Google Cloud PlatformへのSplunk® Enterpriseの導入

Splunk Enterpriseは、オペレーショナルインテリジェンスを取得できる、業界トップクラスのプラットフォームを提供します。Splunkのソフトウェアを利用することで、Webサイト、アプリケーション、サーバー、ネットワーク、センサー、モバイルデバイスなどのマシンから生成されるビッグデータを検索、監視、分析、可視化できます。13,000以上の組織がSplunkのソフトウェアを利用して、ビジネスとお客様に対する理解の向上、サイバーセキュリティリスクの緩和、サービスパフォーマンスの向上、そしてコストの削減を実現しています。Splunk Enterpriseでマシンデータをリアルタイムでインデックスすることで、システム管理者からビジネスアナリストまで組織内の様々な役割の人々が、生成された膨大なマシンデータから速やかにインサイトを得ることができます。

クラウド戦略を採用することで、ビジネスの機敏性を高め、コストの削減、市場投入までの時間短縮、イノベーションの促進が可能になります。Splunk Enterpriseはクラウド環境への導入に最適です。エンタープライズクラスの可用性と拡張性を備えており、オンプレミス、クラウド、ハイブリッド環境のワークロードから1日に数百TBのデータを収集できます。このドキュメントでは、Splunk EnterpriseをGoogle Cloud Platform (GCP)に導入する際のガイドラインを提供します。

Splunk導入環境のコンポーネント

標準的なSplunk導入環境には、Splunkフォワーダー、インデクサー、サーチヘッドという3つのコンポーネントが含まれています。主要なSplunkコンポーネントをすべて1つのインスタンスにまとめてインストールして実行することもできますが、それぞれを別々のインスタンスで実行することもできます。また、インデクサーとサーチヘッドを追加してスケールアウトすることで、ワークロードの増加に対応できます。

Splunkソフトウェアは、任意のハードウェア(物理、クラウド、仮想)上のオペレーションシステムに、わずか数分でインストールできます。パッケージはダウンロードで入手でき、ほとんどのオペレーティングシステムに対応しています。導入インフラに応じて、コンポーネントタイプごとに適切な数のリソースを割り当てる必要があります。

フォワーダーは、Splunk Enterpriseのインスタンス間でデータ転送を行うものです。インデクサーやほかのフォワーダー、サードパーティシステムが転送先となります。ほとんどのフォワーダーは少ないリソースで利用できるため、データを生成するマシン上に無理なくインストールできます。

インデクサーは、ストレージデバイスへのデータ書き込みと、そのデータの検索を行うものです。リソースを多用するため、I/OとCPUを十分に割り当てる必要があります。

サーチヘッドは、サーチ管理機能をつかさどるものです。検索リクエストを複数のインデクサーに振り分け、結果を結合してユーザーに返します。サーチヘッドには十分なCPUとメモリーを割り当てる必要があります。

検索とインデックスを行うために必要なシステムリソースと帯域幅は、インデックスするデータの総量と、任意の時点でのアクティブな同時実行数(スケジュール設定済み検索またはそれ以外)によって異なります。

インデクサーは、ディスクへのデータの迅速な書き込みに加え、ディスクからのデータの読み取り、圧縮ファイルの解凍、ナレッジの抽出、レポート作成など、検索実行に関連する作業の大部分を実行します。ワークロードのほとんどをインデクサーが占めるので、インデックスする量が増えた場合は、インデクサーのインスタンスも追加する必要があります。インデクサーを追加してスケールアウトすると、データ量増加による負荷を分散できるため、検索時のリソースの競合が軽減され、検索パフォーマンスが向上します。

一般的にGCPの環境では、ネットワークストリームとフォワーダーを組み合わせて、Splunkインデクサーにデータを送信します。ソースからデータを収集するのにフォワーダーは必須ではありませんが、フォワーダーを利用することで柔軟性、ロードバランシング、信頼性などのメリットが得られます。Splunkインデクサーにデータを取り込む場合には、データソースからのファイルシステムのマウントもよく使用されます。また、各種APIソースからデータを収集する場合は、モジュール入力(カスタムデータ入力を定義するSplunk Enterpriseの拡張機能)とHTTPイベントコレクタ(大量のデータをSplunkプラットフォームへセキュアに効率よく、直接送信するためのメカニズム)のどちらも使用できます。

