

Cloud Computing for Research

日本マイクロソフト株式会社

パブリックセクター統括本部

クラウドアーキテクト 中田 寿穂

©2016 Microsoft Corporation. All rights reserved.

<https://www.youtube.com/watch?v=TnhZqkLchIM>

バイオ医薬品の有害事象を機械学習で予測、 患者の QOL 向上に貢献する研究のシステム基盤に Microsoft Azure を採用

<https://www.microsoft.com/ja-jp/casestudies/yokohama-cu.aspx>



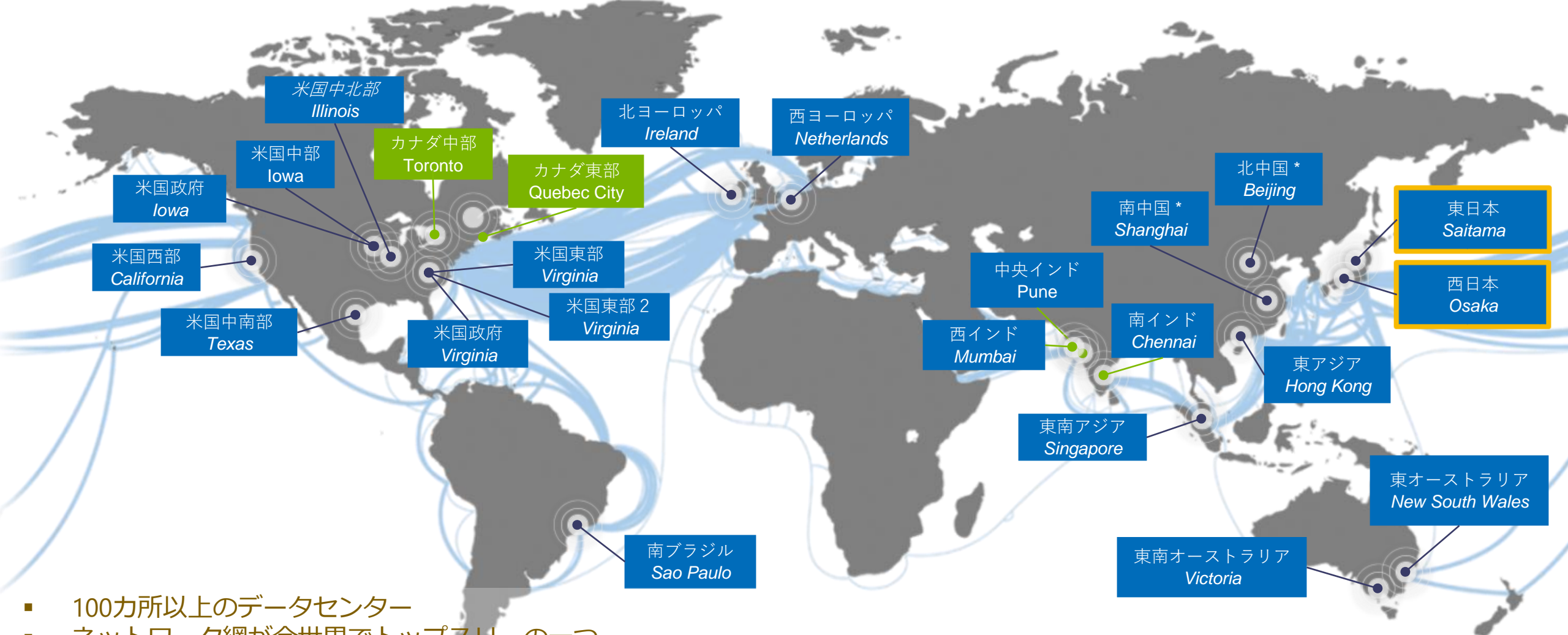
- ✓ オンプレミスと比較して柔軟性が高くコストの低いシステムを実現
- ✓ 初めて利用する学生でも簡単に機械学習の分析モデルを作成可能
- ✓ 大規模データの解析にも、Azure の機能拡張で対応できる



Microsoft Azure とは

世界最大のインフラストラクチャー

22 の地域でサービス中、24 の地域まで拡大予定



- 100カ所以上のデータセンター
- ネットワーク網が全世界でトップスリーの一つ
- AWS の2倍、Google 6倍の地域サポート
- G Series – 最大 VM 提供開始 – 32 コア, 448GB RAM, SSD...

Azure Cloud Platform

Platform Service

Security & Management

- Portal
- Active Directory
- Multi-Factor Authentication
- Automation
- Key Vault
- Store / Marketplace
- VM Image Gallery & VM Depot

Compute

- Cloud Services
- Service Fabric
- Batch
- Remote App

Web and Mobile

- Web Apps
- API Apps
- API Management
- Mobile Apps
- Logic Apps
- Notification Hubs

Developer Services

- Visual Studio
- Azure SDK
- Team Project
- Application Insights

Hybrid Operations

- Azure AD Connect Health
- AD Privileged Identity Management
- Backup
- Operational Insights
- Import/Export
- Site Recovery
- StorSimple

Integration

- Storage Queues
- Biztalk Services
- Hybrid Connections
- Service Bus

Analytics & IoT

- HDInsight
- Machine Learning
- Data Factory
- Event Hubs
- Stream Analytics
- Mobile Engagement

Data

- SQL Database
- SQL Data Warehouse
- Redis Cache
- Search
- DocumentDB
- Tables

Infrastructure Services

Compute

- Virtual Machines
- Containers

Storage

- BLOB Storage
- Azure Files
- Premium Storage

Networking

- Virtual Network
- Load Balancer
- DNS
- Express Route
- Traffic Manager
- VPN Gateway
- Application Gateway

Datacenter Infrastructure
(24 Regions, 22 Online)



Research 向けの Microsoft Azure の機能



Virtual Machine

- ✓ Windows, Linux が利用できます。
- ✓ InfiniBandやGPUを搭載したインスタンスも利用できます。



HDInsight

- ✓ Apache Hadoop の PaaS環境です。
- ✓ Spark、HBase、Storm などが利用できます。



Machine Learning

- ✓ 機械学習を簡単に始められるサービスです。
- ✓ 自由度が高く、「Python」、「R」などを使って新しいアルゴリズムを開発することも可能です。



Event Hub

- ✓ 数百万台の IoT デバイスからデータを収集できるサービスです。
- ✓ 収集したデータをリアルタイムで解析する Azure Stream Analytics なども用意されています。

