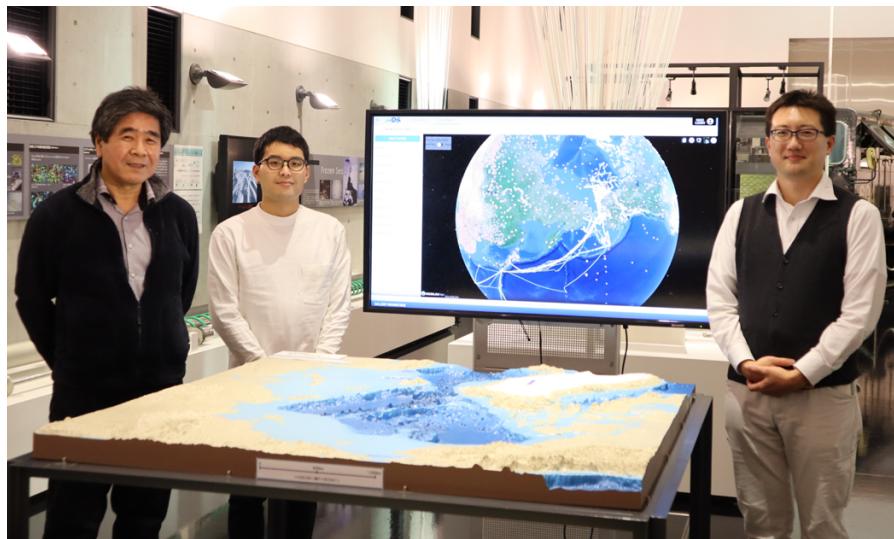


海水分布予報の計算を同一コア数でオンプレの2倍に高速化



大学共同利用機関法人 情報・システム研究機構 国立極地研究所(以下、極地研)では、北極域研究のナショナル フラッグシップ プロジェクトである北極域研究加速プロジェクト(ArCS II:Arctic Challenge for Sustainability II)を推進。このプロジェクトにおける北極海海水分布予測のハイ・パフォーマンス・コンピューティング(HPC)基盤として、Google Cloud が採用されています。極地研の特任研究員、国立大学法人 東京大学(以下、東京大学)大学院新領域創成科学研究科の教授、および特任研究員に話を伺いました。

(利用している Google Cloud サービス)

[Compute Engine](#)、[Filestore](#)、[Cloud Source Repositories](#)、[Google Workspace](#)

ArCS II の海水分布予報のハイ・パフォーマンス・コンピューティング(HPC)基盤に Google Cloud を採用

「1995 年に、北極海航路の国際共同研究プロジェクトでロシアの耐氷貨物船をチャーターし、横浜からノルウェーまで調査航海をしました。当時、海水を避けながら航海するためのデータとして、先行する船からの情報はありました。しかし、目前の海氷を避ける、避けないなどの判断は船長の勘と経験でした。海氷は急激な変化をするため、精度の高い海水分布予報の重要性を痛感しました」東京大学 大学院新領域創成科学研究科 海洋技術環境学専攻 海洋情報基盤学分野 教授の山口さんは、当時をこう振り返ります。

「氷海航行を、昔のような“冒険”には、もうしないという思いから、帰国してすぐにセンシング、モデルと予測、衛星モニタリングなどの技術を組み合わせ

た氷海航行支援システムを設計し、国際シンポジウムで発表しました。25 年かけて、この仕組みを実現することができました」(山口さん)

現在山口さんは、極地研、国立研究開発法人海洋研究開発機構(JAMSTEC)、および国立大学法人北海道大学の 3 機関が中心となり、2020 年 6 月～2025 年 3 月の約 5 年弱で実施される ArCS II に参画しています。

ArCS II では、極地研で天気予報データを取得し、10 日先までの海水分布予報を計算し、衛星を経由して、北極海を航海する JAMSTEC の海洋地球研究船(みらい)に送信する海水分布・海洋予報モデル(IcePOM)を構築。海水分布予報の計算インフラ(HPC)基盤として、Google Cloud が採用されています。

