

NoSQL製品の実力を検証する

—データベースシステムの新しいトレンド—

クラウドコンピューティング（以下、クラウド）上でWebシステムを構成するケースが増えるなかで、柔軟な拡張性を持つNoSQLと呼ばれる新しいオープンソースのデータベースに注目が集まっている。本稿では、NoSQLをその特徴によって分類し、どういう用途に向くのかを整理した上で、NoSQLの機能や性能について紹介する。

NoSQL登場の背景

従来、一般的なWebシステムはWeb3階層と呼ばれる構成が主流である。これは、Webブラウザを用いたインタフェースを司るプレゼンテーション（PL）層、Webアプリケーションを実装するビジネスロジック（BL）層、データを格納するデータベース（DB）層の3階層に分けてシステムを構成することで、柔軟なシステム変更を可能にしたものである。

しかしこのアーキテクチャも、サービス規模の急拡大によって限界に近づいている。PL層、BL層はスケールアウト（サーバーの増設による拡張）が容易なのに対して、既存のRDBMS（リレーショナルデータベース管理システム）によって作られたDB層は基本的にスケールアウトが困難だからである。

RDBMSでも、複数のサーバーから同時にデータベースを操作できる「Oracle」のRAC構成のようにスケールアウト可能なものもあるが、現実問題としては数倍程度が限界である。RDBMSの前に分散キャッシュ（DBの内容を複数のアプリケーションサーバーに分散してキャッシュする仕組み）を配置する方法もあ

るが、それもクラウド規模ともなると同期遅延や運用の複雑さが実用的な範囲を超える。

このような既存データベースの問題から、Amazon社、Facebook社などをはじめとする大規模なサービス事業者が、拡張性の高いデータベースシステムを独自に開発して自社サービスに適用するようになり、これをオープンソースとして公開している。

これらの新しいデータベースはNoSQL（SQLに取って代わる）と呼ばれる。RDBMSがSQLという標準言語によってデータベースを操作するのに対して、新しいデータベースはSQLを用いないためである。しかし現状はRDBMSの得意でない部分で選択的に利用されていることから、“Not only SQL”という意味も持つと考えられるようになっている。

NoSQLの分類

それでは、NoSQLとは具体的にどのようなものだろうか。

NoSQLは、RDBMSの弱点を補完するものとして、広い意味での非関係モデルに属するデータストア（データ保存の仕組み）を利用したデータベース技術とされる。NoSQLとRDBMSの具体的な違いを比較したのが表1



表1 NoSQLとRDBMSの主な違い

	NoSQL	RDBMS
データモデル	キーバリュー型、ドキュメント型、グラフ型、表形式型など	リレーショナルデータモデル
スキーマ	柔軟に変更可能（スキーマレス）	固定的（変更しにくい）
データ一貫性	一時的に一貫性が厳密に維持されていない状態もある（BASE原理、結果整合）	一貫性が厳密に維持される（ACID特性）
拡張性	スケールアウトにより大量アクセスによるパフォーマンス低下が少ない	スケールアップが基本。データ一貫性の保持を厳密に行うためパフォーマンスの低下が目立つ
耐障害性	SPOF（単一障害点）がないものが多く、コストが低い	耐障害性を高めるためのコストが高い
データ操作言語	多様なデータアクセス言語	SQL
インタフェース	シンプルなアクセスインタフェース	多様なアクセス形態

である。NoSQLはRDBMSと比較して拡張性、耐障害性に優れる反面、データの一貫性に関しては多くは結果整合（一時的に一貫性が厳密に維持されていない状態もあること）に緩和されている。

NoSQL製品の種類は非常に多く、明確に分類できるわけではないが、データモデルの違いによって大まかに分類すると次ページの図1のようになる。単純キーバリュー型から始まり、アクセスインタフェースの拡張に伴って、カラム型、ドキュメント型が順次、派生してきた。図では併せて各モデルの特徴や利用シーンなどについてもまとめている。

