ペタバイト級IoTデータを高速に処理するスケールアウト型データベース GridDB

GridDB Scale-Out Database Facilitating Handling of Petabytes of Data at High Speed

服部　雅一 HATTORI Masakazu　　福島　伸之 FUKUSHIMA Nobuyuki　　スヘルマン アンガ SUHERMAN Angga

世界的にIoT（Internet of Things）の適用が広がり，様々なデータを用いてシステムの全体最適化を目指すCPS（サイバーフィジカルシステム）に注目が集まっている。CPSを実現するために，高いレベルの信頼性，拡張性，及び処理能力を備えたデータベース（DB）が求められているが，従来のRDB（Relational DB）やNoSQL（Not Only SQL（Structured Query Language））DBではそれらを満たすことが困難であった。

東芝デジタルソリューションズ（株）は，DBクラスター技術や高速データ処理技術などの独自技術を開発し，信頼性，拡張性，及び処理能力の要件を満たしたスケールアウト型DBであるGridDBを製品化して，提供してきた。これまでに，中小規模から大規模まで，社会インフラ系のアプリケーションを中心に適用されている。GridDBの適用範囲を更に拡大するため，ペタ（P：10^{15}）バイト級のデータへの対応など新たな技術開発を行うとともに，オープンソースソフトウェア（OSS）化も推進している。

In line with recent world trends in application of the Internet of Things (IoT), attention is being increasingly focused on cyber-physical systems (CPS) aimed at the optimization of whole systems by making use of a wide variety of data. In order to realize CPS systems, database systems with high reliability, high scalability, and high processing capacity are essential. However, it is difficult to fulfill such requirements using conventional database systems such as relational databases (RDBs) and Not only Structured Query Language databases (NoSQL DBs).

To rectify this situation, Toshiba Digital Solutions Corporation has developed and released a lineup of GridDB scale-out databases for social infrastructure systems ranging from small to large in scale, utilizing proprietary technologies including a database cluster technology and a high-speed data processing technology. To disseminate GridDB more widely, we are also making efforts to promote the development of technologies to handle petabytes of data as well as activities related to open source software (OSS).

1. まえがき

様々なものがインターネットにつながるIoT化が進展し，大量のデータを活用してシステムの全体最適化を目指すCPSが注目されている。CPSとは，実世界（フィジカル空間）にある多様な情報をセンサーネットワークなどでデータとして収集し，仮想世界（サイバー空間）で大規模データ処理技術などを駆使して分析・知識化を行い，そこで創出した情報・価値をフィジカル空間に戻して，産業の活性化や社会問題の解決を図るものである。CPSを実現するには，フィジカル空間で生成される膨大な観測データをサイバー空間で再現する大規模なDBが必要となる。しかし，次のような厳しい要件を満たす必要があるため，従来のDBではその達成が困難であった。

⑴　高い信頼性　　CPSの多くはミッションクリティカルであり，信頼性が損なわれる可能性があるベストエフォート型の性能は許されない。つまり，高い信頼性が必要不可欠である。

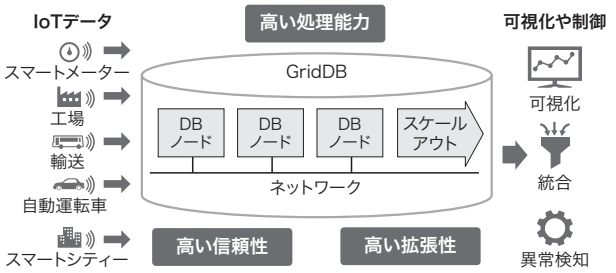


図1．GridDBのコンセプト

GridDBは，高いレベルでの信頼性，拡張性，及び処理能力を備えたデータベースとして開発された。ノードを増やすことで，負荷の増加にも対応できる。

Concept of GridDB

⑵　高い拡張性　　将来にわたって増大し続けるデータ量に対応するため，DBノード（サーバー）の数を増やすことによる処理性能の向上，すなわちスケールアウト性が求められる。

