



绿色计算 产业发展白皮书

(2023版)

▶ 指导委员会：

梅宏、肖然、刘仁辰、陈大纪、窦强、何征宇

▶ 编写组成员：

任翔、赵鑫、钟伟军、李雪莲、李卫忠、胡晓晶、周嘉颖、周绍梁、祝秀秀、王骏超、张伟、何义、曾审聪、赵明惠、王枫、赵世凯、刘魁、陈祖欢、杨统凯、夏俊鸾、徐明灼、郭雅琳、崔安颀、杜伟、杜玉平、朱敏健、罗云、周庆飞、邓学明、曹力、夏天宇、李洋、付佳琪、黄耿、吴灵熙、詹年科、刘一飞、刘玉杰、杨敦峰、白玥、孙焯、季福胜、王鑫、鲍全宇、张楠、徐爽、吕磊、张凯迪、姚秉、周杰、张宇、隋成龙、许刚、任彤、吴平、王丽华、张善伟、陈争胜、张定春、陈新、宋康华、谢津、白岩、戴权、宋梦、戴俊杰、刘宽、李昆、区俊彦、李鸿雁、鞠可一

▶ 指导单位：

中国电子技术标准化研究院

华为技术有限公司

安谋科技（中国）有限公司

飞腾信息技术有限公司

蚂蚁科技集团股份有限公司

▶ 编写单位：

百信信息技术有限公司
麒麟软件有限公司
云宏信息科技股份有限公司
中国长城科技集团股份有限公司
NVIDIA技术服务（北京）有限公司
阿里巴巴集团控股有限公司
百度（中国）有限公司
北京宝兰德软件股份有限公司
北京东方通科技股份有限公司
北京海量数据技术股份有限公司
北京神州数码云计算有限公司
北京神州数码云科信息技术有限公司
北京万里开源软件有限公司
北京星辰天合科技股份有限公司
此芯科技(上海)有限公司
杭州鸿钧微电子科技有限公司
华云数据控股集团有限公司
昆仑太科（北京）技术股份有限公司
南京百敖软件有限公司
深圳市杉岩数据技术有限公司
深圳市遇贤微电子有限公司
四川华鲲振宇智能科技有限责任公司
苏州超集信息科技有限公司
统信软件技术有限公司
武汉长江计算科技有限公司
星环信息科技（上海）股份有限公司

版权声明

本白皮书版权属于绿色计算产业联盟。

使用说明：未经绿色计算产业联盟事先的书面授权，不得以任何方式复制、抄袭、影印、翻译本文档的任何部分。凡转载或引用本文的观点、数据，请注明“来源:绿色计算产业联盟”。

序

FOREWORD

绿色经济、低碳经济是全球共同追求的发展方向。习近平主席在联合国大会庄严宣布，中国将力争在2030年实现碳达峰，努力争取在2060年实现碳中和，为推动全球绿色低碳发展作出了重要的承诺。为实现这一战略目标，我国政府出台了一系列有力措施，工业和信息化部也通过制定相关政策、标准和指南，促进绿色数据中心行业的可持续发展，推动行业转型升级，提高数据中心的能效和环境表现。

绿色计算产业联盟自2016年成立伊始，就明确提出为世界提供绿色算力的选项，在梅宏院士带领下，成员单位的共同努力下，顺应软件开源与硬件开放的发展趋势，立足中国，面向全球，搭建开放共享的平台，在绿色计算产业研究、标准建设、测试认证、人才培养、国际合作、产业推广等方面的工作取得了积极进展，彰显了我国信息技术产业正在融入全球主流产业链和生态链，逐步走向底层技术，发力核心关键领域。

绿色计算产业联盟继成功发布《绿色计算产业发展白皮书》2021、2022版之后，此次发布的2023版白皮书融合了数十家核心会员单位的集体智慧，聚焦当今绿色计算的产业新趋势，总结绿色计算的技术新成果，梳理绿色计算的发展新空间，汇集从处理器、固件、整机、基础软件和行业应用等产业链伙伴的优秀解决方案，尤其是及时归纳总结出自去年“东数西算”工程全面启动和今年人工智能领域掀起“百模大战”大背景下，绿色计算产业发展的最新八大核心趋势，为行业 and 用户提供了重要的参考和指引。

当前，绿色计算产业发展已经驶入快车道，不断涌现出新技术、新业态、新模式，产业实现新突破正当时。绿色计算产业联盟以开放共享为核心理念，继续充分调研和分析当前绿色计算产业的标准化需求，进一步构建完善标准体系；坚持以应用为导向构建产业生态，促进软硬结合、产用融合；持续提升产业创新能力，推动绿色计算成果标准化和产业化，为全球绿色低碳转型作出更大的贡献。

陈大纪

中国电子技术标准化研究院副院长

目录

CONTENTS

引 言.....	1
第一章 绿色计算产业发展趋势.....	11
1.1产业空间.....	12
1.2产业趋势.....	13
1.3技术趋势.....	14
第二章 绿色计算处理器和固件生态.....	22
2.1综述.....	23
2.2处理器代表企业.....	23
2.3固件代表企业.....	35
第三章 绿色计算整机生态.....	39
3.1综述.....	40
3.2代表企业.....	40
第四章 绿色计算软件生态.....	47
4.1综述.....	48
4.2操作系统.....	48
4.3数据库.....	51
4.4虚拟化和容器.....	59
4.5分布式存储.....	62
4.6大数据.....	65
4.7中间件.....	70
4.8高性能计算.....	75
展 望.....	78

引言

INTRODUCTION

《绿色计算产业发展白皮书》2023版是绿色计算产业联盟（GCC）继《绿色计算产业发展白皮书》2021、2022版之后，第三次发布的绿色计算产业发展白皮书。2023版白皮书聚焦今年绿色计算产业的新发展新变化，梳理了绿色计算的最新产业空间、数据中心的最新需求以及处理器技术新的发展趋势，并汇集了处理器、固件、整机、基础软件和行业应用等软硬件生态的优秀应用案例，为行业 and 用户提供参考和指引。

当前，绿色算力产业发展已驶入快车道，不断涌现出新技术、新业态、新模式，产业实现新突破正当时。白皮书2023版总结归纳了以下八大核心趋势：



趋势1： 落实“双碳”战略，绿色计算从可选项走向必选项

习近平主席在2020年向世界庄严承诺：“我们愿承担与中国发展水平相称的国际责任，为全球环境治理贡献力量。力争于2030年前达到峰值，努力争取2060年前实现碳中和。”

2023年2月27日，中共中央、国务院印发了《数字中国建设整体布局规划》，按照“2522”的整体框架进行布局，在数字中国建设这场伟大的变革中，算力是核心的驱动力，算力规模年均增长率保持在30%左右，预计到2025年，我国数据中心耗电量预将增长至4000亿千瓦时，占全国用电量比重高达5.8%，在当前碳中和目标下，数据中心能耗转型迫在眉睫。各部委纷纷出台多项政策落实“双碳”战略，其中就包括对数据中心绿色、低碳、集约、高效提出具体要求，持续推进绿色数字中心建设，加快推进数据中心节能改造。

自2022年2月，“东数西算”工程正式全面启动以来，在京津冀、长三角、粤港澳大湾区、成渝、内蒙古、贵州、甘肃、宁夏等8地逐步布局建设国家算力枢纽节点，并规划设立10个国家数据中心集群，加快构建我国新型绿色算力网络体系。

2023年3月，新一代AI语言大模型GPT-4发布，全球巨头纷纷入局，中国市场也掀起“百模大战”，海量数据集、千亿级参数的AI大模型分布式训练，对绿色算力提出更高的要求。加快建设绿色数据中心，正成为保障资源及环境可持续发展的基本要求，打造绿色智算数字化底座，已从可选项变为必选项。

趋势2： Arm架构能效比优势显著，为世界提供绿色选项

目前，Arm架构处理器的性能已经赶上甚至在某些领域超越x86架构处理器，同时依然能保持较低的功耗，能效比优势十分显著，综合成本较传统服务器有着明显的降低，为数据中心带来绿色新选择。

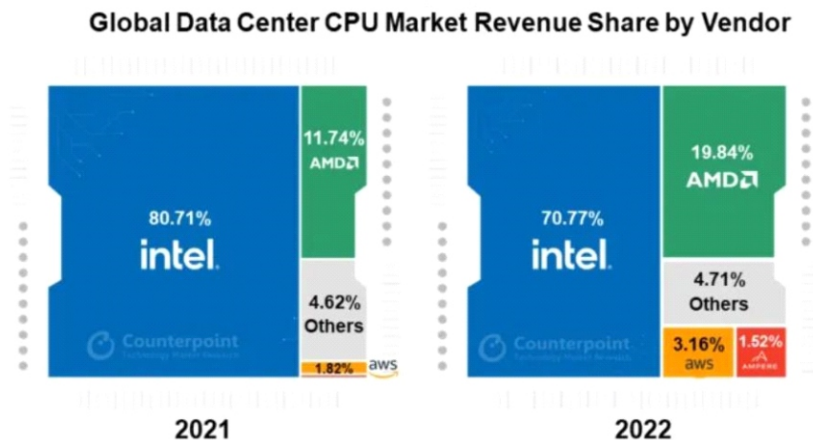
今年2月，国际学术期刊Nucleic Acids Research一篇被评价为“突破性进展”的论文：华中农业大学动物科技学院农业动物遗传育种与繁殖教育部重点实验室主任赵书红教授团队开发的基因组育种大数据计算新工具HIBLUP（天权），首创基于V矩阵的“HE+PCG”策略，并针对鲲鹏算力底座链接了鲲鹏数学库KML，实现了在1小时完成方差组分估计及育种值求解，而其他软件需要长达数周甚至数月的时间。

阿里云g6r是基于Ampere Altra云原生处理器，面向容器、微服务、DevOps等云原生场景的云实例产品。7月，智奥科技宣布完成DolphinDB系统在g6r云实例上的部署，性价比相对传统架构高40%以上。

实例名	ecs.g6.2xlarge	ecs.g6r.2xlarge
CPU	Intel 8269CY	Ampere Altra
核数	8 vCPU	
内存	32 GB	
OS	Alibaba Cloud Linux release 3 (Soaring Falcon)	
IOPS	3万	
费用	2.042 RMB/时	1.642 RMB/时

图表 1云实例对比

Arm架构在应用中的能效比的优势，赢得市场越来越多的正向反馈，第三方调研公司给出了整体的数据。Counterpoint 今年2月发布的2022年全球服务器CPU市场报告显示，英特尔数据中心CPU部门收入与2021年相比下滑16%，AMD的数据中心产品组合虽然同比增长62%，但主要侵蚀的是英特尔的份额，两家合计份额下降至90.6%。而基于Arm的CPU收入首次超过10亿美元，其中亚马逊云科技（AWS）市场份额为3.16%，相比2021年的1.82%提升了近一倍，紧随其后的Ampere Computing市场份额为1.52%。



图表 2全球数据中心CPU市场份额

除了在Counterpoint统计中重点提及的亚马逊云科技和Ampere以外，绿色计算产业联盟会员单位英伟达、华为、飞腾、阿里巴巴、百度、遇贤微电子、鸿钧微电子等在Arm架构产品方面的布局也群星闪耀，实力不容小觑，可以预见，Arm架构服务器芯片市占率的持续增长已经成为大趋势。今年一季度，Gartner数据显示全球Arm服务器出货量占比达到了约7%，IDC数据则显示同期全球Arm服务器出货量占比约为10%。Arm架构正以能效比的显著优势，为世界提供绿色选项。

趋势3： 算力应用环节的计算绿色化，将成为未来行业发展的重要一环

我国高度重视算力基础设施建设，工信部发布《新型数据中心发展三年行动计划（2021—2023年）》，积极推动数据中心向集约化、规模化、绿色化方向发展。但目前，行业内对于绿色数据中心的研究和实践多停留在电力生产/能源使用、算力供给环节，强调从能源使用效率角度优化数据中心的选址、设计、采购、管理，从计算资源角度优化IT基础设施的设计、建设、使用和回收过程，从而减少对于环境的影响。

当前行业重点关注数据中心电能利用效率PUE等指标，并积极应用高效IT设备、高效制冷方案、高效供配电系统、先进储能装置等技术解决方案，但伴随着产业蓬勃发展，PUE值优化空间在逐渐缩小。

与此同时，行业对于通过数据中心所产生算力的利用效率的关注度尚且不够，存在较大优化空间。麦肯锡早期曾做过一个调研，商用和企业数据中心的服务器很少超过6%的利用率，而高达30%的服务器处于“昏睡”状态，一直在耗费电力但没有提供有用的信息服务。面向算力应用环节的计算，起到将硬件算力有效转换成业务价值，直接赋能于上层场景的作用，因此，实现该环节的高效与节能，是实现端到端绿色算力的重要一环，也是下一阶段产学研应共同探索的潜在蓝海。

趋势4： 算力多样性特征广为显现，一云多芯成为云服务标配

当前，数据和应用的多样性驱动不同计算架构齐头并进。Arm、RISC-V等通用计算技术路线全面迸发，GPU、NPU、DPU等异构智算平台繁荣发展，算力的多样性特征已广为显现。

今年4月，为推动多样性算力技术和产业发展，中国移动、华为、清华大学、中国电信、中国联通等11家初创成员共同发起成立多样性算力产业及标准推进委员会，汇聚算力产业上下游组织涵盖了芯片IP，运营商，设备商，整机、部件，基础软件厂家和科研院所。

7月，一云多芯应用创新生态社区宣布成立。中移信息、浪潮云、天翼云、阿里云、中国电子云、联通云、华为云、云轴科技ZStack、华夏银行、农业银行等首批成员参与了启动仪式。

阿里云自研的飞天云计算操作系统可同时支持x86、Arm、RISC-V等多种芯片架构，用一套云端操作系统来管理不同架构的服务器硬件，可将不同架构处理器的运算能力标准化，从根本上解决不同类型处理器共存所带来的多云管理问题；云轴科技ZStack信创云平台作为唯一兼容x86、Arm、LoongArch和Alpha四种架构、八类平台环境的云平台，是首评云平台中支持架构和平台类别最全的云平台；天翼云一云多芯分布式云操作系统主要面向国产化CPU芯片架构，针对超

大规模资源池进行管理与调度，具备高性能分布式云存储、高性能大规模云网络、高性能计算异构算力管理能力，目前已应用于政务、制造、金融等多个行业，落地规模应用600余个，取得了显著的经济和社会效益；浪潮云海OS完成了业界首个“一云多芯”SPEC Cloud基准测试，以X86、Arm、Power等不同架构处理器为基础的产业生态框架已经就绪；易捷行云EasyStack“一云多芯”产品级云产品，支持但不限于X86、Arm等不同CPU架构服务器在一个云内混合部署，解决行业客户对多元算力的需求，同时控制行业公有云运营成本。

很显然，在算力供给侧，旨在纳管不同技术路线、不同技术栈，满足应用多样性需求，一云多芯已成为主流云服务商的标配。



趋势5：

从2+8到更广泛商业领域，绿色计算越来越受到市场青睐

从能用到好用，从2+8核心行业到更广泛的商业领域，基于Arm架构的绿色计算产品在需求侧正越来越受到行业用户的欢迎。尤其是2023年8月，中信银行通用基础设施集成商入围采购项目《入围通知书》发布，根据招标公告，该项目采购的服务器产品预估采购金额为44.21亿元人民币，其中Arm芯片服务器为34.01亿元，占比高达76.9%，表明Arm芯片服务器综合能力不断赢得市场与客户的信赖，市场竞争力持续提升。

鲲鹏处理器已呈现“2+8+N”的发展态势，目前已有11家合作伙伴推出基于鲲鹏主板的服务器产品，2022年整机伙伴发货占比达95%，已有4500多个合作伙伴，13500多个解决方案通过鲲鹏兼容性测试认证，在金融、电力、政府、电信等行业全面落地。

事实上，飞腾在金融、电信、电力等各个行业所取得的成绩与鲲鹏相比也不遑多让。在金融领域，飞腾CPU已支撑某国有银行OA办公系统上线运行，基于飞腾平台的服务器类、终端类和金融机具类产品成功入围六大行，同时也在金融主管机构实现全入围，终端入围160多家金融机构，服务器入围80多家金融机构。飞腾产品实现了金融信创单位全覆盖，包含政策性银行、股份制银行、城商银行、保险、证券等其它金融机构。

飞腾还积极布局电信市场，与三大运营商建立了全面的合作通道并实现规模化应用。搭载飞腾CPU的服务器在三大运营商的云资源池中得到广泛部署，服务了大量国产化项目；基于飞腾CPU的终端也在不同运营商的多个地市营业厅陆续落地；基于飞腾CPU，已有多家生态伙伴推出了“云电脑”产品，正在与运营商开展相关的适配测试工作。5G方面，飞腾已联合京信等头部生态伙伴推出满足运营商技术标准和设备资质的5G云小站、5G核心网和边缘计算三大系列产品，飞腾5G云小站更是通过中国移动测试的唯一信创方案。

在发电、输电、变电、配电等领域，基于飞腾CPU的联合解决方案已全面落地应用或具备应用条件。飞腾CPU已经助力50余家电厂实现自主可控，覆盖火力发电、风力发电、水力发电以及特高压变电站。

在京东 618 购物节期间，飞腾CPU成功提供了核心算力支撑并顺利通过技术考验，继关键基础领域获得大规模应用之后，在电商平台等消费级市场也实现了规模化应用。

趋势6： 评测标准化与认证服务，牵引绿色计算全栈解决方案性能提升

一直以来，Arm 持续不断引领着行业最佳实践标准和认证的推动。Arm SystemReady 计划于 2020 年正式推出，旨在促进行业内协作，实现无差异化软件开发，助力基于 Arm 架构的生态系统实现跨操作系统的快速部署（操作系统包括 Linux、Windows、BSD 和 VMware）。过去三年里，SystemReady 计划取得了重大突破，为计算奠定了通用软件基础，也为海量应用提供了支持。截至2023年上半年Arm SystemReady 计划累计突破 100 张认证，新的里程碑彰显出该计划强劲的发展势头，也成为推进软件“开机即用（Just Works）”愿景的重要一步。

绿色计算产业联盟自成立之初就把标准化工作视为绿色计算产业发展的重要抓手，持续汇聚各方资源、理顺创新链条、优化商业模式，加快构建具有先进性、创新性和国际性的绿色计算生态体系，打造高效率、低成本、低能耗的绿色计算产品。截至目前，绿色计算产业联盟已累计发布19项绿色计算体系标准规范，提升了底层架构的兼容性，促进了Arm 服务器技术的推广应用。

2023年，绿色计算产业联盟已立项在研标准包括《服务器可服务性评价技术要求》《云手机用户体验质量评价技术规范》《移动办公云手机技术规范》；计划立项标准包括《服务器应用场景性能测试方法 超融合》《服务器应用场景性能测试方法 分布式数据库》；拟发布《绿色计算服务器典型应用场景性能对比测试分析报告(2023)》。

为规范建设Arm全栈标准体系，促进产品技术迭代升级，更好地展现Arm架构和算力优势，同时通过官方发布权威认证，支持优秀产品与解决方案品牌提升和市场拓展。绿色计算产业联盟开展了“服务器典型应用场景性能评测行动”（简称：PECA评测行动），旨在推动底层服务器硬件和上层应用软件的联合调优，以体现Arm全栈解决方案的整体竞争力，包括大数据、分布式存储、Web应用、高性能计算、集中式关系型数据库、Arm原生云手机、虚拟化7大服务器典型应用场景。

目前，已有星环科技一站式大数据基础平台Transwarp Data Hub(TDH)、海量数据Vastbase G100v2.2、云宏CNware WinStack虚拟化云平台软件V8.1、东方通应用服务器TongWeb7.0、星辰天合科技天合翔宇分布式系统V5、华为鲲鹏BoostKit解决方案（HPC）、华为Arm原生云手机、

北联国芯Vastbase G100 V2.2、宝兰德BES Application Server Enterprise V9.5.5、爱可生云树数据库软件 ActionDB v1.0、大道云行FASS分布式全闪存储系统 1.5、广电五舟PentaStor分布式存储系统 v2.1.2、宝德计算机PLStor D22000企业级分布式存储系统、万里开源软件万里安全数据库软件 V5.1、深信服信云sCloud 6.8.1、品高软件PolarDB数据管理软件 V2.0、品高软件品高基础架构云资源管理软件 V9.0等17项产品通过了《服务器典型应用场景性能评测》。

趋势7： 中国市场Arm服务器链日益完善，Arm PC产业蓄势待发

目前，我国从芯片、整机、操作系统、云平台、中间件到应用软件等技术领域的创新链、产业链已初步形成。

华为鲲鹏社区旨在推动鲲鹏全栈开放，使能全产业链伙伴创新，覆盖技术支持与开发、生态合作与使能、行业应用与实践、产业沟通与交流等内容，汇聚全产业链力量，打造面向鲲鹏产业链上下游的一站式产业社区。目前已有超过4500家伙伴加入，推出超过13500个鲲鹏认证方案。

飞腾开放实验室（POL）是面向行业用户提供产品迁移指导、适配验证、PoC验证、产品测评、技术培训等服务的企业级开放实验室，基础实验室涵盖5G、AI、安全可信、嵌入式、智能制造等方向，全国分布式部署。目前已有168个成员单位，25个行业生态联合实验室，4个伙伴共建实验室，3个合作测评认证实验室。

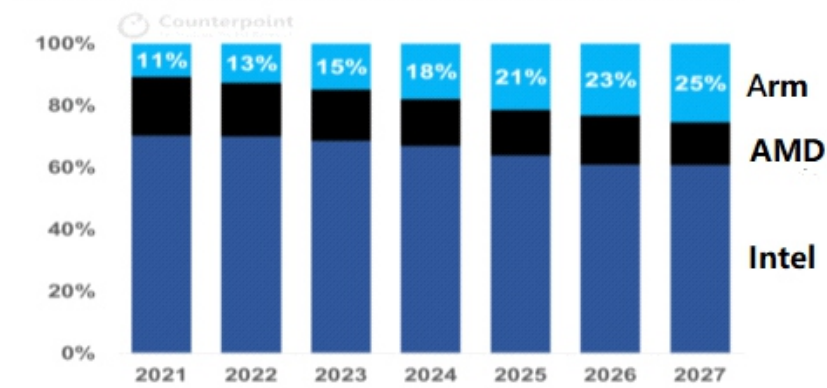
2023年5月，微软宣布原生支持Windows on Arm新的应用程序；6月，甲骨文宣布Oracle Database 19c支持Arm架构，无论在云端还是本地数据中心，都可以将Oracle Database 19c部署在Arm架构之上。这些重要的里程碑事件，显示出Arm架构在关键领域的竞争力。

7月，中国首个BMC根社区——OurBMC社区正式宣布成立。BMC是嵌入在计算机主板上的专用微控制器，以其多功能、易操作、广兼容、更底层的应用优势愈加受到各行业的青睐，成为服务器基础生态的重要组成部分。OurBMC社区由飞腾、昆仑太科、百敖等行业头部企业发起成立，旨在推进BMC技术快速发展，辐射上下游形成产业共振，加速构建繁荣的信息系统软硬件生态。

8月，Arm虚拟硬件（Arm Virtual Hardware）正式上线百度智能云，不仅将该服务通过云平台扩展到本土的个人开发者，同时也首次带来Arm虚拟硬件运行基于Arm架构云服务器的高效性能。当Arm发布全新IP时，开发者不必再等待物理硬件交付即可同步地开展相关的应用开发和验证，提升效率，降低成本。

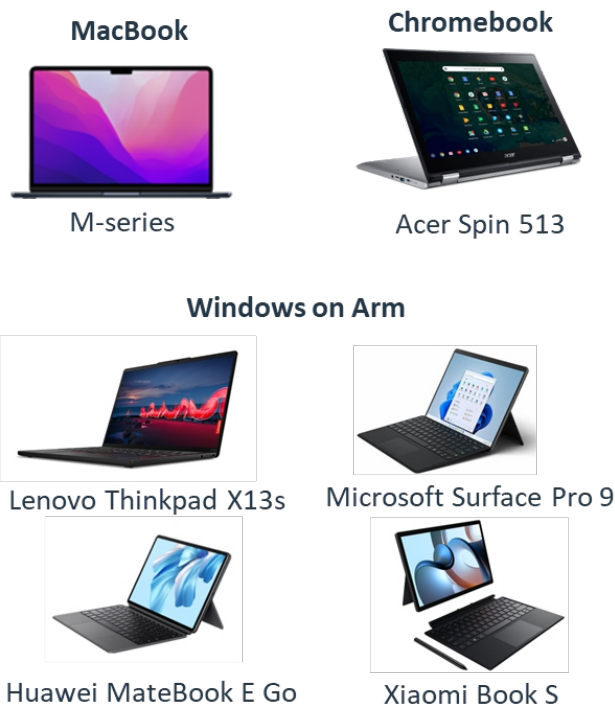
在PC领域，随着Arm架构性能提升和生态的完善，以苹果和高通为首的国际厂商推出可与X86 CPU生产力相媲美的Arm PC SoC，由于Arm架构天然的高能效特性，Arm PC展现出比X86 Pc更优

