



Cloudera および Apache Spark を実装したビッグデータ および分析向け Cisco UCS 統合インフラストラクチャ

ビッグデータのリアルタイム分析を実現

概要



ビッグデータとインメモリ分析のための包括的統合インフラストラクチャ

- ビッグデータおよび分析のための Cisco UCS 統合インフラストラクチャは、高いパフォーマンス、容量、拡張性を、Apache Spark と Cloudera Enterprise に提供します。
- 1 台のラック筐体に収まる規模から複数台のラックを使用する規模に至るまで、実績のある高性能のリニアな拡張性を実現し、アーキテクチャの拡張を容易にします。



導入が簡単

- Cisco UCS Manager は、設定エラーやシステムダウンタイムの減少に役立つポリシーベースの自動化メカニズムによって、インフラストラクチャの導入をシンプル化します。



管理のシンプル化

- ビッグデータ向け Cisco UCS Director Express は、プロビジョニング、インストール、設定をワンクリックで実現し、迅速かつ簡単にビッグデータ インフラストラクチャに導入できます。エンドツーエンドのシステム管理を実現し、エンタープライズ データ ハブのあらゆる部分を詳しく精密に可視化および制御できる包括的なインターフェイスである Cloudera Manager と、このソリューションとを組み合わせることで使用することにより、シンプルでわかりやすいクラスタ管理が実現します。



柔軟なビッグデータ プラットフォーム

- Cloudera Enterprise 上の、ビッグデータおよび分析向け Cisco UCS 統合インフラストラクチャでは、Apache Spark を、スタンドアロン クラスタ モードで導入できます。また、Hadoop および主流の NoSQL と合わせて導入することもできます。



Spark Streaming によるリアルタイムデータ処理

- Hadoop の Spark Streaming と Apache Kafka をビッグデータおよび分析向け Cisco UCS 統合インフラストラクチャに追加することにより、小さなまとまりでデータを取り込んで変換を行うストリーム分析が可能になります。



Apache Spark によるデータ ストリームのバッチ処理とリアルタイム処理

ビッグデータ テクノロジーは長い間、その努力の大半をバッチ処理にのみ費やしてきました。その結果、ビッグデータから価値を引き出すには、通常、数時間から数日の時間がかかっていました。企業は、ビッグデータのバッチ処理の可能性を認識する一方で、さらに多くのデータを収集しようとする状況の中、データからリアルタイムで価値を取得したいと考えるようになりました。センサー、Internet of Things (IoT) デバイス、ソーシャル ネットワーキング、オンライントランザクションなど、これらすべてからデータは生成されます。これを迅速にキャプチャ、監視、処理して、感情分析や探索的分析を提供したり、警告を発したりするための決定を、データベースに基づいて即座に行う必要があるのです。

Apache Spark は、大規模データ処理のための高速汎用エンジンです。Spark とあわせて、より多くの企業が Hadoop を採用し、ストリーミングや機械学習など、以前よりはるかに幅広いワークロードを処理する機能を獲得しています。Spark はオペレーションのほとんどをメモリで処理するので、アプリケーション パフォーマンスが大幅に加速します。この機能によって新しいインタラクティブ アプリケーションが実現され、アナリストやデータ サイエンティストはシミュレーションをより短期間で行って生産性を高めることができます。すでに導入済みの Hadoop に加えて Spark を使用することにより、新しい機能を迅速に導入できます。Hadoop エコシステムと高い互換性があるため、Hadoop 分散ファイル システム (HDFS) のデータを使用することができ、Hadoop 2.0 YARN で動作可能です。Spark のコアは、同じアプリケーションで透過的に使用できる強力なハイレベルのライブラリ式によって補完されます。これらのライブラリには、現在、SparkSQL、Spark Streaming、MLlib (機械学習用) が含まれています。さらに、Spark は、R、Java、Scala など、一般的な開発言語を幅広くサポートしています。Spark の基本的な概念として、Resilient Distributed Dataset が採用されています。RDD は、読み取り専用のフォールトトレラント分散オブジェクト コレクションで、並行して運用することが可能です。