⑶　高い処理能力　　フィジカル空間を制御するには，

秒やミリ秒の間隔で生成される大量のデータを格納し，更に，データ取得直後から，リアルタイムに参照や加工などのデータ処理ができなければならない。つまり，高い処理能力が求められる。

GridDBは，来るべきCPS時代に向けて，これらの厳しい要件に応えるために開発された，新しいコンセプトとテクノロジーを備えたスケールアウト型のDBである（**図1**）。

ここでは，GridDBの技術や機能の特長と，ベンチマーク結果について述べるとともに，OSS化についても述べる。

2. GridDBの特長

2013年，東芝デジタルソリューションズ（株）は，高頻度で大量に生成されるIoTデータやビッグデータの管理に適したスケールアウト型DB GridDBを製品化し，2015年にSQLインターフェース機能の提供，2017年にSQLの分散並列処理化など，機能拡張及び性能強化に努めてきた[1]。2018年以降，Pバイト級のデータへの対応強化やSQL処理の改善などを行っている。

以下に，GridDBの主な特長をまとめる。

2.1 自律型DBクラスター技術ADDA

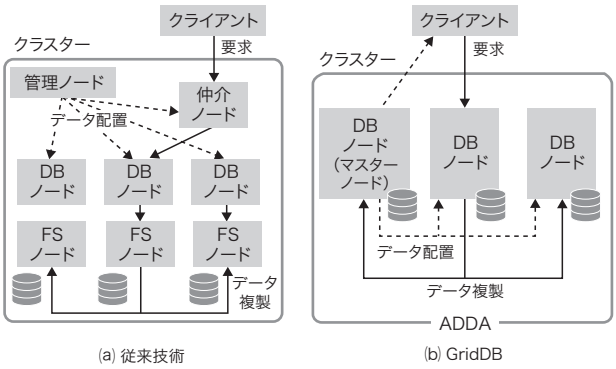
GridDBの高い信頼性と拡張性を確保するため，独自開発の自律型DBクラスター技術ADDA（Autonomous Data Distribution Algorithm）を適用している。

⑴ ノード障害への対応　ADDAは，データの複製（レプリカ）をノード間で互いに持ち合う冗長性を備えている。万一ノードに障害が発生しても，ほかのノードがレプリカを使って障害ノードの役割を引き継ぎ，クライアントは接続先ノードを自動的に切り替えて，処理を継続できる。

⑵ パフォーマンスの向上　従来技術と，ADDAを組み込んだGridDBの構成を，**図2**に示す。従来のスケールアウト型のDBでは，データを分散化するためデータの一貫性が弱くなり，逆に，一貫性を強めるとパフォーマンスが低下するという欠点があった。例えば，マスターノードがクラスターを集中管理するマスタースレーブ方式の場合，データの一貫性を維持しやすいメリットはあるものの，クライアントとDBノードの間に存在する管理ノードや仲介ノードがボトルネックとなり，結果としてパフォーマンスの低下を引き起こす。

これに対して，GridDBには管理ノードや仲介ノードは存在せず，パフォーマンスの低下が起こらない。また，通信やデータ変換のための間接コストがないため，大幅なパフォーマンスの向上が可能となった。

⑶ 安定なノード拡張の実現　ADDAは，ノードの追加・削除や，レプリカの欠損，データ負荷のアンバランスなどの状態変化を検知すると，クラスター内の各ノードへのデータ割り当てとノード間のデータ転送に関する最適プランを計算することで，安定的なデータ再配置を可能にする。これにより，ミッションクリティカルなシステムで必須とされる，無停止でのスケールアウトを実現した。



FS：ファイルシステム

図2. ADDAを利用したGridDBと従来技術のクラスター構成の比較

GridDBでは，ADDAの働きにより，複数のDBノードからマスターノードが選ばれ，そのノードが，ほかのノードにデータ配置やデータ複製を指示する。従来技術のような管理ノードや仲介ノードがないため，パフォーマンスが向上する。