その他にも、デプロイサーバー (設定の配布)、ライセンスマスター (ライセンス管理に使用)、マスターノード(インデックスレプリケーションを管理)といったコンポーネントがあり、以降ではこれらを管理ノードと総称します。

GCPにおけるパフォーマンスの考慮事項

SplunkソフトウェアをGoogle Cloud Platformに導入する際には、パフォーマンスについて考慮すべきことがいくつかあります。特に重要なのは、インスタンス(Compute)とストレージのタイプに関する事項です。

Compute：プリエンティブルインスタンスを使用すればコストを節約できます。しかし、Splunkは永続的なソフトウェアであり、常にデータの収集とインデックスを行うので、通常のインスタンスを使用することをお勧めします。VMインスタンスの最小推奨要件は以下のとおりです(VMクラス**n1-standard-4** に相当)：

- 4つのvCPU
- 15GBのRAM

Splunkソフトウェアはスケールアウトが可能なため、GCPでの使用に最適です。Splunkのインスタンスを追加することで、データ量の要件に応じてパフォーマンスやキャパシティを増強できます。さらに詳細な推奨サイズについては、以降の表を参照してください。

ストレージ：Splunkの構成、OS、インデックスしたデータの保存には、ルート永続ディスクの使用をお勧めします。クラスター環境では、ローカルSSDも選択肢になります。永続ディスクとローカルSSDの主な特徴は以下のとおりです。

- 永続ディスクは可用性、信頼性、耐久性が高く、容量を64TBまで増加できる
- 永続ディスクには、標準のハードディスク(PD-HDD)とSSDもしくはソリッドステートドライブ(PD-SDD)の2つの形態がある
- PD-SDDはホットおよびウォームデータ向きで、PD-HDDはコールドデータ向き
- ローカルSSDはVMインスタンスをホストするサーバーに接続され、1つのインスタンスにつき3TBという制限がある
- PD-HDDやPD-SSDよりもローカルSSDの方がはるかに高スループットで低レイテンシである
- ローカルSSDのインターフェイスはSCSIとNVMeの2種類

PD-SSDまたはPD-HDDを使用する場合、Splunkデータのバックアップには永続ディスクのスナップショットを検討します。インデックスのストレージ要件を計画する際には、Splunkソフトウェアによってデータが圧縮されることを考慮してください。標準的なインストール環境では、Rawデータとそれに関連するインデックスとメタデータを保存する場合の実効圧縮率は2:1となります。つまり、1日に100GBのインデックスを行う場合、1日で約50GBを使用することになります。PD-HDDやPD-SSD、ローカルSSDの数と容量は、保持期間の要件や、1日にインデックスする量の予想に基づいて決める必要があります。

導入のガイドラインと例

Splunkのワークロードにインスタンスをマッピングする際の一般的なガイドラインを以下の表に示します。このガイドラインに加えて、アーキテクチャとサイジングのベストプラクティスも考慮する必要があります。Splunkの負荷は、インデックスと検索の両方で構成されていることに留意してください。

表1：インデクサー

数	インスタンスタイプ	1日あたりの量(GB)
1	n1-standard-16	最大100
1	n1-standard-32	100 ～ 250

表2：サーチヘッド

数	インスタンスタイプ	同時使用ユーザー	パフォーマンス
1	n1-standard-16	最大100	良い
1	n1-standard-32	100 ～ 250	優れている

表3：デプロイサーバー、ライセンスマスター、クラスターマスター

数	インスタンスサイズ(タイプ)	パフォーマンス
1	n1-standard-8 n1-highcpu-8	良い
1	n1-highcpu-16 n1-highcpu-16	優れている

小規模の導入環境

以下の仕様は小規模の導入環境の例を示しています。最大100GB/日のインデックスが可能で、いつでも最大6件の同時検索に対応できます。インデックスする量が数GB/日の範囲では、このようなインスタンスがよく導入されます。