Apache Spark の現在の使用方法の実例を次に示します。

- シンプルで高速な抽出、変換、ロード (ETL) 処理: 途中でディスク書き込み操作をすることなく、必要な形式でデータを処理できます。また、最終的なディスク書き込み操作の前にメモリ内でクリーニングおよび集約が可能です。
- リアルタイムのアクション: 異常な動作をリアルタイムで検出し、それに従ってダウンストリーム アクションを実行します。たとえば、想定とは異なる場所でクレジットカード取引が発生した際に詐欺の警告などのアクションを生成したり、IoT センサーからのデバイス障害データ転送によって障害回復プロセスを始動させたりできます。
- データ統合: キャッシュされた静的データ セットにライブ データを加えることにより、データの情報量がさらに豊富になり、より包括的な機能セットをリアルタイムで使用できます。
- 探索的分析: 特定の時間帯に関連するイベントをグループ化して分析し、データサイエンティストがそれを Spark の機械学習モデルの更新に利用することができます。

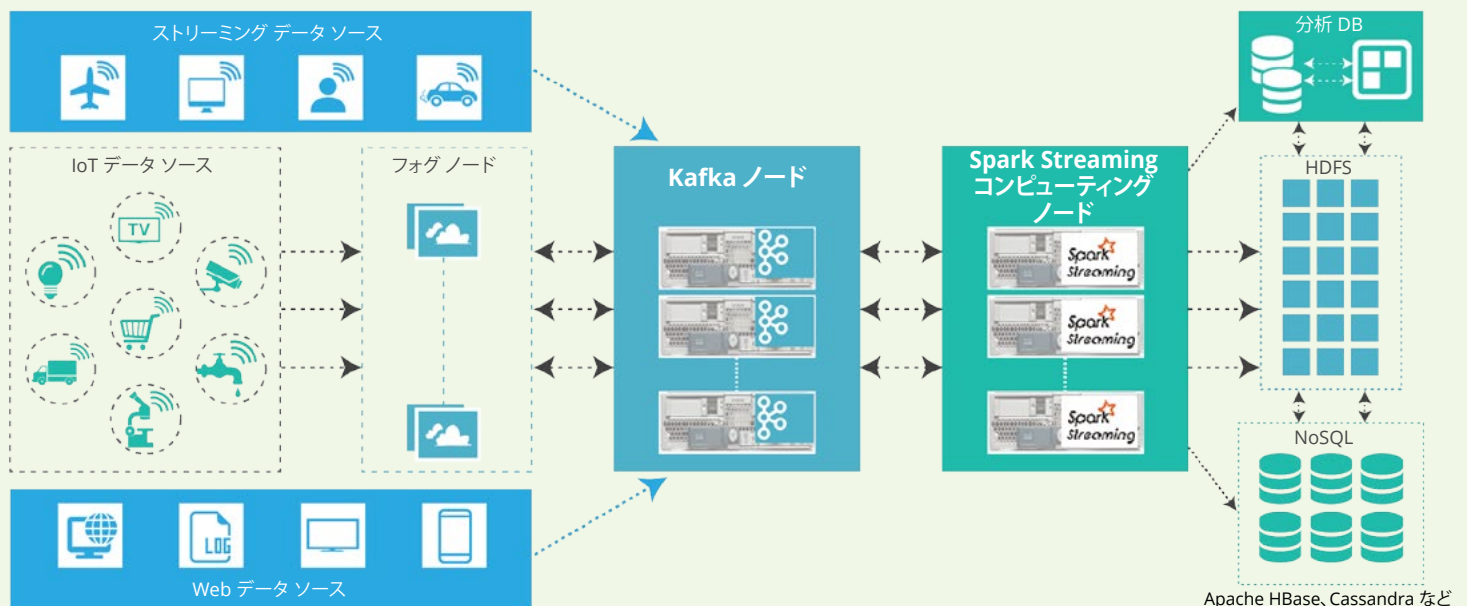
図 1 に、Spark と Spark Streaming の最も代表的な使用例を示します。この図では、さまざまなデータソースからのデータがフォグ ノードと Kafka ノードに流れ、次に Spark、そしてさらに下流の HDFS、NoSQL、SQL データベース、Elastic、Solr およびその他のシステムへと流れ、追加処理が行われる様子を示しています。