単純キーバリュー型は、買い物サイトのショッピングカートのように更新・検索が1つのキーに限定されている場合に利用される。「キー」とは情報の組を一意に識別するための情報で、その情報の内容が「バリュー」である。カラム型はキーが1つまたは2つ程度に固定的である場合に利用される。ドキュメ

ント型は、完全に構造化されていないが類似したデータ構造を持つデータ（非構造化データ）の更新・検索に適している。

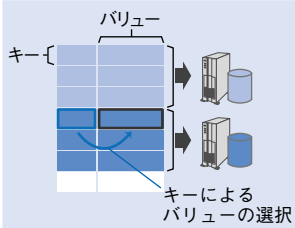
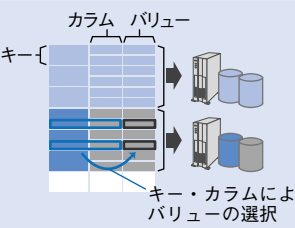
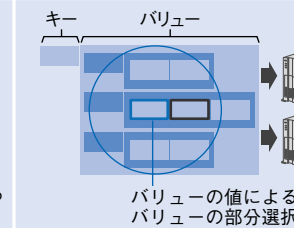
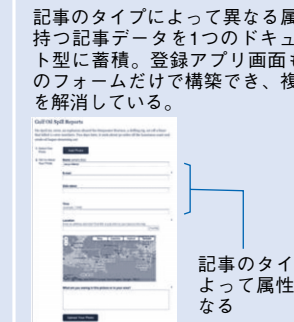
NoSQL製品の中でいま最も注目されており、採用事例も増えているのが「Cassandra」である。「Cassandra」はFacebook社が開発しApacheソフトウェア財団へ寄贈されたオープンソースで、次のような特徴がある。

- ①SPOF（Single Point of Failure：単一障害点。その個所が故障するとシステム全体が障害を受けるような個所）がない
- ②リッチなデータモデル（データ構造が柔軟で複雑なデータ管理が可能）
- ③一貫性制御の自由度（結果整合か一貫性保証かをコントロール可能）
- ④リニアな拡張性（ノード数の増加による性能劣化がない）

