

ORACLE

Gfarm WS 2025

OCI IaaS Update と File Storage Service IO500測定

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

OCI Compute X12 Standard



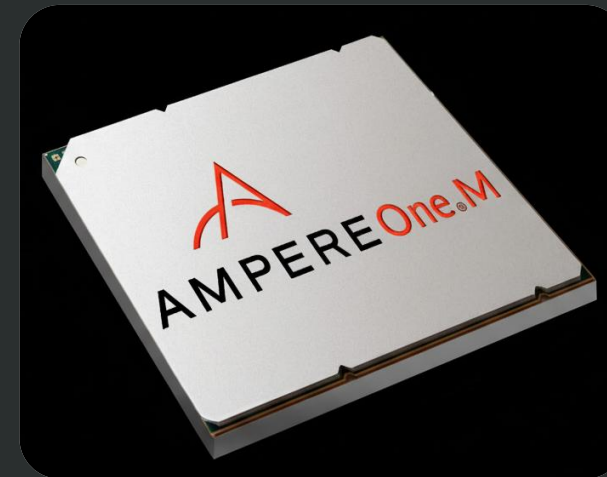
Intel Xeon 6 (Granite Rapids) 120 core
1152 GB DDR5 memory
100 Gbps network



OCI Compute A4 Standard

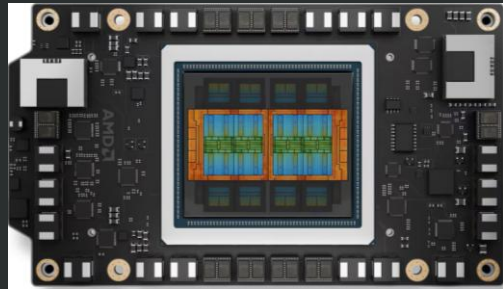


AmpereOne M (Banshee12) 48 core x 2 socket
768 GB DDR5 memory
100 Gbps network





OCI Compute MI355X



フロントエンドネットワーク
400Gbps
バックエンドネットワーク
3.2Tbps



OCI Compute NVIDIA GB300 NVL72



フロントエンドネットワーク
7.2Tbps (NVIDIA BlueField)
バックエンドネットワーク
57.6Tbps (NVIDIA ConnectX-8)

OCI Compute MI455X



フロントエンドネットワーク
400Gbps
バックエンドネットワーク
3.2Tbps

OCI Compute NVIDIA RTX Pro 6000 Blackwell Server Edition



フロントエンドネットワーク
400Gbps (NVIDIA BlueField)
バックエンドネットワーク
3,2Tbps (NVIDIA ConnectX-8)

Oracle Acceleron

ORACLE
CLOUD
Infrastructure

ANNOUNCING



Acceleron

Network

Storage

Security

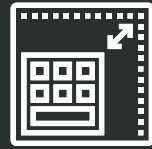
Acceleronは、入出力の保護と高速化を実現するOCIのアーキテクチャ

ネットワーク仮想化をHWで実現するOff-Box をAMD Pensando DPUおよびNVIDIA BlueField-3 DPUに変更

- 部品点数の削減
- ネットワーク帯域の拡大
- NVMe-oFによるブロックボリュームの高速化、耐性強化
- DPUのスイッチ機能を利用しCLOS階層削減
- NICの無停止パッチ適用

OCI Zettascale10

クラウド最大の
AIクラスター



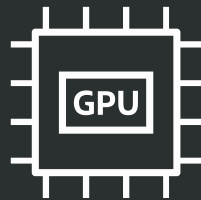
Power
1.5 GW



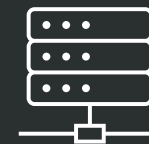
Campus size
1.5 KM diameter

1010
1010

Up to
16 ZFLOPS



800,000+
GPUs
in one OCI
Supercluster



Oracle Acceleron
FE **131 Pb/s**
RDMA **2.1 Eb/s**
Latency **<10 μs**

NVIDIA と Oracle、米国エネルギー省最大の AI スーパーコンピューターを構築

NVIDIA Blackwell x 100,000 GPU が、アルゴンヌ国立研究所の AI を活用した科学の時代を切り開く

- 米国エネルギー省 (DOE) 最大の AI スーパーコンピューターを構築する NVIDIA と Oracle の画期的なコラボレーションを発表。
- Solstice システムは、NVIDIA Blackwell 10万 GPU を搭載し、米国の安全保障、科学、エネルギー アプリケーション全体で技術的リーダーシップを推進するという DOE の使命を加速。また、Equinox システムには、Blackwell 1万 GPU を搭載。
合計 **2,200 EFLOPS**の AI 性能。
- Oracleは高性能ソブリン AI 機能を提供する OCI テクノロジのソフトウェアスタックと運用体制さらにはオンデマンドで利用出来る Hopper GPU と Blackwell GPU を提供。