Google Workspace との連携を評価して Google Cloud を推薦

ArCS II は、2015 年度～2019 年度に実施された北極域研究推進プロジェクト(ArCS)の後継プロジェクトです。ArCS においても、海水分布予報を計算していましたが、最大の違いは計算インフラにオンプレミスのサーバーを利用していました。オンプレミスの課題を山口さんは、次のように語ります。

「大学の研究室でサーバーを所有し続けるのは非常に大変です。トラブルが発生すると、研究者が SE の仕事をしなければなりません。性能が不足すると、新しいサーバーを調達するための時間と労力、コストがかかります。スーパーコンピューターもありますが、ジョブ待ちに入るといつ計算が終わるかわかりません」

明日の海水分布予報を、明後日に送っても意味がありません。そのため海水分

布予報は、10 日先までを 6 時 30 分～14 時 45 分の約 8 時間以内で計算するという時間の制約もありました。山口さんは、「ArCS の担当者からオンプレミスからクラウドへ移行することで 2 倍速くなると聞いたので、極地研 国際北極環境研究センター 特任研究員の照井さんに相談しました」と話します。『ちょうど私の管理している北極域データ アーカイブ システム(ADS)のサーバーのクラウド化について、いくつかのクラウドサービスと比較検討していました。その中で Google Cloud の担当者にも色々と相談していたところでした。今回の計算インフラの相談があり、極地研と東京大学の複数の利用者を想定しなければならないことと、両機関で Google Workspace を使っていたことから、認証と認可を考えて Google Cloud を推薦しました』(照井さん)

Google Cloud による検証環境の短期構築経験を生かし、本番環境を3日で構築

海氷分布予報の計算インフラのクラウド化にあたり、検証環境と本番環境を構築しています。検証環境は、2020年7月より検討を開始し、8月11日～24日で構築。本番環境は、9月8日～10日で構築しています。「検証環境の経験とノウハウを生かせたので、本番環境は3日で構築できました」(照井さん)

検証環境は、仮想マシンにCompute Engine、モデルのデータを保存するためにCloud Filestore、OSにLinuxを使って検証環境を構築。極地研と東京大学では、Google Workspaceを利用していたので、新たにID/パスワードを発行する手間なく、既存のアカウントによる認証と認可でGoogle Cloudへアクセスしています。

「マシンタイプは8CPU/32GBメモリを選択し、複数のCPUのアーキテクチャを使って検証しました。その結果、14時間弱で終わったので、CPUを増やしていけば8時間以内で計算できるという見込みが立ちました。オンプレミスとクラウドで同一コア数の計算速度を比較したところ、アーキテクチャの違いだけで、約2倍早くなったのは驚きました」(照井さん)

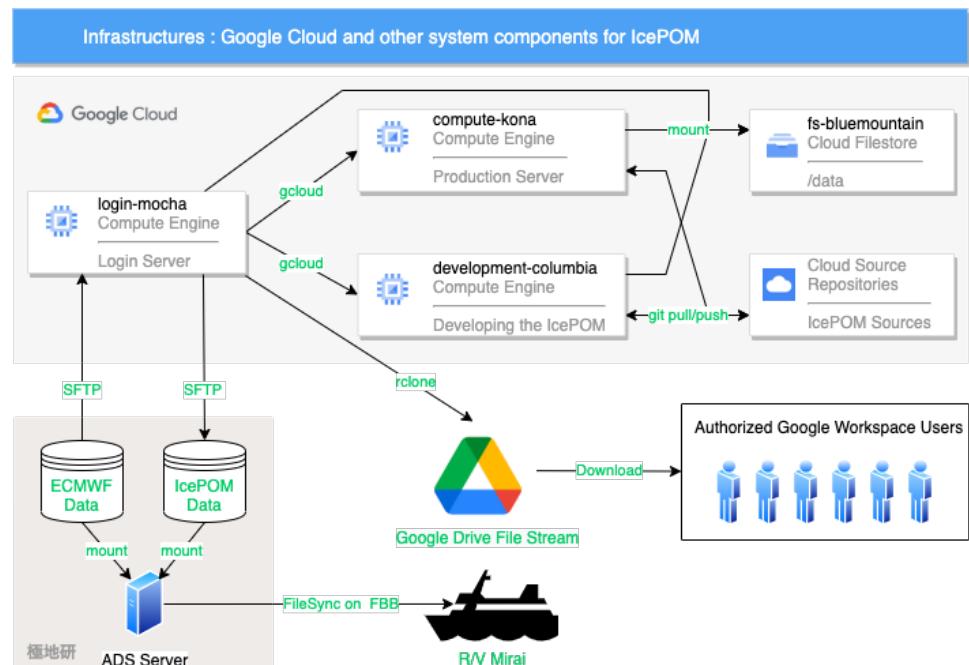
検証の結果を受けて、本番環境を構築。海氷分布予報の計算環境(compute-kona)、また開発/テスト環境(development-columbia)の仮想マシン(VM)として、Compute Engineを利用。Compute Engineは、モデルを計算する時間だけ64CPUを使い、モデルの計算が終わったらシャットダウンする仕組みにしています。

