Ready Solutions for Data Analytics

Splunk Enterprise on Dell EMC Infrastructure

2020 年 2 月

H18106

リファレンス アーキテクチャ

概要

このリファレンス アーキテクチャは、Dell EMC インフラストラクチャ上の Splunk Enterprise でマシン データ分析を行うためのアーキテクチャおよび設計情報を提供します。Dell EMC PowerEdge サーバーおよび PowerSwitch ネットワーク スイッチ上に構築され、Splunk コールド バケット データを保存するための Dell EMC Isilon ストレージも含まれています。

Dell EMC ソリューション

著作権

目次

第 1 章　概要

この章は、次のトピックで構成されています。

はじめに

あらゆる企業、組織、システム、デバイスは、毎日大量のデータ ストリームを生成しています。このデータは、データ センターの中心からネットワークのエッジにいたる、あらゆる場所で発生します。マシンで生成されたこのデータには、以下のような重要な情報が含まれています。

- ユーザーの行動
- セキュリティ上のリスク
- 容量消費量
- サービス レベル
- 不正行為
- カスタマー エクスペリエンス
- その他

マシン データは、データ分析の中で最も急速に成長しているセグメントの 1 つであり、高い価値を持っています。しかし多くの場合、組織の中で最も活用されておらず、過小評価されている資産の 1 つでもあります。マシン データは、次のいずれの分野でも組織が運用インテリジェンスを取得できる重要な資産です。

- IT 運用
- セキュリティ
- ビジネス分析
- IoT
- ビジネスのその他の重要な側面

デジタル トランスフォーメーションを推進するには、マシン データからリアルタイムのインサイトとビジネス バリューを提供する必要があります。ただし、このデータの使用には、大きな課題が伴います。

従来のデータ分析、管理、モニタリング ソリューションは、大量、高速、多様なデータを処理するように設計されていません。マシン データは、ビジネス インテリジェンスやデータ ウェアハウス ツールで処理および分析することが困難な、多くの予測不可能な形式で提供されます。これらの処理および分析方法は、厳格なスキーマを使用した構造化データ向けに設計されています。

ビジネス リーダーは、新しい、または拡張された運用インテリジェンス機能をもたらすように最適化された検証済みのソリューションによって、コストとリスクを削減しながら、イノベーションを加速することを目指しています。Dell EMC、インテル、Splunk は、パートナーシップを通じ、ビジネスにおけるデジタル トランスフォーメーションを支援する拡張性とパフォーマンスを念頭に置いて設計された、標準化リファレンス アーキテクチャを提供しています。これらのソリューションは、Splunk の分析機能と、コスト効率、拡張性、柔軟性に優れたインテルベースの Dell EMC ハードウェアのインフラストラクチャを組み合わせて、組織に最適な運用インテリジェンスをもたらします。

本書では、Dell EMC とインテルのテクノロジーに基づいて構築された、Splunk 向けエンタープライズ インテリジェンス プラットフォームのビジネスレベルの概要を示しています。さまざまなビジネスの規模と種類の、現実的ないくつかの Splunk 導入に向けたリファレンス アーキテクチャ構成について説明しています。

Splunk およびソリューションのメリットの概要

Splunk 上に構築されたエンタープライズ インテリジェンス プラットフォームは、拡張性に優れたハイ パフォーマンスのデータ分析を実現します。これらの分析により、企業やサービス プロバイダーは、マシンおよびイベント データをインサイトに変換し、データ分析のパワーを企業全体のユーザーにもたらすことができます。業界をリードする Splunk の Data-to-Everything プラットフォームとインテルテクノロジーのリファレンス構成を組み合わせることで、組織は Splunk の導入を迅速化し、拡張することができます。この組み合わせにより、未加工データを運用、ビジネス、セキュリティ インテリジェンスに変換できるようになります。

Splunk のパワーは、ビジネスのあらゆる部分でデータを解き放つ能力にあります。データは次のような多くのソースからもたらされます。

- アプリケーション
- デバイス
- ネットワーク
- オペレーティング システム
- IoT センサー
- Web トラフィック
- その他

組織は、データの結合や質問、キーワードでの検索、アクションの実行、ビジネス目標に向けた取り組みが可能になります。結果として得られたインサイトは、セキュリティ上の脅威の特定、アプリケーション パフォーマンスの最適化、お客様の行動の理解に役立つ可能性があります。

Splunk の導入を最大限に活用するために、ユース ケースに関係なく、IT 組織は次のことを行う必要があります。

1. Splunk のワークロードを把握する。
2. インフラストラクチャを最適化して、Splunk ができるだけ効率的に実行されるようにする。

この知識は、低い検索ランタイム、高いデータ取得量、高い同時検索数につながります。複雑になりがちなこのプロセスをシンプルにするために、Dell EMC、インテル、Splunk は連携して、いくつかの異なる構成を使用したリファレンス アーキテクチャを設計しました。これらの構成は、最新のインテル テクノロジーを取り入れた Dell EMC のサーバーとストレージを使用して、さまざまな Splunk ワークロードに対応するように設計されています。

図 1 は、Splunk Enterprise が多くの異なるソースからデータを取得して処理し、実用的なインサイトを生み出して、イベント データやマシン データの価値を最大限に引き出すことを示しています。

図 1　Splunk データの取得、処理、結果

マシン データのパワーを効率的に活用する

アプリケーション、オペレーティング システム、ネットワーク、セキュリティ ソフトウェア、クライアント デバイス、Internet of Things（IoT）などのテクノロジーは、大量のイベント データやマシン データを生成します。マシン データのモニタリングと分析は、Web トラフィックの分析、お客様の行動の把握、財務分析の合理化、カスタマー エクスペリエンスの向上など、多くの問題を組織が解決していく上で役立ちます。しかし、マシン データは複雑であり、企業はそのデータをタイムリーなインサイトに変換する上で大きな課題に直面しています。

Splunk Enterprise は、企業が次のようなことを実行できるようにすることで、これらの課題の解決を支援します。

- 未加工の形式のマシン データやイベント データを調査する。
- IT、セキュリティ、IoT システムにおけるデータの流れを監視する。
- 傾向を分析する。
- 必要な行動を取る。

しかし、どのようなテクノロジーでもそうですが、総所有コストと投資収益率を最大化するのに最も大切なのは効率性です。Splunk Enterprise を効率的に導入するには、Splunk ワークロードの特性と、Splunk のコンポーネントがどのように連携するかについて理解する必要があります。一部のワークロードではデータのインデックスの作成を主に行う必要があり、他のワークロードでは検索の実行に焦点が当てられます。また、データのインデックスの数と検索クエリーの数のバランスがうまく取れているワークロードもあります。

多くの組織が、ワークロードについて理解せずにそのワークロード向けに Splunk インフラストラクチャを構成して、Splunk から最大限のパフォーマンスと拡張性を得ることができないでいます。膨大で急速に増加しているデータ ボリュームに対してタイムリーな分析を行うことは困難です。

データをアクセス可能、使用可能、価値あるものにする

Splunk を基盤とするエンタープライズ インテリジェンス プラットフォームは、最新の「Data-to-Everything」プラットフォームです。このプラットフォームは、広範囲のデータ ソースからもたらされたインサイトを保存、整理、分析、取得する強力な機能を提供します。

Splunk プラットフォームを使用すると、組織は次のような重要なユース ケースに対応することができます。

- **IT ビジネス運用**：Splunk は、リアルタイムのモニタリング、イベント管理、アラート機能を提供し、物理的および仮想的な IT インフラストラクチャの健全性を可視化します。Splunk は、アプリ

ケーション、ビジネス サービス、IT サービスのモニタリング機能も提供します。この広範囲の分析機能は、ダウンタイムを防ぐだけでなく、最適なカスタマー エクスペリエンスと円滑なビジネス運用を実現する上で役立ちます。

- **セキュリティとコンプライアンス**：Splunk は、リアルタイムのモニタリング、履歴分析、膨大なデータセットの可視化により、セキュリティ調査を迅速化します。セキュリティ チームは、包括的なインシデント調査を実施し、アド ホック レポートを数分で作成することができます。一般的なセキュリティのユース ケースには、以下が含まれます。
  - 不正分析と検出
  - 内部関係者による脅威
  - インシデント調査およびフォレンジック
  - 高度な脅威の検出
  - インシデント レスポンス
  - コンプライアンス
  - データ プライバシー
  - セキュリティ オペレーション センターの自動化とオーケストレーション
- **アプリケーション配信**：Splunk はアプリケーション スタック全体でリアルタイムの可視性を実現し、アプリケーション パフォーマンス、トランザクション、ユーザー アクティビティに関するエンド ツー エンドのビューをもたらします。IT 部門は、迅速にアプリケーションを配信し、アプリケーションの品質、パフォーマンス、コストを最適化することができます。
- **ビジネス分析**：Splunk は、複雑なビジネス プロセス、お客様の行動、製品の使用状況、デジタル マーケティング キャンペーンを把握できるようにします。Web サイトやモバイル アプリを通じてさらに収益を上げることを求めている企業は、タイムリーで関連性の高いビジネス インサイトを得ることができます。
- **IoT および業界データ**：Splunk により、運用のモニタリング、使用状況の分析、ビジネス運用のエンド ツー エンド ビューへのインサイトの統合が実現できます。Splunk は、このような高度な作業を、接続デバイス、制御システム、センサー、SCADA（監視制御とデータ取得）システムによって生成されたデータを使用して成し遂げます。