https://prtimes.jp/main/html/rd/p/000000570.000012662.html?fbclid=IwY2xjawNxEtRleHRuA2FibQIxMABicmkETfhYVpPVkxhFYWxhVzdYNFF4AR6NygAOaXrFbjSpl42QF42FkaT_rW2mHBFeTritJOxRZVWEvF3xTPZEYslD0w_aem_2nYdtbo4UkpcuSMdODsVFQ

Gfarm IaC 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のスクリプトサンプルを利用しOCI上に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

The screenshot shows a GitHub repository page for 'oss-tsukuba/incus-auto'. The breadcrumb path is 'oss-tsukuba / incus-auto / example / gfarm-terraform-oci'. A commit by 'takuya-isbs' is shown with the message 'example/gfarm-terraform-oci/README.md: update' from last week. Below is a table of files and folders in the repository.

Name	Last commit message	Last commit date
..		
terraform	example/gfarm-terraform-oci: change keywords of variables.tf	2 months ago
.gitignore	example/gfarm-terraform-oci: terraform/SRC -> terraform/ansi...	3 months ago
Makefile	example/gfarm-terraform-oci: terraform/SRC -> terraform/ansi...	3 months ago
README.md	example/gfarm-terraform-oci/README.md: update	last week

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

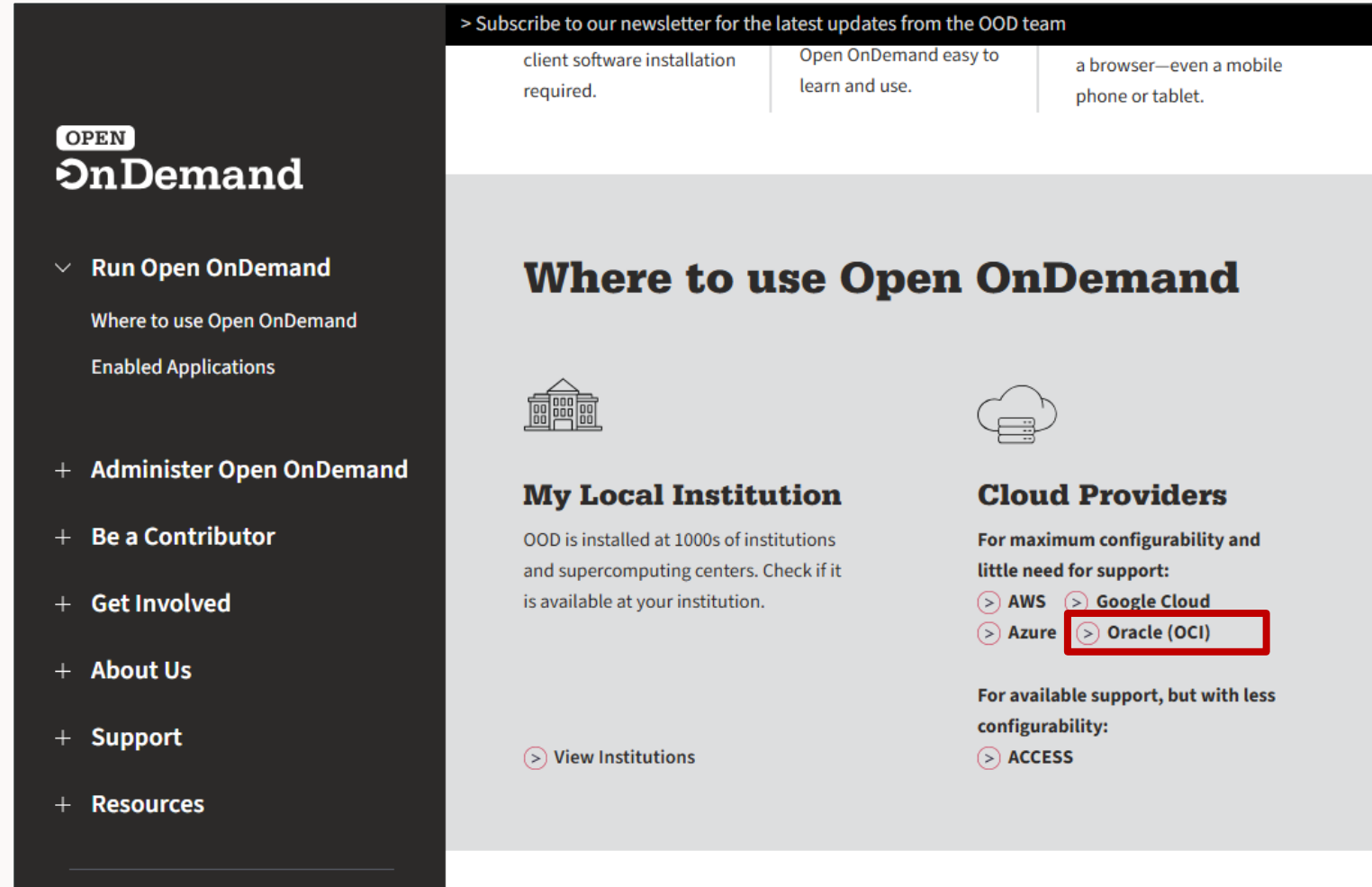


Open OnDemand

- 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Open OnDemand コミュニティWebに利用可能な Cloud Provider として掲載

Open OnDemand インストール済みのマシンイメージも今後提供予定



> Subscribe to our newsletter for the latest updates from the OOD team

client software installation required. | Open OnDemand easy to learn and use. | a browser—even a mobile phone or tablet.

OPEN OnDemand

- Run Open OnDemand
 - Where to use Open OnDemand
 - Enabled Applications
- Administer Open OnDemand
- Be a Contributor
- Get Involved
- About Us
