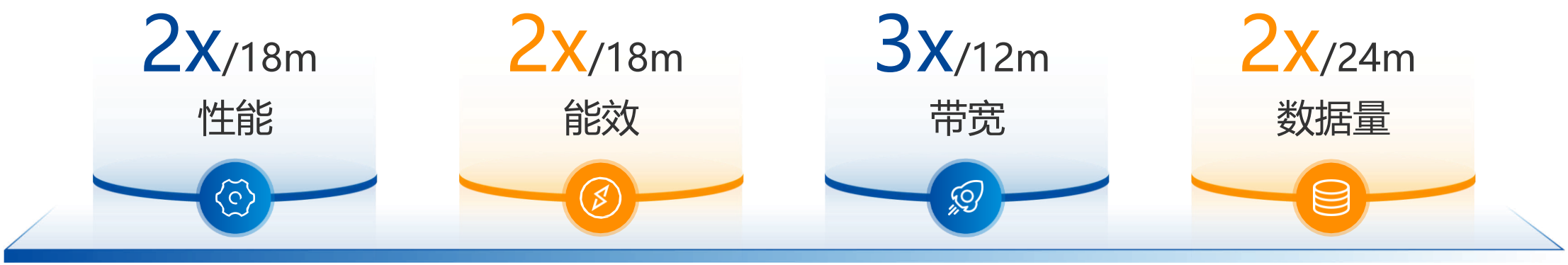
算网融合背景下的技术发展趋势与洞察

新华三集团副总裁、解决方案部总裁 李立

数字化世界正以指数级速度向前发展



摩尔定律 (Moore's Law)

集成电路上晶体管数量每2年翻一番，成本降低一半

库梅定律 (Koomey's Law)

计算机的能源效率每18个月翻一番

吉尔德定律 (Gilder's Law)

通信系统的总带宽每12个月 增长3倍

大数据定律 (Big data's Law)

全球数据量每两年翻一番
2021年，全球IP流量突破278.1 EB/月
从2016年到2021年，繁忙时段的互联网流量将增长4.6倍

算力作为数字经济关键生产力要素，其发展面临诸多挑战

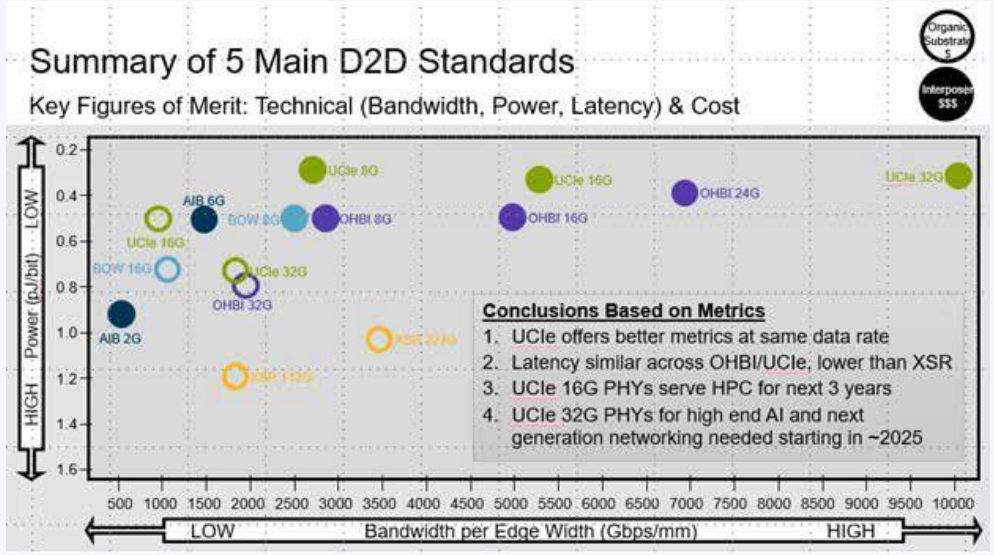
日益增长的行业智能化需求和算力发展的不均衡、不充分的矛盾日益突出

芯片级	面积换性能还是功耗换性能？单芯片Scale-up或多芯片堆叠Scale-out？
设备级	xPU、IO设备提升速度不一，传统冯诺依曼体系的内存墙、功耗墙如何突破？
集群级	数据中心既要高带宽、又要低时延，集群性能提升可否匹配xPU线性增长趋势？
地域级	东数西算跨地域调度，是调度算力还是搬运数据，如何有效实现高效调度？

