

2024开放计算中国峰会

2024开放计算中国峰会

2024开放计算中国峰会

移动云面向算力网络的算力基础设施

中国移动云能力中心计算产品部 王晓辉

2024开放计算中国峰会

2024开放计算中国峰会

2024开放计算中国峰会

2024开放计算中国峰会

2024开放计算中国峰会

2024开放计算中国峰会

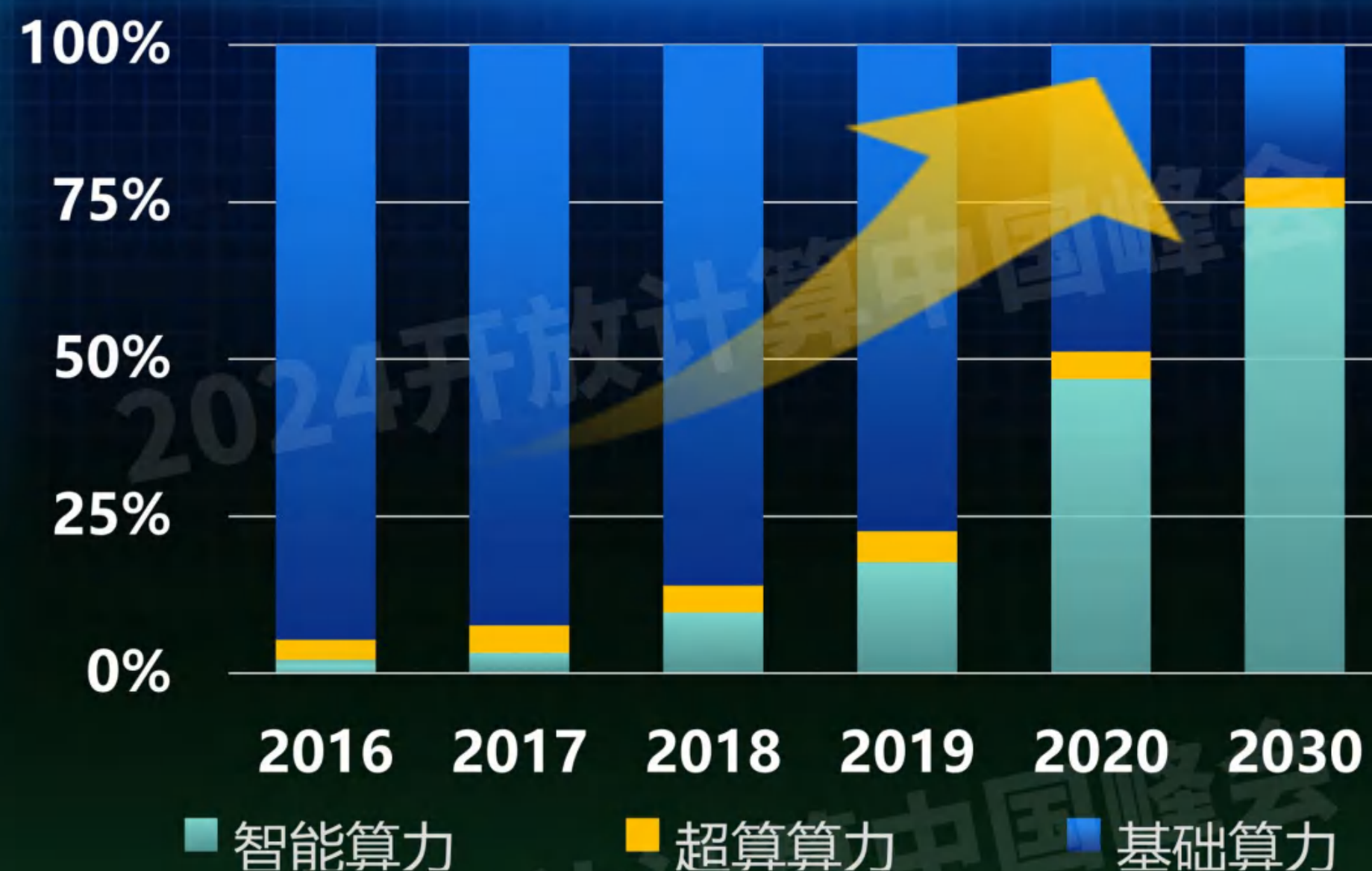
政策市场双轮驱动 推动算力创新发展

算力作为“新质生产力”的核心要素，是推动科技创新的重要动力。据预测，到2025年，中国算力规模将超过300 EFLOPS，市场前景广阔。移动云持续推进算力技术创新和多元算力布局，致力于为算力网络技术发展

