DELL TECHNOLOGIES PARTNER PROGRAM
HEROES

HEROES FORUM 2023 Dec

PowerScaleはAI/GenAI向けのストレージに使えるのか！？

デル・テクノロジーズ株式会社
UDS事業本部
矢ヶ部 謙一

DELL Technologies
PARTNER PROGRAM

 Copyright © Dell Inc. All Rights Reserved.

DELL TECHNOLOGIES PARTNER PROGRAM
HEROES

1. AIワークロードとPowerScale

2. AI関連のアップデート

DELL Technologies
PARTNER PROGRAM

 Copyright © Dell Inc. All Rights Reserved.

AIワークロードとPowerScale

 Copyright © Dell Inc. All Rights Reserved.

AI用データストレージに必要な要件

高性能

エンジンに対して高速大量にデータを送り続ける仕組み

接続性

接続形状の異なるエンジンへの接続

シンプル

簡単にメンテナンスが出来て壊れない仕組み

拡張性

初期段階は小さく徐々に大きくできる仕組み

実績

エンジンとの利用実績

経済性

無駄を少なくデータを貯えることができる仕組み

運用/継続性

データを永続に供給できる仕組み

柔軟性

将来別の用途（機器）に対して供給できる仕組み

Copyright © Dell Inc. All Rights Reserved.

AI用データストレージに必要な要件

高性能
エンジンに対して高速大量にデータを送り続ける仕組み

高スループット

接続性
接続形状の異なるエンジンへの接続

マルチプロトコル

シンプル
簡単にメンテナンスが出来て壊れない仕組み

簡単な運用
ノードレベルでの保護

拡張性
初期段階は小さく徐々に大きくできる仕組み

スケールアウト

実績
エンジンとの利用実績

多くの採用実績
Nvidia認定

経済性
無駄を少なくデータを貯えることができる仕組み

自動階層化
インライン重複排除/圧縮

運用/継続性
データを永続に供給できる仕組み

新旧モデルの混在
データ移行の簡素化

柔軟性
将来別の用途（機器）に対して供給できる仕組み

マルチクラウド対応

PowerScaleはすべて対応！

DELL Technologies PARTNER PROGRAM

Copyright © Dell Inc. All Rights Reserved.

AIワークロード：「トレーニング」時の推定スループット

MLPerf 3.0から具体的な性能、傾向を探る

名称	説明	パラメータ数 モデルサイズ ミリオン/ビリオン	データの種類	トレーニング データ容量	トレーニング サンプル推定サイズ （バイト）	GPUあたりに必要な スループット
Resnet-50	画像分類	23M	画像	~155GB	116KB	5.6K画像/s 634 MB/s
BERT-Large	トランスフォーマーベースの言語モデル	345M	テキスト	~86GB	2KB	1007シーケンス/s 1.9 MB/s
DLRM-DCNv2	クリックスルー推奨モデル	24B	テキスト - クリックストリーム 過去ログ	3.5TB	912B	1.2Mサンプル/s 1065 MB/s
3D U-Net	医療画像のための 3D画像分割	19M	3D映像	300枚の3D画像、約27GB	92MB	57枚/s 5293 MB/s
MaskRCNN	物体検出	44M	画像	121K画像、約18GB	160KB	144画像/s 22.5 MB/s
GPT3	事前に学習された トランスフォーマーベースの言語モデル	175B	テキスト	約305GB	-	582 シーケンス/s 4.5 MB/s

複数のベンダーの様々なシステムサイズのH100 80GB GPUを使用したMlperf v3.0の結果に基づく。デル調べ。

- テスト内容(AIモデル)ごとにデータアクセスのスループット傾向は全く異なる
- GPUバウントのテストのため、GPU数が増えるほど処理性能が向上する
- モデルサイズが少なくトレーニングデータのサンプルサイズが大きいほど、データアクセスのスループットが大きくなる

Copyright © Dell Inc. All Rights Reserved.

