打破能效与成本的边界：
全栈浸没式液冷方案的风液同价实践

上海淼算科技有限公司　副总经理　赵言亮

01

市场与客户痛点分析

02

单相浸没液冷技术优势与解决方案

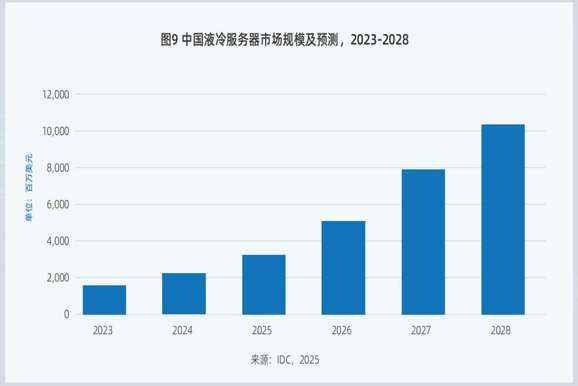
03

风液同价实践路径

01

市场与客户痛点分析

市场规模



来源：IDC 2025

目前***规模法则(Scaling law)正在从预训练扩展到了后训练和推理阶段***，基于强化学习、思维链等算法创新在后训练和推理阶段更多的算力投入，可以进一步大幅提升大模型的深度思考能力。同时基于杰文斯悖论的现象表明，***Deepseek带来的算法效率的提升并未抑制算力需求，反而因更多的用户和场景的加入，推动大模型普及与应用落地，重构产业创新范式，带动数据中心、边缘及端侧算力建设***。

人工智能大模型技术的研发和应用带来了更高的能耗需求，IDC数据显示，2024年全球人工智能数据中心IT能耗(含服务器、存储系统和网络)达到55.1太瓦时(TWh)，2025年将增至**77.7太瓦时**，是2023年能耗量的两倍，2027年将增长至**146.2太瓦时**，**2022-2027年五年年复合增长率为44.8%**，五年间实现六倍增长。2024年，中国占全球数据中心市场的比重约为 **47.2%**。

IT功耗的快速提升，叠加算力需求的爆发，使得**液冷从“可选项”变为“必选项”**。

暖通系统按照**5000元/IT kW**估算，液冷渗透率按30%计算，2027年国内液冷市场规模**1000亿元！**若同步考虑液冷服务器的价值，整个液冷市场将**远超千亿！**

国家政策支持液冷技术发展

中华人民共和国中央人民政府 www.gov.cn

关于印发《绿色数据中心政府采购需求标准（试行）》的通知

财库〔2023〕7号

党中央有关部门，国务院各部委、各直属机构，全国人大常委会办公厅，全国政协办公厅，最高人民法院，最高人民检察院，各民主党派中央，有关人民团体，各省、自治区、直辖市、计划单列市财政厅（局）、生态环境厅（局）、工业和信息化主管部门，新疆生产建设兵团财政局、生态环境局、工业和信息化局：

数字产业绿色低碳发展是落实党中央、国务院碳达峰、碳中和重大战略决策的重要内容。为加快数据中心绿色转型，财政部、生态环境部、工业和信息化部制定了《绿色数据中心政府采购需求标准（试行）》（以下简称《需求标准》），现就有关事项通知如下：

一、采购人采购数据中心相关设备、运维服务，应当有利于节约能源、环境保护和资源循环利用，按照《需求标准》实施相关采购活动。

二、采购人应当加强采购需求管理，根据《需求标准》提出的指标编制数据中心相关设备、运维服务政府采购项目的采购文件，并在合同中明确对相关指标的验收方式和违约责任。

中华人民共和国中央人民政府 www.gov.cn

工业和信息化部等六部门关于印发《算力基础设施高质量发展行动计划》的通知

工信部联通信〔2023〕180号

各省、自治区、直辖市通信管理局，各省、自治区、直辖市及新疆生产建设兵团工业和信息化主管部门、党委网信办、教育厅（局、教委）、卫生健康主管部门，中国人民银行上海总部、各省、自治区、直辖市分行，有关中央企业、中央金融企业，各有关单位：

现将《算力基础设施高质量发展行动计划》印发给你们，请结合实际，认真贯彻落实。

附件：算力基础设施高质量发展行动计划

工业和信息化部 中央网络安全和信息化委员会办公室
教育部 国家卫生健康委员会
中国人民银行 国务院国有资产监督管理委员会
2023年10月8日

中华人民共和国中央人民政府 www.gov.cn

国家发展改革委等部门
关于深入实施“东数西算”工程
加快构建全国一体化算力网的实施意见

发改数据〔2023〕1779号

各省、自治区、直辖市及计划单列市、新疆生产建设兵团发展改革委、数据局、网信办、工业和信息化主管部门、通信管理局、能源局：

算力是数字经济时代的新型生产力。算力网是支撑数字经济高质量发展的关键基础设施，可通过网络连接多源异构、海量泛在算力，实现资源高效调度、设施绿色低碳、算力灵活供给、服务智能随需。为深入贯彻落实党中央、国务院决策部署，加快构建全国一体化算力网，以算力高质量发展支撑经济高质量发展，制定本实施意见。

中华人民共和国中央人民政府 www.gov.cn

国家发展改革委等部门关于印发《数据中心绿色低碳发展专项行动计划》的通知

发改环资〔2024〕970号

各省、自治区、直辖市、新疆生产建设兵团发展改革委、工业和信息化主管部门、能源局、数据管理机构：

为全面贯彻党的二十大精神，认真落实党中央、国务院决策部署，实施全面节约战略，加大节能降碳工作力度，推动数据中心绿色低碳发展，加快节能降碳改造和用能设备更新，支撑完成“十四五”能耗强度降低约束性指标，国家发展改革委、工业和信息化部、国家能源局、国家数据局等部门制定了《数据中心绿色低碳发展专项行动计划》，现印发给你们，请认真抓好落实。

国家发展改革委
工业和信息化部
国家能源局
国家数据局
2024年7月3日

《绿色数据中心政府采购需求标准（试行）》

2023年03月20日

工业和信息化部　财政部　生态环境部

《算力基础设施高质量发展行动计划》

2023年10月08日

工业和信息化部 中央网络安全和信息化委员会办公室　教育部　国家卫生健康委 中国人民银行 国务院国资委

《深入实施“东数西算”工程加快构建全国一体化算力网的实施意见》

2023年12月25日

国家发展改革委 国家数据局 中央网信办 工业和信息化部 国家能源局

《数据中心绿色低碳发展专项行动计划》

2024年07月03日

国家发展改革委 工业和信息化部 国家能源局 国家数据局

客户痛点分析

数据中心

成本模型

- 成本：基建成本、运营成本、电力成本
- 收入：机柜机架出租、电力出售、带宽出售

需求

- 降低PUE：1.5 → 1.3 → 1.2
  自2026年起，北京市将对PUE值高于1.35的数据中心征收差别电价
- 提高面积使用率：机柜密度提升，单位面积租金提升

云服务商

成本模型

- 成本：IT设备电费、机房租赁费、带宽费用
- 收入：算力出租、其他增值服务（MaaS+SaaS）

需求

- 提高算效比：降低IT设备功耗，提高算效比（算力/功耗）
- 提高面积使用率：机柜密度提升，单位算力租金下降
- 提高设备可靠性：降低服务器故障率，提高SLA

解决方案

数据中心

1、单相浸没液冷技术可使PUE降低至1.15（注：考虑大机电的PUE损耗）

浸没液冷技术通过将服务器完全浸泡在冷却液中，彻底去掉服务器高功耗的风扇等零部件，并大幅简化暖通系统（去除了传统暖通系统中的冷水机组、泵房、冷冻站等设施，大幅缩减机房侧精密空调数量及功率），整个浸没液冷系统组成为浸没液冷机柜、CDU（热交换部件）、冷却塔等。

