



東京工業大学
Tokyo Institute of Technology

ソリューション概要

○プロフィール

国立大学法人 東京工業大学 (<http://www.titech.ac.jp/>) は、1881年に創立された135年の歴史を持つ国内最大の理工系総合大学です。2016年4月1日には、研究力のさらなる強化に向けて約180名の研究者を擁する「科学技術創成研究院」を設置。新たなミッションを担う研究所・研究センターを設置するとともに、最先端研究を小規模のチームで機動的に推進する10の「研究ユニット」を立ち上げるなど、新分野や融合分野の研究を加速する柔軟な体制を整えています。

○導入製品とサービス

- ・ Microsoft® Azure

○メリット

- ・ Azure のコンピューティング集中型インスタンスと呼ばれる A8 ~ A11 インスタンスおよび DS14 インスタンスを活用して、大規模な並列計算を実現
- ・ 高速インフィニバンドを備えたノード間連携によって、理論値と実測値の誤差を抑制。数十 ~ 数百ノードを活用した高速計算が手軽に行える環境を実現
- ・ 日本で初めて CS ゴールドマークを取得した高度なセキュリティで、秘匿性の高い情報も安心して運用可能

○ユーザーコメント

「Azure には、A8、A9 インスタンスのリリースや日本で初めての『CS ゴールドマーク』取得といったアドバンテージがあります。膨大なデータを扱うバイオインフォマティクス研究に必要な計算能力を、厳重なセキュリティの下、安心して世界に提供できる最適な選択肢が、Azure であったと考えています。」

国立大学法人 東京工業大学
情報理工学院 情報工学系 教授
情報生命博士教育院 教育院長
工学博士
秋山 泰氏

スパコンと適所で使い分ける HPC 環境を Azure で実現。 CS ゴールドマークを取得した高いセキュリティの下、"スマート創薬" のオープン イノベーションを加速

国内最大の理工系総合大学である、国立大学法人 東京工業大学の秋山研究室では、バイオインフォマティクス分野で利用されるさまざまなソフトウェアを開発・提供しています。本プロジェクトでは、GHOST-MP と MEGADOCK を Microsoft Azure 上に移植することで、全世界のユーザーがこれらのソフトウェアを簡単に実行できる HPC (High-Performance Computing, 高性能計算) 環境を作ることに成功しました。これにより、スーパーコンピューターを所有していない大学や企業でも、Microsoft Azure を利用することで大規模な生命データ解析を行えるようになります。

Microsoft Azure 利用の背景とねらい

バイオインフォマティクスソフトの "ゴールド スタンダード" を目指し、開発したソフトウェアを世界に公開

国立大学法人 東京工業大学は、135年の歴史を持つ国内最大の理工系総合大学として、常に時代の最先端を切り拓く研究活動を行い、数多くの優秀な人材を輩出してきました。そして今、世界中の知性との連携や、産学連携によるイノベーションを加速させるために、1つの試みが進行しています。

それが、バイオインフォマティクス (生命情報科学) 研究における、パブリック クラウド サービスの活用です。

はじめは、情報理工学院 情報工学系 教授 秋山 泰 氏の研究室が開発したソフトウェア……「GHOST-MP」と「MEGADOCK」をオープンソースとして公開したことでした。

■ **GHOST-MP:** 与えられた大量の塩基配列の類似配列検索をアミノ酸配列データベースに対して行う相同性解析ソフトウェア。マルチコア、マルチノードのハイブリッド並列処理を用いることで高速な相同性配列検索を実現。

■ **MEGADOCK:** 高速フーリエ変換を用いたグリッドに基づくタンパク質ドッキングを行う構造バイオインフォマティクス ソフトウェア。ドッキング計算結果に基づいてタンパク質間相互作用を予測し、創薬に役立つ。

秋山 教授は次のように説明します。

「GHOST-MP も MEGADOCK も、私たちが自信を持って開発したソフトウェアです。このソフトウェアをさらに磨き上げ、ライフサイエンス研究における "ゴールド スタンダード" なツールとして確立させるためには、より多くの研究者に利用していただき、機能改善に役立つフィードバックや、新しいアイデアを集めることが必要不可欠です。そのために、2014年からソースコードを含め、全世界に公開してきました。しかし、ここで1つ課題がありました。

