Microsoft AzureへのSplunk Enterpriseの導入

Splunk®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。11,000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さまざまな役割の人々が、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、ビジネスの俊敏性を高め、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Microsoft Azureに導入する際のガイドラインを提供します。Microsoft Azureはオープンで柔軟性の高いクラウドプラットフォームで、分析、コンピューティング、データベース、モバイル、ネットワーク、ストレージ、Webなどの豊富なクラウドサービスが統合されており、機能拡充はさらに続いています。

Splunk導入環境のコンポーネント

Splunkの標準的な導入環境には、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。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。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。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、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。

一般的なAzure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)を使用できます。

その他にも、デプロイメントサーバー (FW設定一括管理)、ライセンスマスター (ライセンス管理)、マスターノード (データレプリケーションにおけるクラスタ構成管理)といったコンポーネントがあります。

Microsoft Azureにおけるパフォーマンスの考慮事項

SplunkソフトウェアをMicrosoft Azure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、Azure仮想マシン(VM)イメージおよびそのサイズと、それを支えるAzure Storageです。

Azure VMイメージ

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。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作成を行うので、予約インスタンスを使用することをお勧めします。

Azure VMのサイズ

Azure VMのサイズは、CPUコアの数、CPUの世代、利用可能なメモリー量、ネットワークの最大帯域幅、接続可能なデータディスクの数によって定義されています。Azure VMの最小推奨要件は以下のとおりです。

- 8個のCPUコア(コンピューティング最適化シリーズ)
- 14GBのRAM

Splunkソフトウェアはスケールアウトが可能であるため、Microsoft Azure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、使い方と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、以下を参照してください。

Azure Storage

Azure VMには、ローカル一時ディスクと仮想ハードディスク(ネットワークに接続された永続ディスク、VHDとも呼ばれる)の2種類のディスクがあります。各VMには、ローカルディスクが1つ、VHDのOSディスクが1つ付属し、さらにVHDのデータディスクを任意の数だけ追加できます。

Splunkのインデックスを保存する場合、一般的にローカルディスクは不向きです。ローカルディスクは一時データを保存するものであり、ハードウェア障害が起きた場合や、VMのサイズ変更または再起動の際にデータが失われる恐れがあるからです。

VHDはAzureのStandardまたはPremiumのストレージアカウントに保存されます。マネージドとアンマネージドのどちらでも使用できますが、Splunkのストレージにはマネージドを使用することをお勧めします。具体的には、Splunkアプリケーションと設定は永続OSディスクに保存し、Splunkのインデックスは複数の永続データディスクに保存します。

マネージドディスクを推奨する理由はいくつかあります。

- マネージドディスクの場合、ストレージアカウントを透過的に管理できます。アンマネージドディスクの場合、ストレージアカウントに含まれるすべてのディスクのIOPSが上限に近付いたときに、ストレージアカウントを追加で作成する必要があります。その上で、IOPSの制限に収まるよう、複数のストレージアカウントで仮想マシンディスクのバランスを再調整する必要があります。マネージドディスクでは、ストレージアカウントを追加作成することなく、こうしたIOPSの制限を効率的に回避できます。
- マネージドディスクの場合、可用性セットの信頼性を高めるため、単一障害点とならないようにディスクが十分に分離されています。そうすることで、可用性セット内のVMが同じストレージスケールユニットに保存されることを回避できます。そのため、可用性セット内の1つのVMがハードウェアまたはソフトウェアの障害でダウンしても、ほかのVMは影響を受けません。
- マネージドディスクとして設定したVHDは99.999%という優れた可用性を発揮するよう設計されています。

ストレージの各種制限についてはMicrosoft社のドキュメントを参照してください。

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。

小規模の導入環境

表1：インデクサー

インスタンスサイズ (タイプ)	1日あたりのインデックス量 (GB)	パフォーマンス
Standard_DS4_v2	最大100	良い
Standard_DS5_v2	100 ～ 500	優れている
Standard_DS15_v2	150 ～ 250	非常に優れている

表2：サーチヘッド

インスタンスサイズ (タイプ)	同時実行ユーザー	パフォーマンス
Standard_DS5_v2	最大8	良い
Standard_DS15_v2	最大16	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

インスタンスサイズ（タイプ）	パフォーマンス
Standard_DS3_v2	良い
Standard_DS4_v2	優れている

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、HTTPイベントコレクタ、ローカルファイル、NFSマウントファイル、SMBファイル共有、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。VHD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。