ビッグデータおよび分析向け Cisco UCS 統合インフラストラクチャ

現代の組織は、基盤となる物理インフラストラクチャの導入、拡張、管理を、ワークロードとビジネス要件の変化に対応した機動的な変更が可能になるような方法で、実施しなければなりません。Cisco Unified Computing System™ (Cisco UCS®) は、統合インフラストラクチャに対する画期的なアプローチによってデータセンターの新たな可能性を広げ、IT のイノベーションおよび高速化へのビジネスニーズに応えます。ビッグデータおよび分析向け Cisco UCS 統合インフラストラクチャは、構造化または非構造化の別を問わず大量のリアルタイム データまたはアーカイブ データを処理するエンドツーエンド アーキテクチャを提供します。同時に、関連する複合機能を透過的に統合し、アプリケーションの要求に応じて高いパフォーマンスと拡張性を備えたエンタープライズクラスのサービスを提供します。

Cisco UCS で Spark に使用する 2 つの共通リファレンス アーキテクチャは、Cisco UCS C シリーズ ラック サーバでも使用できます。1 つのアーキテクチャでは、Hadoop インフラストラクチャに Spark 処理が追加され、もう 1 つは、Kafka または同種のテクノロジーによるストリーム処理を実現します。これらの基本的な構成要素を図 2 に示し、次のセクションで説明します。

図1: Spark、Spark Streaming、Kafka を備えたラムダ アーキテクチャ



Cisco UCS 6200 および 6300 シリーズ ファブリック インターコネクト

Cisco UCS 6300 および 6200 シリーズ ファブリック インターコネクトは、高帯域幅で低遅延の接続をサーバに提供し、すべての接続デバイスを Cisco UCS Manager で統合/統一管理できるようにします。Cisco UCS 6300 シリーズ ファブリック インターコネクトは、Cisco UCS の中核を成しており、低遅延、ロスレスの 10/40 ギガビット イーサネット、Fibre Channel over Ethernet (FCoE)、およびファイバ チャンネル機能を実現し、あわせて、冗長ペアで導入するシステム向けの管理機能も提供します。Cisco® ファブリック インターコネクトによって、ビッグデータ アプリケーションを処理する標準的クラスタに存在する多数のノードをサポートするために必要な、完全なアクティブ-アクティブの冗長性、パフォーマンス、優れた拡張性が実現します。Cisco UCS Manager では、サービス プロファイルによる迅速かつ一貫したサーバ構成が可能で、また、クラスタ全体に対するファームウェア更新などの継続的なシステム保守アクティビティを 1 つの操作として自動化できます。Cisco UCS Manager には高度なモニタリング機能もあり、クラスタ全体の健全性についてアラームを発信し、通知を送信することも可能です。

Cisco UCS C シリーズ ラック サーバ

Cisco UCS C240 M4 ラック サーバは、コンパクトな高密度設計で、コンピューティング、I/O、ストレージ容量に関するさまざまな要求に対応します。このサーバは Intel® Xeon® プロセッサ E5-2600 v4 シリー