- Support
- Resources

Where to use Open OnDemand

My Local Institution

OOD is installed at 1000s of institutions and supercomputing centers. Check if it is available at your institution.

> View Institutions

Cloud Providers

For maximum configurability and little need for support:

- > AWS
- > Google Cloud
- > Azure
- > Oracle (OCI)

For available support, but with less configurability:

- > ACCESS





Globus ツールキット

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提供プロトコル

- 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

OCI は Globus Connector partner です。

The screenshot shows the Oracle Cloud Infrastructure website. At the top, there is a navigation bar with links for Solutions, Resources, Pricing, and Newsroom. Below the navigation bar, there is a "GLOBUS PARTNER" badge with a gear icon. The main heading is "ORACLE Cloud Infrastructure". Below the heading, there is a paragraph of text: "Oracle Cloud Infrastructure (OCI) Object Storage provides scalable, durable, low-cost storage for any type of data. Benefit from 11 nines of durability. Scale storage to nearly unlimited capacity for your unstructured data." Below this text is a link: "VISIT ORACLE.COM FOR MORE INFORMATION" with a right-pointing arrow icon. At the bottom, there is a "Tags" section with a "premium connectors" tag.



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
- ご利用機関様のSINET接続用ルータでBGPおよびIEEE802.1q(tagVLAN)の設定が必要
- 定額の仮想回線サービスであるFastConnectで接続
- 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**転送データ量課金無制限に¥0-**
- 複数のテナントで**SINET接続仮想回線の共用することも可能**
- [必要なお手続き]
 - ご利用開始2週間前までに、NII様へSINETクラウド接続サービスご利用申請
 - ご利用開始1週間前までに、オラクルへ仮想回線サービス FastConnectのOCIDをご連絡



