

クラウド AI アーキテクチャにおける ブートドライブの要件

エンタープライズ・ブートドライブ・ソリューション

ブートドライブは、AI クラウド環境におけるミッションクリティカルなコンポーネントです。サーバーがどれだけ迅速かつ確実にオンラインになり、クラスターに再参加し、ワークロードの処理を開始できるかは、ブートドライブによって決まります。

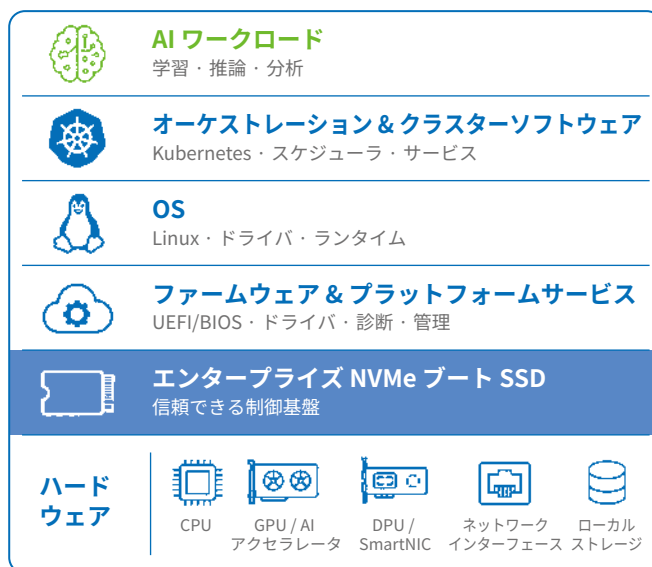
ブートドライブは、見過ごされがちな基盤ストレージレイヤーですが、システムの初期化、トラステッドブート、運用効率において重要な役割を果たすことで、AI インフラストラクチャのパフォーマンスに不可欠な存在となっています。

クラウド・サービス・プロバイダーがヘテロジニアス (異機種混在) の GPU、CPU、およびアクセラレータインフラストラクチャの規模を拡大する中で、ブートストレージには、ボトルネックとなることなく、低レイテンシ、高 IOPS、高い耐久性、および信頼性の高いセキュリティを実現することが求められます。

Silicon Motion のエンタープライズ PCIe NVMe ブート SSD ソリューションは、AI ブートドライブに最適な選択肢です。当社のコントローラおよびファームウェア技術と組み合わせることで、起動時の一貫した動作、電源喪失への耐性、フリート全体の管理性を実現します。

CSP AI アーキテクチャにおけるブートドライブの重要な役割

CSP AI アーキテクチャにおける ブートドライブの重要な役割



ノード全体を支える、
常時稼働で安全かつ
信頼性の高い基盤

図 1: ブートドライブは、AI クラウドノードを初期化し、オーケストレーションサービスやストレージサービスに接続するとともに、ワークロードデータをルートボリュームから分離する、信頼できる制御基盤を提供します。

クラウド AI 環境において、ブートドライブはストレージインフラストラクチャを構成する重要な要素であり、コンピューティングインスタンスの起動と運用を可能にします。

ブートドライブには、ハードウェアを初期化し、AI ワークロード用のソフトウェアスタックを起動するために必要な OS、ブートローダー、コア・プラットフォーム・エージェントが格納されています。また、ネットワーク設定、ID 認証情報、ストレージ・マウント・ポイント、オーケストレーションパラメータなど、インスタンスがクラウドサービスに接続し、必要なコンピューティングリソースとデータリソースにアクセスするための重要な構成データも格納されています。

ブートドライブがサポートするその他の機能として、オーケストレーション、ネットワーク、ストレージスタックの起動があり、学習や推論に必要な通信レイヤーとリソース管理レイヤーを確立します。

クラウド AI 環境では、インスタンスのスピンアップとスピンドアウンによって迅速にスケーリングされることが多いため、ブートドライブの信頼性とパフォーマンスが特に重要です。より高速で信頼性の高いブートドライブは、起動時のレイテンシを短縮し、AI サービスを迅速かつ一貫して稼働可能な状態にするのに役立ちます。

ブートドライブは、一貫性があり再現可能なシステム状態を確立します。標準化により、クラウド AI プロバイダーは、手作業による介入なしに、ヘテロジニアスハードウェアを動作確認済みの構成へ自動的に初期化できます。この抽象化は、運用の一貫性を維持しながら、多様な物理ホストにまたがるフリートをスケーリングする上で不可欠です。