本書で説明されているリファレンス アーキテクチャおよび構成オプションは、これらのユース ケースのいずれにも適用できます。

Splunk Enterprise の概要

Splunk Enterprise は、企業がさまざまなシステム コンポーネントから収集したデータを収集、検索、整理、分析、視覚化できるようにするソフトウェア製品です。Splunk Enterprise は、Web サイト、アプリケーション、センサー、デバイスなどのさまざまなソースのログ データおよびストリーミング データを取得します。Splunk Enterprise は、各データ ソースからデータ ストリームのインデックスを作成し、表示と検索が可能な一連の個々のイベントに解析します。データは Splunk Web インターフェイスを使用して、さらに分析することができます。Splunk の検索言語、ルックアップ、マクロ、サブ検索により、数時間に及ぶ退屈な作業はわずか数秒に短縮されます。タグ、保存済みの検索、ダッシュボードは、運用に関するインサイトと共同作業手段の両方を提供します。

図 2 は、データの解析、インデックス作成、検索を行う Splunk Enterprise コアを示しています。

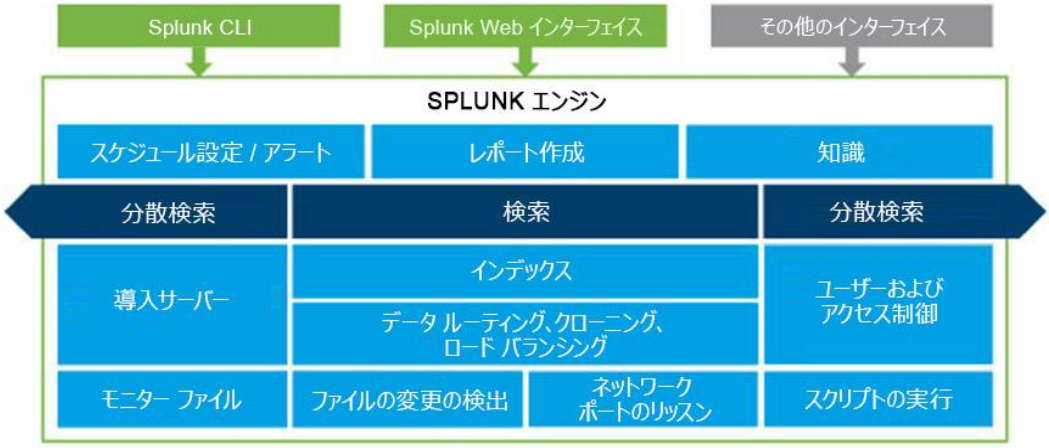

図 2　Splunk Enterprise コア

Splunk Enterprise ソフトウェアは、データ コレクションと分析の分野に新たなバリュー プロポジションをもたらします。従来の抽出、変換、ロード（ETL）システムでは、すべてのデータを構造化してからインサイトを収集する必要があるため、分析プロセスが遅くなります。しかし、Splunk Enterprise は異なります。これは、抽出、ロード、変換（ELT）プラットフォームです。つまり、スキーマオンデマンド（「スキーマオンリード」、「スキーマオンニード」、「スキーマオンユース」とも呼ばれます）をサポートしています。スキーマオンデマンドにより、データを最初に取得し、後でデータに構造化を適用することができます。Splunk Enterprise では、未処理の新しいデータ ソースをいつでも追加できます。

Splunk のメリット

ユース ケースによっては、Dell EMC インフラストラクチャ上の Splunk Enterprise のリファレンス アーキテクチャは、次のようなビジネス バリューを提供できます。

- **ダウンタイムの短縮**：Splunk を使用して、数千もの IoT デバイスからマシン データを収集、分析、報告することで 、予知保全を向上させることができます。Splunk の使用は、機器のアップタイムと顧客満足度の向上にもつながります。
- **継続的な脅威修復**：Splunk によって、侵害につながる可能性のある脆弱性を監視することで、潜在的なインシデント、感染したシステム、またはその両方を特定できます。Splunk では、継続的なセキュリティ モニタリングによって脅威を検出し、セキュリティ対策を決定し、コンプライアンスについて報告することができます。
- **よりスマートな生産のインサイト**：リアルタイムのモニタリングと予測分析によって、機械学習のパワーを最大限に引き出すことができます。
  - パフォーマンスのベースラインを理解する。
  - 偏差を予測する。
  - 生産、資産管理、サプライ チェーン マネジメントに関連するインテリジェントな推奨事項を提供する。
- **アプリケーションのアップタイム**：アプリケーションの開発速度と品質によって、競争上の優位性とカスタマー エクスペリエンスを向上させます。Splunk は、アプリケーション開発のライフサイクル全体にわたってリアルタイムのインサイトを提供します。

対象読者

このガイドは、Splunk Enterprise 環境の評価、購入、管理、保守、または運用を行う担当者を対象としています。例えば、次のような担当者が含まれます。

- IT 管理者
- ストレージ管理者
- 仮想化管理者
- システム管理者
- IT マネージャー

第 2 章　ソリューション コンポーネント

この章は、次のトピックで構成されています。

リファレンス アーキテクチャの概要

このセクションでは、Splunk Enterprise on Dell EMC Infrastructure リファレンス アーキテクチャについて説明します。Dell EMC は、Splunk が推奨するリファレンス ハードウェア上で実行される Splunk Enterprise のパフォーマンスを満たすか、またはそれを上回るように、このアーキテクチャを設計しました。図 3 は、Splunk Enterprise on Dell EMC Infrastructure リファレンス アーキテクチャにおける Splunk Enterprise のハイレベル ビューを示しています。

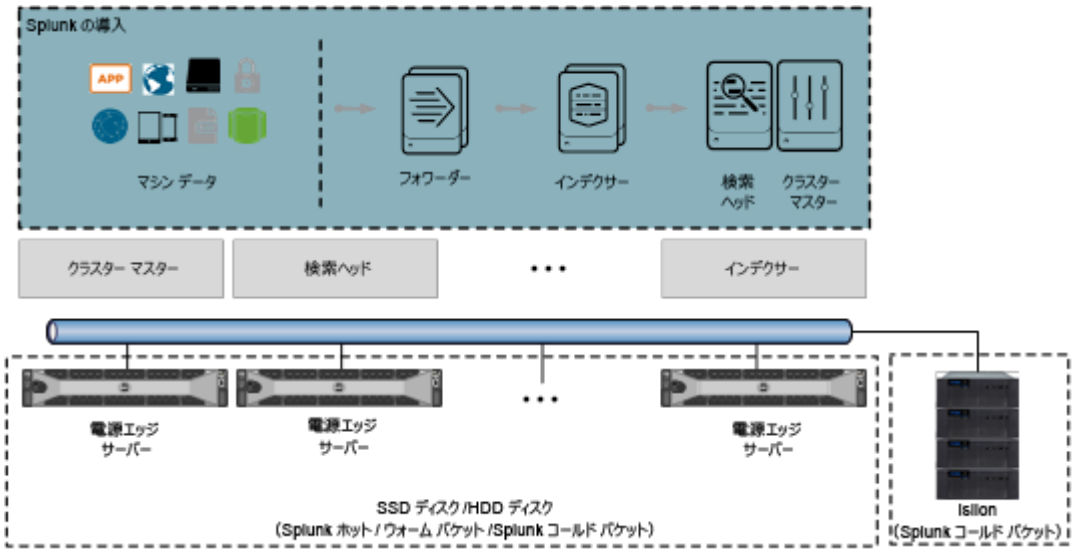

図 3 Splunk Enterprise on Dell EMC Infrastructure **リファレンス アーキテクチャ**

このリファレンス アーキテクチャは、Splunk ホット/ウォーム バケット データに SSD ストレージを使用することで、ハイ パフォーマンスと低レイテンシーの I/O を実現します。また、Splunk コールド バケット データに HDD ディスクまたは Isilon ストレージを使用することで、大容量を実現します。

注：ホット/ウォームおよびコールド バケットの概念の説明については、「Splunk コア アーキテクチャ」を参照してください。

Dell EMC PowerEdge サーバー

PowerEdge サーバーは、IT 組織の業務をサポートするように設計されています。このサーバーは、最も要求の厳しいビジネス アプリケーションに対応できるように設計されており、以下のようなワークロードをより効率的に実行するための特定の機能を備えています。

- ハイパフォーマンス コンピューティング（HPC）
- コラボレーション
- データベース
- エンタープライズ リソース プランニング（ERP）

- ビジネス インテリジェンス
- データ ウェアハウスおよびデータ分析

完全かつ適応性の高いソリューションの基盤として、PowerEdge サーバーは、卓越したパフォーマンスと管理のメリットを提供し、お客様が最も頻繁に実行するビジネス アプリケーションにパワーをもたらします。

PowerEdge サーバーは、革新的な Dell EMC OpenManage Systems Management のポートフォリオと業界をリードするワークロード ソリューションを組み合わせて以下のことを実現します。