AIワークロード：「推論」、「LLMファインチューニング」、「チェックポイント」の傾向

■ 推論

- 推論スループット要件は、テスト内容、モデルサイズ、サンプルサイズによって大きく異なる
- トレーニングと違い重みを更新しないためGPU負荷が小さくなり、トレーニング時より必要なスループットは大きくなる
- Resnet-50は、モデルサイズが小さく（計算量が少なく）、推定サンプルサイズが大きいため、GPUあたりの推論データ帯域幅が非常に大きい。

■ LLMファインチューニング

- パラメータ数が多くなるほど、データアクセスのスループットが小さい

■ チェックポイント

- GPUメモリーにあるモデルの現在の状態を保存し、後でトレーニングを再開できるようにする処理 ＝ 外部ストレージが必要
- 作成時には、ストレージにはWriteの負荷、復元する際にはReadの負荷が発生
- チェックポイントの処理中はトレーニングできない ＝ 短時間で完了するのが望ましい
- モデルサイズが多いほどサイズは大きくなる

DELL Technologies PARTNER PROGRAM

 Copyright © Dell Inc. All Rights Reserved.

AIワークロードの種類とストレージ性能へ影響する要素

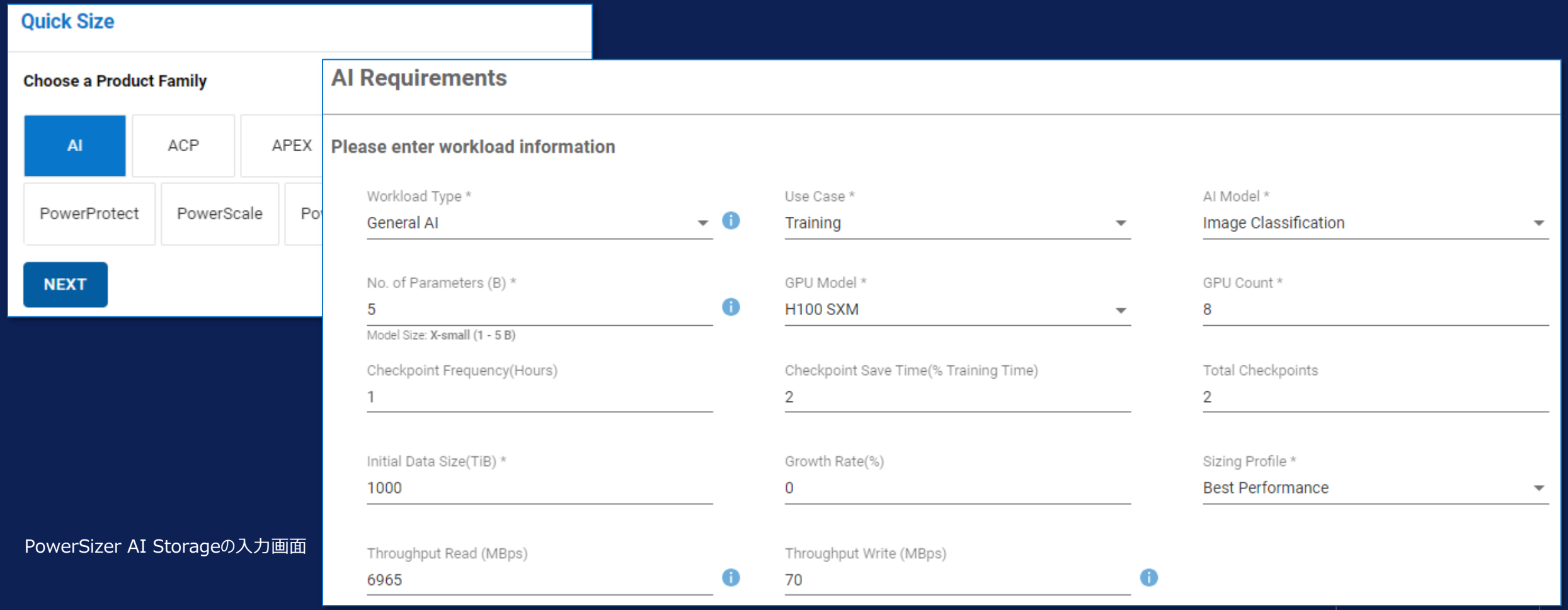
ワークロードの種類	処理内容	AIモデル
AI	• トレーニング • 推論	• 画像分類 • 物体検出 • 医療画像のセグメンテーション • 音声ーテキスト変換 • 自然言語処理 • レコメンデーション etc.
Generative AI	• トレーニング • 推論 • 微調整(ファインチューニング)	• LLM • 画像生成 • 音声生成 etc.

