

情報爆発時代におけるデータベース技術

Database Technology in Information Explosion Era

喜連川 優
Masaru Kitsuregawa

東京大学 生産技術研究所 戦略情報融合国際研究センター
Center for Information Fusion, Institute of Industrial Science, Univ. of Tokyo

本稿は、パネルの趣旨である「データベース分野における現状、問題点、そして今後の目指すべき方向について議論をする」為の素材を提供することを目的としている。

1. データベースにおける核技術

データベースにおける核技術は、トランザクション処理とクエリ処理に集約される。この2つの核技術はデータベース管理システムのみが提供できるものであり、非常に広範に利用される必須のソフトウェアプロダクトと位置づけることができる。

トランザクション処理は、今日の社会における契約の基盤を与えるものであり、コミットにより処理の永続性を保証する。即ち、障害時においても一貫性のある状態への回復可能性を保証する核技術である。基本的な枠組みには変化は無いものの、今後 MicroPayment の多用により当該技術の一層の高度化が進むと期待される。

クエリ処理はストレージウォールに対する処理の高速化を目指すものであり、ディスク装置の性能とプロセッサの性能の隔たりは著しく、各種のアクセス法が種々のデータに対して提案されてきた。通常のデータベースにおけるインデックス、結合手法だけでも、高度な工夫がなされ、とりわけ、View Materialization に見られる如くストレージのコストの大きな低減から、スペースで、性能を得る手法が多様化しつつある。クエリコンパイラの高度化も長期に亘り改良が加えられ、近年では、動的な実行時最適化が導入されるに到っているがマルチコア対応等更なる高度化が継続的になされることは確実であろう。

2. 情報爆発 <http://itkaken.ex.nii.ac.jp/i-explosion/>

UCパークレイ校による調査によれば、人類の創生する情報量は爆発的に増大している。これにマシンやセンサーによるデータを加えるとその勢いはさらに凄まじく、情報爆発時代と呼ぶにふさわしい。特定領域研究「情報爆発IT基盤」では、情報爆発時代における種々の難題に対し、当該グランドチャレンジを克服するための基礎研究を推進する予定である。

必要な情報を「探す」という行為の効率化は知識集約型産業において極めて重要な課題と言える。情報検索、データベース、機械学習、自然言語処理など多くの技術を集約することにより更なる改善を一步一步研究してゆきたい。探索対象であるサイバー情報空間自体の解明、トラストなど、挑戦すべき課題は山積している。

情報量の膨張はそれを運用する情報システムに対し大きなストレスを生み出す。一方、システムからの膨大な管理情報を活用することにより、システムの安定化への展開が期待される。システム技術への期待は性能の高さではなく、レジリエンシ、メンテナンスビリティなど、運用容易性にシフトしており、大規模実験システムを利用した研究を進める予定である。

日々爆発的に増大する未知なる情報への日常的な遭遇は、人々に新しいライフスタイルを強いている。また、実世界に対するサイバー世界の滞留時間比率が増加する中で、人に優しいインタラクションを支える技術開発が望まれ、重要な研究課題と考えている。

適切な情報共有は、組織活性化の肝であり、膨大な情報の適切な配信など、情報ガバナンスの重要性が注目されている。また、新しい情報サービスにおいては、メリットとリスクの両面を考慮した社会制度設計が不可欠であると考えており、実証実験を想定した有用性検証を視野に入れた計画を策定しつつある。

情報が爆発的に急増する今世紀において、当該現象をどのように乗り越えて行くかという課題に対し、上記の視点を中心に、広い分野の研究者間で戦略的に研究を進めることにより従来にないタイムリーな成果を目指している。とりわけ、支援班により、10,000 CPU環境、高度センサーラーム、大規模ウェブアーカイブなどユニークな研究プラットフォームを共有資源として活用することにより、大きなスケールの実験を試みたいと考えている。公募枠を非常に大きく設定しており、参画を是非ご検討頂きたい。

3. データベース分野における技術潮流

3.1 構造から非構造へ：データベースが対象とするデータは構造化データであるが、データ全体に対してその割合は僅かではない。半構造のみならず、非構造データからの価値創出は大きな流れとなる。

3.2 超レジリエンシ：9.11により、米国 SECにより経済機関業種の営業継続性に対する努力が推奨されるに到っている。勘定系だけではなくデータの短時間復旧、RTO短縮、RPOの厳密性が強く求められる。上位アプリにおけるトランザクション特性の実現も課題と言えよう。

3.3 ノブレス：ソフトウェアシステムの宿命とは言え、現行のDBMSはSQL機能の強化、非常に多様なインデックス、アクセス法の導入などにより資源管理が極度に複雑化し、DBMS自体の調整に大きな労力が必要となっている。本質的に非常に難しい課題と言えるが、自己管理機構についての研究が開始されている。

3.4 Data Driven Science/Data Exploration Science：サイエンスにおいて、そのスタイルがdata explorationになりつつあり、その為の強力なツールが求められる時代となる。巨大なデータを効率よく管理可能なソフトウェア基盤としてのDBMSの果たす役割への期待は大きい。

その他、紙面の都合上記載できないが、情報融合、多様なマイニング、ウェブ、サービス生成、センサーネット、Pub/Subストリーム、プライバシーデータ管理、超分散、1000年保存、Lineage、軽量エンジンなど次々とワクワクする研究課題が生み出されている。