异的电池续航能力和更轻薄的特性，吸引大批消费者购买，在短短几年内Arm Pc在全球笔电市场的占比快速提升。



图表 3全球笔记本市场出货量份额

数据来源：Counterpoint



图表 4 Arm架构处理器笔记本示例

Arm的多核异构CPU架构设计，用大核来提供高性能计算力，以小核来满足轻负载计算，实现了性能和功耗的均衡，满足PC用户在不同场景下的需求，提供优异的用户体验。

5月，MAXHUB 正式推出搭载飞腾 CPU 的台式电脑，打造 MAXHUB 办公空间数字化解决方案，实现更快计算能力与更高性能，满足复杂应用场景下用户对性能和信息安全的需求，为专业办公赋能。

6月，飞腾开发者平台联合中国桌面操作系统根社区 openKylin 推出飞腾开发者平台首个联合技术板块——openKylin 专区，在开源、自愿、平等、协作的基础上，共享生态资源，建立软硬件生态兼容高效合作方式，推动 Linux 开源技术及软硬件生态繁荣发展。

8月，此芯科技与安谋科技在上海签署合作备忘录，携手在 Arm PC、AI、生态开发板等领域深化合作，共同推进桌面级智能 Arm v9 生态开发板的商业化落地和 Arm PC 生态建设。

随着端侧 AI 应用的快速增加，Arm SoC 灵活的架构中可以集成高算力 NPU，实现 AI 应用的高效部署和低功耗实现。比如针对视频会议支持新的 AI 特性：背景噪音消除、眼部交互、自动取景、背景模糊等，以及未来 AIGC 的端侧推理的部署。

总之，Arm 的高效异构计算架构以及天生的低功耗特性，使得 Arm SoC 在现在和将来会引领 Arm PC 生态在 PC 产业不断取得发展。

趋势8： 产业发展，人才为本，绿色大赛助力开源生态高质量发展

产业发展，人才为本。随着绿色计算产业的蓬勃发展，绿色计算亟需高质量的人才。

为激发开源创新活力，培养开源实践人才，助力开源生态建设，探索开源教育改革，绿色计算产业联盟持续打造的中国软件开源创新大赛已成功连续举办了5届，累计覆盖1000多所高校，80000多位参赛选手，举办了50多次 Arm 服务器架构和开源软件系统培训。

今年，第六届“中国软件开源创新大赛”在国家自然科学基金委信息科学部的指导下，由中国计算机学会（CCF）主办，西北工业大学、绿色计算产业联盟、CCF 开源发展委员会联合承办，以“小我创造·无限精彩”为主题，围绕“开源项目贡献”“开源任务挑战”“开源案例教学”“开源代码评注”四大赛道展开，助力开源生态建设的高质量发展，各项赛事及交流活动正在全国各地紧锣密鼓地进行。

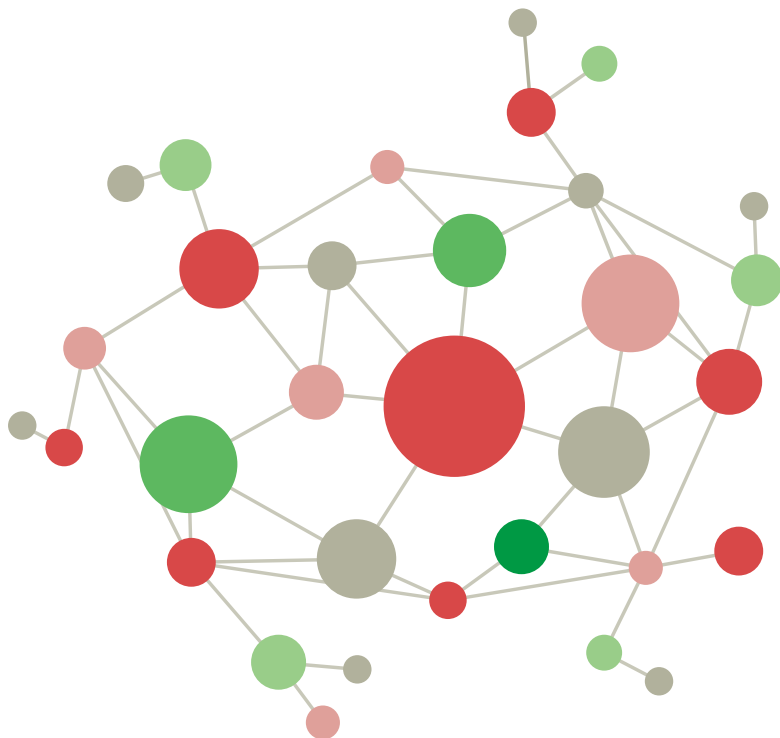
此外，华为、飞腾、蚂蚁等绿色计算产业联盟会员举办的开发者大赛也如火如荼地进行中。自2020年创办至今，鲲鹏应用创新大赛已连续举办三届。5月，鲲鹏应用创新大赛2023正式启动，总奖金池1000万。大赛以“数智未来，因你而来”为主题，由华为技术有限公司主办，鲲鹏生态创新中心承办，并联合中国软件行业协会、绿色计算产业联盟、中国计算机行业协会、“智能基座”虚拟教研室、“101计划”虚拟教研室、武汉光谷超级算力科技有限公司、国家超级计算深圳中心、CCF 武汉分部共同举办。相较往届，鲲鹏应用创新大赛2023在人群覆盖、专业领域和参赛权益等维度再

上新台阶，更进一步赋能开发者人才培养。大赛首次独立设置高校赛和科研赛，面向高校及科研机构的开发者群体开放，与企业赛一起，在开发者人群层面形成更广范围的覆盖。为持续推进数字技术创新，围绕三大赛事设有行业应用创新赛道、HPC应用创新赛道、openEuler开源创新赛道、openGauss开源创新赛道、机密计算创新赛道共计五大赛道，赛道的专业领域得以进一步延伸。

飞腾连续三年在中国研究生电子设计竞赛、全国大学生集成电路创新创业大赛等国家级学科赛事中设立“飞腾杯”赛道，通过设置企业命题、参与命题评审、提供参赛开发板、组织专题讲座、提供技术辅导、冠名企业奖项等多种形式，助力创新人才快速成长，在竞赛平台中脱颖而出。两项赛事中飞腾赛道共有500余支队伍参赛，百余所高校的参赛作品均采用飞腾芯片来展开研究与设计，有效提升高校人才生态建设。8月下旬，飞腾公司与萤火工场联合研发的开源硬件产品——飞腾派正式发布。该产品是首款基于飞腾定制芯片的国产化开源硬件产品，采用飞腾自主设计的定制款嵌入式CPU，以“硬件足够实惠、软件足够便利”为设计目标，具有芯片定制、接口丰富、体积小巧、系统开源四大特点。目前，飞腾派已在教育教学、人工智能、自动化控制等方面实现落地应用。同时，还公布了飞腾派“种子计划”，拟在飞腾派发布一年内，以创新大赛、现场交流会、产品赋能培训会等形式，培育不少于10000名飞腾派开发者，将助力更多客户认可国产方案、采用国产方案，激活和哺育国产开源软硬件生态，为国家信息产业培养更多复合型人才。

2022年，在绿色计算产业联盟指导下，蚂蚁集团主办了“绿色计算大赛”，成功打造了一场聚焦绿色计算主题的行业大赛。蚂蚁集团开放了自身几年来探索绿色计算技术的经验和实际案例，并设计成相关赛题，从系统软件层、云原生技术层、算法层等全方位考察算力利用效率与绿色低碳技术能力，吸引了超2000名的国内外高校学子和企业工程师参赛角逐。

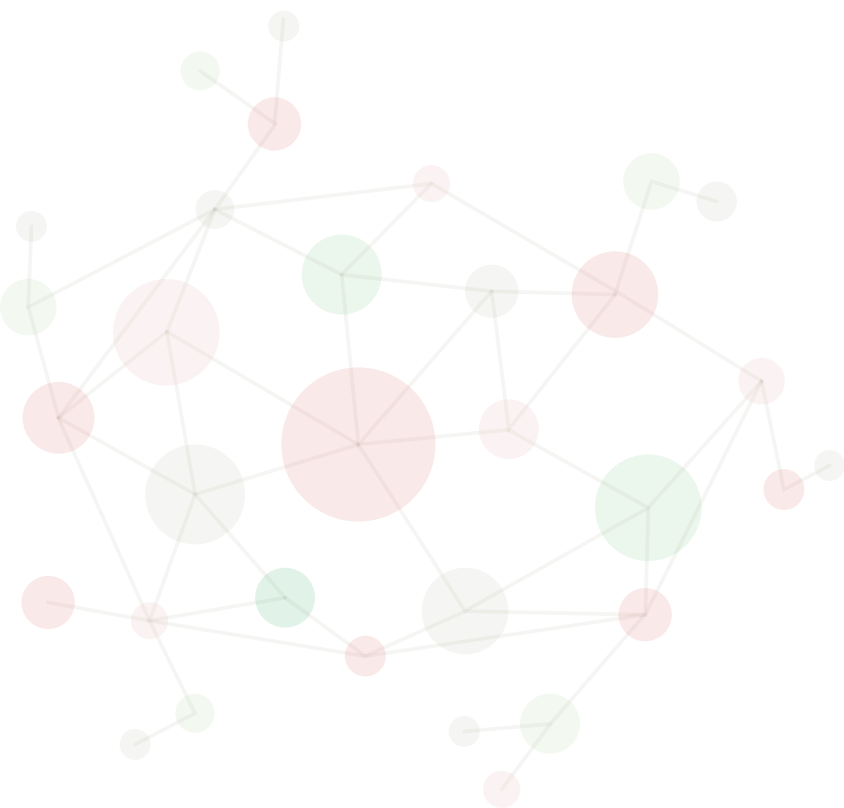
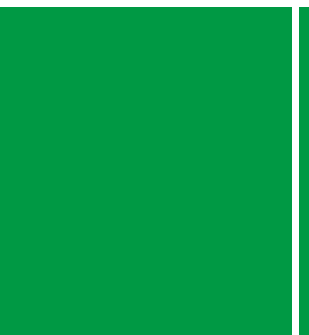
展望未来，绿色计算产业联盟将持续凝聚行业优势力量，引领产业从“行业渗透”向“生态繁荣”进阶，促进产业升级变革。聚沙成塔，积水成渊，发挥全方位、多角度优势，为打造绿色计算产业发展格局创造新机会，以标准评测为引领，共建新型绿色零碳生态圈，激发会员各自所长，实现能力拼图，一起碳路中国，共赢未来。



第一章

CHAPTER 1

绿色计算产业发展趋势



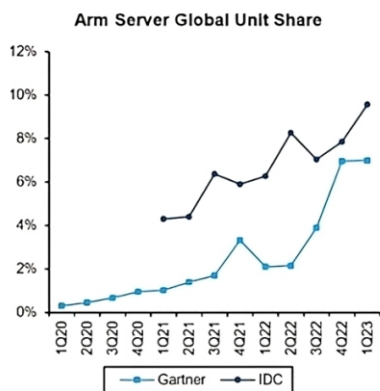
伴随东数西算、算力网络、一体化大数据中心、智能计算中心等战略工程的推进，我国计算产业迈入新的高速发展阶段。新阶段带来了许多挑战，同时也给整个计算产业界带来更多机遇，新技术、新产品、新业态、新商业模式正在大量涌现。

1.1 产业空间

绿色计算以基于开放授权Arm处理器技术为基础，继在控制器和移动计算领域的成功之后，以节能和安全为特点的Arm处理器技术逐渐成为计算基础设施的新标准。

全球范围内，Arm架构CPU在服务器市场处于快速上升期。Bernstein Research的一份研究报告显示，目前全球近10%的服务器已经用上Arm架构，其中多达40% Arm服务器位于中国，中国市场Arm服务器市场份额占比已经超过15%。

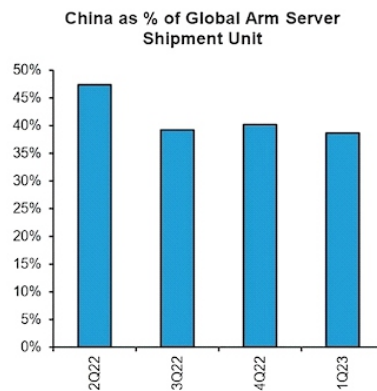
EXHIBIT 4: Estimates from market researchers differ but the rise of Arm CPU in servers is clear and consistent.



Source: Gartner, IDC and Bernstein analysis

图表 5 Arm服务器全球份额

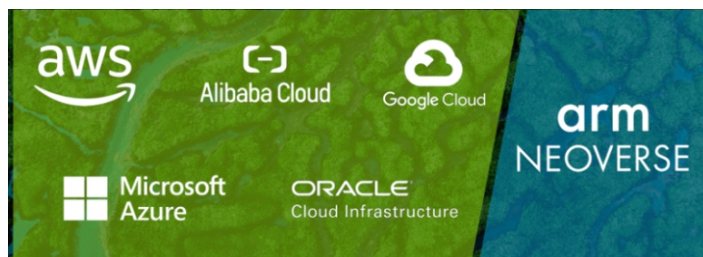
EXHIBIT 5: We estimate nearly 10% of servers now have Arm CPUs, with roughly 40% used by China.



Source: Gartner, Bernstein analysis

图表 6 全球Arm服务器中国出货占比

主要的公有云服务都提供基于Arm的实例，包括AWS、Azure、谷歌云、阿里云等。其中以AWS为首的Graviton最具市场规模，Bernstein Research报告2022年期，Graviton占20%左右的AWS CPU实例，AWS数据中心部署了全球50%以上的Arm服务器CPU。从而刺激其他主要的云端服务商跟进。另据Canalys分析师预测，2026年Arm架构将占据云服务器市场一半以上市场份额。



图表 7 主要公有云服务商提供基于Arm实例

在全球PC市场，Arm架构CPU同样处于快速上升期。Bernstein Research报告估计，2022年Q4出货的个人电脑（不包括Chromebook）中，Arm芯片约占12%，其中绝大多数是Mac电脑。另据Canalys分析师预测，2026年Arm架构将占据PC市场30%的份额。

1.2 产业趋势

云计算、人工智能、元宇宙、5G网络、物联网、区块链等新技术的广泛采用，不断地丰富人们的生活和提高社会生产力，同时也带来日益的数据量爆炸和计算能力需求的提高。

目前基础设施转为定制化打造，从SSD到HDD，从DPU到视频加速器，服务器CPU算是最后的标准产品，将不会作为通用型产品继续发展。与此同时，计算工作负载正极力增长，而且愈加复杂。ML和AI正在发挥取代作用。另外一个问题则是功耗问题，目前大型互联网公司的电力支出仅微次于电信网络运营商。为了满足这些需求，未来的云基础设施将需要应对即将到来的数据爆炸和日益复杂的工作负载的有效处理，同时提高能效并最大限度地减少碳足迹。这就是为什么业界越来越多地寻求更高性能、高能效、专业处理和工作负载加速，以重新定义和改变计算基础设施。

1. 从数据中心的需求到处理器的技术发展，绿色计算产业对于性能、能效和安全性提出更高的要求

数据中心的算力需求持续增长，AI成为重要计算任务。根据IDC数据显示，2020-2025年全球产生或复制数据量以CAGR 23%保持增长，并预计每四个月对算力的需求就会翻一倍。在“双碳”目标大背景下，减少碳排放，节能减排是重中之重。数据中心是数字基础设施的重要载体，作为典型的耗能大户，面临的节能任务更加艰巨与紧迫。随着网络与计算业务的快速发展，越来越多的关键性服务和高价值数据被迁移到了云端和边缘，与传统的本地数据保护策略不同，对数据的保护也变得更加复杂。从欧盟GDPR到我国个人信息保护法，数据隐私监管保护的范畴愈加扩大，力度日益增强，对关键数据和业务进行安全保护，是合规遵从的关键因素。

随着摩尔定律走向终结，解决方案提供商正在寻求专业化处理能力。Armv9-A架构，在安全性、能效和性能方面都有提升，通过越来越多的芯片产品将Neoverse的卓越性能、每瓦性能和TCO优势带到基础设施领域。

构建端到端的绿色算力链路，已成为我国算力行业发展的大势所趋。我国高度重视算力基础设施建设，工信部发布《新型数据中心发展三年行动计划（2021—2023年）》，积极推动数据中心向集约化、规模化、绿色化方向发展。但目前，行业内对于绿色数据中心的研究和实践多停留在电力生产/能源使用、算力供给环节，强调从能源使用效率角度优化数据中心的选址、设计、采购、管理，从计算资源角度优化IT基础设施的设计、建设、使用和回收过程，从而减少对于环境的影响。根据信通院《面向算力应用环节的计算绿色化白皮书》，算力行业应从端到端的视角出发，构建从电力生产、算力生产到业务应用各个环节的绿色算力，以实现算力产业从节能向绿色、增效并

重的延伸，实现绿色数据中心向端到端面向业务应用的全流程转型，实现单一产品视角转向软硬协同一体化的视角。

2、绿色计算产业的标准工作应该重点从性能、能效和安全角度出发，持续推动相关标准和评测规范的制定和落地

标准体系是指引产业方向、推动产业各方达成产业共识、促进商用进程的重要力量。标准和评测工作首先应根据业务的诉求，满足数据中心应用（例如人工智能、大数据分析、高性能计算、移动应用上云、容器等）对软硬件基础设施在性能、能效和安全等方面不断提升的需求。针对多样性算力的不同应用和产品，存在着不同的算力评价方法和工具，导致用户在评价和选型时难以适从，亟需完善多样性算力的评价体系和工具。

联盟和标准协会是聚拢各方力量，建立和完善标准的重要组织形式，目前绿色计算产业联盟GCC、边缘计算产业联盟ECC、金融开放联盟、中国电子工业标准化技术协会（CESA）、欧洲OEHI（Open Edge+HPC Initiative）、机密计算联盟（CCC）等联盟和组织都在积极推动产业上下游共同制定开放的绿色计算产业标准。

3、产业链上下游企业、开发者和用户需求的真正满足，亟需繁荣和多样性的绿色计算软硬件生态

计算产业是一个充分开放的体系，构建一个更加成熟的绿色计算产业，需要产业链上成千上万的上下游企业、上千万的开发者以及千行百业的用户共同努力和相向而行才能成功。中国每年近千万高校毕业生的人才基础，以及众多硬件和服务提供商，充分具备发展绿色计算产业生态的基本要素，但仍存在生态碎片化、条块化的挑战，各个领域都需要产业领先者来聚合产业资源、引领产业发展方向。

当今的国际形势对计算产业的稳定供应带来极大不确定性，世界比任何时候都需要多元化稳定可靠的算力基础设施供应，以支持各行业业务平稳发展。通过绿色算力创新部署，构建多样性算力平台，可实现算力特点和业务特点的更好匹配，以达到更高的业务质量和安全，满足千行百业数字化转型算力需求。

1.3 技术趋势

1.3.1 数据中心客户需求和趋势

1、算力需求持续增长，AI成为重要计算任务

根据 IDC 数据显示，2020-2025 年全球产生或复制数据量以 CAGR 23% 保持增长，并预计每四个月对算力的需求就会翻一倍。海量数据增长、数据频繁交互以及各类应用激增带来的巨大流量增长，加速数据中心网络向更高性能和更高灵活性方向演进。数据中心网络打破传统端口速率十倍增长的十年代际演进周期，加速向 100 Gbps、200 Gbps、400 Gbps，甚至 800 Gbps、1.6 Tbps 发展。根据华为预测，2030年通用算力对比2020年增长10倍，AI算力对比2020年增长500倍。TrendForce集邦咨询预估2023年AI服务器（包含搭载GPU、FPGA、ASIC等）出货量近120万台，同比增长38.4%，占整体服务器出货量近9%，至2026年将占15%。2022~2026年AI服务器出货量年复合增长率同步上调至22%，AI芯片2023年出货量将增长46%。

多样性算力需求突出：异构计算，GPU&DPU满足数据为中心的多样化应用场景的计算需求

全球数据量激增促使数据中心网络从“以计算为中心”转向“以数据为中心”，加速数据传输、提升用户体验，需要最小化后端请求的尾部延迟。网络带宽和连接数的剧增进一步推进算力需求爆炸式增长，而摩尔定律日趋极限，CPU 性能增长速度逐渐放缓。寻求更符合多样化应用需求的计算芯片已成为业界共识。

云化基础设施平台基于 CPU 完成网络数据转发的传统模式出现瓶颈。CPU 串行计算模式难以在高并发网络数据转发工作中发挥最大计算能力，另外网络带宽的增长速度远超 CPU 的算力增长速度，网络数据转发占用的 CPU 资源不断增大。数据处理单元（DPU）用于处理云计算的幕后工作：数据包检查和虚拟机监控程序管理等严格且对时间敏感的安全任务。网络、存储和安全任务现在可以消耗实例所需资源多达服务器CPU周期的30%。使用针对这些任务优化的专用处理器，能够将CPU集中用于工作负载而不是基础设施任务，可在电力和资本支出方面获得TCO好处。

2、片间互连异构集成，Chiplet模式加速兴起

进入后摩尔时代，异构集成（HI）和系统级封装（SiP）是重要出路。Chiplet模式采用不同于SoC设计的方式，将大尺寸的多核心的设计，分散到较小的芯片，再通过先进封装的形式以一种类似搭积木的模式实现整合，更能满足现今高效能计算处理器的需求；而弹性的设计方式不仅提升灵活性，也能有更好的良率及节省成本优势，并减少芯片设计时程，加速芯片上市的时间。**综合而言，相对于SoC，Chiplet将有设计弹性、成本节省、加速上市等三大优势。**

英特尔与AMD、Arm、高通、微软、谷歌、Meta、台积电、日月光、三星等十家行业巨头正式成立UCIe（通用芯粒高速互连）产业联盟，意欲共同打造Chiplet互连标准，携手推动Chiplet接口规范的标准化。国内厂商方面，包括芯原、超摩科技、灿导、芯和等多家半导体企业已经陆续加入。阿里巴巴也成为首家加入UCIe董事会的中国大陆企业。

3.低功耗，节能减排需求

在“双碳”目标大背景下，减少碳排放，节能减排是重中之重。数据中心是数字基础设施的重要载体，作为典型的耗能大户，面临的节能任务更加艰巨与紧迫。最新数据显示，我国数据中心用电量已经占到了全社会用电量的2.6%左右。随着数据流量呈指数增长，数据中心设备功耗也在不断攀升，电能消耗数目弥巨和运营成本居高不下的问题愈发凸显，其能耗问题将成为重点整治对象。数据中心能源侧的碳排放主要来自IT设备及基础设施的电力消耗，数据统计，在数据中心能耗结构中，主设备耗电占45%~55%，空调设备占35%~45%，电源及其他设备占10%~15%。当前大型互联网公司的电力支出占到总拥有成本(TCO)的30-40%，仅微次于电信网络运营商，功耗问题不容小觑。节能降耗无疑是数据中心的重中之重。

企业积极践行碳减排工作，开始综合考虑IT产品能耗比和碳排放量

头部企业开始不断探索节能减排新方式，例如，在IT层面采购低能耗的软硬件产品，并利用技术手段进行实时碳足迹监控和碳排放的核算。“绿色节能”已逐步成为除价格、业务性能之外的新考量点。

企业面临降本增效压力，精细化IT治理成为发展刚需

企业对扩大IT支出的态度愈发谨慎，从以“可用”为目标对相关IT解决方案进行高投入，转向以“性价比”为目标综合评估业务投入产出比进行IT支出。这些因素都驱动着企业向降本增效，精细化IT运营方式的方向转型。在硬件层面通过利旧、提升产品效能等方式，在软件层面通过IT成本监控和治理、提升面向业务应用的算力资源利用效率等方式，实现精细化的IT成本管控。其中，算力资源利用效率起到将硬件算力转换，从而直接赋能于更多数字业务发展的作用，因此具有巨大关注价值和提效空间。

4.安全，机密计算需求

在计算的过程中，数据存在三种状态：传输中、静止和使用中。数据于网络中传输属于“传输中”状态，存储的数据属于“静止”状态，正在处理的数据属于“使用中”状态。传统的安全技术对静态存储或网络传输状态的数据可以进行有效保护，但解决不了的个人敏感信息或者重要业务数据使用阶段的安全问题和风险，包括云计算服务中多租户数据安全问题。机密计算通过在可信硬件基础上配套软件构建加密、隔离、可证明的计算环境，在该环境中执行计算来保护运行中的数据 and 代码，它可以从服务器芯片级保障 OS kernel、Hypervisor 等特权软件在已经遭到破坏甚至本来就是恶意的情况下，仍然可以保证系统运行时敏感数据和代码的机密性和完整性。