- インテリジェントかつシンプルなテクノロジーを提供する。
- 最も複雑な環境における処理能力を向上させる。

最新世代のサーバーは、以下の分野でお客様のニーズに応えます。

- **メモリー容量と拡張性**：大容量のメモリー フットプリント
- **仮想化パフォーマンス**：より多くのプロセッサー コアとメモリー密度の向上
- **システム管理**：Lifecycle Controller を搭載した Integrated Dell Remote Access Controller（iDRAC）の使用と、OpenManage Essentials を使用したモニタリングおよび更新機能による包括的なライフ サイクル管理
- **エネルギー効率**：Dell EMC OpenManage Power Center を含む、包括的な最適化
- **インフラストラクチャの柔軟性**：ネットワーク インターフェイス コントローラーなどの革新による、より優れた、より多くの I/O オプションの提供
- **信頼性**：ほとんどのサーバーでのフェールセーフ ハイパーバイザー オプションを含む、より多くのリモート アクセス サービス（RAS）機能

このリファレンス アーキテクチャは、Dell EMC PowerEdge R640 サーバーと Dell EMC PowerEdge R740xd サーバーを使用しています。

Dell EMC ネットワーキング

Dell EMC PowerSwitch ネットワーキングは、あらゆる規模のデータ センターに適した、柔軟でパワフルなトップオブラック（ToR）スイッチを提供します。これらのスイッチは、オープン ネットワーキング時代に向けた最新のワークロードとアプリケーションを導入するように設計されています。これらは、ハードウェアとソフトウェアの冗長性により、低レイテンシー、卓越したパフォーマンス、高密度を実現します。

このリファレンス アーキテクチャは PowerSwitch S4148F-ON および S3048-ON モデルを使用していますが、他のスイッチ モデルも使用できます。

Dell EMC Isilon

本書で説明する構成は、以下のデータ保持機能を提供します。

- SSD を使用した、インデクサー サーバー上での 30 日間分のホットおよびウォーム バケットのデータ ストレージ

- SAS ドライブを使用した、120 日間または 365 日間のコールド バケット保存

Dell EMC Isilon スケールアウト ネットワーク接続型ストレージ（NAS）は、追加のコールド バケット ストレージおよびフローズン バケット ストレージに使用できます。

Isilon H-Series は、大容量およびハイ パフォーマンスの柔軟で包括的なストレージ製品です。

Isilonストレージではインテリジェントなソフトウェアを使用して多くのコモディティー ハードウェア ユニット間でデータを拡張し、パフォーマンスと容量の急速な増加に対応します。Isilon の革新的なストレージ アーキテクチャである OneFS オペレーティング システムは、1 つのクラスタ化されたファイル システムを提供します。

OneFS は、オペレーティング システムの詳細レベルに並列性を組み込むことによって価値をもたらします。仮想的には、システムは複数のハードウェア ユニットに分散されています。この並列性により、インフラストラクチャの拡大に伴い、OneFS はあらゆる次元で拡張できるようになります。複数の冗長性レベルを提供することにより、システムには単一障害点がなくなります。その結果、OneFS は、従来のシステムよりも優れた信頼性を維持しながら、複数ペタバイトの規模にまで拡張できます。

OneFS は、Isilon スケールアウト NAS ハードウェアで実行されます。これにより、コモディティー ハードウェアのコストと効率の継続的な向上からメリットを得ることができます。OneFS では、クラスターへのハードウェアの追加とクラスターからのハードウェアの削除がいつでもできます。データはハードウェアの変更から保護されます。この機能によって、データ移行やハードウェアの更新のコストと負担が軽減されます。

Splunk Enterprise

Splunk Enterprise は、IT インフラストラクチャ内のさまざまなソースから収集されたマシン生成データの収集、インデックス作成、可視化を可能にするソフトウェア プラットフォームです。これらのソースには、アプリケーション、ネットワーキング デバイス、ホストおよびサーバーのログ、モバイル デバイスなどが含まれます。

Splunk は、データのサイロを運用上のインサイトに変換し、IT インフラストラクチャ全体のエンドツーエンドの可視性を提供し、迅速な問題解決と、データに基づく決定を可能にします。

Splunk コア アーキテクチャ

図 4 は、Splunk システム アーキテクチャの概要を示しています。Splunk Enterprise インスタンスは、サーチ ヘッド、インデクサー、または小規模な導入に向けて、その両方の役割を果たすことができます。1 日のデータ取得量または検索負荷が、統合インスタンス環境のサイジングに関する推奨値を超えた場合、Splunk Enterprise ではインデクサーとサーチ ヘッドをさらに追加することで、水平方向に拡張します。詳細については、『Splunk キャパシティ プランニング マニュアル』を参照してください。

図 4　Splunk アーキテクチャの概要

Splunk Enterprise インデクサーがデータを受信すると、イベントのタイム スタンプに基づいて、未加工データを個別のイベントに解析します。次に、これらを適切なインデックスに書き込みます。Splunk は、ホット/ウォームおよびコールド データ バケットによるストレージ階層を実装し、新たにインデックスが作成されたデータのパフォーマンスを最適化します。このオプションは、古いデータをより大容量のストレージで長期間保持します。

新たにインデックスが作成されたデータはホット バケットに配置されます。これにより、Splunk はアクティブに読み取りと書き込みを行います。ホット バケットは、次の場合にウォーム バケットに移動されます。

- ホット バケットの数が指定のしきい値に達した場合、または
- ホット バケット内のデータのサイズが指定のしきい値を超えた場合。

ウォーム バケットは、ホット バケットと同じ階層のストレージに配置されます。唯一の違いは、ウォーム バケットは読み取り専用であることです。ホット/ウォーム データ用のストレージは、最速のストレージ階層である必要があります。その理由は、それが Splunk Enterprise の導入パフォーマンスに最も大きな影響を与えるためです。

ウォーム バケット数またはボリューム サイズがしきい値を超えた場合、 データはコールド バケットに移動されます。これはストレージの別の階層に含めることができます。レイテンシーが 5 ミリ秒未満（理想）で、200 ミリ秒を超えない場合、コールド データはネットワーク ファイル システム（NFS）マウントに配置できます。NAS テクノロジーは、パフォーマンスと 1TB あたりのコスト削減の許容可能な組み合わせを提供します。このためこのテクノジーは、コールド データの長期保存に適した選択肢となります。

データはアーカイブまたは凍結することもできますが、そのようなデータは Splunk サーチ ヘッドによって検索できなくなります。データを Splunk Enterprise バケットに戻して検索可能にするには、ユーザーの手動操作が必要になります。コンプライアンスの保存要件を満たすために、フローズン(凍結)バケットを使用することもできます。しかし本書では、Isilon の優れた拡張性と競争力のある所有コストにより、コールド バケットにより多くの検索可能なデータを保存できる方法を示しています。図 5 は、Splunk バケットの概念についての詳細情報を示しています。

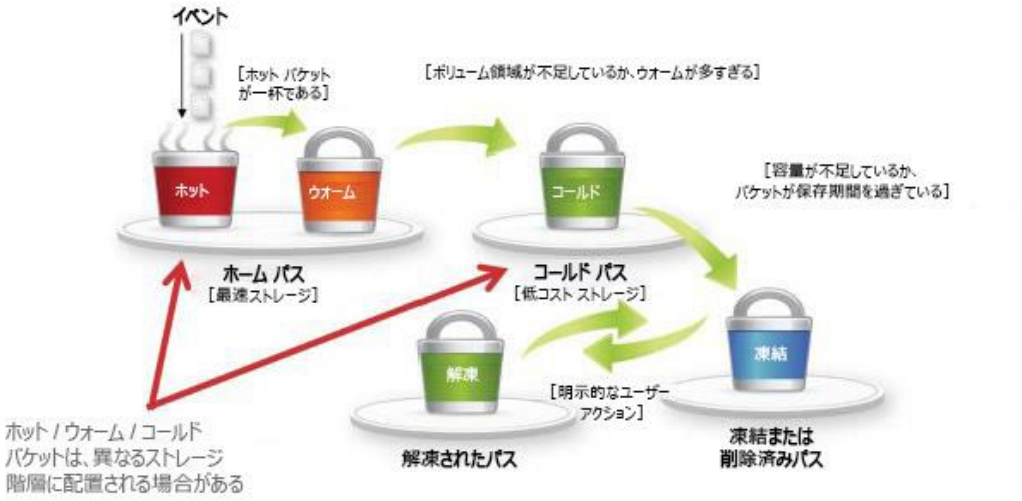

図 5　Splunk インデックス バケット

構成のサイズ設定の概要

Splunk Enterprise on Dell EMC Infrastructure のリファレンス アーキテクチャは、多くのお客様の Splunk 本番稼働環境における実績に基づいて設計されています。アーキテクチャには、Splunk Enterprise を本番稼働環境に導入し、管理するために必要なハードウェア、ソフトウェア、リソース、サービスがすべて含まれています。このリファレンス アーキテクチャは、さまざまなお客様のニーズに対応する 3 つの構成に対する、Dell EMC on Dell EMC Infrastructure について記述しています。

このリファレンス アーキテクチャでは、シングル インスタンス モード、分散モード、またはインデクサー クラスター モードで Splunk を導入できます。構成と最小サーバー数の概要については、表 1 を参照してください。

