



フラッシュ - データベースを超えて広がる用途

著者: *George Crump*、リード・アナリスト
作成日: 2016 年 8 月

半導体メモリーを搭載したストレージ・システム、すなわちオールフラッシュ・ストレージ・システムと呼ばれるものは、IT でデータベース環境の高速化を目的に、これまでよく使用されてきました。入出力サイズが小さく、トランザクションが高い頻度で行われるデータベースの処理特性は、フラッシュの属性にぴったり合致しています。しかし、オールフラッシュ・ストレージ・システムの大幅な価格低下や、データを効率的に扱うテクノロジーとの組み合わせにより、仮想化のような他のワークロードにも、オールフラッシュ・ストレージ・システムが適合するようになりました。ただし、最近では、フラッシュ・コストの低下傾向や、フラッシュ・メディアの密度向上のおかげで、フラッシュはファイル・ベースまたはオブジェクト・ベースのワークロードでも真の価値を発揮しています。

ファイル・ベース・データの状況

ファイル・システム、特にオブジェクト・ストレージ・システムは、多くの場合、データセンターにおけるデジタル・データの蓄積場所と考えられています。しかし、ファイル・ストレージ・システムの利用方法は、ビッグデータの出現により近年大幅に変化しました。多くの企業では、モノのインターネット (IoT) デバイスが生成するデータを分析するために、Hadoop、SAS、Spark、または Splunk のようなプロジェクトが進行中です。さらに、一部の企業では、従業員または顧客にメディア・ストリーミングを提供する必要があります。最終的に、フラッシュは従来のアプリケーション群に広く採用され、企業は、今日主流のハードディスクやテープより高速なデータ保護戦略を実行する必要性を感じています。

従来、これらの セカンダリー・ストレージ・システムはハードディスクを搭載していましたが、オールフラッシュ・ストレージへの移行も検討する機会が訪れています。

オールフラッシュによるセカンダリー・ストレージの実現性

セカンダリー・ストレージ・システムは、ビッグデータ、DevOps プロジェクトや従来のファイル・ストレージに代表される、主にファイル・ベースのワークロードを保管するように設計されています。セカンダリー・ストレージは「重要ではない」という意味ではないことを理解することが大切です。セカンダリー・ストレージは、多くの企業にとって不可欠ですが、プライマリー・ストレージのワークロードに必要なパフォーマンスや可用性を必要としません。ただし、ファイル・ストレージの利用方法が増え続けるのと同じように、パフォーマンス向上に対する要求も増え続けます。

このような、従来はデータがセカンダリー・ストレージに置かれていたワークロードに対して、フラッシュを検討する主な理由が 2 つあります。最初の理由は、これらのワークロードに、フラッシュ・ストレージ・システムならではのパフォーマンスに対する確かなニーズがあることです。

Hadoop、Spark および Splunk のような環境では、高いパフォーマンスのメリットが直ちに役立ちます。これらの環境では、データ分析が高速かつ頻繁であるほど、意思決定のスピードが改善されます。高いパフォーマンスを発揮するセカンダリー・ストレージに対するもう 1 つの重要な要素は、個人情報への対応です。企業は、顧客が使用するアプリケーションが、カスタマイズによりユーザーが最も関心があるコンテンツが提示されるという顧客体験を通じて、顧客とのつながりが強化されることを認識しています。

したがって、フラッシュ・システムのメリットの大部分が待ち時間の短縮や非常に高い IOPS に関連している一方で、セカンダリー・ストレージ市場に提供されるフラッシュ・システムは、途方もない帯域幅も提供する必要があります。これは、大量の負荷がかかる状況やデータ保護のために有益です。

フラッシュを検討する 2 番目の理由は密度です。正しく設計されたフラッシュ・ストレージがラックあたりに提供できる容量は、データ効率化を行わない場合であっても、標準的なハードディスク搭載ストレージ・システムよりも、はるかに多くなります。フラッシュ・ストレージ・システムの購入コストが増える可能性がある一方、データセンターの設置スペースのコスト、消費電力の削減、冷却能力の節約、運用を行う人材と運用費の削減を考慮に入れると、フラッシュ・ストレージ・システムは (特に密度が高い場合)、ハードディスクを搭載したビッグデータ向けのシステムよりも安価になります。