芯片级，异构芯片集成技术发展迅猛，片间互联标准初现

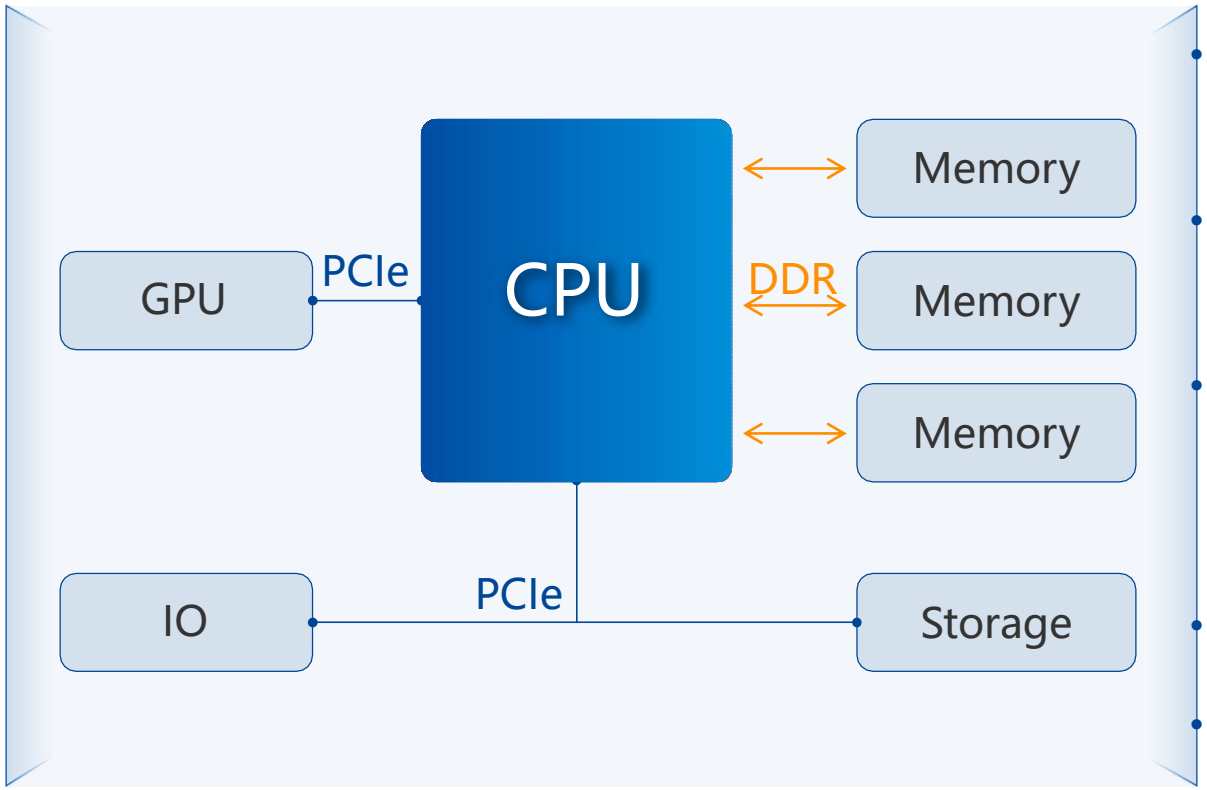
Chiplet

- 新的封装工艺的发展推动了芯片架构的升级，Chiplet使用2.5D或3D堆叠技术，将不同功能的芯片封装在一起，通过Die级复用形成较高的成本优势和开发优势，是未来芯片封装技术的主流趋势
- UCIe(Universal Chiplet Interconnect Express)标准联 盟成立，开放Chiplet生态系统起步
- CCITA在中电标协立项《小芯片接口总线技术要求》，Chiplet互联国标呼之欲出



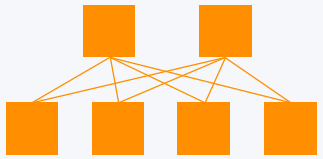
相对于传统Die-to-Die技术， 从各项指标对比看， UCIe具备一定优势

设备级，冯诺依曼体系下，主内存成为性能增长瓶颈



- CPU内核数量平均每2年增加一倍，主内存容量平均每3年增加一倍，意味着平均内存用量/CPU core每2年下降 30%
- 2022年1月，PCIe6.0正式发布，带宽升级至256GB/s(x16)，同期主内存升级至DDR5（6400 Mhz 51.2GB/s）
- 为了弥补主内存带宽不足和缓存未命中的性能损失，扩大CPU的L2/L3 cache（Intel/AMD）或者直接集成主内存（Apple M1），成为新一代CPU设计选择
- 片外内存功耗已达设备功耗40%~50%，也无法关停
- 需要突破冯诺依曼体系，寻找新的优化方向