表 1.　3 つの構成パターン

構成パターン	リファレンス	ミッドレンジ	ハイ パフォーマンス
日次データ取得量	200 GB	250 GB	300 GB
ホット/ウォーム保存期間（日）	30	30	30
コールド保存期間（日）	120 または 365	120 または 365	120 または 365
サーチヘッド	スタンダード x1	スタンダード x1	スタンダード x1
管理サーバー	スタンダード x1	スタンダード x1	スタンダード x1
インデクサー	200 GB x1	250 GB x1	300 GB x1
Isilon コールド ストレージ	オプション	オプション	オプション

Splunk でのインデックス作成は、圧縮アルゴリズムと似ています。共通の文字列がデータに含まれており、情報へのポインタが生成されます。結果として得られるデータはディスクに保存され、次のメリットをもたらします。

- ストレージの節約
- 圧縮データの検索機能

結果として得られる圧縮比（圧縮サイズまたは元のサイズ）は通常 1 未満で、受信データによって異なります。ランダムなバイナリ データを圧縮しても、サイズは節約できません。ログ ファイルなどの頻繁に反復されるデータは、圧縮比が大幅に高くなる可能性があります。

このような圧縮コストとパフォーマンスのばらつきは、ソリューションのパフォーマンスとストレージ要件（データ取得量など）に大きく影響します。また、ITSI、エンタープライズ セキュリティ、ユーザー行動分析などのプレミアム アプリケーションを使用している場合も異なります。プレミアム アプリケーションは、セキュリティ モニタリングやユーザーの行動分析などの追加機能を 1 つのパッケージで提供します。

ソフトウェア コンポーネント

表 2 は、Splunk Enterprise on Dell EMC Infrastructure に推奨されるソフトウェアとバージョンを示しています。

表 2　ソフトウェアに関する推奨事項

ソフトウェア	バージョン
Splunk Enterprise	8.0.1
OneFS	8.2.0 以降

第 3 章　ソリューション アーキテクチャ

この章は、次のトピックで構成されています。

概要

この章では、Splunk Enterprise on Dell EMC Infrastructure の設計と構成について詳しく説明します。ここでは、日次データ取得量に基づいた 3 種類の異なる構成を見ていきます。

- **リファレンス**：1 日あたり最大 200 GB のデータ取得、オンボード コールド ストレージでの 120 日または 365 日の保存
- **ミッドレンジ**：1 日あたり 200～250 GB のデータ取得、オンボード コールド ストレージでの 120 日または 365 日の保存
- **ハイ パフォーマンス**：1 日あたり最大 300 GB のデータ取得、オンボード コールド ストレージでの 120 日または 365 日の保存

コンピューティング設計

Splunk Enterprise は、次の 3 種類の導入をサポートしています。

- シングル インスタンス導入
- 分散導入
- クラスタ化導入

本書では主に、以下で構成される分散導入について説明します。

- サーチ ヘッド サーバー
- 管理サーバー
- 1 つまたは複数のインデクサー

本書では、3 種類のデータ取得量のシナリオのシステムについて説明し、2 種類の異なる保存期間のオプションを提供します。特定のワークロードのサイジングの詳細については、次のような追加の要因についての検討が必要になります。

- クラスターのユーザーの総数
- 実行予定の同時クエリーの数

これらの要因は、Splunk Enterprise on Dell EMC Infrastructure を効率的に実行するためのインデクサーとサーチ ヘッドの数に影響を及ぼします。

追加のシナリオに対応するためのクラスター構成のガイドラインについては、表 1 を参照してください。

ネットワーク設計

クラスター ネットワークは、冗長性とシステム管理機能へのアクセスを実現しながら、ハイ パフォーマンスで拡張性の高いクラスターのニーズを満たすように設計されています。

このアーキテクチャは、Dell Networking 10 GbE ToR スイッチと、BMC システム管理ネットワーク用の 1 GbE スイッチを使用した高可用性（HA）設計となっています。

クラスターでは、表 3 で説明する 2 つのネットワークが使用されています。

表 3　クラスター ネットワーク

論理ネットワーク	接続	スイッチ
クラスター データ ネットワーク	10 GbE (ボンディング)	ToR スイッチ
BMC ネットワーク	1 GbE	ラックごとの専用スイッチ

クラスター ネットワークは、図 6 に示すように設計されています。

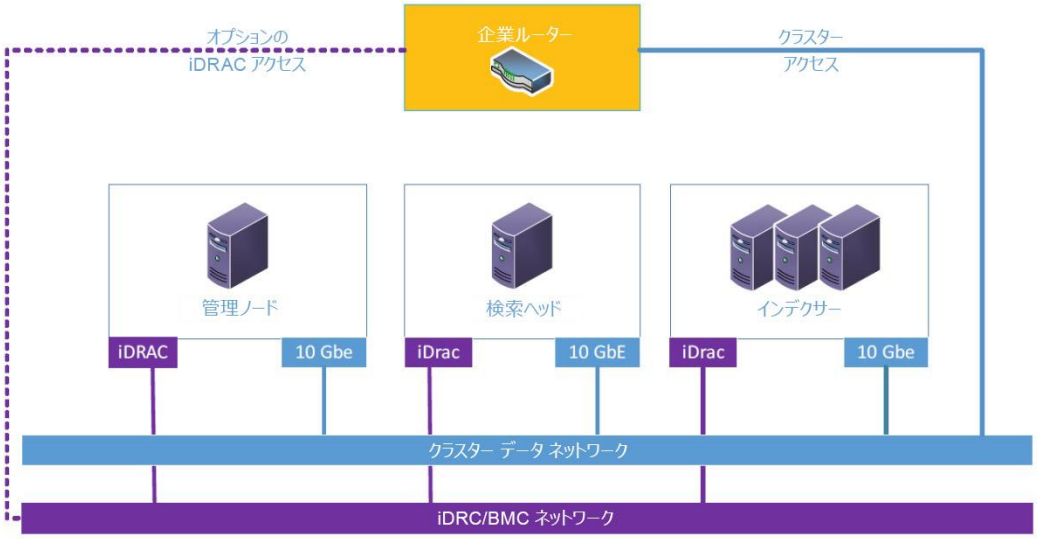

図 6　クラスター ネットワークの設計

10 GbE ToR スイッチ

各クラスターは、仮想リンク トランキング（VLT）機能を使用して高可用性を実現するように構成された、Dell Networking 10 GbE スイッチのペアを使用しています。VLT は、ボンディングされた NIC をサーバーが使用して、1 つではなく、2 つの異なるスイッチで LAG インターフェイスを終了できるようにします。このアーキテクチャは、スイッチに障害が発生した場合やメンテナンスが必要な場合に、ラックに冗長性を提供します。また、アクティブ/アクティブの帯域幅も使用できるようにします。

データ ネットワーク スイッチへのサーバー接続はボンディングされています。これらは IEEE 802.3 リンク アグリゲーション制御プロトコル（LACP）を使用して、ロード バランシング構成でアクティブ/アクティブ LAN アグリゲーション グループ（LAG）を使用します。Linux では、このボンディングは 802.3ad（またはモード 4）ボンディングと呼ばれます。接続が Pod スイッチのペアに対して行われ、ポート、ケーブル、またはスイッチに障害が発生した場合に冗長性を提供します。図 7 を参照してください。

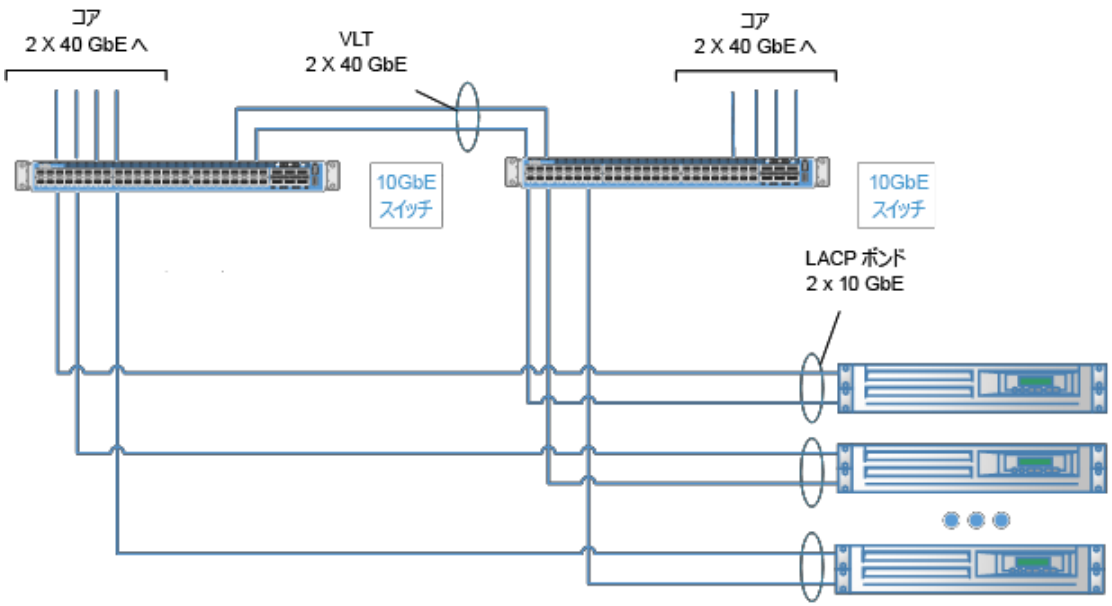

図 7　データ ネットワーク接続

ストレージ設計

1 日あたり 200 GB、250 GB、および 300 GB の構成には、以下が含まれます。

- 各インデクサー ノード上にあり、30 日間保存できるようにサイズが設定されているホットおよびウォーム バケット ストレージ
- 120 日または 365 日間保存できるようにサイズが設定されているコールド バケット ストレージ