Comparison of configuration of clusters of conventional database and GridDB

2.2 NoSQL・SQLデュアルインターフェース

GridDBは，NoSQL DBで用いられてきたキーとバリューから成るデータモデルを発展させて，カラムとレコードから成るテーブルをバリューとして表現する独自のデータモデルを採用している。インターフェースを通して，RDBで使われてきたSQL DBの利便性と，NoSQL DBの高速性とを，シームレスに利用できる。

⑴ NoSQLインターフェース　キーで識別されたレコードに対して，登録や，更新，削除，参照などの操作を行える。Javaや，C，Python，Go，Node.jsなどのプログラム言語からアクセスできるように，各種プラグインを提供している。

⑵ SQLインターフェース　バリューであるテーブルをRDBのリレーションとみなしてアクセスできる。ANSI-92（米国規格協会規格 92）のSQL機能をサポートし，ODBC（Open Database Connectivity）やJDBC（Java Database Connectivity）などのRDBへの接続インターフェースを，提供している。ユーザーにとって，DB言語として最も普及しているSQLを使えるメリットは大きく，またBI（ビジネスインテリジェンス）ツールやETL

（Extract，Transform，Load）ツールとの連携も容易である。

2.3 高速データ処理技術

⑴ CPU性能を最大限に引き出す高速化技術　従来のRDBでは，メモリーを大容量化しても，クエリー処理や，バッファー処理，リカバリー処理などに大きなオーバーヘッドが発生する。そのため，本質的なデータ処理にCPUリソースの僅か10％前後しか割り当てられず，CPUパワーを十分に発揮できないことが知られている。GridDBでは，大容量化されたメモリーを前提に，バッファー処理の軽量化，リカバリー処理の軽量化，及びデータ処理時のロックフリー化を行うことで，従来のRDBで発生していたオーバーヘッドを最小化した。また，CPUのマルチコア・メニーコア化を前提に，データ受信やタイマーといったイベントをトリガーとして，非同期的なデータ処理を絶え間なく実行するイベント駆動方式を開発した（**図3**）。このアーキテクチャーの利点は，マルチコア化されたCPUパワーを，最大限に引き出せることである。

⑵ 巨大テーブルに対応した分散並列SQL処理技術

タスク，データ，及びパイプラインの3層にわたる分散並列SQL処理技術と，巨大テーブルをより小さな複数の内部テーブルに分割するテーブルパーティショニング機能により，単一ノードでは扱えなかった巨大データを短時間で処理できるようになった（**図4**）。

また，分散DBは単一のDBとは異なり，最適化に必要なテーブル情報をノードから集め続けることは困難である。そこで，ノード間にまたがるグローバルな最適化と，ノード内に閉じたローカルな最適化という，粒度の異なる2段階のSQL最適化を行えるようにした。

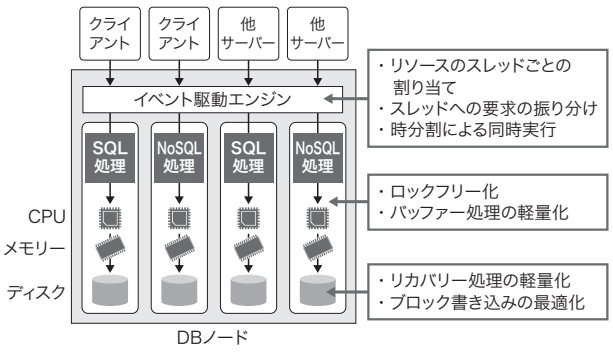


図3. イベント駆動方式の概要

CPUパワーをフル活用するため，オーバーヘッドの少ないイベント駆動エンジンを組み込んでいる。

Outline of event-driven architecture

2.4 Pバイト級データとノードへの対応

近年のIoTシステムでは，デバイスの数とともに，長期稼働を経て蓄積するデータサイズも増えており，Pバイト級のデータ管理が求められるケースも珍しくない。その場合，数十台のノードから成る大規模クラスターシステムの構築・運用コストは非常に高くなるため，より少ないノード数から成るクラスターが望ましい。