Windows HPC Server (HPC Pack) とその歴史

ユーザーフレンドリーであるのみならず大規模クラスタにも対応

HPC Pack (旧称: Compute Cluster Pack)

- オンプレミス・クラウドを統合管理できるジョブスケジューラー
- MPICH2 ベースの MPI ライブラリ (MS-MPI)
- 使いやすい GUI 管理ツール
- 効率的なコマンドライン管理ツール
- Excel 高速化機能
- 最新版は HPC Pack 2012 R2 Update 2 (2015年7月リリース)



2006年 Compute Cluster Pack (HPC v1)

- 三菱UFJ証券様のクラスタがTop 500にランクイン。(1760コア、6.52TFlops)
<http://www.top500.org/system/174885>

2008年 HPC Pack 2008 (HPC v2)

- 上海スーパーコンピューティングセンターのDawning 5000AがTop500で11位にランクイン。(30,720コア、180.6TFlops)
<http://www.top500.org/system/176118>

2010年 HPC Pack 2008 R2 (HPC v3)

- 東工大のTSUBAME 2.0で初のペタフロップス越え。1.13PFlops. Top500の5位相当の記録。

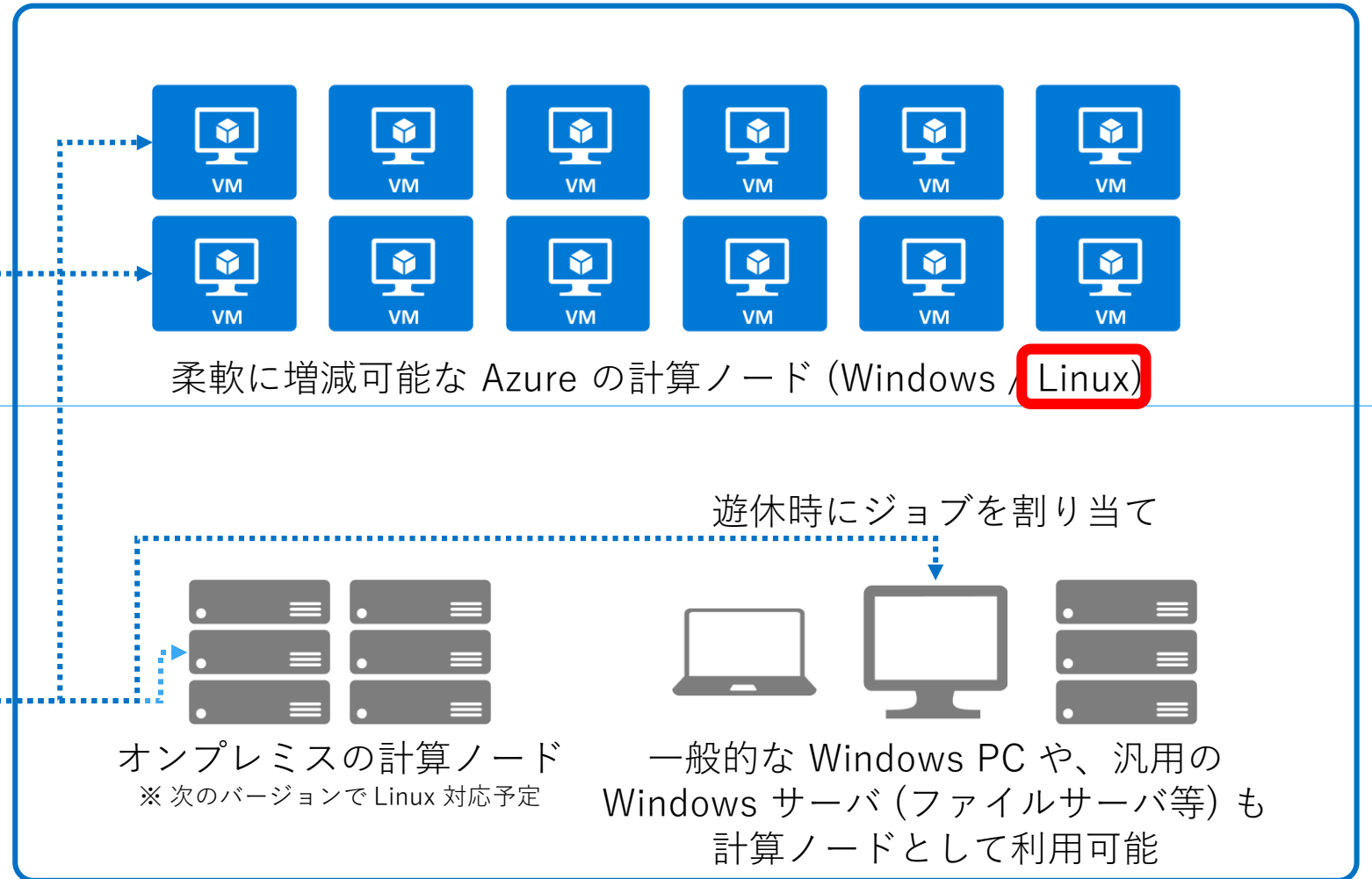
2012年 HPC Pack 2012 (HPC v4)

- HPC用AzureインスタンスでTop500にランクイン (8064コア, 151.3 TFlops) <http://www.top500.org/system/177982>

オンプレミス + クラウドの統合クラスタ

PC, サーバー, クラウド. 様々なコンピューターを計算ノードに

- 社内とクラウドの計算ノードを「一つのクラスタとして」統合管理可能
- クラウドへのノード追加・削除は、数百ノードレベルでも10分程度で完了
- スケジュールに従って自動的にノードを追加・削除することも可能



クラウド (Azure)

社内



管理ツール

ジョブ投入

ジョブの
割り当て

利用者端末

ヘッドノード
(Windows Server)

オンプレミスの計算ノード
※ 次のバージョンで Linux 対応予定

遊休時にジョブを割り当て

一般的な Windows PC や、汎用の
Windows サーバ (ファイルサーバ等) も
計算ノードとして利用可能

高性能インスタンス (A8 - A11)