2、单相浸没液冷技术可使单柜密度最大提升至144kW（注：A级机房条件下）

单个浸没液冷机柜尺寸为1600mm*1200mm*2010mm（长宽高），占地面积为1.92m²，最多可放置12台睿启NVL16整机服务器（按A级机房2N冗余设计），单台服务器功耗最大为12kW，则单个浸没液冷机柜功耗为144kW。

传统风冷6kW机柜占地面积为0.6m²（600mm长*1000mm 宽），单位面积功耗为10kW/m²，而浸没液冷机柜单位面积功耗为75kW/m²，面积使用率提升7.5倍！

风冷环境下，机房选址条件严苛，位置偏远；浸没液冷环境下，机房可在任何位置。

云服务商

1、单相浸没液冷技术可减少单台服务器功耗至少600W（注：以NVIDIA 4090 八卡机为例）

以NVIDIA RTX4090八卡整机服务器为例，其风冷机型功耗约4.6-5kW，而对于浸没液冷机型，因去除了服务器风扇、电源风扇、GPU风扇等，同时服务器处于更低温的环境中，整机功耗仅4kW，单台服务器功耗至少下降13%（600W）。**以1000台服务器为例，假设电费为0.7元/度，则5年运营周期内，服务器电费节约费用为1840万元。**

2、单相浸没液冷技术可使服务器可靠性提升50%

对于高性能GPU服务器而言，传统风冷服务器的风扇长期处于高速运转中，易损坏，且高温易导致GPU出现降频现象，内存、硬盘和网卡寿命也会因为高温大幅减少，易出现故障告警，按照芯片10℃法则，温度每提高10℃，元器件寿命约降低50%。浸没液冷可以将服务器GPU核温最低控制在**70℃**以内，远低于风冷环境下**85℃**的核温。

风冷环境下，服务器的生命周期为5年；浸没液冷环境下，服务器的生命周期为8年。

PUE对比（单相浸没液冷VS风冷）

以3MW 机柜功率为例

冷冻水机组

◆ **传统风冷数据中心（CPU为主）功耗占比：IT设备（67%）+制冷（27%）+供配电（5%）+照明及其他（1%）**

$$PUE=\frac{总功耗}{IT功耗}=\frac{IT功耗+制冷功耗+电气功耗+照明及其他功耗}{IT功耗}=1+\frac{制冷功耗+电气功耗+照明及其他功耗}{IT功耗}=1+\frac{27\%+5\%+1\%}{67\%}=\mathbf{1.49}$$

PUE 1.49=1.0（IT设备）+0.4（制冷）+0.075（供配电）+0.015（照明及其他）

AHU（蒸发冷却机组）

◆ **新型风冷智算中心（GPU为主）功耗占比：IT设备（76%）+制冷（18%）+供配电（5%）+照明及其他（1%）**

$$PUE=\frac{总功耗}{IT功耗}=\frac{IT功耗+制冷功耗+电气功耗+照明及其他功耗}{IT功耗}=1+\frac{制冷功耗+电气功耗+照明及其他功耗}{IT功耗}=1+\frac{18\%+5\%+1\%}{76\%}=\mathbf{1.31}$$

PUE 1.31=1.0（IT设备）+0.237（制冷）+0.066（供配电）+0.013（照明及其他）

CDU+闭式冷却塔

◆ **浸没液冷智算中心（GPU为主）功耗占比：IT设备（86.5%）+制冷（9%）+供配电（4%）+照明及其他（0.5%）**

$$PUE=\frac{总功耗}{IT功耗}=\frac{IT功耗+制冷功耗+电气功耗+照明及其他功耗}{IT功耗}=1+\frac{制冷功耗+电气功耗+照明及其他功耗}{IT功耗}=1+\frac{9\%+4\%+0.5\%}{86.5\%}=\mathbf{1.15}$$

PUE 1.15=1.0（IT设备）+0.1（制冷）+0.046（供配电）+0.006（照明及其他）

PUE对比（单相浸没液冷VS风冷）

以3MW 机柜功率为例

CDU+闭式冷却塔

◆浸没液冷智算中心（GPU为主）功耗占比：IT设备（86.5%）+制冷（9%）+供配电（4%）+照明及其他（0.5%）

$$PUE=\frac{总功耗}{IT功耗}=\frac{IT功耗+制冷功耗+电气功耗+照明及其他功耗}{IT功耗}=1+\frac{制冷功耗+电气功耗+照明及其他功耗}{IT功耗}=1+\frac{9\%+4\%+0.5\%}{86.5\%}=\mathbf{1.15}$$

PUE 1.15=1.0（IT设备）+0.1（制冷）+0.046（供配电）+0.006（照明及其他）

注：PUE计算这块，因为浸没液冷服务器功耗比风冷服务器功耗低，即同样服务器数量条件下，PUE计算公式的分母（IT功耗），浸没版本比风冷版本低不少，PUE计算较为吃亏。

智算中心电气损耗的PUE

变压器 → UPS → 列头柜 →机柜PDU→ IT设备PSU → 芯片

10kV → 12V（经历多次交直流变化）

◆电气设备电能转换效率造成的PUE损失约0.045-0.08，叠加照明及其他造成的PUE损失，常规智算中心的PUE值理论最低1.1以上。

浸没液冷系统PUE < 1.1的情况

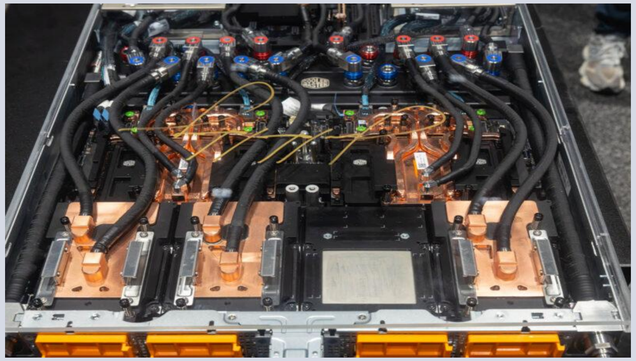
◆**pPUE。**即总电量取自列头柜或者小母线始端箱处，**仅考虑小机电**条件下的浸没液冷设备的PUE数值，未考虑大机电条件下的电量损耗；

◆**区块链。**（1）因区块链耗电量巨大，且多为集装箱部署形式，因此会采取**就近部署变压器和开闭所的形式进行供电**，对于电力可靠性的要求不如智算中心的要求高（智算中心多数要求电力可靠性要达到国标A级甚至更高），因此**电气架构更为简单**，电能损耗率低。（2）对于制冷架构来说，多数区块链项目所用服务器耐热度更高，因此常**使用干冷器直接进行散热，无CDU设计**。（3）区块链所用服务器**芯片常超频**使用，IT实际功耗比额定设计功耗更大，因此PUE值可以更低。

02

单相浸没液冷的技术优势与解决方案

液冷技术路线对比



单相冷板式液冷

- ✓**运维改动小：**服务器运维方式与风冷类似，机房运维人员上手速度快；
- ✓**兼容性好：** 醇类冷却液兼容广泛，且避免芯片及其他部件与液体直接接触产生兼容性问题；
- ✓**产业链成熟：**市场接受度高，技术经验积淀较多