- 1 x サーチヘッド/インデクサー (インスタンスタイプ：n1-standard-16)
  - ストレージオプション1*：PD-SSD(ホット/ウォームおよびコールド用)
  - ストレージオプション2*：PD-SSD(ホット/ウォーム用)、PD-HDD(コールド用)
- N – ユニバーサルフォワーダー (データソース)

* ストレージはコストを考慮して選択。

この例では、1つのSplunkインスタンスでインデックスと検索を行うアーキテクチャを使用しています。このシステムへのデータ送信には、Splunkフォワーダー、ローカルファイル、リモートシステム、syslog、HTTPイベントコレクタ、モジュール入力などを使用できます。

用意するストレージの合計サイズは、保持期間の要件や、1日にインデックスする量の予想に基づいて決める必要があります。

中規模の導入環境

以下の仕様は中規模の導入環境の例を示しています。500GB/日のインデックスが可能で、8 ～ 16ユーザーの検索負荷に対応できます。

- 3 x インデクサー (インスタンスタイプ：n1-standard-32)
  - ストレージオプション1*：PD-SSD(ホット/ウォームおよびコールド用)
  - ストレージオプション2*：PD-SSD(ホット/ウォーム用)、PD-HDD(コールド用)
- 1 x サーチヘッド(インスタンスタイプ：n1-standard-32)
  - ストレージオプション：PD-SSD
- 1 x 管理ノード(インスタンスタイプ：n1-highcpu-8)
  - ライセンスマスター、DMCなど
- N – ユニバーサルフォワーダー (データソース)

* ストレージはコストを考慮して選択。

この例では、5つのSplunkインスタンスで構成される、オーソドックスな分散構成のアーキテクチャを使用しています。

このうち3つのインスタンスがインデクサー、1つがサーチヘッド、もう1つが管理ノードとして使用されます。用意するストレージの合計サイズは、保持期間の要件や、1日にインデックスする量の予想に基づいて決める必要があります。

大規模な導入環境

以下の仕様は大規模な導入環境の例を示しています。1TB/日のインデックスが可能で、16ユーザーの同時検索負荷に対応できます。先に述べたとおり、Splunkはスケールアウトが可能です。この構成のキャパシティやパフォーマンスを増強したい場合は、インデクサーやサーチヘッドを必要なだけ追加します。

- 5 x インデクサー (インスタンスタイプ：n1-standard-32)
  - ストレージオプション1*：PD-SSD(ホット/ウォームおよびコールド用)
  - ストレージオプション2*：PD-SSD(ホット/ウォーム用)、PD-HDD(コールド用)
- 1 x サーチヘッド(インスタンスタイプ：n1-standard-32)
  - ストレージオプション：PD-SSD
- 1 x 管理ノード(インスタンスタイプ：n1-highcpu-8)
  - ライセンスマスター、DMCなど
- N – ユニバーサルフォワーダー (データソース)

* ストレージはコストを考慮して選択。

この例では、サーチヘッドが1つ、インデクサーが5つあるアーキテクチャを使用しています。任意の数のSplunkフォワーダーを使用して、5つのインデクサーにデータを分散できます。用意するストレージの合計サイズは、保持期間の要件や、1日にインデックスする量の予想に基づいて決める必要があります。