ズ CPU をデュアル構成で使用し、最大 768 GB のメイン メモリに加えて、ハード ディスク ドライブ (HDD) およびソリッドステート ディスク (SSD) の多様なオプションをサポートしています。パフォーマンス最適化オプションでは、24 基の小型フォーム ファクタ (SFF) ディスクドライブがサポートされ、キャパシティ最適化オプションでは、12 基の大型フォーム ファクタ (LFF) ディスクドライブがサポートされています。このサーバでは、Cisco UCS 仮想インターフェイス カード (VIC) 1227 または 1387 を使用できます (使用するファブリック インターコネクトによって異なる)。VIC 1227 は、高帯域幅で低遅延のクラスタ接続を実現する最適な設計になっています。VIC 1387 は、デュアルポート拡張 Quad Small Form-Factor Pluggable (QSFP+) 40 ギガビット イーサネットおよび FCoE を、モジュール型 LAN-on-motherboard (mLOM) フォーム ファクタで提供します。

ラムダ アーキテクチャに対応したビッグデータおよび分析向け Cisco UCS 統合インフラストラクチャ

Apache Kafka と連携した Spark Streaming の使用例では、Kafka を 4 ~ 8 の追加ノードに導入することができます (ストリーミングのデータ要件によって異なる)。Kafka ノードは Cloudera Manager で管理できますが、Hadoop クラスタに含める必要はなく、他の Hadoop サービスを実行する必要はありません (図 3)。

図2: ビッグデータおよび分析向け Cisco UCS 統合インフラストラクチャ

Cisco UCS 6296UP ファブリック インターコネクト X 2

Cisco UCS C240 M4 サーバ X 64

10 ギガビット リンク X 16

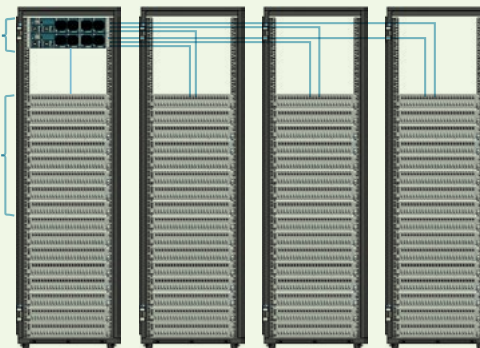
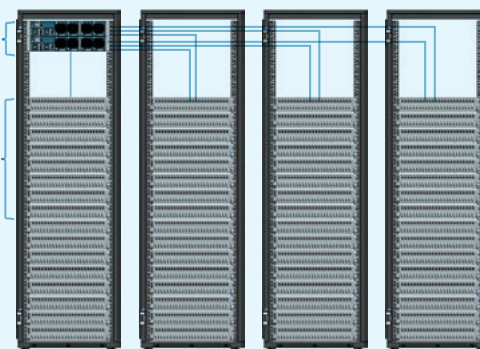


図3: Apache Kafka を実装したビッグデータおよび分析向け Cisco UCS 統合インフラストラクチャ

Cisco UCS 6296UP ファブリック インターコネクト X 2

Kafka ノード Cisco UCS C240 M4 X 8

10 ギガビット リンク X 16



Hadoop ノード Cisco UCS C240 M4

Hadoop クラスタ

(NameNode、リソース マネージャ、DataNode、Spark 実行プログラム)

Kafka を使用しながらストリーミング アーキテクチャを拡張するには、表 1 のスケーリングおよびサイジングのガイドラインに従ってください。表には、サーバドライブおよび複製に関するファクタやネットワーク帯域幅の条件がさまざまに異なる Kafka ストレージのガイドラインを記載しています。