- ✓**设备改动大：**风冷散热器改冷板，机箱开孔走软管，风冷机柜增加manifold和UQD；
- ✓**漏液风险高：**焊接点与接头较多，醇类冷却液易脏污
- ✓**造价成本高：**冷板服务器做加法，相比风冷服务器，增加了冷板套件、快接头、manifold等

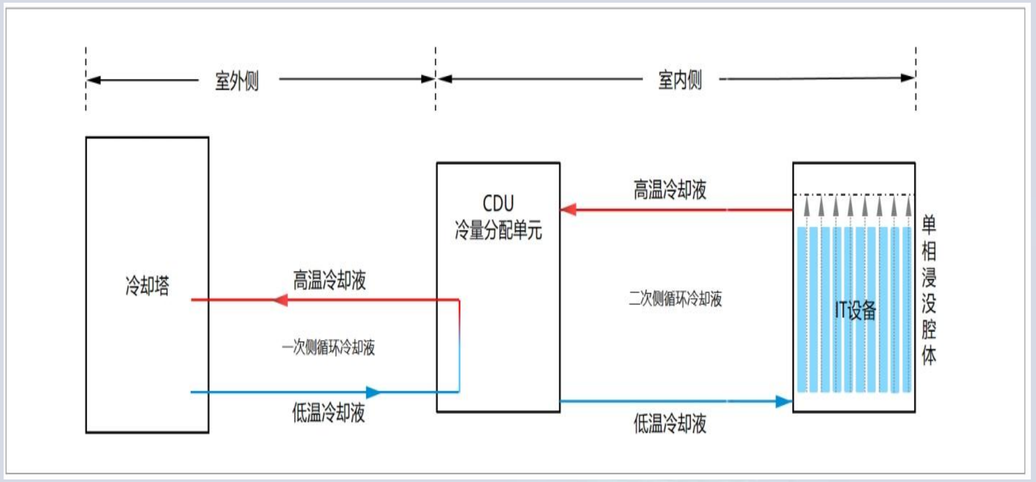
VS

单相浸没式液冷

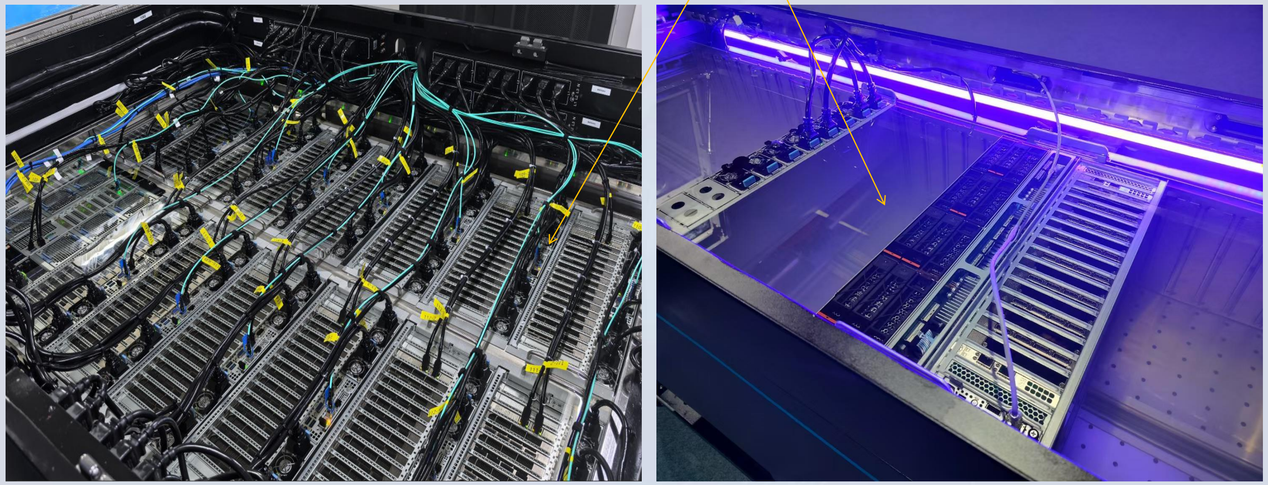
- ✓**设备改动小：**将风冷散热器改为浸没散热器（硅脂改为铟片），拆除风扇，BIOS及BMC更新等；
- ✓**部署密度高：**单柜功耗80kW无压力，最大功耗可达144kW（A级机房标准）；
- ✓**散热效果好：**芯片温度控制在70℃以内，故障率低；
- ✓**造价成本低：**浸没服务器做减法，合成油方案下，现场部署难度小，TANK无需严格密封，结构设计简单，承重12kN/m2，风冷机房结构无需更改，自然冷源利用率高（进出液温度40/50℃）。

- ✓**市场认知低：**落地的浸没液冷方案不多，市场认知低。

单相浸没液冷技术原理



单相浸没式液冷原理图



单相浸没式液冷服务器

单相浸没液冷技术优势——9个“减”

减 服务器结构 去风扇

- 去除GPU、电源、机箱风扇；
- 去除机箱外壳（睿启模块化服务器）

减 IT功耗 ↓ 600W

- 单台GPU服务器，浸没版本比风冷版本减少600W功耗。1000台5年节约电费1840万元。

减 PUE ↓ 1.2

- PUE可从1.5将至1.2以内，pPUE可降至1.1以内。

减 占地面积 ↓ 80%

- 同等算力条件下，浸没液冷占地面积较风冷下降80%。

减 机电规模 ↓15%

- 同等算力条件下，浸没液冷可节约15%以上的机电规模投资。1000台服务器机电节约1000万元。

减 运维频次 ↓85%

- 浸没液冷环境下，服务器可靠性大幅增加，硬件故障频次大幅下降（平均运维周期降低至1年2-3次）

减 部署时间 ↓75%

- 浸没液冷系统较架构较传统风冷架构大幅简化，平均施工周期从2个月缩短至2-3周

减 网络成本 ↓15%

- 浸没液冷机柜密度高，服务器数量多，网络设备布线距离短，成本低

减 投资造价 ↓30%

- 通过优化整体架构设计，达到初投资风液同价的目的，叠加运营成本较风冷降低30%以上，整体TCO降低30%

便宜的算力

智算中心：方案一 VasPod方案

单模块部署图

规模化部署图

1个TANK=百卡集群!

