

Relationship Builder

リレーションシップビルダー

非構造化データ時代に
ストレージの選び方を再考する

年々増え続けるITサービスの中で、飛び交うデータ量は明らかに右肩上がりに急増しており、5年、10年後は想像もできない時代を迎えた。そんな現状を鑑み、ここでは、データを格納するストレージの選び方に着目する。



株式会社NTTデータ
基盤システム事業本部 システム方式技術ビジネスユニット
第三技術統括部 第三技術担当
課長代理 榎本 圭氏

急増する「非構造化データ」

10年前と今のITサービスを比べたときに、やり取りされるデータ量が急増していることは疑う余地のない事実だ。増加するデータの内訳を考えると、「構造化データ」「非構造化データ」という分類がある。前者は、簡単に言えば、長さが決まっているデータであり、例えばDBに格納されるユーザ名だ。一方後者は、長さが不定なデータであり、画像や動画、オフィス文書が該当する。両者のうち、今後も増え続けるのは非構造化データであると言われている。

ITサービスに登録するユーザ数が増えても、ユーザ名だけでデータ量はさほど増えない。他方で、スマートフォンの普及により動画配信サービスなどが増えているが、動画一つ当たりのデータ量が大きいことを考えると、全体として増え続けるのは非構造化データであろう。

データを格納するストレージの二極化

この状況においてストレージ技術も二極化している。一方は高速性を追求する場合だ。例えば、全英オープンテニスのウェブサイトは、ライブストリーミングや画像を駆使して情報配信するが、試合数や時間が決まっており、データ量の見積りが比較的容易である。反面、ピーク時は世界中から一日数千万アクセスが集中するため、ストレージには厳しい性能要件が想定される。

もう一方で、拡張性やコストを追求する場合がある。例えばFlickrやSlideShareのように、写真や発表資料を預かり公開するサービスを考えると、利用者から大量のアップロードがある。サービス提供者からすれば、事前の容量見積りが困難であるため、不足したら都度拡張する設計が望ましいし、ストレージ自身のコストも抑えたい。そのため拡張性やコストが求められる。

拡張性重視の
クラウドストレージ

このように、拡張性やコスト重視の要求を受けて、クラウドストレージと呼ばれる技術は生まれた。多く

が図のような構成をとり、専用ストレージ機器を使う代わりに、汎用サーバを寄せ集めて一つの大きなストレージに見せる技術だ。利用者からは、一つのストレージに見えるが、内部ではデータは汎用サーバ群に分散して配置される。この方法では、専用ストレージ機器よりコストが抑えられ、汎用サーバを増やせば容量も増加する。また一般に、ペタバイト以上にまで拡張できる設計となっている。具体的な例を挙げれば、パブリックオンラインストレージサービスとしてAmazon S3、自社にクラウドストレージを構築する場合はOpenStack Swiftがある。

前者はパブリックサービスを利用して運用コスト削減したい場合に適し、後者は、機密の問題から自社保管したい場合に適する。いずれにしても、拡張性やコストを求める場合には、クラウドストレージは一考すべき選択の一つといえよう。

●お問い合わせ先●
株式会社NTTデータ
広報部
TEL：03-5546-8051

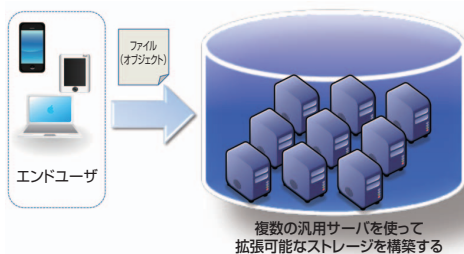


図 クラウドストレージの構成イメージ