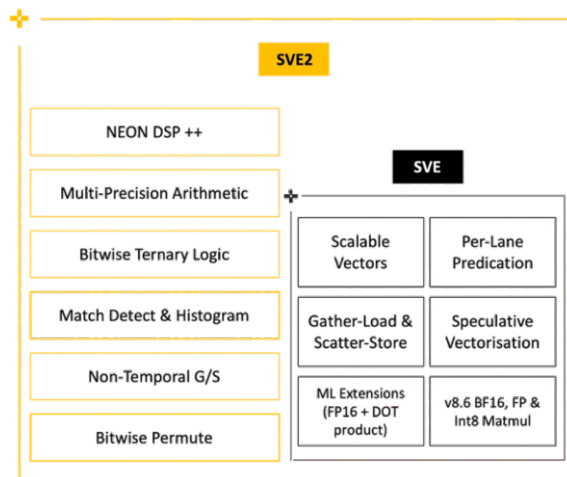
产业界对机密计算的需求来自多个方面，其一，是来自行业与用户的互信需求，需要在不可信的环境中保护用户数据安全及隐私；其二，来自行业自身的防御需求，在不可信的环境中通过有效手段，如增加攻击复杂性、实施深度防御等，来保护企业数据安全；其三，来自机构之间数据共享的需求，通过机密计算保障互信不足的合作方仍然能进行数据合作。

1.3.2处理器技术趋势

随着摩尔定律走向终结，解决方案提供商正在寻求专业化处理能力。越来越多的基于Arm Neoverse芯片产品将卓越性能、每瓦性能和 TCO 优势带到基础设施领域。

1、单核性能提升

采用更宽的微架构，以容纳更多的指令，来支持高性能和超级计算需求。宽深架构加上可伸缩矢量扩展SVE，能够更好提高单核性能，并为SoC设计者提供了实现灵活性。SVE 和 SVE2 支持可变长度矢量寄存器，允许开发人员部署和运行128位到2048位的代码，而无需重新编译，从而加快上市时间并为客户提供更多选择。SVE可以支持气候建模、疫苗研究和材料研究等科学工作负载，支持向量长度无关 (VLA) 编程解决自动矢量化传统障碍。SVE2 扩展了 SVE 指令集，支持包括 HPC 和 ML 之外的数据处理领域，包括加速计算机视觉、多媒体等多种应用使用的常见算法



图表 8 Arm 可伸缩矢量扩展SVE->SVE2提升性能

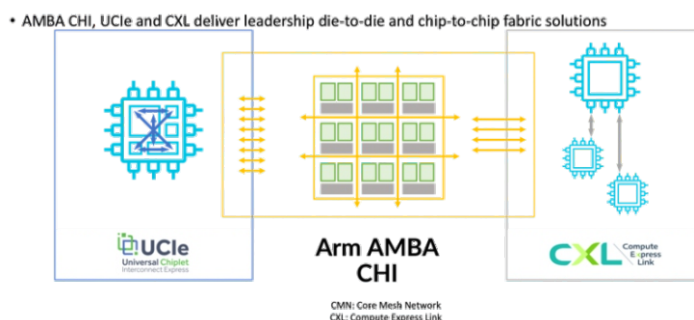
SME是在可伸缩向量扩展（Scalable Vector Extensions, SVE和SVE2）的基础上建立的，并增加了有效处理矩阵的能力，主要功能包括：

- 矩阵tile的存储
- 存取，插入，提取 tile向量，包括on-the-fly 转置
- 计算SVE向量的外积（Outer product）
- Streaming SVE 模式

2、多核异构网状互连

应用所需的性能无法通过纯粹以CPU为中心的解决方案实现服务器架构，而是需要基于高效异构的解决方案。包括各类加速器，例如图形处理单元、FPGA、智能网卡/DPU和许多其他特定域的可编程设备。

计算性能越来越高的同时，系统间互连的挑战越来越大，Arm 系统 IP CMN对于实现可扩展效率至关重要，它允许合作伙伴从低核心数扩展到最多 256 个核心，并以每秒近 4 兆字节的带宽实现。此外，CMN支持AMBA CHI、DDR5、UClc、PCIe、CXL等主流外围设备和芯片间互操作标准，确保兼容性。



图表 9 Arm芯片互联接口技术介绍

3、性能与功耗平衡

Arm开发了新的电源管理和低延迟工具，使系统软件能够在TDP限制定义的预定义电源和热包线内进行电源性能权衡。这些低延迟工具完全可编程，以扩展动态电压频率缩放（DVFS）的典型功能。其中包括最大功耗缓解机制（MPMM），它是一个可调的电源管理系统，允许客户以尽可能高的频率运行高核数处理器，而调度限制（DT）系统可以在某些工作负载和高IPC时降低功耗，比如矢量化工作。

4、安全可靠

安全在基础设施中的重要性怎么强调都不为过，多年来云服务提供商已经提供加密服务来对在存储以及在传输过程中的数据进行保护，机密计算通过保护正在使用的数据（在处理或运行中的数据），从而消除了数据计算时的安全漏洞，ArmV8-A架构下可以基于TrustZone TEE来实现机密计算，ArmV9-A CCA能更好满足机密计算所需的更高隔离和内存动态分配要求。Arm也提供了许多新安全功能来帮助解决基础设施的安全需求：

1) 指针身份验证（PAC）和分支目标指令（BTI）

防止面向返回的编程跳转的代码重用攻击，提供使恶意参与者可以利用的可用小工具最小化的机制。在glibc上，启用PAC和BTI将可用小工具的数量减少了97%以上，而代码大小的损失仅为1%到-2%。

2) 内存标记扩展（MTE）

内存安全问题占安全漏洞的70%以上，MTE提供了一种检测内存安全冲突的机制。MTE通过提高测试和模糊化的有效性，帮助在部署之前检测潜在的漏洞。MTE还可以帮助在部署后检测大规模的漏洞，易于部署的内存安全违规检测和缓解措施可以防止一大类安全漏洞被利用。

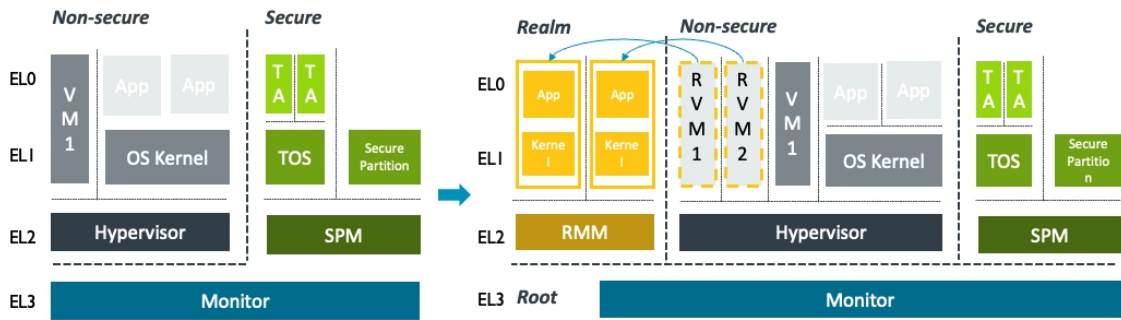
3) Secure EL2

多个供应商提供的多个可信应用都运行在安全世界中，将这些应用相互隔离变得越来越重要。Secure EL2扩展增加了对安全世界中虚拟化的支持，并启用安全分区管理器。安全分区管理器支持安全的空间分区，为分区提供所需的安全访问，同时将它们彼此隔离。

4) Armv9 CCA机密计算架构

为了解决保护数据的技术挑战，Armv9架构引入了Arm机密计算架构（CCA）。机密计算通过在基于硬件的安全环境中执行计算，使部分代码和数据在使用时不受非法访问或修改，甚至不受特权软件的影响。

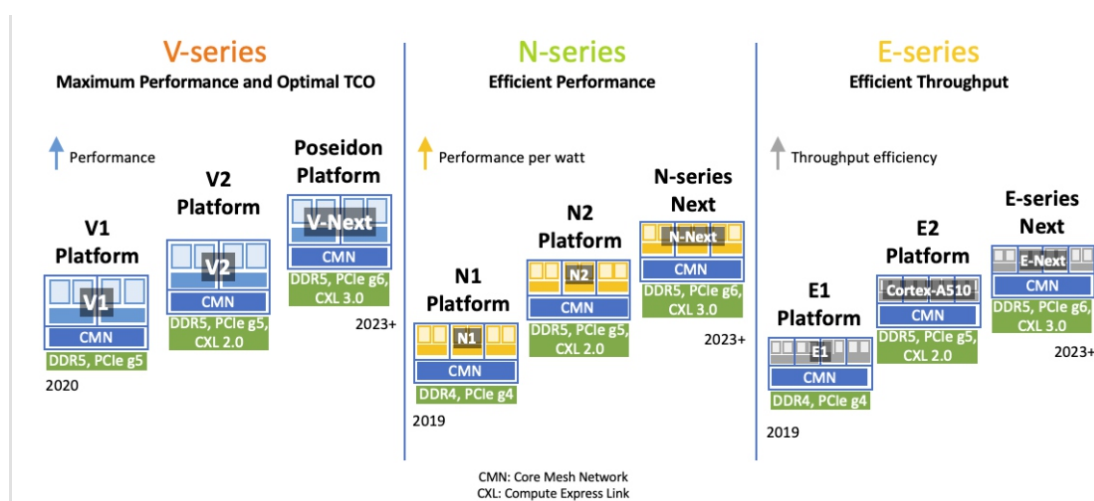
Armv8架构基于TrustZone TEE实现机密计算面临两个方面挑战，一是无法隔离安全世界TEE，使得可信应用必须信任设备，二是不支持动态分配安全内存。Armv9 CCA将引入动态创建域的概念，所有应用程序都可以使用这个概念，这个域与安全世界和非安全世界都是隔离的，而且可以动态分配域使用的内存。例如，在业务应用程序中，域可以在使用、静态和传输过程中保护商业敏感数据和代码不受系统其他部分的影响。



图表 10 从Armv8-A TrustZone TEE 到 Armv9-A RME机密计算架构

1.3.3 Arm Neoverse产品路标

Arm 在整个基础设施市场中正快速迭代创新，其路线图包括应用于云、高性能计算（HPC）和人工智能（AI）领域的追求极致性能、优异的总拥有成本 V 系列；应用于云、5G、网络和边缘领域，注重平衡性、能效比N 系列；以及应用于 5G、网络和基础设施边缘领域，主攻高能效的 E 系列。



图表 11 Arm Neoverse平台路标

1、Neoverse V系列

Neoverse V2提供了市场领先的单线程性能，V2平台采用最新的64位的Armv9-A指令集和广泛部署的CMN-700 mesh互连技术，带宽高达4TB/s，同时系统末级缓存容量最高512MB，是上代的四倍。增加了2MB的专用L2缓存，比V1平台翻一倍；采用SVE2可以帮助满足更多非HPC ML类型的工作负载，将矢量引擎重构为4通道的128位，并对微架构进行了调整，以提高其有效吞吐量。V2之后的下一代V系列产品，代号Poseidon平台，将会加入PCIe 6.0、CXL 3.0总线，支持RME的机密计算安全架构。

2、Neoverse N

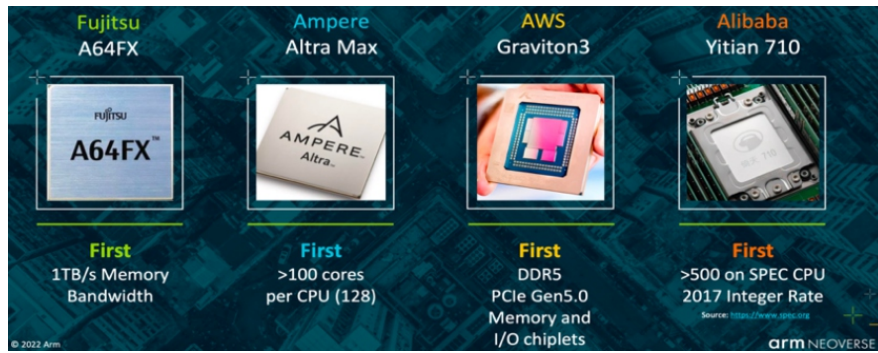
N系列方面继续保持市场领先的性能功耗比，已有Intel, Marvel等多家厂商发布了基于N系列的DPU产品。Neoverse N2是N系列第一个基于Armv9架构的平台，在安全性、能耗以及性能方面都有全面的提升。相比于N1，N2的单线程性能提升了40%。新一代N系列产品将在性能和效率方面实现代际提升。

3、Neoverse E

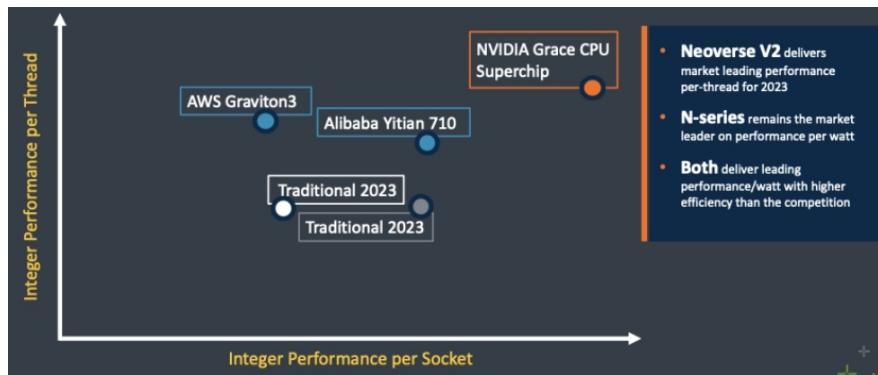
Neoverse E系列正广泛应用于数据平面处理、5G RAN、边缘网络和加速器等领域。Neoverse E2平台，结合Arm Cortex-A510 CPU与可扩展的Neoverse CMN-700，条件相对受限的应用也能充分利用诸如可扩展的核心数量范围、Arm System Ready兼容性等云技术。

1.3.4芯片产品进展

近年来，Arm 及合作伙伴一直在推动整个行业的多项新创新。推动了许多与基础设施相关的行业第一，包括第一个实现每秒 TB 级内存带宽芯片、第一个超过 100 核芯片、第一个实现 DDR5 和 PCIe Gen 5、第一个实现 LPDDR 5 和第一个 DPU都基于Arm实现。

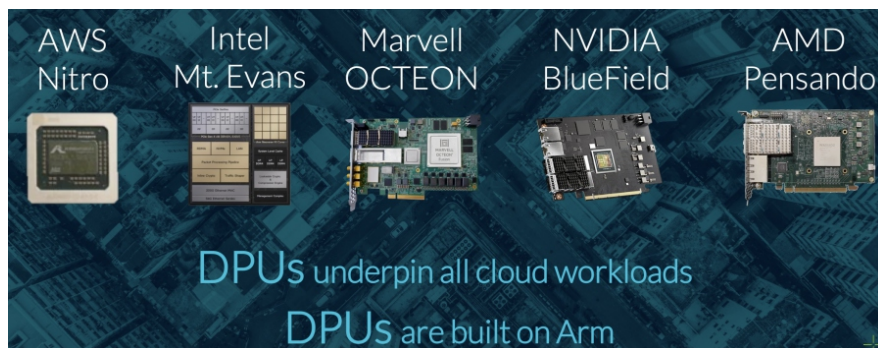


图表 12 基于Arm的基础设施芯片创新



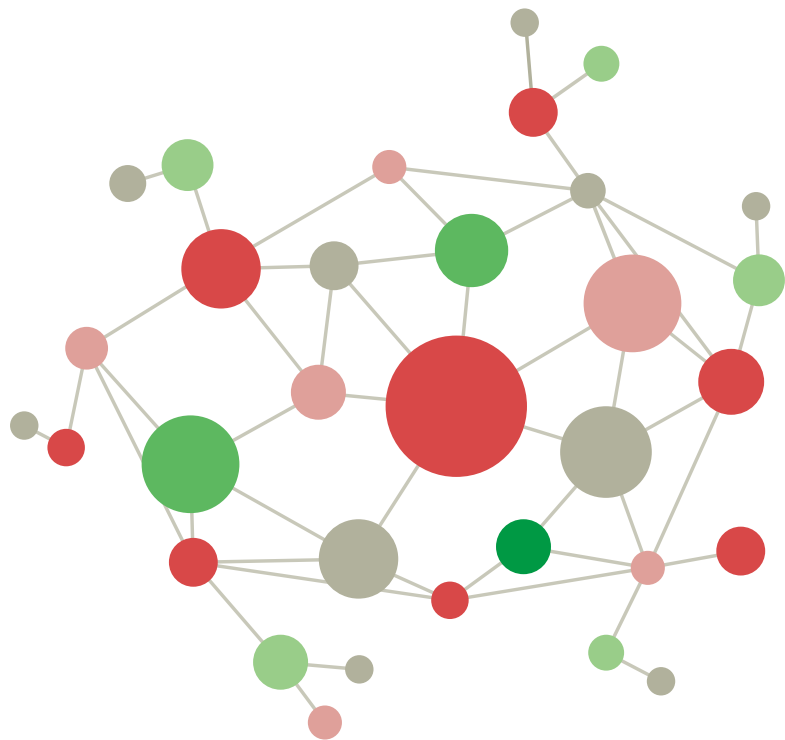
图表 13 基于Arm的基础设施芯片性能提升

除了 CPU 之外，全球主要的超大规模数据中心供应商还利用 基于Arm DPU 来重建其基础设施。目前市场上的多种 DPU 均基于 Neoverse 平台；包括 AWS Nitro、Marvel OCTEON、Intel Mount Evans 和 AMD Pensando。



图表 14 基于Arm的DPU芯片

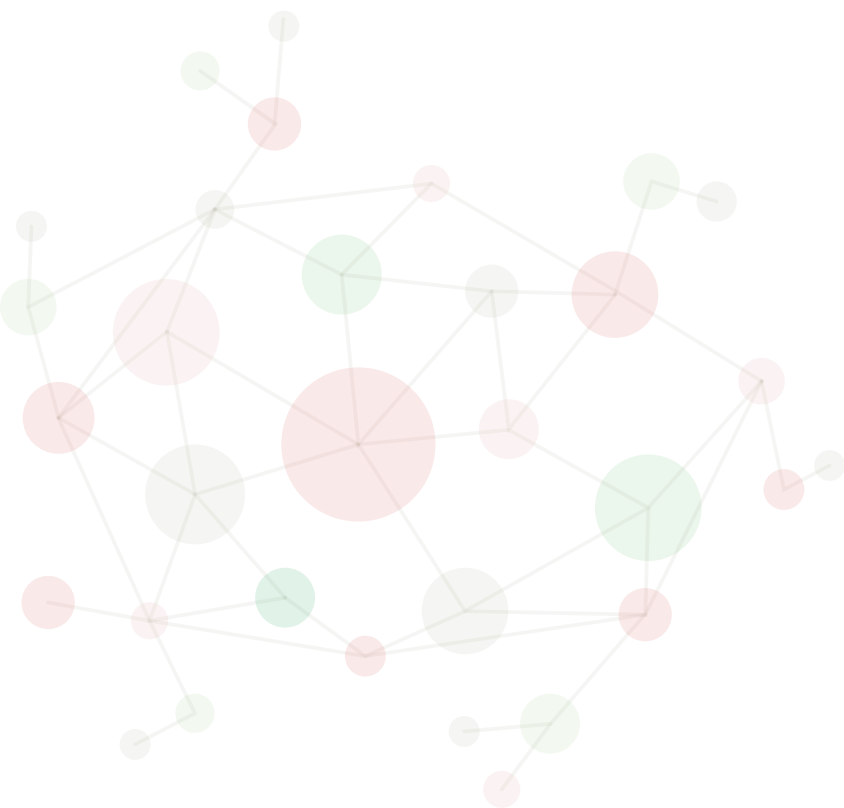
此外，Arm、软银和 NVIDIA 之间的最新合作旨在重塑 5G 数据中心。作为合作的一部分，Grace Hopper Superchip 正在为人工智能数据中心提供支持，该数据中心旨在提高投资回报率 (ROI)，同时为长期存在的问题提供解决方案，包括 Open RAN 面临的挑战以及边缘人工智能支持的承诺 5G 和物联网。



第二章

CHAPTER 2

绿色计算处理器和固件生态



2.1 综述

随着大数据、物联网、人工智能等技术快速发展，数据呈现爆炸式增长的态势，这也对于云端的计算力提出了更高的要求，推动了全球服务器及服务器CPU市场的持续增长。

曾几何时，Intel在全球服务器CPU市场可谓是一家独大，但随着摩尔定律的放缓，以及云计算、端云协同对于计算架构需求的变化，云端计算架构体系开始重构，Arm架构在服务器CPU市场渐成气候。

Arm架构跨越性能鸿沟，Arm阵营全方位推进：原有数据中心服务器算力的提升，主要是依赖于摩尔定律推动之下的CPU的半导体制程工艺的提升和CPU内核架构的提升。2018年，Arm推出了针对数据中心服务器及边缘服务器市场的全新Neoverse平台，包括Neoverse V/N/E系列，并确认每一代Neoverse平台都将带来30%的性能提升。随后众多企业推出了基于Neoverse的服务器CPU，例如亚马逊已经发布了三代Graviton处理器，并大范围运用到亚马逊的云中；Nvidia则是发布了应用于超算和AI场景的Grace CPU；微软、谷歌也正在研发基于相应的服务器CPU。

云计算、移动互联网加持，推动Arm软件生态发展：随着IT架构“云化”转型的进一步深入，Arm的众核并行计算优势开始凸显。在此背景之下，不少软件厂商开始基于成熟的开源软件打造不依赖于x86架构的商业化解决方案。同时，随着移动互联网、5G技术和边缘计算的发展，推动了算力空间开始发生结构性变化，云端计算的协同发展是大趋势。目前移动终端和IoT市场超过90%的CPU是基于Arm架构，终端侧Arm的软件生态极其丰富，这也使得云端侧的软件开发开始兼顾Arm生态，支持Arm相关生态的底层系统和上层应用也逐渐增加，比如三大Linux发行版（Ubuntu、Suse和Redhat）都已经直接有支持Arm Server的版本。据相关数据统计，今年一季度Arm在全球服务器市场的出货量份额已经达到了10%，其中40%出货是在中国市场。虽然目前大多数服务器仍然是基于x86处理器，但越来越多的应用开始运行在基于Arm架构的服务器CPU上。

2.2 处理器代表企业

2.2.1 华为

1、鲲鹏计算介绍

数字经济时代，ICT技术的蓬勃发展加速计算应用创新多元化的应用场景和数据类型，驱动计算架构向多样性发展。其中，以鲲鹏为代表的基于Arm架构的计算平台，因其多核、低功耗、端云同构等特点在场景多样化计算时代具备显著优势。目前，越来越多的行业如电信、金融、政务、能源等，均选择将应用部署在鲲鹏平台之上。

2019年华为正式发布鲲鹏计算产业。鲲鹏计算产业是基于鲲鹏处理器的基础硬件设施、行业应用及服务，涵盖从底层硬件、基础软件到上层行业应用的全产业链条。鲲鹏计算产业是基于鲲鹏计算平台的基础硬件设施、行业应用及服务，涵盖从底层硬件、基础软件到上层行业应用的全产业链条。华为向业界开放鲲鹏全栈能力，包括进一步开放鲲鹏主板，使能伙伴差异化创新，推出场景适配性更好的整机产品；开源操作系统openEuler、企业级开源数据库openGauss、使能合作伙伴推出商业发行版，繁荣基础软件生态；开放鲲鹏应用使能套件（鲲鹏BoostKit，英文名：Kunpeng BoostKit）和鲲鹏开发套件（鲲鹏DevKit，英文名：Kunpeng DevKit），使能伙伴和开发者快速创新行业应用。

硬件方面，鲲鹏主板走向更加开放的局面，华为自身仅聚焦“CPU+内存”最小计算单元，通过天池架构全量组件化创新，实现从使能伙伴创新走向伙伴主导创新的转变。

基础软件方面，开源操作系统openEuler，开创性的提出全场景操作系统理念，通过全栈原子化解耦，实现版本灵活构建、服务自由组合。通过一套操作系统架构，实现了对服务器、云计算、边缘计算和嵌入式等场景的支持。企业级开源数据库openGauss5.0在2023年3月份上线，架构内核持续创新，支撑企业核心业务。

应用软件方面，提供鲲鹏DevKit和鲲鹏BoostKit两个使能套件，使能极简开发、极致性能。持续升级鲲鹏开发工具DevKit，在原有迁移、原生开发能力持续增强基础上，新增提供组件化、服务化能力，让开发更便捷。2023年发布鲲鹏DevKit23.0，通过工具组件化，使得开发者以前需要手动提交适配之后的代码到开发流水线，到现在，可以将DevKit按需集成到现有DevOps流水线中，每天自动扫描、修改、提交代码，非常便捷的发布版本；而且，还可以一键开通云开发服务，随时随地迁移、开发。

性能优化上，发布鲲鹏BoostKit23.0，升级“5大”场景化（大数据场景、HPC场景、机密计算场景、数据库场景、分布式存储场景）能力，同时持续贡献、迭代基础软件openEuler和openGauss内核能力，通过软硬协同，充分释放鲲鹏算力，使能行业应用极致性能。

2、生态进展

目前，已有11家合作伙伴推出基于鲲鹏主板的服务器产品，2022年整机伙伴发货占比达95%；已有4500多个合作伙伴，13500多个解决方案通过鲲鹏兼容性测试认证，在金融、电力、政府、运营商等行业全面落地；国内外主流操作系统伙伴均已发布openEuler商业发行版本，数十家数据库伙伴均发布openGauss商业发行版本；在全国建设24个鲲鹏生态创新中心；发展超过200万鲲鹏开发者，面向硬件、软件和应用全栈创新。鲲鹏当前已广泛服务与政府、金融、运营商、电力等行业，成为行业的首选技术路线。

