

ORACLE



Oracle Cloud Infrastructure (OCI) の HPC と AI

PCクラスタワークショップ in 柏 2025

日本オラクル株式会社
クラウド営業統括 松山 慎



Safe harbor statement

As of August 2020, for most presentations, a safe harbor statement is no longer required, and you will not need to use this slide.

The forward looking statement safe harbor slide is required ONLY for external presentations that pertain to: product launches, executive presentations and keynotes at major events, and any event attended by financial analysts.

The notice is available at the [Safe Harbor Confluence Page](#)
Contact for Forward Looking Statement: quietperiod_ww_grp@oracle.com



Oracle Cloud Infrastructure (OCI) 概要



Oracle Cloud Infrastructure の日本国内展開

国内規格準拠や契約形態など

政府のクラウド施策対応	ISMAP登録済み、ガバメントクラウド対象クラウドサービス
準拠法	日本国の実体法と手続法を適用
裁判管轄	東京地方裁判所を第一審の専属的合意管轄裁判所とする
決済方式	契約地の現地通貨(日本円) および請求書ベースの支払い -ご利用リージョンが海外であっても適用されます。-
単価	ワールドワイドで同一の単価 (月10TB超のインターネットアウトバウンドデータ量を除く) 円建て単価 (為替リスクをオラクルが持ちます)

リージョン	大阪 (Japan Central)	東京 (Japan East)
開業年	2020年2月	2019年5月
SINETクラウド接続(*) 	○ 	○ 
クラウド相互接続  	×	○

(*) 通信データは上り/下りとも**無制限に転送データ量課金無し**



Oracle Cloud Infrastructure の SINET接続

お問い合わせ先: sinet-oracle_jp_grp@oracle.com



- 東京リージョン、大阪リージョンとも接続回線敷設済み
- 有償の仮想回線サービスであるFastConnectが必要
- SINET DCとOCIリージョン間の物理回線はOracleが敷設し、**別途ご契約不要、ご負担は¥0-**
- FastConnect/SINET経由のPrivate Peering は上り/下りとも**転送データ量はIngress/Egressともに無制限に¥0-**



圧倒的なコストパフォーマンス

Oracle Cloud Infrastructureは圧倒的低価格

	Oracle の強み	他社クラウド	Oracle	
Compute 	✓ 圧倒的低価格 ✓ CPU/メモリの高い柔軟性 (フレキシブル・シェイプ)	¥173.6868/時 仮想マシン (16vCPU, 64GB, Linux)	¥57.04/時 Compute (16vCPU相当, 64GB, Linux)	 67% 低価格
Storage 	✓ 高性能ストレージ ✓ IOPS設定 ✓ 性能に対するSLA設定	¥288,840/月 ブロック・ストレージ (1TB, 25K IOPS)	¥6,588/月 Block Volume (1TB, 25K IOPS)	 98% 低価格
Network 	✓ AD間無償 ✓ 10TB/月までデータ転送無償 ✓ 閉域網接続時は、データ転送無償	¥625,246/月 接続サービス (1Gbps, 100TB) *閉域網接続	¥24,506/月 FastConnect (1Gbps, 100TB) *閉域網接続	 96% 低価格