注: 1 つのサーバがいっぱいになるまでの時間 = $\sim (\text{総ストレージ容量} / \text{ネットワーク帯域幅}) / 3600$

Apache Spark と連携した Cloudera Enterprise

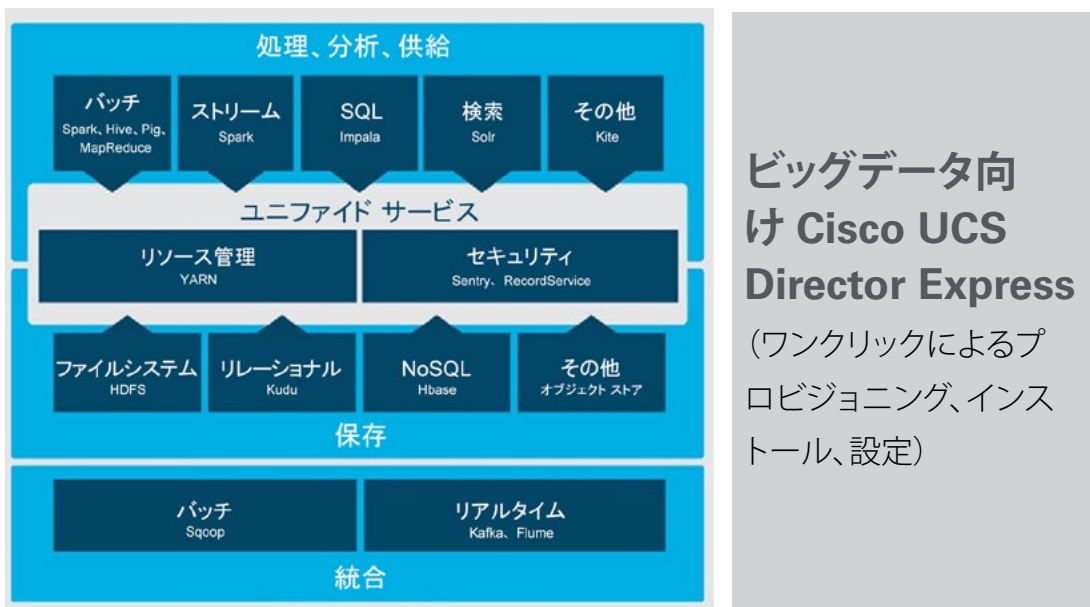
Cloudera は、エンタープライズ対応のビッグデータ ソフトウェアおよびサービスを提供する業界トップクラスのプロバイダーです。Cloudera

Enterprise には、市場をリードする Hadoop ディストリビューション (Apache Hadoop を含む Cloudera ディストリビューション (CDH))、高度な管理ツール (Cloudera Manager)、ネイティブのエンドツーエンド データ ガバナンス ソリューション、包括的なセキュリティ ツールセット、およびテクニカル サポートが含まれています (図 4)。シスコは、Cloudera との提携により、エンタープライズ対応データ管理プラットフォームと、エンタープライズ アプリケーション エコシステムとの管理の統合を提供します。両社はシームレスに連携して、Hadoop ベースのアプリケーション向けに、独自の機能を備えた業界トップクラスのアーキテクチャ プラットフォームを提供します。

表 1: Kafka ノードのストレージ用スケーリングおよびサイジングのガイドライン

ネットワーク帯域幅	サーバタイプ	使用可能な総ストレージ容量	1 つのサーバがいっぱいになるまでの時間	サーバ総数 (3 ウェイ複製データおよびネットワークのフル利用)
10 Gbps (1.25 GBps)	Cisco UCS C240 M4 (SFF) (1.8 TB ドライブ搭載)	~ 40,800 GB	~ 9 時間	1 日分のデータの保管に ~ 9 台のサーバ
40 Gbps (5 GBps)	Cisco UCS C240 M4 (SFF) (1.8 TB ドライブ搭載)	~ 40,800 GB	~ 2.3 時間	1 日分のデータの保管に ~ 30 台のサーバ
10 Gbps (1.25 GBps)	Cisco UCS C240 M4 (LFF) (6 TB ドライブ搭載)	~ 72,000 GB	~ 16 時間	1 日分のデータの保管に ~ 6 台のサーバ
40 Gbps (5 GBps)	Cisco UCS C240 M4 (LFF) (6 TB ドライブ搭載)	~ 72,000 GB	~ 4 時間	1 日分のデータの保管に ~ 18 台のサーバ

図 4: Cloudera Enterprise



ビッグデータ向け Cisco UCS Director Express (ワンクリックによるプロビジョニング、インストール、設定)