オールフラッシュによるセカンダリー・ストレージ・システムの要件

非常に高い密度

セカンダリー・ストレージ・オールフラッシュ・アレイの重要な要件は、非常に高密度であることです。IT の専門家は、3U で 1/2 ペタバイトの容量を期待します。これは、数十もの回転ドライブによる熱と振動を心配しなければならないハードディスク搭載システムとは異なり、フラッシュ・モジュールを非常に緊密にストレージ・エンクロージャーに詰め込むことができますからです。さらに、フラッシュの筐体は小さく、ハードディスクの機械部分によって囲い込む必要がありません。これらの部分の密度も、セカンダリー・ストレージ層の電力供給と冷却のコストを削減し、関連したサーバー・ファームのコストにプラスの影響を与えます。

信頼性と保護

セカンダリー・オールフラッシュ・システムは、オールフラッシュ・アレイと同じパフォーマンスと高可用性の要件を満たす必要はありません。ただし、従来のハードディスクを搭載したセカンダリー・ストレージ・システムよりもパフォーマンスを改善する必要があります。同時に、プライマリー・ストレージのオールフラッシュ・アレイによる 24 時間 365 日稼働に対応する必要がない一方で、高い信頼性と高いデータ保全性を提供する必要があります。

ソフトウェアからの独立

一般的なアプリケーションのタイプがわかっているプライマリー・ストレージのワークロードとは異なり、セカンダリー・ストレージ・システムはさまざまなワークロードとアプリケーションをサポートする必要があります。Spark、Splunk、Hadoop、Couchbase および Cassandra などのセカンダリー・ストレージ・システムでデータにアクセスする一般的なアプリケーションの大部分は、5 年前には存在すらしていませんでした。これらのプラットフォームで一層のイノベーションを期待し、今後数年間にさらに多くの新しいアプリケーションが登場すると予想するのは自然なことです。

その結果、ストレージ・ハードウェアはソフトウェアに対して独立でなければなりません。つまり、さまざまなストレージ・ソフトウェアやファイル・システムがそのハードウェアを利用できなければなりません。ストレージ・ソフトウェアは、NFS のような既存プロトコルと、Hadoop File System (HDFS) や Object Storage のような新規プロトコルの両方を提示できるように適応する必要があります。

また、ソフトウェアは、データ・コピーがコンポーネント障害後も存続できるようにし、今日書き込まれたデータが、数年後にも同じように読み取れるようにすることによって、データ保護とデータ保全性の責任の一部をストレージ・システムに依存しないようにする必要があります。

コスト効率

最後に、これは セカンダリー・ストレージ層に対する見解であり、近年はこの層のパフォーマンスの重要性がますます高まっている一方で、ハードディスクを搭載したストレージと比較されています。密度と電力コストの節約、および大幅なパフォーマンス向上の効果を考えれば、オールフラッシュ・ストレージ・システムがディスクの価格を下回る必要はありません。ただし、近づく必要はあります。IT 専門家は、GB あたり 1 ドル未満でこれらのシステムを購入することを期待しています。

余計な荷物になる「付属」ソフトウェアをシステムから切り離すことで、さらなるコスト節約が可能です。セカンダリー・データ市場は非常に流動的なので、組み込まれているどのソフトウェアも数年後には使用できなくなる可能性があることを覚えておいてください。代わりに、ストレージ・ベンダーは、ソフトウェアと競合するのではなく、ソフトウェアを補完するように設計された、高品質、高密度のストレージ・システムの提供を重視する必要があります。

まとめ

セカンダリー・ストレージ・システムをオールフラッシュへと移行する機会が訪れました。セカンダリー・ストレージ・システムは、従来のアプリケーションが引き出せるだけのメリットを、フラッシュのパフォーマンスから得ることができます。データを即時に分析する機能は、途方もないメリットを企業にもたらします。ただし、この新しいフラッシュ・プラットフォームは従来とは異なるものでなければなりません。対応するプライマリー・ストレージよりも密度が高くなければなりません。密度が高ければ、データセンターの設置スペースや電力コストが削減されるだけでなく、データをはるかに詳しく調査できます。これは、ハードディスク搭載ストレージの場合よりも、様々な切り口での分析を、より短い時間で実行できるためです。最後に、セカンダリー層は引き続き進化し、人間や機械が生成するデータに対応する必要があります。このプラットフォームには、ソフトウェア主導のアプローチが最適です。

本書の制作は、**IBM** の支援金により行われました。本書は **IBM** を含むさまざまなベンダーの公開資料を使用していますが、本書で扱う課題に関してそのベンダーの立場を反映しているとは限りません。