※ 東京リージョン 2024/9 時点の比較



OCI の HPC と AI

Oracle Cloud エコシステム全体にわたるAI

Applications

Fusion Applications

NetSuite

Fusion Analytics

Industry Applications

3rd Party Applications

AI Services



OCI Generative AI



OCI Generative AI Agents



Digital Assistant



Speech



Language



Vision



Document Understanding



Logging Analytics

ML for data platforms



OCI Data Science



AI Vector Search in Oracle Database



MySQL HeatWave Vector Store



OCI Data Labeling

Data

AI infrastructure



Compute bare metal instances and VMs with NVIDIA GPUs



OCI Supercluster with RDMA networking



Block, object, and file storage; HPC filesystems

Oracle AI Partners

Generative AI Managed Service

エンタープライズ・クラスのセキュリティ、柔軟性、ユーザーチューニングを実現した生成AIクラウドサービス

高性能の大規模言語モデルをフルマネージドで提供

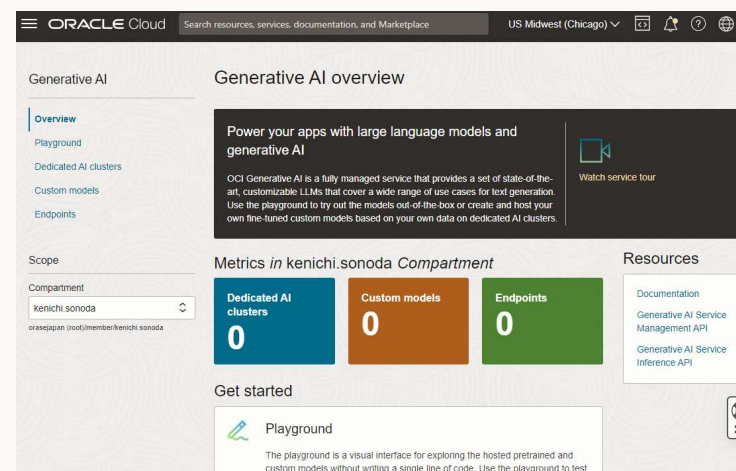
- 大手LLMプロバイダーCohere社、Meta社、xAI社のモデルをフルマネージドのPaaSとして提供
- ローコード、Web UIベースのシンプルなインタフェース

エンタープライズクラスのセキュリティとプライバシー

- 入力データや学習データがLLMプロバイダーと共有されたり、他の契約者に漏れることはありません
- 入力テキストはOCIには保存されず、マーケティングやその他の調査などに使われることはありません

専用環境でのファインチューニング

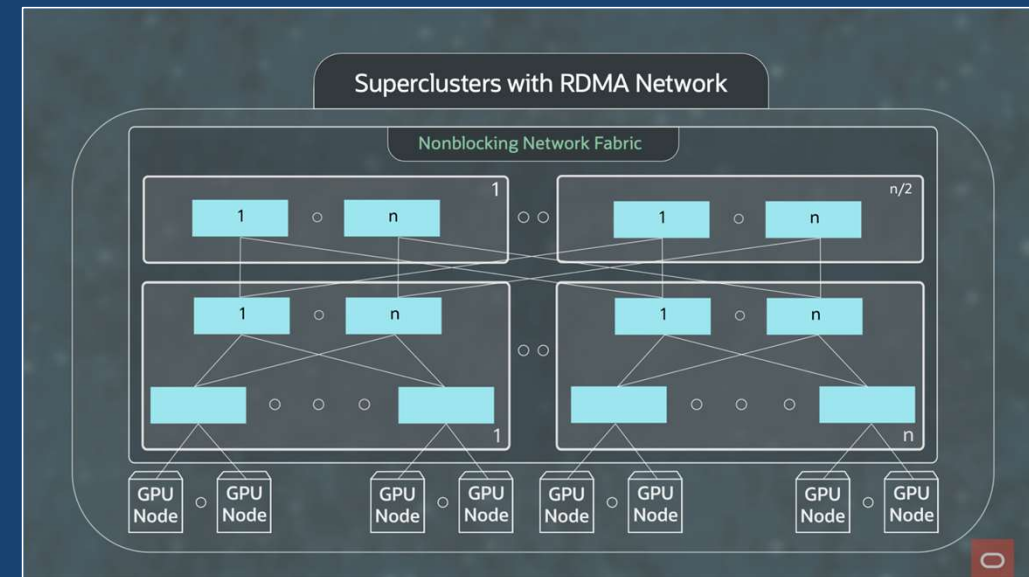
- オラクル独自のAI SuperclusterアーキテクチャによりGPUインフラの高性能を実現
- 占有環境のため、パフォーマンスと価格のプランが容易



OCI の HPC / AI インフラ

OCI Supercluster

- Super-Spine/Spine/Leafトポロジによるフルバイセクションバンド幅を備える低遅延の大規模RDMA(RoCEv2)ネットワーク



ゼタスケール_(FP4)のクラウドGPUクラスタ

NVIDIA GB300もしくはB300 x **131,072GPU** 搭載のOCI Super Clusterの受注開始

GPU タイプ	OCI Supercluster	A社	B社	C社
NVIDIA A100	32,768	10,000	公開情報無し	公開情報無し
NVIDIA H100	16,384	20,000	14,400	公開情報無し
NVIDIA H200	65,536	公開情報無し	公開情報無し	公開情報無し
NVIDIA B200	131,072	公開情報無し	公開情報無し	公開情報無し
NVIDIA GB200	100,000+	20,736	公開情報無し	公開情報無し
NVIDIA B300	131,072	公開情報無し	公開情報無し	公開情報無し
NVIDIA GB300	131,072	公開情報無し	公開情報無し	公開情報無し