GridDBのVer.4.3では，DB内部のデータ管理構造を最適化してリソース使用量の大幅な削減を図り，1ノード当たりの最大DBサイズをPバイト級まで増加させた。この強化とともに，DBのバッファー制御機能と，クラスター内のデータ配置機能，更には，複合索引などの機能を強化し，大規模データに対する処理性能を向上させた。その結果，スケールアウトだけでなく，スケールアップ（サーバーそのものの増強による処理性能の向上）も可能にして，Pバイト級のデータを実時間で処理できるようになった。

2.5 ファストデータとビッグデータの処理統合

従来のビッグデータ処理では，大規模データに対するバッチ的な分析処理が主流で，バッチ処理であるために最新データを使えないという欠点があった。そのため，リアルタイム性の高い処理（ファストデータ処理）と組み合わせることでこれを解決しようとするラムダアーキテクチャーが提唱されている。しかし，ラムダアーキテクチャーには，構築・運用コストの上昇やシステムの複雑化という新たな問題が発生する。

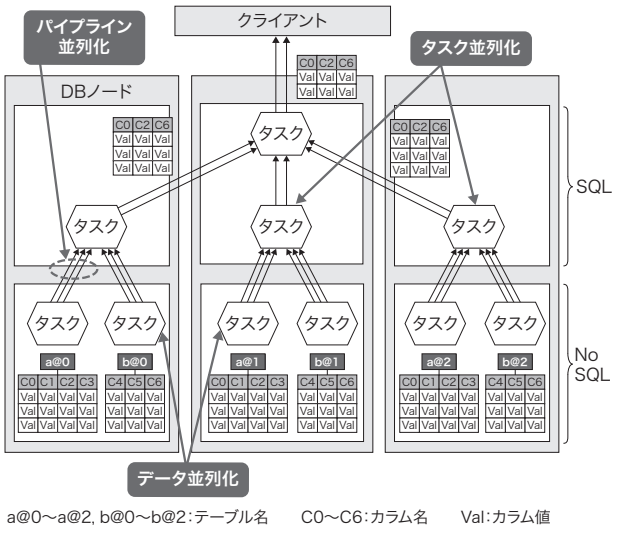


図4. 分散並列SQL処理の概要

タスク，データ，及びパイプラインの3層にわたる分散並列処理により，SQLが最大限に高速化される。

Outline of distributed parallel Structured Query Language (SQL) processing

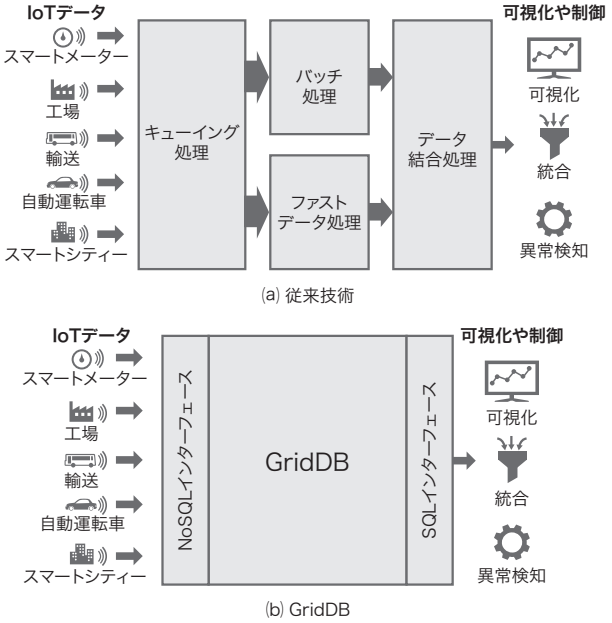


図5. ファストデータとビッグデータの統合

従来は複数のサービスやコンポーネントを組み合わせていたが，GridDBはそれらを統合できる。

Integration of fast data and big data

GridDBは，リアルタイム性の高い登録や参照にはNoSQL，大規模なデータ集約・加工にはSQLという使い方により，ファストデータ処理とビッグデータ処理を一つのDBで扱える（**図5**）。