2.2.2 飞腾

1、OurBMC社区介绍

BMC全称为基板管理控制器（Baseboard Management Controller），是服务器的管理控制单元，是用于监控和管理服务器的专用控制器，有设备信息管理、服务器状态监控管理、服务器远程控制管理、维护管理等功能。BMC以其多功能、易操作、广兼容、更底层的应用优势愈加受到各行业的青睐，成为服务器基础生态的重要组成部分。

飞腾联合昆仑太科、百敖等行业头部企业发起成立BMC开源技术根社区OurBMC，旨在推进BMC技术快速发展，辐射上下游形成产业共振，加速构建繁荣的信息系统软硬件生态。OurBMC社区成立大会暨理事会第一次会议于2023年7月12日在北京成功举行，标志着中国首个BMC根社区——OurBMC社区正式成立。OurBMC社区将充分发挥各成员在BMC领域的技术专长，以及多年的开源社区的开发和贡献经验，秉持开放、平等、协作、创新的原则，将社区建设为由基础软硬件企业、第三方机构、高等院校、个人开发者等共同参与贡献的社区。社区将致力于为BMC开发者提供国内首个开源BMC技术平台，涵盖BMC软件、系统内核、启动固件、硬件等BMC全栈软硬件，用于使用、交流和创新BMC技术，立足开放合作和技术创新，共同推进BMC技术和相关软硬件生态的繁荣发展。

飞腾基于自研芯片飞腾腾珑E2000实现了对OpenBMC的适配，为客户提供全方位的BMC软硬件解决方案，并开放E2000 BMC板卡的DDR4及Gen-Z硬件接口规范、硬件图的参考设计；提供可视化的固件配置工具，且支持安全启动。目前已实现共飞腾一期首批软硬件产品落地，100%国产化设计，飞腾服务器配套飞腾BMC，兼容性更强；重点完成了redfish、IPMI、KVM、UOL等关键核心功能开发与验证；软件方面从uboot、kernel、filesystem到tools，自底向上开源支持。同时飞腾还积极投身OurBMC社区建设，在社区贡献飞腾E2000 BMC的uboot、内核、板级BSP的全套源码，协助客户实现BMC产品的迭代升级，共建BMC生态。

2、基础算法库PhyLibs介绍

飞腾秉持绿色计算的理念，在持续推进处理器设计革新的同时，也高度重视从硬件算力到软件算能的萃取。目前，以高性能计算、多媒体处理、安全、人工智能等热点领域为牵引，飞腾对信号处理库、线性代数库、基础数学库、密码算法库、图像处理库、AI加速库以及网络加速库等基础软件进行了深度优化，增强底层实现与飞腾处理器硬件架构的亲合性，重构算法设计突破算力效能瓶颈，实现了飞腾高性能基础算法库PhyLibs，核心算能均有显著提升。

PhyLibs信号处理库符合vsipl规范，提供向量运算、矩阵运算、快速傅里叶变换、滤波函数、希尔伯特变换、卷积等高性能计算接口。相比公版vsipl，向量运算模块加速比在1.18-30.38之间、加速比2.0以上函数占比55.2%，矩阵运算模块加速比在1.17-508.00之间、加速比2.0以上函数占比56.3%，信号处理模块加速比在1.36-45.021之间、加速比2.0以上函数占比51.3%。

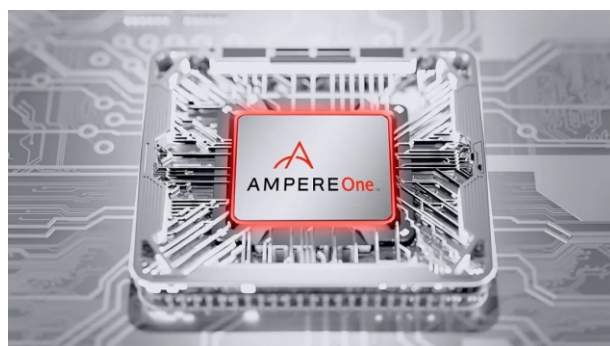
3、生态进展

截至2023年6月底，飞腾的生态伙伴数量已经近5500家，包括集成商合作伙伴300+家、硬件合作伙伴1230+家、软件合作伙伴3900+家。目前，飞腾已联合数千家国内软硬件厂商，支撑了3630多款飞腾平台设备上市，已经和正在适配的软件超过11200种，与近2万款开源软件产品完成了适配，兼容200万级移动APP应用，构建起了国内最完善、最庞大的从端到云的信息化建设全栈生态体系。

2.2.3 安晟培

1、产品介绍 全新AmpereOne 系列处理器

5月，Ampere Computing 宣布推出5nm全新 AmpereOne 系列处理器，该处理器拥有多达192个单线程 Ampere 核，内核数量为业界最高。这是第一款基于 Ampere 新自研核的产品，由 Ampere 自有 IP 全新打造。



图表 15 AmpereOne系列处理器示例

AmpereOne 具备大型云优化私有缓存以及针对 AI 等高增长的云应用的新功能。这些新功能，如网格拥塞管理（Mesh Congestion Management）、细粒度电源管理（Fine Grained Power Management）和内存标签（Memory Tagging）等，为高性能、高利用率的多用户环境（如云）提高了性能一致性、可管理性和安全性。通过增加 8 通道的 DDR5 内存和128 通道的 PCIe Gen5 I/O，整个平台实现了扩展。

这些技术创新为 Ampere 客户的云原生工作负载提供了最高的整体性能、可扩展性和密度。此外，得益于低延迟和高吞吐量，Ampere 云原生处理器在 AI 推理方面较其他 CPU 有 2 倍以上的性能优势。

2、生态进展

Ampere 的客户群包括谷歌云、微软 Azure、甲骨文云、阿里巴巴和腾讯等领先的云服务提供商（CSP），以及 HPE 和 Supermicro 等领先的原始设备制造商（OEM），这也体现了客户对 Ampere 产品组合的信心。

2.2.4 英伟达

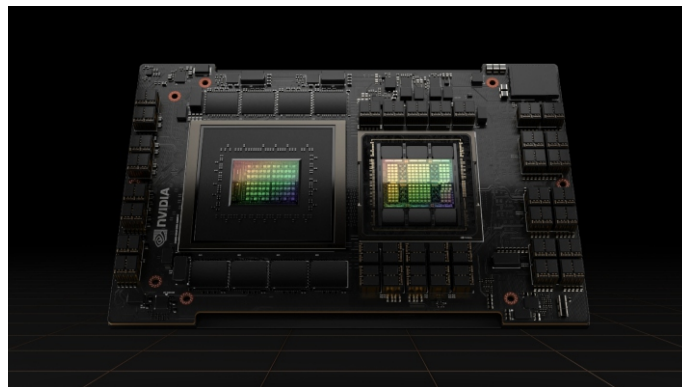
1、产品介绍 NVIDIA Grace Hopper 超级芯片

NVIDIA Grace Hopper 超级芯片是一款专为大规模 AI 和高性能计算应用打造的加速 CPU，可以为运行 TB 级数据的应用提供高达 10 倍的性能。该芯片通过 NVIDIA NVLink-C2C 互连技术将 Grace 和 Hopper 架构相结合，为加速 AI 和 HPC 应用提供 CPU+GPU 相结合的一致内存模型。

技术规格：

- Grace Arm Neoverse V2 CPU
- 高达 480GB 的 LPDDR5X 内存
- 高达 96GB 的 HBM3
- NVLink-C2C: 900GB/s 一致性接口，比 PCIe 5.0 快 7 倍
- 节能 1,000W 模块（CPU, GPU 和内存）
- 可运行完整的 NVIDIA 软件栈和平台，包括 NVIDIA HPC SDK、NVIDIA AI 和 NVIDIA

Omniverse



图表 16 NVIDIA Grace Hopper 超级芯片示例

在今年 SIGGRAPH 大会上，NVIDIA 发布新一代 GH200 Grace Hopper 平台，该平台基于全球首款搭载 HBM3e 处理器的 Grace Hopper 超级芯片，专为加速计算和生成式 AI 时代而构建。新平台专为处理大语言模型、推荐系统、矢量数据库等全球最复杂的生成式 AI 工作负载而构建，将提供多种配置选择。该平台采用双配置——提供的内存容量和带宽比当前产品分别增加了 3.5 倍和 3 倍——包括一个拥有 144 个 Arm Neoverse 内核、8 petaflops 的 AI 性能和 282GB 最新 HBM3e 内存技术的单个服务器。



图表 17 NVIDIA GH200 Grace Hopper™ 平台示例

借助 NVIDIA MGX 服务器，任何系统制造商都可以快速、经济地将 Grace Hopper 超级芯片添加到 100 多款服务器机型中。

NVIDIA Grace CPU 的每瓦性能是传统 x86-64 平台的两倍，是目前最快的 Arm 数据中心 CPU，专为实现高单线程性能、高内存带宽和出色的数据移动功能而设计。该设计实现了带宽、能效、容量和成本的最佳平衡。

2、生态进展 NVIDIA对Arm软件开发者的支持

NVIDIA 为开发者提供包括库、SDK等在内的工具，帮助他们实现云和数据中心、高性能计算、边缘AI和机器人、PC系统和应用。

- NVIDIA HPC SDK 包含针对 Arm CPU 的性能优化的编译器，提供向量化和多核支持。
- CUDA-X 基于 NVIDIA CUDA 构建，是 Arm 系统上经过优化的各库的集合。
- 适用于 Arm 的 CUDA 工具套件提供了开发环境，用于在 Arm 上创建 GPU 加速的高性能应用。
- NVIDIA NGC 是 GPU 优化软件的中心，包含框架容器、带示例代码的演示等，有助于在基于 Arm 的 NVIDIA Jetson 设备上加速开发边缘 AI 应用。

2.2.5 阿里云

1、产品介绍 倚天处理器

在2021杭州·云栖大会上，阿里巴巴发布首款通用芯片倚天710，这是一款为云而生的芯片，针对云计算的特点做了大量优化。

架构层面，倚天710采用最新Armv9架构，多达128核，主频最高3.2GHz，可同时兼顾性能和功耗。同时，集成了业界最领先的DDR5、PCIE5.0等技术，能有效提升芯片的传输速率，并且可适配云的不同应用场景。



图表 18 倚天710云原生处理器芯片

2、产品介绍 飞天云计算操作系统

云计算操作系统作为面向互联网上应用的核心基础软件，向下管理数据中心资源，向上提供公共服务及编程接口。飞天（APsara）是阿里云自主研发、服务全球的超大规模的云计算操作系统，突破关键核心技术，达到国际先进水平，实现按需+高效输出部署，降低架构复杂性和保证架构稳定性，同时实现了多元异构数据同时在线，具备高可靠、高可用、高性能和大规模的应用特点。



图表 19 飞天云计算操作系统

3、标杆案例 绿色数据中心

阿里云一直重视环保和可持续发展，将绿色计算作为业务发展重点考虑之一，提出基础设施碳中和战役，加强节能减排。阿里云采取了多种措施来降低能源消耗和减少碳排放，例如优化数据中心基础设施、采用可再生能源等

- 加强节能减排：采取了多种措施来降低能源消耗和减少碳排放，例如优化数据中心基础设施、使用液冷技术降低PUE等。
- 推广绿色技术：积极推广可持续发展的技术解决方案，例如采用可再生能源、高效节能的服务器等。
- 推进数据中心可持续发展：通过加强数据中心能源管理、优化温室气体排放等措施，推动数据中心的可持续发展。
- 促进数据安全和隐私保护：在数据中心中加强数据安全和隐私保护，保护客户数据的安全和隐私。



图表 20 阿里云绿色数据中心

2.2.6 遇贤微电子

1、公司及产品介绍

遇贤微电子成立于2020年,致力于为云计算数据中心提供芯片、平台和系统，目前是LF Edge、OpenEuler、GCC等成员。第一代产品规格：160核数据中心CPU，3.2GHz和620+ SPECint2017。

高性能数据中心CPU产品简介：

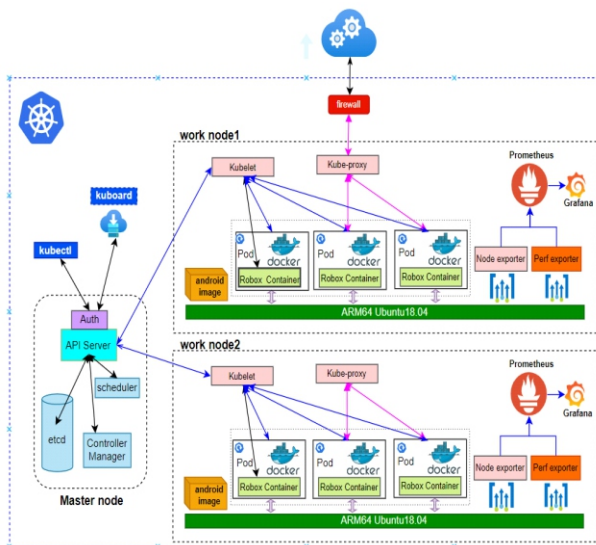
- 关键特性：
 - CPU:基于Arm Neoverse N2 架构，比Neoverse N1架构提升 35~45%。核心功耗比前两代芯片节省30%；支持2x128b SVE/NEON、MPAM、虚拟化扩展等技术；
 - 核心：单片160 N2核心，支持cache一致性片间CCIX/CXL互联，核心支持CMN互联；
 - 总线：支持经过优化的高速片间互联，实现SOC片间核间访问性能提速；
 - 外设：支持PCIe5.0，支持DDR5控制器；
 - 功耗：约为 200 ~ 300W；支持DVFS；
 - 算力：SPECint 2017 160 核 620分；
 - 安全：支持Secure EL2；
 - RAS: V8.2+ RAS技术。

2、生态建设

多核高性能Arm服务器助力云原生&边缘计算云游戏

遇贤微电子和Linux基金会Akraio社区合作基于Arm架构的云原生平台并联合发布《基于Arm高性能CPU的边缘计算云游戏白皮书》，该平台支持Android云游戏和 VR/AR 等 Arm 应用，主要有以下特点：

- 计算的分布式部署提高整体并发性
- 云服务架构带来运营及维护的高效率



Akraio White Paper Edge Cloud Game based on Arm architecture high performance CPU

<1st edition - November 2022>



图表 21 《基于Arm高性能CPU的边缘计算云游戏白皮书》

基于算力服务原生CPU服务平台

遇贤微电子和行业伙伴支持中国联通对算力服务原生的定义内涵、总体架构、关键技术进行了分析，通过发展算力服务原生技术，实现“一套代码全网通用”的目标，合作共建、联合推动算力服务原生统一的标准体系。遇贤微电子将推出超多核（160核）高性能CPU，支撑运营商进行算力网络研究与建设。与业界合作共建、联合推动算力服务原生统一的标准体系，繁荣开源生态。

该方案主要有以下特点：

AI场景加速计算方面：遇贤CPU支持众多人工智能与机器学习硬件加速特性，如SVE2可变长向量处理、BF16浮点数等。

异构算力下统一软硬件生态方面：引入抽象层级，如对芯片封装、PINMAP等物理形态进行抽象，实现同一架构多个芯片能够支持单板层面兼容。

算力服务原生的安全可信方面：采用以硬件保护软件的新型安全防护技术，提供硬件主动防御，弥补以往“用软件保护软件”的软肋。



图表 22 《中国联通算力服务原生白皮书》

2.2.7 鸿钧微电子

1、公司及产品介绍

杭州鸿钧微电子科技有限公司创立于2021年，致力于开发基于Arm架构的面向数据中心的通用CPU，为业界提供“更高效能、更易部署”的服务器CPU和系统解决方案，推动全球数据中心的变革！

公司开发中的服务器CPU拥有超过80个CPU内核，并支持DDR5以及PCI Gen5高速外设接口，产品预计于2024年上市。

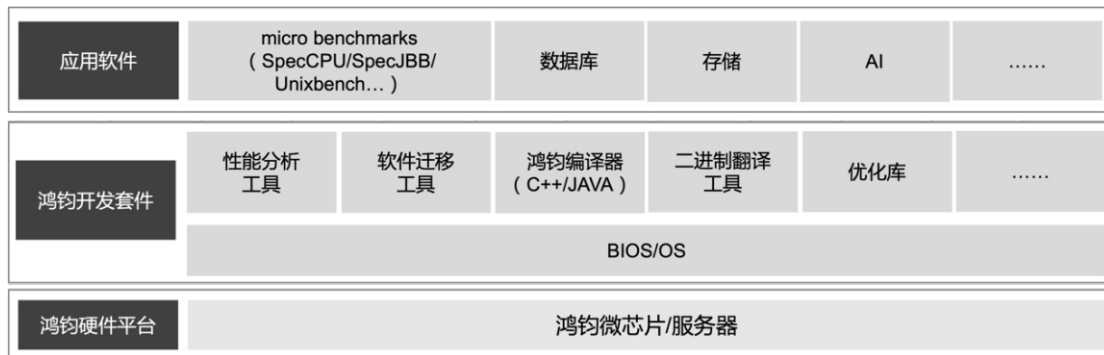
2、生态建设

鸿钧微电子将提供编译器工具链的优化工具，通过编译器优化提升实际应用环境的软件性能，为用户的应用部署提供协助。主要包括高性能编译增强：通过内存优化增强，自动向量化等，提升指令和数据吞吐量；加速指令集：优选NEON/SVE等指令，运行时库优化，发挥芯片算力。



图表 23 编译器工具链优化工具

鸿钧微电子还将提供基础软件栈的适配与优化套件，包括基础软件增强套件，例如针对OS/BIOS/BMC基本功能和增强功能，提高可靠性和运维能力；同时提供软件升级和适配工具套件，主要包括各类软件迁移工具和性能分析工具集，支持适配高版本基础软件，获取更高性能。



图表 24 基础软件开发套件

3、生态应用——MySQL在Arm架构平台上的优化

MySQL是一款开源的关系型数据库管理系统。它被广泛用于各类业务场景中，是最受欢迎的数据库之一。MySQL具有可靠性高、性能强、可扩展性好、安全性好等特点。为充分发挥Arm服务器CPU的优势，公司团队开展了MySQL在Arm平台上的优化工作。

优化方法论：业界对于MySQL性能的评价指标认知统一，即吞吐量和延迟。因此，性能优化工作的重点均落在提高系统单位时间吞吐量和降低延迟两方面。而优化过程则遵循“应用层→系统层→微架构层”的顺序，逐层完成工作负载的特征刻画与瓶颈定位，进而开展优化工作。

优化工作介绍：针对应用层将重点围绕三个方面给出优化方案，分别包括典型场景识别、重点代码实现理解以及社区调研；针对系统层的优化重点则是分析资源的使用，包括针对MySQL Workload性能瓶颈分析，重点监控包括CPU使用率、内存使用率等的硬件指标，通过分析这些指标定位到资源瓶颈，然后结合MySQL的业务特点进行优化；针对微架构层，可以参考Arm官方文档，明确Arm核心有关的微架构设计以及相关性能指标，明确微架构性能瓶颈的定位和优化方法。

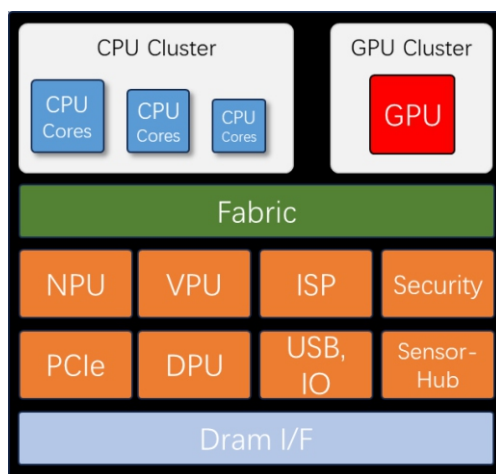
优化结论：基于Arm平台，分别针对应用层、系统层和微架构层开展优化，每个优化措施均能带来一定的效率提升，结合具体的应用场景，有的能有着50%的性能提升。而具体可选择的优化措施，则更多依赖于用户的实际使用场景。

2.2.8 此芯科技

1、公司及产品介绍

此芯科技是一家专注于开发智能SoC芯片以及高效算力解决方案的科技企业。公司于2021年10月成立，由国内外知名芯片和IT企业的核心技术和管理人员创立，公司拥有业界资深的智能计算架构和全建制研发设计团队，在CPU内核研发，SoC，全栈软件开发和系统设计等领域具备雄厚的技术积累，致力于开发兼容Arm指令集的高能效计算解决方案，为智能芯片2.0时代贡献力量。

此芯科技的第一颗在研高效能Arm SoC采用高达两位数的CPU内核架构，内置高性能GPU并可通过高速PCIe总线外接独立显卡，满足PC用户图形计算要求。该SoC除了支持PC级高速存储接口之外，还采用高带宽和高容量内存总线，不仅可以支持CPU和GPU计算所要求的高速内存数据交换，也可以满足内置高算力NPU在端侧大模型部署中对大容量内存数据的读写要求。此芯科技的首颗SoC，支持Arm V9最新安全特性并集成自研安全引擎，满足国密算法和安全需求。基于该芯片的Arm PC方案在软件架构，系统设计和生态建设上，也全面支持Arm Pc的技术要求和发展趋势。



图表 25 此芯科技首颗SoC架构框图

2、生态进展

此芯科技自去年加入Linaro Windows工作组之后，发起成立了Client PC合作项目，旨在推动基于UEFI + ACPI标准的Arm PC启动架构标准化，通过统一的系统固件支持Windows和Linux等多种操作系统。这将有助于整机厂商快速推出Arm PC整机产品,也是为Windows on Arm生态的推进发挥关键的作用。

今年春季， Arm软件生态全球峰会Linaro Connect在英国伦敦举办，此芯科技联合Arm和Linaro的技术专家在峰会上发表了主题为“ACPI for Arm Client PC”的技术演讲，对Linaro Client PC项目的工作进行了总结，并与来自全球不同厂商的开发者共同探讨如何实现针对Arm PC的ACPI标准化支持。Linaro CTO Grant Likely在其主题演讲中也介绍了由此芯科技发起成立并重点参与的Client PC项目，强调了通过统一的固件来支持Windows和Linux等多种操作系统对于Arm Pc生态发展的重要性。

另外，此芯科技携手安谋科技和国产操作系统厂商共同推动Arm PC生态建设。此芯科技已加入 deepin 深度社区和openKylin社区，联合社区生态伙伴共同推动Arm PC的软硬件标准化，加速Linux桌面图形系统和AI软件框架的技术发展，旨在打造Linux桌面操作系统和Arm架构高能效智能芯片融合发展的创新生态。

2.3固件代表企业

2.3.1 昆仑太科

1、产品介绍 新一代昆仑智能绿色固件产品

新一代昆仑智能绿色固件系列产品，包括新一代昆仑BIOS固件、昆仑BMC固件、昆仑卫士终端运维管理系统、昆仑卫士数据中心运维管理系统四大系列产品，将针对通用计算和智能计算等多样化算力需求，突破智能计算、绿色节能、统一通信等关键技术，适用于通用终端、服务器、边缘计算、特种计算等各类型计算设备，支撑构建更加完善的绿色计算设备开发生态、集成生态和用户生态。新一代昆仑智能绿色固件系列产品如图所示。



图表 26 新一代昆仑智能绿色固件系列产品

新一代昆仑BIOS固件，具有异构硬件平台统一支撑、操作系统统一接口，支持安全启动、可信计算、国密算法，支持各类主流外设板卡在终端、服务器、智能计算设备中适配应用，提供BIOS固件层服务能力。新一代昆仑BIOS固件可高效解决计算设备硬件平台、板卡、系统兼容适配复杂、设备形成能力周期长等痛点问题，为客户提供高效率、低成本BIOS固件解决方案。

新一代昆仑BMC固件，突破设备智能故障预测、安全可信启动、跨架构BMC硬件支撑等核心技术；基于硬件/软件加解密算法，可根据应用需求提供不同等级的密码计算和安全验证机制，提升服务器系统整体安全能力；可适配多类型BMC硬件，构建高安全BMC解决方案和商用BMC解决方案。昆仑BMC固件全面满足包括通用服务器、边缘计算服务器和特种服务器等在内的多样化算力服务器对于带外管理应用需求，广泛应用于金融、通信、能源、互联网等重点行业领域中。

昆仑卫士终端运维管理系统，基于终端设备BIOS固件持久化守护技术，可支持信创终端计算设备和商用计算终端设备，实现跨硬件平台、不依赖操作系统运行的终端设备集中维护管控和安全管理，为各类型终端计算设备提供自设备底层到应用层的可见、可控、可恢复能力。昆仑卫士终端管理系统，解决集团用户海量终端计算设备运维保障和安全管理实现复杂、操作困难、维护成本高等痛点问题，保障复杂网络环境下海量计算终端稳定可靠、安全受控运行，目前已经在教育、医疗、云终端租赁等行业领域取得了规模应用。

