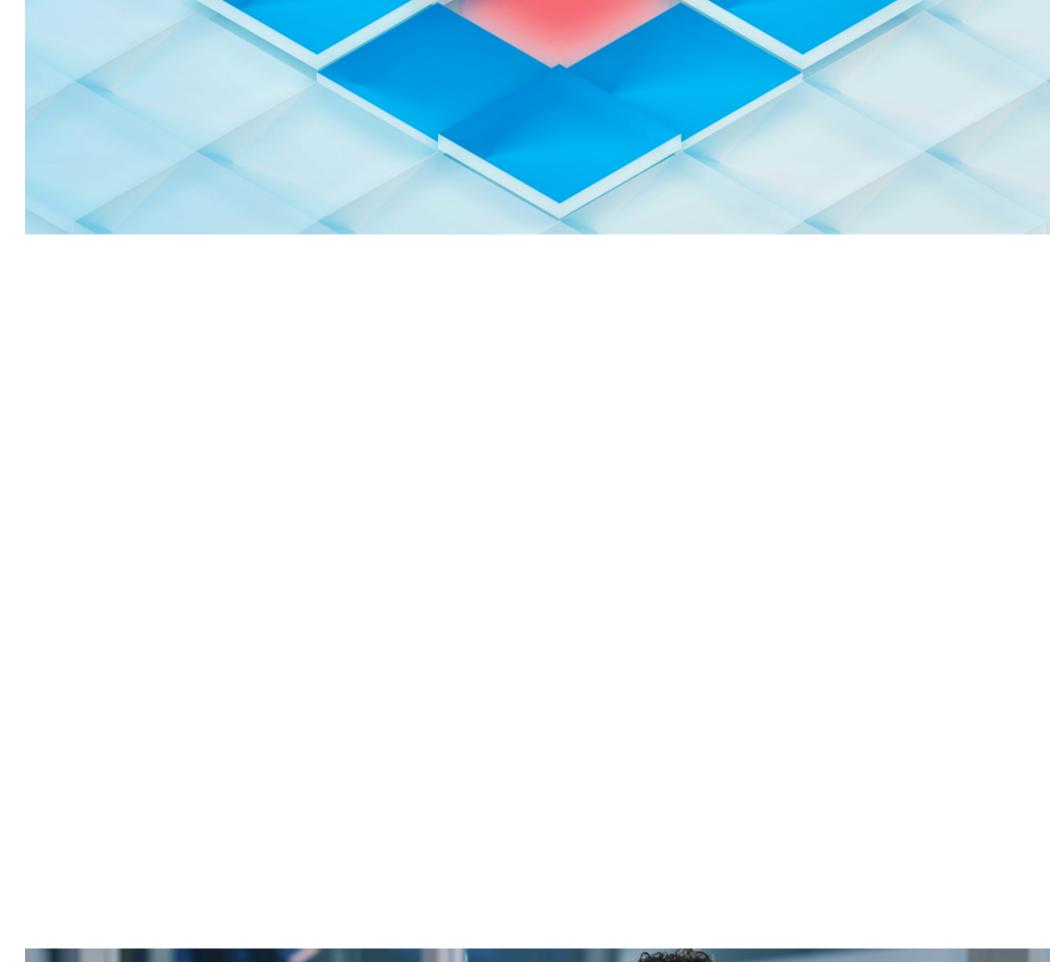


AIワークロードをIBM Power上で実行すべき5つの理由

AI戦略を強化するための
信頼できる基盤



1 Powerはデータが存在する場所にAIを導入

オンプレミス、クラウド、エッジと、データがあらゆる場所に存在する世界にあっては、スマートなだけでなく戦略的なAIソリューションが必要です。PowerはAI推論を、データが生成され、存在する場所にもたらします。ネットワーク経由で機密データを移動させたり、効果で複雑なGPUを待つたりする必要はもうありません。Powerは、ミッションクリティカルなアプリケーションがまさに存在する場所にAI推論を組み込むことで、レイテンシー、セキュリティ、リスク、追加コストを削減できます。PowerのIBM Spyre AcceleratorとIBM Watsonx.dataにより、最新のデータレイクハウスを高度なオフチップ・アクセラレーションと統合することで高速な洞察が獲得でき、信頼できるデータと強力な推論機能の架け橋が構築されます。

2 Powerの24時間365日セキュリティ

AIワークロードは価値のあるものですから、脅威に対して脆弱であってはなりません。Powerは、透過的なメモリー暗号化を用いてスタックの各レイヤーにセキュリティを組み込み、パフォーマンスに影響を与えることなくインサイトを保護します。生成AIのような複雑なタスクのために、パフォーマンスに関する不安なしに、自身を持って推論を拡張できます。

Powerサーバーでは、IBM Power Cyber Vaultにより1分未満でのランサムウェア検知¹、最大99.9999%のアップタイム²が保証され、レジリエンツであらゆる用途に備えた常時稼働のAIが実現できます。

1分未満 99.9999%

脅威検知の保証¹

アップタイムの信頼性²

3 Powerはハイブリッドの柔軟性でスムーズ

AIワークロードのデプロイに関しては、ハイブリッドITの柔軟性が非常に重要です。Powerサーバーは、ハイブリッドクラウド・エクスペリエンスを円滑に実現するための完全に最適化されたエンタープライズ・ソフトウェアを提供し、オンプレミス・システムとIBM Power Virtual Serverとの間におけるワークロードの移動を容易にします。クラウドにおけるモデルのトレーニングも、オンプレミスでの推論も、Powerではより多くのことを柔軟に行えます。さらに、130以上のオープンソースAIツールやパッケージをサポートするため、チームは障害にぶつかることなく構築、デプロイ、拡張できます。

4 Powerはサステナブルなパフォーマンスをオンデマンドで提供

AIワークロードではサステナビリティ要件は満たせないなどと言うのは誰でしょうか。Power11は、x86プロセッサーのサーバーと比べてワットあたりの性能が2倍なので、同じワークロードを低消費電力で実行できます。³そしてPower11サーバーの新たなエネルギー効率モードにより、最大パフォーマンス・モードと比較して電力効率を28%改善できます。⁴

↑ 28%

電力効率に優れる

(X86 Serverとの比較)⁴

↑ 42%

処理量が増加 (X86 Serverとの比較)

5 Powerは妥協なき高速化を実現

既存のワークロードを中断せざることなく、大規模のAIをリアルタイムで実行したいとお考えでしょうか。IBM Powerサーバーは、まさにそのため構築されました。高度な並列処理、大容量メモリー、オンチップ・アクセラレーションにより、ワークフローにAIを直接組み込むよう設計されています。プラットフォームを使用するためにデータサイエンティストがリファクタリングを行う必要はなく、AIワークロードをそのまま実行できます。さらに、大規模言語モデルを実行する場合は、同時にユーザー数40人というピーク負荷時でも、Power S1022はx86プロセッサー搭載サーバーと比較して毎秒42%多くバッチ・クエリー処理を実行⁵でき、レイテンシーは1秒未満。⁶つまり、迅速な洞察、スムーズな操作、1秒未満の推論が実現できるということです。

↑ 42%

Power S1022で毎秒のバッチ・クエリー

処理量が増加 (X86 Serverとの比較)

AIとIBM Powerのテクノロジーをより深く掘り下げる →