3. ベンチマーク

2章で述べたように，様々な技術や手法を取り入れて改善を重ねたGridDBの，ベンチマーク結果を記載する。

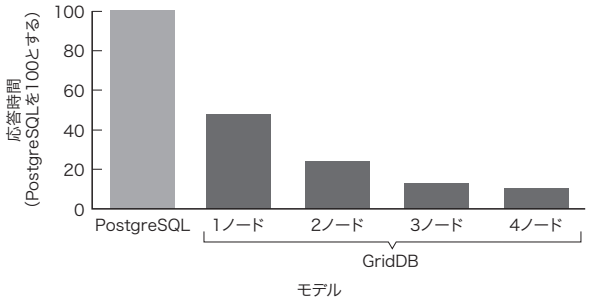
3.1 SQLベンチマークTPC-H

TPC-H（TPC Benchmark H）(注1)を利用した，主要なRDBのPostgreSQLとパフォーマンスを比較した結果を，**図6**に示す。

データ規模を示すスケールファクターを100とし，ノード台数を増やしながらSQLに対する総応答時間を計測した。PostgreSQLは1ノード上での実行であり，縦軸にPostgreSQLを100としたときの各応答時間を示した。

この図から分かるように，GridDBは1ノードでもPostgreSQLより高速であり，ノード台数の増加によって応答時間が短くなり，ほぼリニアーなスケーラビリティーを示している。また，2.4節で述べた大規模データに対する性能改善により，

（注1） TPC（Transaction Processing Performance Council：トランザクション処理性能評議会）が運営する意思決定支援システムの標準的なベンチマーク。



使用したモデル：PostgreSQL 9.6, GridDB AE 4.0
CPU：8-core Intel® Xeon® E5-2620 v4 2.10 GHz
メモリー：64 Gバイト
HDD（ハードディスクドライブ）：SAS（Serial Attached SCSI（Small Computer System Interface））12 T（テラ：10^{12}）バイト
基本ソフトウェア：CentOS 7 with kernel 3.10.0-514.el7.x86_64
ネットワーク：1 Gビット Ethernet
データセット：TPC-H（SF 100），Q1-Q8

図6. SQLベンチマーク結果

TPC-Hにおける，八つのクエリーの総応答時間で比較している。GridDBは，ほぼリニアーなスケーラビリティーを示している。

Changes in response time accompanying increase in number of nodes in GridDB cluster

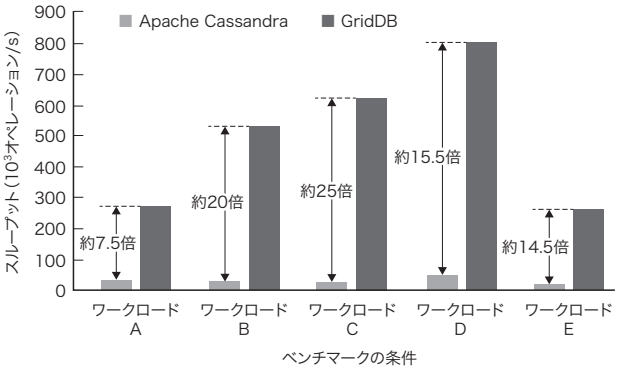


図7. NoSQLベンチマーク結果

様々なワークロード条件で測定した結果，GridDBはApache Cassandraに比較して非常に高いスループットを示した。

Comparison of performance of NoSQL DB (Apache Cassandra) and GridDB

これまでの性能[1]に比べて1ノードの応答時間を2倍近く高速化した。

3.2 NoSQLベンチマークYCSB

YCSB（Yahoo! Cloud Serving Benchmark）を利用し，主要なNoSQL DBであるApache Cassandraと性能を比較した（**図7**）。登録，更新，及び参照の比率を変えた五つのワークロード条件A～Eを使って計測した。どのシナリオにおいても，GridDBは，Apache Cassandraと比較して非常に高いスループットを達成した[2]。

