩展。PD Server：存储集群元数据,集群的调度和负载均衡。TiKV Cluster：分布式 KV 存储引擎；基于 raft 协议的多副本管理。

2019 年的第十届中国数据库技术大会，评选的年度最佳创新产品中，**PingCAP** 公司是入选的唯一一家创业公司，其它入选的公司有 3 家：阿里、华为和腾讯。

PingCAP 此前曾在 2015 年获得经纬中国的天使轮融资，2016 年获得云启资本领投 A 轮融资，2017 年获得华创资本领投 B 轮融资，2018 年获得复星、晨兴资本领投的 C 轮融资，2020 量获得 2.7 亿美金的 D 轮融资。**TiDB 长期占据墨天轮国产数据库流行度排行榜的霸主地位。**

TiDB 已有准生产测试用户 1400 余家，其中 300 余家企业已经将其应用在

实际生产环境中，涉及互联网、银行、证券、保险、政府、电信、制造业、新零售等。主要用户如下：

Tencent 腾讯	小米	美团点评	头条 今日头条	爱奇艺	搜狗	饿了么	去哪儿网	携程旅游	58同城
mobike 摩拜单车	凤凰网 IFENG.COM	Bilibili	VIP KID	小红书	易果	百草味	乐视云	智联招聘	人人车
hulu	SEA	秒拍	蘑菇街 mogujie.com	鬼族 MEIZU	熊猫直播	趣头条	G7 汇通天下	LinkDoc	SUNMI
雪球	七牛云	二维火	聚时城	富友云	哗啦啦	一亩田	CDS 首都在线	优居优住	andon九安
meitu美图	婚礼纪 Hunliji.com	春雨医生	猿辅导	一面数据 yimian.com.cn	美团	Mobikola	艾润 iRain	车置宝	卖座
HUAWEI	新东方	Lenovo	Haier	中国万达集团	永辉超市	澳优 Ausnutra	丰巢 HIVE	华润医疗 CR Healthcare	特来电
国家税务总局	中国移动 China Mobile	中国电信 China Telecom	中国联通服务 CHINA COMMUNICATIONS SERVICE	PMS品茗	HCR	金蝶云	Terren	安天	中兴技术
北京银行 BANK OF BEIJING	中国工商银行	平安产险	方正证券	国联证券	51信用卡	网宿科技	PING++	功夫贷	元素征信
360金融	马上消费金融	BEIKE FINANCE	量化派	凡普金科	美利金融	挖财	团贷网	黄金钱包	连连科技
盖娅互娱	YOUZU.COM 游族网络	西山居	EGLS	铜猴互动	FUNYOURS	小满	出门问问	美行科技	小美助教

四、研究组老师与项目介绍

并行数据存储研究组负责人**万继光**，男，教授，博士生导师，**华为云存储技术创新中心主任**，在美国中佛罗里达大学访问学习 1 年。主要从事计算机系统结构和计算机存储与网络存储系统研究。并获得了**华为基础服务产品部部长奖**，湖北省科学技术发明奖二等奖，及工业和信息化部项目鉴定：“高性能、多协议加固型磁盘阵列”。



1. 在研项目：

- 1) 华为与华科的“**云存储技术创新中心**”（华为共有云的核心存储系统）



（主持）；

- 2) 华为与华科的“**变革性存储技术创新中心**”（华为私有云的核心存储系统）（主持其中**分布式文件系统**项目组）；



- 3) 国家基金面上项目“横向扩展的智能混合式键值存储系统结构与原理研究”（主持）；
- 4) PingCAP 项目 1 项（基于 Intel 新存储介质的键值存储系统）（主持）；
- 5) PingCAP 项目 1 项（基于云平台的智能弹性云键值存储系统）（主持）；
- 6) 商汤项目 1 项（AI 平台的分布式存储系统优化）（主持）；
- 7) 深圳市项目 1 项（大数据键值存储系统）（主持）；
- 8) 研究所项目 1 项（存储+AI）（主持）；
- 9) 国家基金委创新群体项目 1 项（参加）；
- 10) 国家重大研发计划项目 1 项目（参加）；

2. 实验室负责人介绍：

实验室负责人**谢长生**，男，教授，博士生导师，原计算机学院存储所所长，原“信息存储系统”教育部重点实验室主任，原武汉光电国家实验室（筹）副主任，兼任中国计算机学会外围设备专业委员会副主任，信息存储专委会常委，国

际网络存储工业协会（SNIA）中国区专家委员会副主席。国家领导人参观实验室照片：

- 1) 温家宝总理参观实验室：最右边的是谢老师。



- 2) 李克强总理参观实验室，在旁边讲解的是谢老师。



- 3) 吴邦国参观实验室，在跟吴邦国交流。