集群级，追求Scale-out 性能线性增长



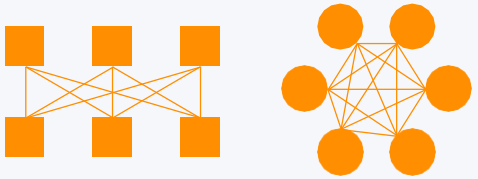
Spine-Leaf

数据中心主流网络架构，Spine+Leaf两层扁平组网，访问延迟低，同时具有良好扩展性

Dragonfly+

超算中心主流组网模型，Dragonfly+一个组为一个Spine-Leaf架构，组间形成全互联组网

DragonFly易于扩展成大规模组网，可应用于大规模算力中心



其他研究

EFLOPS： 网络设备采用背靠背组网，数据交互路径更短，但大规模扩展受网络设备端口限制

Aquila：基于Dragonfly，用定制的内部ASIC和通信软件，解决RDMA用于多租户场景

业界持续研究最优拓扑架构，实现Scale-out 性能线性增长，保障高带宽低延时

地域级，东数西算拉开算力跨域调度建设序幕

东数西算八大枢纽十大集群提供不同级别的计算能力，各算力中心跨区域进行互联形成算力网络

- **东部枢纽** 处理工业互联网、金融证券、灾害预警、远程医疗、视频通话、人工智能推理等计算速度快，网络延时低的业务
- **西部数据中心**处理后台加工、离线分析等计算量大，网络要求不高的业务
- 目前跨地域算力调度多为**数据调度**，将数据传输至算力所在位置进行计算

内蒙古枢纽 和林格尔集群	**宁夏枢纽** 中卫集群	**甘肃枢纽** 庆阳集群
成渝枢纽 天府集群	**贵州枢纽** 贵安集群	**粤港澳枢纽** 韶关集群
京津冀枢纽 张家口集群	**长三角枢纽** 长三角生态绿色一体化发展示范集群 芜湖集群	