Dell EMC Isilon スケールアウト ネットワーク接続型ストレージ（NAS）は、追加のコールド バケット ストレージおよびフローズン バケット ストレージに使用できます。

コールド バケット ストレージを追加するか、またはインデクサー ノードから分離するかを決定する際は、次の点を考慮してください。

- Splunk の保存期間は、取得量、検索期間、高可用性によって異なります。
- Splunk のコールド階層は、検索可能なデータ階層を提供します。この階層におけるデータ検索の性質と頻度により、ホットおよびウォームの階層のようなドライブのパフォーマンスは重要ではありません。
- 使用するオプションは、必要なコールド ストレージの量と、使用されているインデクサーの数によって決まります。
- Isilon を使用したコールド ストレージの分離により、適切な量のコールド ストレージを実現するためにインデクサーをさらに追加する必要はなくなります。
- ホットとウォームの保存ニーズが高まった場合は、コールド ストレージの分離を行うことが理にかなっています。コールド ドライブ ベイを再配置すると、ホットおよびウォーム階層用に追加の領域が解放されます。

Isilon の構成では、以下も推奨されます。

- すべての Isilon ノードで SmartPools 設定を有効にし、ランダム読み取りを高速化するために L3 キャッシュとして SSD を使用します。
- SmartConnect を有効にして、自動クライアント接続のロード バランシングとフェールオーバー機能を提供します。
- SmartCache を有効にしてライト パフォーマンスを向上させます。
- 同時データ アクセス パターンに最適化を使用します。
- 40 Gb/s の外部ネットワークを使用してデータを接続します。
- ネットワーク MTU を 9000（ジャンボ フレーム）に増加します。

Splunk および Dell EMC では、Isilon ストレージを含む NFS ストレージをホット/ウォーム データではなく、コールド データおよびフローズン データにのみ使用することを推奨しています。システム要件の詳細については、『Splunk Enterprise インストール マニュアル』を参照してください。

インフラストラクチャの検討事項

このリファレンス アーキテクチャで概説されているソリューションは、シングル インデクサー ノードに基づいています。このノードは、記載されているデータ量ですべてのインデックス作成とバックエンドの検索を実行します。ソリューションは、効率性を高めるためにこのように指定されています。クラスタリングを使用して複数のノードにインデックス作成作業を分散すると、ストレージ オーバーヘッドが大きくなりますが、容量も増加します。

複数の低コストのインデクサーとは対照的に、1 台の高コストのインデクサーを購入すると、ソリューション全体のコストを大幅に節約できる場合があります。運用コストはノード数が少ないほど低くなります。クラスタ化されたソリューションは高可用性をもたらすため、高いコストを正当化できる可能性があります。

より多くのコールド ストレージが必要な場合は、インデックス作成の速度や検索速度を高めるのではなく、Isilon のような NAS を追加すると、多くの場合、コスト効率が向上します。コールド ストレージには NAS ソリューションの使用が適しています。そのパフォーマンスの要件がホットおよびウォーム ストレージよりも低いためです。構成のサイジングの詳細については、Dell EMC 担当者にお問い合わせください。

構成

このセクションでは、各 PowerEdge サーバー構成に必要なハードウェアについて説明します。構成には以下が含まれます。

- リファレンス構成
- ミッドレンジ構成
- ハイ パフォーマンス構成

これらの構成は、Splunk のリファレンス ハードウェア ホストの仕様に基づいているか、それ以上のものです。以下の構成表は、リファレンス、ミッドレンジ、ハイ パフォーマンスの構成が、類似の 3 ノード アーキテクチャに基づいていることを示しています。このアーキテクチャでは、サーチ ヘッド、管理サーバー、インデクサーに個別のノードを使用します。インデクサー サーバーの仕様は、取得量とコールド バケット ストレージの保存量によって異なります。

一部の設計では複数のインデクサーを使用する場合があります。別の方法でサイズを設定する必要が生じる場合もあります。サイジングは、容量、レプリケーション、または導入オプション（分散、クラスタ化など）などの要因によって異なります。詳細情報または相談が必要な場合は、Dell EMC 担当者にお問い合わせください。連絡先情報については、第 5 章「リファレンス」を参照してください。

リファレンス構成

リファレンス構成には以下が含まれます。

- SSD を使用した 30 日間のホット/ウォーム保存
- 1 日あたり最大 200 GB のデータ取得と、10 K RPM SAS ドライブを使用した 120 または 365 日の保存

サーチ ヘッド

表 4　サーチ ヘッドの構成

PowerEdge R640 サーバー
• インテル Xeon Gold 6242 2.8 G x 2、16C/32T • 16 GB RAM x 6 • BOSS コントローラー カード + 2 M.2 スティック 480 GB（OS）RAID 1 • 480 GB SSD SAS Mixed Use 12 Gbps RAID 1 x 2 • 1 GbE x 2、10 GbE x2 インテル X710 NIC • PowerEdge R640　x8 ドライブ

管理サーバー

表 5　管理サーバー構成

PowerEdge R640 サーバー
• インテル Xeon Gold 6226 2.7 G x 2、12C/24T • 8 GB RAM x 6 • BOSS コントローラー カード + 2 M.2 スティック 480 GB（OS）RAID 1 • 960 GB SSD SAS Mixed Use 12 Gbps RAID 1 x 2 • 1 GbE x 2、10 GbE x2 インテル X710 NIC • PowerEdge R640 x8 ドライブ

1 日あたり 200 GB のインデクサー、120 日間保存

表 6　1 日あたり 200 GB のインデクサー（120 日間保存構成）

PowerEdge R740xd サーバー
• インテル Xeon Gold 6234 3.3 G x 2、8C/16T • 16 GB RAM x 6 • 480 GB SSD SATA Mix Use 6 Gbps（OS）RAID 1 x 2 • 960 GB SSD SAS Mix Use 12 Gbps（ホット/ウォーム）RAID 6 x 6 • 1.8 TB、10 K RPM SAS 12 Gbps（コールド）RAID 6 x 9 • 1 GbE x 2、10 GbE x2 インテル X710 NIC • R740xd 24 2.5 インチ ドライブ ベイ シャーシ

1 日あたり 200 GB のインデクサー、365 日間保存

表 7　1 日あたり 200 GB のインデクサー（365 日間保存構成）

PowerEdge R740xd サーバー
• インテル Xeon Gold 6234 3.3 G x 2、8C/16T • 16 GB RAM x 6 • 480 GB SSD SATA Mix Use 6 Gbps（OS）RAID 1 x 2 • 960 GB SSD SAS Mix Use 12 Gbps（ホット/ウォーム）RAID 6 x 6 • 2.4 TB、10 K RPM SAS 12 Gbps（コールド）RAID 6 x 18 • 1 GbE x 2、10 GbE x2 インテル X710 NIC • R740xd 24 2.5 インチ ドライブ ベイ シャーシ

ミッドレンジ構成

ミッドレンジ構成には以下が含まれます。

- SSD を使用した 30 日間のホット/ウォーム保存
- 1 日あたり最大 250 GB のデータ取得と、10 K RPM SAS ドライブを使用した 120 または 365 日の保存

サーチ ヘッド

表 8　サーチ ヘッドの構成

PowerEdge R640 サーバー
• インテル Xeon Gold 6242 2.8 G x 2、16C/32T • 16 GB RAM x 6 • BOSS コントローラー カード + 2 M.2 スティック 480 GB（OS）RAID 1 • 480 GB SSD SAS Mixed Use 12 Gbps RAID 1 x 2 • 1 GbE x 2、10 GbE x2 インテル X710 NIC • PowerEdge R640 x8 ドライブ

管理サーバー

表 9　管理サーバー構成

PowerEdge R640 サーバー
• インテル Xeon Gold 6226 2.7 G x 2、12C/24T • 8 GB RAM x 6 • BOSS コントローラー カード + 2 M.2 スティック 480 GB（OS）RAID 1 • 960 GB SSD SAS Mixed Use 12 Gbps RAID 1 x 2 • 1 GbE x 2、10 GbE x2 インテル X710 NIC • PowerEdge R640 x8 ドライブ

1 日あたり 250 GB のインデクサー、120 日間保存

表 10　1 日あたり 250 GB のインデクサー（120 日間保存構成）

PowerEdge R740xd サーバー
• インテル Xeon Gold 6226 2.7 G x 2、12C/24T • 16 GB RAM x 6 • 480 GB SSD SATA Mix Use 6 Gbps（OS）RAID 1 x 2 • 960 GB SSD SAS Mix Use 12 Gbps（ホット/ウォーム）RAID 6 x 6 • 1.8 TB、10 K RPM SAS 12 Gbps（コールド）RAID 6 x 11 • 1 GbE x 2、10 GbE x2 インテル X710 NIC • R740xd 24 2.5 インチ ドライブ ベイ シャーシ