InfiniBand 搭載

高速 CPU, 大容量メモリ, 高速インターコネク

サイズ	コア数	メモリ容量	プロセッサ	ネットワーク 1	ネットワーク 2
A8	8	56 GB	Xeon E5-2670 2.6 GHz	10 Gbps イーサネット	QDR InfiniBand (w/ RDMA)
A9	16	112 GB			無し
A10	8	56 GB			
A11	16	112 GB			

A8,A9 = MPI による並列計算に最適化

- TOP500 にランクインしました (2012年11月)
 - 151.3 TFLOPS (効率 90.2%) で 165 位
 - A9を504ノード, 8064 コアで実施
<http://www.top500.org/system/177982>
- MSMPI と Intel MPI に対応。

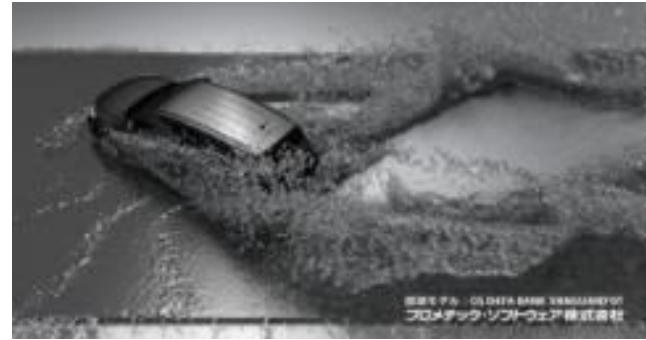
(A8,A9) - InfiniBand = A10,A11

- 「A8,A9 の速い CPU や大きなメモリは必要だが、MPI は使わないので InfiniBand 無しモデルを」という要望を受けて追加。
- InfiniBand の有無以外は A8, A9 と同一。
<https://msdn.microsoft.com/library/azure/dn689095.aspx>

検証事例: Particleworks on Azure

実施内容

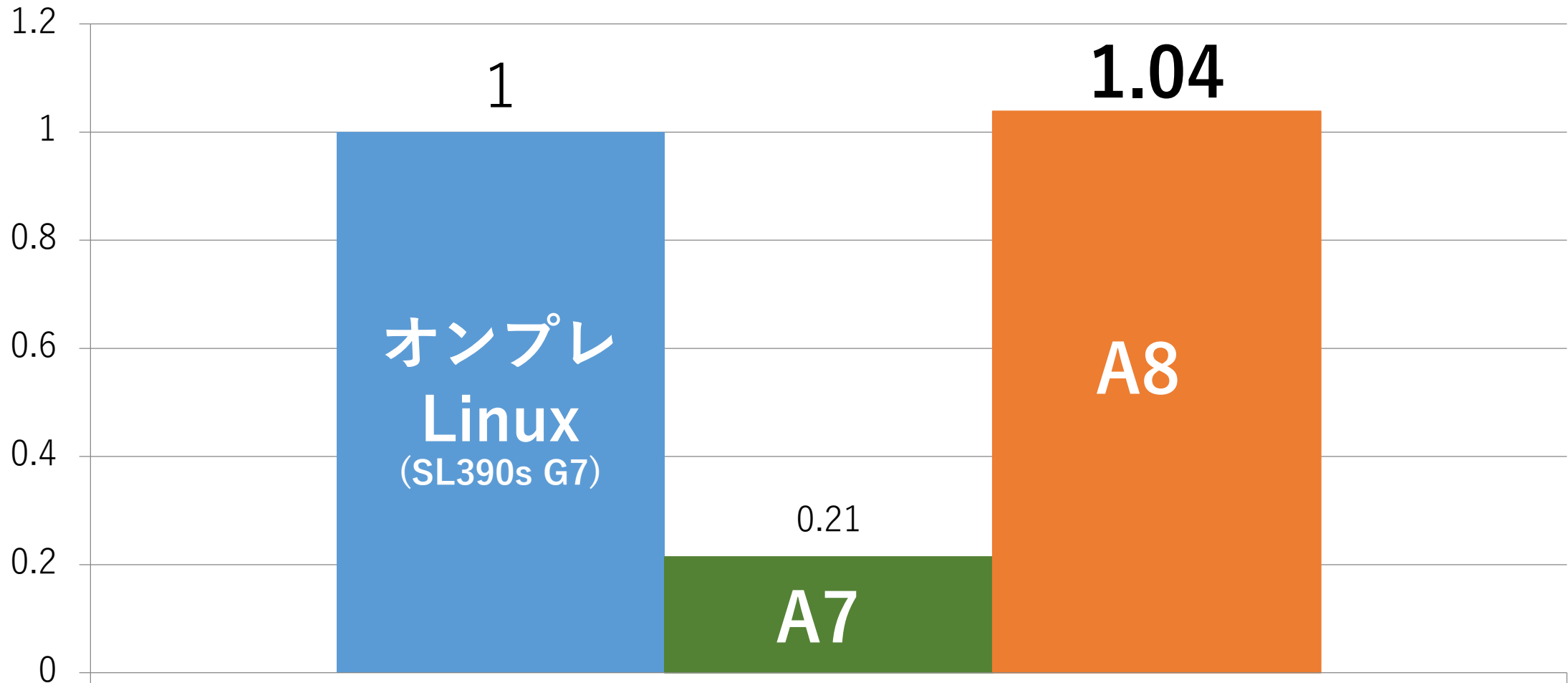
- 4000万粒子規模の解析
- 内容: 自動車の水はね



オンプレミスのLinuxマシンとの比較を実施

- オンプレミスの Linux クラスタと、AzureのA7,A8,A9インスタンスで同じ解析を実施し、実行時間を比較。
 - 機種: ProLiant SL 390s G7 x 4 ノード (計48コア)
 - CPU : Intel Xeon X5675 3.06GHz 6 cores × 2
 - RAM : 4GB × 12 = 48 GB
 - QDR InfiniBand 40Gbps × 2

オンプレミスの物理マシンと遜色ない性能



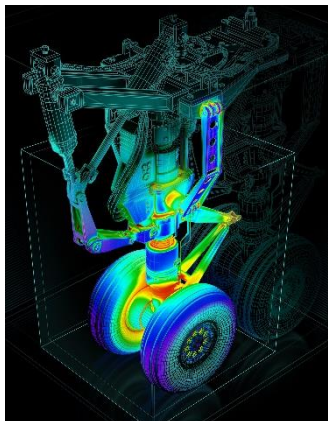
オンプレミスの InfiniBand 付き物理マシンの性能を 1 とした場合の比較
A8/A9 は物理マシンと同等、A7 との比較では 5 倍の性能を記録
(流体計算アプリケーションでの性能検証結果)

