

IBM BigInsights for Apache Hadoop を実装したビッグデータ向け Cisco UCS 統合インフラストラクチャ



ソリューション概要

概要



企業のビッグデータ環境向けに最適化

ビッグデータ向け Cisco UCS® 統合インフラストラクチャは、ニーズに応じて小規模から非常に大規模なレベルにまで拡張可能なモジュラフレームワークである IBM BigInsights for Apache Hadoop 向けに、バランスの取れたコンピューティング能力、I/O 帯域幅、およびストレージ容量を提供します。



Cisco UCS 上に構築するメリット

- このソリューションは、ユニファイド ファブリック、統合管理、および高度なモニタリング機能を提供します。
- Cisco UCS サービス プロファイルを利用する迅速で一貫性のある導入により、優れたパフォーマンスを即座に利用できます。
- 簡素化されたポリシーベース管理が可能です。
- グローバルなインベントリ ビュー、ワンクリックでのシステム ソフトウェア管理、およびワンクリックでの設定変更などが可能です。



リスクを軽減

- 事前検証、緊密な統合、およびパフォーマンスの最適化により、統合および導入に伴うリスクが軽減されます。
- ソフトウェア ディストリビューションについて広範なテストと検証が実施済みであるため、ソリューションを安心して導入できます。



データ処理および分析の速度と効率性を向上

- Cisco® と IBM による共同ソリューションは、パフォーマンス、容量、拡張性に加えて、広範な管理機能と可視化機能を提供します。
- 広範な開発者ツールと強力な分析機能が提供されます。
- Hadoop 向けの高性能 SQL エンジンである Big SQL により、分析時間が短縮されます。
- IBM BigSheets の働きにより、Hadoop 上でのデータの準備や閲覧をコーディングなしに実行できます。
- IBM BigInsights Big R を使用してビッグデータ規模で実行されるネイティブな R 関数により、大規模データに対する機械学習のパフォーマンスが自動的に調整されます。

シスコと IBM が実現する、ビッグデータ分析を促進する業界トップクラスのソリューション

IBM BigInsights for Apache Hadoop 向けの Cisco Unified Computing System™ (Cisco UCS) ソリューションは、ビッグデータ向け Cisco UCS 統合インフラストラクチャを基盤としており、業界トップクラスのコンピューティング、ネットワーク、および管理機能が、ビッグデータ ワークロード向けに最適化されたファブリックベースのユニファイド アーキテクチャに統合されています。知見の迅速な提供と総所有コスト (TCO) の削減に重点を置いて最適化された、この共同開発ソリューションは、データ内に存在するインテリジェンスを開放して、持続可能かつ強力なビジネス上の強みを生み出すことが可能です。

ソリューション: IBM BigInsights を実装したビッグデータ向け Cisco UCS 統合インフラストラクチャ:

- Cisco UCS によるプログラム可能なインフラストラクチャや優れたオープンソース ソフトウェアなどのイノベーションと、IBM BigInsights for Apache Hadoop によるエンタープライズクラスの機能が組み合わせられて提供されます。
- シスコと IBM が共同で設計、テスト、および検証し、詳細に文書化した環境は、予測どおりの導入が可能です。ワークロードのニーズに合わせた拡張も容易です。
- シスコと IBM により、データの操作や管理タスクを容易にするための、透視的なインフラストラクチャ管理機能や、Hadoop とその他の情報管理ソリューションを統合するための機能が提供されます。
- Enterprise SQL on Hadoop for Businesses による高いパフォーマンスと拡張性により、データサイエンスや分析が加速され、データを深く掘り下げることが可能になります。
- シスコ、IBM、およびパートナー企業から、業界をリードするワールドワイドなサポートとサービスが提供されます。

提携:



IBM BigInsights for Apache Hadoop 向けの Cisco UCS ソリューション

このソリューションは、Cisco UCS 6200 シリーズ ファブリック インターコネクタと、IBM BigInsights for Apache Hadoop 向けに最適化された Cisco UCS C シリーズ ラック サーバを使用して、Cisco UCS インフラストラクチャ上に構築されています (図 1 参照)。

Cisco UCS 6200 シリーズ ファブリック インターコネクタ

Cisco UCS 6200 シリーズ ファブリック インターコネクタは、システム全体の接続と管理を一元化します。サーバに高帯域幅で低遅延の接続を提供し、すべての接続デバイスを Cisco UCS Manager で一元管理できるようにします。通常、ビッグデータ アプリケーションに対応しているクラスタには大量のノードが存在します。冗長ペアで展開される Cisco® ファブリック インターコネクタは、それらのノードのサポートに必要なアクティブ-アクティブの完全冗長性、パフォーマンス、および優れた拡張性を提供します。Cisco UCS Manager は、