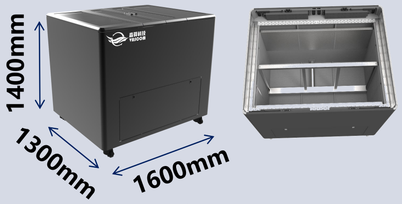
- **单模块设备组成：**列头柜*2（A/B路）、网络机柜*1、CDU*2（1+1冗余）、64U TANK*4，冷却液为**合成油**；
- **服务器数量：**CDU一拖四设计，单模块可提供4*64U空间，可满足**64台**4U服务器放置；
- **散热设计：**单个TANK散热量设计为**72kW**，通过高效流场换热设计，可满足高功率芯片散热需求，芯片核温控制在70℃以内；
- **占地面积：**较传统风冷方案**下降80%**；
- **承重要求：**机房结构承重要求**12kN/m2**，原风冷机房可直接利用；
- **高度要求：**机房净高要求不低于**3.1m**，原风冷机房可直接利用；
- **冷源要求：**室外冷源可选用闭式冷却塔，也可选用原风冷机房内的一次侧冷冻水系统
- **投资造价：****初投资风液同价（3MW IT功率规模下）**，三年运营成本较风冷**下降30%。 pPUE低至1.09。**

Tips:

- **场景1：**客户指定服务器形态，由我方采购相关液冷服务器，并提供服务器售后维保服务；
- **场景2**：客户已采购服务器，由我方提供风冷服务器液冷改造，并由我方提供该服务器的售后维保服务。

产品矩阵：智算中心 VasPod方案

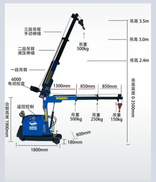
浸没液冷设备



浸没TANK

液冷CDU

运维车



清洗车

清洗剂

浸没液冷服务器

浸没服务器前端

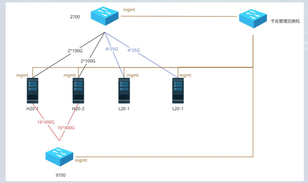
浸没服务器后端

液冷服务器维保

◆浸没服务器改造内容：

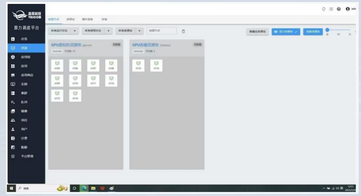
- 电源改造；
- 散热改造；
- 软件适配。

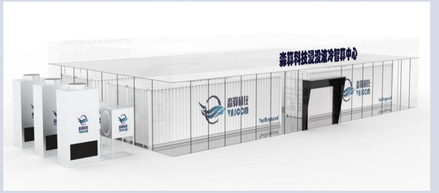
浸没液冷增值服务



IB/RoCE组网

液冷云管理平台


算力调度平台



IDC规划建设

智算中心：方案二 VasRack方案

单模块部署图

规模化部署图

1个TANK=百卡集群!

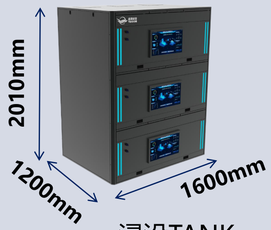
- **单模块设备组成：**列头柜*2（A/B路）、网络机柜*1、CDU*2（1+1冗余）、96U TANK*4，冷却液为**合成油**；
- **服务器数量：**CDU一拖四设计，单模块可提供4*96U空间，最多可满足**60-72台**8卡模块化服务器放置；
- **散热设计：**单个Rack散热量设计为**72kW-144kW不等**，通过高效流场换热设计，可满足高功率芯片散热需求，芯片核温控制在70℃以内；
- **占地面积：**较传统风冷方案**下降80%**；
- **承重要求：**机房结构承重要求**12kNg/m2**，原风冷机房可直接利用；
- **高度要求：**机房净高要求不低于**3.1m**，原风冷机房可直接利用；
- **冷源要求：**室外冷源可选用闭式冷却塔，也可选用原风冷机房内的一次侧冷冻水系统
- **投资造价：****初投资风液同价（3MW IT功率规模下）**，三年运营成本较风冷**下降30%。pPUE低至1.09。**

Tips:

◆ 服务器宽度可根据需求为2U-10U不等，具体跟客户选择8卡机型（双宽八卡PCIe版本）还是16卡机型（单宽16卡NVLINK版本），以及管理服务器和存储服务器有关。

产品矩阵：智算中心 VasRack方案

浸没液冷设备

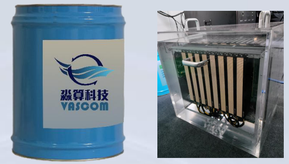


浸没TANK

液冷CDU

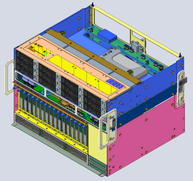
运维车

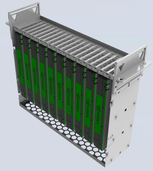
清洗车

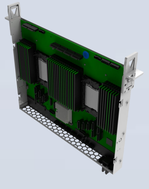

清洗剂

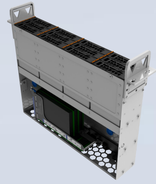
浸没液冷服务器

双宽8卡机型


单宽16卡机型

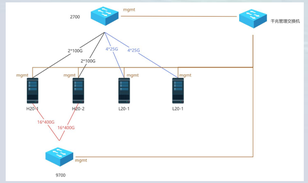

GPU模块


CPU模块

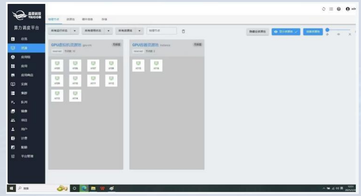

存储模块

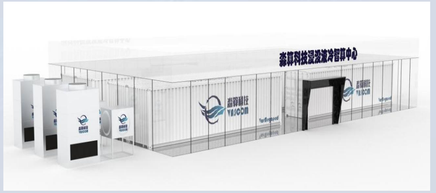
液冷服务器维保

浸没液冷增值服务


IB/RoCE组网

液冷云管理平台

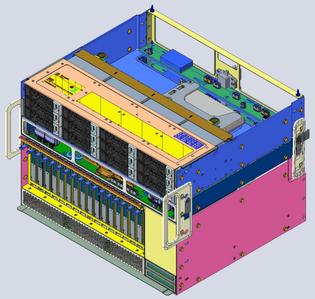

算力调度平台



IDC规划建设

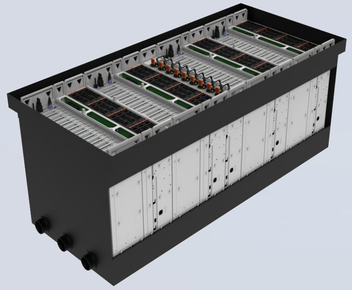
智算中心：方案二 VasRack方案 睿启模块化服务器（浸没液冷版本）

双宽8卡机型

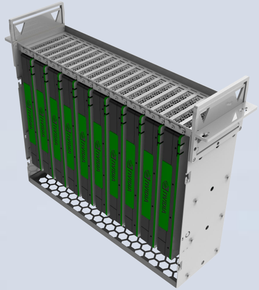


单宽16卡机型（NVL16）

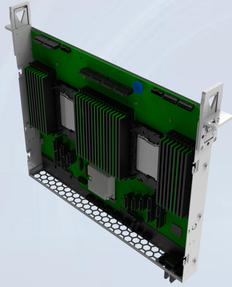
睿启模块化服务器，采用创新的积木式架构设计，将核心硬件（CPU、内存、GPU）独立模块化，支持客户根据需求自由选配与组合。产品以轻量化设计**（整机仅30kg，传统服务器约60kg）（服务器深度仅370mm，传统服务器为850mm）**和灵活扩展性为核心优势，结合Busbar集中供电技术，重新定义了高效、便捷的服务器运维体验。



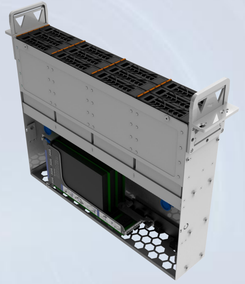
浸没服务器部署图



GPU模块



CPU模块



存储模块

370mm

优点

- **◆灵活自由：**打破传统服务器固化配置的限制，赋予用户硬件定义主动权。
- **◆降本增效：**轻量化设计降低物流与运维成本，Busbar供电提升能源利用率。
- **◆未来就绪：**模块化架构无缝适配技术演进，保护客户长期投资价值。