昆仑卫士数据中心运维管理系统，采用包括Redfish、IPMI、SNMP、RESTful API、Web等行业标准协议和接口，实现与设备平台、供应商无关的统一数据中心带外管理解决方案。昆仑卫士数据中心管理系统不仅可提供数据中心设备和设施实时运行健康状况监控和警报，同时具有能耗智能管理控制、智能容量规划部署、设备负载自动分配等能力，解决大中型数据中心大规模设备/设施设备降本节能、绿色减碳、稳定可靠运行现实急需和痛点，并实现数据中心部署容量灵活可扩展，增容增效。目前，昆仑卫士数据中心管理系统已在金融、能源、电力等行业领域得到应用。

2、生态进展

2022年，昆仑太科公司与龙芯、飞腾、申威、海光、兆芯、麒麟、普华等数十家产业生态合作伙伴，成立了OpenKunlun开源固件社区，本着“开放共享、技术自主，丰富生态、助力产业”的理念，致力于中国固件技术生态自主演进发展，支撑中国计算设备产业健康快速发展。2023年，昆仑太科公司与飞腾、百敖、浪潮等合作伙伴，共同成立了中国首个BMC根社区——OurBMC社区，旨在推进中国BMC技术快速发展，加速构建自主、安全、全面的BMC技术链路。昆仑太科公司还是OpenKylin社区、龙蜥社区等开源社区成员单位，积极为中国计算产业生态贡献自己力量。

同时，昆仑太科积极与包括显卡、网卡、内存、可信模块等板卡/部件厂商技术合作，加强昆仑固件产品基础适配能力，助力各厂商加强其生态适配能力，更高效的开拓市场、取得效益；与各大整机、主板厂商进行深入技术生态合作，为各类终端、服务器计算设备提供包括固件和设备管理整体解决方案，支撑厂商更好的与客户应用需求结合，提升用户接受度，扩大产品销售量。

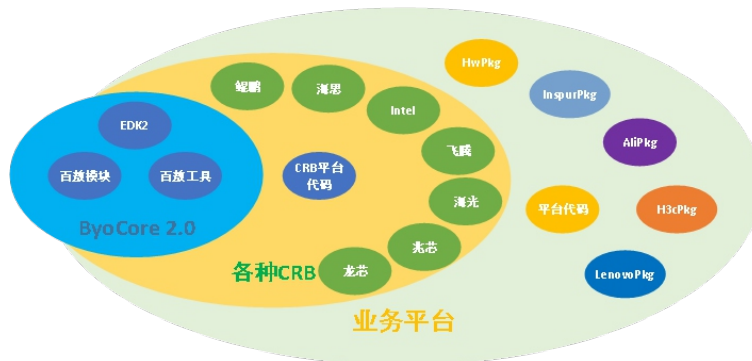
2.3.2 百敖软件

1、产品介绍 固件核心框架及产品ByoCore 2.0

基于多年的固件开发积累，南京百敖软件发布了固件核心框架及产品ByoCore 2.0，产品支持各种主流CPU架构，并在众多信创整机和国际通用整机中得到了落地。

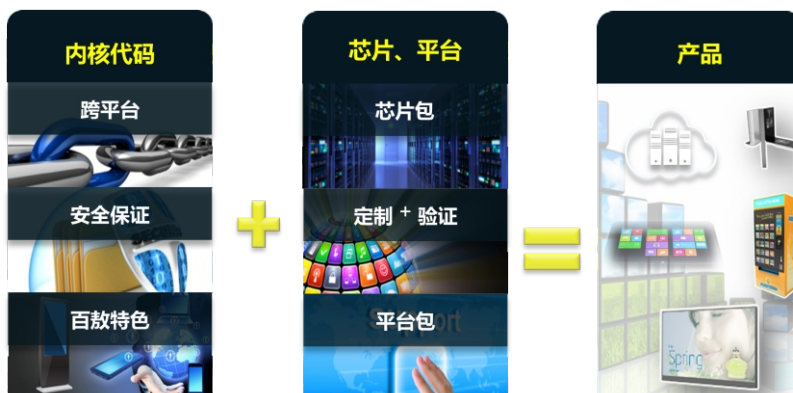
该产品在符合最新国际国内标准的同时汇集多个百敖特色功能模块，在一套代码架构下支持所有国内与国际主流处理器的不同形态整机。产品支持海光、兆芯、飞腾、海思/鲲鹏、龙芯、赛昉和英特尔、AMD等厂家的全系列CPU，形成全品类优势，全面提升通用性，降低开发难度，缩减开发周期。

百敖ByoCore 2.0创新性地采用“分仓+清单”式的代码管理方法，“积木+云端”式的固件开发模式，将先进的大型系统软件研发管理流程引入到固件开发领域，形成固件开发框架。



图表 27 百敖ByoCore 2.0核心框架

在此固件框架平台的基础上，兼容最新国际标准和国内标准的同时，汇集多个百敖特色功能模块（如ByoSafe、特色RAS功能等），整合国密算法、国密签名、国密信任链重建等重点安全需求，向上打通了芯片，向下链接统信和麒麟等国产操作系统，完成安全闭环，从而用一套代码框架和工具链支持所有主流CPU，以及其下游整机产品，并支持各类定制化需求功能。



图表 28 百敖ByoCore 2.0采用积木式开发

2、生态进展

2021年，百敖召开ByoCore 2.0发布会，邀请了浪潮、华为、兆芯、同方、飞腾、海光、联想、中兴、小米等数十家国内外知名厂商参加；近两年，百敖积极加入到“龙蜥社区”“欧拉社区”“龙芯生态伙伴计划”“飞腾基础软件联合实验室”“openKylin社区”“OurBMC社区”等社区和平台，开展生态共建，贡献技术力量，落实应用推广。

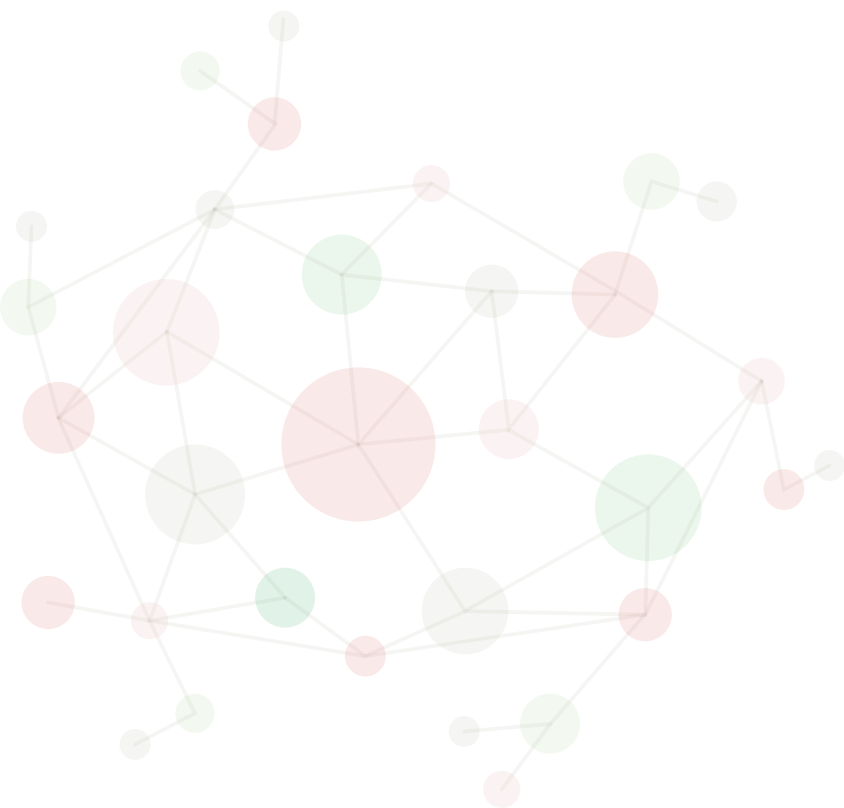
2022年和2023年，百敖ByoCore 2.0全面落地，搭载ByoCore 2.0的固件产品已经全面落地到海光、鲲鹏、兆芯、飞腾、龙芯和赛昉等国内芯片厂商和Intel、AMD等国际厂商相关参考代码和整机产品中，大部分产品已经实现量产上市。



第三章

CHAPTER 3

绿色计算整机生态



3.1 综述

据IDC发布的《2022年中国服务器市场跟踪报告》，2022年中国服务器市场规模为273.4亿美元，同比增长9.1%。中国服务器市场规模在全球占比24.5%。

基于Arm架构的绿色计算产品在需求侧正越来越受到行业用户的欢迎，在中国服务器市场的市场份额快速提升，尤其是运营商、政府和金融等行业。同时，从能用到好用，也逐步扩展到了更广泛的商业领域。

3.2 代表企业

3.2.1 中国长城

1、产品介绍

2U 双路旗舰服务器--擎天RF6260 V5

标准2U双路机架式服务器，支持领先的DDR5、PCIe5.0技术，全面的管理特性开发，满足行业客户核心应用，适用于大数据、云计算、分布式存储、虚拟化、人工智能和数据库等多种应用场景

2U 双路中端服务器--擎天RF5260 V5

标准2U双路机架式服务器，平衡性能、扩展性与经济性均衡设计，覆盖主流应用场景

4U 存储服务器--擎天RF6280 V5

4U双路存储优化服务器，兼顾高存储容量、强大计算性能和极致IO扩展能力，非常适用于温/冷数据存储、视频监控存储、大数据存储、云存储、备份归档等应用场景

6U AI服务器--擎天RF7290H V5

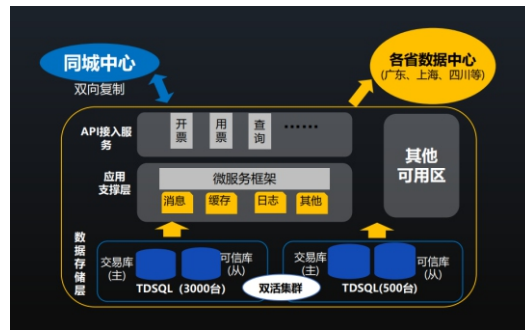
6U机架式高端AI服务器，可应用于AI训练、推理、HPC等场景，支持集群大模型训练。模块化设计，兼容OAM及PCIe标卡，可选风冷及液冷解决方案

2、标杆案例 政府电子发票服务平台

中国长城作为国产服务器的领导者，能够为行业合作伙伴提供端到端的全栈式IT产品交付能力，已助力多个地区、多个行业客户实现国产化转型

政府电子发票服务平台

政府电子发票服务平台服务中国4500万企业，2.8亿非企业用户，是中国覆盖用户最多，规模最大的电子政务系统。系统要求能够有效防范非法攻击，避免未经授权的访问，保障关键信息数据的稳定传输、存储、交换安全，保护纳税人信息安全，确保系统运行安全。硬件和软件系统保证7×24小时在线免费为纳税人提供电子发票申领、开具、交付、查验等服务，一年内系统平稳运行的可用性大于99.99%。



图表 29 政府电子发票服务平台方案架构

在上述安全保密性以及高可用性的严格指标规范下，中国长城提供2U双路服务器作为专有云底座，承载多种业务，通过单品垂直打通、系统优化、深度适配，实现整体系统能力最优，为用户提供最佳性能。

3.2.2 神州数码

1、产品介绍 KunTai D526-2D信创云终端

KunTai D526-2D信创云终端是神州数码专为国家教育信创事业打造的基于飞腾腾锐D2000 8核处理器和国产桌面操作系统、整合主流硬件、兼容主流软件应用，丰富接口、轻巧便捷，人机工学设计，支持多款信创多媒体教学软件，为师生提供国产化软硬件一体的信创教学实训环境，全面助力国家信创人才培养。

基于 Arm 架构，可实现状态实时检测、风扇双重控制，达到系统安全、可靠、稳定运行。自行开发GPU指令转换、前端显示、外设虚拟化、Linux与Windows 应用文件共享、窗口管理、界面融合等组件，实现在国产统信和麒麟操作系统上也能使用Windows应用和外设，具有双系统界面融合、大部分X86应用免适配即可用等优点。



图表 30 KunTai D526-2D信创云终端

2、标杆案例 职业院校信创实训室方案

客户痛点：

产业的变革对职业院校人才培养提出了新的挑战，如何将信创产业的岗位需求、信创专业核心课程、世界职业院校技能大赛和信息技术应用创新赛项中的考核技能点相结合，同时使学生具备考取工信部信创专项认证的能力，是一个亟待解决的问题。

解决方案：

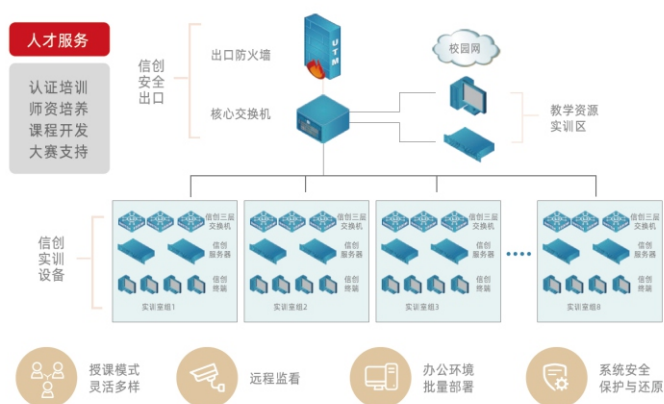
神州数码推出“岗课赛证”融通的信创实训室方案

“岗”，针对信创产业新增的人才需求，如信创系统运维工程师、信创售前工程师。

“课”，根据实际岗位需求，在合作院校计算机大类相关专业，通过新增信创课程完成专业转型升级。

“赛”，世界职业院校技能大赛“信息技术应用创新”赛项主要面向全世界职业院校信息技术类相关专业国内外的学生,打通技术壁垒,实现相互赋能,促进信创产业国际化人才生态布局,加速服务信创装备与相关技术标准建设。

“证”，神州数码与工业和信息化部教育与考试中心联合开展集成适配工程师(初级)、信息系统运维工程师(初级)、信息安全工程师(中级)、信创规划管理师(高级)、信创办公软件应用工程师(中级)的人才评价与培养工作,通过培训考试颁发考试合格证书并规范考试合格证书管理机制。



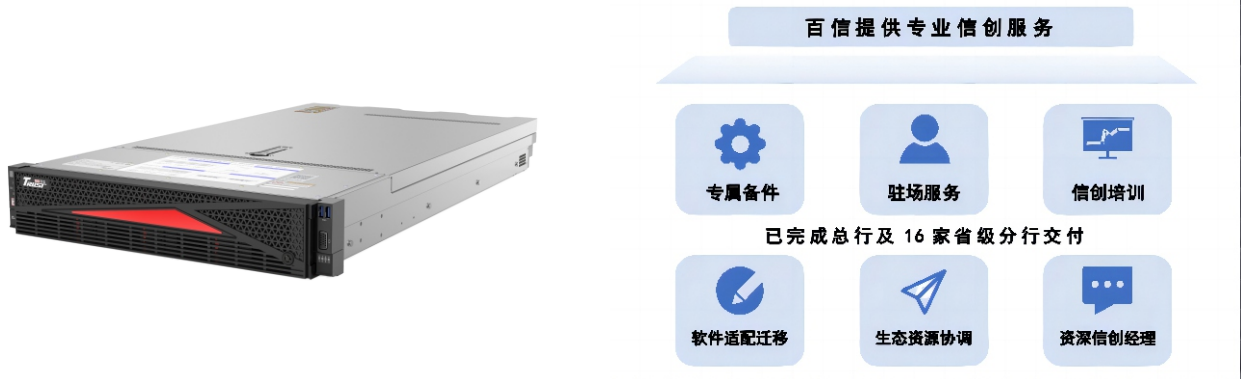
图表 31 信创实训室方案

应用效果：

通过丰富的信创课程资源，功能全面的信创教学平台、安全稳定的信创计算中心和信创终端机房等各方面支撑，使职业院校能够针对信创行业从业需求，培养出适用于信创行业的复合型人才。

3.2.3 百信信息

1、产品介绍 百信恒山TS02F-F30服务器



图表 32 百信恒山TS02F-F30服务器

百信恒山TS02F-F30服务器是一款基于鲲鹏920处理器开发的双路机架服务器，系统最高能够提供128核、3.0GHz主频的计算能力。该服务器采用了国产处理器和操作系统、自主安全BIOS/BMC和自研OCP网卡的设计，有效适配多种国产部件，并对部分设计进行了优化，使得服务器的国产化率进一步提升，性能和可靠性都得到了很好的保障，同时也能够满足客户的多样化需求。

该服务器面向互联网、分布式存储、云计算、大数据、企业业务等领域，具有高性能计算、大容量存储、低能耗、易管理、易部署等优点。经过多项国家级认证，包括CCC、节能、能效和RoHS等，具有节能、环保、安全等特点；同时，百信服务器经过30W小时MTBF测试，质量更优、稳定可靠。在日常办公领域，该服务器已经完全具备替换Windows系统的能力，可以为用户提供高效、稳定和可靠的计算服务。

2、标杆案例 C银行国产化硬件替代

我国银行信息化至今经历了三十余年的发展历程，数字化转型是不可逆转的潮流，银行数字化转型的成功有助于提高银行业务的效率和客户满意度，为银行创造更多的商业机会。

百信为金融行业的国有六大行提供多年的持续耕耘和优质服务。恒山服务器兼容主流云平台，适用于支付、票据、风控等应用场景，为金融客户提供稳定算力，极大地提升了业务效率，较传统方案扩展性更高，便于客户后续扩容、降低成本、提升效率。此外，百信服务器还兼容华为IAAS、腾讯TCE、腾讯TBDS、南大通用MPP数据库、东方通中间件、主流KVM虚拟化软件等应用，为C银行的五大数据中心（北京、上海、合肥、西安、内蒙）以及各省分行提供服务。百信根据客户需求，完成了总行及16家省级分行的交付工作，累计交付超过2000台服务器，实现了C银行国产化硬件替代，同时提升了客户数据使用效率，保障了业务数据的信息安全。

3.2.4 华鲲振宇

1、产品介绍

天擎TG225 B1:

华鲲振宇推出的一款2U双路旗舰级服务器，具备三大方面优势：

- (1) 极致性能：支持2颗鲲鹏920处理器，采用原生全闪设计，CPU直出NVMe链路，支持16块NVMe SSD
- (2) 灵活扩展：支持11个PCIe 4.0标准扩展插槽，规格提升37%，支持10+种Riser卡模组，配置灵活；最大支持9张xPU加速卡(x8)，AI计算性能提升15%；支持内置2块m.2 SSD盘，支持硬件RAID0/1，节省槽位占用，提供更多扩展能力
- (3) 安全可信：支持基于鲲鹏TPCM可信解决方案，满足行业可信计算3.0需求；防篡改、防破解、防安全攻击

天智AT800 (Model 9000) A2:

华鲲振宇推出新一代人工智能训练服务器，产品升级，全面支持FP32算力精度满足更多场景，其特点：

- (1) 整机可提供2.5PFlops FP16算力，为业界同等的1.1倍，可提供0.65PFlops FP32算力，为业界同等的3倍，较上一代芯片提升17.5倍；
- (2) 8颗NPU芯片之间通过HCCS全互联，互联带宽可达392GB/s，持平业界；
- (3) NPU芯片直出200GE网络端口，支持RoCE

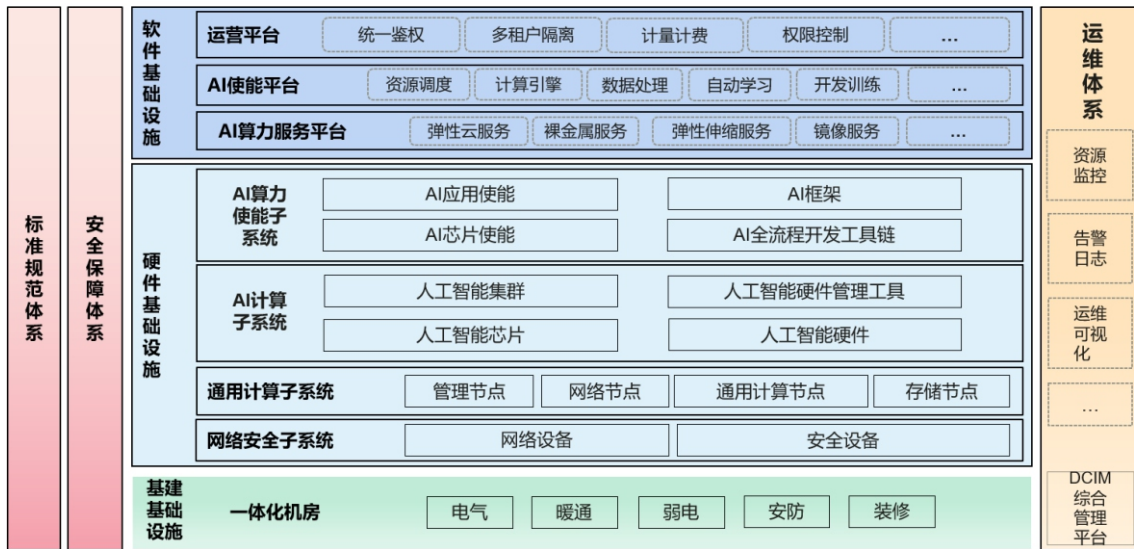
天极HC6000:

华鲲振宇推出的浸没式液冷解决方案，其中国产鲲鹏、昇腾浸没式液冷服务器为全球首发，整体解决方案具备三大方面优势：

- (1) 极致节能：全浸没式液冷，流速精确控制；功耗降低10%~15%，PUE小于1.1
- (2) 极稳运行：液体低介电系数，SI参数优化设计；可长期低温运行，故障率降低50%
- (3) 轻松运维：无风扇设计，噪音相比风冷降低50dB

2、标杆案例

智算存一体化大数据中心解决方案是面向新型数据中心场景的综合性解决方案，着力以智算存一体化大数据中心为依托，解决AI算力资源稀缺且昂贵、AI高端人才缺乏、产学研转化不足等问题。方案以昇腾全栈AI基础软硬件作为底座，选用绿色节能预制化数据中心产品技术和超强人工智能算力集群，通过能基、网络、存储、计算、云平台、算力平台构筑完整的软硬件方案，形成从设计、交付、运营、运维端到端服务体系，提供智算存集成的训推一体化算力服务、计算服务、存储服务，为政府、企业、高校、科研机构等用户建设城市智脑平台、全球领先（E级）人工智能计算平台（普惠AI+科研创新专用）和全球智能数据存储与机器视觉科研创新平台提供支撑。



图表 33 智算存一体化大数据中心解决方案架构

目前，此方案成功应用于成都、福州、青岛、天津、南昌等智算中心项目。其中成都总共投入建设261个机柜、1套人工智能集群1套、120台通用服务器、2套高端全闪存存储、2套海量高密存储2套、179台网络及安全设备，实现首期提供307P人工智能计算能力；采用高效全液冷解决方案，提供60+kW/cabinet的散热和供电能力，PUE低至1.2；并通过智慧化运营，放大算力中心的边际效应，孵化区域典型AI场景解决方案，端到端打通“产”“学”“研”“用”全产业链，发展产业生态，推进产业数字化升级和转型，实现算力利用率超过90%，注册用户90家，累计适配企业伙伴80家，已承接3个科研项目，同时为多个项目提供支撑。

3.2.5 超集信息

1、产品介绍 封闭式高性能液冷机架式服务器

由超集信息自主研发的封闭式高性能液冷机架式服务器——ServMAX® GL414-H3，可实现单路Arm架构处理器及四张全尺寸专业级GPU搭载。在液冷技术的高效散热下，整机满载运行下CPU最高温度为77℃，GPU温度为70℃（环温30℃，12H压力测试），可保证关键芯片高效稳定运行，避免降频风险。

同时，GL414-H3对机房环境的整体依赖度更低，无需进行整个机房的一次侧改造，可在客户现有机房进行直接上架，开机即用，纵享高效算力升级带来的全新体验。面对非标准化机房客户（整体散热、降噪未达到机房标准），亦可实现有效部署。

并且，GL414-H3通过封闭式设计，无需外接CDU，机房空间要求更小，可实现有限机房空间内的更高密度部署，以应对日益提升的算力扩展需求。



图表 34 GL414-H3封闭式高性能液冷机架式服务器

2、标杆案例 上海某大学非标机房液冷升级

上海某大学大数据项目组为实现大数据技术的进一步创新，其需导入更大量数据以进一步优化数据模型。但由于项目组仍沿用着较早期建设的非标准化机房，机房整体空间、散热、降噪能力均有限，关键芯片降频及噪音等问题也是设施扩展中难以避免的重点难题。

考虑到非标准机房的限制条件，超集信息为其提供了基于ServMAX® GL414-H3服务器的液冷解决方案，整机的单路Arm芯片及四张专业级GPU通过液冷散热，有效规避了降频风险，可实现2640 TFLOPS混合算力的稳定输出，构建了CPU+GPU的高效处理流。