政策驱动算力规模不断壮大

- 2023 六部委印发《算力基础设施高质量发展行动计划》，提出到2025年智能算力占比达到35%
- 2023 国家信息中心印发《智能计算中心创新发展指南》，夯实数字基础设施和数据资源体系“两大基础”
- 2023 国务院印发《数字中国整体建设规划》，推动公共算力新型基础设施创新驱动

市场驱动算力结构多元演进



数据来源: IDC、Gartner、TOP500等

面向未来的算力思考

- 如何面对算力规模增长与算力异构融合的双重挑战?
- 如何实现弹性扩展、高效灵活、生态自主的泛在算力?

亟需算力技术创新

面向算力网络 构建新型算力基础设施

算力技术多元创新 升级算网基础设施



软件定义硬件加速 打造计算三层架构



全新计算产品

- 凝芯聚魂，全新发布新一代计算产品
- 异构创新，加速AI应用低成本上云



异构算力 (COCA)

- 软硬一体，释放硬件最高性能
- 端网协同，超万卡智算集群互联
- 算力原生，AI跨芯架构无缝迁移



算网原生硬件

- 一机多芯，支撑多样算力
- DPU芯片国内领先，提升算力能效

移动云一机多芯服务器 打造多元化算力平台

一机多芯磐石服务器支持4种通用处理器 + 多种国内外GPU，结合磐石DPU形成多元化的基础设施解决方案

新一代磐石核心理念：一机多芯算力平台

异构芯片全量支持

服务器模块全量复用



全量化组件复用 算力平台灵活切换

一机多芯磐石服务器采用模块化设计，只更新核心计算板，其它组件完全复用，满足跨平台部署需求



系列化算力灵活切换

- 一套硬件平台支持多个CPU计算板，不同算力灵活切换

场景化机型快速交付

- IO接口/扩展板卡完全复用，加深物料储备水库，提升供货安全 and 生产速度，快速满足各类场景需求

标准化硬件生态构建

- 基于主板、部件、机箱、IO接口模块化/标准化，繁荣硬件生态

各项性能全面提升 助力上层业务加速

计算平台升级

60%↑

计算性能提升

新增针对不同业务负载加速器

DDR5内存升级

87%↑

有效数据带宽提升

全面支持CXL

IO扩展升级

20*PCIe

扩展槽位提升54%

全面升级PCIe5.0

链路信号优化

20%↑

采用线板及阻抗匹配方案

最优器件布局, 减少信号衰减及串扰

安全可靠

24w+

稳定运行时间

BMC/BIOS芯片级冗余

内存健康恢复



开机故障内存自检修复
主动内存故障预警, 故障自愈

系统宕机率比业界水平**低69%以上**
减少停机时间及不必要内存更换升级

前出线设计



便于前置运维, IO部件高效散热
增强数据中心单边维护效率

机柜后端温度可达55°C-65°C
机柜前端温度在25°C左右
智能网卡等热敏部件寿命**提升3倍**

计算/管理分离式设计



DC-SCM标准化管理模块
实现计算/管理解耦

加快安全管理芯片和CPU的迭代速率
实现**跨平台兼容**
简化运维 降低换板风险

BIOS / BMC芯片冗余



独家芯片级BMC及BIOS冗余

当主用芯片损坏启动失败时,
自动回复备用芯片
保障**远程运维安全**

BIOS seamless



实现固件秒级无感升级, 过程中可
选择快速热重启或不重启生效

实现FW**秒级热升级**
提升客户体验

智能诊断



CPU、内存、网卡等核心部件全面监控
基于丰富故障诊断库一次定位