サービス プロファイルによる迅速で一貫性のあるサーバ構成を実現し、クラスタ全体のファームウェア アップデートの操作を一元化するなどして、継続的なシステムメンテナンス作業を自動化します。さらに、クラスタ全体の状態に関するアラームや通知などのオプションを備えた、高度なモニタリング機能を提供します。

Cisco UCS C シリーズ ラック サーバ

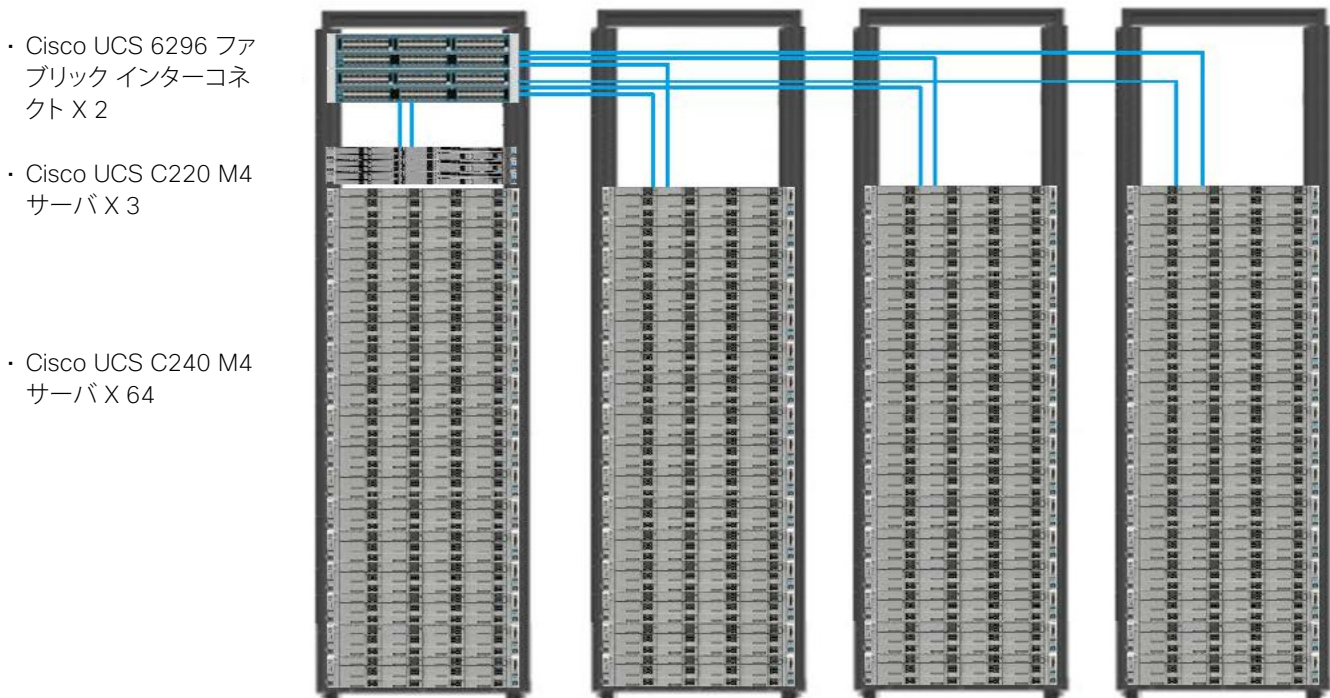
Cisco UCS C シリーズ ラック サーバは、業界標準のラック筐体でユニファイド コンピューティングの機能を提供できるため、総所有コストの軽減と俊敏性の向上に役立ちます。このシリーズの各モデルは、処理、メモリ、I/O、内蔵ストレージリソースのバランスを取ることで、処理負荷にまつわるさまざまな課題に対応しています。

Cisco UCS C240 M4 ラック サーバは、コンパクトな設計でありながら、コンピューティング、I/O、およびストレージ容量に関する広範なニーズに対応可能です。サーバの中核となるのは Intel® Xeon® プロセッサ E5-2600 v3 シリーズで、12 Gbps の SAS スループットにより、旧世代のサーバよりはるかに優れたパ

フォーマンスと効率が実現されています。このサーバは Intel Xeon プロセッサ E5-2600 v3 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、広範なハード 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) 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、キャパシテイ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ます。さらに 2 基の 1 ギガビット イーサネット組み込み LAN-on-motherboard (LOM) ポートが用意されています。Cisco UCS 仮想インターフェイスカード

(VIC) 1227 は、Cisco UCS C シリーズ ラック サーバの M4 世代向けに設計されています。VIC は、高帯域幅で低遅延のクラスタ接続用に最適化されており、Cisco UCS Manager によりオンデマンドで構成される最大 256 台の仮想デバイスをサポート可能です。