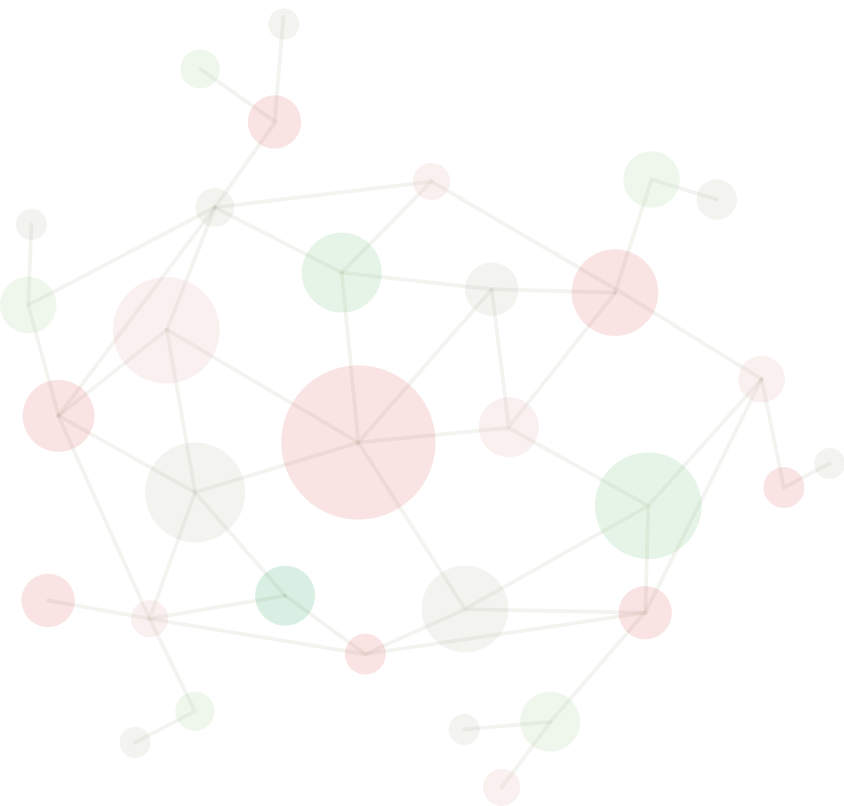
依托搭载Arm架构处理器的超集信息封闭式高性能液冷机架式服务器ServMAX® GL414-H3，客户不仅实现了非标准化机房下的算力高效扩展，同时加速了ETL、数据清洗等数据准备工作，并且和后期的模型训练Job无缝对接，形成整个从raw data到results的pipeline。



第四章

CHAPTER 4

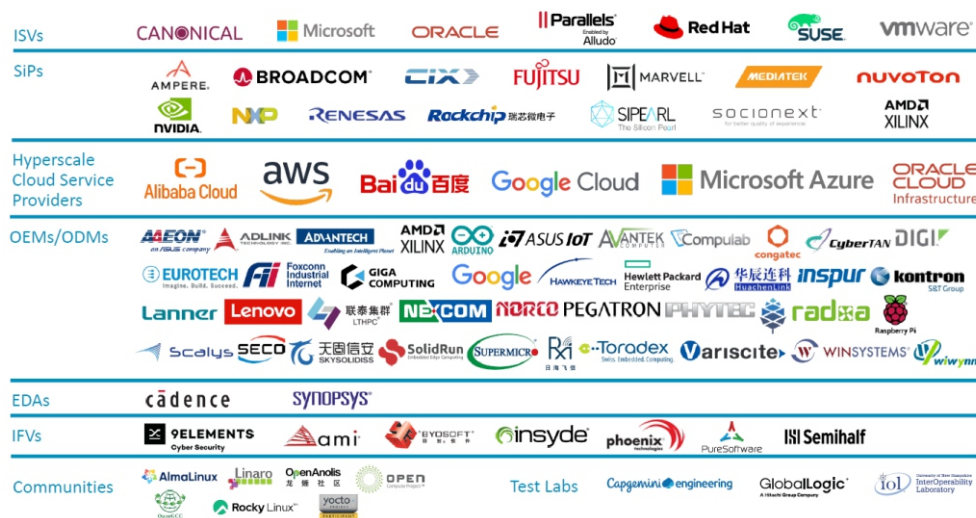
绿色计算软件生态



4.1 综述

Arm 多年来一直专注于软件开发，包括确保对所有主要语言、数据库、工具和操作系统的支持和兼容性。因此，大约有 1500 万开发者在 Arm 上进行开发，并且这一趋势正在迅速加速，尤其是在 Armv9 开发中。例如，NVIDIA 于 2019 年宣布其 CUDA 开始支持基于 Arm 的 AI 和 HPC 应用程序，目前 CUDA 已支持超过 3000 个应用程序。世界上主要的超大规模数据中心已采用 Arm 并建立了用例，同时对软件的持续大量投资确保它们适合客户采用。这种应用类似于滚雪球效应，并且不断增长。

Arm 在 2020 年推出 SystemReady 计划，提供了一套专用于服务器和云数据中心的新系统架构标准，并包含更广泛的操作系统规范和认证，以确保软件的互操作性。Arm SystemReady 计划已完成超过 100 项认证，其中 Microsoft Azure 和 Red Hat 均拥有经过认证的 SystemReady-VE 解决方案。



图表 35 Arm SystemReady 支持者

4.2 操作系统

4.2.1 麒麟软件

1、产品介绍

银河麒麟高级服务器操作系统 V10 SP3 具备内生安全、云原生支持、国产平台深度优化、高性能、易管理的新一代自主服务器操作系统；可支撑构建大型数据中心服务器高可用集群、负载均衡集群、分布式集群文件系统、虚拟化应用和容器云平台等，部署于物理服务器和虚拟化环境、私有云、公有云和混合云环境。与上一版本相比，银河麒麟高级服务器操作系统在功能/性能表现、安全能力、软硬件兼容等方面提升明显，在兼顾系统性能与系统安全的同时，引入新技术、新特性稳定地支撑应用，解决企业数字化转型业务上云的核心需求。

银河麒麟桌面操作系统V10 SP1 2303适配国产软硬件平台并深入优化，提供最优的硬件兼容性。新版本实现界面风格和交互设计的全新升级，注重移动设备协同并引入安全可信计算体系。与上一版本相比，银河麒麟桌面操作系统V10 SP1 2303在功能易用性、交互体验一致性、系统安全性以及应用兼容性等方面均有显著提升，并紧紧围绕体验好用、安全好用、生态好用、行业好用这四个核心目标，为用户带来最佳的使用体验同时，满足用户跨场景协同办公的需求。

星光麒麟操作系统1.0（Starkylin 1.0）具有“弹性扩展、智能互联、内生安全、统一框架”的产品特性，已全面适配、覆盖国产主流移动芯片，基于Linux+H5兼容AOSP应用生态。星光麒麟操作系统首个1.0版本将覆盖平板、大屏、手持设备和可穿戴设备业务需求。星光麒麟以安全为基，打造软硬结合的内生安全体系，满足各行业面向万物智联时代的安全应用需求。未来，星光麒麟会持续拓展至工控、物联网以及边缘计算等设备，支持海量业务系统互联互通。

2、标杆案例:银河麒麟操作系统助力某证券金融信息化创新项目建设

银河麒麟操作系统助力某证券金融信息化创新项目的建设，实现了其核心业务系统和传统应用均可在国产桌面操作系统上稳定使用的需求。用户无需切换操作系统来使用传统C/S架构应用，保证了客户的核心业务连续可靠运行，降低了用户的学习成本，具有示范效应，为金融行业国产化创新提供了一条可行的过渡技术路线。

某证券公司为某省金融信息化创新项目首家试点单位，现已完成办公系统（办公OA系统、电子邮件系统、云桌面系统）、核心系统（网上交易信创系统、理财商城系统）、麒麟应用虚拟化等业务系统的建设，并稳定运行。

项目成果：

针对某证券在短期内需要改造适配国产操作系统的难题，联合麒麟软件组建联合创新小组共同就内部的C/S应用适配开展测试评估，对C/S应用进行分析后，制定了以应用虚拟化为主的路线，将原操作系统上的应用程序生成为可跨平台发布的虚拟应用。通过多方面协作配合展开工作，项目已完成86套服务器系统，450套桌面系统，1套软件商店，450节点应用虚拟化等相关底层系统的部署工作。

技术架构：



图表 36 项目技术架构

桌面：同方、兆芯、银河麒麟桌面操作系统V10、阿里专有云桌面、麒麟软件商店（私有化版）；

服务器：华为（鲲鹏920）、海光、银河麒麟高级服务器系统V10；

办公软件生态：wps、奇安信浏览器、dbeaver等；

麒麟应用虚拟化：SAP、PB、HSRCP_ADMIN、恒生柜台、Trade、MobaXterm、Navicat、pl/sql、postman。

项目成果：

本次应用虚拟化方案实施完成后，确保了核心系统顺利迁移，成为该省金融行业首个应用虚拟化落地案例。通过银河麒麟应用虚拟化管理软件，用户不需要依赖虚拟机，应用服务直接运行在高性能服务器或服务器集群上，与原生系统及硬件之间不再有“虚拟化中间层”，可以极大降低性能损耗。同时应用程序可以便捷地下发到用户终端桌面，不再受用户终端CPU架构和操作系统限制。

4.2.2 统信软件

1、产品介绍

统信UOS云原生版（统信有燕）是基于统信UOS服务器版V20衍生而来，面向云原生场景打造的轻量、快速、安全、不可变操作系统，通过深度结合Kubernetes和容器技术，助力客户轻松搭建企业级容器云PaaS平台，满足企业业务上云需求。



图表 37 统信UOS云原生版OS的优势

统信容器云管理平台（统信有雀）是以容器技术为核心，面向应用的企业级容器云PaaS平台。平台整合了业内主流的DevOps、微服务、服务网格、CI/CD等新理念和新技术，为企业容器化的应用提供从开发、测试、部署到运维的全生命周期管理功能，真正实现应用构建交付环境一致化、资源管理自动化、应用运维智能化，全方面提升企业竞争力，助力企业顺利完成“云改数转”。

容器云平台自主化替换

完全兼容OKD、Kubernetes集群环境，无需更改配置，即可将业务系统平滑迁移至自主化容器云平台，助力企业提高业务运行可靠性，满足国内用户自主可控、安全合规的需求。

CentOS迁移及容器化改造

提供深度优化、安全加固的不可变系统，支持根据用户需求进行容器化和微服务改造，提升业务应用可移植性，同时实现应用容器化，助力云原生改造。

云基础设施平台构建

支持快速搭建单节点和多节点云基础设施平台，提供弹性伸缩机制，将硬件资源抽象成资源池，作为云主机资源供上层应用统一调度，从而有效降低IT运维成本，提升业务敏捷性。

2、标杆案例：科研集团容器云解决方案国产替换

拥有数以万计用户量的某科研集团，计划进行容器云解决方案国产替换，持续提升业务效率：

项目概况：

某科研单位积极推进国产化进程，将原有Intel x86环境中使用的Kubernetes集群替换为国产化产品，在满足信创安全的同时，进一步提升自身业务快速创新的需要，提高业务迭代整体效率。

项目挑战：

- 1) 原有Kubernetes集群对应用的生命周期管理较薄弱，Pod升级时会影响业务连续性，客户体验较差，需要支持业务的平滑升级。
- 2) 研发体系面向DevOps从无到有的转型挑战，需要借助业界优秀实践，完成体系转型。

解决方案：

- 1) 安装部署“统信有雀”，完全兼容客户原有Kubernetes集群，无需更改配置，并通过流量控制等微服务治理方式，保障在完成平台中的Pod升级时，业务平滑运行。
- 2) 通过分析用户研发和运维流程，基于自动化、可移植性、持续集成、滚动发布等特性，提供整体方案，帮助用户完成DevOps转型。

项目成果：

- 1) 将企业业务顺利迁移至“统信有雀”上，圆满完成国产化替代工作。
- 2) 通过完成优秀的DevOps实践，提升企业自身研发效率，能够快速创新和交付验证，使科研效能再上新台阶。

4.3 数据库

4.3.1 蚂蚁集团旗下OceanBase

1、产品介绍 OceanBase

OceanBase 已助力400+行业客户实现关键业务系统升级，应用于超过 1/4 国内头部金融机构。现已服务工商银行、建设银行、中国人保、中华财险、招商证券、山东移动、中石化、海底捞、GCash 等，助力金融、政府、运营商、零售、互联网等多个行业的客户实现核心系统升级。4.1 版本在 4.0 版本的基础上进一步增强性能、稳定性及高级特性，内核能力在分布式事务和存储方面进行大量优化。未来的版本中，OceanBase 将继续提升OLAP性能，通过列存功能来进一步增强对复杂查询的支持能力，同时提升极端场景下的应急处理能力。

此外，OB Cloud实现跨基础设施的混合云部署，可以根据客户需求将OceanBase数据库一键式的部署到阿里云、腾讯云、AWS等云厂商上，支持包年包月和按量付费模式；支持从14G到128核800G等不同规格的实例，能够支持用户“从小到大”的全生命周期数据库需求，帮助用户更好地业务创新及降本增效。

依托OceanBase的“LSM-Tree数据高级压缩技术”、“高效的分布式事务处理引擎”等绿色计算能力，经中环联合认证中心（CEC）测算，2022年OceanBase助力蚂蚁集团合计减排 4392吨二氧化碳当量。

2、标杆案例:云南红塔银行新一代核心系统

OceanBase 已助力400+行业客户实现关键业务系统升级，应用于超过 1/4 国内头部金融机构。云南红塔银行就是一个典型的核心系统全栈分布式升级案例。

客户痛点：

云南红塔银行原有IBM小机+DB2+高端存储组成的集中式数据库系统成本昂贵，其核心系统的系统架构、数据架构都已保持不变运行多年，虽然运行相对平稳，但在当下数据量猛增，并发不断增大的业务场景下，已经难以支撑业务需求。

解决方案：

云南红塔银行启动了新一代核心系统项目群建设，采用“云计算+微服务+OceanBase分布式数据库”的技术底座，向云原生分布式架构转型。

- 化繁为简，降低成本：利用OceanBase的分布式数据处理能力和分布式事物处理能力，不采用单元化方案，不使用分布式事物中间件，简化应用的开发部署。
- 同城双活，灵活高效：利用OceanBase分布式数据库的高可用能力，采用两地三中心五副本架构，“主城市”包含两个生产数据中心，同时应用系统对等部署在两个机房，实现同城双活。
- 异地容灾，安全无忧：异地灾备数据中心部署主集群的一个仲裁副本，同时部署备集群提供异地容灾能力，备集群通过日志回放进行异步的数据同步。



图表 38 云南红塔银行分布式数据库部署架构图

应用效果：

OceanBase数据库上线后，云南红塔银行具备服务亿级客户、每天亿级交易量的“双亿”能力，实现了平均每分钟处理约50万笔联机交易、交易平均响应时间不高于300毫秒的卓越性能。核心应用性能压测达到国内领先水平，满足未来8-10年的业务增长需求。

- 硬件投入成本是传统集中式数据库在相同吞吐量下的1/10，系统运维成本降低数百万元。
- 大幅提升了业务创新效能，实现同城双机房全功能对称双活，在高可用、易扩展、高性能等方面得到全面提升。
- 故障恢复时间缩短到30秒内，业务连续性高达99.99%。

4.3.2 蚂蚁集团旗下TuGraph

1、产品介绍

高性能图数据库TuGraph由蚂蚁集团和清华大学共同研发，经过蚂蚁万亿级业务的实际场景验证，并通过国际图数据库基准性能测试（LDBC-SNB），测试结果3次刷新世界纪录。TuGraph在功能完整性、吞吐率、响应时间等技术指标上均达到全球领先水平，为用户提供高效、易用、可靠的平台管理和分析复杂关联数据。

2、标杆案例：某大型国有商业银行资金关联分析案例

TuGraph应用于蚂蚁内部的金融风险、会员关系、数据治理等300多个场景，为支付宝在2021年实现资产损失率小于亿分之一发挥了作用。同时也面向金融、政务、能源、工业等多个行业的客户提供服务。

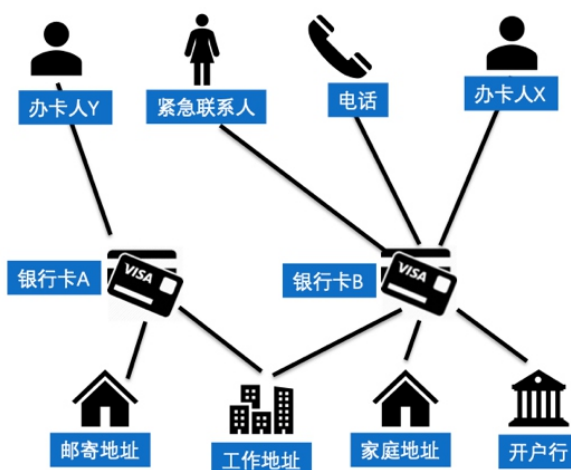
客户痛点：

作为国内最大的信用卡发卡行之一，某大型国有商业银行的用户长期以来面对无数作案团伙想法设法钻营风控漏洞，实施金融盗窃或诈骗，非法牟利的挑战。用户信用卡申请中的团伙欺诈占总欺诈的比例较高，作案方式隐蔽，传统分析方法无法有效识别风险。近年来，涉信用卡犯罪呈现出犯罪团伙分工专业、犯罪手法隐蔽多样、复合型犯罪多发等新特点，如何挖掘犯罪欺诈团伙是反诈迫切需要解决的问题。

解决方案：

TuGraph基于图模型的形式，将信息中的实体以及实体之间的关联关系，分别抽象表达成为“顶点”以及“顶点间的边”这样的数据结构。通过将电话、邮件、地址、介质等数据信息构建申请关联关系图，基于社区发现算法（Louvain、广度优先遍历等）挖掘人、交易关系、转账关系、电话、邮件、地址、介质之间的群体关系以及聚集关系，进一步分析查找其中可能的异常申请信息，并挖掘出有相关性的作案团伙，使违规申请团伙无所遁形。

信用卡申请关系图



图表 39 信用卡申请关系图

应用效果：

业务方面，通过先进的图数据库技术，根据已知的黑种子进一步挖掘信用卡申请关系，发现了黑产所拥有的信用卡、涉嫌非法申请信用卡套现等，为银行降低了资产上的损失。

技术方面，原有采用的技术手段是某开源的图系统，但是计算一次信用卡申请关系的时间要用10个小时。采用TuGraph后，整体运算时间从10小时降到10分钟，60倍的性能提升，大大提高了黑产挖掘的效率，降低银行资产的损失。

基于TuGraph，用户实现了绿色可持续的大规模风险分析计算，效率和能效均远超传统方案，助力银行实现高效、环保的大规模风险管理。

4.3.3 海量数据

1、产品介绍 海量数据库Vastbase G100

海量数据在数据库的运维服务领域和自主研发领域均有超过10年的丰富经验，在数据库内核及周边产品、以及各大核心行业应用系统的理解程度上，均处于国产数据库厂商的头部位置。

海量数据库Vastbase G100是海量数据基于开源openGauss内核开发的企业级关系型数据库。融合了多年对各行业应用场景的深入理解，Vastbase除了具有openGauss极致的性能以外，还在其原有功能基础上增加了大量对Oracle、MySQL与PostgreSQL的兼容特性、提供了极佳的业务连续性保障、并遵循EAL4+标准进行了大幅度的安全增强、同时整合GIS空间组件等一系列专业应用领域的企业级功能，为用户数据库选型提供更高性价比。

2、标杆案例:中华联合人寿保险核心业务系统

客户诉求：

作为金融行业在信创产业的标杆客户，中华联合人寿保险急需寻找一款国产数据库产品，在功能性、稳定性、可靠性、安全性和兼容性等方面，可以完全满足保监会对于保险行业核心业务系统的要求。具体要求：

（1）团险新核心业务系统：专注于承保、保全、理赔及收付费等寿险核心业务的处理。要求系统具备相对的独立性以及开放的架构、高效的信息吞吐和数据处理能力。

（2）高可靠和强一致：保险核心系统属于交易型业务，对数据的安全性和可靠性要求极高，不能出现数据的错误和数据的丢失。

（3）高性能：流量增大时，要求数据库实现复杂保单业务大批量数据在线验证、导入与查询功能。

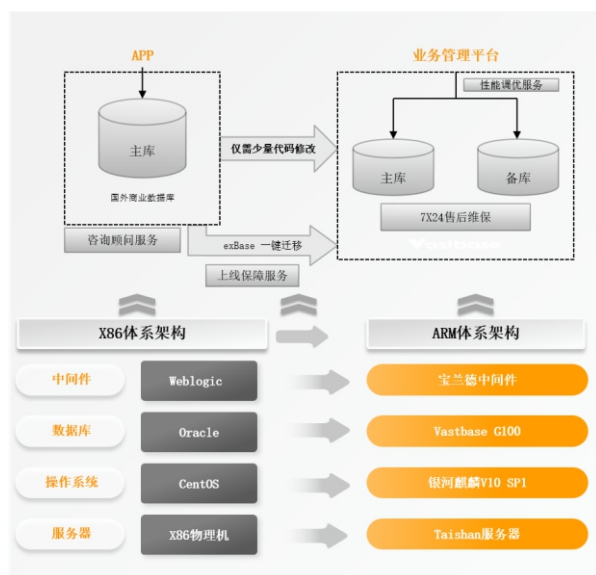
（4）高安全：投保客户的用户数据属于隐私数据，需要100%保证数据的安全。

解决方案：

（1）数据库产品Vastbase G100，提供了高稳定、高性能、高可靠、高安全、高兼容的数据库平台，完全满足保险行业核心业务系统对数据库的高标准要求。

（2）数据迁移产品exBase提供了专业的数据库和应用系统兼容性评估报告，实现了异构数据库平台一键式自动迁移及转换，帮助客户一次性迁移大量的核心系统数据库。

（3）三权分立的权限认证机制，以及多种数据加密方式和审计功能，全方位保护客户隐私数据的安全。



图表 40 中华联合人寿保险解决方案架构图

客户收益：

(1) 全栈国产：

海量数据库Vastbase G100与华为鲲鹏服务器、银河麒麟V10（openEuler版）操作系统、宝兰德中间件、中科软开发商等全栈国产软硬件产品深度适配及融合，构建出了全新的信创IT架构体系。

(2) 性能提升：

在批量导入、批量加人、保全更换被保险人等场景中，新体系处理性能相对原国际商业数据库平台平均提升2倍。

(3) 高安全易维护：

完善的安全功能，帮助客户达到《个人信息保护法》的要求。自动化、可视化的运维平台，帮助客户达到降本增效的目标。

4.3.4 万里数据库

1、产品介绍 万里安全数据库软件GreatDB

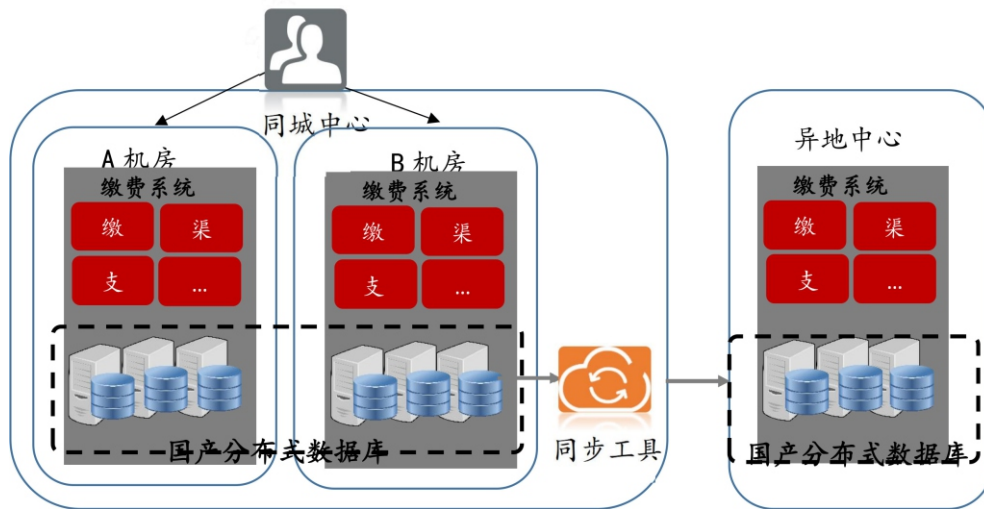
万里安全数据库软件GreatDB，是一款自主研发的关系型数据库产品，可根据用户需求采用集中式或分布式部署。它采用Shared-Nothing的计算存储分离分布式架构的设计，具有极强的可扩展性，满足用户按需动态扩缩容的需求；基于经典的多数派一致性协议实现数据库多副本的强一致性，满足金融级数据高可用需求；数据自动均衡分布，兼容主流SQL标准，对应用开发完全透明，数据可平滑迁移；采用了线程池、并行计算、计划下推等多种优化策略，深度适配多种平台，满足了高并发、低响应延迟的场景需求。

2、标杆案例：某国有股份制银行便民缴费系统

项目背景：

便民缴费系统之前使用的是Oracle RAC 2+2架构。在使用过程中针对灾备风险、存储风险、成本风险、扩展性风险、自主可控风险和业务需要、高吞吐低延迟等问题。从2019年开始逐步切换至国产分布式数据库架构产品支撑便民缴费业务。截至2022年末，便民缴费接入缴费项目超过1.5万项，全年缴费金额突破5000亿元，缴费笔数突破20亿笔，服务活跃用户突破4.4亿户，直联客户突破1亿户，最大可支持并发交易量为1.5万笔/秒，最大日交易量超过2亿笔。便民服务输出至微信、美团、中国银联、金融同业等650多家大型机构，基本实现国内主流服务平台全覆盖。业务峰值TPS约为8800，业务交易平均响应时间为57ms。目前单副本：10TB左右。

