

### Content

### AI 系统 + 大模型全栈架构图 时事 RAG 具身智能 自动驾驶 生成式推荐 智能体 RAG 热点 07. 模型应用 & 热点 大模型& 06. 模型算法与数据 AI 应用 视觉大模型 视频生成大模型 文生图大模型 多模态大模型 语言大模型 6. 大模型数据&算法 LLM VLM SoRA (ViT) Stable Diffusion MLM 大模型 数据&模型评估 大模型算法 TERESTA 加速库& 训推 Prompt 工程,模 Scaling Law, 三方框架 型评估算法和测评 Transform 结构, Megatron-LM DeepSpeed Colossal-Al HuggingFace LLamaIndex AutoChain RAGFlow Milvus LangChain 体系 LLM/MLM 模型 Caffe AI 训练& 4. 计算架构 推理框架 PyTorch JAX Caffe TensorFLow MindSpore VLLM MNN Xinference 编译 传统编译器 AI 编译器 前端优化 04. AI 训练 05. AI 推理 计算 传统编译器 AI编译器发展与 前端优化(算 架构 GCC NVIDIA. CUDA. 8 GCC与LLVM 架构定义, 未来 子融合、内存 CANN AI 编译& 挑战与思考 优化等) 计算架构 TVM GLOW XLA TC Triton LLVM GCC CUDA Al编译器 传统编译器 集群管理运维 3. AI 集群 田 Q K8s集群运维、K8s容器、集群监控等工具 Extra POD 超节点 Super POD K8s ubuntu 硬件 等操作系统 容器化部署 AI 硬件& 体系 I. AI 芯片 体系结构 03. AI 集群 结构 AI 计算体系 AI 芯片基础 英伟达GPU 2 E@ AI 计算模式 CPU, GPU, 英伟达GPU 与计算体系架 NPU等芯片体 TensorCore, TPU CPU RDMA IB Ethernet нвм 存储阵列 构 基础原理 NVLink剖析 01. AI 芯片 (算) 02. 互联 (网) 02. 存储 (存)



### Content

- 1. 风冷机柜 Air cooled cabinet
- 2. 板冷机柜 Board cooling cabinet
- 3. 全液冷机柜 Fully liquid cooled cabinet







### 3D 爆炸图大品鉴

- <a href="https://info.support.huawei.com/computing/server3D/index\_zh.html">https://info.support.huawei.com/computing/server3D/index\_zh.html</a>
- https://servers.asus.com/products/servers/
- https://www.gigabyte.com/Enterprise/GPU-Server





### 风冷机柜:外部结构

外部结构:包括机柜、电源框、电池框、计算节点、交换节点、线缆转接托盘、假面板组成



编码	部件名称
1	风冷机柜
2	管理模块
3	电源框
4	电池围框
5	电池框
6	假面板
7	计算节点
8	交换节点

# 风冷机柜: 内部结构

• 包括 Cable 背板,提供节点网络传输;Busbar 提供节点电力传输

# 校冷机连

### 3D 爆炸图大品鉴

- <a href="https://info.support.huawei.com/computing/server3D/index\_zh.html">https://info.support.huawei.com/computing/server3D/index\_zh.html</a>
- https://servers.asus.com/products/servers/
- https://www.gigabyte.com/Enterprise/GPU-Server



# 板冷机柜:外部结构

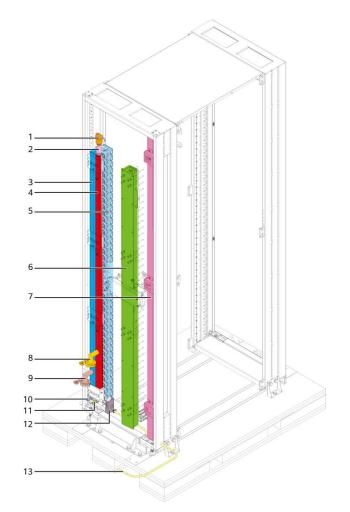
板冷服务器的机柜使用风冷机柜,增加 Manifold 集流管等组件



编码	部件名称
1	风冷机柜
2	管理模块
3	电源框
4	电池围框
5	电池框
6	假面板
7	计算节点
8	交换节点

### 板冷机柜: 内部结构

• 包含Manifold 集流管、Cable背板、Busbar,及漏液传感器,液冷供水、回水、排水等管路。



编码	部件名称
1	Manifold 排气阀
2	Manifold 球阀
3	Manifold 供水管
4	Manifold 回水管
5	导流管
6	Cable 背板
7	Busbar
8	Manifold 回水管接口
9	Manifold 供水管接口
10	Manifold 供水管液冷调测接头
11	Manifold 回水管液冷调测接头
12	光电式漏液传感器
13	Manifold 排水管