NVIDIA Tesla M60 あるいは K80 を備える「Nシリーズ」

サイズ	コア数	メモリ容量 (GB)	SSD (TB)	InfiniBand with RDMA	GPU
N1	6	64	~0.5	-	M60 x 0.5
N10	6	64	~0.5	-	K80 x 0.5
N11	12	128	~1.0	-	K80 x 1
N12	24	256	~2.0	-	K80 x 2
N21	24	256	~2.0	有り	K80 x 2

GPU インスタンスの想定用途

- ゲームなどのクラウドレンダリング
- 動画のエンコーディング
- デスクトップアプリでの OpenGL / DirectX サポート
- GPGPU (CUDA / OpenCL)



NCBI BLAST on Windows Azure

<http://research.microsoft.com/en-us/projects/ncbi-blast/>

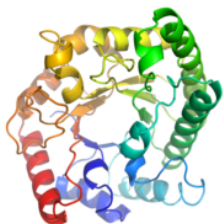
Microsoft Research

Our research Engage with us Careers About us

All Downloads Events Groups News People **Projects** Publications Videos

NCBI BLAST on Windows Azure

Making bioinformatics data more accessible to researchers worldwide



Built on Windows Azure, NCBI BLAST on Windows Azure enables researchers to take advantage of the scalability of the Windows Azure platform to perform analysis of vast proteomics and genomic data in the cloud.

BLAST on Windows Azure is a cloud-based implementation of the Basic Local Alignment Search Tool (BLAST) of the National Center for Biotechnology Information (NCBI). BLAST is a suite of programs that is designed to search all available sequence databases for similarities between a protein or DNA query and known sequences. BLAST allows quick matching of near and distant sequence relationships, providing scores that allow the user to distinguish real matches from background hits with a high degree of statistical accuracy. Scientists frequently use such searches to gain insight into the function and biological importance of gene products.

BLAST on Windows Azure extends the power of the BLAST suite of programs by allowing researchers to rent processing time on the Windows Azure cloud platform. The availability of these programs over the cloud allows laboratories, or even individuals, to have large-scale computational resources at their disposal at a very low cost per run. For researchers who don't have access to large computer resources, this greatly increases the options to analyze their data. They can now undertake more complex analyses or try different approaches that were simply not feasible before.

BLAST on Windows Azure in Action

One of the major challenges for many bioinformatics laboratories has been obtaining and maintaining the very costly computational infrastructure required for the analysis of vast proteomics and genomics data. NCBI BLAST on Windows Azure addresses that need.

Download

- [NCBI BLAST on Windows Azure](#)

Resources

- [Windows Azure](#)
- [Windows Azure team blog](#)

Project Links

- [Azure Research Engagement](#)
- [Research Projects on Windows Azure](#)

Related Links

- [National Center for Biotechnology Information](#)
- [Microsoft Brings Bioscience "BLAST" to the Windows Azure Cloud](#)
- [SC10: The Future of Discovery and the Power of Simplicity](#)
- [Technical Computing in the Cloud: Big Solutions for Big Challenges](#)
- [eXtreme Computing Group](#)

Contact Us

- [Send email to XCG Engagement](#)

Azure お試しプラン

<https://azure.microsoft.com/ja-jp/free/>

無料の Azure アカウントを今すぐ作成しましょう

¥20,500 のクレジット付きで始めて、無料オプションを引き続き使えます

[今すぐ開始](#)

または今すぐ購入する ▶

よく寄せられる質問 ▶

More questions? Call us 0120-952-593

¥20,500 のクレジットを使うと…

- ✓ 1 か月に最大で 14 台までの仮想マシン、40 個までの SQL データベース、または 8 TB までのストレージをプロビジョニングできます
- ✓ Redis Cache や Search、Content Delivery Network を使用する Web アプリ、モバイル アプリ、API アプリを作成できます
- ✓ Machine Learning、Streaming Analytics、Hadoop を使ってビッグデータを活用できます
- ✓ 監視と異常検出を利用してリアルタイムの IoT アプリを作成できます

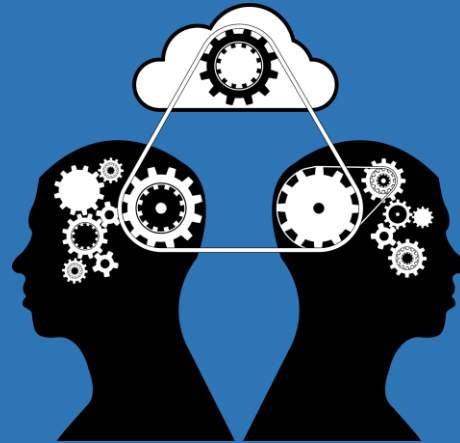
Azure なら、引き続き無料で使い続けられます

最初の試用期間のみ無料なのが一般的ですが、Azure はサブスクリプションの種類によらず、引き続き無料でご利用いただけます。

- App Service**
任意のプラットフォームやデバイス上で最大 10 個までの Web アプリとモバイル アプリを迅速に作成してホストできます。
- Machine Learning**
機械学習の実験を作成すれば、クラウドでの高度な分析の構築を今すぐ始められます。
- Visual Studio Application Insights**
ライブ アプリの使用状況とパフォーマンスを監視します。アプリを実行するホストやデバイスの数に制限はありません。
- Azure IoT Hub**
Get up to 3,000 free messages per day allowing you to monitor and control up to 10 of your Internet of Things (IoT) devices.