ORACLE

IO500によるOCI File Storage Serviceの 性能比較

日本オラクル株式会社



OCI File Storage Service の IO500測定

本資料ではOCIの共有ファイルストレージとして以下の3種類のサービスのIO500の結果を比較する。

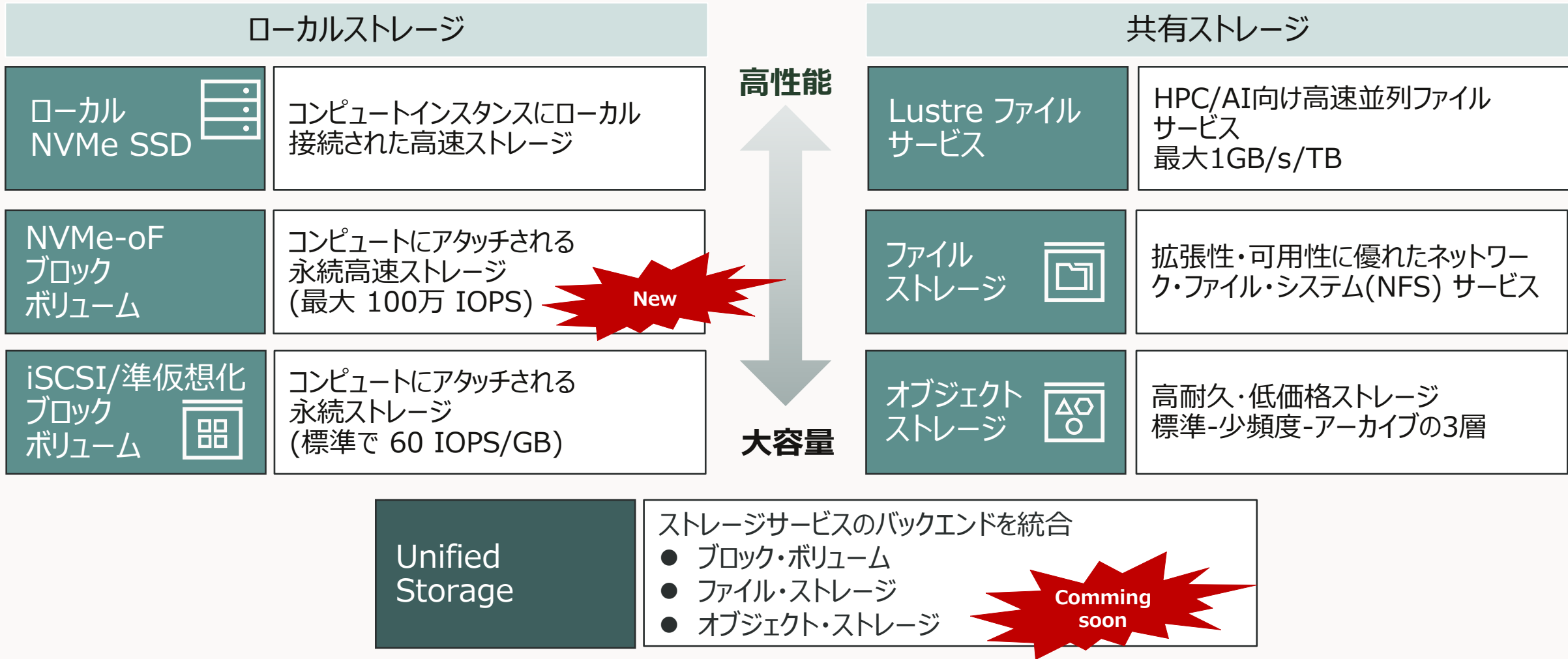
- File Storage Service (nfs v.3)
 - High Performance Mount Target 20Gbps (2.5GB/s)
 - High Performance Mount Target 80Gbps (10GB/s)
- File Storage Service with Lustre
 - 125MB/s/TB、31.2TB (3.9GB/s)

本計測では IO500 version io500-sc24_v1-11-gc00ca177071b (standard) を利用し、各サービスはデフォルトの状態ですべての計測をしています。また利用できるリソースの都合上本結果は最大パフォーマンスを計測するものではありません。



OCI ストレージ・サービスのラインナップ

用途に応じたストレージを選択可能、全サービスデバイスはSSDベースで構築



File Storage Service

フルマネージドで、簡単にデプロイ可能な共有ファイルストレージ

- 数キロバイトから8エクサバイトまで
- ストレージプールをマルチテナントで共有
- 保存データ量課金 (¥46.50 /GB/month)
- 性能要件に応じてマウント・ターゲットのタイプを選択可能