1 日あたり 250 GB のインデクサー、365 日間保存

表 11　1 日あたり 250 GB のインデクサー（365 日間保存構成）

PowerEdge R740xd サーバー
• インテル Xeon Gold 6226 2.7 G x 2、12C/24T • 16 GB RAM x 6 • 480 GB SSD SATA Mix Use 6 Gbps（OS）RAID 1 x 2 • 960 GB SSD SAS Mix Use 12 Gbps（ホット/ウォーム）RAID 6 x 6 • 2.4 TB、10 K RPM SAS 12 Gbps（コールド）RAID 6 x 18 • 2.4 TB、10 K RPM SAS 12 Gbps（コールド）RAID 6（ミッドプレーン）x 4 • 1 GbE x 2、10 GbE x2 インテル X710 NIC • R740xd 24 2.5 インチ ドライブ ベイ シャーシ

ハイ パフォーマンス構成

ハイ パフォーマンス構成には以下が含まれます。

- SSD を使用した 30 日間のホット/ウォーム保存
- 1 日あたり最大 300 Gb のデータ取得と、10 K RPM SAS ドライブを使用した 120 または 365 日の保存

サーチ ヘッド

表 12　サーチ ヘッドの構成

PowerEdge R640 サーバー
• インテル Xeon Gold 6242 2.8 G x 2、16C/32T • 16 GB RAM x 6 • BOSS コントローラー カード + 2 M.2 スティック 480 GB（OS）RAID 1 • 480 GB SSD SAS Mixed Use 12 Gbps RAID 1 x 2 • 1 GbE x 2、10 GbE x2 インテル X710 NIC • PowerEdge R640 x8 ドライブ

管理サーバー

表 13　管理サーバー構成

PowerEdge R640 サーバー
• インテル Xeon Gold 6226 2.7 G x 2、12C/24T • 8 GB RAM x 6 • BOSS コントローラー カード + 2 M.2 スティック 480 GB（OS）RAID 1 • 960 GB SSD SAS Mixed Use 12 Gbps RAID 1 x 2 • 1 GbE x 2、10 GbE x2 インテル X710 NIC • PowerEdge R640 x8 ドライブ

1 日あたり 300 GB のインデクサー、120 日間保存

表 14　1 日あたり 300 GB のインデクサー（120 日間保存構成）

PowerEdge R740xd サーバー
• インテル Xeon Gold 6252 2.1 G x 2、24C/48T • 16 GB RAM x 12 • 480 GB SSD SATA Mix Use 6 Gbps（OS）RAID 1 x 2 • 1.92 TB、SSD SAS Mix Use 12 Gbps（ホット/ウォーム）RAID 6 x 5 • 1.2 TB、10 K RPM SAS 12 Gbps（コールド）RAID 6 x 18 • 1 GbE x 2、10 GbE x2 インテル X710 NIC • R740xd 24 2.5 インチ ドライブ ベイ シャーシ

1 日あたり 300 GB のインデクサー、365 日間保存

表 15　1 日あたり 300 GB のインデクサー（365 日間保存構成）

PowerEdge R740xd サーバー
• インテル Xeon Gold 6252 2.1G x 2、24C/48T • 16 GB RAM x 12 • 480 GB SSD SATA Mix Use 6 Gbps（OS）RAID 1 x 2 • 1.92 TB、SSD SAS Mix Use 12 Gbps（ホット/ウォーム）RAID 6 x 5 • 2.4 TB、10 K RPM SAS 12 Gbps（コールド）RAID 6 x 19 • 2.4 TB、10 K RPM SAS 12 Gbps（コールド）RAID 6（ミッドプレーン）x 4 • 1 GbE x 2、10 GbE x2 インテル X710 NIC • R740xd 24 2.5 インチ ドライブ ベイ シャーシ

Splunk Enterprise の導入

Splunk Enterprise の導入設計

Splunk Enterprise は、シングル インスタンス導入、分散導入、クラスタ化導入の 3 種類の導入をサポートしています。これらの導入オプションはすべて有効ですが、ここで説明する構成に最も類似しているのは分散導入です。

図 8 は、インデクサーと検索ヘッドを 1 つのサーバー ノードに統合した Splunk Enterprise シングル インスタンス導入を示しています。

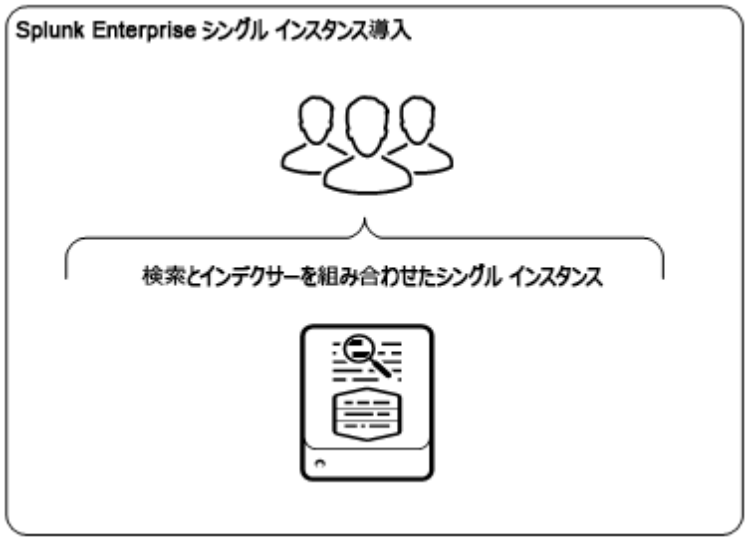

図 8　Splunk Enterprise シングル インスタンス導入

図 9 は、1 つのサーチ ヘッド、1 つのインデクサー、および 1 つのマスター ノード（管理サーバー）を使用した Splunk Enterprise 分散導入を示しています。ここでは、インデクサー データが一度保存され、複数のインデクサーがある場合は使用可能なインデクサー全体に分散されます。

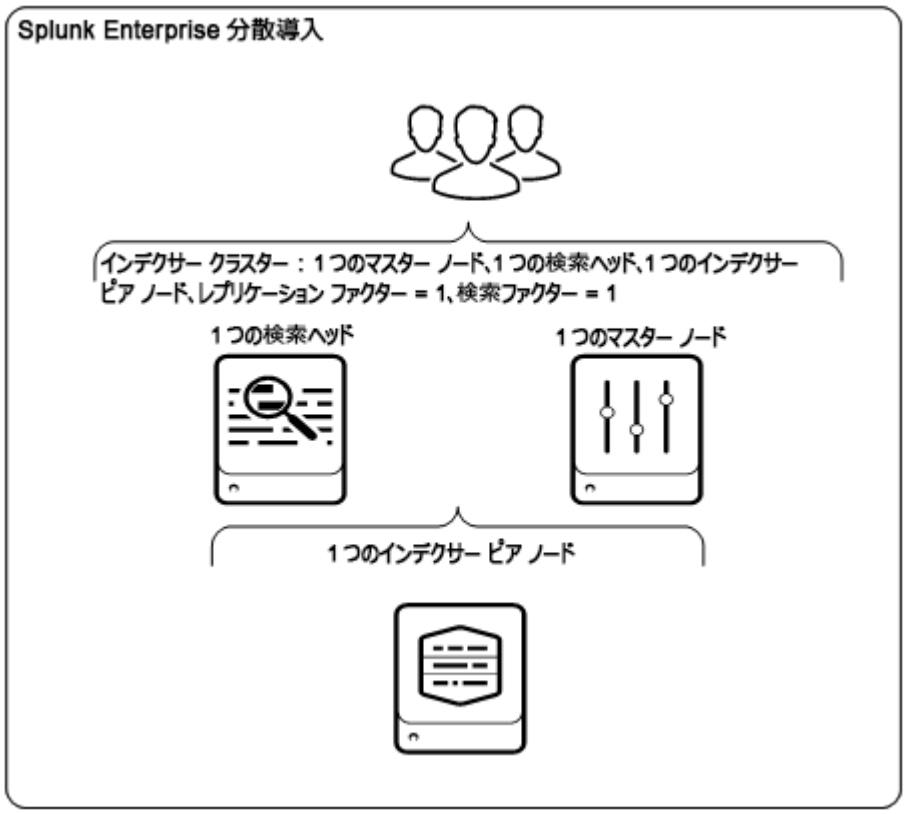

図 9 Splunk Enterprise 分散導入

図 10 は、1 つのサーチ ヘッド、2 つのインデクサー、1 つのマスター ノード（管理サーバー）を使用した Splunk Enterprise クラスタ化導入を示しています。このタイプの導入では、インデクサーは互いのデータを複製するように構成されているため、高いデータの可用性を必要とするクライアントがターゲットになっています。

