

サイバーエージェント

AI/ML 向け GPU 基盤の大幅強化を 800GbE ネットワークで支える

導入企業

株式会社サイバーエージェント

業種

インターネット広告、メディア&IP、ゲーム

所在地

〒150-0042
東京都渋谷区宇田川町 40 番 1 号
Abera Towers

URL

<https://www.cyberagent.co.jp/>

使用製品

QFX5240 Switch
400G & 800G Solution



「QFX5240 によって、AI 環境の基礎となるロスレス・高性能・低遅延な 800GbE ネットワークを構築できました。GPU クラスタの大規模化を支えながらも、消費電力などコストの大幅な削減を実現できたのは、非常に大きな成果です。事前の技術検証や実際の構築まで、ジュニパーの親身なサポートでスピーディにプロジェクトを推進できました」

株式会社サイバーエージェント
CIU Platform Div NW チーム
ネットワークエンジニア

小障子 尚太郎氏

サイバーエージェントは、社内 AI 開発の急速な拡大に対応するため、GPU 基盤を大幅に強化しました。この基盤を支えるネットワークは、従来の 400GbE から 800GbE への強化を検討。ジュニパーネットワークスの 800GbE スイッチ「QFX5240」を選定し、事前の共同研究・検証もあってスピーディな導入に成功しました。GPU 環境の大規模化を果たしながら、Pod あたりのポート収容数を 4 倍に向上し、ネットワークの電力効率も 50%改善。将来的な AI ビジネスの発展にも寄与する強固なネットワーク基盤を構築しました。

Challenge

GPU クラスタの大幅増強が求めるネットワークの進化

サイバーエージェントは、メディア & IP 事業、インターネット広告事業、ゲーム事業を中心に事業を展開している企業です。同社は近年、AI 技術の開発に注力しており、サービス強化や業務効率化など積極的な活用を進めています。

サイバーエージェント グループ IT 推進本部 CIU Platform Div NW チームの小障子尚太郎氏は、AI の活用例について「たとえばインターネット広告であれば、クリエイティブから効果の予測、評価まで活用しています。また一般業務でも、スケジューラーに AI を組み込んで自動的に予定を最適化するなどの活用方法を見いだしています。全社的に AI 活用に取り組んでいるのです」と説明します。

同社は生成 AI サービスが社会的に注目される以前から AI 技術への投資を進めており、プライベートクラウド「Cycloud」をベースに独自の AI 基盤を構築し、社内の AI プロジェクトへサービスとして提供してきました。「パブリッククラウドよりも長期的なコストメリットがあり、いつでも使いたいときに GPU を使える強力な AI 基盤を目指して構築しました」と小障子氏は述べます。しかし、この目標が大きな課題になっていたと続けます。というのも、社内の AI への注目度が急速に上昇し、大規模な環境にもかかわらず、常に GPU は予約でいっぱいという状況でした。

サイバーエージェントでは、この需要拡大を受け、2025 年に GPU 基盤を数倍に増強する計画を立てました。そこで新たな課題として浮上したのがネットワーク基盤です。AI 基盤では多数の GPU で並列分散処理するため、GPU クラスタを安定的に接続し、ロスレス・広帯域・低遅延を維持できるネットワークが欠かせません。既存のネットワーク基盤は、400G Ethernet スイッチをベースとしたスパイン・リーフ構成を採っていましたが、大幅な拡張には課題がありました。

「400G リーフスイッチの台数を増やしてスケールアウトする方式では、多くの GPU 間通信がスパインスイッチを介するようになり、帯域を圧迫してパケットロスを誘発し、パフォーマンスが低下する恐れがありました。1 台あたりのポート密度を高められるスケールアップ方式のほうが望ましいと考えました」(小障子氏)

Transformation

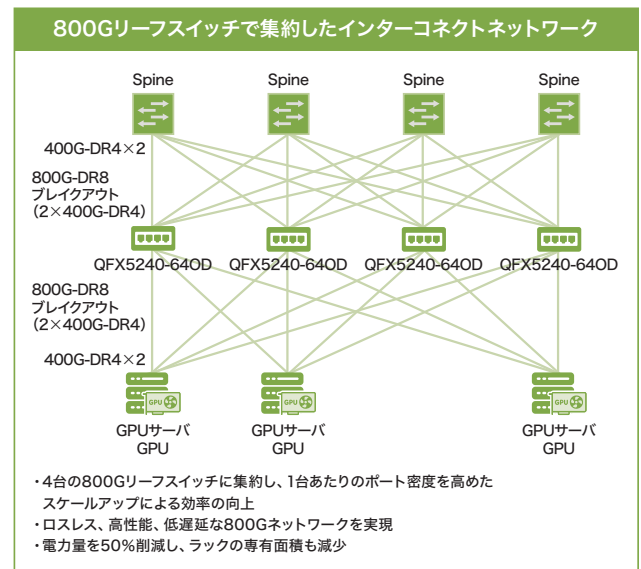
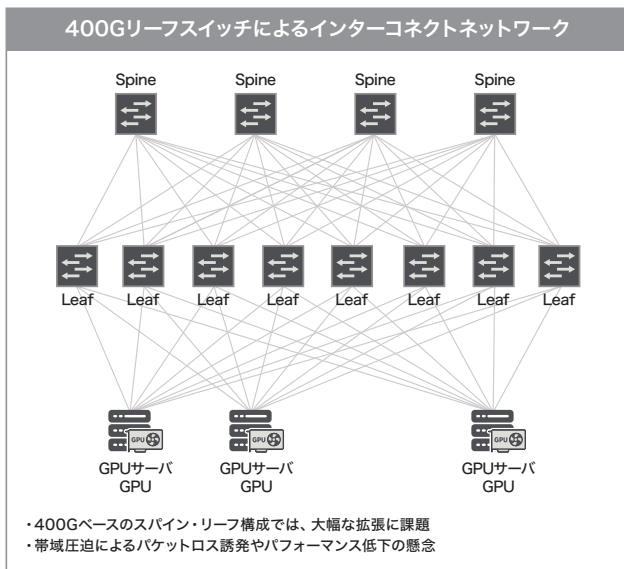
事前研究が先進的なネットワークの迅速な実現に寄与