全液冷机柜

### 3D 爆炸图大品鉴

- <a href="https://info.support.huawei.com/computing/server3D/index\_zh.html">https://info.support.huawei.com/computing/server3D/index\_zh.html</a>
- https://servers.asus.com/products/servers/
- https://www.gigabyte.com/Enterprise/GPU-Server



### 全液冷机柜: 前视图

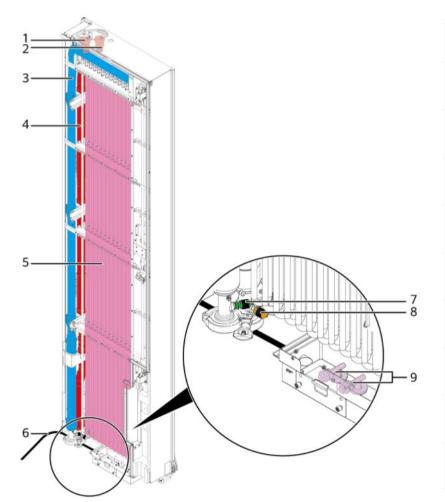
• 全液冷机柜比板级液冷外部组件区别,多一个风液换热器,其他结构一致(KunTai PoD 7210)



编码	部件名称
1	液冷机柜
2	风液换热器
3	电源模组 (电源框、电池围框、电池模块)
4	管理面交换机节点
5	计算节点
6	参数面交换机节点
7	业务面交换机节点
8	线缆转接托盘

### 全液冷机柜: 风液换热器组件。 https://support.huawei.com/enterprise/zh/doc/EDOC1100313980/cb51627f

• 风液换热器组件包括冷热水管、排气阀、冷热水管针阀、换热芯体、排水管、浮子漏液传感器



编码	部件名称
1	风液换热器冷水管排气阀
2	风液换热器热水管排气阀
3	风液换热器冷水管
4	风液换热器热水管
5	换热芯体
6	风液换热器排水管
7	风液换热器冷水管针阀
8	风液换热器热水管针阀
9	浮子式漏液传感器

# 全液冷服务器区别对比

https://support.huawei.com/enterprise/zh/doc/EDOC1100313980/cb51627f

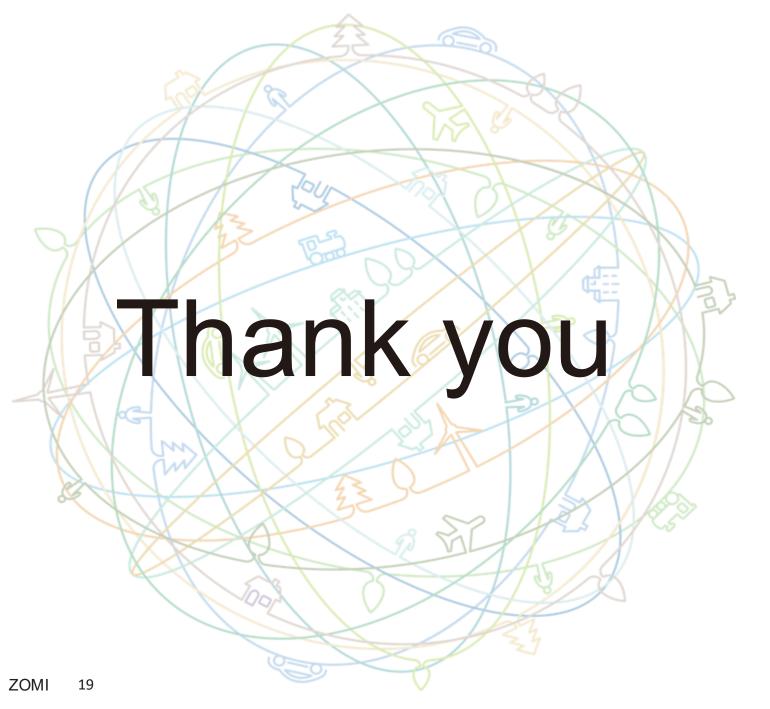
模块	传统服务器	AI 服务器
结构	电源框 + 服务器 + 交换机 + 线缆转接托盘	电源框 + 服务器 + 交换机 + 线缆转接托盘
	<ul><li>X86 架构处理器计算节点</li><li>支持 CPU + 液冷散热</li></ul>	<ul><li>ARM 架构处理器</li><li>支持 XPU + CPU 液冷散热</li></ul>
交换节点	<ul><li>风冷散热</li><li>线缆转接托盘</li><li>服务器节点数据交换</li></ul>	<ul><li>液冷散热模块</li><li>支持主芯片 + CDR + 光模块液冷散热</li><li>Cable 背板服务器节点数据交换,不需要连接线缆</li></ul>
液冷管路	服务器节点Manifold盲插液冷管路	服务器节点和交换节点Manifold盲插液冷管路