算网融合的核心：解决从芯片到广域的IO不均衡问题

高集成，存算一体芯片架构，借助封装技术平衡IO

冯诺依曼架构存在内存墙和功耗墙问题，存算一体芯片架构通过将计算单元和内存单元按比例封装到同一芯片中，平衡计算和内存的配比，缩短数据搬运路径，降低数据搬运功耗

近期

近存储计算

Processing Near Memory

计算单元和存储单元相互独立，利用Chiplet封装技术进行芯片级封装，缩短数据迁移路径

Apple M1

基于混合键合的堆叠芯片

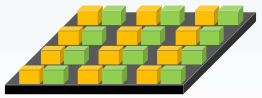
将逻辑单元与DRAM单元通过TSV键合在一起，大大减少访问外部内存的次数

中期

内存储计算

Processing In Memory

计算操作由位于存储芯片内部，存储单元和计算单元形成一个计算核心



基于片上缓存的存算一体

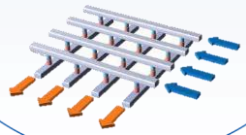
每个独立内核包含了计算核心与独立SRAM的单元，众多独立内核构成了一个独立芯片

远期

存算一体

In-Memory Computing

存储芯片内部的存储单元完成计算操作，存储单元和计算单元完全融合，写入为数据，读取为计算好的结果



基于数模混合的存算一体化

处于“百家争鸣”阶段，RRAM忆阻器最被看好，AI涉及大量矩阵运算，适合基于器件阵列来计算，速度快、能耗低

分布式，借助扩展总线技术，以Scale-Out模式提升IO

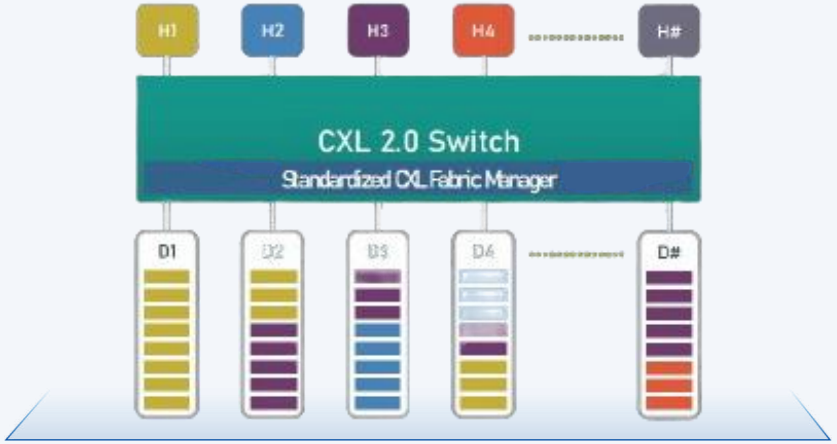
当前服务器IO跟不上CPU密度增长，借助CXL扩展总线技术，将内存、计算加速器件部署在服务器以外IO设备以Scale-Out方式对服务器进行扩展，平衡数据IO和计算密度

计算加速扩展：CXL Single Logical Device

计算加速设备统一池化管理，设备通过CXL总线挂载到服务器，打破服务器设计中计算加速设备槽位规划的限制

内存扩展：CXL Multiple Logical Device

纯内存设备池化管理后再进行细粒度的切分，内存通过CXL总线分配给服务器，打破服务器设计对内存总容量的限制

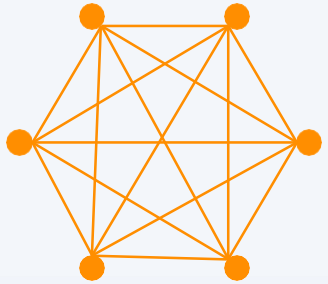


智能化，AI on AI，借助AI+在网计算保障最优算力拓扑

AI训练需要对海量数据和模型参数进行实时交互，需要高带宽、低时延的网络传输，而不同组网拓扑架构网络性能不同
AI技术赋能AI训练，采用AI技术，通过仿真方式对业务场景流量进行采集、分析、学习、规划，以期获得最优算力拓扑