Azure Machine Learning

ブラウザだけで、すぐに始められる機械学習環境



Azure Machine Learning の特徴

簡単

- ✓ ブラウザ (ML Studio) だけで すぐに始められる
 - サーバー等の環境準備/設定不要
- ✓ 複雑なモデルを GUI 操作だけでも実装可能
- ✓ 作成したモデルをボタン 1つで Web サービス化

連携可能

- ✓ R / Python での実装も可能
- ✓ 各種ストレージ、データベースを入力、出力に
 - Azure Blob/Azure Table/Azure SQL Database/HiveQL/Web URL via HTTP/OData

安価

- ✓ 1時間あたり ¥102 (ML Studio)/ ¥204 (API) の従量課金
 - 実行時間に対してのみ課金される
 - サブスクリプションなしで始められる Free Tier あり

Azure ML における開発の流れ

1. トレーニングデータ（実績データ）の準備

予測モデルを作成する為のトレーニングデータ（実績データ）を準備

例：顧客属性によるお勧め商品の提示（リコメンデーション）を行いたいのであれば、こういった属性情報のお客様が、どういう商品を買ったのかという実績データが必要

2. 予測モデルの開発と評価

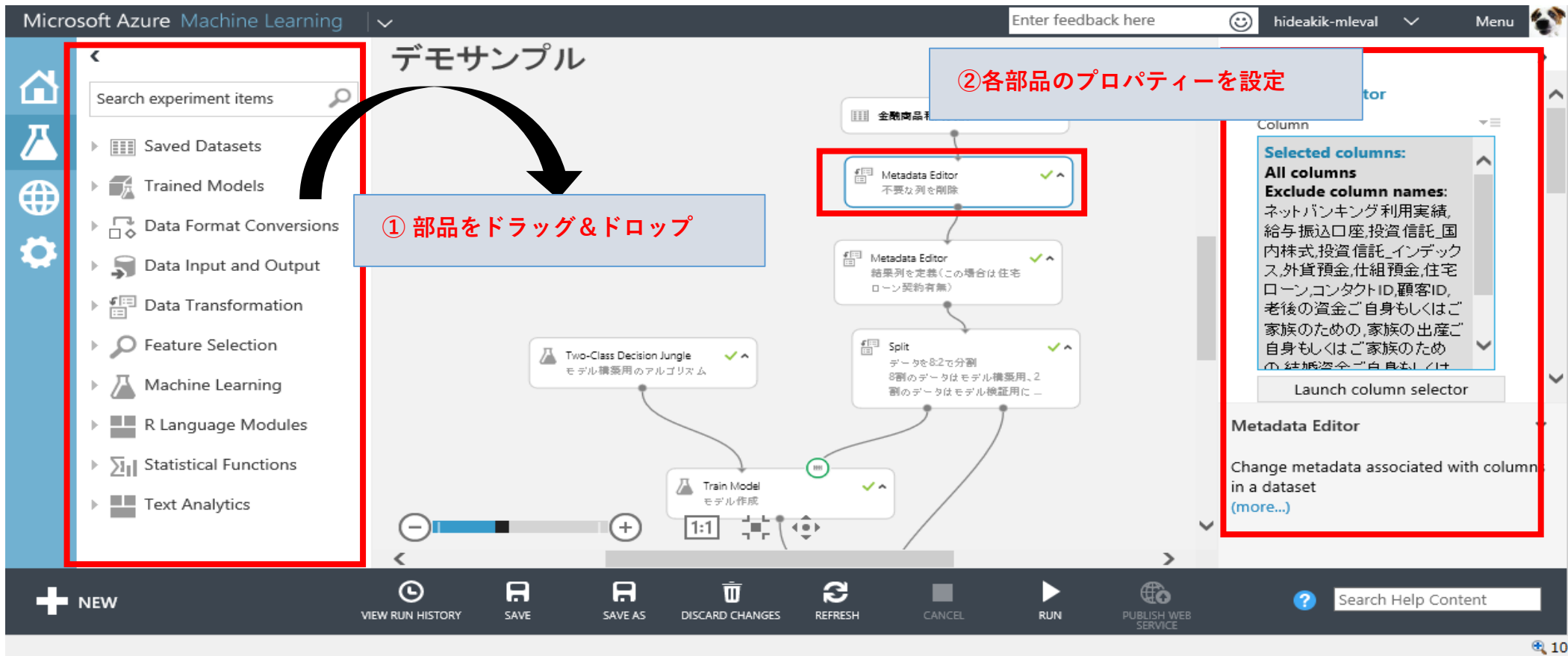
Web ベースの開発ツール「 Azure Machine Learning Studio（ Azure ML Studio ） 」を用いて、予測モデルの開発と評価を実施

3. 予測モデルの公開（Web サービス）

作成した予測モデルは、非常に簡単な操作でWebサービスとして公開可能。

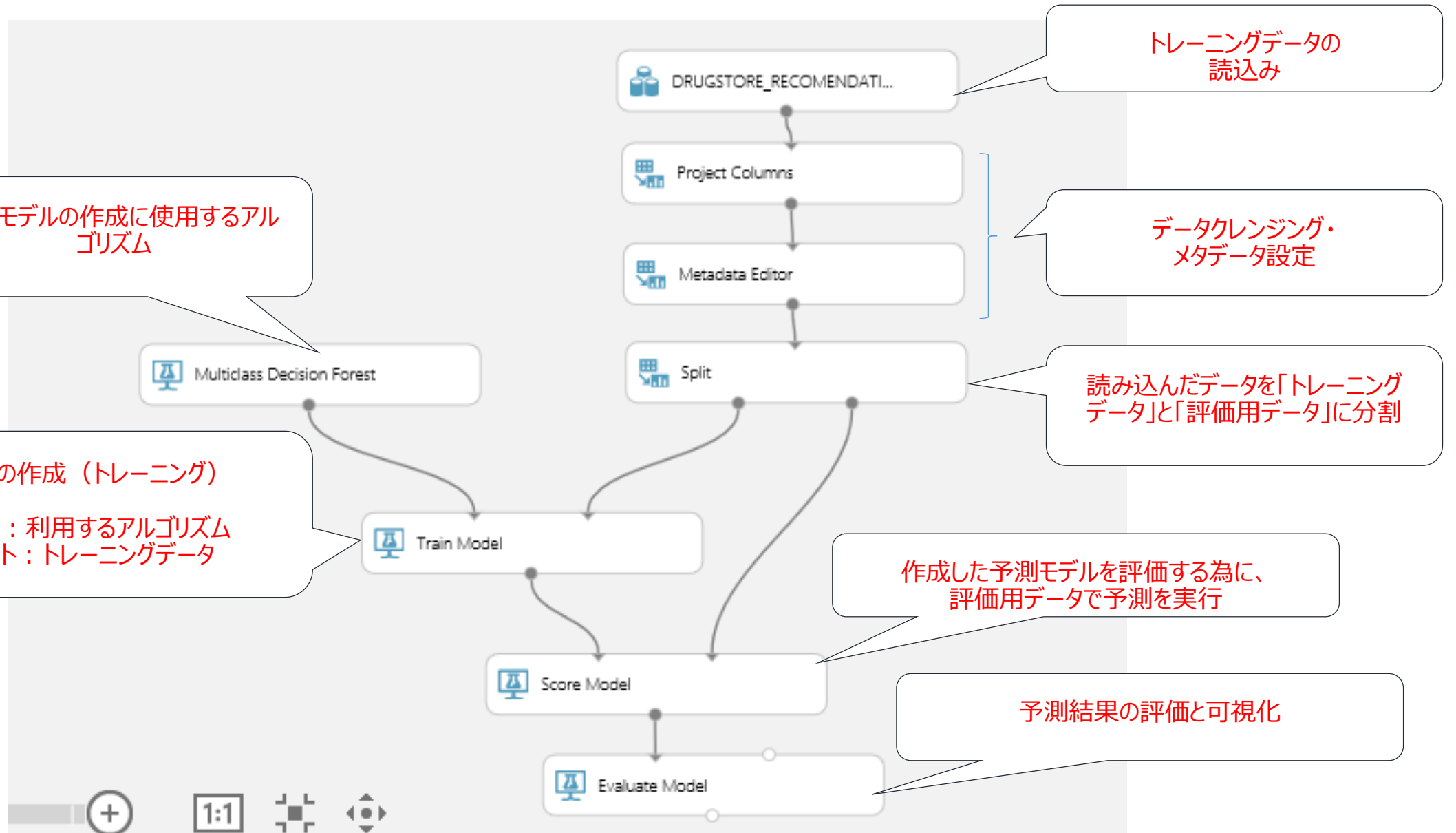
Azure Machine Learning Studio (Azure ML Studio)