3.3 時系列ベンチマークYCSB-TS

YCSBの時系列データ版であるYCSB-TS（Time Series）を利用し，主要な時系列DBのInfluxDBと性能を比較し

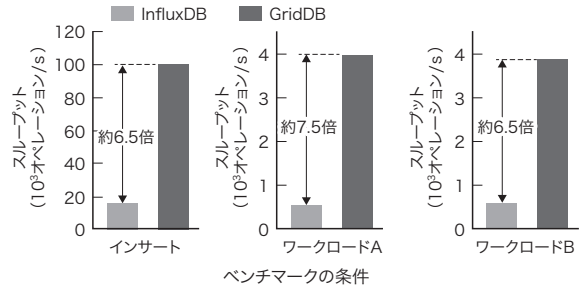


図8. 時系列ベンチマーク結果
GridDBは，InfluxDBと比較して高いスループットを達成した。
Comparison of time-series benchmarks of InfluxDB open source time series database and GridDB

た（**図8**）。インサート，ワークロードA，及びワークロードBは，それぞれ，データ登録，ランダムな読み取り操作，データのスキャンなどの参照操作を意味している。GridDBは，InfluxDBと比較して高いスループットを達成した[(3)]。

4. OSS化

2016年，オープンイノベーションによる新しい価値を目指し，GridDBをOSSとしてソースコードを広く公開した。また，GridDBデベロッパーサイトやSNS（Social Networking Service）を通じて，GridDBの開発者向けにホワイトペーパーやチュートリアルビデオなどの情報発信を行っている。その結果，ダウンロード数は海外を中心に増加傾向にある。

また，商用版だけで使用可能であったSQLインターフェースを，2020年にOSS化した。これにより，外部ソフトウェアとの連携やアプリケーション開発の利便性が，格段に向上すると考えられる。

5. あとがき

当初，GridDBはメガソーラー発電のデータ管理向けに開発されたが，その適用事例は年々増加し，産業や社会の幅広いシーンで導入されてきた。電力自動化システムや，スマートコミュニティシステム，車両管理システム，設備監視システム，製品や製造プロセスでのデータを管理するデジタルツイン・システムなど，社会インフラ系を中心に様々な分野で利用されている。また，ソリューションパッケージMeister DigitalTwinや，インダストリアルIoTサービスであるTOSHIBA SPINEXなどの基幹DBとしても採用されており，適用範囲が拡大している。

文　献

(1) 服部雅一，幸田和久．増え続けるIoTデータの管理に最適なスケールアウト型データベース GridDB．東芝レビュー．2018，**73**，3，p.45–49．<https://www.toshiba.co.jp/tech/review/2018/03/73_03pdf/a11.pdf>，（参照 2020-05-07）．
(2) フィックスターズ．GridDBとCassandraのパフォーマンスとスケーラビリティ Microsoft Azure環境におけるYCSBパフォーマンス比較．フィックスターズ，2017，33p．<https://www.griddb.net/ja/docs/Fixstars_NoSQL_Benchmarks_ja.pdf>，（参照 2020-05-07）．
(3) フィックスターズ．GridDBとInfluxDBを使用した時系列データベースのパフォーマンス比較．フィックスターズ，2018，19p．<https://griddb.net/ja/docs/TimeSeries_Database_Benchmark_GridDB_InfluxDB.pdf>，（参照 2020-05-07）．


服部　雅一　HATTORI Masakazu
東芝デジタルソリューションズ（株）
ソフトウェアシステム技術開発センター
日本データベース学会会員
Toshiba Digital Solutions Corp.


福島　伸之　FUKUSHIMA Nobuyuki
東芝デジタルソリューションズ（株）
ソフトウェアシステム技術開発センター
ソフトウェア開発部
Toshiba Digital Solutions Corp.


スヘルマン アンガ　SUHERMAN Angga
東芝デジタルソリューションズ（株）
デジタル人材開発・技術管理部
Toshiba Digital Solutions Corp.