要素	内容
パラメータ数	• パラメータ数
GPU	• GPU数 • GPUの種類
チェックポイント	• 取得頻度 • 1回の保存時間 • 取得総数
データ容量	• 初期容量 • 増加率

Copyright © Dell Inc. All Rights Reserved.

AI PowerSizer

- 「Compute」と「Storage」をそれぞれサイジング
- 「Compute」はGPUサーバのサイジング、「Storage」は、必要なストレージをPowerScaleで算出



PowerSizer AI Storageの入力画面

 Copyright © Dell Inc. All Rights Reserved.

AI PowerSizer

OVERVIEW | POWERPOINT

Cluster

Node Pool

PowerScale Cluster

BEST PERFORMANCE

Node Pools
1

Backend Networking Type
Ethernet

Usable Capacity at 100%
1108 TiB

Usable Capacity at 85%
941.8 TiB

PERFORMANCE

Max Throughput Read
52.22 GiB/s

Max Throughput Write
12.39 GiB/s

Workload / Use Profile
Database

Benchmark
Spec SFS 2014 (NFS3)
This benchmark models a standard file environment with high file protocol operations and small, mostly random, reads and writes.

CAPACITY & DISKS

NodePool

6 x F600	**8 x 30.72 TB**	QLC-SSD

POWER DETAILS

Typical power consumption @ 240v:	2,802 watts
Max power consumption @ 240v:	4,308 watts
Typical thermal rating:	9,561 BTU
Max thermal rating:	14,700 BTU

ADDITIONAL DETAILS

SIZER ID

Storage AI Sizing Inputs

Input	Value
Solution Type	Storage
Workload Type	General AI
Use Case	Training
AI Model	Image Classification
No. of parameters (B)	5
Sizing Profile	Best Performance
GPU Model	H100 SXM
GPU Count	8
Initial Data Size(TiB)	1000
Growth Rate(%)	-
Checkpoint Frequency(Hours)	1
Checkpoint Save Time(% Training Time)	2
Total Checkpoints	2
Throughput Read(MBps)	6965
Throughput Write(MBps)	70

Copyright © Dell Inc. All Rights Reserved.