また、ブートドライブをデータストレージから分離することで、アーキテクチャの最適化も可能になります。ルートボリュームは、OS やプラットフォームサービス向けに最適な容量となるように設計され、チューニングされます。一方、データディスクやオブジェクトストレージは、容量、耐久性、ワークロード固有の I/O パターンを考慮して設計されます。アプリケーションデータによってルートボリュームがいっぱいになったり破損したりすると、OS に障害が発生し、ダウンタイムにつながる可能性があります。アプリケーションデータを別のボリュームに保持することで、リソースの競合を防ぎ、プラットフォームサービスの完全性が保護できます。

クラウド AI アーキテクチャにおける AI の普及拡大に伴い、より大きな容量、高いスループット、低レイテンシ、より高い耐久性、および優れた電力効率が求められています。

AI 需要が CSP ブートドライブへの要求をさらに高めている

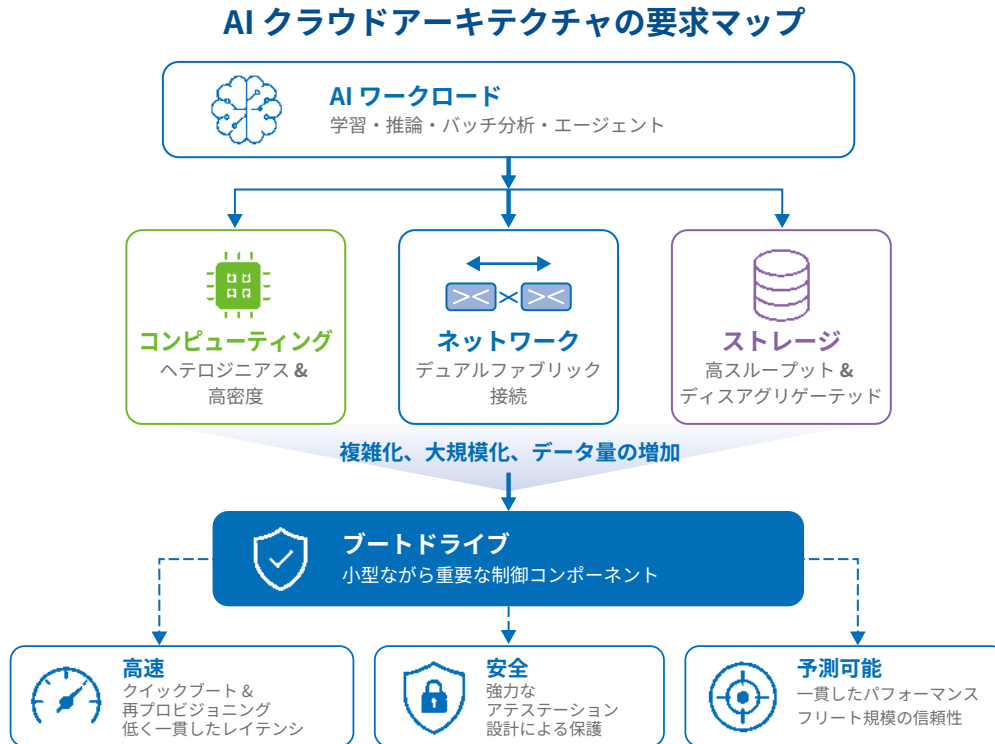


図 2: AI ワークロードにより、クラウドインフラストラクチャはヘテロジニアスコンピューティング、デュアルファブリックネットワーク、高スループットストレージへと移行しつつあり、フリート規模の運用において、高速で安全かつ予測可能なブートドライブの重要性が高まっています。

クラウド・サービス・プロバイダーは、コンピューティング、ネットワーク、ストレージ全体にわたる AI 需要に対応するため、データセンターのアーキテクチャを再構築しています。現在の高密度アクセラレータは、より広いメモリ帯域幅、チップレットまたはマルチチップモジュール、およびハードウェアとソフトウェアのより緊密なコデザイン (協調設計) を中心に構築されており、液冷による高電力環境で効率的に動作するとともに、単一サーバー内だけでなくラック間でもスケーリングできるようになっています。

