

# 「GPUさえ用意すれば」という誤解 AIインフラ構築は「超カオス状態」 打開するカギは「三種の神器」

企業競争力に直結する AI 活用では、機密性やコスト効率から独自の AI インフラ構築が向く。しかし、ハードウェアやネットワーク、運用で安易にインフラを選ぶとコスト増や性能不足、セキュリティリスクといった「落とし穴」にはまる。

総務省が発表した「令和7年版 情報通信白書」によると、大企業の約56%が生成AIの活用方針を策定している。生成AIサービスの利用率は2023年度調査と比較して約3倍に伸びたほどだ。

AIの利用が本格化するにつれて、企業がAIに求める要素は大きく変化した。当初は議事録の作成など特定のタスクの効率化が中心だったが、今では研究開発や設計支援、コンタクトセンター業務、IT運用といった、企業の競争力強化に直結するビジネスプロセスへの組み込みが求められている。

この変化に伴い、機密データの安全な利用や既存システムとの連携、長期的なコスト効率といった観点から、自社で独自にAIインフラを構築する需要も高まりつつある。しかし、自社でAIインフラを構築したり運用したりするのは容易ではない。自社のAIインフラの効用を最大化しつつ、落とし穴を回避するにはどうすればよいのか、有識者に聞いた。

自社の事業を拡大、成長させるためのAIインフラの構築においては、パブリッククラウドサービスの活用やデータセンターの

活用、大容量・低遅延ネットワークの活用などさまざまなデザインが考えられる。

機密データを学習させて独自アルゴリズムでシミュレーションを実行する場合や、工場のロボットを遠隔制御してリアルタイムで映像解析するといった用途ではパブリッククラウドが適さない場合がある。そのような場合には自前のAIインフラ構築に向けてハードウェアやネットワーク、ソフトウェア、セキュリティの検討が欠かせない。

ハードウェア面ではGPUサーバやストレージの設計、高性能な冷却システムの検討が不可欠だ。ソフトウェア面では最適なパフォーマンスを引き出す緻密なチューニングが欠かせない。リアルタイム性を強く求めるのであれば、大容量・低遅延のネットワークが前提になる。AIエージェントの自律的な動作によってトラフィックの変化は予測が難しくなるため、ネットワーク構成は柔軟に調整できなければならない。AIをビジネスプロセスに組み込むためには、パブリッククラウドやオンプレミスで稼働する既存システムとの連携も必要だ。

AI学習に必要なデータを供給するデータマネジメント基盤や、AIエージェントの稼働を考慮した新しいサイバーセキュリティ対策も重要になる。これらの技術的な要件だけでなく、都市部での電力枯渇や建設コストの増大、災害対策の観点などからAIインフラの分散化という新たなニーズも現れた。



NTTドコモビジネス 安藤博樹氏（クラウド&ネットワークサービス部 担当部長）

これらの検討・構築には多大な時間と手間がかかる上に広範な分野に精通した専門家も不足しており、多くの企業にとって深刻な課題だ。NTTドコモビジネスの安藤博樹氏(プラットフォームサービス本部 クラウド&ネットワークサービス部 担当部長)が「超カオス状態だ」と表現するこの状況を打開するにはどうすればよいのか。

## AI時代の羅針盤 「AIインフラソリューション」とは

NTTドコモビジネスの「AIインフラソリューション」は、このようなAIインフラの検討・構築時に生じる「超カオス状態」に一石を投じるものだ。同社は長年にわたりクラウドやネットワーク、データマネジメント、データセンターといったサービスを提供しており、各分野で豊富な実績とノウハウを持っている。

そのノウハウをベースに大規模言語モデルやサーバ、ストレージ、ネットワーク、データセンター、運用保守や監視など、AI活用のためのインフラに必要な全ての要素をワンストップで提供できるのがAIインフラソリューションの大きな強みだ。

AIインフラソリューションは複数のサービス群から成る(図1)。顧客専用型IaaSでGPU基盤を提供する「Smart Data Platform(SDPF) GPUプラットフォーム」(以下、SDPF GPUプラットフォーム)や超省エネ型データセンターサービスの「Green Nexcenter」、高速・低遅延通信を提供する「docomo business APN Plus powered by IOWN」(以下、docomo business APN Plus)、セキュリティ統合型NaaS「docomo business RINK」、軽量でありながら高い日本語処理性能を持つNTT独自開発の大規模言語モデル「tsuzumi」、運用保守・監視サービスの「X Managed」や「WideAngle」などだ。

