

クラウド環境におけるデータベースのあり方

- 雲の中のデータベース KVS の可能性 -

アブストラクト

1. 研究の背景

クラウドコンピューティング(以下、クラウド)とはネットワーク、特にインターネットを経由したサービスの提供並びに利用の形態である。近年、クラウドが企業の間で盛んに利用されるようになってきたのは、クラウドの形態にすることでコスト削減が期待できることが最大の理由である。

KVS(Key Value Store)はクラウドを支える最も重要な技術の1つであり、クラウド環境下においてスケーラブルな大規模分散 DB を実現する。図表 1 は分科会が独自に調査したクラウド事例から、KVS 利用事例 12 件の KVS の採用理由とサービス分類の内訳を示したものである。

KVS 採用の最大の理由は、やはりコストであり、スケーラビリティとレスポンスがそれに続く。また、KVS 採用事例のサービス分類に共通するのは、サービスの利用拡大に伴ってデータサイズが増大していくということと、基幹業務データを直接扱わないということであった。

さらに、これらの事例を詳細に調査すると、KVS を使用したクラウドサービスの自社による構築事例は 1 件だけで、残り 11 件は大手クラウドサービス提供元のクラウドサービスを利用した SaaS(Software as a Service)の事例であることも判明した。以上より、次のような KVS の現状が浮かび上がってくる。

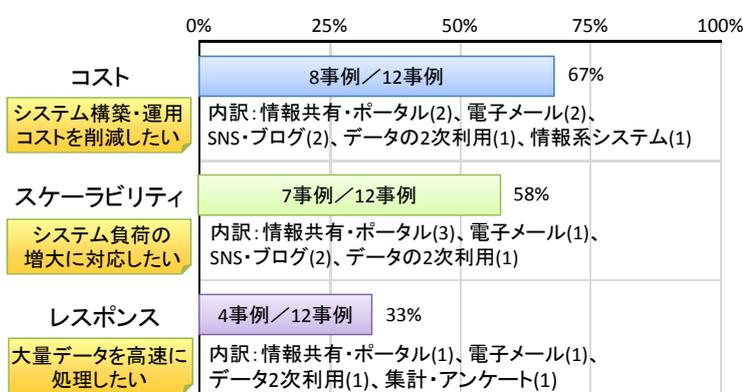
- (1) クラウドサービスにおける KVS の利用が認知され始めている。
- (2) 特に、スケーラビリティが要求される分野において利用が進んでいる。
- (3) しかし、基幹業務データを扱う分野での利用は進んでいない。
- (4) また、KVS を使用するクラウドサービスを自社で構築するといった積極的な利用は少ない。

(3)および(4)については、RDB のような十分に認知されている技術とは異なり、KVS の情報や実績が (RDB に比較して) 少ないことによるリスク回避という、KVS 黎明期ゆえの問題であると考えられるが、もし KVS が本当に優れた技術であるにも関わらず、それだけの理由で忌避されているとするならば、それは社会全体にとって大きな損失である。そこで、分科会では、「KVSこそクラウド時代にふさわしいデータベースである」として、その可能性について研究することにした。

2. 検証項目

KVS の可能性について検証するにあたり、検証項目を 3 つに絞り込んだ。まず、①業務・サービスへの適用性である。事例から既に多くの分野で KVS が採用されていることはわかったが、機能的な観点から適用性についてさらに検証する。次に、②スケーラブルなシステムの構築・運用の容易性である。実際に企業が KVS を採用した理由の内、スケーラビリティと構築・運用コストに注目し、実機を用いて検証する。最後に、③DBMS としての基本的な機能性として、RDB で動くアプリケーションを KVS に移行する作業を実際に行い、KVS の機能に不足や問題点が無いかを検証する。

