

オープンソースの機械学習 PyTorch の進化

PyTorch は、CircleCI を活用して、900 人以上の
オープンソースコントリビュータを管理しています

Facebook 社のオープンソース開発者である Eric Nakagawa 氏は、次のように述べています。「CircleCI により、より柔軟な方法で作業を進めることが可能になりました。私たちのチームが直接取り組んできた最大の課題の 1 つは、Web サイトでチュートリアルを作成することでした。このプロセスにこれまで 4~6 時間かかっており、少しでも調整しようとする、あまりにも時間を取られることに、フラストレーションがたまることがありました」。CircleCI により、Nakagawa 氏のチームは複数のインスタンスタイプで並行してビルドを実行でき、ビルド時間を 75% も短縮できました。さらに重要なことは、マシンを調整するという毎日の悩みから解放されたことです。これにより、プロジェクトの中核、つまりコミュニティに注力することが可能になりました。

「インフラストラクチャの管理に労力を費やすことができなくなり、より高い価値を提供できると我々が考える分野にリソースを集中できるようになりました。これは、CircleCI と連携することで得られた大きな成果の 1 つです。」

課題

PyTorch チームには、900 を超えるオープンソースコントリビュータと 6 人のメンテナンス担当者に加えて、約 100 人の主要メンバー (Facebook 社内外のメンバー) がいます。

PyTorch ライブラリは、数百に及ぶダウンストリームの学術的および商業的なプロジェクトにつながっています。PyTorch プロジェクトを進めている Facebook 社の製品マネージャーである Joe Spisak 氏は、次のように述べています。「我々が携わっている分野は、量子化や低精度演算ニューラルネットワークなど、現在も活発に研究が進められている分野です。今日、翻訳やコンピュータービジョンのような技術については、文字通り何兆もの予測を低精度で実行していますので、これは当然のことかもしれません。しかし、これらの分野では未だに多くの論文が発表されており、画期的な新しい方法が発見されています」。PyTorch に関するチームの業務は、最新の研究・調査を行っているユーザーが必要とするツールの一助となっています。

「インフラストラクチャの管理に労力を費やすことができなくなり、より高い価値を提供できると我々が考える分野にリソースを集中できるようになりました。これは、CircleCI と連携することで得られた大きな成果の 1 つです。」



Eric Nakagawa 氏、
PyTorch のオープンソース部門の主任



CircleCIにより、Nakagawa 氏のチームは複数のインスタンスタイプで並行してビルドを実行でき、ビルド時間を75%も短縮できました。

オープンソースにすることの大切さ

恐らく、営利目的の企業では、オープンソースコミュニティのニーズを管理することは難しいことかもしれません。しかし、PyTorch は、サービスを提供する対象こそが重要でした。Spisak 氏は、次のように述べています。「オープンソースは、我々の DNA に最初から組み込まれていました。Facebook が 2004 年にスタートしたときから、Linux、Apache、MySQL、PHP などのオープンソーススタックで構築されており、オープンソースは常に Facebook のエンジニアリング部門における最重要の文化の一部となっています」。Facebook のオープンソースへの取り組みは同社とともに成長しています。PyTorch チームや AI 部門の他のメンバーは世界各国の研究者に重要なツールを提供しており、膨大な数のプロジェクトネットワークが相互に連携し構築されています。

Spisak 氏はさらに次のように述べています。「我々はほぼすべてをオープンソースにするように取り組んでいます。プロジェクトが安定しており、コミュニティや研究に役立つと思った場合には、オープンソース化しています。我々は、これらのメンテナンスを行い、PyTorch のような多くのプロジェクトのコミュニティを作り出しています。」

大規模な OSS のための CI/CD の実行

900 人を超える PyTorch コントリビュータが、CircleCI を使用してソフトウェアデリバリワークフローを管理しています。このチームのスタッフには、CircleCI、Netlify などの CDN サービス、GitHub が含まれます。これにより、プロジェクトのコントリビュータの数に関わらず、コミットまたはプルリクエストに対する一貫したプロセスが実現されます。Facebook の OSS プログラムのデベロッパーアドボケートである Joel Marcey 氏は、次のように述べています。「このようなシステムでは、小規模なプロジェクトが大規模化していきます。プロジェクトに参加するすべてのユーザーは、一貫したエクスペリエンスの元で作業でき、自分の PR (プルリク) がパスしたかどうかを確認できます。

Facebook 社のオープンソース部門の主任を務める Eric Nakagawa 氏は、次のように述べています。「CircleCI により、より柔軟な方法で作業を進めることが可能になりました。私たちのチームが直接取り組んできた最大の課題の 1 つは、Web サイトでチュートリアルを作成することでした。このプロセスにこれまで 4~6 時間かかっており、少しでも調整しようとする、あまりも時間を取られることに、フラストレーションがたまることがありました」。CircleCI により、Nakagawa 氏のチームは複数のインスタンスタイプで並行してビルドを実行でき、ビルド時間を 75% も短縮できました。さらに重要なことは、マシンを調整するという毎日の悩みから解放されたことです。これにより、プロジェクトの中核、つまりコミュニティに注力することが可能になりました。

PyTorchについて

PyTorch は、Facebook の AI 研究グループが作成したオープンソースのディープラーニングプラットフォームです。NumPy と同様に、PyTorch はテンソル演算が可能なライブラリですが、GPU やその他のハードウェアアクセラレーションをサポートし、AI 研究者がさまざまな分野を探究できるようにする効率的なツールが追加されています。PyTorch は、Python ベースの Lua Torch フレームワークの後継として開発が始まりましたが、研究用のプラットフォームだけでなくデプロイメントプラットフォームへと、その開発範囲が拡大されています。スタンフォード大学や Salesforce などの組織は、PyTorch を使用して、機械学習の研究を促進し、機械学習を活用した製品・サービスの開発を進めています。また、PyTorch は、CircleCI を使用し、多くのオープンソースのコントリビュータ向けにスムーズで信頼性の高い PR プロセスを作成し、これまで使用されていたソリューションと比較して、より迅速で少ないオーバーヘッドでコードをリリースしています。

CircleCIについて

CircleCI はソフトウェア開発においてソースコードのテストやビルドを自動化する CI (継続的インテグレーション)、リリースを自動化できる CD (継続的デプロイメント) のプラットフォームを提供しています。