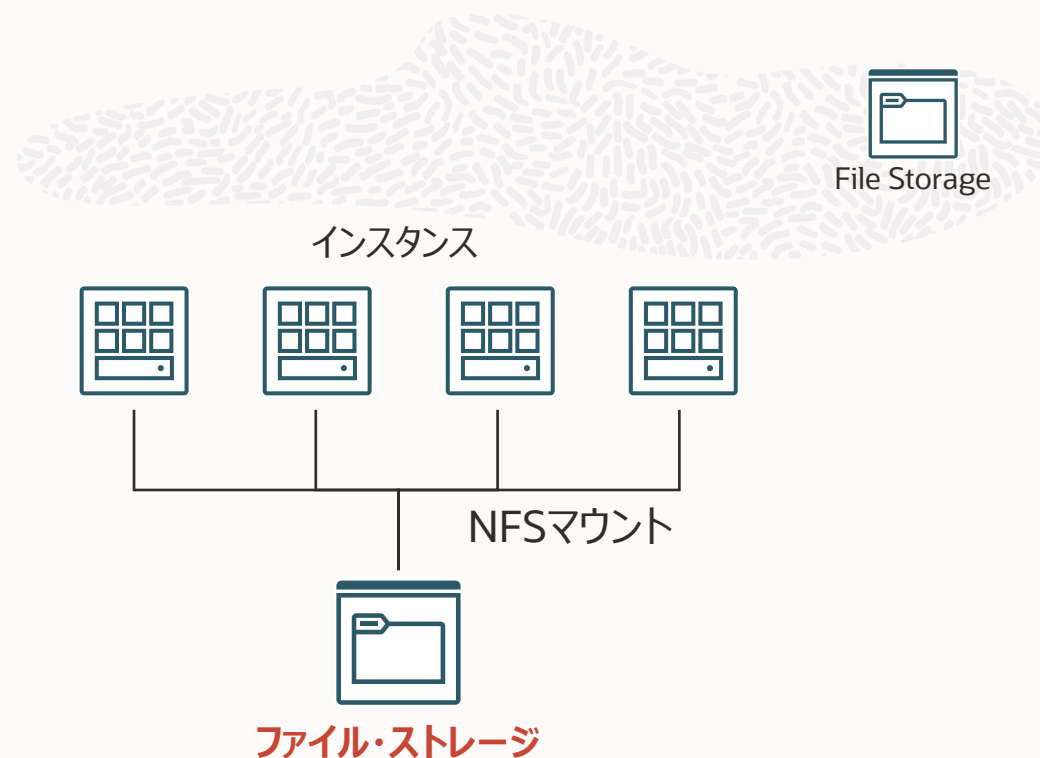
機能

- NFS v3、POSIX準拠
- スナップショット機能、クローン機能
- レプリケーション機能
- デフォルトで暗号化
- 管理コンソール、API/CLIアクセス

高性能マウントターゲットを利用する場合、スループットに応じて最低容量分の30日間の課金コミットメントが必要

例)

- 標準タイプの場合、1000GB利用していれば ¥46.50 x 1,000
- 高性能20Gbpsの場合、20,000GB以内の利用はすべて ¥46.50 x 20,000



マウント・ターゲット・タイプ	読取りスループット	最低容量
標準	1 Gbps (125MB/s)	なし
HPMT-20	20 Gbps (2.5GB/s)	20TB
HPMT-40	40 Gbps (5GB/s)	40TB
HPMT-80	80 Gbps (10GB/s)	80TB



File Storage Service with Lustre

フルマネージド Lustre 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
- 容量 31.2TB~20 PB、確保容量課金 ※ 容量単価は性能層による
※サービス制限初期値: ADごと総容量200TB、総帯域25GB/s
- 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

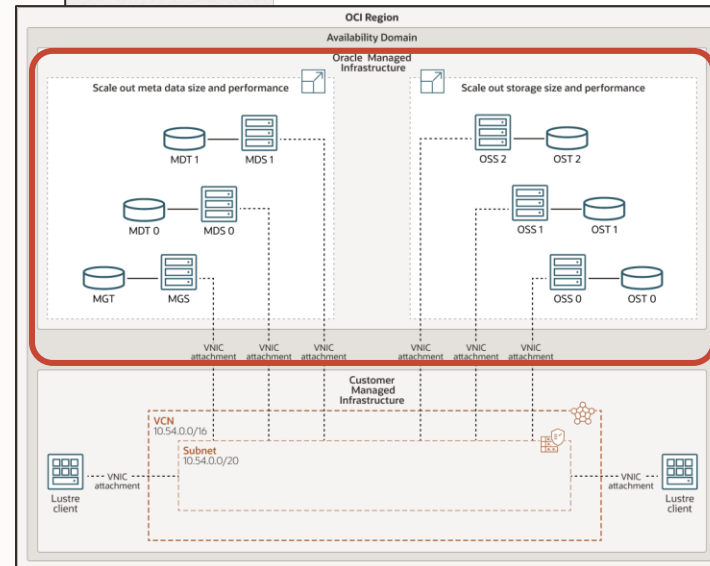
125 MB/s/TB (1Gbps/TB)
250 MB/s/TB (2Gbps/TB)
500 MB/s/TB (4Gbps/TB)
1 GB/s/TB (8Gbps/TB)