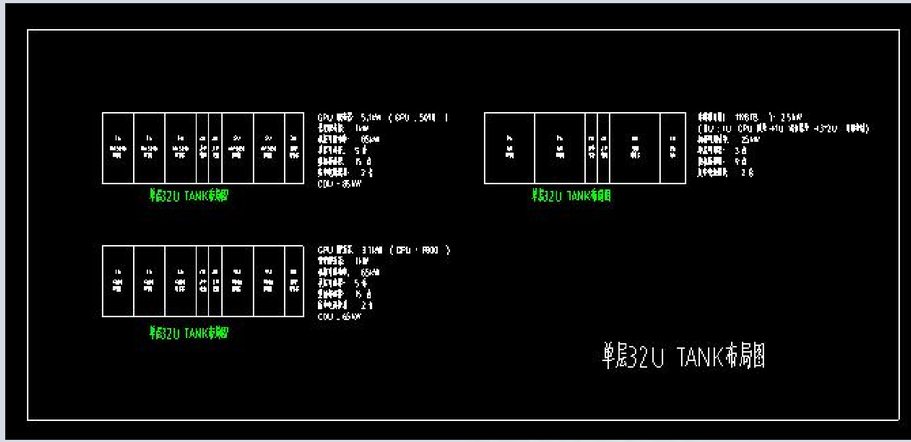
智算中心：方案二 VasRack方案 睿启模块化服务器（浸没液冷版本）

序号	模块种类	模块描述	模块高度
1	CPU模块1	2个Intel EagleStream CPU	1U
2	CPU模块2	1个Intel BirchStream AP CPU，2个PCIe插槽，2个M.2 SSD，1U	1U
3		AMD Turin or Genoa CPU, 2个PCIe插槽，2个M.2 SSD，1U	1U
4		Ampere OneX CPU,2个PCIe插槽，2个M.2 SSD，1U	1U
5	CPU模块3	2个Intel EagleStream CPU AI专用主板，1U	1U
6	CPU模块4	2个AMD Turin or Genoa CPU ，AI专用主板，1U	1U
7	存储模块1	12个3.5英寸插槽，可以支持SATA、SAS、NVMe 2U	2U
8	存储模块2	24个2.5英寸插槽，可以支持SATA、SAS、NVMe 2U	2U
9	GPU模块1	13个PCIe插槽，可以支持8个双宽GPU，5个网卡，含2个144Lans PCIe SW,3U	3U
10		12个PCIe插槽，可以支持8个双宽GPU，4个网卡，直通式,3U	3U
11	GPU模块2	支持16个单宽NVL16 GPU，4U	4U
12	混合模块1	支持6个全高卡 或 6个2.5寸盘 或 2盘4卡 或 4盘2卡，1U	1U
13	混合模块2	支持6个全高卡和4个2.5寸盘，2U	2U
14	电源模块	10个CRPS电源， 最大支持30KW,2U	2U

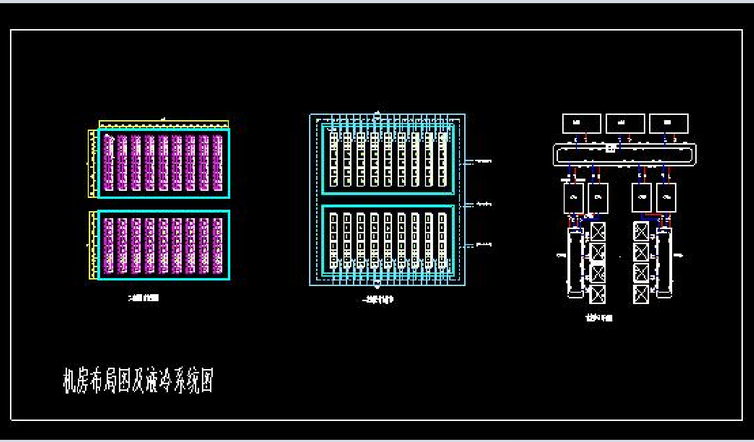
智算中心：方案二 VasRack方案 睿启模块化服务器（浸没液冷版本）

序号	服务器种类	模块1	数量	模块2	数量	模块3	数量	模块4	数量	服务器配置	高度
1	高性能计算服务器	CPU模块2	1							单路 AMD or Intel处理器，2个PCI插槽，2个M.2 SSD	1U
2	通用服务器1	CPU模块1	1	混合模块1	1					双路Intel至强处理器，4个PCIe插槽，2个U.2SSD	2U
3	通用服务器2	CPU模块1	1	混合模块1	1	存储模块1	1			双路Intel至强处理器，6个PCIe插槽，12个3.5寸	4U
4	通用服务器3	CPU模块1	1	混合模块1	1	存储模块2	1			双路Intel至强处理器，6个PCIe插槽，24个U.2	4U
5	通用服务器4	CPU模块2	1	混合模块1	1	存储模块1	1			单路 AMD or Intel处理器，8个PCIe插槽，12个3.5寸	4U
6	存储服务器（24盘）	CPU模块2	1	混合模块1	1	存储模块1	2			单路 AMD or Intel处理器，8个PCIe插槽，24个3.5寸	6U
7	存储服务器（36盘）	CPU模块2	1	混合模块1	1	存储模块1	3			单路 AMD or Intel处理器，8个PCIe插槽，36个3.5寸	8U
8	存储服务器（48盘）	CPU模块2	1	混合模块1	1	存储模块1	4			单路 AMD or Intel处理器，8个PCIe插槽，48个3.5寸	10U
9	AI推理服务器（单路）	CPU模块2	1					GPU模块1	1	单路 AMD or Intel处理器，6个双宽GPU，2个M.2 SSD	4U
9	AI推理服务器	CPU模块1	1			存储模块1	1	GPU模块1	1	双路Intel至强处理器，8个双宽GPU 直通或PCIe SW，12个3.5寸	6U
10	AI训练服务器	CPU模块3	1			存储模块1	1	GPU模块2	1	双路Intel至强处理器，16个单宽GPU NVL16，12个3.5寸	7U

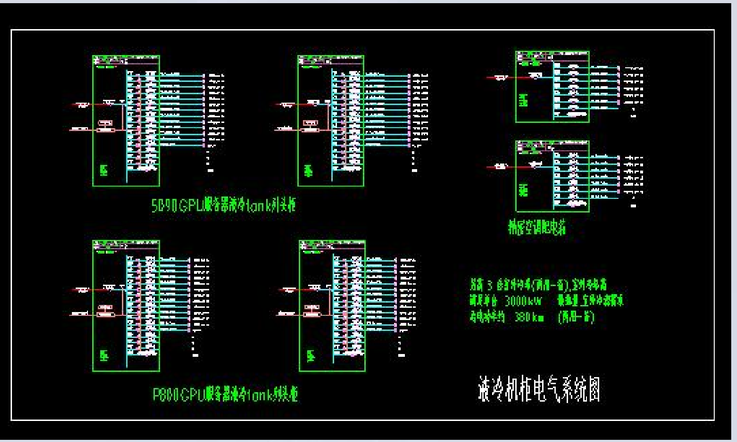
全浸没液冷方案设计



浸没服务器布局图
（NV+国产+管理+存储）

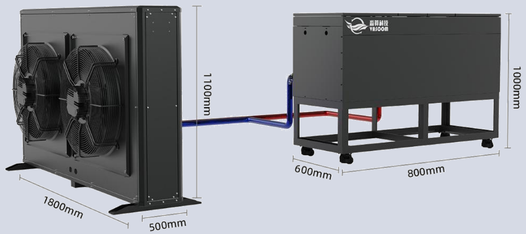


浸没暖通系统图



浸没电气系统图

边缘计算：浸没一体机方案



干冷器

单层浸没式液冷TANK

极速部署　极致静音　风液同价

淼算科技(VASCOM)联合睿启服务器(比亚迪电子旗下服务器品牌)推出针对本地化部署的漫没式液冷静音机柜，专为中小型企业、初创企业、科研高校等打造，集高效散热、超静音运行、即装即用于一体，实现**“风液同价、性能更优”**的突破。产品采用单层浸没式液冷架构，搭配睿启模块化服务器轻量型机身、简易安装，无需专业团队即可快速部署，**4小时内可安装开机运行**，有效解决传统服务器设备噪音大、散热难、安装复杂的问题，轻松享受极速、稳定、安静的计算体验。