安藤氏はNTTドコモビジネスの最大の強みはこれらの幅広いサービス群をワンストップでサポートできることだと語る。

「AI活用のユースケースごとに最適なインフラ構成を、運用保

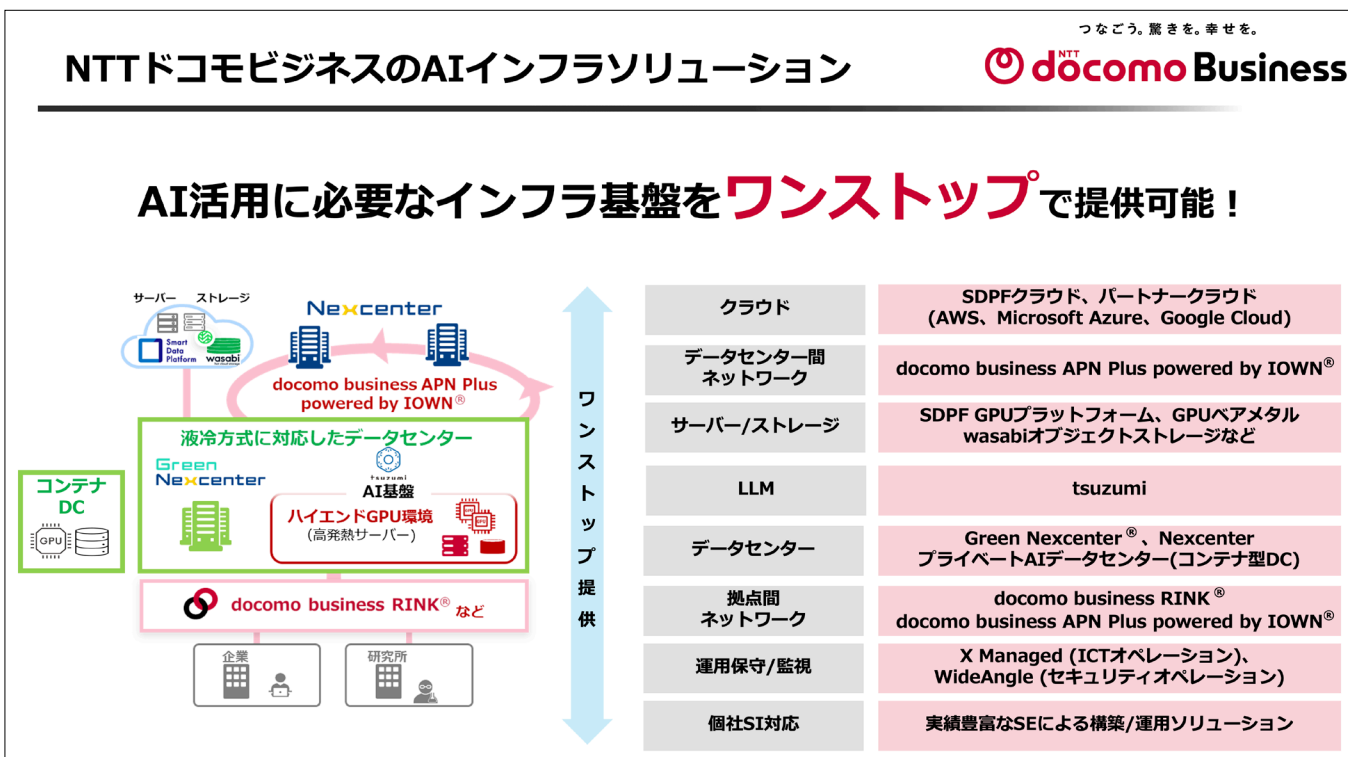


図1 AIインフラソリューションはさまざまなサービスを内包する(提供: NTTドコモビジネス)

守も合わせて柔軟に提供できます。お客さまは自社のビジネスのニーズに応じて必要なサービスを組み合わせ利用可能です。それぞれのサービスを個別に調達して構築する場合よりも低コストで迅速にインフラを構築できます」