ビッグデータおよび分析向け Cisco UCS 統合インフラストラクチャ

業界をリードする Cloudera の製品とソリューションは、Apache Hadoop および関連プロジェクトの導入と管理、データの処理と分析、安全で保護されたデータ保持を実現する機能を企業に提供します。

Cloudera が提供する製品およびツール:

- **Cloudera Enterprise:** Cloudera Enterprise には、Spark をはじめとして、Apache Hadoop や関連するその他のオープンソース プロジェクトの Cloudera ディストリビューションが含まれています。Cloudera Enterprise はまた、多数のハードウェアおよびソフトウェア ソリューションと統合し、セキュリティも確保しています。
- **Apache Spark:** Cloudera Enterprise に組み込まれている Spark は、インメモリ データを柔軟に処理してバッチ分析、リアルタイム分析、高度な分析を行うためのオープン スタンドードです。Cloudera は、分析ワークロードのデフォルト データ処理エンジンとして Spark を採用することを確定しています。
- **Cloudera Manager:** この高度なアプリケーションは、Cloudera の導入、管理、監視、および Cloudera 導入に関する問題の診断

に使用します。Cloudera Manager には、シンプルでわかりやすい企業データ管理を実現する Web ベースのユーザ インターフェイスである、Admin Console が含まれています。

- **Cloudera Navigator:** Cloudera Enterprise プラットフォーム用のエンドツーエンド データ管理ツール。管理者、データ マネージャ、アナリストが使用して Hadoop の大量のデータを探索することができます。Cloudera Navigator の強力な監査、データ管理、システム管理、ライフサイクル管理機能により、企業は厳しいコンプライアンスおよび規制の要件を順守することができます。

リファレンス アーキテクチャ

ビッグデータおよび分析向け Cisco UCS 統合インフラストラクチャは、表 2 に示すように数種類の構成を提供して、さまざまなコンピューティングおよびストレージ要件に対応します。

