

SLA を満たすようにアプリケーションを設計する

100 XP

7 分

Tailwind Traders 社では、Special Orders アプリケーションに対して許容できる SLA が 99.9 パーセントであると判断しました。これにより、週あたり 10.1 分間のダウンタイムが推定されることを思い出してください。

次に、このアプリケーションの SLA を考慮しながら、Azure でこのアプリケーション用の効率的かつ信頼性の高いソリューションを設計する必要があります。必要な Azure 製品およびサービスを選択し、それらの要件に従ってクラウド リソースをプロビジョニングします。

現実では、障害が発生します。ハードウェアの障害が発生することがあります。ネットワークには間欠的なタイムアウト期間が発生する可能性があります。サービスまたはリージョン全体で中断が発生することはまれですが、このようなイベントについても計画しておく必要があります。

Tailwind Traders 社が使用するプロセスに従って、この会社によるテクノロジーの選択がアプリケーション SLA を満たしていることを確認してみましょう。

ワークロードを確認する

"ワークロード" とは、ビジネス ロジックおよびデータ ストレージの要件に関して、他のタスクから論理的に分離された個別の機能またはタスクです。各ワークロードでは、可用性、スケーラビリティ、データの一貫性、およびディザスター リカバリーに関する一連の要件が定義されます。

Azure 上で、Special Orders アプリケーションには以下が必要になります。

- 2 つの仮想マシン。
- Azure SQL Database の 1 つのインスタンス。
- Azure Load Balancer の 1 つのインスタンス。

基本的なアーキテクチャを示す図を次に示します。



Special Orders アプリケーション



複数の SLA を組み合わせて複合 SLA を計算する

Special Orders アプリケーションに含まれる個々のワークロードの SLA を確認すると、それらの SLA がまったく同じではないことに気付くかもしれません。これは 99.9 パーセントという全体的なアプリケーション SLA 要件にどのような影響を与えるのでしょうか。この問題を解決するには、いくつかの計算を行う必要があります。

複数の SLA を組み合わせるプロセスは、一連のサービスに対する "複合 SLA" を計算するのに役立ちます。複合 SLA を計算するには、個々のサービスの SLA を乗算する必要があります。

「サービス レベル アグリーメント」から、必要な各 Azure サービスの SLA を確認できます。これらは次のとおりです。

サービス	SLA
Azure Virtual Machines	99.9 パーセント
Azure SQL データベース	99.99 パーセント
Azure Load Balancer	99.99 パーセント

したがって、Special Orders アプリケーションの複合 SLA は次のようになります。

$$\begin{aligned} & 99.9\% \times 99.9\% \times 99.99\% \times 99.99\% \\ &= 0.999 \times 0.999 \times 0.9999 \times 0.9999 \\ &= 0.9978 \\ &= 99.78\% \end{aligned}$$

2 つの仮想マシンが必要であることを思い出してください。そのため、この数式には 99.9 パーセントという Virtual Machines の SLA を 2 回含めます。

個々のサービスにはすべてアプリケーションの SLA と同等以上の SLA が設定されているにもかかわらず、それらを組み合わせると、全体的な数値が必要な 99.9 パーセントより "低く" なることに注意してください。なぜですか? 複数のサービスを使用すると、新たな複雑さのレベルが追加され、エラーのリスクがわずかに増加するためです。

ここでは、99.78 パーセントの複合 SLA が、必要な 99.9 パーセントの SLA を満たしていないことがわかります。チームに戻り、これが許容できるかどうかを確認できます。または、その他の戦略を設計に実装して、この SLA を向上させることもできます。

複合 SLA がニーズを満たしていない場合の対処

Special Orders アプリケーションでは、複合 SLA が必要な 99.9 パーセントの SLA を満たしていません。Tailwind Traders 社で検討される可能性があるいくつかの戦略を見てみましょう。

必要な SLA に適したカスタマイズ オプションを選択する

以前に定義した各ワークロードには固有の SLA があり、各ワークロードのプロビジョニング時に選択したカスタマイズはその SLA に影響を与えます。次に例を示します。