## NTTドコモビジネスの「AIインフラソリューション」を支える「三種の神器」

AIインフラを構成する主なサービスはGPU基盤とデータセンター、ネットワークだ。AIワークロードを支えるGPUの計算資源を効率良く稼働させるには、熱の放出量が大きいGPUサーバを安定的に収容できる環境と、それらを結ぶ高速・大容量の

通信基盤が欠かせない。

NTTドコモビジネスのSDPF GPUプラットフォームはGPUサーバや高性能ネットワーク、ストレージなど、GPU基盤の構築に必要な全ての環境を提供可能だ。特に製薬会社や半導体メーカーなど、機密性の高い自社データを活用して独自のAI環境を構築することで競争優位性を確立したい企業にとっては、このような専用型サービスが不可欠だろう。

Green Nexcenterは液冷方式のサーバ機器に対応した冷却環境を提供するデータセンターサービスだ。サーバの冷却に必要な電力を大幅に削減可能であり、ラック単位での柔軟な導入にも対応する。

ネットワーク接続を支えるのは NTT が推進する IOWN 構想（最先端の光関連技術および情報処理技術を活用したネットワーク・情報処理基盤の構想）に基づく APN（オールフォトニクス・ネットワーク）技術を取り入れた「docomo business APN Plus powered by IOWN」（以下、docomo business APN Plus）だ。AI 活用の拡大に伴う利用形態の変化や多様性に応じ、大容量、低遅延かつ揺らぎのない通信品質の特長に加え、柔軟かつ迅速に帯域や経路変更ができるネットワークを実現できる。

これらのサービスを組み合わせれば、AI 活用に最適なインフラ構成を実現できると安藤氏は説明する。

「今後の AI 学習では、GPU サーバの高性能化に伴ってこれまで以上の電力消費と発熱に対応する必要があるため、Green Nexcenter の液冷環境が不可欠です。学習用のデータ転送では大容量通信が求められますが、推論においてはレスポンス性能が重要になります。これらのネットワーク要件については、docomo business APN Plus の大容量かつ低遅延な特性が効果を発揮します」

昨今の AI 用の GPU 基盤は、電力や冷却設備、床荷重といった物理的な制約からデータセンターの分散配置を余儀なくされるケースも増えた。BCP やデータ保管場所の要件に対応するために、意図的に分散するニーズもある。これらのニーズに対応す

るために、NTT ドコモビジネスは「GPU over APN」というコンセプトの下、docomo business APN Plus で接続した GPU 基盤間での分散 AI 学習の実証実験を試み、成功した（図 2）。これによって物理的な制約や災害リスクに対応しつつ、AI インフラの柔軟かつシームレスな拡張の実現を目指す。

## AI活用をさらに加速する 「AI-Centric ICTプラットフォーム構想」

NTT ドコモビジネスは 2025 年 6 月、AI 時代に最適な次世代 ICT プラットフォームを目指す「AI-Centric ICT プラットフォーム構想」を始動させた。この構想は、企業が AI を活用して生産性を抜本的に改善したり競争力強化やビジネスモデル変革を進めたりすることを目標とした際、インフラに期待される「分散」「柔軟」「安全」「リーズナブル」を実現する ICT 環境の提供を目指すものだ。

安藤氏はネットワーク機能をサービスとして提供する「NaaS」（Network as a Service）を中心に置いた点がこの構想の特徴だと説明する。

「企業の AI 活用が本格化し、コンピューティング環境が分散する中で、ネットワークの柔軟性はますます重要になります。ビジネスの展開や AI 技術の進化に応じて頻りにネットワーク構成を変えなければならない事態が想定できます。AI エージェントなどが自律的に動くようになれば、トラフィック自体が爆増して予測も困難になります。Software-Defined による柔軟な帯域変更やセキュリティ機能を持つ NaaS は、そのようなニーズに対応するために不可欠な存在です」

AI-Centric ICT プラットフォーム構想はこの NaaS を中心として、Green Nexcenter や GPU のクラウドサービス、他社のクラウドサービス、コンテナ型データセンターなどを接続して、AI 活用に最適化したインフラを実現する。これらを統合管理してコストの最適化を図るために、AIOps / FinOps を取り入れた AI