表 2: ビッグデータおよび分析向け Cisco UCS 統合インフラストラクチャのオプション

パフォーマンス最適化オプション 1 (UCS-SL-CPA4-P1)	パフォーマンス最適化オプション 2 (UCS-SL-CPA4-P2)	パフォーマンス最適化オプション 3 (UCS-SL-CPA4-P3)	キャパシティ最適化オプション 1 (UCS-SL-CPA4-C1)	キャパシティ最適化オプション 2 (UCS-SL-CPA4-C2)
<p>接続性:</p> <ul style="list-style-type: none"> • Cisco UCS 6296UP 96 ポート ファブリック インターコネクト X 2 <p>拡張性:</p> <ul style="list-style-type: none"> • Cisco Nexus 9000 または 7000 シリーズ スイッチを使用してサーバー数千台まで拡張可能 <p>Cisco UCS C240 M4 ラック サーバ (SFF) 16 台 (サーバ構成内容は以下):</p> <ul style="list-style-type: none"> • Intel Xeon プロセッサ E5-2680 v4 CPU X 2 (各 CPU に 14 のコア) • 256 GB のメモリ • Cisco 12 Gbps SAS Modular RAID コントローラ (2 GB フラッシュベース書き込みキャッシュ (FBWC) 搭載) • 1.2-TB 10000 rpm SFF SAS ドライブ X 24 (合計 460 TB) • ブート用 240 GB 6 Gbps 2.5 インチ Enterprise Value SATA SSD X 2 • Cisco UCS VIC 1227 (10 ギガビットイーサネット SFP+ ポート X 2 搭載) 	<p>接続性:</p> <ul style="list-style-type: none"> • Cisco UCS 6296UP 96 ポート ファブリック インターコネクト X 2 <p>拡張性:</p> <ul style="list-style-type: none"> • Cisco Nexus 9000 または 7000 シリーズ スイッチを使用してサーバー数千台まで拡張可能 <p>Cisco UCS C240 M4 ラック サーバ (SFF) 16 台 (サーバ構成内容は以下):</p> <ul style="list-style-type: none"> • Intel Xeon プロセッサ E5-2680 v4 CPU X 2 (各 CPU に 14 のコア) • 256 GB のメモリ • Cisco 12 Gbps SAS Modular RAID コントローラ (2 GB フラッシュベース書き込みキャッシュ (FBWC) 搭載) • 1.8-TB 10000 rpm SFF SAS ドライブ X 24 (合計 691 TB) • ブート用 240 GB 6 Gbps 2.5 インチ Enterprise Value SATA SSD X 2 • Cisco UCS VIC 1227 (10 ギガビットイーサネット SFP+ ポート X 2 搭載) 	<p>接続性:</p> <ul style="list-style-type: none"> • Cisco UCS 6332 ファブリック インターコネクト X 2 <p>拡張性:</p> <ul style="list-style-type: none"> • Cisco Nexus 9000 または 7000 シリーズ スイッチを使用してサーバー数千台まで拡張可能 <p>Cisco UCS C240 M4 ラック サーバ (SFF) 16 台 (サーバ構成内容は以下):</p> <ul style="list-style-type: none"> • Intel Xeon プロセッサ E5-2680 v4 CPU X 2 (各 CPU に 14 のコア) • 256 GB のメモリ • Cisco 12 Gbps SAS Modular RAID コントローラ (2 GB フラッシュベース書き込みキャッシュ (FBWC) 搭載) • 1.8 TB 10000 rpm SFF SAS ドライブ X 24 (合計 691 TB) • ブート用 240 GB 6 Gbps 2.5 インチ Enterprise Value SATA SSD X 2 • Cisco UCS VIC 1387 (40 ギガビットイーサネット SFP+ ポート X 2 搭載) 	<p>接続性:</p> <ul style="list-style-type: none"> • Cisco UCS 6296UP 96 ポート ファブリック インターコネクト X 2 <p>拡張性:</p> <ul style="list-style-type: none"> • Cisco Nexus 9000 または 7000 シリーズ スイッチを使用してサーバー数千台まで拡張可能 <p>Cisco UCS C240 M4 ラック サーバ (LFF) X 16 (サーバ構成内容は以下):</p> <ul style="list-style-type: none"> • Intel Xeon プロセッサ E5-2620 v4 CPU X 2 (各 CPU に 8 つのコア) • 128 GB のメモリ • Cisco 12 Gbps SAS Modular RAID コントローラ (2 GB フラッシュベース書き込みキャッシュ (FBWC) 搭載) • 6 TB 10000-rpm LFF SAS ドライブ X 12 (合計 1152 TB) • ブート用 240 GB 6 Gbps 2.5 インチ Enterprise Value SATA SSD X 2 • Cisco UCS VIC 1227 (10 ギガビットイーサネット SFP+ ポート X 2 搭載) 	<p>接続性:</p> <ul style="list-style-type: none"> • Cisco UCS 6296UP 96 ポート ファブリック インターコネクト X 2 <p>拡張性:</p> <ul style="list-style-type: none"> • Cisco Nexus 9000 または 7000 シリーズ スイッチを使用してサーバー数千台まで拡張可能 <p>Cisco UCS C240 M4 ラック サーバ (LFF) X 16 (サーバ構成内容は以下):</p> <ul style="list-style-type: none"> • Intel Xeon プロセッサ E5-2620 v4 CPU X 2 (各 CPU に 8 つのコア) • 256 GB のメモリ • Cisco 12 Gbps SAS Modular RAID コントローラ (2 GB フラッシュベース書き込みキャッシュ (FBWC) 搭載) • 8 TB 7200-rpm LFF SAS ドライブ X 12 (合計 1536 TB) • ブート用 240 GB 6 Gbps 2.5 インチ Enterprise Value SATA SSD X 2 • Cisco UCS VIC 1227 (10 ギガビットイーサネット SFP+ ポート X 2 搭載)