図 1: IBM BigInsights を実装したビッグデータ向け Cisco UCS 統合インフラストラクチャ: 64 データ ノード クラスタ



IBM BigInsights for Apache Hadoop: 完全な Hadoop プラットフォーム

IBM BigInsights for Apache Hadoop は、Hadoop 向けの新しい分析機能やエンタープライズ機能、たとえば IBM InfoSphere BigInsights、Big R、IBM Big SQL などの機能拡張を使用する機械学習や、最新のオープンソース Apache パッケージを提供することにより、データサイエンティスト、アナリスト、および管理者によるデータサイエンスタスクを迅速化します。

Big SQL

Big SQL は、SQL エンジンの分野で IBM が持つ強みを生かして、データが Hadoop またはリレーショナルデータベースのどちらに存在しているかにかかわらず、Java データベースコネクティビティ (JDBC) またはオープンデータベースコネクティビティ (ODBC) を介して、Hadoop からあらゆるシステム上のデータに対する透過的な ANSI SQL アクセスを実現します (図 2 参照)。

SQL プログラミング言語に習熟している開発者であれば、新しい言語やスキルを習得する必要なしに、Hadoop 内のデータにアクセスできます。

Big SQL を使用すると、あらゆるビッグデータに SQL を介してアクセスできるようになります。Big SQL は、利用可能なリソースに合わせて最適な処理戦略を使用して、既存データの構造化されたビューを提供します。Big SQL は、複雑なデータセットに対しては必要に応じて MapReduce パラレルizmを使用できる一方で、MapReduce によりパフォーマンスが阻害される場合には、ダイレクトアクセスによる、より簡潔で低遅延なクエリを実行します。

Big SQL が提供する主な機能は、次のとおりです。

- ・ 大規模並列処理 (MPP) テクノロジーによる低遅延クエリ。
- ・ クエリリライト最適化およびコストベースの最適化。

- ・ Hive および HBase データソースの統合。
- ・ SQL ANSI 標準の優れたサポート。
- ・ IBM DB2、Oracle、Teradata、および ODBC ソースに対するフェデレーテッドクエリアクセス。

Big SQL は、データウェアハウスのモダナイズや次世代論理データウェアハウスの構築に関する、一般的なあらゆる使用例をサポートしています。

- ・ 既存のデータウェアハウスからデータやワークロードをオフロード。
- ・ Hadoop 内にクエリ可能なアーカイブを作成し、使用頻度の低いデータを高コストなデータウェアハウスから移動。
- ・ ビジネスインテリジェンスレポートのプロトタイプを作成を迅速化。
- ・ 既存の SQL スキルを使用して、Hadoop の迅速な導入を確実にサポート。

図 2: IBM Big SQL の機能



BigSheets

BigSheets は、構造化されたデータベースの管理にとどまらず、非構造化データの管理にも対応することで、セルフ方式での分析を現実のものとし、全体像の把握が容易になることで、組織内のあらゆるレベルで意思決定が改善されます。

BigSheets は、Web ベースのスプレッドシート形式のビューを Hadoop 内のファイルコレクションに提供します。ユーザは、データ変換、データのフィルタリング、ビューの作成などを、大規模に実行できます。BigSheets によって、スプレッドシートアクションが MapReduce プロセスに変換されて、Hadoop クラスタのコンピューティングリソースが使用されるため、コーディングは必要ありません。そのため、アナリストはデータ内に存在する価値を迅速かつ容易に見出すことができます。

BigSheets はモデルの拡張機能として、次のようなメリットを提供します。

- Web ベースのリポジトリ内の非構造化データを、ギガバイト、テラバイト、またはペタバイト規模で統合。
- ユーザ定義のシード URL から広範な非構造化 Web データを収集。
- 選択した非構造化情報管理アーキテクチャ (LanguageWare、OpenCalais など) を使用して、収集データを抽出および強化。
- これらのデータを、ユーザ定義の個別のコンテキスト (ManyEyes など) 内で、検索および閲覧可能。

さらに、BigSheets には次のようなメリットもあります。

- データのエスカレーションに対応するための新しいアプローチを、ビジネスユーザに提供します。BigSheets はデータを構造化することにより、追加ストレージを必要とすることなく、ペタバイト規模のデータマイニングを可能にします。

- データを利用しやすい状況固有のリファレンスフレームに分割するための新しいアプローチを、ビジネスユーザに提供します。この機能は、組織が利用されていない非構造化データ (その多くが未知の Web データ) を、実用的なインテリジェンスに変換するうえで役立ちます。
- Hadoop クラスタのあらゆるコンピューティングリソースを活用して、クラスタ内から知見やビューを提供します。データの抽出は必要ありません。