根因定位准确度**提升至95%**
数据中心故障运维效率**提升80%**

计算存储资源丰富 支撑多场景应用

一机多芯磐石服务器采用计算、存储、IO资源最大化设计，满足各类云上应用场景

通用计算型

覆盖大部分业务场景，高效通用算力

IO密集型

高物理部件冗余性&高物理空间利用率

存储密集型

全闪配置优化，极限利用存储密度

异构加速型

AIGC风头正盛，AI算力需求日益增长

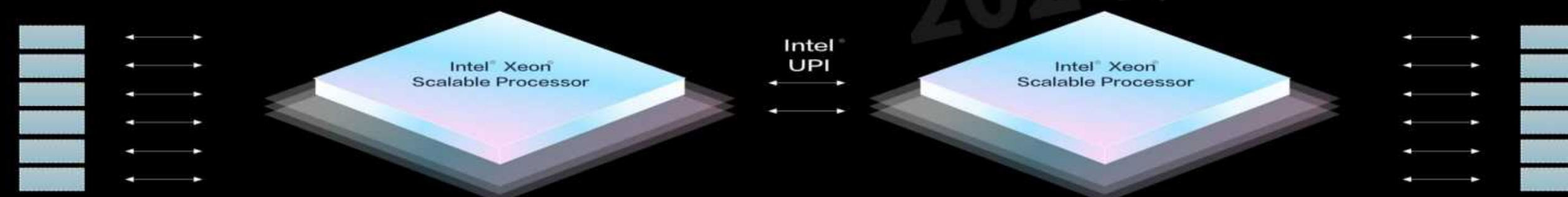
创新前IO型

前出线架构设计，增强数据中心单边维护效率

冷板液冷型

年PUE可降低至1.1-1.2，整机功耗下降可达70%

整机最大128核



整机最多24xNVMe SSD

24

x



云主机

裸金属

数据库

云电脑

云硬盘

算网原生算力基础设施 支撑全场景业务应用

基于ASIC芯片，研发新一代磐石DPU，结合一机多芯磐石服务器，打造磐石系列算网原生硬件体系

云主机

裸金属

数据库

云电脑

云存储

全场景

100G

4*25G网络接口

200G

2*100G网络接口

400G

2*200G网络接口

OpenStack

Kubernetes

Zabbix

Prometheus

Grafana

多形态

150M PPS

网络转发

4M IOPS

数据读写

400Gbps

网络带宽

5μs

RDMA时延

软件堆栈

x86架构



ARM架构

高性能

ASIC芯片

ARM+RISC-V+硬件pipeline

可编程

定制指令集

全卸载

计算、存储、网络、安全、管控

硬件平台

RDMA

SR-IOV

NVMe-oF

CT

QoS

VirtIO

PCC

P4

网络

存储

计算

管控

安全

新架构



加速引擎



全新算力基础设施 充分释放“算”能力

移动云第六代企业级计算产品

定制版新一代海光®处理器或第五代英特尔®至强®可拓展处理器

开启高效计算新纪元

磐石DPU 4.0

计算产品存储与网络IO性能加速器，为数据驱动型应用提供澎湃动力

RDMA弹性网络

为严苛网络性能应用场景而生，实现高吞吐量与低延时的完美融合

通用算力型云主机

专为中小型应用量身定制，以低成本高性能算力助力企业轻松上云

80%

计算型c6e
整机性能

2.9GHz

计算型c6e/c6he
主频

5200MT/s

计算型c6e
内存频率

55%

通用算力型u1
性价比

5000万

计算型c6e/c6he
最大网络转发

100万

计算型c6e/c6he
最大存储IOPS

200Gbps

智能计算型ci2k
RDMA网络带宽

5μs

智能计算型ci2k
RDMA网络延时

共建算网原生算力基础设施生态

一机多芯



全场景布局



期待与更多行业伙伴携手，推动移动云算网原生基础设施持续优化，助力算力网络发展

2024开放计算中国峰会

2024开放计算中国峰会

2024开放计算中国峰会
中国移动 China Mobile | 移动云

2024开放计算中国峰会

2024开放计算中国峰会

2024开放计算中国峰会

THANKS

2024开放计算中国峰会

2024开放计算中国峰会

2024开放计算中国峰会

2024开放计算中国峰会

2024开放计算中国峰会

2024开放计算中国峰会