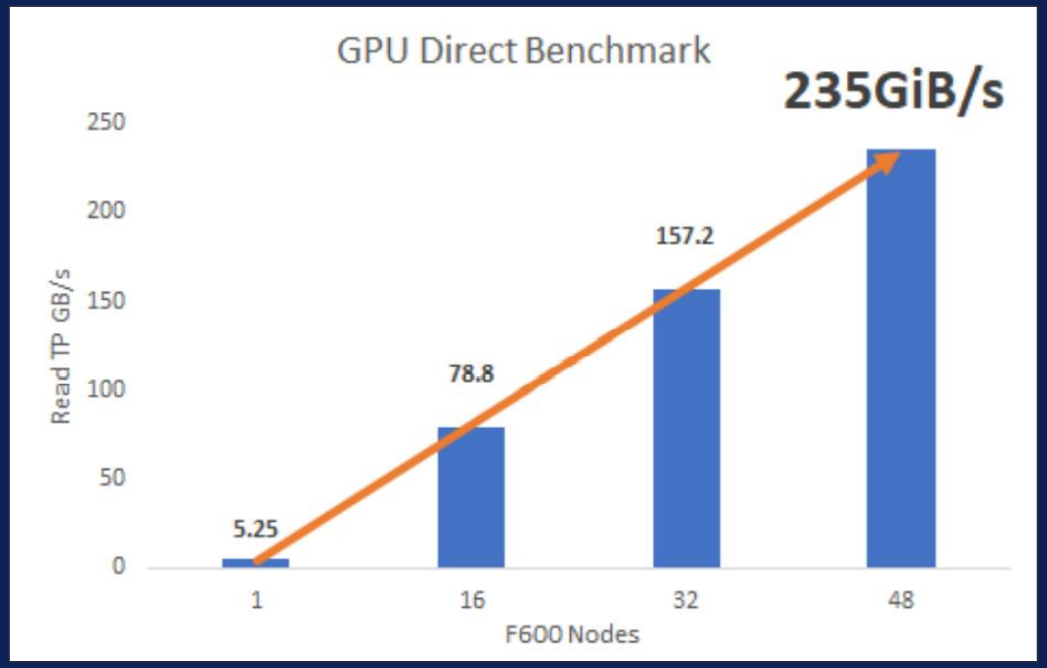
DELL Technologies PARTNER PROGRAM

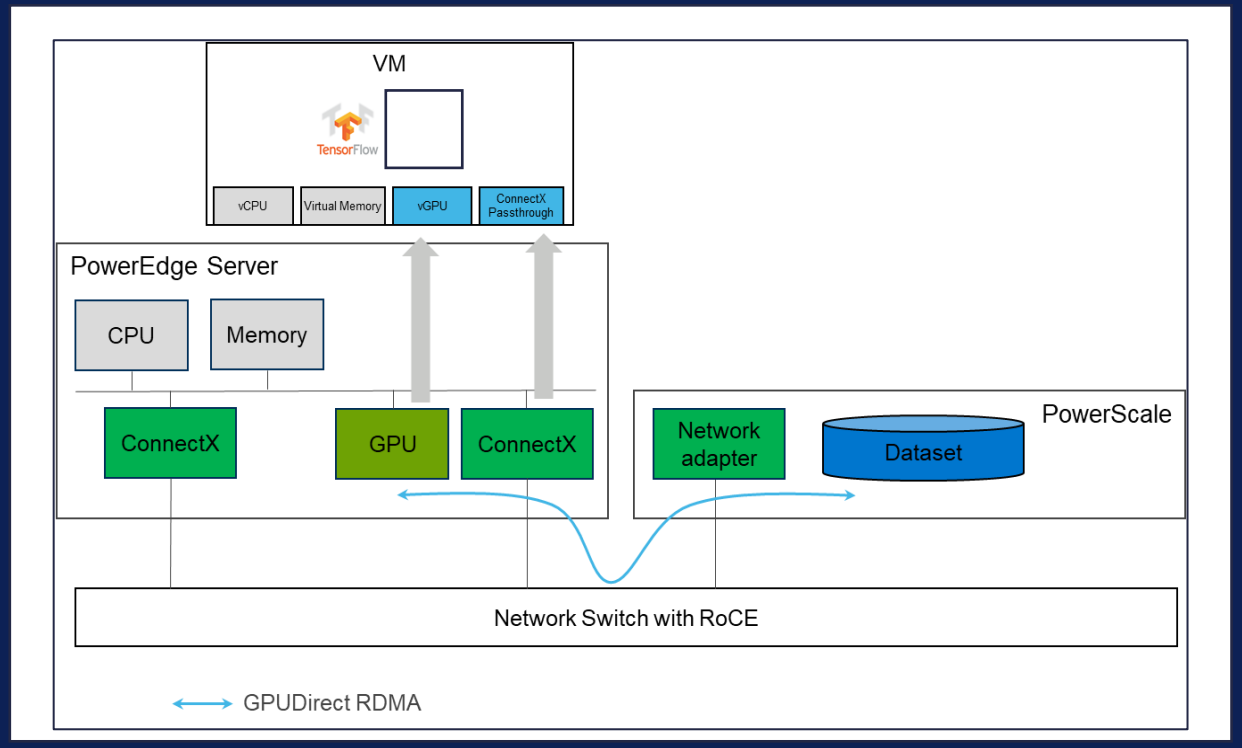
Dell PowerScaleで対応できる

GPUDirect Storage & NFS over RDMA with Nvidia GPUs and Dell PowerScale storage
ストレージシステムのメモリーとアクセスクライアントのメモリーの間で直接データをやり取りできる

■ ストレージ要件

- ペタバイトクラスの膨大な量のストレージ容量が必要
- ソース データセットの高速な読み取り
- パラメータ チェックポイントをできるだけ速く書き出す/読み込む





■ Dell PowerScale

- 高性能フラッシュ ストレージを使用して AI ワークロード向けに最適化
- ハイブリッド SSD/HDD システムも提供しており、SSD と HDD 間でデータを透過的に移動
- NFS over RDMAでアクセスが可能
- 252GB/秒ものリード性能
- CSI Driverへの対応

Source: Dell EMC PowerScale and NVIDIA GPUDirect Performance Report (delltechnologies.com)

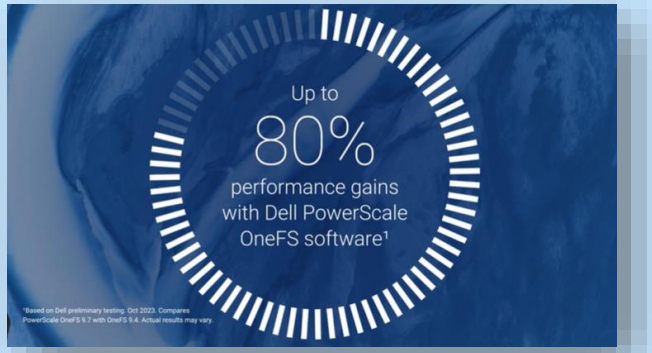
Copyright © Dell Inc. All Rights Reserved.