AI ワークロードでは、より急激なバーストや大きなスキューも発生し、需要が GPU リッチなリージョンやクラスターに集中することも少なくありません。そのため、特に学習ジョブ、推論スパイク、バッチ分析の負荷が急激に増減する場合には、需要予測や負荷分散がより困難になります。これに対処するために、クラウドプロバイダーは、学習や推論のさまざまな要件に対応できるよう、CPU に GPU、DPU、FPGA を組み合わせたヘテロジニアスノードを採用しています。

これらの変化に対応するには、GPU 間トラフィック向けの高帯域幅 East-West パスと、一般的なトラフィック向けの North-South パスで構成されるデュアルファブリックネットワークが必要です。また、アクセラレータを継続的に稼働させるための高スループットストレージや、各ジョブを適切なリソースに割り当てることができるスケジューラも必要です。同時に、数百メートルから数百キロメートルにわたって低レイテンシ性能を実現するスケールアクロス接続により、地理的に分散したデータセンターが単一の AI ファクトリーとして扱われるようになります。

統合されたスケールアウト / スケールアクロス設計では、ハードウェアスタックやソフトウェアスタックの共通化が進んでおり、データセンター内およびデータセンター間のトラフィック全体にわたり、ネットワークリソー

スを柔軟にプールできるようになっています。これにより、エッジクラウド・ノード間に広がるエージェンティック AI ワークロードを含む分散推論や、マルチサイト GPU クラスタがサポートされます。

また、ストレージも AI に合わせて再設計されています。プロバイダーは、NVMe フラッシュ SSD を採用するとともに、ストレージをコンピューティングから分離し、データを GPU メモリに直接移動することで、スループットを最大化し、遅延を削減しています。チェックポイント処理は、モデルの大規模化によってテラバイト規模のチェックポイントが作成されること、並列ジョブの増加によって書き込み負荷が高まること、さらにサービス目標の厳格化によってスナップショットと復元のさらなる高速化が求められることから、大きなボトルネックとなっています。

さらに、データライフサイクル管理は、収集、クレンジング、ガバナンス、再利用、モデルへの迅速なフィードバックを中心とする、継続的なクラウドネイティブパイプラインになりつつあります。そのためには、AI システムが最新の信頼できるデータに基づいて動作できるよう、リアルタイムの共有、より強力なトレーサビリティ、より強固なセキュリティが必要です。その結果、ブートドライブは、小型化、高速化、多くの場合はステートレス化へと向かっています。

クラウド AI がブートドライブに新たな要求をもたらしている

AI インフラストラクチャ全体のブートドライブ

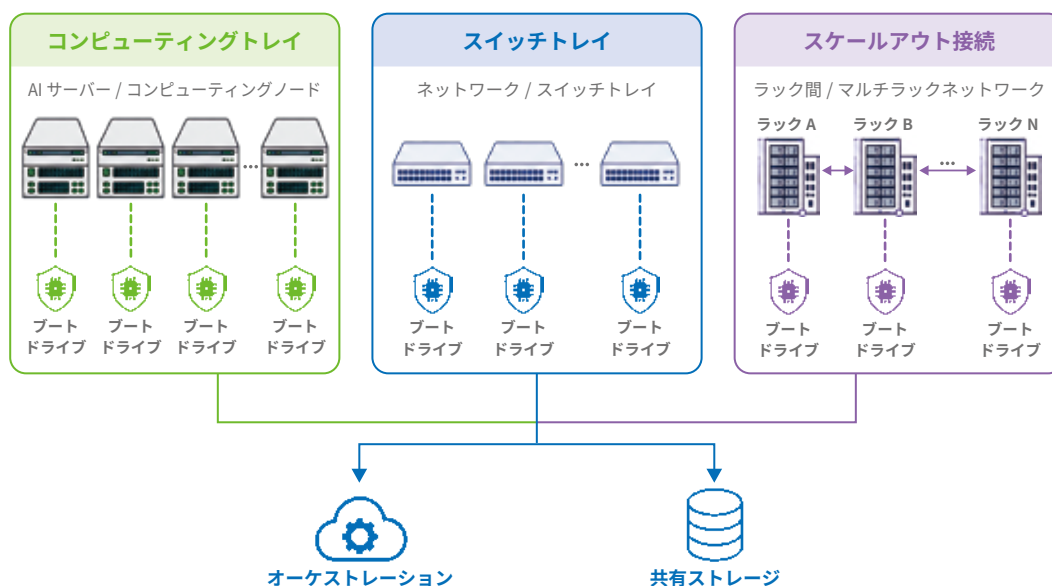


図 3: ブートドライブは、コンピューティングトレイ、スイッチトレイ、スケールアウト接続全体にわたり、デバイスを初期化し、信頼できる構成を確立するとともに、インフラストラクチャをオーケストレーションサービスや共有ストレージサービスに接続するコンパクトな制御の要として機能します。