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

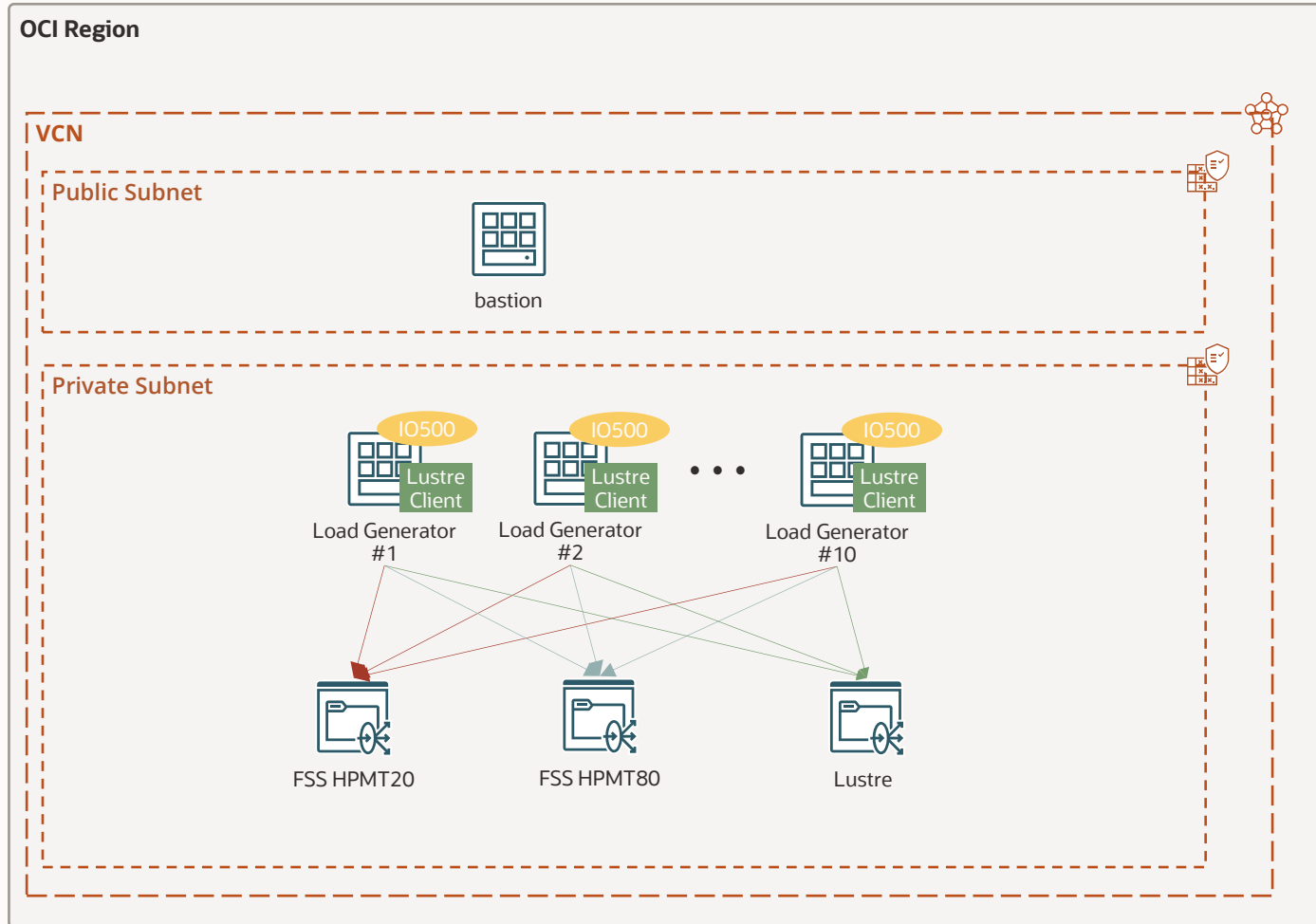


各サービスの月額最低コスト

File Storage ...	Estimated Monthly Cost ¥930,000.00	FSS HPMT 20Gbps
Storage - File Storage	ESTIMATED MONTHLY COST ¥930,000.00	
Performance level: 20 Gbps	Storage capacity [GB]: 20,000	Max IOPS: 100000 Max Throughput (Gbps): 20
File Storage ...	Estimated Monthly Cost ¥3,720,000.00	FSS HPMT 80Gbps
Storage - File Storage	ESTIMATED MONTHLY COST ¥3,720,000.00	
Performance level: 80 Gbps	Storage capacity [GB]: 80,000	Max IOPS: 400000 Max Throughput (Gbps): 80
File Storage with Lustre ...	Estimated Monthly Cost ¥920,700.00	マネージド Lustre
File Storage with Lustre Service	ESTIMATED MONTHLY COST ¥920,700.00	
Performance level: 125 MB/s/TB	Storage capacity [GB]: 31,200	



ベンチマーク環境構成図



Load Generator スペック

- VM.Standard.E5.Flex (仮想マシン)
 - CPU: AMD Genoa (EPYC 9J14)
10物理コア
 - Memory: 32GB
 - OS: Oracle Linux 8.10
 - NIC 10Gbps
- 上記構成を10インスタンス