また、計算結果を保存するファイルストア(fs-bluemountain)にはCloud Filestoreを、ソースコード管理のリポジトリにはCloud Source Repositoriesを利用。さらに24時間アクセスできる最小のVM(login-mocha)からgcloudを使うことで、Google CloudのVMの制御や、Google ドライブとのデータ共有に利用しています。研究者は、Google ドライブから海氷分布予報をダウンロードして、各自の研究に利用できます。

海氷分布予報の計算結果は、login-mochaからSFTPでADSのサーバーにダウンロードされ、衛星通信により、(みらい)に転送されます。

東京大学 大学院新領域創成科学研究科 特任研究員の藤原さんは、「海氷分布予報のためのシミュレーションは、すでに作っていたプログラムを活用しています。ADSから受け取った天気予報データを加工して、シミュレーションで使えるようにする前処理から毎日決まった時間にVMを立ち上げ、計算が終わったらシャットダウンする処理、および出力された海氷分布予報をADSに送信するまでの一連のプロセスをGoogle Cloud上で自動的に行うシステムを開発しました」と話します。

Google Cloudのサポートについて照井さんは、「Google Cloudの担当者には、かなり初期の段階から、秒単位の課金やアーキテクチャの違いがモデル計算のパフォーマンスに与える影響、gcloudコマンドでVMをオン/オフする方法、ログの取得な方法など、マニアックな問い合わせも対応してもらいました」と話しています。



計画停電にも影響を受けない Google Cloud には安心感

Google Cloudの導入効果を照井さんは、次のように話します。「VMのスクラップ&ビルトが非常に楽です。本番用のVMができるまでに何とか失敗もありましたが、10台程度のVMを5分～10分程度で準備ができるため、試行作業が容易です。感動したのは、コスト計算が容易なこと。オンプレミスでは、64CPUのサーバーの購入に数百万円の予算が必要で、予報計算の計算単価を見積もるのが困難です。Google Cloudは、1回あたりの予報計算に必要な予算を1,000～2,000円程度と見積もることができました」

また藤原さんは、「より精度の高い予測のための研究に、引き続きGoogle Cloudを利用しています。Google Cloudは、CPUやメモリの設定、アーキテクチャを柔軟に変更でき、開発もテストも簡単にできるので非常に魅力的で

す。また、(みらい)航海期間中に東京大学で計画停電があり、オンプレミスのサーバーが止まったのですが、Google Cloudはそのような影響を受けないで、心理的に非常に安心感がありました」と話します。

今後の計画について山口さんは、「2021年も、(みらい)の航海は予定されているので、IcePOMの改善を続けていきます。また北極海を運航している海運会社に使ってもらい、足りない情報や要望などを研究者にフィードバックしたいと思っています。現在は海氷の予報だけですが、波の情報と組み合わせるなど、モデル計算の改良も進めていく計画なので、今後もGoogle Cloudのサポートに期待しています」と話しています。

Google Cloudを活用することで、ビジネスの将来に注力できるようになります。インフラストラクチャの管理やサーバーのプロビジョニング、ネットワークの構成などに起因する負担を軽減することができます。つまり、イノベーターもプログラマーも、自分の本来の仕事に集中することができます。

お問い合わせはこちから
<https://goo.gl/CCZL78>