私たちは情報工学を専門とする研究者なので、これらのソフトウェアを動かす環境を作るのにそれほど苦労することはないのですが、そうではない研究者がこれらのソフトウェアを利用



国立大学法人 東京工業大学



国立大学法人 東京工業大学
情報理工学系 情報工学系 教授
情報生命博士 教育院 教育院長
工学博士
秋山 泰氏



国立大学法人 東京工業大学
情報理工学系 情報工学系 助教授
博士 (工学)
大上 雅史氏

するための環境を作ろうとしてもなかなかうまくいきません。そのたびに『うまくシステムが構築できないのでどうすればいいのか?』という問い合わせが数多く寄せられ、対応に苦慮していました。

こうした課題を解消するために秋山研究室では、「パブリック クラウド サービス上にシステムを構築し、これを利用してもらう」方法を検討し始めたのだといいます。

「GHOST-MP、MEGADOCK は、東工大が誇るスーパーコンピューター『TSUBAME 2.5』や理研神戸の『京』の上でも実行できますが、外部の利用者が気軽に利用できる体制にはなっていません。パブリッククラウド上のシステムであれば、全世界どこからでも利用できますし、私たちの作ったシステム環境を自分が購入した Microsoft Azure 上にコピーすれば、ライブラリやデータベースのバージョンの違いに悩まされることもなく、必ず動き、必ず同じ性能を再現できるはず。こうすれば、私たちへの問い合わせも減りますし、利用者も簡単に大規模情報解析の環境を手に入れることができます。」

しかし、パブリッククラウドを利用する上で、セキュリティも考慮する必要がありました。さまざまな研究機関のデータを取り扱う「共通プラットフォーム」として成立させるためには、非常にハイレベルなセキュリティを施すことが重要となるなど、技術的な要求が非常に高かったのです。

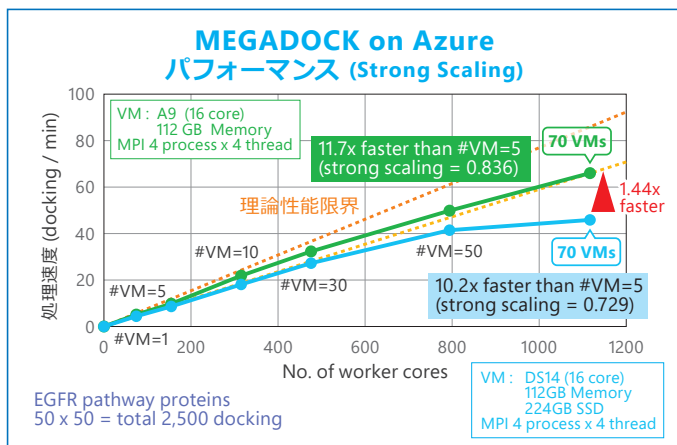


図1 MEGADOCK (A9、DS14で70ノードまで)

そして、これらすべての要件を満たすクラウド サービスとして選ばれたのが、マイクロソフトが提供する Microsoft Azure でした。Azure が選ばれたポイントは大きく2点。それが「必要十分な計算能力」の確保と、「日本初のCSゴールドマーク取得」に裏付けられた安全性と信頼性でした。

「検討の結果 Azure を選択したのは、テクノロジーに対する期待値が、ほかのサービスよりも大きかったということです。実は、マイクロソフト本社に移籍した知人から、『Azure のテクノロジーは、HPC に適している』ということ、をかなり前から聞かされてもいました。実際、VM (Virtual Machine) のノード間を高速インフィニバンドでつなぐ A8、A9 インスタンスのリリースによって、HPC 分野での Azure 活用が現実味を帯びてきました。加えて、2016年2月には、Azure が日本で初めてクラウド情報セキュリティ監査制度※に基づく、『CSゴールドマーク』を取得しています。