File Storage with Lustre: 125MB/s/TB の31.2TBデプロイ時のLustre構成

```
ubuntu@kazuito-client01:~$ lfs df -h
```

UUID	bytes	Used	Available	Use%	Mounted on
lustrefs-MDT0000_UUID	563.4G	976.6M	512.4G	1%	/lustre[MDT:0]
lustrefs-MDT0001_UUID	563.4G	2.1G	511.3G	1%	/lustre[MDT:1]
lustrefs-OST0000_UUID	2.4T	1.5T	788.2G	67%	/lustre[OST:0]
lustrefs-OST0001_UUID	2.4T	1.5T	797.2G	67%	/lustre[OST:1]
lustrefs-OST0002_UUID	2.4T	1.5T	798.3G	67%	/lustre[OST:2]
lustrefs-OST0003_UUID	2.4T	1.5T	799.8G	67%	/lustre[OST:3]
lustrefs-OST0004_UUID	2.4T	1.5T	798.1G	67%	/lustre[OST:4]
lustrefs-OST0005_UUID	2.4T	1.5T	802.3G	67%	/lustre[OST:5]
lustrefs-OST0006_UUID	2.4T	1.5T	799.3G	67%	/lustre[OST:6]
lustrefs-OST0007_UUID	2.4T	1.5T	794.8G	67%	/lustre[OST:7]
lustrefs-OST0008_UUID	2.4T	1.5T	800.8G	67%	/lustre[OST:8]
lustrefs-OST0009_UUID	2.4T	1.5T	792.6G	67%	/lustre[OST:9]
lustrefs-OST000a_UUID	2.4T	1.5T	799.3G	67%	/lustre[OST:10]
lustrefs-OST000b_UUID	2.4T	1.5T	807.3G	66%	/lustre[OST:11]
filesystem_summary:	29.4T	18.4T	9.4T	67%	/lustre