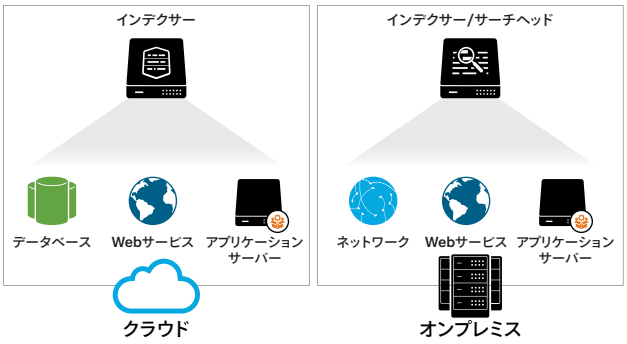
クラスターを使用した導入環境

以下の仕様は、インデックスレプリケーションを使用する大規模な導入環境の例を示しています。インデックスレプリケーションでは、インデクサーのデータのコピーを複数作成して管理するため、万が一Splunkインデクサーが停止したときでもデータのコピーをすぐに使用できます。この導入環境では1TB/日のインデックスが可能で、最大16ユーザーの同時検索負荷に対応できます。1つ前の例と同様に、インデクサーやサーチヘッドを適宜追加することで、パフォーマンスやキャパシティを増強できます。

- 8 ～ 10 x インデクサー (インスタンスタイプ：n1-standard-32)
  - ストレージオプション1*：ローカルSSD(ホット/ウォーム用)、PD-HDD(コールド用)
  - ストレージオプション2*：PD-SSD(ホット/ウォームおよびコールド用)
  - ストレージオプション3*：PD-SSD(ホット/ウォーム用)、PD-HDD(コールド用)
- 1 x サーチヘッド(インスタンスタイプ：n1-standard-32)
- 1 x 管理ノード(インスタンスタイプ：n1-highcpu-8)
  - マスターノード、ライセンスマスター、DMCなど
- N – ユニバーサルフォワーダー (データソース)

* ストレージはコストを考慮して選択。

この例では、Splunkインデクサーが5つ、Splunkサーチヘッドが1つあるアーキテクチャを使用しています。これらコンポーネントはすべて、レプリケーションとライセンス取得のために、クラスターとライセンスマネージャーのインスタンスと通信します。1つ前の例と同様に、サーチヘッドが5つのインデクサーすべてに検索を分散します。これはクラスターマスターからの情報に基づいて行います。保持期間を長くする場合、キャパシティを増強する場合、もしくはその両方を行う場合には、インデクサーやストレージを追加します。



ハイブリッド環境

上の図は、Splunk Enterpriseがオンプレミスとクラウドにインストールされたハイブリッド環境を示しています。Splunkソフトウェアの分散サーチ機能により、1つのインターフェイスから両方の環境を参照できます。

その他の考慮事項

- Splunkユニバーサルフォワーダーを使用すれば、既存のシステムからもデータを収集できます。
- デプロイサーバーを使用して、Splunk Appや設定ファイルを一元管理することが可能です。
- インデックスレプリケーション機能により、複数のSplunkシステム間でインデックスされたデータの可用性を高められます。ストレージによって可用性を確保する従来の方式と異なり、Splunkソフトウェアのレイヤーで可用性を管理します。

まとめ

Splunk EnterpriseをGoogle Cloud Platformに導入して最高のパフォーマンスを実現するためには、推奨されたインスタンスサイズおよびタイプを使用し、予想される1日あたりのボリューム要件に沿って計画を立てる必要があります。GCPはスケールアウトに優れているため、Splunkインスタンスを追加導入することで、キャパシティやパフォーマンスを増強できます。

GCPにてSplunkソリューションをぜひご利用ください

Splunk Enterprise：Splunk Enterpriseは無料でダウンロードできます。Splunk Enterpriseライセンスを60日間利用でき、1日あたり最大500MBのインデックスが可能です。60日間の終了後、もしくは終了を待たず任意のタイミングで、永久無料ライセンスに移行するか、Enterpriseライセンスを購入できます。https://www.splunk.com/ja_jp/talk-to-sales.htmlからSplunkへお問い合わせください。

Splunk Add-on for Google Cloud Platform：Splunk Add-on for Google Cloud Platformを使用すれば、Google Cloud Platform APIを使用してプラットフォームのイベント、ログ、メトリクス、請求データを収集できます。

splunk>

お問い合わせはこちら：https://www.splunk.com/ja_jp/talk-to-sales.html
〒100-0004 千代田区大手町1-1-1 大手町パークビルディング 8階

www.splunk.com/ja_jp
splunkjp@splunk.com

1st-2019-Splunk-SEC-DeployingSplunkGoogleCloud-A4-TB-202007