マイクロソフト、Bingの対話型検索にOracle Cloud Infrastructureを活用

Press Release

マイクロソフト、Bingの対話型検索にOracle Cloud Infrastructureを活用

テキサス州オースティン—2023年11月9日

（本資料は米国2023年11月7日にオラクル・コーポレーションより発表されたプレスリリースの抄訳です）

オラクルは本日、AIサービスの急激な成長を支えるため、マイクロソフトと複数年にわたる契約を結んだことを発表しました。マイクロソフトは、Microsoft Bingの対話型検索向けに最適化されたAIモデルで推論を行うために、Microsoft AzureのAIインフラストラクチャとともに「Oracle Cloud Infrastructure (OCI)」のAIインフラストラクチャを使用しています。マイクロソフトは、「Oracle Interconnect for Microsoft Azure」を活用することで、Azure Kubernetes Service (AKS) などのマネージド・サービスを使用した大規模な「OCI Compute」のオークストレーションが可能となり、Bingの対話型検索の需要増加をサポートできます。

Bingの対話型検索では、Bingの推論モデルが検索結果を評価し分析するための、強力なコンピューティング・インフラストラクチャのクラスタが必要となります。

Oracle Cloud Infrastructure担当シニア・バイスプレジデントのカラン・バッタ（Karan Batta）は、次のように述べています。「生成AIは画期的なテクノロジーの飛躍です。オラクルは、マイクロソフトをはじめとする何千もの企業が、OCIのAI機能を使って新しい製品を構築し、実行できるようサポートしています。この度、マイクロソフトとの協力関係がさらに強化されたことで、世界中のより多くの人々に新しいエクスペリエンスを提供できるようになります。」

マイクロソフト Search & AIマーケティング担当グローバル責任者のディヴィヤ・クマール（Divya Kumar）氏は次のように述べています。「Microsoft Bingは、最新のAIテクノロジーを活用し、格段に優れた検索エクスペリエンスを世界中の人々に提供しています。オラクルと協業し、「Oracle Cloud Infrastructure」をMicrosoft AzureのAIインフラストラクチャとともに使用することで、より多くの顧客へのアクセスが可能になり、多くの検索結果のスピードを向上させることができます。」

推論モデルには、何千ものコンピュート・インスタンスとストレージ・インスタンス、そして並列動作が可能な何万ものGPUが必要です。これらは、マルチテラビットのネットワーク上で単一のスーパーコンピュータとして機能します。

OCI Superclusterについて

「OCI Superclusters」には、「OCI Compute Bare Metal」インスタンス、超低レイテンシRDMAクラスタ・ネットワークング、および選択可能なHPCストレージが含まれています。「OCI Superclusters」は、32,768個のA100 GPUまたは16,384個のH100 GPUを搭載し、最大4,096個の「OCI Compute Bare Metal」インスタンスまで拡張できます。また、超並列アプリケーションを効率的に処理するために、ペタバイト級のハイパフォーマンス・クラスタ・ファイル・システム・ストレージまで拡張できます。

Oracle Interconnect for Microsoft Azureについて

「Oracle Interconnect for Microsoft Azure」は、お客様にマルチクラウド環境へのシンプルなパスを提供します。お客様は、OCIのAIインフラストラクチャ、「Oracle Exadata Database Service」、「Autonomous Database」、「MySQL HeatWave」などの「Oracle Database」の機能をご利用いただけます。また、シームレスな相互運用性により、OCIとAzure両方の長所を活かしてのイノベーションが可能となります。Interconnectでは、2つの主要なクラウド・プロバイダー間に低レイテンシのプライベート接続を提供して、投資収益率を最大限に拡大しながら、柔軟なイノベーションを実現します。Interconnectの価格はポート単位で、消費した帯域幅に対する追加料金は発生しません。





NVIDIA、DGX Cloudの基盤に、Oracle Cloud Infrastructureを採用

Press Release

NVIDIA、AIサービスにOracle Cloud Infrastructureを採用

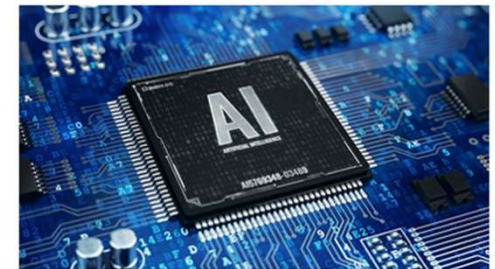
NVIDIA ネットワークで接続されたOCI Supercluster™上でNVIDIA AI FoundationsとDGX Cloudが稼動