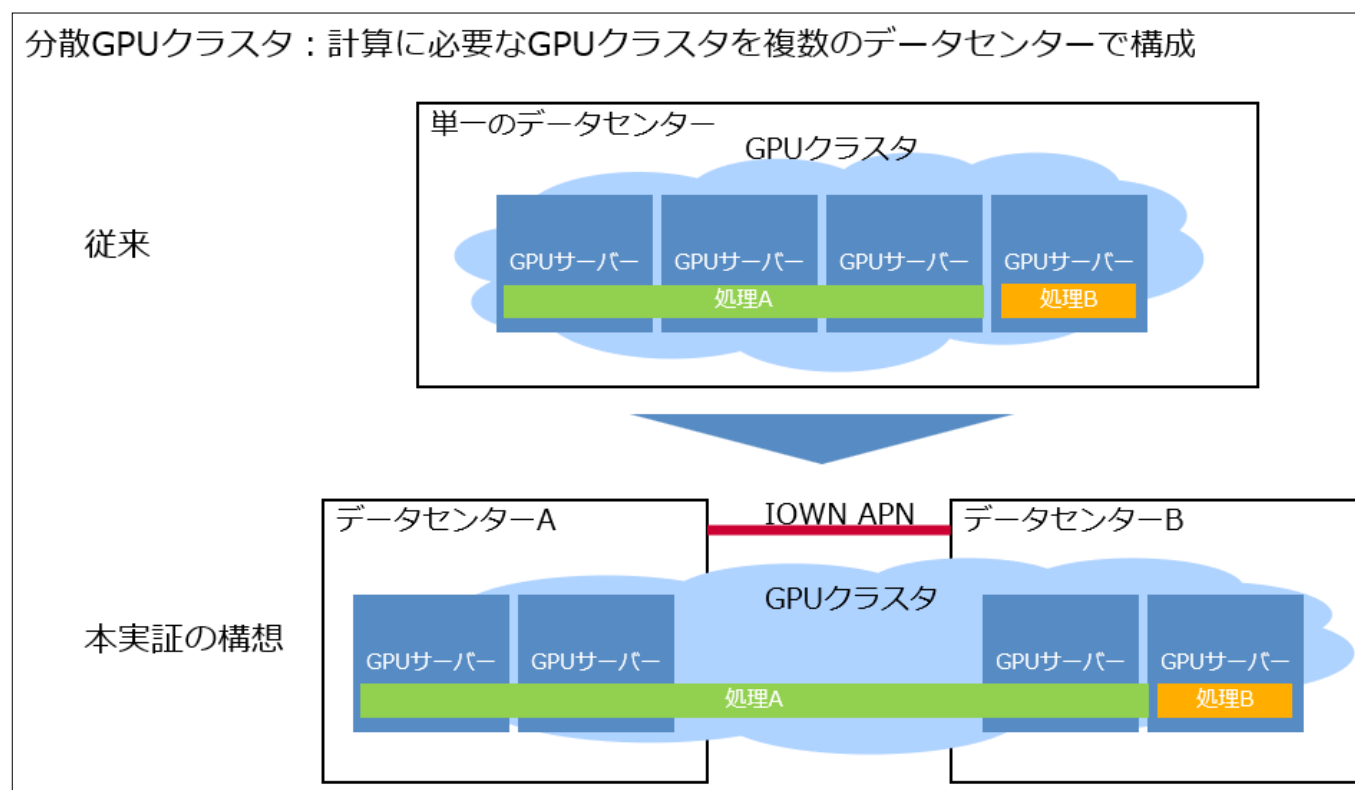


図 2 IOWN APN を用いた GPU 分散クラスタ（提供：NTT ドコモビジネス）

マネージドサービスも提供する。

NTTドコモビジネスは今後さまざまなパートナーと連携・提携してAI-Centric ICTプラットフォーム構想を拡張する。企業のAI活用を強力に後押しすると同時に、地方創生といった社会

課題の解決にも貢献するという。

「AIを活用して自社の競争力を高めるには、インフラ領域でも単にGPUサーバを用意するだけでなく、さまざまな要素を考慮する必要があります。企業が自社単独でそれを実現するのは容

易ではありません。私たちは、お客さまのビジネスに合わせてAIインフラの構築に必要な要素を最適な構成で提供できます。幅広い領域で実績を持つNTTドコモビジネスをビジネスパートナーとしてぜひ活用してください」