AI ワークロードを多く扱うクラウド・サービス・プロバイダー環境において、ブートドライブには、各ノードを初期化し、AI ワークロードを実行できるようにする OS と重要なシステムソフトウェアを格納するという、限定的ながらも重要な役割があります。ブートドライブは、データセット、モデルチェックポイント、学習トラフィックに使用される高性能ストレージレイヤーとは分離されています。

実際には、ブートドライブには通常、起動のための最小限の基盤を提供する軽量の OS イメージが格納されています。ノードは、小さなローカルイメージを起動するか、OS をメモリに読み込んだ後、NVMe-over-Fabrics などのネットワークストレージやその他の共有システムに接続して、実際のワークロードを実行できます。

これらの環境は通常、コンピューティングトレイ、スイッチトレイ、スケールアウト接続という3つのシナリオに分類されます。いずれも基本的な要件は共通していますが、GPU サーバー、ファブリックスイッチ、または大規模な分散サーバーフリートのいずれを起動するかによって、運用上の重点は異なります。

コンピューティングトレイでは、専用の NVMe デバイス (多くの場合、M.2 または E1.S フォームファクタ) が OS イメージを格納するため、ノードは外部ストレージに依存することなく初期化できます。起動後、トレイはクラスター管理フレームワークに参加し、所定のソフトウェア状態になってから分散型 AI 運用に加わります。

スイッチトレイの場合、ローカル・ブートドライブにスイッチの OS、構成ファイル、および管理ソフトウェアが格納されます。GPU サーバーが低レイテンシで通信するには、ファブリックが確実に起動している必要があるため、これは重要です。ブートドライブは、スイッチの迅速な復旧と円滑なスケールアップにも役立ちます。

スケールアウト接続では、ブートドライブは容量を担うストレージというよりも、制御の要として機能します。ブートドライブは、ファブリックに参加するために必要な OS、コンテナランタイム、ドライバ、およびネットワーク構成を読み込みます。一方で、データやモデルのパラメータの大半は、独立してスケールする共有ストレージシステムに保持されます。

ブートドライブは、フリート規模での管理性とセキュリティも支えています。標準化されたイメージ、セキュアブート、迅速な再プロビジョニング、ローカル認証情報、ファームウェア、および構成ベースラインは、いずれも予測可能なローカルストレージに依存しています。そのため、障害が発生したノードや転用されたノードも、システム全体に影響を与えることなくリセットして再展開できます。

つまり、クラウドプロバイダーは、ブートドライブを容量ではなく、IOPS、レイテンシ、信頼性、および管理性の観点から最適化し、クラスターのセキュリティ基盤となる信頼の起点 (Root of Trust) の一部として扱う必要があります。これらのシナリオは、動的な AI 環境において SSD がブートドライブとして優れた選択肢である理由を示しています。

AI クラウド環境のブートドライブに SSD が最適な理由

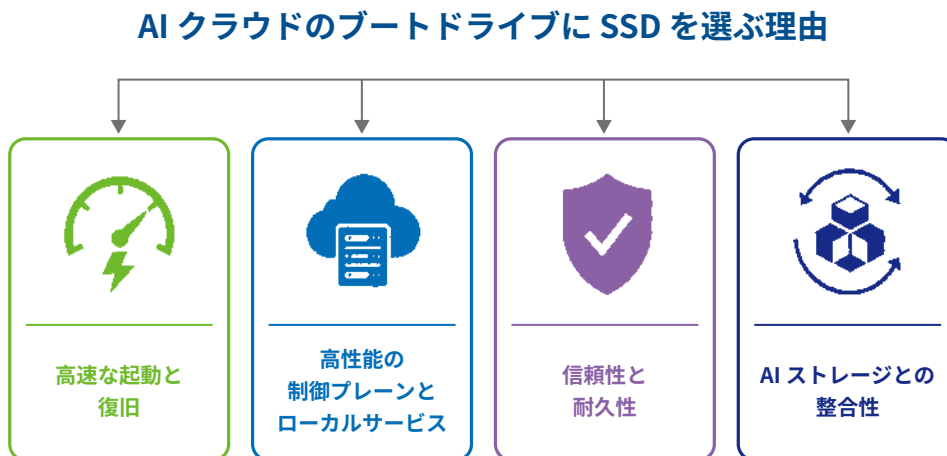


図 4: 機械式 HDD と比較して、エンタープライズ PCIe NVMe SSD は、AI クラウドインフラストラクチャを大規模に起動、復旧、管理するために必要な低レイテンシ、高 IOPS、高い耐久性、および一貫した動作を実現します。