- Web ベースの開発ツール
- グラフィカルな GUI で予測モデルの作成・評価を実施可能



Azure ML Studio の基本的な使い方

基本的な予測モデルの作成例



Microsoft R Server

Microsoft R Open

Microsoft R 製品：2016/01/01から製品をリブランド

Microsoft R Open

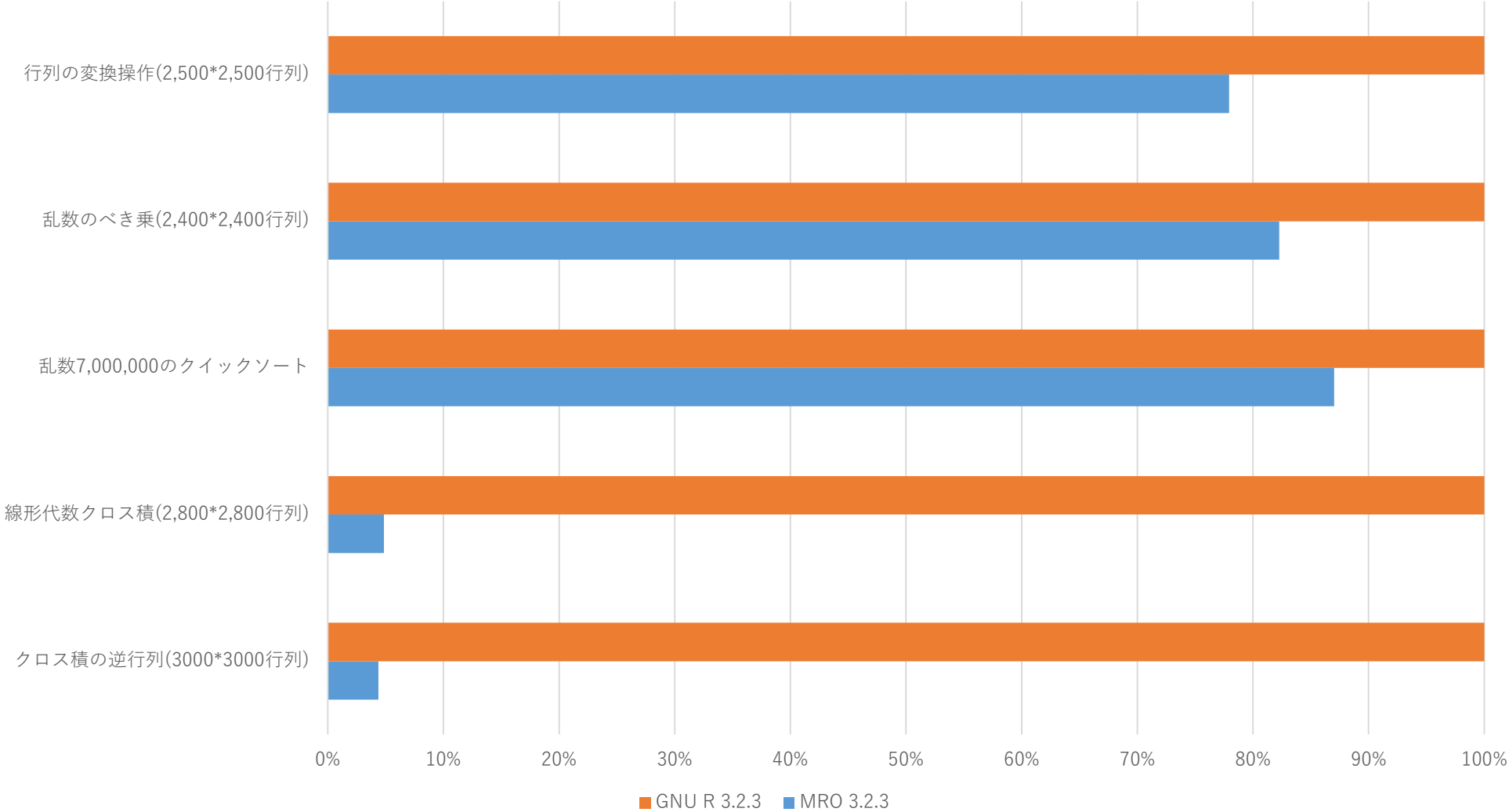
- マイクロソフトにより開発・提供されるオープンソースライセンス(GPLv2)の R ディストリビューションです。
 - GNU R と 100 % の互換性を保ちつつ、GNU R より高速※¹に動作します。
 - R のマルチスレッドよりに対応します。
 - Windows、Mac OS、Linux プラットフォームで動作します。
- ※1: インテルの数値演算ライブラリの活用により高速化を実現