分散導入環境とAzure可用性セット

可用性セットとは、Azureで使用する論理グループです。この論理グループによって、Azureデータセンターに導入される際に、グループ内のVMリソースが互いに分離するよう配置されます。Azureでは、可用性セット内の複数のVMが、複数の物理サーバー、コンピューターラック、ストレージユニット、ネットワークスイッチをまたがって実行されます。1つの役割を担うために複数のVMを使用する場合には、その役割に可用性セットを使用することをお勧めします。たとえば、導入環境で複数のインデクサーを使用する場合は、それらのインデクサーを可用性セットに入れます。サーチヘッドクラスタリングを使用する場合は、それらのサーチヘッドを別の可用性セットに入れてください。

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ユーザーの同時検索負荷に対応できます。

- 5 – Standard_DS5_v2とVHD使用ストレージを1つの可用性セットで構成(インデクサー)
- 1 – Standard_DS5_v2とVHD使用ストレージ(サーチヘッド)
- 1 – Standard_D(S)3_v2 (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、6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5つのインスタンスがインデクサーとして、1つがサーチヘッドとして使用されます。この導入環境では、Splunkソフトウェアのスケールアウト可能な拡張性が活かされています。VHD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 – Standard_DS15_v2とVHD使用ストレージを1つの可用性セットで構成(インデクサー)
- 1 – Standard_DS15_v2とVHD使用ストレージ(サーチヘッド)
- 1 – Standard_D(S)3_v2 (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、1つのサーチヘッドが5つのSplunkインデクサーに検索を分散し、任意の数(N)のSplunkフォワーダーがこれらのインデクサーにデータを分散するアーキテクチャを使用しています。VHD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。

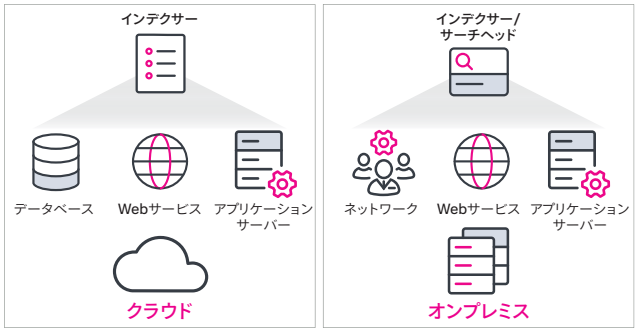
クラスターを使用した導入環境

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。この導入環境では500GB/日のインデックスが可能で、最大8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 – Standard_DS5_v2とVHD使用ストレージを1つの可用性セットで構成(インデクサー)
- 1 – Standard_DS5_v2とVHD使用ストレージ(サーチヘッド)
- 1 – Standard_D(S)3_v2 (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認証のために、クラスター及びライセンスマスターと通信します。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、インデクサーまたはストレージ、もしくは双方を追加します。

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。
- 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。ストレージによって可用性を確保する従来の方式(RAID構成のVHDなど)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。
- SSHやRDPのコンソールにアクセスするためのジャンプボックスとして、Azure VMをプロビジョニングすることを検討してください。ジャンプボックスのセキュリティ対策として、安全なパブリックIPアドレスからの接続のみを許可するNSGルールを追加します。

まとめ

Splunk EnterpriseをAzure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Azure VMサイズおよびAzure Storageボリュームを使用し、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。Azure VMとAzure Storage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。

Microsoft AzureにてSplunkソリューションをぜひご利用ください

Splunk Enterpriseは無料でダウンロードでき、Azure上の単体インスタンスまたは分散クラスターとして短時間で導入できます。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 Microsoft Cloud Services：**Splunk Add-on for Microsoft Cloud Services**を使用すれば、Office 365とAzureのさまざまなサービスを活用しながら、運用の可視性とセキュリティを確保できます。

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

© 2020 Splunk Inc. 無断複写・転載を禁じます。 Splunk, Splunk>, Data-to-Everything, D2EおよびTurn Data Into Doing は、米国およびその他の国におけるSplunk Inc.の商標または登録商標です。他のすべてのブランド名、製品名、もしくは商標は、それぞれの所有者に帰属します。

3rd-20-13235-Splunk-ES on Microsoft Azure-A4-TB-JA-202007