膨大なデータを扱うバイオインフォマティクス研究に必要な計算能力を、厳重なセキュリティの下、安心して世界に提供できる最適な選択肢が、Microsoft Azure であったと考えています。」(秋山 教授)

※「クラウド情報セキュリティ監査制度」は、JASA(特定非営利法人 日本セキュリティ監査協会)が、総務省および経済産業省と共に取り組んだ制度で、クラウド サービスプロバイダーが提供する情報セキュリティマネジメント要件(基本言明要件)の設計・実装・運用に対する監査を通じて、基本言明要件に対して約1,500項目の管理基準に基づいた情報セキュリティ監査を通じて、セキュリティ対策の信頼性を示すものです。

システム概要とパフォーマンス

すでに 100 ノード × 16 コアまで Azure の能力を実証済み。十分なハイパフォーマンスで、バイオインフォマティクス研究を加速

東京工業大学では、2015年から、Azure A9 インスタンスおよび DS14 インスタンスを活用して、VM 上に GHOST-MP と MEGADOCK の実行環境の構築を開始しました。

計算能力に関しては、環境構築後すぐに 30 から 50 の VM をつないだベンチマークを測定。VM の数に対する計算速度向上率の、理論値

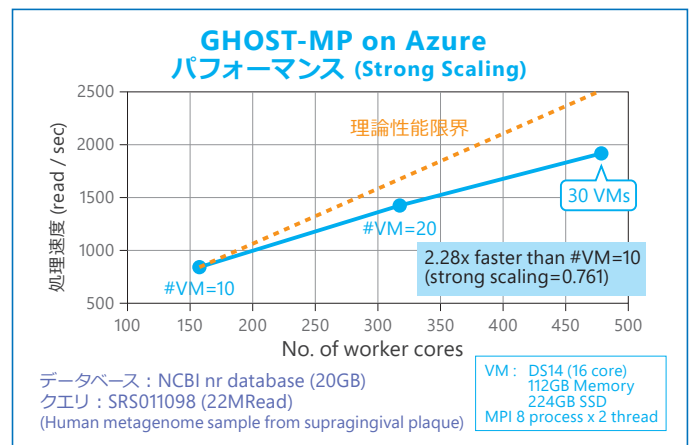


図2 GHOST-MP (DS14で30ノードまで)

に迫るスケールアップが確認されています (図1、2 参照)。

秋山 教授と共に研究を行っている、助教の大上 雅史 氏は、次のように説明します。

「現在までに、Azure 上で 100 ノードを使った計算まで試験し、検証を行ってきました。TSUBAME 2.5 の計算ノードのコア数は 12 ですが、Azure A9 の CPU は TSUBAME 2.5 が構築された当時よりも性能が上がっていますし、コア数も 16 です。このため、同じノード数でも Azure の方が優れたパフォーマンスを得られることとなります。ス

パソコンは CPU などのアーキテクチャを常に最新型に維持することはとても難しいのですが、クラウドならその心配ありません。Azure によって、最新の HPC 環境で、バイオインフォマティクス研究を進めることができるのです。」

とは言え、抜本的なイノベーションを目指し、数百～数万ノードを必要とするような超大規模計算を行うプロジェクトにおいては、「TSUBAME 2.5」や「京」などのスパコンの重要性は揺るぎません。

秋山 教授も、「2014 年に Azure の A9 などがリリースされていたのに、なかなか本学におけるクラウド活用に至らなかったのも、学内に圧倒的なスーパーコンピューターが存在していたため、クラウドに目が向いていなかった経緯もある」と振り返ります。

「クラウド上で、10～20 ノードまでの構成であれば、今は誰でも、比較的容易に入手できます。しかし、膨大な計算を必要とするバイオインフォマティクス研究において、その程度の処理能力では物足りないのが実情です。しかし、TSUBAME などのスパコンを利用するためには、大学や機関の厳格なルールに従って承認を受ける必要があります。

一方、Azure を活用した外部環境であれば、民間企業にも活用しやすいルールで提供することが可能になりますし、『広範囲な研究に役立てることができる数十～数百ノードの処理能力』も容易に構築・提供できると期待しています。