Microsoft R Server, SQL Server 2016 R Service

- 複数ノードでの分散・並列処理(Scale R、Distributed R)を行うことができます。
- SQL Server、Hadoop、Teradata を使用した並列処理ができます。
- R の統合開発環境 Develop R がバンドルされます。

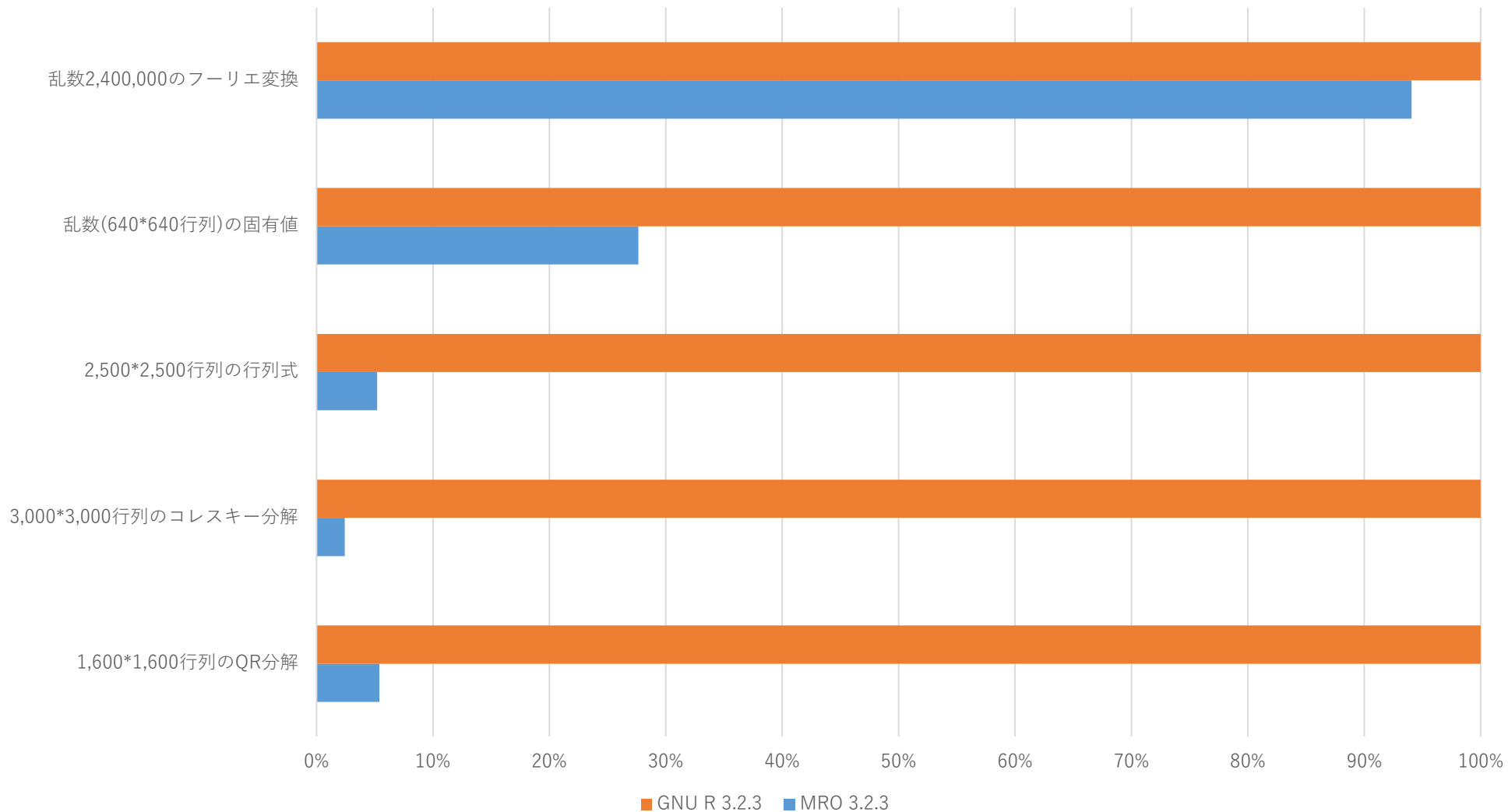
Microsoft R のパフォーマンス

Matrix calculation



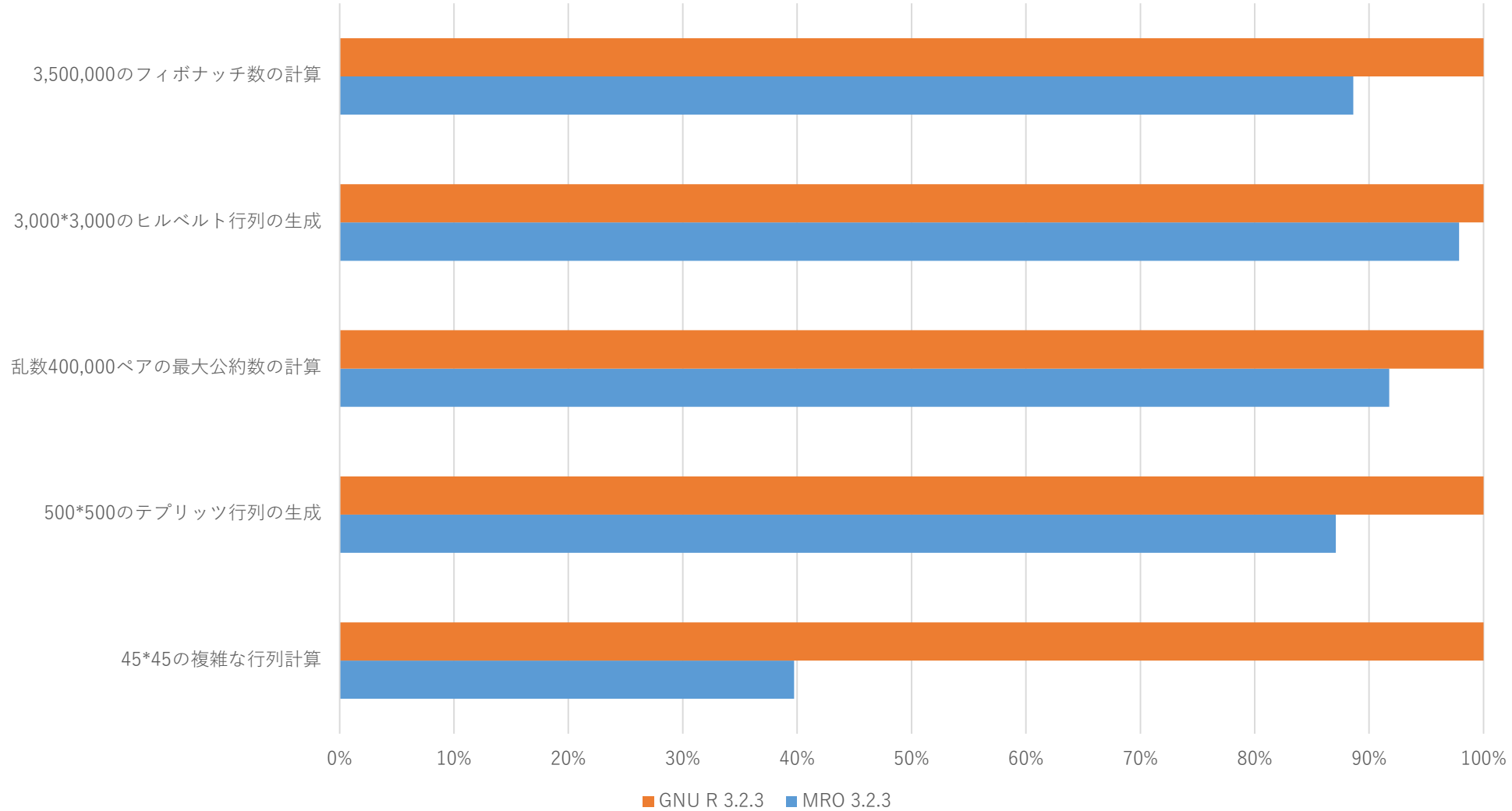
Microsoft R のパフォーマンス

Matrix functions



Microsoft R のパフォーマンス

Programmation



GNU R vs. Microsoft R Open vs. Microsoft R Server

	GNU R	Microsoft R Open	Microsoft R Server
データ処理	インメモリ	インメモリ	インメモリ or ディスク
分析スピード	シングルスレッド	マルチスレッド	マルチスレッド 1:N のサーバーに対する 並列処理
サポート	コミュニティ	コミュニティ	コミュニティ+ 商用サポート
分析のための パッケージ	7500 を超えるパッケージ (CRAN)	7500 を超えるパッケージ (CRAN)	7500 を超えるパッケージ (CRAN) + 商用の高速並列関数
ライセンス	オープンソース	オープンソース	商用ライセンス



GNU R で分析した時の課題

- メモリに収まる範囲のデータしか分析できない

GNU R は基本的にすべてのデータをオンメモリで扱うため、大規模なデータの処理や分析には適しません。

- マルチスレッド処理できない

GNU R は基本的に 1 つのコアを使用して 1 つの処理を行うため、最新のコンピュータ(マルチコア)の計算資源を有効に活用できません。

- 商用サポートなし

企業利用の場合、問題が発生しても解決が難しい場合も相談する窓口がない。

Microsoft R Server はこれらの課題を解決します。


Microsoft R Server 今後の機能拡張(予定)

- ✓ 並列分散実行環境として [Azure HDInsight](#) が利用可能に
- ✓ データソースとして [Azure Data Lake Store](#) が利用可能に
- ✓ [Azure Machine Learning](#) の R 実行環境として利用可能に
- ✓ [R Tools for Visual Studio](#) の提供で VS 上で R 開発/デバッグ可能に

- ✓ Microsoft R Server Developer Edition の提供
- ✓ Microsoft R Server への機械学習アルゴリズム追加

Microsoft R Open のダウンロード

<https://mran.revolutionanalytics.com/download/>

MRAN [About R](#) [Microsoft R Open](#) [Community](#) [Download](#) 

Download Microsoft R Open 3.2.3, the enhanced R distribution

1 Review Requirements **2** Download Software **3** Next Steps

[News File](#) | [Install Docs](#) | [Past Releases](#)

Microsoft R Open: Operating Systems Requirements

Microsoft R Open 3.2.3 is the latest release of Microsoft R Open. This release is supported on 64-bit platforms only.