解决方案：



图表 40 解决方案图

价值体现：

- 1) “便民缴费”平台依托金融科技，统一缴费供应商、缴费用户、缴费渠道、支付渠道，降低了收缴费中的管理成本；
- 2) “便民缴费”业务涉足企业缴费市场，运用的数字化技术提升了企业的缴费效率，为企业提供了完善的缴费服务；
- 3) “便民缴费”业务真正从实处构建“金融+生活+服务”的普惠便民生态圈；
- 4) “便民缴费”深入创新发展，深刻服务新基建平台的持续化建设。在“后疫情时代”更加体现其业务优势，便捷的线上化服务符合未来“新基建”的发展方向，创新开放的服务模式也更能助力打造新型数字生活平台。

4.3.5 星环科技

1、产品介绍

星环科技是一家企业级大数据基础软件开发商，是截至目前国内唯一独立上市的基础软件企业。星环科技于2013年开始自研大数据基础平台，随后将业务逐步拓展到全产业链，基于分布式技术、多模型技术、容器云技术等逐步推出了分布式分析型数据库ArgoDB、交易型数据库KunDB、图数据库StellarDB、向量数据库Hippo、搜索引擎Scope、时序数据库Timelyre、时空数据库Spacture等一系列国产分布式数据库产品，拥有覆盖业内全部主流数据库类型的产品。星环科技坚持“自主原创，领先一代”的技术发展路线，用创新的技术去替代旧的技术，替代的对象包括传统关系型数据库Oracle、MySQL、IBM DB2、Teradata，搜索引擎Elasticsearch、图数据库Neo4j、时序数据库InfluxDB等，给用户提供更好的技术架构、更高的性能、更好的易用性、满足更多的场景等，帮忙企业实现数据平台升级和数字化转型。公司已经助力金融、政府、能源、制造、交通等行业多个客户实现了国外多类数据库产品的国产化升级。



图表 41 星环科技产品介绍

2、标杆案例：某集团构建多模型联合分析平台

某大型集团为了实现不同业务需求，部署了多个不同的产品，包括关系型数据库、图数据库、搜索引擎等。这些不同产品的接口标准不一致，开发成本高，同时数据存储在各自生态中难以互通，也难以保证数据的一致性和实效性。因此该集团采用星环科技分布式多模型数据库ArgoDB，替代了之前多个平台，基于多模型统一技术架构提供统一接口、统一计算引擎、统一存储管理、统一资源管理，能够同时支持关系型、图等多种数据模型。例如，当需要研究特定人群消费习惯时，过去需要分别到图数据库、关系型数据库、搜索引擎中查询，业务人员需要同时了解多种数据库技术，业务流程复杂。而基于星环多模型统一技术架构，实现了多种类型数据统一存储管理，不需导入导出，用户只需用一句SQL就能同时访问这3种存储模型进行联合分析，一次操作完成了之前三次操作才能完成的业务，大大简化了用户操作。

4.4 虚拟化和容器

4.4.1 云宏信息

1、产品介绍 CNware WinStack虚拟化云平台

CNware WinStack虚拟化云平台定位于构建可信赖的数字化基础信创云平台，聚焦为软件定义云数据中心的计算、存储、网络等基础设施提供产品级的、高度可靠的虚拟化单元，提供虚拟化&云原生双栈服务能力，具有模块化柔性演进、轻量便捷、按需应变等特点，助力实现开放解耦、高弹性、高安全、高可靠服务水平的数字化基础设施。

面向多样化算力和调度需求，WinStack具有独到的创新优势。

从算力供给的角度，WinStack的核心是通过虚拟化技术解决基础设施弹性、高可靠、绿色增效问题，扩展云原生运行时底座能力，支持超融合或存算分离等部署形态，奠定应用上云的基础；与其他云平台区别之处在于，WinStack十年磨一剑，原生基于服务器虚拟化产品为核心，因此产品化、标准化、成熟度更高，能够做到架构简洁、可靠、稳定、问题少。

从算力多样化的角度，WinStack是真正的独立第三方虚拟化软件，业内领先明确适配6大信创芯片、，融合多种xPU（包括CPU、GPU、DPU等），不挑硬件或品牌，不限制芯片路线、服务器品牌、网卡型号，技术开放度高客户议价能力更强，推动异构计算、算力融合发展。从算力效率的角度，基于积木式架构、迭代式建设的精益设计理念，WinStack通过控制5%的虚拟化损耗、5%的管理组件损耗、3节点最小规模，实现算力最大化利用、单位算力成本更低；面向火热的AI加速领域，容器运行时底座能够统一管理GPU算力资源池，提供秒级动态分配、弹性聚合与超分、细粒度监控、远程调用等能力，按需实现任务占用与退出的高效调度场景。

2、标杆案例：某大型股份制银行信创项目

客户背景：

信息数据存在安全隐患，业务故障率高，IT系统运行效率低，服务器异构资源繁杂，信息系统建设与IT资源的矛盾突出，为实现银行整体信息化的自主可控、提升资源效率和管理效率，某大型股份制银行拟采用国产软硬件实现本质安全、彼此良好兼容、增强管理便捷性需求，逐步过渡到基于国产芯片的云计算平台、持续稳定保障业务发展。

建设方案：

结合自主可控要求，采用多技术路线并行发展的方案，提供X86、Arm两种高性能计算平面保障算力供给、供应链安全；云宏CNware同位替代VMware虚拟化，打破“VMware+X86”体系，并在Arm低功耗、多核架构下实现能耗降低40%-60%、设备成本降低30%以上；支持无缝集成已有的NoF等存储，平滑构建具有先进性、支撑数字化升级要求的国产虚拟化平台。同时，CNware预留一云多芯的扩展建设能力，支持多种新型的国产CPU、GPU、DPU的选型验证，充分为中长期规划发展奠定基础。

客户收益：

- 性能高于目标，体验高于预期，使用户对国产软硬件发展的强化认识和信心；
- 分批改造、逐步迁移，科学合理地实现全栈自主可控；
- 提升业务可靠性、连续性，提升业务服务质量；
- 通过智能监控组件实现资源有效监控、高效利用，减少故障率，提升满意度；

某大型股份制银行信创项目



图表 42 某大型股份制银行信创项目案例

4.4.2 华云数据

1、产品介绍

华云数据通过为用户打造的“全芯全栈”自主创新云平台、超融合解决方案、公有云平台、工业互联网平台以及智能办公平台等解决方案，推动政府、企业数字化转型，区域信息化能力提升和中国数字经济整体发展。目前，华云数据业务覆盖全国及海外市场，公司遍布中国华北、华东、华南、西北、西南等区域以及北美和日本；在国内，华云数据为党政、金融、医疗、教育、电信、电力、交通和制造等各行业和领域的30多万用户提供上云服务，也为中国企业的信息化发展和区域数字经济发展贡献自主创新“云力量”。

2、标杆案例：某知名证券企业—云多芯方案

客户痛点：

该证券公司从2008年开始探索虚拟化技术，但随着证券业务迎来爆发式增长，对应的系统构建、运维成本也越发变得不容忽视。一方面，此前使用的多款国内外超融合解决方案，相关软、硬件成本持续高涨而服务能力却不尽如人意；另一方面，为应对客户容量、功能扩展、性能优化等一系列新需求，尽快上线一套安全可靠、高可用、可扩展的超融合解决方案已迫在眉睫。

解决方案：

华云数据在充分调研用户业务诉求的基础上，协助其构建了包含整机型裸机、虚拟机和容器的信创云资源池，整体架构如下：



图表 43 异构计算资源池逻辑架构

信创云计算资源管理平台承担计算资源池的资源管理任务，虚拟化资源池及裸金属资源池都接入管理平台中，并通过其对资源实例进行可视化操作，包括虚拟资源的调度、镜像管理等全生命周期管理操作。资源管理平台也会对资源池进行扩容及升级过程进行可视化管理。

通过本信创云架构，可有效精简不同国产操作系统、处理器之上所运行云平台版本的功能差异性，降低后续运营、扩容等诸多层面的复杂度，并为信创云之上运行的应用、数据库和中间件提供良好的云基础设施环境。

成果与亮点：

案例实施之后，该证券公司实现了一云多芯的云环境架构，通过优化云平台整体架构，使得相应的运营、部署工作更加便捷，并为后续推进从内核到外延的全面信创转型打下了坚实的基础。同时，本案例也入选国内知名研究咨询机构爱分析发起的“2023爱分析·信创产品及服务创新奖”。

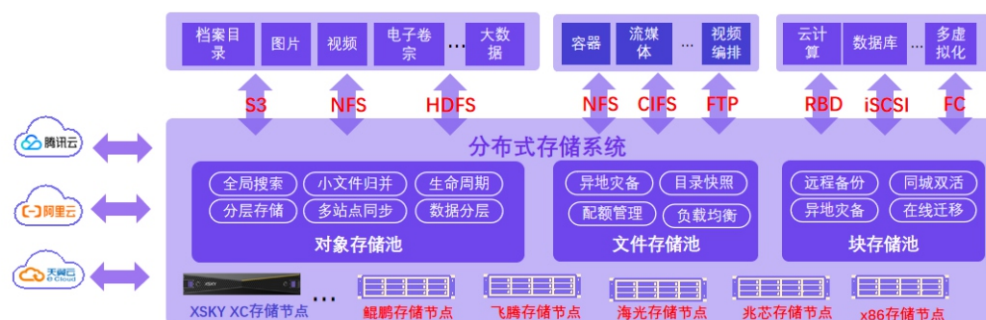
4.5 分布式存储

4.5.1 星辰天合

1、产品介绍：天合翔宇分布式存储系统

XSKY公司的天合翔宇分布式存储系统是可以异构多CPU生态的存储产品，实现“一池多芯”的异构能力，可在一个分布式存储平台下异构支持多CPU生态：Intel、AMD、鲲鹏、飞腾、海光、兆芯、申威等CPU生态的融合部署，将多生态CPU的代码进行移植、融合，并能保证存储系统能稳定运行。

天合翔宇是XSKY星辰天合拥有100%自主知识产权的产品。该产品在多CPU生态的存储节点中可以构建统一的数据资源池，不仅全协议支持，同时提供块、文件和对象服务，也可以实现关键业务高可用、场景一体化交付，同时还具备丰富的企业数据服务功能：快照、复制、双活、多站点、上云、数据策略流动、QoS、生命周期管理、WORM等，满足企业数据存储管理的各种需求。可以全业务场景适配，端到端的场景化解决方案对接，适配大数据、容器、两地三中心、私有云、数据备份、数据归档等各种应用场景。实现跨生态的数据生命周期管理；按需定制，弹性扩展。



图表 44 天合翔宇“一池多芯”分布式存储架构图

该产品帮助用户解锁对硬件CPU平台的依赖，降低后续的供应风险。同时可以实现业务无感知的将数据从非信创存储池迁移至信创存储池的数据迁移，最终帮助用户实现无业务中断的信创平滑替换。

2、标杆案例：新疆农信一池多芯云资源池

客户痛点：

对于新疆维吾尔自治区农村信用社联合社（简称“新疆农信”）来说，要实现底层存储平台资源、云管理平台、应用、数据等多层的信创支撑改造与建设，首先要面临解决适配兼容、融合部署、数据迁移的问题。

解决方案：

星辰天合提供给新疆农信专为信创环境研发的天合翔宇，构建 200 节点的存储集群，实现云平台 and 影像数据等全业务场景匹配，提供基于信创架构的超融合平台，承载银行各类关键业务，包括内网开发测试云平台、互联网开发测试云平台、内网生产 KVM 云平台、互联网生产 KVM 云平台，以及地州生产 KVM 云平台等。



图表 45 新疆维吾尔自治区农村信用社联合社信创云平台案例

应用效果：

星辰天合分布式存储平台与银行现有云平台深度耦合，通过提升存储资源的交付效率，支撑企业新应用上线效率的提升：一方面，支持硬盘级、节点级、机柜级或资源池级的不同粒度横向扩展；另一方面实现应用需求驱动的基础架构灵活变更和应用无感知的硬件生命周期更替，无需数据迁移；

进一步简化了信创平台基础架构，通过分布式集群的统一管理，替代原来多套存储管理，简化运维复杂度，提升效率，为客户节省 TCO 约 40%；

借助天合翔宇的一池多芯能力，新疆农信云数据中心可以实现硬件利旧，以及未来可以选购异构平台的硬件节点；使用星辰天合产品，数据可在新疆农信新旧设备中自动同步，也不会有业务中断的风险，无需进行数据迁移，相比采用传统架构存储组建云资源池的方案，可以节约可能高达百万元的数据迁移费用。

4.5.2 杉岩数据

1、产品介绍 杉岩海量对象存储

杉岩海量对象存储（杉岩MOS）是面向企业级海量非结构化数据的全分布式存储产品，采用业界先进的全分布式和无共享（Share Nothing）架构设计理念，将多台Arm/x86等服务器组成横向扩展型集群，构建统一的虚拟存储资源池，为企业数据中心提供云存储服务。杉岩MOS核心引擎能够帮助企业轻松应对EB级数据容量和千亿级文件数量的挑战，支持跨地域的数据中心主备、双活以及多站点容灾部署，为用户提供高可靠和高可用服务。

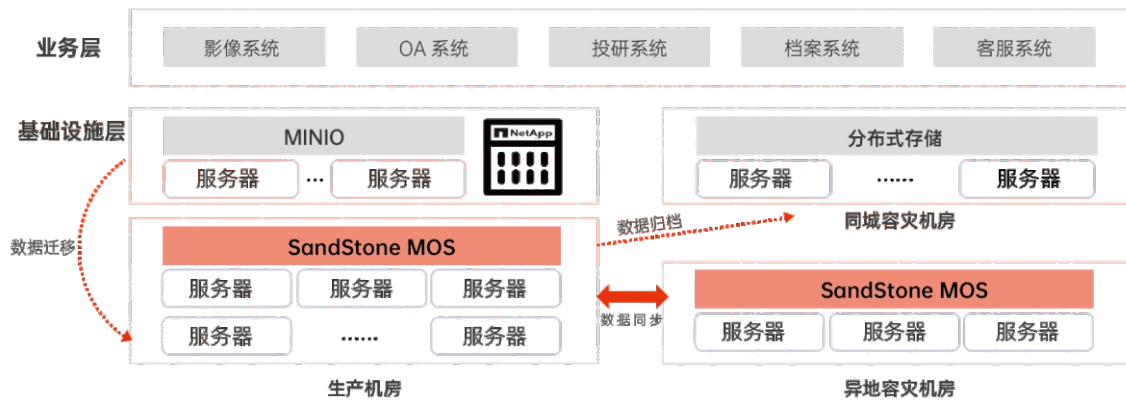
应用层	 电子档案	 金融影像	 电子证照	 流量分析	 视频监控	 标签平台	 执法记录
应用层	S3/OSS			NFS/CIFS/FTP/SFTP			
服务层	副本 / EC 策略		静态加密		全文检索		生命周期管理
	WORM		蓝光归档		多版本 / CDP		多站点 / 汇聚分发
MOS 核心引擎	DHT 数据分布			强一致性读写			
	小文件在线合并			智能元数据缓存			
	集群故障自愈			并行数据重建			
硬件层	磁盘控制卡管理		S.M.A.R.T/ 磁盘错误检测		慢盘检测 / SSD 寿命管理		
	通用服务器		硬盘错误在线修复		SAS/SATA/NVMe		
	Ethernet 网络						

图表 46 杉岩海量对象存储产品架构图

杉岩MOS提供海量存储、智能管理、数据价值三位一体的存储解决方案，兼容适配鲲鹏、飞腾、华为、海光等自主硬件平台，以及麒麟、统信等国产操作系统，提供安全可信、开放生态的智能存储解决方案及服务。已通过中国信息通信研究院首批分布式存储评测、工信部第五研究所中国赛宝实验室《软件类产品-存储备份（软件）》评测等，入选信创工委产品图谱、深圳市创新产品推广应用目录等，为新型基础设施建设海量数据应用提供存储底座。

2、标杆案例：杉岩海量对象存储助力某证券公司加速信创转型

杉岩数据为某证券公司建立了一套基于Arm架构的鲲鹏CPU和杉岩对象存储的统一存储平台，实现非结构化数据的统一存储、统一管理，消除信息孤岛，同时满足未来对非结构化数据的内容挖掘与利用、数据迁移等。在日常的业务中，某证券公司的业务系统逐步采用AI技术，诸如客户证件、远程开户录像、合同扫描件、客服中心语音、企业相关资料等非结构化数据每年增长40-50%。新建非结构化数据统一管理逐步把业务系统的非结构化数据，从原来各种型号的NAS、MongoDB、本地文件系统统一存储到杉岩对象存储上。



图表 47 项目方案图

全栈信创替代，满足信创需求：在国产化服务器上部署杉岩海量对象存储软件，全线产品兼容国产化软硬件，并支持国产化操作系统，满足信创需求。

丰富接口，业务系统快速接入：支持 S3、NFS、SMB、FTP、CSI、HDFS 等丰富的接口，实现云平台、大数据平台、备份软件等多种业务系统快速接入；可灵活对接 SNMP 和 Zabbix 等运维监控系统，提高运维效率。

全面安全方案，为业务保驾护航：实现多站点异步复制、总部分支机构汇聚分发，支持多站点智能 DNS 解析，实现多站点多集群统一监控管理，采用一套运维便利、架构简洁、高效灵活的非结构化数据保护解决方案。

项目上线之后，基于国产化服务器搭建了多套存储集群，实现现网存储数据的平滑迁移，应用于更多业务系统，加速 IT 基础架构升级。

4.6 大数据

4.6.1 蚂蚁集团

1、产品介绍：大数据云原生湖仓一体产品

云原生湖仓一体化产品实现海量数据规模支持，采用统一SQL、元数据、权限体系，通过多种特有技术，结合强大的弹性调度、serverless能力。可根据业务特性动态优化执行计划，实现计算、存储单元独立弹性调度及亲和性部署。其中MaterializeView自动识别提升重复计算复用率。Adaptive Query Execution技术实现执行计划动态调整，在减少数据shuffle次数、降低数据shuffle量、加少sort的应用等大数据关键问题上起到极大作用。利用serverless和亲和化部署能力，极大降低了容器成本的同时提升了本地计算覆盖场景。

高性能分布式执行引擎Yak - 完全基于自研的JIT的本地计算引擎。基于自研的代码编译和生成，编译时间通常只需要几毫秒，相比LLVM的方式缩短1~2个数量级，而且Yak采用whole stage的方式编译而不是每个算子编译整体的执行上会更快。本地计算采用Morsel-Driven Parallelism的方式，会根据机器的核数，不同算子设置不同并发度，避免中间数据网络传输。数据计算场景经常会需要scan大量的数据并参与计算，这时候数据拷贝成为系统计算的瓶颈之一，为了减少数据拷贝以及传输的性能损耗，Yak会和存储引擎部署在同一台机器上通过Shared Memory进行数据访问，大大提升系统性能。

2、标杆案例：蚂蚁AI系统接入湖仓一体

客户痛点

蚂蚁的AI体系中，特征数据包括入仓、准备、加工等一系列操作均依赖外部数据体系。一方面成本巨大，另一方面技术体系不可控，使得无法针对蚂蚁AI体系的数据特性进行计算、存储等方面针对性的优化。随着AI场景的持续增加，以及对数据应用要求的逐渐复杂，原有系统的响应速度越来越不符合我们的要求。采用蚂蚁自主可控的大数据系统为AI系统提供数据支撑的必要性非常突出。

解决方案

湖仓一体通过Calcite统一SQL层兼容了Spark/Flink/Explorer等多场景引擎的SQL标准的同时搭建统一的元数据管理平台，实现多场景快速接入。

引擎通过执行计划优化扩展及动态图优化结合数据特性，替换常规RBO/CBO优化方式，扩大了MapJoin、HashJoin使用范围。完善HashAgg/ObjectHashAgg算子替换了SortAgg场景，降低SortAgg带来spill to disk的压力。另外，通过近端结合远端的shuffle方式提升了shuffle service整体吞吐能力和资源利用率。

通过结合具备弹性调度和亲和性部署能力的统一调度系统，实现容器资源的serverless模型。同时，计算、存储亲和性部署方式大大增加了本地计算的概率，对网络开销和shuffle开销起到了极大的优化。



图表 48 湖仓一体架构图

应用效果

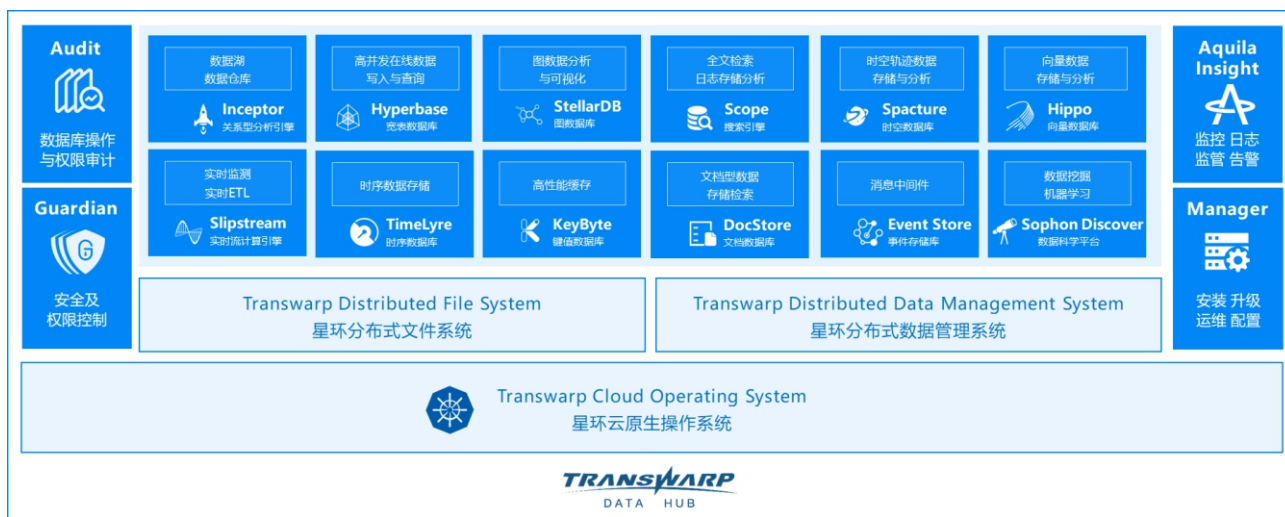
针对蚂蚁AI体系中特征数据日均20w+、10w core+计算任务，均已完成任务兼容验证、数据核对和性能比对。稳定支持2个机房级日均10w+任务已切换至湖仓一体平台。其中，约30%任务（如udf/udaf/udtf,动态优化后使用map join及HashAgg）场景具有10倍+性能提升。

4.6.2 星环科技

1、产品介绍：大数据基础平台Transwarp Data Hub

星环科技自主研发的企业级一站式多模型大数据基础平台Transwarp Data Hub（简称“TDH”），基于领先的多模型技术架构支持统一的接口、统一的计算引擎、统一的存储管理和统一的资源管理，支持关系型、图数据、向量、时空、时序等11种数据模型，提供高性能的查询搜索、实时分析、统计分析、预测性分析等数据分析功能，广泛应用在批处理、实时数据处理、高并发在线数据服务、向量检索、图计算、湖仓集一体等各类大数据业务场景。

TDH是全球首个通过TPC-DS基准测试与官方审计的产品，比开源产品具有更先进的技术架构、更高的性能、更好的稳定性与易用性，与国产主流软硬件兼容适配等优势，并成功替代了国外传统数据库和大数据平台产品，在政府、金融、能源等十多个行业应用，帮助企业更全面、更便捷、更智能、更安全地运用数据。



图表 49 大数据基础平台Transwarp Data Hub产品介绍

2、标杆案例：某农商行海量多模型数据统一存储管理

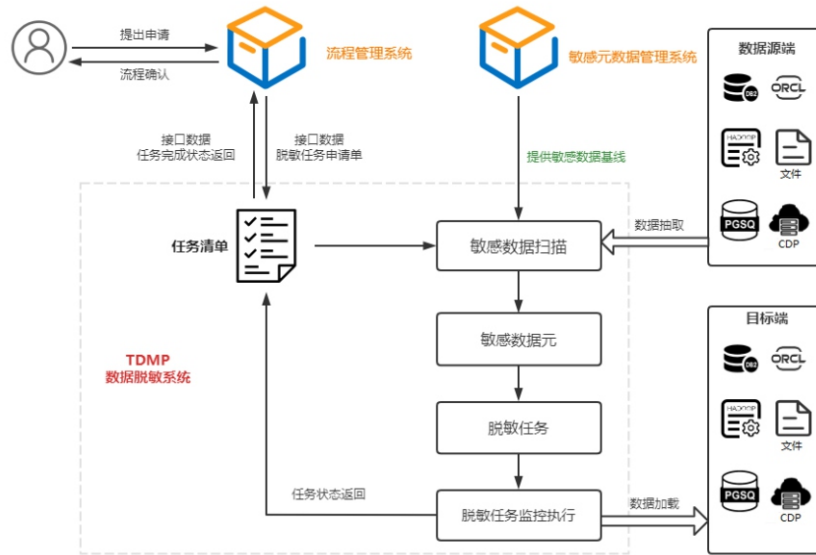
随着互联网金融生态的发展，某农商行业务数据内容不断丰富，数据规模不断增长。该农商行最初构建了MPP数据仓库、Oracle小数据平台以及Hadoop大数据平台。烟囱式的构建方式导致架构复杂，运维成本高，三个平台同质化，数据存储冗余严重。并且传统数据仓库无法兼顾实时营销与实时风险管理等高并发、低延迟场景需要，无法满足业务团队的数据挖掘与业务建模需求。