ZOMI

# 总结与思考

# 机柜形态对比

	风冷	板冷	全液冷
工作原理	通过风扇或空调将冷空气吹入服务器内部, 带走热量并排出热空气	使用冷板(覆盖CPU、GPU等高功耗部件)通过 液体循环散热,其余部件仍依赖风冷辅助	所有发热部件(如CPU、内存、电源、SSD等)均 通过液冷系统直接散热,无需风扇
散热效率	散热效率较低,依赖空气比热容 (比热容小)	散热效率较高,液体比热容大且可精准冷却高热 部件	散热效率最高,液体直接接触所有发热部件
PUE值	PUE较高(通常为1.5-2.0),依赖传统空调	PUE<1.1 (纯液冷) 或<1.2 (风液混合)	PUE<1.1 (浸没式液冷) 或更低
能耗	能耗高 (空调和风扇功耗大)	能耗显著降低(空调能耗降低70%,服务器风扇 功耗降低70%-80%)	能耗最低 (完全取消风扇,液冷系统效率更高)
维护成本	维护成本低(技术成熟,无需特殊防护)	维护成本中等 (需处理液冷管路泄漏风险)	维护成本高(需专业维护液体循环系统,防止泄漏)
适用场景	适用于低密度部署(传统数据中心)	适用于中高密度部署 (兼顾兼容性和散热能力)	适用于超大规模/高密度部署(如AI、超算中心)
兼容性	完全兼容现有服务器架构	基本兼容风冷设计(硬盘、光模块无需改动)	需定制化服务器架构(取消风扇,增加液冷接口)
噪音水平	噪音高 (风扇和空调运转)	噪音中等 (风扇减少,液冷泵噪音较低)	噪音极低 (无风扇运转)
初期建设成本	成本低 (无需复杂冷却系统)	成本较高 (需安装液冷管路和冷板)	成本最高 (需全系统液冷改造)



把 Allnfra 带入每个开发者、每个家庭、 每个组织,构建万物互联的智能世界

Bring AI Infra to every person, home and organization for a fully connected, intelligent world.

Copyright © 2025 Infrasys-Al org. All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Infrasys-Al org. may change the information at any time without notice.



**GitHub** github.com/Infrasys-AI/AIInfra Book infrasys-ai.github.io

### 引用与参考

- 1. <a href="https://navitassemi.com/nvidias-grace-hopper-runs-at-700-w-blackwell-will-be-1-kw-how-is-the-power-supply-industry-enabling-data-centers-to-run-these-advanced-ai-processors/">https://navitassemi.com/nvidias-grace-hopper-runs-at-700-w-blackwell-will-be-1-kw-how-is-the-power-supply-industry-enabling-data-centers-to-run-these-advanced-ai-processors/</a>
- 2. <a href="https://www.eet-china.com/mp/a357098.html">https://www.eet-china.com/mp/a357098.html</a>
- 3. <a href="https://info.support.huawei.com/computing/server3D/index\_zh.html">https://info.support.huawei.com/computing/server3D/index\_zh.html</a>
- 4. <a href="https://servers.asus.com/products/servers/">https://servers.asus.com/products/servers/</a>
- 5. <a href="https://www.gigabyte.com/Enterprise/GPU-Server">https://www.gigabyte.com/Enterprise/GPU-Server</a>
- 6. <a href="https://support.huawei.com/enterprise/zh/doc/EDOC1100313980/cb51627f">https://support.huawei.com/enterprise/zh/doc/EDOC1100313980/cb51627f</a>
- 7. <a href="https://support.huawei.com/enterprise/zh/category/ascend-computing-pid-1557196528909?submodel=doc">https://support.huawei.com/enterprise/zh/category/ascend-computing-pid-1557196528909?submodel=doc</a>

PPT 开源在: <a href="https://github.com/Infrasys-AI/AlInfra">https://github.com/Infrasys-AI/AlInfra</a>