OCI Superclusterは、世界最大規模のスーパーコンピュータを上回り、32,768のGPU、4,096コンピュート・ペアメタル・インスタンスまで拡張可能に

テキサス州オースティン—2023年3月30日

(本資料は2023年3月21日にオラクル・コーポレーションより発表されたプレスリリースの抄訳です)

オラクルは、NVIDIAとのコラボレーションを拡大し、新しい **Oracle Cloud Infrastructure (OCI) Supercluster™** 上での戦略的NVIDIA AIアプリケーションの実行を可能にしました。NVIDIAは、AIスーパーコンピューティング・サービスであるNVIDIA DGX Cloud™を大規模に提供する最初のハイパースケール・クラウド・プロバイダーとしてOCIを選択しました。また、NVIDIAは、DGX Cloudを通じて提供する新しいジェネレーティブAI・クラウド・サービスである「NVIDIA AI Foundations」をOCIで稼働させています。



Oracle Cloud Infrastructureのエグゼクティブ・バイスプレジデントであるクレイ・マゴウイク (Clay Magouyrk) は次のように述べています。「OCIは、あらゆる業界の何千ものお客様にAIスーパーコンピューターを大規模に提供する最初のプラットフォームです。今後より多くの組織がビジネス独自のAIユースケースにコンピューティング・リソースを必要とするようになることを考えると、これは大変重要な機能です。このようなお客様の需要をサポートするために、オラクルはNVIDIAとのコラボレーションを拡大し続けています。」

NVIDIAのエンタープライズ・コンピューティング担当バイスプレジデントであるManuvir Das氏は次のように述べています。「AIによるイノベーションの可能性は無限大であり、AIはさまざまなビジネスの変革に大きな影響を与えています。NVIDIAとOracle Cloud Infrastructureのコラボレーションにより、NVIDIAのアクセラレーテッド・コンピューティング・プラットフォームの驚異的なスーパー・コンピューティングの性能を、あらゆる企業が利用できるようになります。」

OCIの新しいSupercluster

NVIDIA DGX Cloudと、その上で稼働する NVIDIA AI Foundations サービス は、NVIDIAからDGX Cloudの高い基準を満たすための認定を受けたOCI独自のSuperclusterを活用しています。

OCIのSuperclusterには、OCI コンピュート・ペアメタル、NVIDIAネットワークに基づく超低レイテンシRoCEクラスタ、および選択可能なHPCストレージが含まれています。NVIDIAによってデプロイ・検証され、超並列アプリケーションを効率的に処理できる数千のOCI コンピュート・ペアメタル・インスタンスをサポートしています。OCI Superclusterネットワークにより、32,768のA100 GPU、4,096個のOCI コンピュート・ペアメタル・インスタンスまで拡張可能になりました。現在、NVIDIA H100 GPUs を搭載したOCI コンピュート・ペアメタル・インスタンスは限定的に提供されています。

また、NVIDIAは、オラクルがNVIDIA BlueField-3 DPUをネットワーク・スタックに追加することを発表しています。

カスタム・エンタープライズ・モデルを構築するためのジェネレーティブAIサービス

NVIDIA AI Foundationsのモデル構築サービスは、言語、画像、ビデオ、3D、さらに生物学にまで及びます。企業は、NVIDIA NeMo™言語サービスとNVIDIA Picasso画像、ビデオ、3Dサービスを利用して、インテリジェントなチャットやカスタマーサポート、プロフェッショナル・コンテンツ作成、デジタル・シミュレーションなどのための、独自のドメイン特化型ジェネレーティブAIアプリケーションを構築することができます。また、NVIDIA BioNeMo™は、生命科学分野でのAIモデル構築を支援し、タンパク質構造予測、遺伝子発現解析、薬剤発見などの研究に役立ちます。