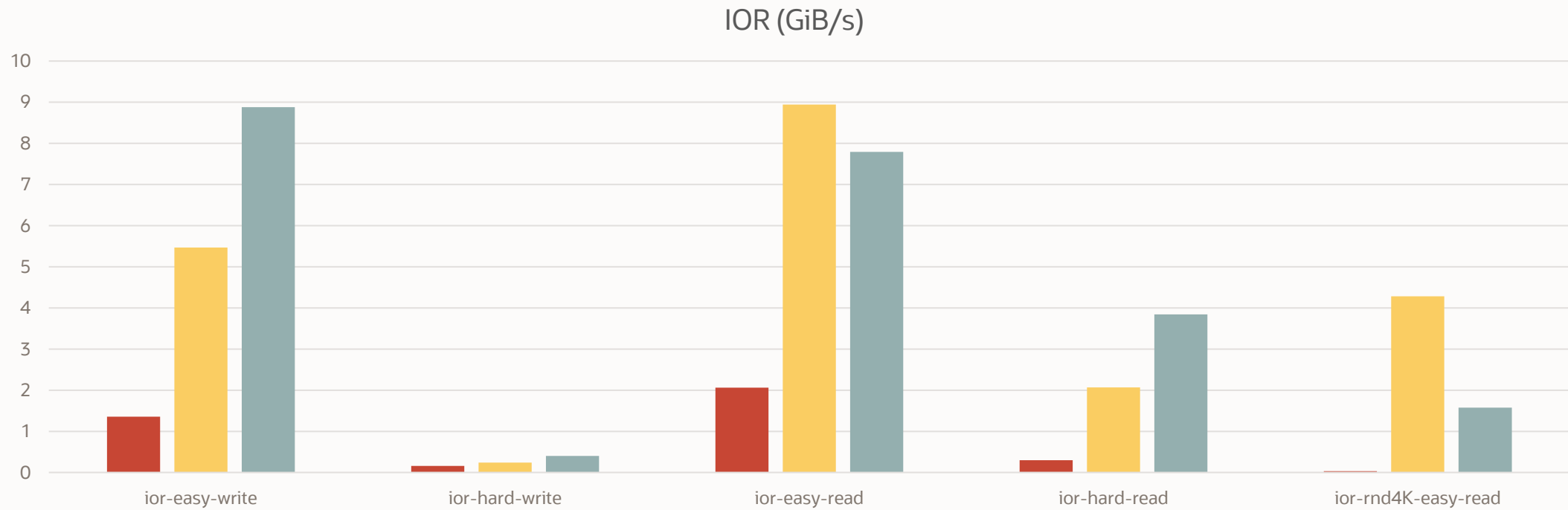
- MDTは2つ
- OSTは2.4TB x 12個で構成



結果比較

IOR

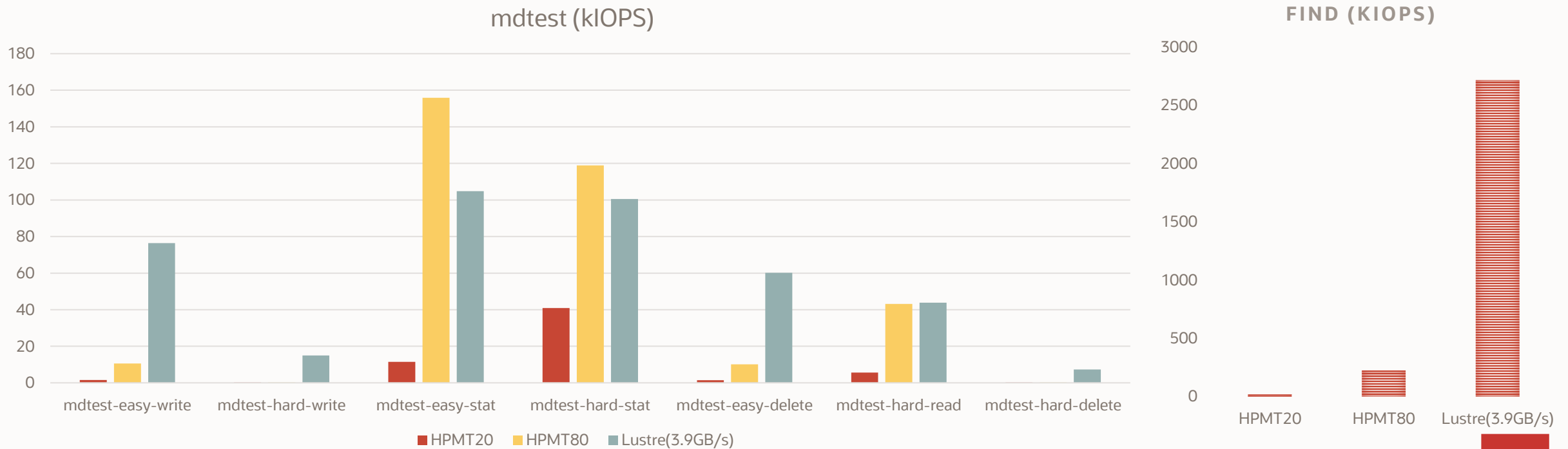
(GiB/s)	HPMT20 (2.5GB/s)	HPMT80 (10GB/s)	Lustre (3.9GB/s)
ior-easy-write	1.354433	5.466822	8.876323
ior-hard-write	0.158835	0.23891	0.400416
ior-easy-read	2.062217	8.942844	7.791247
ior-hard-read	0.300082	2.065734	3.840034
ior-rnd4K-easy-read	0.032853	4.280853	1.577535



結果比較

mdtest、find

(kIOPS)	HPMT20 (2.5GB/s)	HPMT80 (10GB/s)	Lustre (3.9GB/s)
mdtest-easy-write	1.548055	10.532364	76.442342
mdtest-hard-write	0.198371	0.179323	14.917965
mdtest-easy-stat	11.523253	155.873526	104.829492
mdtest-hard-stat	40.898082	118.853495	100.478115
mdtest-easy-delete	1.469713	10.149441	60.178641
mdtest-hard-read	5.629901	43.200073	43.786014
mdtest-hard-delete	0.190594	0.165603	7.334138
(kIOPS)	HPMT20	HPMT80	Lustre(3.9GB/s)
find	16.777146	222.794123	2713.621202



結果比較

Summary

Summary	FSS HPMT20 (2.5GB/s)	FSS HPMT80 (10GB/s)	FSS Lustre (3.9GB/s)
Bandwidth (GiB/s)	0.604046	2.216307	3.211233
IOPS (kIOPS)	2.804709	12.419541	70.779799
TOTAL Score	1.301604	5.246477	15.076153
# of clients	10	10	10
# of proc/client	10	8	8
Total # of procs	100	80	80
IO500 SC25 - Full #	ランク外	301	266
" - Production #	ランク外	ランク外	42
" - 10Node Prod. #	ランク外	ランク外	30



【お問い合わせ先】

日本オラクル株式会社
クラウド事業統括 公共営業本部
松山 慎 (まつやま まこと)
E-Mail: Makoto.Matsuyama@oracle.com
電話: 080-1289-8315

お気軽にご連絡ください!



ORACLE