因此，该农商行基于星环科技湖仓集一体平台，替代了之前的三个平台，将海量多模型数据统一存储管理，降低了原同质化平台的运维开发成本和数据冗余。对外提供统一的SQL接口，既能面向业务实现高并发、高性能的历史数据、实时数据的查询服务，又能承载分析报表、批处理等分析型数据集市业务。行内大数据量批处理效率提升2倍以上，综合成效比提升14倍，并建立了行内统一的数据管控和治理体系，从整体上提升了行内对业务发展的数据支持能力。

4.6.3 神州数码

1、产品介绍：TDMP数据脱敏系统V3.0

神州数码自主设计、研发的TDMP数据脱敏系统V3.0是一款基于B/S架构，采用最流行的Spring MVC和Mybatis框架技术的专业级数据脱敏系统，具有数据源配置、数据脱敏规则、数据脱敏过程、任务管理、安全管理、系统管理六大功能模块，能够实现从咨询评估、实施部署、定制开发到系统维护、战略合作的全生命周期项目管理，在保护敏感数据安全、合法合规的同时，最大限度保证数据可用性及可挖掘价值。



图表 50 TDMP数据脱敏系统V3.0

标杆案例：某农商行数据脱敏平台项目

用户需求：

作为数据密集型行业，银行业在客户画像构建、风险防控、效率优化等业务都需要数据的支撑，保障数据的安全是银行业务发展的底线，推进落实企业数据安全势在必行。然而，农信社旗下的农村金融机构网点众多、分布广泛，在各乡镇与大量本地机构、客户开展广泛业务合作。相较于大型商业银行，农村金融机构数据所处的业务环境更加开放，业务生态更加复杂，数据安全更加复杂、开放、多源。

解决方案：

神州数码自主设计、研发的TDMP数据脱敏系统V3.0是一款基于B/S架构，采用最流行的Spring MVC和Mybatis框架技术的专业级数据脱敏系统，能够为企业的隐私保护提供高安全性、高可用性、高可靠性、高稳定性、高效率的专业数据安全方案。该产品同时支持静态脱敏和动态脱敏算法，拥有强大的敏感数据自动发现功能和并行处理引擎，具有灵活的脱敏策略、精准的隐私数据发现、多样化的脱敏方式、高效的处理算法、严谨的流程审批、详实的审计报告、完善的监控体系等特点，可针对全业务场景需求提供变形、屏蔽、替换、加密等多种敏感数据处理方式，并能对数据脱敏服务器资源进行实时监控，为用户提供可追溯的安全审核。

应用效果：

经由神州数码TDMP数据脱敏系统V3.0处理后的业务数据保有高仿真性、一致性和源数据分布特征的继承性，从而继续满足关联分析、机器学习、即时查询等不同应用场景的使用需求。

4.7 中间件

4.7.1 东方通

1、产品介绍：东方通中间件新品

东方通名称配置服务器TongNCS：

东方通名称配置服务器TongNCS是一款高效可靠、安全易用的国产分布式服务注册发现和配置管理中间件产品，提供了一组简单便捷的特性集，帮助开发者快速实现动态服务注册与发现、服务集中配置、服务元数据及流量管理等能力。

TongNCS满足高并发与高性能要求，能够支撑10万级的客户端和50万级的服务实例；采用云原生架构设计，支持spring、springboot、springcloud、dubbo等不同技术框架，支持java、Python、GO、Rust、C++、.Net、Lua、Node.js、PHP等多种开发语言，并适配了主流国产数据库系统。

东方通中间件管理平台TongMMP：

东方通中间件管理平台TongMMP基于东方通多年对中间件管理、运维所积累的经验，自主研发和创新的一体化云原生中间件管理平台产品，集中间件运维、监控为一体的统一解决方案。主要用于构建企业运维运营一体化的PaaS平台，提供了集中告警、统一登录、统一纳管、集中监控、租户模式、可视化操控、日志平台等能力。帮助IT运维人员低成本、高效能的构建企业IT运维体系。

TongMMP品类纳管丰富，可纳管Tong系列、开源、商用中间件；可同时对部署在虚拟机、云平台、混合云等多种环境中的中间件进行集中管理，并实现全栈国产化，全面支持基于云原生的分布式环境。

云原生消息中间件TongLINK/Q-CN：

东方通TongLINK/Q-CN是一款云原生分布式消息中间件，助力为实时事件和历史事件构建实时应用。计算与存储分离的架构设计，使得它具备极好的云原生和Serverless特性。它拥有原生Java、C++、Python、Go等多种API，同时支持HTTP协议，可为分布式应用系统提供异步解耦合削峰填谷的能力，具备互联网应用所需的海量消息堆积、高吞吐、可靠重试等特性。

2、标杆案例：某运营商业务系统云原生消息服务项目

用户需求：

某运营商省级公司推进IT系统向PaaS云升级，其中消息服务需要具备云原生服务能力，以便满足各类核心业务系统在功能、性能、稳定性、可靠性、安全性，运维便捷性以及兼容性等方面的要求。其中，在功能方面，要求支持多租户、多种消息消费模式、消息延迟处理等能力；在性能方面，话单量不少于30亿/天，业务单环节处理峰值不低于TPS4.5万条/秒；在稳定性和可靠性方面，要求系统保障消息不丢失、支持中断消息重新消费等能力；在安全性方面，要求国产化产品，支持链路加密、数据加密等多种安全防护手段；在运维方面，要求提供便捷易用的图形化监控管理平台以及全面的监控管理API接口；在兼容性方面，要求消息服务系统需要与其使用的容器云平台适配，满足容器云平台对于消息服务系统的资源隔离、统一管控等要求。

解决方案：

东方通通过提供云原生消息中间件TLQ-CN为客户构建全面符合各项目需求的消息服务系统。首先，通过产品定制化改造，满足了客户特定容器云平台对于兼容适配的要求。其次，针对客户业务特点，优化监控管理平台，使消息服务系统能够根据客户输入的业务数据（如业务数据量、数据保留时间等），自动测算出所需使用的消息服务资源，并自动进行相关计算资源、存储资源的分配，并提供消息传输服务。并且通过充分的测试验证，东方通云原生消息中间件TLQ-CN在功能、性能、稳定性、可靠性、安全性等各方面，也均可满足业务系统要求。

应用效果：

在此项目中，东方通基于云原生消息中间件TLQ-CN为客户构建了一套适用于PaaS云的云原生消息服务系统，全面满足了客户构建PaaS云时对于消息传输服务的各项需求。后续，随着客户各业务系统向PaaS云迁移升级，此套云原生消息系统将全面承担起业务整体的消息通信服务。

4.7.2 宝兰德

1、产品介绍：宝兰德中间件新品

宝兰德中间件统一管理平台BES MCP

BES MCP提供统一的中间件管理入口，帮助企业实现对容器和非容器中间件的快速部署、配置、监控、运维及全生命周期管理，解决中间件管理中的版本复杂不统一、自动化程度低、运维监控不完善、管理分散等问题，平台支持宝兰德商用及主流开源中间件等十余种类型中间件，有效提升中间件运维管理效率。

宝兰德Web服务器软件BES WebServer

宝兰德BES WebServer顺应对日益增长的流量进行灵活管理的需要，适应云原生、微服务技术架构和平台服务化的趋势，提供10万+高并发连接下超高速请求响应能力。产品基于主备部署模式提升服务高可用保障能力，灵活的轮询/权重/会话保持等多种负载均衡策略，内置SQL注入/DDOS攻击等常见安全攻击防护能力，通过这些关键能力的支持，对外提供更优质的安全流量负载管理服务。

宝兰德分布式缓存软件BES CacheServer

BES CacheServer是分布式高性能KV存储数据库，完全兼容Redis协议标准。支持基于内存和文件的持久化存储，保证数据的安全可靠；使用标准C编写，提供高性能内存数据结构服务；支持包括字符串（strings），散列（hashes），列表（lists），集合（sets），有序集合（sorted sets）在内的多种数据结构；支持单机、主从、集群等多种部署架构；支持数据分片，数据分布在多个实例中；客户端支持Python、Java、Lua、PHP、Ruby、Perl等多种语言。可用于解决高并发、大数据量场景下的数据访问性能问题。

宝兰德消息队列平台服务软件BES MQPaaS

BES MQPaaS分布式消息平台旨在构建统一的分布式消息平台，对业务提供统一的消息服务，支撑新业务对消息服务的使用需求，同时提供统一的消息服务管理、配置、监控、运维体系。对业务应用提供统一、编程风格一致的SDK及API，对业务应用使用消息服务提供统一接入、统一鉴权、黑白名单、负载均衡等能力。

2、标杆案例：某大型证券公司国产化中间件替换项目

用户需求：

该证券公司经营范围包括证券经纪、证券投资咨询、与证券交易、证券投资活动有关的财务顾问、证券承销与保荐、证券自营和证券资产管理等，随着中信建投证券信息技术应用创新工作的推进，需要将大量的系统迁移至国产自主可控的软硬件平台之上。宝兰德高性能中间件作为基础软件平台连接起了操作系统和上层应用，减轻应用迁移的成本，具有十分重要的作用。为更好的推进信创相关工作当下面对的主要难点为应用软件的适配和数据的稳定迁移，由于证券类项目最重要问题是数据的安全传输和高并发下的数据正确传输。

作为头部券商，在此次金融信创的项目中先行试点，已在OA、CoreMail邮件系统、FMS、CFE等系统成功部署宝兰德应用服务器软件产品，2021年至2023年持续推进更多业务系统的生产环境割接，同时采用BES WebServer作为安全流量接入管理的模块替换原有开源方案。

解决方案：

该证券公司信创环境IaaS层采用华为Arm架构服务器，以宝兰德公司BES WebServer产品完成了对Nginx的替换，实现流量高性能、安全接入管理。



图表 51 BES WebServer特性

同时使用了宝兰德BES AppServer、BES CacheServer等产品。当前已经在全信创生产环境部署340余台主机，成功上线20余类办公及一般业务系统，其中OA系统、FMS文件管理系统、CFE中央流程系统(偏核心)共28个节点，Coremail邮件系统共22个节点。同时在培训系统、运管平台等成功上线了以镜像、微服务等架构模式的服务系统，预计今年将实现600+节点的替换。

系统上线前期采用信创和非信创双保险环境支撑，XC环境中采用宝兰德中间件，非XC环境中采用Tomcat、Nginx、Redis，在2022年结束前已切换为全XC环境支撑。



图表 52 项目解决方案架构

应用效果：

该证券公司是国内头部券商，在此次项目中，成功完成 OA、CoreMail 邮件系统、FMS、CFE 等 20 余类办公及一般业务系统迁移到全国产环境中，宝兰德凭借其丰富的国产化建设经验和金融行业的项目经验积累，协助该证券公司圆满完成 XC 试点，取得了阶段性胜利，在 2022、2023 年持续推进更多业务系统的 XC 改造，拓展信创中间件种类，率先投产运营宝兰德负载均衡 BWS 达 100+ 节点，宝兰德 BES 应用中间件达 140+ 节点及分布式缓存 BCS 中间件达 50+ 节点。

4.7.3 蚂蚁集团

1、产品介绍：分布式中间件 SOFA

蚂蚁分布式中间件（SOFA）的产品发展路径，一直秉承着引领和拥抱业界先进标准和实践的理念，同时满足传统金融架构的平滑迁移和融合适配，以稳妥应对业务升级变更，并积极应对金融交易系统所面临的服务和数据扩展性、事务一致性、秒级容灾、弹性供给与调度等关键技术的挑战。

产品的核心能力包括双模微服务、分布式事务、消息队列、数据访问代理、任务调度、API 网关等，如下图所示：



图表 53 蚂蚁集团分布式中间件产品

2、标杆案例：某省级农村信用联社数字化升级

客户痛点：

某省级农村信用联社在数字化升级过程中，现有的IT系统面临着极大的挑战，已有的传统架构面临极大的业务压力，系统快速响应能力受限，业务连续性诉求也不断提高。在这种情况下，该行采购了蚂蚁分布式中间件系列产品。

解决方案：

该省级农村信用联社遵循“安全可靠、自主可控”的总体目标，逐渐采用自主、安全、可控的蚂蚁分布式中间件技术平台和架构方案，以保证IT架构的稳定与信息安全，实现移动互联网等新兴业务平台的快速发展，并进一步推动市场化改革，促进业务发展和创新。通过使用蚂蚁分布式中间件产品，该行得以实现业务应用系统全面云化、核心系统大机下移，快速响应业务要求，保证农信社业务稳定、快速成长，满足监管部门业务高可用和安全性要求，从而使农信社可以进入同业头部梯队，解决更多的农村、农业、农民问题，实现普惠金融。

应用效果：

蚂蚁分布式中间件产品，帮助客户建立了面向分布式架构的开发平台和运行平台，结合分布式数据库、分布式缓存、微服务框架、服务管理和发布流程等方面做统一的集成和接口封装，支持分布式应用的开发和管理，降低了分布式应用项目的开发成本。

在应用架构上，借助蚂蚁强大的分布式中间件产品，实现了从应用到数据的分布式多活架构，该行建成的下一代IT架构，对标行业内的先进实践，并保证架构的持续迭代和进化能力；多活机制覆盖从接入层、应用层到数据层，在接入层建立相应的流量调配机制，基于多活架构建设相应的容灾切换机制，保障了业务的连续性。

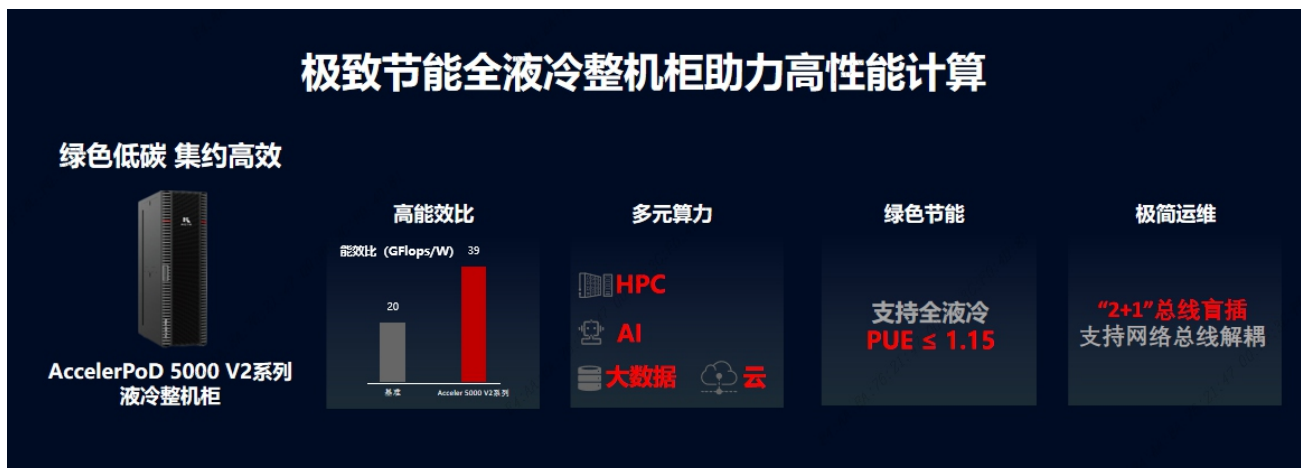
4.8高性能计算

4.8.1 长江计算

1、新产品介绍：全液冷高性能一体化整机柜AccelerPoD 5000 V2系列

作为算力基础设施国家队，武汉长江计算自成立伊始，就瞄准算力和运力进行产业布局和科技创新，依托中国信科集团在ICT领域的垂直创新整合能力，与鲲鹏、昇腾AI产业展开合作，推动计算产业发展，构筑自主创新产业高地。

2023年8月11日在武汉东湖算力大会上，长江计算全新升级发布了基于鲲鹏+昇腾的全液冷高性能一体化整机柜AccelerPoD 5000 V2系列。该系列产品较长江计算第一代液冷整机柜，在继承全液冷散热，实现极致PUE≤1.15的优势下，补充了适用于AI算力场景的AccelerPoD 510G以及通用算力场景的AccelerPoD 500K全液冷整机柜产品，同时在产品本身原有三总线盲插的特性下，结合需求创新发布兼容二总线盲插技术，支持网络总线解耦，在业务部署上拥有更多、更灵活的选择。



图表 54 AccelerPoD 5000 V2系列

全新升级的AccelerPoD 5000 V2系列液冷整机柜较之前的HCP、科研、渲染等领域拓展了AI、云计算等场景。

2、标杆案例：武汉人工智能计算中心

武汉人工智能计算中心是扎根武汉，面向华中的，由武汉市政府统筹建设，国企运营的大型高性能计算中心。融合了AI与HPC两种算力资源。在技术领先性和服务便利性方面达到“国内一流，国际领先”。当前已经扩容到400P的算力。围绕武汉人工智能计算中心，打造了“公共服务算力”、“应用创新服务”、“产业聚集服务”、“科研人才培养”四大生态平台，以AI算力集群赋能产业集群。

解决方案：

长江计算AccelerPod 510G算力集群产品具有极致算力密度、极高AI能效、极简交付部署等特点。该基础单元广泛应用于深度学习模型开发和训练,适用于人工智能计算中心、智慧城市、科研探索、运营商等需要重算力的场景。47U高度可最大提供25.6 PFLOPS/FP16 AI算力，支持机柜单元扩展，最大可拓展至3.2.EFLOPS,全节点200G网络互联。业界首创供电、互联、液冷三总线架构,计算节点全盲插。

方案价值：

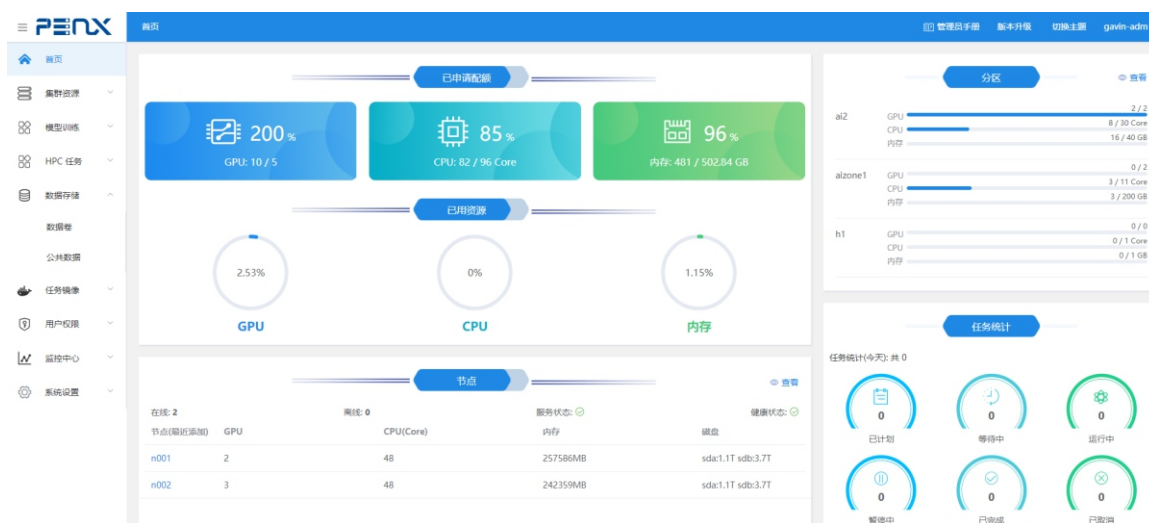
支持全液冷环境，免冷机、免空调。支持通用算力和AI算力的混合部署，能效比最高可达25.6PFLOPS/475kW，PUE≤1.15。为算力中心提供高能效和高计算性能的解决方案。

4.8.2超集信息

1、产品介绍：PlatforMax智算融合平台

PlatforMax是超集信息自主研发的面向于高性能计算（HPC）和人工智能（AI）场景的智算融合平台，采用先进的技术架构，由同一个管理节点纳管，可以将计算节点资源划分为HPC和AI两个分区，分区资源既相互独立又相互融合，满足用户在不同时间段对于HPC及AI业务的不同资源需求，为用户提供了一套既可以进行HPC作业，又可以进行AI训练的环境，助力HPC+AI业务的发展。

PlatforMax是定位于AI+HPC的容器云平台，能够实现异构资源的高效管理、调度和监控；除提供AI模型开发、AI模型训练到AI部署的完整功能外，也适用于生命科学、气象预报、量子力学、油气勘探等HPC科学计算领域。可以广泛应用于教育、科研、金融、医疗、能源等适用AI或HPC以及AI和HPC融合领域，降低AI和HPC的开发和应用成本，提高AI和HPC领域的研发和生产效率。



图表 55 PlatforMax智算融合平台

2、标杆案例：浙江某知名院校

项目特点：部署简易化，集群管理智能化

院校科研水平向前迈进同时，学院对计算能力需求与日俱增，如何充分利用现有设备，高效匹配科研计算需求进行合理的资源管理和分配，提升整体资源利用率，是客户当前的最主要诉求。

由超集信息自主研发的PlatforMax凭借清晰的管理界面、简单的使用操作、安全可靠的监控管理、灵活方便的维护方案，帮助客户实现了异构资源的高效管理、调度、分配和监控，满足了不同项目组的资源使用和隔离需求

客户价值：

在超集信息PlatforMax智算融合平台软件帮助下，客户通过直观的分区分区管理，更清晰地了解到各分区状态信息，通过全方位监控及调度，实现了20%以上的资源利用率提升。并且，通过Slurm作业调度系统和便捷的AI工具及预设镜像，实现了HPC和AI的高度结合。

展望

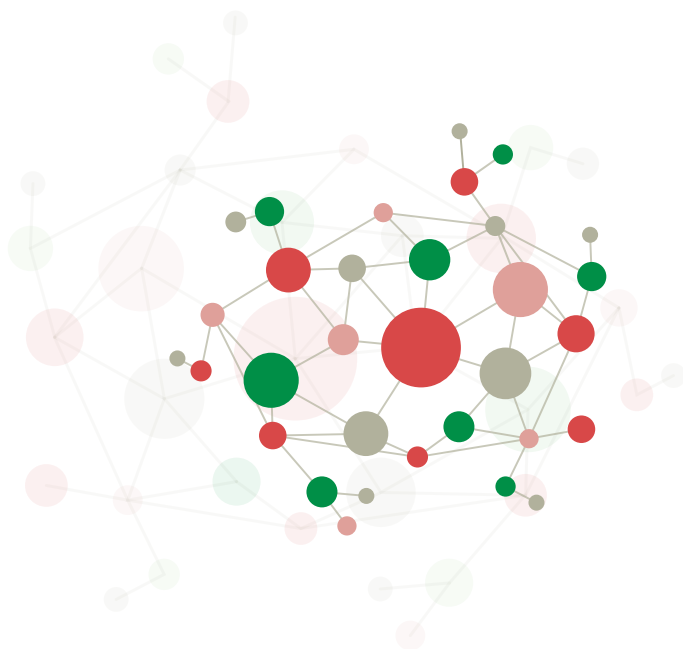
PROSPECT

在全球“碳达峰、碳中和”大背景下，站在算力全产业链视角、推进软硬协同的绿色计算已是产业关注重点和未来发展趋势。当前，产业界积极探索高能效、低能耗的硬件基础设施的建设，已形成一定的成熟度和应用规模，并有效作用于降低数据中心PUE值、减少企业碳排放。相比之下，在软件层面通过提升算力利用效率实现应用环节的计算绿色化，仍处于初期阶段，其产学研相关工作仍在探索期。而根据行业数据显示，大量数据中心服务器未被有效利用，意味着该环节存在着巨大的提升优化空间，是下一阶段实现绿色计算的重点工作方向。

在此背景下，绿色计算产业的发展呼唤包括政府机构、行业联盟、产业智库、技术厂商、应用企业等生态各方进一步协同，共建软硬协同的绿色计算生态体系，并重点关注算力应用环节的计算绿色化的标准构建和产业标杆落地，实现立足于未来的绿色计算长远发展。

绿色计算产业联盟联合政、产、学、研、用等各界伙伴，将共同持续推出《绿色计算产业发展白皮书》，力图勾勒出绿色计算产业发展的全貌，并以标准、评测、产业研究为抓手，推动处理器、固件、整机、基础软件、应用软件等产业链生态的成熟，以成功案例为标杆，促进绿色计算应用在各行各业的广泛落地。

让我们携起手来，构建新格局，重塑新时代，共同推动绿色计算产业的加速发展，为“碳达峰、碳中和”的绿色世界贡献出自己的力量。





绿色计算产业联盟
www.opengcc.org.cn