Amazon Web ServicesへのSplunk® Enterpriseの導入

Splunk Enterpriseは有意義な成果を生むためのインサイトをリアルタイムで取得できる、業界トップクラスのプラットフォームです。物理、仮想、クラウドなどの環境を問わず、ITシステムやテクノロジーインフラから生成されたマシンデータを取得し、解決策を導きます。

Splunkソフトウェアはマシン生成データをリアルタイムでインデックスします。そうすることで、サーバーチームからビジネスユーザーまで、組織のあらゆるレベルのユーザーが必要に応じてデータを詳細にドリルダウンしたり、強力な統計分析を利用したり、リアルタイムのダッシュボードやビューを表示したりすることができます。市販のハードウェアを増やすことでスケールアウトできるため、さまざまなソースから生成される膨大なデータから、速やかに相関関係を見い出すことができます。

ビジネスの機敏性を高め、コスト削減、市場投入までの時間短縮を目指すIT組織にとって、クラウド導入はオーソドックスな戦略です。実際に多くの組織が、クラウドの使用を前提とした環境導入ポリシーを掲げています。そのためSplunk Enterpriseはクラウドとハイブリッド環境への導入に力を入れています。ハイブリッド環境とは、オンプレミスとクラウドインフラの混在環境です。このような混在環境を使用する組織が、それまで得られなかった可視性を実現できます。このドキュメントでは、Splunk EnterpriseをAmazon Web Services (AWS)に導入する際のガイドラインについて説明します。

Splunk導入環境のコンポーネント

Splunk導入環境を構成する標準的なコンポーネントとして、Splunkフォワーダー、インデクサー、サーチヘッドがあります。Splunk Enterpriseは1つのパッケージにつき、いずれか1つのコンポーネントの役割を担うのが通常ですが、それに加えて複数の役割を担うこともできます。Splunk® Enterpriseは、AWSの任意のハードウェア(物理、クラウド、仮想)上のオペレーティングシステムにわずか数分でインストールして導入できます。パッケージはパブリックAMI (Amazon Machine Image)として提供されているほか、ほとんどのオペレーティングシステムに対応したダウンロード版も入手できます。主要なSplunkコンポーネントをすべて1つのクラウド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、それぞれを別々のクラウドインスタンスで実行することもできます。導入するインフラ環境に応じて、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、データ収集、データ転送、データのロードバランシングを行います。通常はデータの読み取りと送信にはわずかなオーバーヘッドしか発生しないため、少ないリソースで実行できます。ユニバーサルフォワーダーはSplunkソフトウェアに含まれている軽量なパッケージです。フォワーダー機能のすべてではありませんが、ほとんどを実行できます。

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。

サーチヘッドはインデクサーの情報を検索するものであり、十分なCPUとメモリを必要とします。

検索とインデックスのパフォーマンスを高める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。

EC2導入ではほとんどの場合、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では、マシンデータをさまざまな方法で簡単に取得できます。SplunkのHTTPイベントコレクタ(HEC)にデータを直接ポストする方法、APIでクエリを実行する方法、ファイルを監視する方法、ネットワークデータをリッスンする方法などがよく使われます。

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(ライセンスを管理)、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ます。

AWSにおけるパフォーマンスの考慮事項

SplunkソフトウェアをAmazon Web Services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。具体的には、AWS EC2インスタンスのサイズ、AWSストレージのタイプ、Amazon Machine Imageの選択です。

AWSインスタンス：スポットインスタンスやオンデマンドインスタンスは、使用しないときのコストを節約できます。ただし、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、リザーブドインスタンスを使用することをお勧めします。EC2インスタンスの最小推奨要件は以下のとおりです。

- 4つのvCPU
- 8GBのRAM

Splunkソフトウェアはスケールアウトが可能なため、AWS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、表2〜4を参照してください。

AWS S3を使用したSmartStore：Splunk導入環境では以下のメリットを得られるため、SmartStoreの使用を推奨します。

- ストレージコストを削減できます。導入環境において、高コストなローカルストレージの代わりに、低コストのS3互換のリモートオブジェクトストアを利用できます。
- リモートオブジェクトストアを介して、高可用性機能とデータ回復機能を利用できます。
- コンピューティングリソースとストレージリソースを別々に拡張できるため、リソースを効率的に使用できます。
- インデックスごとに、シンプルかつ柔軟に設定できます。

インデックスのストレージ要件を検討する際には、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および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10GBのインデックスを行う場合、1日でストレージを約5GB使用することになります。SmartStoreのキャッシュサイズには、検索回数が多かった数日分の平均値を反映する必要があります。SmartStoreの導入環境では、Splunkを最適化するために、以下の表1のように追加設定が必要です。