File Storage Service with Lustre

オープンソース並列分散ファイルシステム Lustre のフルマネージド・サービス

並列分散アーキテクチャにより大容量・高性能・高可用性を実現

- 並列分散アーキテクチャで構築され、VCN 上の数千のクライアントからの並列アクセスに対しても高スループットを実現。必要に応じて容量 (最大 20 PB) やそれに伴う性能を拡張可能。
 - ※サービス制限初期値: ADごと総容量200TB、総帯域200Gbps
- トレーニングデータセット、研究モデル、チェックポイント、分析用大規模データなど、AI/ML/HPCのデータ保存に効果的
- ファイルシステム作成時に性能層を指定 (作成後変更不可)

選択可能な性能層

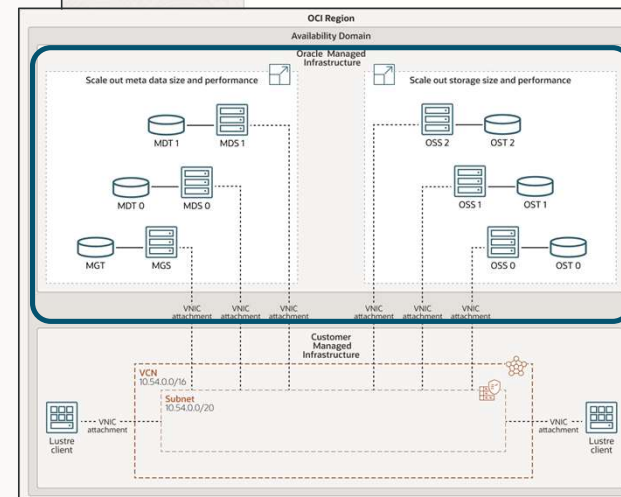
125 MB/s/TB (1Gbps/TB)
250 MB/s/TB (2Gbps/TB)
500 MB/s/TB (4Gbps/TB)
1 GB/s/TB (8Gbps/TB)

フルマネージド・サービス

- ストレージサーバー、メタデータサーバー、データボリュームといったLustreの基盤コンポーネントのデプロイとメンテナンスを自動化。

クライアント対応: Lustre 2.15.5 クライアントのデプロイが必要

- Ubuntu 22.04 (5.15.xカーネルを使用)
- Oracle Linux 8 (Red Hat Compatible Kernel (RHCK) 4.8 を使用)