Supported Platforms. Microsoft R Open is built and tested on the following **64-bit** platforms:

- Windows®:
 - Windows® 7.0 (SP1), 8.1, 10, Windows Server® 2008 R2 (SP1) and 2012
- Linux:
 - Ubuntu 12.04, 14.04, 15.04
 - CentOS / Red Hat Enterprise Linux 5.8, 6.5, 7.1
 - SUSE Linux Enterprise Server 10 & 11 & OpenSUSE 13.1
- Mac OS X:
 - Mavericks (10.9), Yosemite (10.10), El Capitan (10.11)

To see all system requirements, [read the Installation Guide](#).

Start [downloading Microsoft R Open](#).

[What's Next?](#) [Download Software](#)



ANNOUNCEMENT: Revolution R Open is now known as Microsoft R Open.

Microsoft R Server / RRE Users:

- Get [R Open for R Server/RRE 2016](#)
- Get [RRO 8.0.3](#) for RRE 7.4.1

Microsoft R Open, formerly known as Revolution R Open, is **the enhanced distribution of R** from Microsoft Corporation. Microsoft R Open is a complete open source platform for statistical analysis and data science, which is free to [download](#) and use.

The current version, Microsoft R Open 3.2.3, is based on (and 100% compatible with) the statistical language, R-3.2.3, and includes additional capabilities for performance, reproducibility and platform support. [Learn more...](#)

研究者のためのクラウド

Microsoft Azureが大学生協でお支払い可能になりました！

10,200円単位の
プリペイド形式です。

有効期限12か月です。 使い切れる
ようにお求めください。

Volume License Service Center
(VLSC) を通じて提供されます。



© 2011-2012 Microsoft Corporation All Rights Reserved.

新しい購入方法は
**クレジットカード
決済不要**
10,200円単位の
プリペイド形式です。

料金計算ツール

<http://azure.microsoft.com/ja-jp/pricing/calculator/>

サービス専用お問い合わせ、ご質問窓口

Tel.03-5665-8651 (平日9時～17時まで) 担当 竹内

※大学出張セミナーは実施回数に限りがございますので、上記窓口へ事前にお問い合わせください。

ご注文は、生協店舗まで

@2016 Microsoft Corporation. All rights reserved.