表1

server.conf	設定項目	S2の推奨値
バケツのサイズ	maxDataSize	auto
S2キャッシュのローカルディスク容量	minFreeSpace (MB)	5000
	eviction_padding (MB)	5120

indexes.conf	設定項目	S2の推奨値
ホットからウォームへの遷移またはデータアップロードの頻度	maxHotBuckets	3
	maxHotIdleSecs	0
	maxHotSpanSecs	777600
インデックスごとのキャッシュ設定	hotlist_recency_secs	86400
	hotlist_recency_filter_recency_hours	360

AWS AMI：Splunk Enterpriseは、Windowsや*NIXプラットフォームなど、広く使用されているオペレーティングシステムのほとんどで実行できます。サーチヘッドやインデクサーのOSには、64ビットアーキテクチャを強くお勧めします。AWS Marketplaceでは、Splunk Enterpriseを64ビットのLinux Amazon OSで実行するパブリックAMIが提供されています。

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。

i3インスタンスタイプではエフェメラルストレージを使用することに注意してください。そのため、i3インスタンスタイプを使用する場合には、クラスタリングする必要があります。

表2：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	c5.4xlarge	最大100
1	c5.9xlarge	最大300
1* (SmartStoreを使用する場合)	i3.4xlarge	最大100
1* (SmartStoreを使用する場合)	i3.8xlarge	最大300

表3：サーチヘッド

数	インスタンスサイズ (タイプ)	同時使用ユーザー	パフォーマンス
1	c5.4xlarge	最大8	良い
1	c5.9xlarge	最大16	優れている

表4：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	c5.2xlarge	良い
1	c5.4xlarge	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。インデックスの量が数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 – c5.4xlargeと、EBS-Backedのオペレーティングシステムと Splunk (SmartStoreを使用しない場合)
- N – ユニバーサルフォワーダー(データソース)

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、NFSマウントファイル、スクリプトによる呼び出し、モジュール入力を使用できます。EBSボリュームの数とサイズは、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 – c5.9xlargeと、EBS-Backedのオペレーティングシステムと Splunk (SmartStoreを使用しない場合) (インデクサー)
- 1 – c5.9xlargeとEBS-Backedストレージ(サーチヘッド)
- 1 – c5.2xlarge (ライセンスマスター)
- N – ユニバーサルフォワーダー(データソース)

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つがライセンスマスターとして使用されます。インデクサーのEBSボリュームの数とサイズは、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。

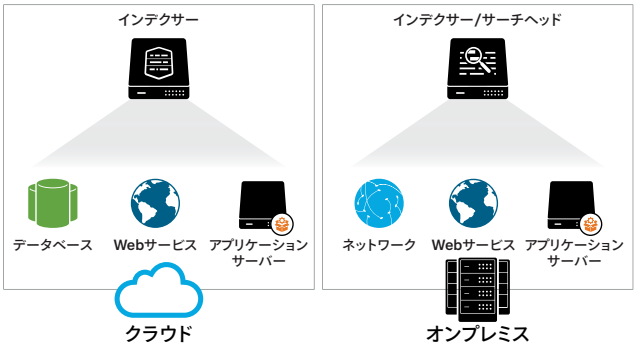
- 5 – c5.9xlargeと、EBS-Backedのオペレーティングシステムと Splunk (SmartStoreを使用しない場合) (インデクサー)
- 1 – c5.9xlargeとEBS-Backedストレージ(サーチヘッド)
- 1 – c5.2xlarge (ライセンスマスター)
- N – ユニバーサルフォワーダー(データソース)

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(N)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。EBSボリュームの数とサイズは、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、インデックスのバケツ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。この機能のメリットの1つは、各インスタンスでエフェメラルストレージを使えることです。これによってインデックス間で複製されたデータをSplunkで管理できるようになります。インデックス用にEBS-Backedストレージを使用しない場合にはこの方法を使用できます。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。

- 5 – i3.8xlarge (SmartStoreを使用する場合)(インデクサー)または
- 5 – c5.9xlargeと、EBS-Backedのオペレーティングシステムと Splunk (SmartStoreを使用しない場合)(インデクサー)
- 1 – c5.9xlargeとEBS-Backedストレージ(サーチヘッド)
- 1 – c5.4xlarge (ライセンスマスターとマスターノード)
- N – ユニバーサルフォワーダー(データソース)



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で両方の環境からインサイトを得られます。

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。
- Splunkデプロイサーバーを使用して、Splunk Appsや設定ファイルをSplunkインスタンスから一元管理し、伝播できます。
- 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。
- クラスターを使用しない環境を自動化するには、AWS CloudFormationのテンプレートをダウンロードしてご利用いただけます。

まとめ

Splunk EnterpriseをAmazon Web Services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。AWS EC2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。

AWS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 AMIをAWS Marketplaceで検索してください。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 Cloud：Splunk Enterpriseをサービスとして提供するSplunk Cloudにご登録いただけます。

Splunk App for AWS：Splunk App for AWSを利用して、AWS環境の運用の可視化とセキュリティを実現しましょう。

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

2019-Splunk-SEC-Deploying SE on AWS-A4-TB-JA-202007