Deepseek满血版本

GPU 服务器：1 台
CPU：2*Intel 8480+ 处理器
内存：32*64GB DDR5 4800MHz(2048GB)
系统盘：2*960G SATA SSD
数据盘：4*3.84T NVMe SSD
网卡：**16*400G BF3网卡（内置）**+1*200G BF3网卡 +1* 板载 RJ45 千兆网口
GPU：16*H20 120GB 浸没式
集中式电源：8*3200W CRPS AC/240HVDC
大存储管理服务器：1 台
CPU：2*Intel 6426Y(2.5~3.3G/37.5M/16C/32T/185W)
内存：8 * 32GB DDR5（256GB）
系统盘：2* 480GB SATA SSD
存储盘：10* 3.84T U.2 NVMe SSD
网卡：1*200Gb IB NDR/Ethernet 适配卡
网卡：1* 双口 25GE 光接口 OCP3.0 网卡
BMC 管理板模块

Deepseek-R1-32B

GPU 服务器：1 台
CPU：2*Intel Xeon Gold 6530 32C/64T 2.1GHz 160MB 270W
内存：16*32GB DDR4 (512GB)
系统盘：2*480G SATA SSD
数据盘：2*3.84T NVMe SSD
网卡：1* 双口 25GB 网卡 +1* 双口板载 RJ45 千兆网口
GPU：2*L20 48GB 浸没式
集中式电源：2*3200W CRPS AC/240HVDC
大存储管理服务器：1 台
CPU：2*Intel6426Y(2.5~3.3G/37.5M/16C/32T/185W)
内存：4 * 32GB DDR4（128GB）
系统盘：2* 480GB SATA SSD
存储盘：10* 3.84T U.2 NVMe SSD
网卡：1* 双口 25GE 光接口 OCP3.0 网卡
BMC 管理板模块

边缘计算：浸没一体机方案

方案特点

该TANK与传统TANK设计不同，通过PowerShelf（集中电源）与外部配电系统相连后，将380V/220V电压转为54V电压集中供应给TANK底部的busbar供电系统，再通过busbar与模块化服务器底部的PowerClip（电源取电夹子）相连，将54V电压的电供应至各芯片使用。

优势亮点

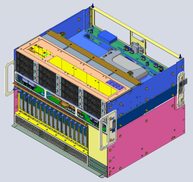
- ◆**高度大幅降低：**TANK高度将由传统的1.2-1.4m下降至0.6-1m以内，可不借助运维车，单人直接上下架服务器，大大方便运维人员运维，亦可本地化部署；
- ◆**支持弹性扩容：**根据客户需求，可对CPU、GPU、存储模块进行任意搭配，并可将GPU服务器、通用服务器、存储服务器均放置于TANK内；
- ◆**本地化部署便捷：**无CDU设计，TANK内冷却液经过循环泵的驱动直接至室外干冷器内进行散热，系统部署简便，且支持TANK叠加部署。

产品矩阵：边缘浸没一体机方案

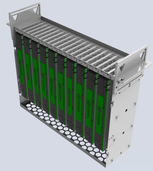
浸没液冷设备

边缘TANK

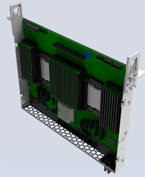
双宽8卡机型



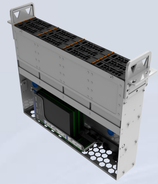
单宽16卡机型



GPU模块



CPU模块



存储模块

MaaS+SaaS

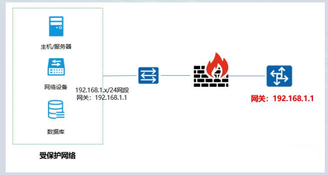
知识库

稠密模型
（阿里Qwen）
32B以下 微调

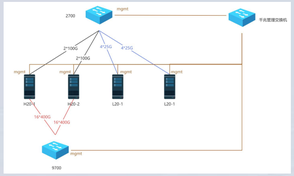
智能问答

MoE模型
（阿里Qwen/DeepSeek）
671B高智商

浸没液冷增值服务



安全防护设备



局域网搭建

UPS选型与安装

客户案例一：Telstra（太平洋电信）天津武清机房一期浸没液冷改造工程

浸没液冷改造全景图

16台（128卡）浸没液冷服务器部署图

客户&机房信息

太平洋电信股份有限公司成立于2008年，是国家级高新技术企业，也是国内唯一拥有互联网数据中心、国内互联网虚拟专用网、互联网接入服务、国内多方通信服务、云协作等全面的通信业务牌照的IDC企业。

该数据中心位于京津高村科技创新园，是太平洋电信有限公司在天津打造的首个世界级软件定义网络（SDN）数据中心，按照**国际T3+级别**设计。项目占地300亩，建筑面积38万平方米，包括占地160亩的四座数据中心和占地140亩的研发总部办公大楼，总投资100亿元。

一期液冷改造内容

◆提供一套浸没液冷系统（冷却液为**合成油**，冷源利用机房原冷冻水系统）；
◆提供**128卡高性能浸没式液冷GPU服务器及其维保服务**；
◆提供25G RoCE组网及算力调度平台服务；
◆提供液冷系统云管理平台服务（远程监控、智能运维、故障提醒等）

稳定运行半年
服务器0故障报警

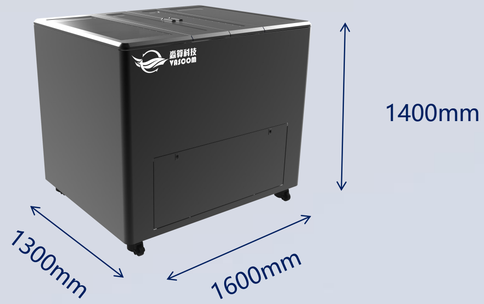
预期目标

➢ pPUE降至1.2以内（因冷源为客户原有的冷冻水系统，故PUE值较高）；
➢ 服务器故障率降低50%以上，服务器功耗下降10%以上；
➢ 机房噪音值降至50dB以下；