内部的にデプロイされているコンポーネントの例



Gfarm laC for OCI

筑波大学 / NPO法人つくばOSS技術支援センター

TerraformおよびAnsibleでlaCが整備され OCI環境にGfarmを構築可能

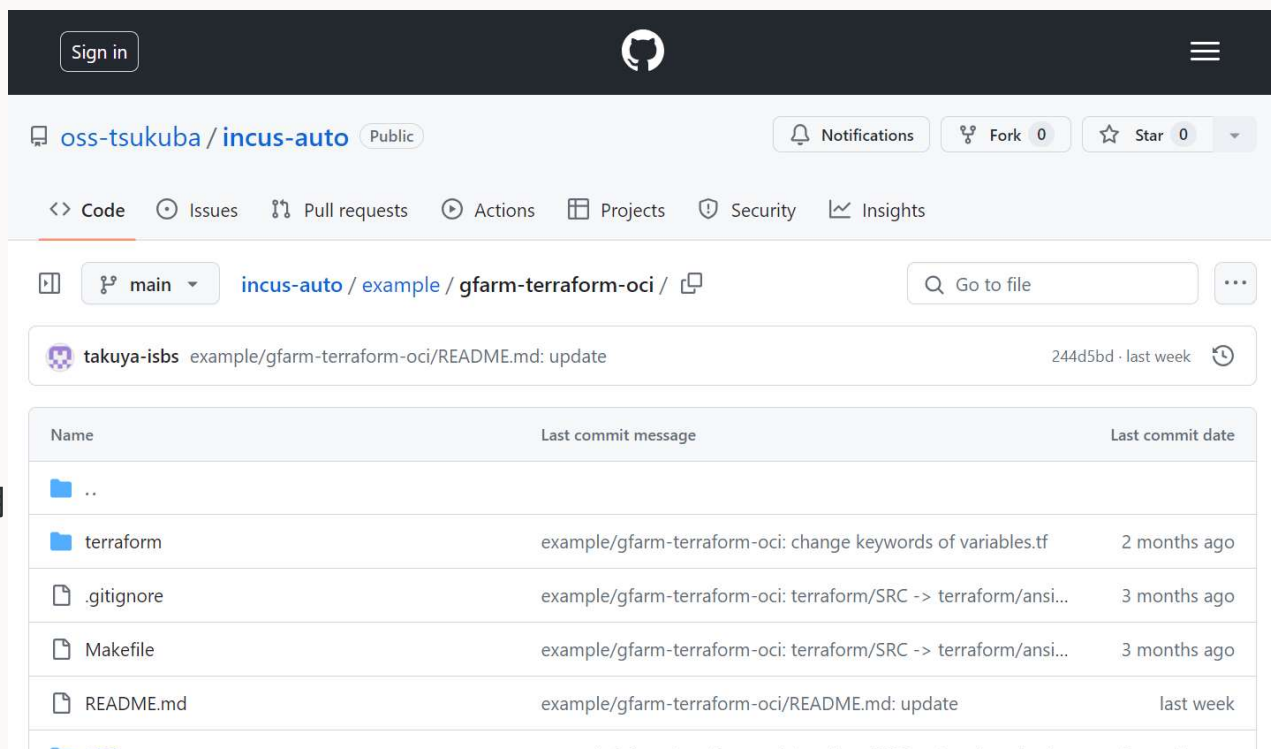


Gfarmファイルシステム：筑波大 建部教授がオープンソースソフトウェアとして開発し、NPOつくばOSS技術支援センターなどがサポートする分散ファイルシステム。

Gfarm の特徴：

- 広域分散共有ファイルシステム
- 複数同時アクセスの負荷分散
- 遠隔レプリケーション

HPCI共用ストレージやJLDGで大規模に採用



<https://github.com/oss-tsukuba/incus-auto/tree/main/example/gfarm-terraform-oci>



Open OnDemand

- Open OnDemandはHPCの容易な利用を可能にするオープンソースWebポータル
- Ohio Supercomputer Centerが開発
- アイコンをクリックするだけで、アプリケーションの実行やファイルのアップロード等が可能

OCIは利用可能なCloud Providerです。



A screenshot of the Open OnDemand website. The left sidebar is dark with the "OPEN OnDemand" logo and a list of links: "Run Open OnDemand", "Administer Open OnDemand", "Be a Contributor", "Get Involved", "About Us", "Support", and "Resources". The main content area has a light blue background. At the top, there's a newsletter subscription bar. Below it, three columns of text describe the ease of use: "client software installation required.", "Open OnDemand easy to learn and use.", and "a browser—even a mobile phone or tablet." The main heading is "Where to use Open OnDemand". Under this, there are two sections: "My Local Institution" with a building icon and "Cloud Providers" with a cloud icon. The "Cloud Providers" section lists "AWS", "Google Cloud", "Azure", and "Oracle (OCI)", with "Oracle (OCI)" highlighted by a red box. At the bottom of the "Cloud Providers" section, it says "For available support, but with less configurability: ACCESS".



Globus

Globusによって開発されたグリッド・コンピューティングを構成するためのオープンソースミドルウェア

Globusの機能

- 資源管理 (グリッド資源管理プロトコル : GRAM、Grid Resource Management Protocol)
- 情報サービス (監視と発見サービス : MDS - Monitoring and Discovery Service)
- データ移動と管理 (二次記憶への広域アクセス : GASS - Global Access to Secondary storage)
 - GridFTP: グリッドコンピューティング用ファイル転送プロトコル
 - Globus Connect :さまざまなストレージを接続してデータを転送・共有