Cloudera Enterprise 用の Cisco UCS リファレンス アーキテクチャは、Hadoop Enterprise の導入に必要な大規模な拡張性をサポートします。ここで説明した構成は、96 ポート Cisco UCS ファブリック インターコネクトのペアを使用することで、最大 80 台のサーバをサポートするように拡張可能です。Cisco Nexus® 9000 または 7000 シリーズ スイッチの使用により、複数の Cisco UCS ドメイン (最大でサーバ数千台規模) をサポートできます。

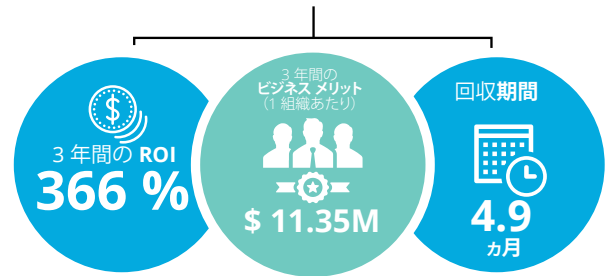
まとめ

ビッグデータおよび分析向け Cisco UCS 統合インフラストラクチャに実装された Apache Spark は、業界をリードするパフォーマンスを発揮する包括的なプラットフォームを提供します。このソリューションにより、組織は、Cisco UCS のシンプルな導入モデルを使って Spark を簡単に導入し、オン デマンドでソリューションを拡張して、Spark による強力なビッグデータ分析処理をサポートすることができます。

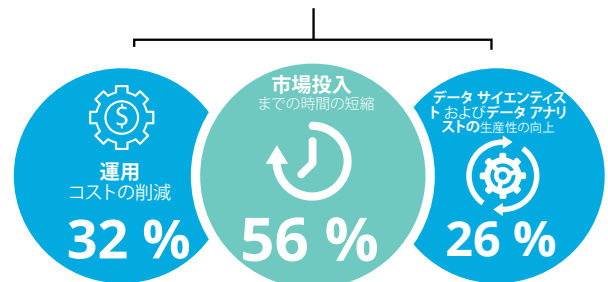
関連情報

- Cisco UCS ビッグデータ ソリューションの詳細については、<http://www.cisco.com/jp/go/bigdata/> を参照してください。
- ビッグデータ向け Cisco UCS 統合インフラストラクチャの詳細については、<http://blogs.cisco.com/datacenter/cpav4/> [英語] を参照してください。
- Cloudera に実装した Spark の詳細については、<https://www.cloudera.com/products/apache-hadoop/apache-spark.html> [英語] を参照してください。

ビッグデータ向け Cisco UCS 統合インフラストラクチャのビジネス価値のまとめ



ビッグデータ向け Cisco UCS 統合インフラストラクチャのビジネス価値のまとめ



ビジネス価値上のメリット: Cisco UCS サーバ 1 台あたりの平均年間利益

ビジネス生産性 \$ 29,654	IT スタッフの生産性 \$ 3,861	IT インフラストラクチャの コスト削減 \$ 123
-----------------------------	--------------------------------	--

©2016 Cisco Systems, Inc. All rights reserved.

Cisco、Cisco Systems、および Cisco Systems ロゴは、Cisco Systems, Inc. またはその関連会社の米国およびその他の一定の国における登録商標または商標です。

本書類またはウェブサイトに掲載されているその他の商標はそれぞれの権利者の財産です。

「パートナー」または「partner」という用語の使用は Cisco と他社との間のパートナーシップ関係を意味するものではありません。(1502R)

この資料の記載内容は2016年10月現在のものです。

この資料に記載された仕様は予告なく変更する場合があります。



シスコシステムズ合同会社

〒107-6227 東京都港区赤坂9-7-1 ミッドタウン・タワー
<http://www.cisco.com/jp>

お問い合わせ先