客户案例一：Telstra（太平洋电信）天津武清机房一期浸没液冷改造工程

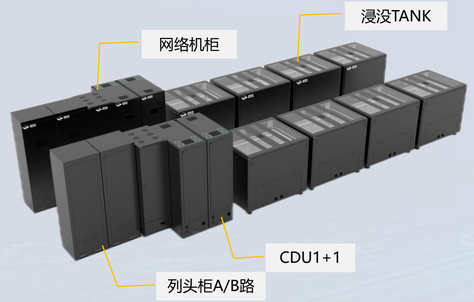
浸没液冷改造全景图

16台（128卡）浸没液冷服务器部署图



单TANK占地面积**2平米**

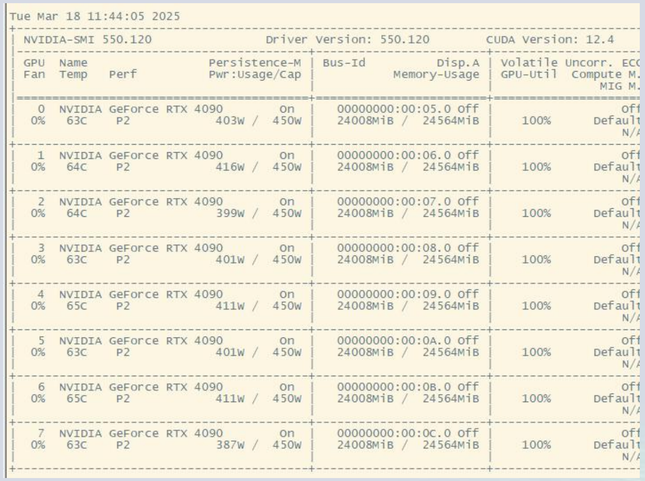
1*TANK=百卡集群（128卡）！



千卡集群占地面积**80平米**
（含运维空间）

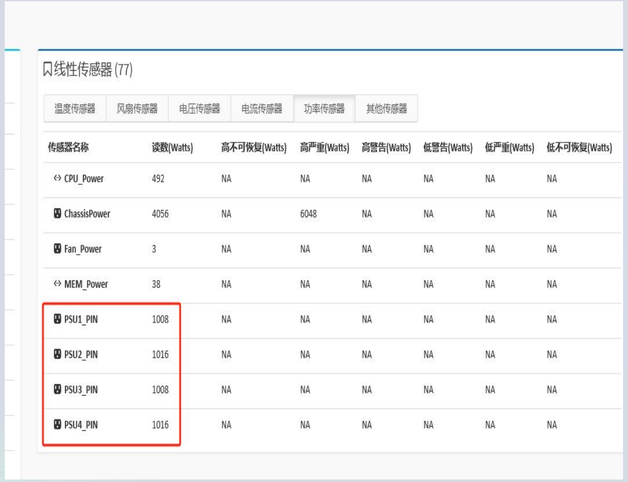
8*TANK=千卡集群（1024卡）！

客户案例一：Telstra（太平洋电信）天津武清机房一期浸没液冷改造工程



```
Tue Mar 18 11:44:05 2025
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 550.120                Driver Version: 550.120      CUDA Version: 12.4 |
|-------------------------------+----------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id          Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |           Memory-Usage | GPU-Util  Compute M. |
|                                         |                        |               MIG M. |
|===============================+======================+======================|
|   0  NVIDIA GeForce RTX 4090      On  | 00000000:00:05.0 Off |                  Off |
|  0%   63C    P2             403W / 450W |  24008MiB / 24564MiB |     100%      Default |
|                                         |                        |                  N/A |
|-------------------------------+----------------------+----------------------|
|   1  NVIDIA GeForce RTX 4090      On  | 00000000:00:06.0 Off |                  Off |
|  0%   64C    P2             416W / 450W |  24008MiB / 24564MiB |     100%      Default |
|                                         |                        |                  N/A |
|-------------------------------+----------------------+----------------------|
|   2  NVIDIA GeForce RTX 4090      On  | 00000000:00:07.0 Off |                  Off |
|  0%   64C    P2             399W / 450W |  24008MiB / 24564MiB |     100%      Default |
|                                         |                        |                  N/A |
|-------------------------------+----------------------+----------------------|
|   3  NVIDIA GeForce RTX 4090      On  | 00000000:00:08.0 Off |                  Off |
|  0%   63C    P2             401W / 450W |  24008MiB / 24564MiB |     100%      Default |
|                                         |                        |                  N/A |
|-------------------------------+----------------------+----------------------|
|   4  NVIDIA GeForce RTX 4090      On  | 00000000:00:09.0 Off |                  Off |
|  0%   65C    P2             411W / 450W |  24008MiB / 24564MiB |     100%      Default |
|                                         |                        |                  N/A |
|-------------------------------+----------------------+----------------------|
|   5  NVIDIA GeForce RTX 4090      On  | 00000000:00:0A.0 Off |                  Off |
|  0%   63C    P2             401W / 450W |  24008MiB / 24564MiB |     100%      Default |
|                                         |                        |                  N/A |
|-------------------------------+----------------------+----------------------|
|   6  NVIDIA GeForce RTX 4090      On  | 00000000:00:0B.0 Off |                  Off |
|  0%   65C    P2             411W / 450W |  24008MiB / 24564MiB |     100%      Default |
|                                         |                        |                  N/A |
|-------------------------------+----------------------+----------------------|
|   7  NVIDIA GeForce RTX 4090      On  | 00000000:00:0C.0 Off |                  Off |
|  0%   63C    P2             387W / 450W |  24008MiB / 24564MiB |     100%      Default |
|                                         |                        |                  N/A |
+-------------------------------+----------------------+----------------------+
```

GPU满载运行

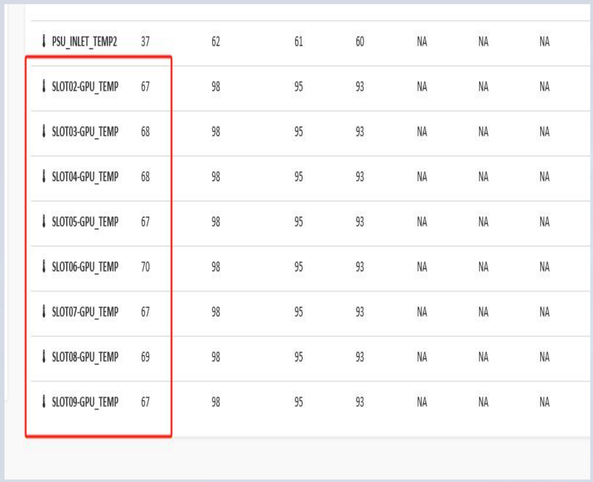


线性传感器 (77)

温度传感器 | 风扇传感器 | 电压传感器 | 电流传感器 | 功率传感器 | 其他传感器

传感器名称	读数(Watts)	高不可恢复(Watts)	高严重(Watts)	高警告(Watts)	低警告(Watts)	低严重(Watts)	低不可恢复(Watts)
CPU_Power	492	NA	NA	NA	NA	NA	NA
ChassisPower	4056	NA	6048	NA	NA	NA	NA
Fan_Power	3	NA	NA	NA	NA	NA	NA
MEM_Power	38	NA	NA	NA	NA	NA	NA
PSU1_PIN	1008	NA	NA	NA	NA	NA	NA
PSU2_PIN	1016	NA	NA	NA	NA	NA	NA
PSU3_PIN	1008	NA	NA	NA	NA	NA	NA
PSU4_PIN	1016	NA	NA	NA	NA	NA	NA

浸没整机服务器功耗 **4kW**
（同配置下的风冷服务器功耗约**4.6kW**）



PSU_INLET_TEMP2	37	62	61	60	NA	NA	NA
SLOT02-GPU_TEMP	67	98	95	93	NA	NA	NA
SLOT03-GPU_TEMP	68	98	95	93	NA	NA	NA
SLOT04-GPU_TEMP	68	98	95	93	NA	NA	NA
SLOT05-GPU_TEMP	67	98	95	93	NA	NA	NA
SLOT06-GPU_TEMP	70	98	95	93	NA	NA	NA
SLOT07-GPU_TEMP	67	98	95	93	NA	NA	NA
SLOT08-GPU_TEMP	69	98	95	93	NA	NA	NA
SLOT09-GPU_TEMP	67	98	95	93	NA	NA	NA

