



# The New York Times におけるデータ アーキテクチャの進化

The New York Timesでは、新聞紙面、アプリ、サイトで何億人もの読者に24時間365日ニュースを届けています。同社のニュースルームは、報道内容を支える巨大なCOVID-19(新型コロナウイルス感染症)データベースの構築が2020年に必要になるとは想定していませんでした。そして、大統領選挙では過去最高のトラフィック量を記録しました。

## 課題:

The New York Timesは、基本的にはデータ企業ではありませんが、ビジネスの新たな需要を満たすために、データ企業のような考え方や運用方法を取り入れる必要がありました。たとえば、読者の好みに応じてコンテンツを作成して整理しながらフィードをパーソナライズする、サイトやアプリの利用状況に関する行動分析を学ぶことでユーザーの登録と購読を促す要因や機能を検討する、アナリストがキャッシュメカニズムの構築に注力する必要がないように指標を追跡するためのデータレポートツールを構築する

といったさまざまな取り組みを通じて、ニューノーマル以降の課題に対処してきました。



2020年第4四半期には全世界で2億7,300万人の読者を獲得し、他の選挙週間と比較しても、4日間連続で過去最高のトラフィックを記録しました。それにもかかわらず、データ損失が出なかったことを誇らしく思いました。”

**Ed Podojil 氏**

The New York Times データプロダクト  
担当エグゼクティブディレクター

## ソリューション:

Google Cloud のデータベースと分析プロダクト ([BigQuery](#)、[Dataflow](#)、[Pub/Sub](#))、そして [BigQuery BI Engine](#) により、The New York Times は無制限のスケーリングとパフォーマンスを実現し、次のことが可能になりました。

- 迅速に対応する。アナリストはビジネスパートナーに高品質の成果物をすばやく提供できます。

## メリット:

Google Cloud のオープンで柔軟なアプローチにより、The New York Times は読者が増加しても高品質なニュースの配信を続けてきました。具体的には、次のようなメリットがあります。

- アナリストの日常的なニーズを満たさなかった古いデータ アーキテクチャやツールを BigQuery に置き換えることで、数時間分あるいは数日分ものエンジニアリング時間を節約しながら、パフォーマンスとコストを改善しています。
- 多彩なサービスから選択し、さまざまな関係者のニーズに合わせてプロジェクト要件を自由に調整できます。
- BigQuery とデータ分析のロードマップを把握することで、自社開発の重点領域を特定しやすくなりました。

- リアルタイムの選挙データをストリーミングして分析する。
- データレポート API を作成し、インメモリ分析サービスの BigQuery BI Engine を使用して BigQuery からデータを直接読み取る。

- The New York Times のエンジニアは、Google Cloud を利用して、米国有数の [包括的な COVID-19 データセット](#) を構築しました。



上司が初めて BigQuery でクエリを実行してコーヒーを飲みに行こうとしたら、椅子から立ち上がる前にクエリが完了したことをいまだに覚えています。当社のアナリストは今でもこの話をします。”

**Ed Podojil 氏**

The New York Times データプロダクト担当エグゼクティブディレクター

## お客様への3つのヒント



テクノロジーはかつてないほど急速に変化しています。そのため、最新動向の把握が難しい場合もあります。貴社のニーズにすぐ対応し、オープン性への取り組みに信頼が置けるパートナーを選びましょう。そうすれば、必要に応じて柔軟に軌道修正できます。



情報セキュリティに優先順位を付け、堅牢なセキュリティとガバナンス体制を備えたパートナーを選びましょう。Google Cloud の全データは、転送中も保存中も、そして Confidential VMs のメモリ内でもデフォルトで暗号化されています。



AI を初めて導入する場合は、要件を明確にしたうえで、数十年にわたる AI の研究とイノベーションに裏打ちされた Google Cloud を活用して最大の問題を解決しましょう。

## Google Cloud のオープン性に対する取り組み:

- ハイブリッド環境とマルチクラウド環境 (BigQuery Omni、Dataproc on Kubernetes) でデータを分析できます。
- さまざまな処理、分析エンジン、オープンソース ツール、ML サービスから選択できます。
- 使い慣れた MySQL や Postgres と標準の接続ドライバをサポートしているため、Google が推奨するツールを利用し、コミュニティで提供される最新の拡張機能を常に活用できます。
- オープン API によってロックインを排除し、移行、ポータビリティ、データアクセスを容易にします。
- オープン スタンダードとオープン API により、取り込み、保管、処理、分析などのさまざまなサービス (Spark や Presto など) の間ですぐに相互運用できます。
- 統合型 AI プラットフォームにより、堅牢な ML モデルと OSS フレームワークを実現します。
- Kubernetes、Istio、Kubeflow、TensorFlow 向けに Google が提供するオープンソースを活用すれば、インフラストラクチャやモデルの導入と管理が容易になります。
- Apache Beam を使用して一度ビルドすれば、どこでも実行できます。

詳しく知りたい方は、



[Data Cloud Summit](#) にご参加ください。

9月7日(火)~10日(金)