TSUBAME などのスパコンと適材適所で使い分けて、より柔軟な研究開発体制を作ることができるようになったことが、パブリック クラウド活用のポイントでしょう。」

導入効果と今後の展望

Azure を共通プラットフォームとすることで産学連携をスムーズに。"スマート創薬" の実践で、オープンイノベーションを促進

GHOST-MP と MEGADOCK を、クラウド上に展開し、「共通プラットフォーム」として広く公開したことで、東京工業大学が取り組むオープ

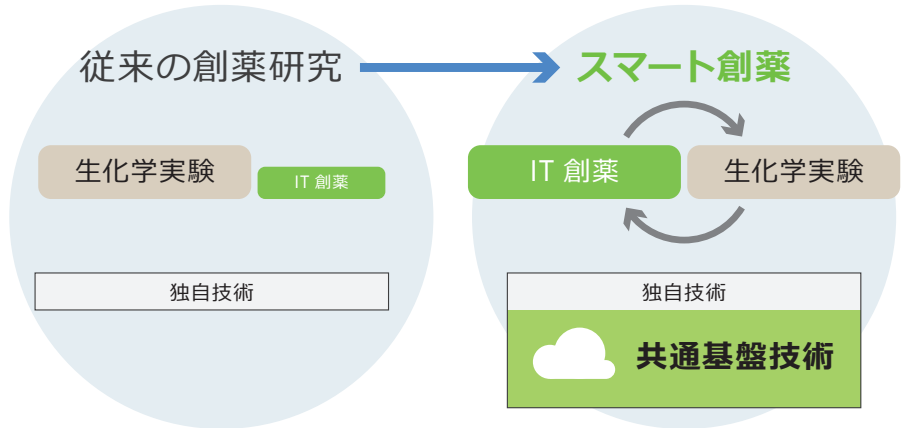


図3: スマート創薬の概念図

ンイノベーションも加速しようとしています。

その代表例が、2016年4月1日付で設置された、約180名の研究者を擁する「東京工業大学 科学技術創成研究院」の中に作られた10の研究ユニットの1つ、「スマート創薬研究ユニット」です。

スマート創薬とは、従来「IT創薬」と呼ばれてきた手法において、ITが「生化学実験を支える補助的手段」にとどまっていたことに対し、「TSUBAMEを用いた分子シミュレーションと、バイオインフォマティクス解析などをIT側のエンジンとして中心に据え、生化学実験と相互補完的に融合させる」ことを意図しています (図3 参照)。

「スマート創薬研究ユニット」を通じて、創薬における産官学連携が加速すれば、社会への多大な貢献を達成できると、秋山 教授と大上 助教は声を揃えます。

「通常、1つの薬を開発するまでに、約3,000億円もの開発費がかかると言われていました。しかも、従来は情報の秘匿性ゆえに、研究開発にかかるすべての工程を各製薬会社が独自に抱え込み、苦勞してきました。しかし、高度なセキュリティと、各国の法律を遵守してデータのプライバシーを守る Azure であれば、情報の秘匿性も守られます。私たちの提供するツールを共通プラットフォームとして提供することで、創薬コストの低減とスピード化に貢献できます。」(秋山 教授)

「薬は、病気の諸症状を引き起こす原因となるタンパク質と結合して作用します。一見簡単なくみのように思えますが、創薬の成功確率はわずか3万分の1。この確率を上げるためには、タンパク質各々が持つ性質や振る舞いの違い、私たちの細胞内の約10万種類のタンパク質それぞれとの関係(相互作用)などを考慮した、非常に複雑な予測計算が求められます。Azure を活用することで、タンパク質間相互作用予測をはじめとする、スマート創薬技術に必要な処理能力を広く共有し、製薬開発の効率化に役立てられるようになったことは、とても大きな変



化だと思えます。」(大上 助教)

東京工業大学のスマート創薬研究ユニットでは今後、企業コンソーシアムの発足や、オープン参加型の「IT 創薬コンテスト」の継続的開催を予定しています。

計算機で予測された薬のタネとなる化合物を実際に実験して有効性を検証する「創薬コンテスト」について、「Azure 活用の舞台が広がるかもしれない」と大上 助教は話します。