AI インフラストラクチャには、機械式ハードドライブでは実現できない超低レイテンシ、高 IOPS、および持続的なパフォーマンスが求められます。ブートドライブは、ボトルネックになることなく、数千台のサーバーの同時起動、OS の頻繁な再イメージング、迅速な障害復旧をサポートする必要があります。

SSD は、ブートドライブに求められる主要な機能を実現するための重要な特長を備えています。

- **高速な起動と復旧** : ブートドライブは、障害発生後にサーバーを迅速にオンライン状態へ復帰させ、クラスターのダウンタイムを最小限に抑え、GPU の利用率を維持する必要があります。大規模環境でも確実に復旧できる必要もあります。わずかな遅延や再構築の繰り返しでも、運用のボトルネックにつながる可能性があるためです。
- **高性能な制御プレーンとローカルサービス** : AI クラウド環境でオーケストレーション、認証、監視、およびその他のスタートアップサービスを迅速かつ確実に立ち上げるには、ブートドライブに超低レイテンシかつ高 IOPS のストレージが必要です。ローカル SSD に近いブートストレージは、レイテンシが大幅に低いため、起動時のボトルネックを回避するのに役立ちます。
- **信頼性と耐久性** : ブートドライブの障害は、インフラ全体にわたる即時のサービス停止につながる可能性があります。これらのシステムは、頻繁な書き込みや長期にわたる運用ライフサイクルに依存しているため、データ完全性やアップタイムを損なうことなく、持続的なワークロードに耐える必要があります。
- **AI ストレージ割り当てとの整合性** : I/O の不整合は、パフォーマンス低下やレイテンシのばらつきを引き起こします。SSD は、信頼性の高い OS 読み込み、ファームウェア実行、および継続的なテレメトリ運用に不可欠な、一貫したパフォーマンス、安定した電力挙動、強固なデータ完全性を実現します。

SSD は、GPU インフラストラクチャを高速に起動し、迅速に復旧させ、大規模環境でも中断なく運用を継続するために必要な速度、一貫性、レジリエンスを提供します。

Silicon Motion のエンタープライズ PCIe NVMe ブート SSD ソリューションは AI ブートドライブ向けに最適化されています

Silicon Motion のブート SSD テクノロジスタック



図 5: Silicon Motion のエンタープライズ PCIe NVMe ブート SSD ソリューションは、コントローラおよびファームウェア技術に IntelligentSeries™ 機能を組み合わせることで、大規模 AI クラウドインフラストラクチャ向けに、高速性、予測可能性、セキュリティ、および電力障害への耐性を備えたブートストレージを提供します。

AI クラウド環境では、ブートストレージはもはや重要性の低いコンポーネントではありません。クラスター全体のアップタイム、稼働率、および運用効率に直接影響を及ぼす重要な管理点です。

Silicon Motion は、効率性、一貫性、セキュリティに加え、大規模運用時の起動リスクと運用コストを低減する独自のコントローラおよびファームウェア技術を組み合わせることで、AI サーバーに求められるブートドライブ要件を満たします。その主な利点は、高速性能だけではありません。何千台もの AI サーバーを迅速に起動し、セキュリティを確保しながら、本番環境での消費電力を抑えて運用できる点にあります。