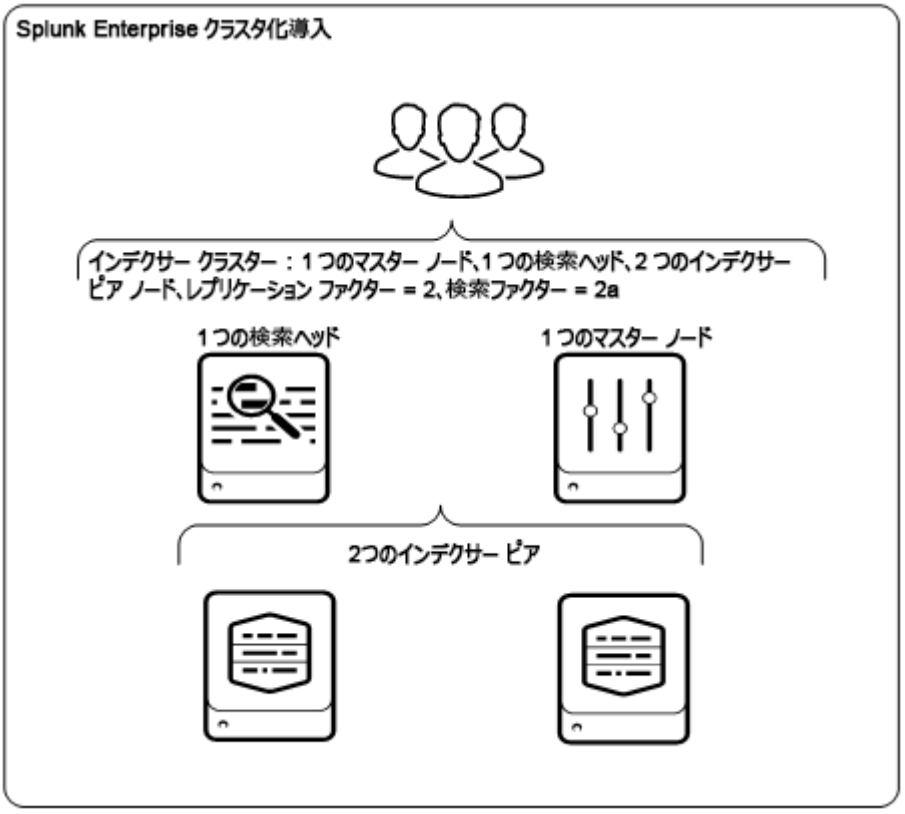

図 10 Splunk Enterprise クラスタ化導入

第 4 章　まとめ

本書では、Dell EMC とインテルのテクノロジーに基づいて構築された、Splunk 向けエンタープライズ インテリジェンス プラットフォームのビジネスレベルの概要を示しています。さまざまなビジネスの規模と種類の、現実的ないくつかの Splunk 導入に向けたリファレンス アーキテクチャ構成について説明しています。

本書のまとめ

Splunk は、組織のあらゆる部分にデータのインサイトをもたらすように設計されたエンタープライズ インテリジェンス プラットフォームです。Splunk Enterprise on Dell EMC Infrastructure のリファレンス アーキテクチャは、Splunk の導入に関して経験に基づいたガイダンスを提供します。このガイダンスにより、パフォーマンスを最適化し、導入までの時間を短縮し、IT 運用、およびサポート対象のビジネスのリスクを軽減することができます。

このリファレンス アーキテクチャが適用されるユース ケースは数多くあります。

- IT ビジネス運用
- セキュリティとコンプライアンス
- アプリケーション配信
- ビジネス分析
- IoT および産業データ

本書の目的：

- Splunk Enterprise の主要な概念とアーキテクチャについて説明する
- ソリューションの主要な Dell EMC コンポーネントの概要を説明する
- さまざまなビジネスの規模と種類のいくつかの現実的な Splunk の導入に関するリファレンス アーキテクチャ構成の詳細な推奨事項を提供する

本書では、類似しているサーバー、ネットワーキング、ストレージ コンポーネントを含む共通のアーキテクチャに基づいて、3 種類の異なる構成について詳しく説明しています。各構成には、オンボードのコールド データ ストレージ保存のための 2 つの異なるオプションがあります。構成には以下が含まれます。

- **リファレンス構成**：1 日あたり最大 200 GB のデータ取得、オンボード コールド ストレージでの 120 日または 365 日の保存
- **ミッドレンジ構成**：1 日あたり 200～250 GB のデータ取得、オンボード コールド ストレージでの 120 日または 365 日の保存
- **ハイ パフォーマンス構成**：1 日あたり最大 300 GB のデータ取得、オンボード コールド ストレージでの 120 日または 365 日の保存

このリファレンス アーキテクチャには、すべての構成にオンボード コールド バケット ストレージが含まれています。また、Dell EMC Isilon スケールアウト NAS をオフボードまたは追加のコールド バケット ストレージ、およびアーカイブ済みのコールド バケット ストレージに対して使用する方法とタイミングついても説明します。

Splunk 向けに最適化された Dell EMC とインテルのテクノロジーに基づいて構築されたエンタープライズ インテリジェンス プラットフォームにより、組織はデータ分析を迅速化および強化することができます。また、Splunk Enterprise の統合された柔軟な製品スイートのメリットを享受することができます。

情報の速度と品質は、データ分析ソリューションの最も重要な 2 つの要素ですが、このエンタープライズ インテリジェンス プラットフォームはこれら両方を提供します。Splunk Enterprise on Dell EMC Infrastructure のリファレンス アーキテクチャは、機能が豊富で拡張性に優れたハイ パフォーマンスのソリューションで、現在および将来のほとんどのニーズに合わせて拡張できます。

第 5 章　リファレンス

追加の情報は、「 Dell EMC InfoHub for Big Data Analytics(英語)」から入手できます。追加のサービスや実装のサポートが必要な場合は、Dell EMC のセールス担当者にご連絡ください。

Dell EMC ドキュメント

次の Dell EMC ドキュメントには、追加情報および関連情報が記載されています。これらのドキュメントにアクセスできるかどうかは、お使いのログイン認証情報によって決まります。アクセスできないドキュメントがある場合は、Dell EMC 担当者までお問い合わせください。

- Dell EMC PowerEdge R740xd スペック シート
- Dell EMC PowerEdge R640 の文書
- Dell EMC PowerEdge R740xd の文書
- Isilon サイト準備および計画ガイド

Splunk Enterprise ドキュメント

Splunk Web サイトにある次のドキュメントには、追加情報と関連情報が記載されています。

- Splunk Enterprise キャパシティ プランニング マニュアル
- Splunk Enterprise インストール マニュアル
- オンプレミスでの Splunk Enterprise 使用のシステム要件

Dell EMC Customer Solution Centers

専用の Dell EMC カスタマー ソリューション センターのグローバル ネットワークは、世界レベルの IT 専門家がお客様や、導入を検討されているお客様と協力して、ベスト プラクティスの共有、ブリーフィング、ワークショップ、または概念実証（PoC）を使用した効果的なビジネス戦略の掘り下げた話し合い、ビジネスの成功と競争力の向上を支援する、信頼できる環境です。Dell EMC カスタマー ソリューション センターは、新しいテクノロジーへの投資に関連するリスクを軽減し、実装のスピードを向上させるための支援を行います。

カスタマー ソリューション センターのすべてのサービスは、Dell Technologies のすべてのお客様に無償で提供されています。今すぐアカウント チームに連絡して、エンゲージメント リクエストを送信してください。

Dell Technologies Info Hub

Dell Technologies Info Hub(英語) には、Dell EMC ソリューションおよびネットワーキング製品に関する最新情報がすべて集められています。 新しい資料が絶えず追加されているため、頻繁にアクセスして、拡大し続ける最先端の製品やソリューションのポートフォリオの最新情報を常に入手してください。

その他の情報

サイジングのガイダンス、技術的な質問、または販売支援などの詳細については、splunk.ninjas@dell.com（英語対応）にお問い合わせいただくか、Dell EMC または認定パートナーのセールス担当者までお問い合わせください。

付録 A ソリューションのサイジング

この付録には、以下のトピックが含まれます。

概要

Splunkの実装は、特定の環境に合わせてサイジングする必要があります。この付録では、この文書で説明する構成を決定するために使用したアプローチと計算方法について説明します。これらに加えて、さらなるサイジングの計算を考慮する必要が生じることもあります。特定の設定に必要なハードウェアを特定するための主要な入力変数は次のとおりです。

- データ取得量と保存期間
- 1 つのノード ソリューションまたはクラスターを使用するかどうか

データ取得量

データ取得量は、インデックスを作成するデータの量を、Splunk が達成する予測圧縮比と掛け合わせたものです。通常、ソース データは 1 日あたりの GB として表されます。

圧縮比

圧縮比は、ソース データによって大きく異なる場合があります。ソース データの実際の圧縮比を判断する唯一の方法は、Splunk でインデックスを作成することによって、それを実証的に判断することです。使用する標準的な数値は、0.50、つまり 50%の縮小です。

保存期間

Splunk には、2 種類のデータ バケット保存期間があります。

- ホット/ウォーム
- コールド

ホット/ウォーム バケットは、ライブ受信データに使用されます。このデータには、最速の検索パフォーマンスが必要です。保存期間は通常、日数で表されます。このバケットは、新しい受信データが格納される場所です。

コールド バケットは、検索パフォーマンスを低下させても差し支えない、古いデータに使用されます。ここでのデータは、ホット/ウォームバケットから移動されたものです。

クラスター対シングル ノード

シングル インデクサー ノード ソリューションは、最も効率的に実行されます。重複するストレージを用意する必要はありません。マルチ インデクサー ノード クラスターは、追加のストレージを必要としますが、可用性と集約的パフォーマンスが向上します。

シングル ノード ストレージ

シングル ノード ストレージ構成に必要なストレージは次のように計算します。

[データ取得量] * [保存日数] * [圧縮比] = ストレージ

例えば、次のようなパラメーターがあるとします。

- 1 日あたり 100 GB のデータ取得量
- 0.5 の圧縮比
- 2 種類の保存日数：
  - 30 日間のホット/ウォーム保存
  - 180 日間のコールド保存