「コンテストの上位入賞者に、副賞として Azure を活用する権利を与えるといったことも考えられます。こうした発想ができるのも、柔軟にリソースを確保できるパブリック クラウドならではのメリットですね。」

クラウドが救う、計算環境の分断

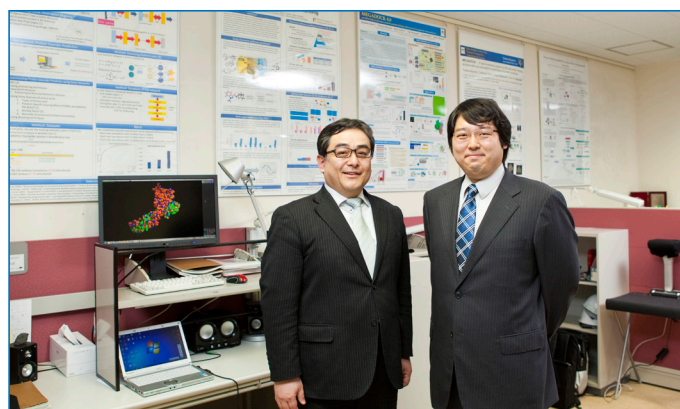
クラウドによって加速するオープン イノベーションの波は、さらに広がり、神奈川県川崎市を舞台とした産官学連携の取り組み「殿町国際戦略拠

点 キング スカイフロント」でも活かされていくだろうと、秋山 教授は言います。

「羽田空港のすぐ側に位置する川崎市殿町では、今、ライフ サイエンス分野を中心に、世界最高水準の研究開発から新産業を創出するオープン イノベーション拠点として、キング スカイフロントが立ち上げられています。本学も、この一端に参加しています。Azure 活用の場は、ここにも広がっていくのではないのでしょうか。」

最後に、秋山 教授はクラウド活用の成果について、次のように強調します。

「クラウド活用のもっとも重要な点は、特定分野のイノベーションを推進するために集まった大学や企業が、同じ計算環境を共有して活用できることにあります。現在では、私たちが選択できる計算環境はあまりに細分化されすぎました。バイオインフォマティクスは多数のツールを使うため、計算環境の差で結果が再現できないことが、しばしばありました。いわば、最初は全人類が共通の言語で会話できていたのに、それが通じなくなった旧約聖書の「バベルの塔」の混乱のようなイメージです。クラウドはこの点を救うことができると思います。『十分な計算能力』と『世界屈指のセキュリティ』および、『各国の法を厳格に順守したプライバシーポリシー』が実現している共通基盤の上で、各界の知識と技術が交流し、イノベーションを加速させていく…。私たちとしても、今後の可能性に非常に期待しているところです。」



導入についてのお問い合わせ

本ケース スタディは、インターネット上でも参照できます。 <http://www.microsoft.com/ja-jp/casestudies/>
本ケース スタディに記載された情報は制作当時 (2016 年 5 月) のものであり、閲覧される時点では、変更されている可能性があることをご了承ください。
本ケース スタディは情報提供のみを目的としています。Microsoft は、明示的または暗示的を問わず、本書にいかなる保証も与えるものではありません。

製品に関するお問い合わせは次のインフォメーションをご利用ください。

■インターネット ホームページ <http://www.microsoft.com/ja-jp/>
■マイクロソフト カスタマー インフォメーションセンター 0120-41-6755

(9:00 ~ 17:30 土日祝日、弊社指定休業日を除く)

※電話番号のおかけ間違いにご注意ください。

* Microsoft、Microsoft ロゴは、米国 Microsoft Corporation の、米国およびその他の国における登録商標または商標です。

* その他記載されている、会社名、製品名、ロゴ等は、各社の登録商標または商標です。

* 製品の仕様は、予告なく変更することがあります。予めご了承ください。

日本マイクロソフト株式会社 〒108-0075 東京都港区港南 2-16-3 品川グランドセントラルタワー