AI関連のアップデート

 Copyright © Dell Inc. All Rights Reserved.

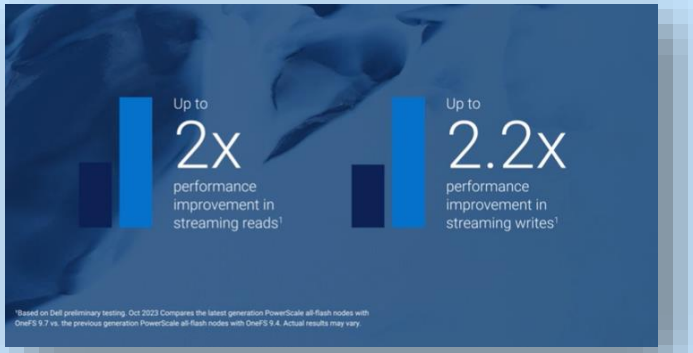
AIのためのDell PowerScaleの進化

NEW！OneFSのパフォーマンス向上



- ストレージソフトウェアの最適化により、**最大80%のパフォーマンス向上を**実現。
- セキュリティを強化し、AIを含む負荷の高いワークロードに対してより高速なパフォーマンスを実現
- APEX File Storage for AWSが強化され、クラスタ容量の拡大、地理的な可用性の拡大、OneFS機能の追加が行われました。
- **OneFS 9.7をリリース済み**

NEW！次世代オールフラッシュ・ノード



- **ストリーミングReadのパフォーマンスが最大2倍向上**し、モデルのトレーニングや微調整のためにGPUへのデータ供給が加速。
- AIモデルのチェックポイントを最適化し、GPUのダウンタイムを抑えます。
- ラックユニットあたりの密度が向上し、TCOが向上します（TBあたりの価格が下がる）。
- **発表予定： 2024年第1四半期**

NEW! NVIDIA DGX SuperPOD認定

- Dell PowerScaleは、**NVIDIA DGX SuperPODで検証された最初のイーサネット・ストレージ・ソリューションとなる見込み**です。
- AIストレージの柔軟性、スピードの向上
- PowerScaleご利用中のお客様がNVIDIA DGX SuperPODソリューションを活用できるようにします。
- **発表予定： 検証完了は2024年第1四半期の予定**

Copyright © Dell Inc. All Rights Reserved.

AIのためのDell PowerScaleの進化

NEW！Cloud対応の拡大

[AWS]

- サポートするインスタンスタイプを追加しました。これにより、日本では大阪リージョンでもご利用可能になりました。
- 最大物理容量を1.6PiBに拡大しました。
- ご利用できなかったいくつかの機能をサポートしました。QoS、SmartLock、Large File Support、Hadoopなどの機能が利用可能になりました。

[Azure]

- APEX File Storage for Azureを間もなくリリース予定です。
  発表予定：2024年前半

NEW！ Ecosystem

- Databricks社とのコラボレーションを促進します。Databricks MosaicMLのライブラリーを使い、「APEX File Storage」に格納している独自のデータで基礎モデルを再トレーニングして、マルチクラウド環境全体に優れた柔軟性を提供することが可能になります。
- **発表予定：AWSは提供中、Azureは2024年前半**

NEW! モダン データ レイクハウス

- Starburstと提携し、オープンかつモダンなデータレイクハウス・ソリューションを提供します。
- Starburst社のアナリティクス ソフトウェアと、「PowerEdge」サーバーおよび「PowerScale」、「Dell ECSエンタープライズ向けオブジェクト ストレージ」、「Dell ObjectScale」のファイル／オブジェクト ストレージ システムの統合によって、データが置かれている場所に関係なく、インサイトを引き出すことが可能になります。
- **<u>発表予定： 2024年前半</u>**

Copyright © Dell Inc. All Rights Reserved.

DELL TECHNOLOGIES PARTNER PROGRAM

HEROES