OCI は Globus Connector partner です。

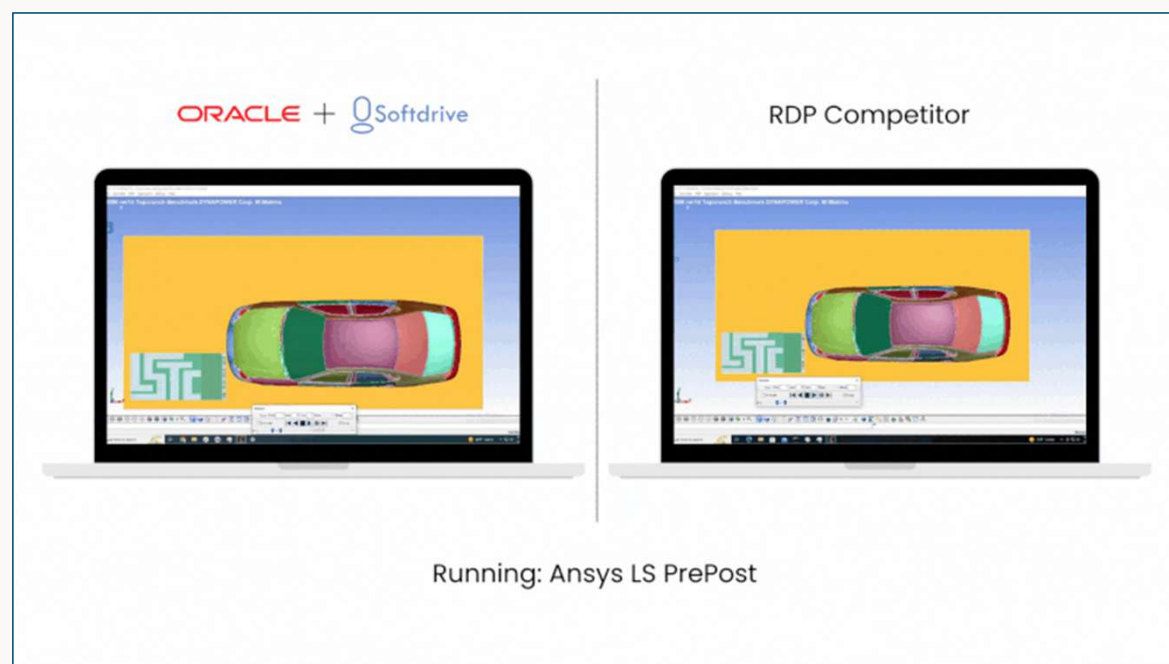


A screenshot of the Globus website. The header is dark blue with the Globus logo and the text 'a uchiicago non-profit service'. A menu icon is in the top right. The main content area is white and lists 'Current partners in our S3-compatible storage ecosystem'. The list includes: Cloudian HyperStore, DataCore Swarm, Dell ECS, Oracle Cloud Infrastructure (OCI) Object Storage (highlighted with a red box), Pure Storage FlashBlade, Storj, and Wasabi. Below the list is a section for 'globus connector' with a logo and text: 'If you are interested in using other storage systems with the Globus S3-compatible storage systems'.

低遅延/高速描画クラウドデスクトップ Softdrive



- 低遅延で快適なユーザエクスペリエンスをクラウドデスクトップで実現
 - OCI上で仮想GPU PCの最適化
 - 独自通信プロトコルとUDP通信による遅延隠蔽
- 柔軟な構成でコスト削減と運用効率化を実現
 - 仮想GPU PC台数課金
 - 仮想GPU PC共有利用が可能
- 強固なセキュリティ実装
 - SOC 1&2を取得済み
 - End2Endの暗号化と多要素認証



Oracle Cloud Lift Services

- OCI ... Oracle Cloud Infrastructure
- ExaDB-D ... Oracle Exadata Database Service
- OCVS ... Oracle Cloud VMware Solution

サービス概要

- お客様のクラウド移行をご支援する**無償**サービス
- 現行システムのクラウドへの移行検討で妨げとなる課題に直面しているお客様をご支援

お客様のベネフィット

1. クラウド移行に精通したオラクルの**専任エンジニア**が、移行のご支援をいたします
2. OCIのノウハウがないお客様でも、環境を**迅速に構築**し、ご利用頂くことができます
3. お客様がクラウド環境の準備に伴う**初期費用や時間を抑える**ことができます
4. 併せてお客様プロジェクトメンバーに対して、クラウド環境の利用・運用するための知識を持てるよう、**スキルトランスファー**のご支援をさせていただきます

主なサービス内容

クラウド移行に向けた“**ケーススタディ支援**”

クラウドへの移行事例を共有し、新たな観点でのクラウドメリットを可視化

クラウド移行に向けた“**フィジビリティスタディ支援**”

現行システムの評価を検証し、クラウド移行の課題とロードマップを可視化

クラウド移行に向けた“**実機検証 (PoC) の支援**”

PoCを通じて、クラウド移行の問題やリスクに対するソリューションを検証

プロジェクト初期フェーズの“**早期立ち上げ支援**”

クラウドサービスを利用して、検証・開発環境の導入、立ち上げをご支援

主なワークロード

Oracle Database On Cloud

Oracle DBを
OCI の PaaSに移行する

VMware on Cloud

VMware環境を
OCI の OCVS に移行する

High Performance Computing applications on Cloud

HPC アプリケーションを
OCI の HPC環境に移行する

※その他ワークロードに関しましては、担当営業にご相談ください



ORACLE