図表 1 KVS の採用理由



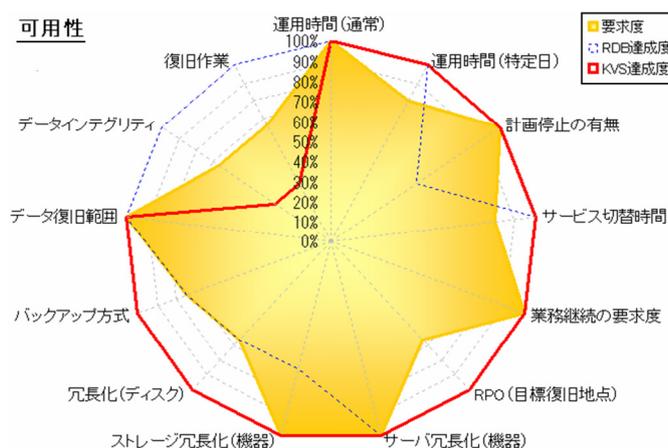
3. 業務・サービスへの適用性の検証

DBMS に求められる機能・非機能 117 項目について評価尺度を設定し、KVS と RDB の達成度を評価した「適用チェックシート」を作成した。適用チェックシートは、任意の業務・サービスについて、要求される達成度(要求度)を入力することで、KVS と RDB の達成度を比較・分析できるツールである。

適用チェックシートの出力の一例として、達成度のレーダーチャートを図表 2 に示す。この図から、ある業務・サービスの可用性に対する要求(塗りつぶし)に対して、RDB(点線)では計画停止の有無とストレージ冗長化への要求を満たさず、KVS(実線)ではデータインテグリティと復旧作業への要求を満たせないことがわかる。

適用チェックシートを利用して、業務・サービス 24 分類について KVS と RDB の適用度を求めたところ、KVS はいずれの分類においても要求達成項目数ベースで 90%前後の高い適用度を示すことを確認した。

図表 2 適用チェックシートで求めた達成度の例



4. スケーラブルなシステムの構築・運用の容易性の検証

メンバー全員で PC を持ち寄り、スケーラブルなシステムの構築・運用の容易性について 2 日間に渡って検証した。

検証の結果、クラスタの構築・運用が非常に容易であることを確認できた。そのことを示す 1 つのデータとして、検証作業の作業時間を図表 3 に示す。1 日目、2 日目ともに 6 ノードのクラスタがわずか 95 分で構築でき、さらに手順改善の余地があった。ただし、その容易さの一方で、設定に多少の誤りがあったても一見正常に動作し、問題が発生した場合に原因に気づきにくいなど、注意を要する点も明らかになった。

図表 3 検証作業の作業時間(実績値)

日程	作業内容	開始	終了	時間
1日目	検証用PC接続・設定 HBaseクラスタ構築(6ノード)	13:05	14:40	95分
	検証作業	14:55	18:20	205分
	クラスタ解体・撤収	18:20	19:00	40分
2日目	HBaseクラスタ再構築(6ノード)	10:10	11:45	95分
	検証作業	13:15	18:20	305分

5. DBMS としての基本的な機能性の検証

オープンソースの CMS(Contents Management System)である MediaWiki の RDB を HBase に移行する作業を通して、HBase の DBMS としての基本的な機能について問題が無いかを検証した。時間的な制約から、ページデータを取得・表示する処理だけを移行したが、KVS の機能で実現できることが確認できた。

検証作業の成果物の概要と規模を図表 4 に示す。作業に要した工数は 14 人日であり、不慣れな作業であることを考慮すれば、生産性は高いと言えるだろう。また、使いやすさ、習得のしやすさといった要件についても、導入が容易である、API が複数の言語に対応している、API がシンプルで理解しやすい等、良好な感触を得た。

図表 4 成果物の概要と規模

成果物	作業分類	概要	ファイル数	修正行数
HBase版 MediaWiki	修正	ページデータ読み込み等のロジック部	5	130
		不要機能無効化	7	14
	新規作成	HBaseアクセスモジュール	1	345
運用ツール	新規作成	HBaseテーブル定義、データ移行ツール	8	461
合計			21	950

6. まとめ

以上の検証の結果、KVS は 業務・サービスへの適用範囲が広いこと、 スケーラブルなシステムを容易に実現できること、 基本的な機能に大きな問題が無いことを確認した。ただし、KVS は管理ツールや情報が RDB のように充実していないことから、現状では利用しづらい側面もある。しかし、これは KVS が発展途上の技術であるが故の問題であり、今後、解消されていくと考えられる。

KVS はネットワーク上での分散性とスケーラビリティを低コストで実現できる、現状では唯一のデータベースである。本研究で示した通り、KVS の構築・運用・アプリケーション開発は決して敷居の高いものではなく、その特性を理解しさえすれば恐れるに足りない。企業や技術者の方々には、積極的に KVS に取り組んで頂き、来るべきクラウド時代のビジネスにおける技術的な強みとして頂きたい。