## AI-Centric ICTプラットフォーム

- 企業がAIインフラに期待する「分散」「柔軟」「安全」「リーズナブル」という4つのニーズに対応
- 日本企業のAIによる競争力強化や、地方企業の先進AIを活用したビジネス拡大(地方創生)に貢献

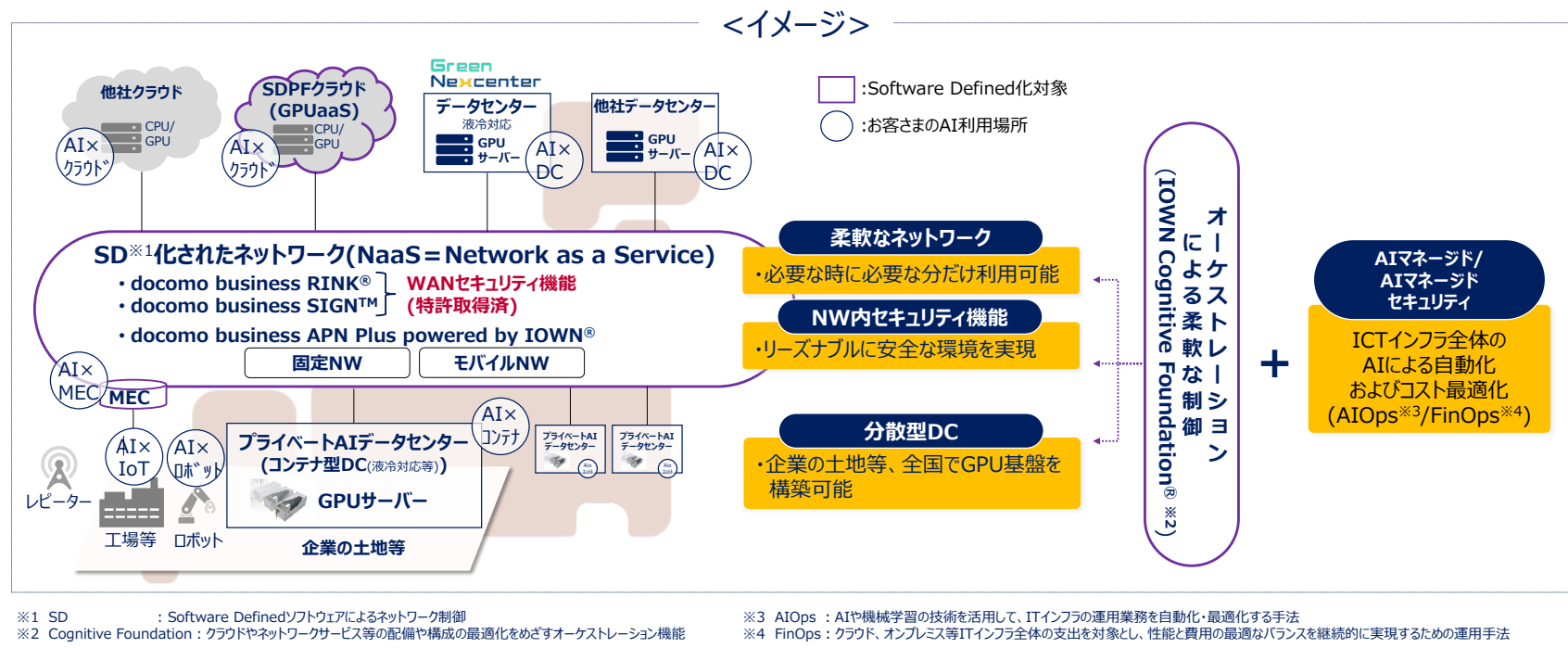


図3 AI-Centric ICTプラットフォーム構想のイメージ (提供: NTTドコモビジネス)

NTTドコモビジネス株式会社 プラットフォームサービス本部 クラウド & ネットワークサービス部

AIインフラソリューションに関する詳細はこちら <https://www.ntt.com/business/lp/aiinfra.html>

お問い合わせ [https://www.mkt.ntt.com/jp\\_PS\\_inq\\_AI-Infra\\_25\\_223\\_wtc\\_reg.html](https://www.mkt.ntt.com/jp_PS_inq_AI-Infra_25_223_wtc_reg.html)