



特集

クラウドを支える データストレージ技術



編集にあたって

横田治夫●東京工業大学, 天笠俊之●筑波大学

1024 ワードから始まった

コンピュータは「計算機」と呼ばれることも多いが、すでに十分に認識されているように、単に計算するだけの装置ではなく、さまざまな機能を提供するもので、コンピュータによる情報処理機能は今や生活に欠かせない基盤の1つとなっている。単純な数値計算能力を超える大きな要因の1つは、コンピュ

タが情報を記憶できることによる。それは黎明期のコンピュータである EDSAC, EDVAC の頃から変わらない。大きく変わったのは、その容量と性能と、さらにネットワーク環境である。

EDSAC, EDVAC が開発された 1940 年代後半では、主記憶装置として、超音波遅延線を用いていた。EDSAC の主記憶容量は、18 ビットを 1 ワードとして 1024 ワードであった。その後、いく



つかの変遷を経て、不揮発性記憶装置として磁気ハードディスク (HDD) が登場することになるが、1956年に製造された最初のHDDであるIBM 305 RAMAC (Random Access Method of Accounting and Control) システムは、直径24インチ(610mm)のアルミ板50枚を使い、5MB (5×10^6 B)であったという。今や1TB (1×10^{12} B)を超えるコンパクトなHDDを安価に購入することができ、その記憶容量はRAMACの200,000倍以上となっている。さらに、ネットワークを介して、より大容量のデータストレージに容易にアクセスすることもできるようになり、そのアクセス性能も格段に改善されている。

このような記憶容量の増加や利用形態の変化、さらにはアクセス性能の向上は、関連する情報処理の技術だけでなく、我々の日々の生活にもさまざまな影響を与えている。特に昨今、情報処理の基盤としてクラウドが注目されているが、クラウドにおけるストレージの役割は非常に大きい。本特集では、種々の面から、クラウドを支えるデータストレージ技術を紹介する。

本特集で取り上げる内容

* ストレージサービスとその標準化

クラウドはネットワークを介して多様な情報処理資源をサービスとして利用する概念であり、提供するサービスの内容により異なった形態をとる。その中で、クラウドが広く認知されるようになったのは、AmazonのS3 (Simple Storage Service) に代表されるストレージサービスが最初と言ってもよいかもしれない。利用者は、情報をローカルに蓄積する代わりにクラウドストレージに蓄積することで、ネットワークを介してさまざまな場所からその情報を利用・共有することができるようになった。

そこで、まず「1. クラウドストレージ標準化の最新動向」では、そのようなストレージサービスを取り上げ、クラウドとしてストレージサービスを提供する場合の資源の操作や管理に関する標準化であるCDMI (Cloud Data Management Interface) を中心

とした標準化の最新動向について紹介する。

* 人間の記憶のサポート

クラウドのストレージサービスをさらに発展させていったときに、これまで同様にデータを蓄積するだけではなく、個人的に体験したことに関する記憶を忘れないように蓄積し、思い出させてくれる「記憶のストレージ」のサービスが生まれた。Evernoteは、クラウド上にパーソナルデータを集約し、いろいろなプラットフォームのクライアントからそれを引き出すことができるサービスを提供している。「2. 万人のためのクラウドデータベース、Evernote その仕組みと展望」では、Evernoteのユーザの視点からの機能と、サービス提供側から見た仕組み、さらにそのビジネスモデルと未来への展望を述べる。

* キーバリュ型データストアの実例

クラウドでサービスを提供するためには、その実現のための情報を蓄積する手段が重要である。クラウドの情報蓄積手段として、これまでのようにSQL問合せを受け付ける関係データベースを前提としたものだけでなく、NoSQLデータベースとも呼ばれるキーバリュ型のデータストアが注目されている。単純なキーとそれに関連する値のペアの参照・更新に特化することで、システムの拡張性を高めるアプローチをとっている。「3. ROMAの概要とその利用事例について」では、楽天トラベルの「最近見た宿泊施設一覧」等で実際に利用されているキーバリュ型のデータストアであるROMAの特徴と利用事例を紹介する。

* 分散データストレージとMapReduce

クラウドにおけるサービスでは、データを大量に蓄積すると同時に、蓄積された大規模なデータに対して高速な処理を行うことも重要である。さらに、規模拡大のためには、データストレージとそのデータに対する処理を分散化することが求められる。「4. データクラウド研究の潮流と最新動向」では、NoSQL系の分散データストレージの構成方法とそ



クラウドを支えるデータストレージ技術

の機能、および大規模データを対象とした分散処理系の代表的な技術である MapReduce とその高速化を中心に関連する研究動向を紹介する。

* クラウドストレージの省電力化

省電力化は、コスト削減のためだけでなく、CO₂の排出量を削減して地球環境を守るためにも取り組むべき重点項目であり、クラウドのデータセンタにおいてもきわめて重要な課題となっている。「5. クラウド時代を支えるグリーンデータセンタのストレージ技術動向」では、ディスクアレイから分散データストアまで、クラウドでデータを格納するシステムに焦点を絞り、グリーンなデータセンタを目指した省電力化技術を紹介する。

* クラウドとストレージネットワーク

クラウドのデータセンタにおけるストレージ装置は、ネットワークを介してサーバと接続される形態になる。そのようなストレージネットワークとして、機器の種類が豊富でコストパフォーマンスが良い汎用のネットワーク技術をストレージ接続に活かすことを前提に、イーサネットの TCP/IP をベースとする方法が注目されている。「6.iSCSI と FCoE によるストレージ構築～ストレージネットワークの進化～」では、TCP/IP によるサーバとストレージが接続する方式である iSCSI (Internet Small Computer System Interface) と FCoE (Fibre Channel over Ethernet) に関する最新の技術動向を紹介する。

* クラウドのプライバシー保護

クラウドストレージに蓄積された個人情報の利活用では、プライバシー保護は避けては通れない課題である。「7. クラウドストレージにおけるプライバシー保護とデータ利活用」では、入力プライバシーと出力プライバシーという2つの観点から、クラウドストレージに保管される個人情報の漏えいリスクをコントロールし、安全に利活用するためのさまざまな技術を紹介する。

データストレージのこれから

コンピュータシステムのデータストレージに情報を蓄積し、それを活用する枠組みは、現代の社会生活において、必須となっている。今後、クラウドに限らず、まったく新しい形態の処理モデル等が出てきたとしても、その枠組みが大きく変わることはないだろう。本特集で対象とした、ネットワークを介したクラウドにおいて情報を蓄積し、活用するためのアプローチと技術、さらにその課題を俯瞰することが、データストレージ技術のさらなる発展につながることを期待する。

最後に、多忙な中今回の執筆を引き受けていただいた著者の方々と、本特集を編集するにあたって貴重なご助言とご尽力を頂戴した東京大学喜連川優教授に、この場を借りて深く感謝申し上げます。

(平成 23 年 3 月 24 日)