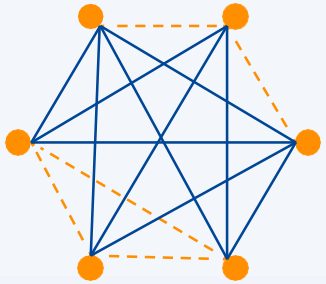
数据采集分析

搜集组网拓扑网元的设备配置、队列深度、带宽吞吐、流量统计、ECN/PFC统计等信息，对网络信息进行特征工程化处理



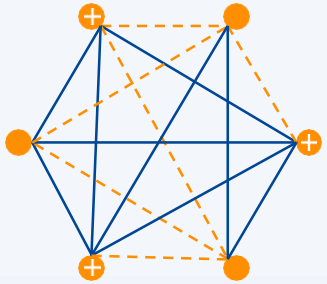
最优拓扑规划

利用强化学习等AI技术，针对业务流量模型，对实际组网拓扑进行规划，寻找最优算力拓扑



在网计算加速

当业务发生变更后，可重新规划调整算力拓扑，同时利用网络设备在网计算能力，加速保障算力拓扑始终处于最优状态



确定性，解决广域互联及可确定性算力调度

普通广域网络采用尽力而为转发，时延大、抖动范围宽，可靠性低，通过确定性网络相关技术，可为算力业务提供端对端确定性的传输保障

确定性网络技术体系

一至三层自底向上、跨层协同、精确可控

业务负载

业务负载

TCP/UDP协议

3层：DetNet/DIP

2层：TSN/5GDN/DetWiFi

1.5层：FlexE技术

新华三研究算网底层技术，助力客户降本增效

资源配制最优

统筹考虑算力因子网络因子，获得最优解

服务动态部署

灵活匹配与动态调度，服务灵活动态部署

紧跟标准

跟进运营商、国际国内标准，参与讨论和制定

场景落地

推动运营商、行业、企业客户算力网络方案落地

分布式算网大脑

带宽 抖动 网络因子 时延 可靠性

网络控制器

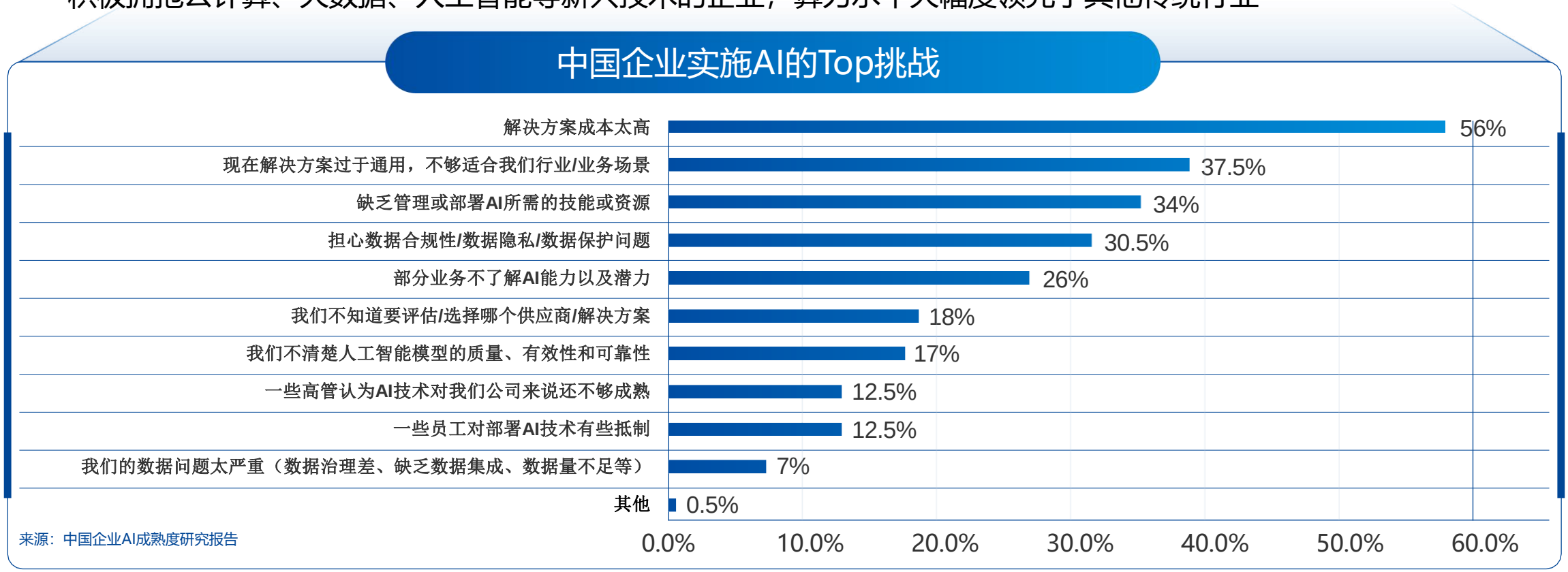
算力 存储 算力因子 地域 成本

云管平台

算力网络通过统筹考虑算力、网络、成本等多维属性
为客户提供最佳性价比的服务，提升体验

提升行业算力发展水平任重道远，行业智能化差距显现

- 算力的指数级发展，已开始形成头部聚集效应，各行业的算力应用水平及发展速度不一，行业智能化差距开始显现
- 积极拥抱云计算、大数据、人工智能等新兴技术的企业，算力水平大幅度领先于其他传统行业



行业算力发展指数Industry Computing Development Index

新华三计划推出行业算力发展指数ICDI，希望从多个维度评估行业算力发展水平，找出影响行业算力发展的关键阻碍，助推百行百业智能化，进而实现行业算力发展的“共同富裕”

行业算力发展指数指标体系（ICDI）

一级指标	二级指标	单位	指标说明
算力	通用算力	EFLOPS	•服务器算力规模，以行业内通用服务器算力规模统计
	AI算力	EFLOPS	•AI算力规模，以行业内AI服务器算力规模统计
	超算算力	EFLOPS	•超算算力规模，以行业内包含的全球Top500超级计算量统计
	边缘算力	EFLOPS	•边缘算力规模，以行业内可提供边缘算力的设备算力规模统计
能效	算效	TFLOPS/W	•平均每瓦功耗所产生的算力
	数据中心PUE	数值	•行业内数据中心平均PUE
网络	带宽	Gbps	•网络平均接入带宽
	网效	数值	•平均每TFLOPS算力的可用网络加权值（综合带宽、时延、抖动等）
	节点数	万个	•行业内节点设备接入总数
数据	数据量	PB	•行业内数据总量
	算力数据量	PB	•行业内可供算力使用的数据以及计算产生的数据量总和

新华三泛在算力赋能百行百业，人人都能享受AI红利

感 谢 聆 听

新华三集团
www.h3c.com