

毎年1月にラスベガスで開催される世界最大級のデジタル技術見本市

「CES 2025」取材レポート

AIを裏側で支える冷却技術

松本 信幸

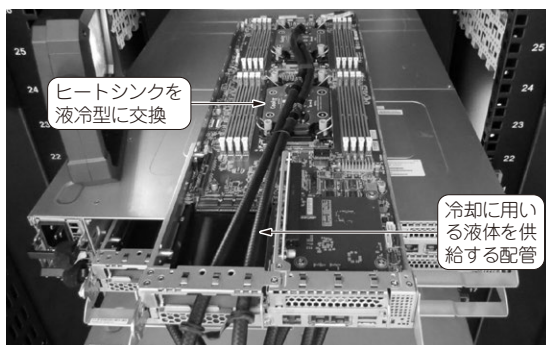


写真1 GPUサーバの冷却方法①…コールド・プレート冷却
従来までのヒートシンクを液冷型に交換するもの。ピンポイントで強力に冷却できる

2025年1月5日～10日に、米国ラスベガスでハイテク機器の展示会CESが開催されました。

展示の内容はAIが主で、AI活用によって実現されるであろう未来の生活が多く描かれていました。しかし、現時点で提供されるAIサービスは力任せに実現している状況であり、膨大な電力を必要としています。膨大な電力が消費されると、膨大な熱が発生するので、これをどう処理するかが問題となっています。今回取材したCESでは、AIサービスを下支える裏方技術が多く見受けられました(写真1、写真2)。

● AIの学習には強力なプロセッサが必要

一般的なAIサービスは、データ・センタなどに保管されている膨大な情報や、特定作業を何度も繰り返して得られるデータなどをもとに、推論を行うことで提供されています。身近なAIサービスとしては、生成AIによる画像や映像、または文章の作成、コール・センタ業務の代替、工場などにおける検査の自動化などがあります。

こうしたAIサービスは、膨大な情報から基本パターンの生成を行う「学習」と、学習で得られたパターン・ファイルを用いて実作業を行う「推論」で構成されるのが一般的です。

学習には強力なプロセッサが必要で、主に大規模なデータ・センタに用意されています。その反面、推論は小型なプロセッサでも対応可能でラズベリー・パイなどの機器でも運用可能です。



写真2 GPUサーバの冷却方法②…液浸ラック
デモ用で水槽のような形状をしている

● 学習を担当するデータ・センタは「熱」に悩まされる

データ・センタでは、AIサービスの提供に伴い、ICT (Information and Communication Technology) 機器の電力消費が急激に増大しています。ウェブ・ページの提供といった従来までのサービスでは、ICT機器を搭載するサーバ・ラックにおける電力消費は3kW程度でした。

しかし最近のAIサービス、特に生成AIで使用されるICT機器を搭載するサーバ・ラックは、60kWくらいの電力が消費されるようになっていて、この先120kWになるとも言われています。AIサービスでは、従来のサーバ・ラック40本分相当の電力を1本で消費することになります。これは40倍の熱を出すということでもあります。

かつてのサーバ・ラックは、サーバ・ルームと呼ばれる部屋をエアコンで冷やす空冷で運用されていました。しかし、40倍の熱量だと空冷では冷やしきれません。このため、AIサービスを提供するICT機器を搭載するサーバ・ラックでは、液体を用いる冷却方法が考えられています。