GPU芯片温度**70℃**以内

服务器配置：

CPU：2*Intel 6530
内存：8*64G DDR5
系统盘：1*960G SSD
数据盘：1*3.84T NVMe SSD
网卡：1*双口25G+1*双口千兆
电源：2700W 白金电源（2+2）
GPU：8*NVIDIA RTX4090 24G
主板架构：直通

浸没服务器配置

备注：按1000台服务器计算，5年生命周期内浸没液冷服务器减少的电费为（电价按0.7元/度）：1840万元

客户案例二：北京蓝厅云景延庆机房一期浸没液冷建设工程

浸没液冷部署实物图

VasRack平时状态

VasRack运维状态

客户&机房信息

位于北京延庆的新一代智算中心项目是由蓝厅云景集团投资的**北京市重点项目**，旨在建设京津冀地区乃至全国的大数据和智能算力核心枢纽。该项目以**国际T3+标准**进行建设，拟建设万卡算力集群，一期总发电量100MW。该项目为**北京首个绿电示范区**，园区内企业使用 100%绿电。项目占地面积约44.9 亩，总建设面积为 6.1万 平方米。

本项目一期土地、建设基础设施总投资约20亿元，已于2024 年底全部交付。算力服务器、网络组网建、算力云平台搭建等投资约80 亿元(2.5万算力卡规模，2万P算力能力)。

一期液冷改造内容

◆提供一套浸没液冷系统（冷却液为合成油，冷源为自建闭式冷却塔）；
◆提供**15台模块化浸没式液冷GPU服务器及其维保服务**；
◆提供400G IB 组网及算力调度平台服务；
◆提供液冷系统云管理平台服务（远程监控、智能运维、故障提醒等）

预期目标

➢ pPUE降至1.1以内（可全年利用自然冷源）；
➢ 服务器故障率降低50%以上，服务器功耗下降10%以上；
➢ 机房噪音值降至40dB以下；

03

风液同价的实践路径

如何实现风液同价（初投资）

合成油冷却液

- 合成油相比氟化液成本大幅降低；
- 合成油不挥发，无需定时补液，TANK设计也无需高度密封，降低设计制造成本；
- 液冷机房结构承重与风冷一致

高密度部署

- 1个TANK即百卡集群（128卡），占地仅不到2平米；
- 单位面积功耗最大可达75kW/m2；
- CDU一拖四设计，高密度部署；

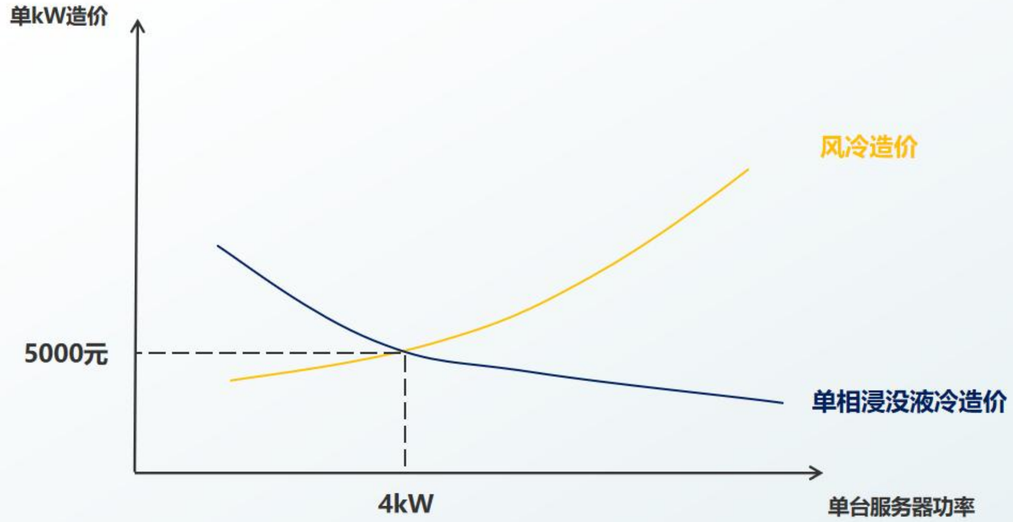
与源头厂家合作

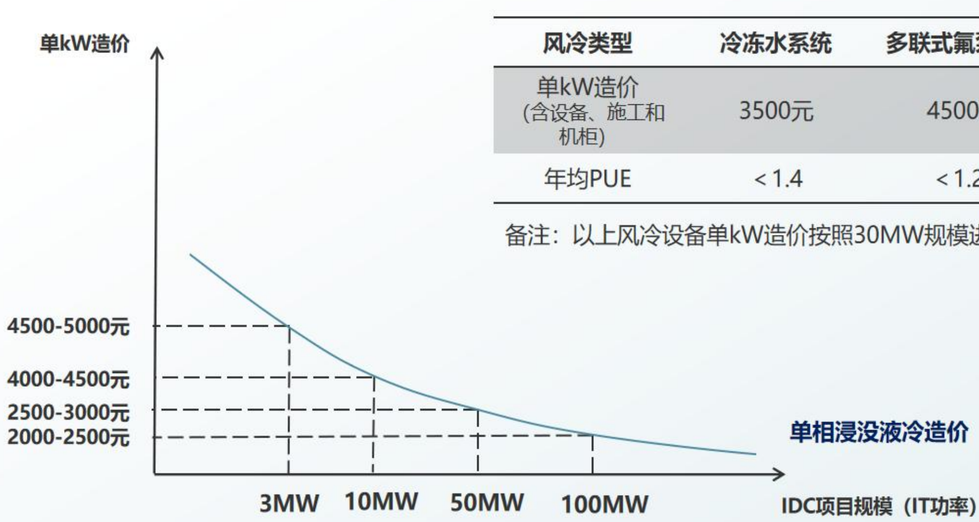
- 浸没液冷服务器出厂时，即取消风扇安装；
- GPU显卡组装时，即按照液冷散热器组装，降低整体改造成本

大幅降低机电投资

- 以1000台NV 4090八卡服务器为例（DDR5平台），风冷条件下，机电系统配电按1000*4.6=4.6MW IT功率考虑。而浸没液冷条件下，机电系统配电按1000*4=4MW IT功率考虑。
- **机电成本平均按照1.7万/IT kW估算，机电成本节约 1.7*600=1020万元。**

风液同价(初投资)





风冷类型	冷冻水系统	多联式氟泵系统	AHU间接蒸发冷
单kW造价（含设备、施工和机柜）	3500元	4500元	3000元
年均PUE	< 1.4	< 1.25	< 1.3

备注：以上风冷设备单kW造价按照30MW规模进行测算。

备注：服务器功率4kW指浸没液冷环境下的功率，对应风冷环境下的服务器功率为4.6kW（典型服务器型号为NVIDIA RTX 4090 8卡服务器）

备注：以上单kW造价包括全套浸没液冷设备、合成油、清洗剂、运维设备、运输、调试、施工安装费等

多一种液冷选择

便宜的算力

难而正确的路，我们一直在不断前行

THANKS

公众号

视频号

电话：+86-136 3660 3945

业务咨询：contact@vastcompute.net

地址：上海市宝山区沪太路2388号贺海创新中心4层409