サイバーエージェントは、新しい AI 基盤のネットワークにジュニパーネットワークスの 800GbE スイッチ「QFX5240-64OD」を採用しました。その理由として、同社は 4 つのポイントを挙げています。

まず、ジュニパーネットワークスが早期に 800GbE スイッチをリリースしていたこと、技術的先進性が評価されました。次に、Junos®の運用の高さについて、サイバーエージェント グループ IT 推進本部 CIU Platform Div NW チームの正田紅樹氏は「コンフィグがコミット方式になっていることが最大のポイントです。過去のコンフィグとの差分を抽出できる点も気に入っています。運用していて助かるのは、ドキュメントが充実しているところ。ジュニパーのソリューションを選ぶ大きな理由になりますね」と説明します。

選定・導入を決定づけたのは、将来的な AI 環境のインフラ強化を見据えて、以前からサイバーエージェントとジュニパーが共同で 800GbE ネットワークの検証を進めていたことでした。サイバーエージェントは迅速なビジネス展開・意志決定が特長で、短期間での AI 基盤強化を求められたのです。QFX5240 であれば問題なく稼働することが検証済みだったので、選定もスムーズに決まりました。

さらに、このスピード導入を支えたのはジュニパーの技術支援です。ジュニパーとの共同検証の結果をベースにコンフィグを作成したことが、効率的な導入作業に寄与しました。正田氏は、「ジュニパーとは密に連携できる関係性があり、親身なサポートに助けられています。ジュニパーのさまざまなノウハウが、導入・運用に役立っています」と高く評価しています。



また、サイバーエージェントがネットワーク技術としてEthernetを積極的に選択している点について、「サイバーエージェントはデータセンター運用を通じてEthernetの知見がありますし、技術情報が豊富という点も重要です。ベンダーロックインはありません。将来的にマルチテナント環境を視野に入れば、Ethernetは理想的と言えるでしょう」と、足田氏は説明します。

Outcomes

GPU 基盤の大幅拡張

3倍

AI 開発を加速する強力な
ネットワーク基盤

最先端のネットワーク技術

800GbE

AI の進化に必要な性能と機能、
将来性を確保

消費電力の改善

50%

密度の向上でトータルコストの
大幅削減

Solution and implementation

QFX5240 Switch

400G & 800G Solution

先端技術で高性能・低遅延・低コストな新ネットワーク基盤

大規模なGPUクラスター強化を支えつつ、消費電力はむしろ低減

性能の大幅強化に反してコストは大幅削減

従来は8台の400GbEリーフスイッチで構成していた環境から、GPUを3倍に拡張しながらも4台の800GbEリーフスイッチへと集約できました。この密度の向上によって、電力量は50%ほど削減でき、ラックの専有面積も減少。大幅な効率改善を達成できました。

パフォーマンス向上とシステム安定性

新しいAI基盤は安定的に稼働しており、パフォーマンスやリソースの問題も報告はありません。エンドユーザーがGPUを効率的に利用できる環境が整備され、AI開発・活用を推進できています。

将来のニーズに備えた先端ネットワーク基盤

800GbEという先端技術の採用により、次世代GPU環境への対応力も確保できました。「今後もGPUクラスターの増強や新しいGPUサーバへの対応を求められるでしょう。ネットワークもしっかり追随できるよう強化していく計画です。ジュニパーにも、新しい技術をいち早く実用化してくれるように期待しています」(小障子氏)



東京本社

〒163-1445 東京都新宿区西新宿3-20-2

東京オペラシティタワー 45階

電話：03-5333-7400 FAX：03-5333-7401

西日本事務所

〒530-0001 大阪府大阪市北区梅田2-2-2

ヒルトンプラザウエストオフィスタワー18階

電話：03-5333-7400 FAX：03-5333-7401