データサイエンティスト向けに、BigInsights Data Scientist モジュールには Big R および高度なテキスト分析機能などが組み込まれています。

Big R

Big R は、データサイエンティストがネイティブな R 関数を使用して、R 環境内からビッグデータに対する検索、閲覧、変換、およびモデリングを行うことを可能にします。データサイエンティストは、広範なアルゴリズムを使用してスケーラブルな機械学習アルゴリズムを実行できる他、R ライクな構文を使用して新しいアルゴリズムを作成したり、既存のアルゴリズムをカスタマイズすることが可能です。Big R を実行する BigInsights for Apache Hadoop は、クラスタ全体のメモリの使用、ディスクへのスpill、数千ものモデルの並列処理などを実行できます。

Big R の主なメリットは、次のとおりです。

- オープンソース R とのエンドツーエンドの統合
- Hadoop 上での透過的な処理
- Big R により提供される広範かつスケーラブルな機械学習アルゴリズムへの透過的アクセス
- 非構造化データから有意な情報を取り出すためのテキスト分析機能

テキスト分析

開発者は、BigInsights 独自の高度なテキスト分析機能を使用することで、複数言語で記述されたテキストを処理する高品質アプリケーションを簡単に作成して、さまざまな形式の膨大なネイティブテキストデータから知見を取得できます。

管理者向けには、BigInsights Enterprise Management モジュールにより、管理コンソールと組み込みのセキュリティ機能が提供されます。

管理コンソール

BigInsights に組み込まれている包括的な Web ベースのインターフェイスは、クラスタ管理、サービス管理、ジョブ管理、およびファイル管理を簡素化します。

管理者およびユーザは、同一インターフェイスを共有して、アプリケーションを起動できる他、さまざまな構成可能なレポートやダッシュボードを利用できます。

組み込みのセキュリティ

BigInsights はセキュリティを重視して設計されており、Kerberos 認証をサポートするとともに、監査およびモニタリング機能によるデータプライバシー、マスキング、詳細なアクセス制御を提供することで、環境の安全性維持をサポートします。

リファレンス アーキテクチャ

あらゆるビッグデータに対応可能な Cisco UCS 統合インフラストラクチャの最新バージョンは、IBM BigInsights のコンピューティングおよびストレージ要件に基づいて、次に示す構成で提供されま
す(表 1 参照)。

表 1:

IBM BigInsights を実装したビッグデータ向け Cisco UCS 統合インフラストラクチャ

接続性:

Cisco UCS 6296UP 96 ポート ファブリック インターコネクト X 2

拡張性:

- ・ドメインごとに最大 80 台のサーバ
- ・Cisco Nexus® 7000 および 9000 シリーズ スイッチを使用して、数千台のサーバにまで拡張可能

IBM BigInsights 用管理ノード(クラスタあたり):

以下を搭載した Cisco UCS C220 M4 ラック サーバ X 3:

- ・Intel Xeon プロセッサ E5-2680 v3 CPU X 2
- ・256 GB のメモリ
- ・600 GB 10,000-rpm SAS HDD X 8
- ・Cisco 12 Gbps SAS Modular RAID コントローラ(2 GB フラッシュベース ライト キャッシュ(FBWC) 搭載)
- ・Cisco UCS VIC 1227 (10 ギガビット イーサネット X 2 搭載)

注: IBM BigInsights 4.x については、管理サービス(ネームノード、リソース マネージャなど)を高可用性構成で実行するために 6 つの管理ノードが必要です。

データ ノード(ラックあたり):

以下を搭載した Cisco UCS C240 M4 ラック サーバ(LFF) X 16:

- ・Intel Xeon プロセッサ E5-2680 v3 CPU X 2
- ・128 GB のメモリ
- ・Cisco 12 Gbps SAS モジュラ RAID コントローラ(2 GB FBWC 搭載)
- ・6 TB 7200-rpm LFF SAS ドライブ X 12 (合計 768 TB)
- ・ブート用 120 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)

詳細情報

シスコのビッグデータ ポータル: www.cisco.com/jp/go/bigdata/

ビッグデータ向け Cisco UCS 設計: http://www.cisco.com/go/bigdata_design [英語]

IBM BigInsights を実装したビッグデータ向け Cisco UCS 統合インフラストラクチャ:

- ・シスコ検証済みデザイン(CVD): http://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/Cisco_UCS_Integrated_Infrastructure_for_Big_Data_with_IBM.html [英語]
- ・シスコ データセンター ブログ: <http://blogs.cisco.com/datacenter/biginsights> [英語]

IBM BigInsights for Apache Hadoop: <http://www-01.ibm.com/software/data/infosphere/hadoop/enterprise.html> [英語]

©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。(1602R)

この資料の記載内容は2016年2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