- ディスク

Virtual Machines を使用する場合、Standard HDD マネージド ディスク、Standard SSD マネージド ディスク、または Premium SSD か Ultra Disk から選択できます。1 つの VM の SLA は、ディスクの選択に応じて、95 パーセント、99.5 パーセント、99.9 パーセントのいずれかになります。

- レベル

一部の Azure サービスは、Free レベルの製品と、Standard の有料サービスの両方として提供されています。たとえば、Azure Automation では、Azure 無料アカウントで 500 分のジョブランタイムが提供されていますが、SLA によるサポートはありません。Azure Automation の Standard レベルの SLA は 99.9 パーセントです。

購入内容を決定する際は、選択する Azure サービスの SLA への影響を必ず考慮してください。こうすることで、その SLA が必要なアプリケーション SLA を満たすようにすることができます。

この場合、Tailwind Traders 社では、仮想マシンに対して Ultra Disk オプションを選択することにより、アップタイムの向上を保証することができます。

可用性の要件を設計に組み込む

基になるクラウド インフラストラクチャに関連して使える、アプリケーション設計上の考慮事項があります。

たとえば、アプリケーションの可用性を向上させるには、単一障害点を回避します。そのため、さらに仮想マシンを追加する代わりに、同じ仮想マシンの 1 つ以上の追加インスタンスを、同じ Azure リージョン内の異なる可用性ゾーンにわたってデプロイできます。

"可用性ゾーン" は、Azure リージョン内の一意的な物理的な場所です。それぞれのゾーンは、独立した電源、冷却手段、ネットワークを備えた 1 つまたは複数のデータセンターで構成されています。各ゾーンでは異なるスケジュールが使用されます。そのため、1 つのゾーンが影響を受けても、他のゾーン内にある仮想マシン インスタンスは影響を受けません。

Azure 仮想マシンの複数のインスタンスを複数の可用性ゾーンにデプロイする場合、仮想マシンの SLA は 99.99 パーセントに向上します。この Virtual Machines SLA を使用して前述の複合 SLA を再計算すると、次のアプリケーション SLA が得られます。

$$\begin{aligned} &99.99\% \times 99.99\% \times 99.99\% \times 99.99\% \\ &= 99.96\% \end{aligned}$$

この 99.96 パーセントという修正された SLA は、99.9 パーセントという目標を上回っています。

Virtual Machines の SLA について詳しくは、「仮想マシンの SLA」を参照してください。

冗長性を追加して可用性を高める

高可用性を確保するために、アプリケーションに複数のリージョンにわたって重複するコンポーネントを持たせることを計画する場合があります。これは "冗長性" と呼ばれます。逆に、重要ではない期間中のコストを最小限に抑えるために、アプリケーションを 1 つのリージョンでのみ実行することもできます。Tailwind Traders 社では、特定の月や季節の間に特別な注文の割合が非常に高くなる傾向がある場合、これを考慮する必要があります。

アプリケーションの可用性を最大限にするには、アプリケーションのすべての部分に冗長性を追加します。この冗長性には、アプリケーション自体に加えて、基になるサービスとインフラストラクチャも含まれます。ただし、その実行は困難でコストがかかる可能性があり、多くの場合、必要以上に複雑なソリューションになることに注意してください。

冗長性を追加する前に、要件にとって高可用性がどの程度重要であるかを検討してください。アプリケーション SLA を満たすより簡単な方法が存在する場合があります。

非常に高いパフォーマンスの実現は困難

99.99 パーセントを超えるパフォーマンス目標を実現するのは非常に困難です。99.99 パーセントの SLA は、週あたり 1 分のダウンタイムを意味します。人間が、99.99 パーセントを上回る SLA のパフォーマンス目標を満たすほどすばやく障害に対応するのは困難です。代わりに、アプリケーションが自己診断を行い、障害発生時に自己回復を行えるようにする必要があります。