- **高速で予測可能な起動パフォーマンス** : Silicon Motion のエンタープライズ PCIe NVMe ブート SSD ソリューションは、専用のブートストレージ設計に、一貫したファームウェア動作、安定した電源制御、および強固なデータ完全性を組み合わせることで、信頼性の高い OS の読み込みと初期化を実現します。
- **常時稼働する AI ノードを支える信頼性** : Silicon Motion のエンタープライズ PCIe NVMe ブート SSD ソリューションは、予測可能なブート動作、電源喪失保護、およびデータ完全性を重視しており、ログの頻繁な書き込みや継続的なランタイム処理中でも、OS の読み込みやファームウェアの実行を安定して維持します。
- **電源喪失や過酷な環境への耐性** : Silicon Motion は、ハードウェアとファームウェアのメカニズムによる高度な電源喪失保護機能を備えており、予期しない停電時には転送中のデータを不揮発性フラッシュメモリへ安全に書き込むとともに、マッピングテーブルの整合性を維持します。
- **セキュリティとトラステッドブート** : Silicon Motion のエンタープライズ PCIe NVMe ブート SSD ソリューションは、ハードウェアベースの暗号化に加え、認証機能を備えたファームウェア保護、セキュアブート、および TCG Opal などの規格に対応したコントローラ統合型のセキュリティ機能を備えており、ブートボリュームを改ざんや不正アクセスから保護します。

Silicon Motion 独自の IntelligentSeries™ テクノロジーは、次のようなさまざまなメカニズムにより、ブートドライブ SSD をさらに強化します。

- **IntelligentLog™** は、SSD の健全性を包括的に記録することで、システム開発者が SSD の現在の状態を把握し、適切に対処できるようにします。また、主要なイベントログを記録することで、デバッグの迅速化と効率化を実現します。
- **IntelligentScan™** は自動スキャンと修復を実行し、保護領域に保存されたデータの正確な読み書きを確保すると同時に、データ保持期間を延長し、SSD の寿命を延ばします。
- **IntelligentThermal™** は、ブートドライブの温度を監視し、一定のしきい値を超えた場合に冷却メカニズムを作動させます。ホスト制御による温度管理とドライブ制御による温度管理を適切に使い分けることで、温度を制御し、高温環境下での SSD の損傷を防止します。
- **IntelligentShield™** は、転送中のすべてのデータを専用の保護領域へ確実に保存し、停電によるデータ消失を防ぎます。電源の不安定さを検出すると、ドライブは直ちにこのメカニズムを起動します。

- IntelligentFlush™ は、一定速度でデータを即座に移行し、安全なストレージブロックへ迅速に書き込むことができます。
- IntelligentGuard™ は、ブートドライブのファームウェアにおける認証機能と保護機能を強化することで、ブートセキュリティを維持します。
- IntelligentImage™ は、組み込みシステム向けのプリロード機能を提供し、表面実装前にデータを安全なブロックへ事前書き込めるようにします。これにより、リフローはんだ付け時の高温など、製造工程におけるリスクからデータを保護するとともに、生産ラインの効率を大幅に向上させます。
- IntelligentZone™ は、マルチネームスペース設計をサポートし、システム設計者ごとの要件に合わせて SSD を最適化できるようにします。SSD を複数のストレージブロックに分割し、設計に基づいてデータを適切なブロックに配置することで、データ完全性を確保するとともに、SSD 容量を効率的に活用し、経済性を最大限に高めます。

AI クラウド環境における重要な管理点としてのブートドライブ

ブートドライブは容量こそ小さいものの、ダウンタイムが稼働率、コスト、サービス品質に直結する AI クラウド環境では、パフォーマンス、信頼性、セキュリティに与える影響は極めて大きいといえます。ブートドライブは、もはや単なる補助コンポーネントではなく、AI スタック全体を支える管理点です。

クラウドプロバイダーが分散 GPU フリートの拡張を進める中、迅速な起動、予測可能な復旧、そして信頼できるシステム状態の維持は、バックグラウンド機能ではなく、中核的な運用要件となっています。PCIe NVMe SSD は、こうした要求に最適なストレージであり、AI プラットフォームの継続的な稼働と障害への耐性を支えるために必要な低レイテンシ、高 IOPS、耐久性、電力効率を提供します。

Silicon Motion のエンタープライズ PCIe NVMe ブート SSD ソリューションは、独自の IntelligentSeries™ テクノロジーにより、一貫した動作、電源喪失保護、セキュアブート機能に加え、クラウド AI インフラストラクチャへの大規模導入を支えるインテリジェントな管理機能を提供することで、この利点をさらに高めます。



画像 1: Silicon Motion のエンタープライズ PCIe NVMe ブート SSD ソリューションは、BGA や M.2 をはじめとする多様なフォームファクタに対応しています。

詳細については、www.siliconmotion.com をご覧いただくか ferri@siliconmotion.com までメールにてお問い合わせください