処理性能が高いNoSQL

NoSQL製品の処理性能評価を行うツール

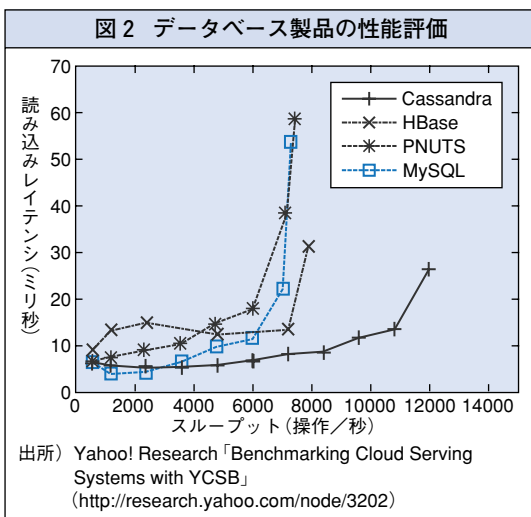
図1 NoSQLデータベースの分類と特徴

	単純キーバリュー型	カラム型	ドキュメント型
データストアの方式	キーとそれに対応するバリューをペアとして保持	キー・バリューのペアと追加キー(カラムファミリー)を保持	ドキュメント内にスキーマを動的に自由に配置
代表的なオープンソース	Voldemort Redis	Cassandra Hypertable	CouchDB MongoDB
アクセス/IF	キーを選択すると、対応するバリューを返すのみ	キー範囲とカラム指定によるバリューの選択と、キーとカラム指定によるバリューの選択が可能(カラム指定は省略可能)	バリューの値を条件としたバリューの部分選択が可能
実装イメージ			
利用形態	更新・検索方法が1つに限定されている場合に適する	更新・検索方法は複数だが固定である場合に適する	データが完全に構造化されず、更新・検索方法に自由度が必要なデータを扱うのに適する
利用事例	Amazonのショッピングカート (Amazon Dynamo)  ユーザーが購入しようとしているチェックアウト前の買い物リスト	Digg (ニュースリンク投稿)のユーザーデータ管理 (Cassandra)  自分が投稿した特定のニュースリンクに対して投票済みの友達を検索する	New York Timesの記事登録 (MongoDB)  記事のタイプによって異なる属性を持つ記事データを1つのドキュメント型に蓄積。登録アプリ画面も1つのフォームだけで構築でき、複雑さを解消している。 記事のタイプによって属性が異なる

として、米国Yahoo!社が開発し公開している「Yahoo! Cloud Serving Benchmark」がある。クラウドで使われる主なNoSQL製品のベンチマークが行えるツールである。

図2は、3つのNoSQL製品と、オープンソースのRDBMS「MySQL」について、処理量とレイテンシ(処理要求を出してから結果

が返されるまでの遅延時間)の関係を計測した結果である。「Cassandra」の処理性能は、TPC-C(業界団体のTPCが策定したベンチマークの1つ)に準じた負荷条件で秒間1万件の処理、レイテンシは読み込み・書き込みともに10ms程度である(図では読み込みのみ示した)。「MySQL」が秒間5千件程度を



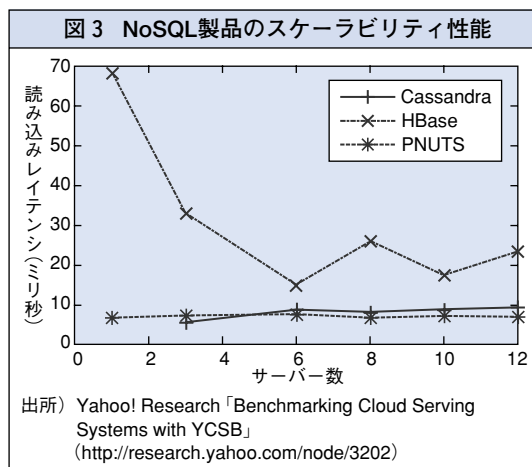
境に処理性能が急激に悪化するのに対し、「Cassandra」はほぼ一定であり、大量処理でも性能劣化が軽微であることが分かる。

図3はNoSQL製品のスケラビリティに関するベンチマーク結果である。12ノードまでの評価結果だが、「Cassandra」は「HBase」とともに性能劣化がなくリニアな拡張性を持つことが分かる。

野村総合研究所 (NRI) は社内で独自に性能評価を実施したが、レイテンシ、スケラビリティともに同様の結果を得ている。そのため今後クラウド向けのサービス構築にNoSQLを活用していく予定である。

オープンソースから生まれる最新技術

NoSQLは、クラウド時代の大規模分散環境でデータベースの主流になると見なされてきたが、最近になって、インメモリーデータベース (DBを高速なDRAM上で稼働する)



でオーバーヘッド (1つの処理に伴う付加的な処理) を除去し、データの一貫性を維持しつつ大きな性能向上とスケラビリティをオープンソースで実現した、SQLベースの「VoltDB」も発表され、NoSQLと次世代DBの座を争っている。スケラブルで高速処理が可能な新しいデータベースがいずれもオープンソースによって開発されていることは注目に値する。

これまでのオープンソースは、OS (基本ソフト) の「Linux」もデータベースの「MySQL」も、コスト削減、ベンダー固定からの解放、サポート選択の自由度などを目的に利用されてきたのであり、技術や機能が商用製品より優れていたとは必ずしもいえない。しかしこれからは、オープンソースから最新の技術や機能が提供される時代になっていくだろう。今後ますます重要度を増すであろうNoSQLを含めて、クラウド発のオープンソースの動向を注意深く見守っていく必要がある。 ■