上記の計算式を適用すると、次の結果が得られます。

- 100 GB/日 * 0.5 圧縮比 * 30 日間のホット/ウォーム = 1500 GB
- 100 GB/日 * 0.5 圧縮比 * 180 日間のコールド = 9000 GB

クラスター ストレージ

Splunk クラスターには、必要なストレージのサイズを設定する際に考慮すべき追加の要素があります。ここではレプリケーション ファクター、つまりデータのコピーの数を選択します。N 個のレプリカには、N * 上記のストレージが必要であり、N-1 のノードに障害が発生するまでデータは失われません。レプリケーション ファクターが 2（オリジナル + 1 つのレプリカ）のクラスターをインストールする場合は、上記で計算したものの 2 倍のストレージが必要になります。

ソリューションのストレージ性能

Splunk では、受信データのインデックスを作成して検索を実行するにあたり、ある程度のパフォーマンスを必要とします。

インデックスの作成

入力データのインデックスの作成は、ストレージ レイヤーにとってそれほど過酷な操作ではありません。1 日全体で平均すると、日次データ取得量が大きくても、小さい数字になります。例：

1000 GB 取得/日 /（24 時間 * 60 分 * 60 秒）= 12.1 メガバイト/秒

ノートパソコンの SATA ドライブでも、その速度を維持することができます。検索をタイムリーに実行するには、高いパフォーマンスが求められます。Splunk では、検索時の平均パフォーマンスを維持するために、少なくとも 800 IOPS を推奨しています。IOPS が高いほど、検索のパフォーマンスは向上します。また Splunk では、NVME または SSD ドライブのいずれかにホット/ウォーム ストレージを配置して、最適な検索体験を実現することを推奨しています。表 16 は、一般的なドライブ タイプの IOPS パフォーマンスの比較を示しています。

表 16 ドライブの IOPS パフォーマンスの比較

ドライブ タイプ	IOPS	読み取り/書き込み
NVME	最大約 620 K	最大 3500 MB/s
SATA SSD	5 K～100 K	最大 520 MB/s
15 K HDD	188～203	91.5 MB/s～126.3 MB/s
10 K HDD	142～151	58.1 MB/s～107.2 MB/s
7.2 K HDD	73～79	43.4 MB/s～97.8 MB/s

詳細については、https://docs.splunk.com/images/6/67/Splunk-8.0.0-Capacity_ja-JP.pdf を参照してください。

Splunk インデクサーの例

このセクションでは、3 つのインデクサーのサイジング例を示します。

- 例 1：リファレンス
- 例 2：ミッドレンジ
- 例 3：ハイ パフォーマンス

例 1：リファレンス

この例は、1 日あたり 200 GB の取得、30 日間のホット/ウォーム保存、120 日または 365 日のコールド保存機能を備えたシングル ノード ソリューションです。

この構成で必要となるストレージは次のように計算します。

- 30 日間の保存（ホット/ウォーム）

  200 GB * 0.50 圧縮比 * 30 日 = 3000 GB

- 120 日間の保存（コールド）

  200 GB/日 * 0.50 圧縮比 * 120 日 = 12000 GB

- 365 日間の保存（コールド）

  200 GB * 0.50 圧縮比 * 365 日 = 36500 GB

ディスクの書き込みは次のとおりです。

200 GB /（24 日 * 60 時間 * 60 分）~ 24 メガバイト/秒

960 GB の SSD ディスクを使用している 30 日間のホット/ウォーム構成では、以下が必要です。

3000 GB / 960 GB = 3.125 ディスク（または 4 ディスク）

RAID 6 を実現するために 2 台のディスクを追加した場合は、6 台のディスクが必要になります。この構成では SSD ディスクを使用しているため、IOPS は最小要件を超えています。

RAID 6 構成で 1.8 TB 10 K RPM SAS ディスクを使用した 120 日の保存は、次のように計算されます。

12000 GB / 1800 GB ディスク = 6.67 ディスク（または 7 ディスク）

RAID 6 を実現するために 2 台のディスクを追加した場合は、9 台のディスクが必要になります。この構成の IOPS は約 1050 です。これは、800 IOPS の最小要件を上回ります。

7 ディスク * 150 IOPS = 1050 IOPS

RAID 6 構成で 2.4 TB 10 K SAS ディスクを使用した 365 日間保存のコールド ストレージでは、以下が得られます。

36500 GB / 2400 GB ディスク = 15.2 ディスク（または 16 ディスク）

RAID 6 を実現するための 2 台のディスクを追加した場合、18 台の 2.4 TB 10 K RPM のディスクが得られ、これは 800 IOPS の最小要件を上回ります。

18 ディスク * 150 IOPS = 2700 IOPS

例 2：ミッドレンジ

この例は、1 日あたり 250 GB の取得、30 日間のホット/ウォーム保存、120 日または 365 日のコールド保存機能を備えたシングル ノード インデクサーです。

この構成に必要となるストレージは次のように計算します。

- 30 日間の保存（ホット/ウォーム）

  250 GB * 0.50 圧縮比 * 30 日 = 3750 GB

- 120 日間の保存（コールド）

  250 GB * 0.50 圧縮比 * 120 日 = 15000 GB

- 365 日間の保存（コールド）：

  250 GB * 0.50 圧縮比 * 360 日 = 45625 GB

ディスクの書き込み速度は次のとおりです。

250 GB /（24 日 * 60 時間 * 60 分）~ 3 メガバイト/秒

960 GB の SSD ディスクを使用している 30 日間保存のホット/ウォーム構成では、以下が必要です。

3750 GB / 960 GB = 3.9 ディスク（または 4 ディスク）

RAID 6 を実現するために 2 台のディスクを追加した場合は、6 台のディスクが必要になります。この構成では SSD ディスクを使用しているため、IOPS は最小要件を超えています。

RAID 6 構成で 1.8 TB 10 K RPM SAS ディスクを使用した 120 日の保存は、次のように計算されます。

15000 GB / 1800 GB = 8.3 ディスク（または 9 ディスク）

RAID 6 を実現するために 2 台のディスクを追加した場合は、11 台のディスクが必要になります。この構成の IOPS は約 1350 です。これは、800 IOPS の最小要件を上回ります。

11 ディスク * 150 IOPS/ディスク = 1350 IOPS

RAID 6 構成で 2.4 TB 10 K SAS ディスクを使用した 365 日間保存のコールド ストレージでは、以下が得られます。

45625 GB / 2400 GB ディスク ＝ 19.01 ディスク（または 20 ディスク）

RAID 6 を実現するための 2 台のディスクを追加した場合、22 台の 2.4 TB 10 K RPM のディスクが得られ、これは 800 IOPS の最小要件を上回ります。

20 ディスク * 150 IPOS/ディスク = 3000 IOPS

例 3：ハイ パフォーマンス

この例は、1 日あたり 300 GB の取得、30 日間のホット/ウォーム保存、120 日または 365 日のコールド保存機能を備えたシングル ノード ソリューションです。

この構成に必要となるストレージは次のように計算します。

- 30 日間の保存（ホット/ウォーム）

  300 GB * 0.50 圧縮比 * 30 日 = 4500 GB

- 120 日間の保存（コールド）

  300 GB * 0.50 圧縮比 * 120 日 = 18000 GB

- 365 日間の保存（コールド）

  300 GB * 0.50 圧縮比 * 365 日 = 54750 GB

ディスクの書き込み速度は次のとおりです。

300 GB /（24 日 * 60 時間 * 60 分） = 3.55 メガバイト/秒

1.92 GB の SSD ディスクを使用している 30 日間のホット/ウォーム構成では、以下が必要です。

4500 GB /（1.92 * 1024）GB = 2.28 ディスク（または 3 ディスク）

RAID 6 を実現するために 2 台のディスクを追加した場合は、5 台のディスクが必要になります。この構成では SSD ディスクを使用しているため、IOPS は最小要件を超えています。

RAID 6 構成で 1.2 TB 10 K RPM SAS ディスクを使用した 120 日の保存は、次のように計算されます。

18000 GB / 1200 GB ディスク = 15 ディスク

ディスク ファイル システムにはわずかにオーバーヘッドがあるため、ディスク合計数を 16 に増加します。

RAID 6 を実現するために 2 台のディスクを追加した場合は、18 台のディスク、つまりそれぞれ 1.2 TB 10 K RPM が必要になります。

この構成の IOPS は約 2700 です。これは、800 IOPS の最小要件を上回ります。

18 ディスク * 150 IOPS/ディスク = 2700 IOPS

RAID 6 構成で 2.4 TB 10 K SAS ディスクを使用した 365 日間保存のコールド ストレージでは、以下のように計算されます。

54750 GB / 2400 GB = 22.8 ディスク（または 23 ディスク）

RAID 6 を実現するための 2 台のディスクを追加した場合、25 台の 2.4 TB 10 K RPM のディスクが得られ、これは 800 IOPS の最小要件を上回ります。

23 ディスク * 150 IOPS/ディスク = 3450 IOPS

ディスク スロットが制限されているため、このソリューションでは、365 日の完全な保存期間に対応できず、23 台の 2.4 TB ディスクを使用しています。

R740XD には、24 台を超えるディスクを使用できるようにミッドプレーン ディスク領域が必要です。
