IoT eラーニング

IoTとデータベース②
（IoTにおけるデータベースの活用事例）

国立大学法人　琉球大学

目次

Amazonのマネージドサービス

- **Amazon DynamoDB**

  Amazon DynamoDBは可用性が高くスキーマレスなサービスをAWS（Amazon Web Services）上で提供するNoSQLデータベースサービスである。

  Amazon.com内におけるシステム開発時に、RDBMSでは限界があることから開発されたNoSQLがベースとなっている。

  DynamoDBはKVS（Key Value Store）に分類されるNoSQLデータベースである。

  KVSとは、KeyとValueを組み合わせる単純な構造からなっている。Keyが指定されると、Keyに関連付けられたValueが呼び出される。

Amazonのマネージドサービス

特徴としては、
◆ シンプルなデータモデル
◆ スケールアウトのしやすい構造
◆ データを高速に読み書きできる
◆ 分散処理に適している
◆ リレーションやトランザクションの概念がない
があげられる。

一貫性よりも速度や可用性を求めたサービスである。

他のマネージドサービスと比べてメンテナンスウィンドウがないことも特徴といえる。

Amazonのマネージドサービス

- **Amazon Redshift**

AWS（Amazon Web Services）のデータウェアハウスサービスとして2013年に提供を開始している。

Amazon Redshiftは超並列演算（MPP）によるデータとクエリの負荷を自動的に分散することで、分散処理による高速なクエリのパフォーマンスの維持とあわせて拡張性も実現している。

データ分析に用いられるデータベース製品には次の２つに分類できる。

- シェアードディスク型
  複数のサーバは単一のストレージかクラスタファイルシステムを操作する。共有ストレージのため並列処理時の競合やサーバー間の通信がボトルネックと言える。

- シェアードナッシング型
  すべてのサーバがストレージを持ち操作する。並列処理時に理論上競合が発生しないがサーバーの数だけストレージが必要となる。

Amazon Redshiftはシェアードナッシング型にあたり、ハードウェアのセットの増加でI/O性能が向上し多くの読み込みが実行しやすくなることから、SELECTに特化したデータベースと言える。

Microsoftのマネージドサービス

- **Azure SQL Data Warehouse**

Azure SQL Data Warehouseはリレーショナルか非リレーショナルかに関わらず、大規模な処理ができるデータベースサービスである。高い拡張性とパフォーマンスを備えており、数百テラバイト、数ペタバイトクラスのデータ分析を可能としている。

Azure SQL Data Warehouseのエンジンは、SQL ServerのRDBMSを基盤としている。SQL Serverとの互換性が高いため、SQL Serverを扱ったことのある技術者であればストレスなく操作することができる。

大量のデータを取り扱うため、アーキテクチャや機能が実装されておりSQL Serverとは異なる点もある。主な機能の１つとして超並列分散処理（MPP）があげられる。複数ノードによる分散処理を行うことにより、高速な処理を実現している。

Azure SQL Data Warehouseでは3層構造のアーキテクチャを持っており、
◆ 接続とクエリの受け取りを制御するコントロールノード
◆ コントロールノードの配下にあり処理を受け取り分散処理を行うコンピューティングノード
◆ データを分散して格納するストレージ
の3層構造となっている。
コンピューティングノードは最大60台まで増やすことができ並列処理の向上を図ることができる。

Googleのマネージドサービス

- **Google Cloud Datastore**

  2008年にGoogle App Engineの一機能としてサービス提供が始まり、2013年にGoogle Cloud Datastoreという名称でGoogle Cloud Platformのサービスの１つとして提供が開始された。

  Google Cloud Datastoreは、ドキュメント指向に分類されるNoSQLデータベースである。

  Google Compute Engineなどの上でAPIを使ってアプリケーションはアクセスすることができる。また、バッチ処理ではGoogle Cloud Dataflowなどと連携することでデータの入出力を行う。

  特徴の一つとして、NoSQLデータベースであるが、トランザクションやSQLライクなクエリ言語が提供されており多機能なデータサービスである。

  ポケモンGOやSnapchatなどで採用されている。

Googleのマネージドサービス

- **Google Cloud Bigtable**

  Googleが社内で開発・利用しているキーバリュー型のNoSQLデータベースである。
  Google検索などGoogleが提供するサービスで利用されている。
  2016年8月にGoogle Cloud Platformの中のサービスとして提供が開始された。

  Google Cloud Bigtableはテーブル間のJOINやトランザクションなどの一般的なRDBMSにあるような機能は提供されていない。

  大量（1TB以上など）のデータを扱い、特定のクエリーに特化したとき最大限の効果を発揮するデータベースサービスとなっている。

  IoTなどで扱われるリアルタイムであり大量のデータを格納・分析する用途などが想定される利用場面となる。

Googleのマネージドサービス

- **Google BigQuery**

  Google BigQueryはGoogleが提供する分析用データウェアハウス型マネージドサービスである。

  もともとは、Google社内で開発・利用していたデータ解析用システム「Dremel」を外部ユーザ向けに改良しサービス化した経緯がある。

  Google BigQueryでは数億行あるいは数100GBの大規模データセットに対して数秒・数十秒で結果が得られる高速処理性能がある。この高速性を実現している仕組みとして次の２つがあげられる。

  - カラム構造データストア
    カラム指向の構造をしており、１件のレコードをカラムごとに分割し別々のストレージに分散配置することにより、クエリ実行時の高速データ参照を実現している。

  - ツリーアーキテクチャ
    クエリを受け付けると大規模な分散処理へ展開するツリーアーキテクチャの構造を持っている。クライアントからクエリが要求されると、root serverから実際にクエリ実行を担当する多数のleft serversに対し、クエリが木構造で広がることによって大規模分散処理を実現している。

データベースの活用事例

- **電力小売自由化対応**

  - 電力小売自由化とスマートメーターの導入

    電力の小売自由化は2000年3月にスタートした。

    最初は20kV以上の「特別高圧」を利用する大規模な工場やオフィスビルなどで電力会社を自由に選べるようになった。この際に、新規に参集した小売電気事業者からの購入も可能となった。

    その後、段階的に条件が変わり2016年4月には小売り全面自由化へと進んだ。
    この小売自由化により、独占的であった市場に競争原理が働くようになり、各事業者は利用者の利便性を考えたサービスを提供する必要に迫られた。

    そのような背景から、利用者のニーズを捉えるべく「スマートメーター」が導入された。これは電気の使用量を30分ごとに検針し集約装置へ送信する仕組みで、検針作業の自動化や遠隔操作による操作の簡素化だけでなく、電気使用量の見える化などの新しいサービス提供にもつながっている。

データベースの活用事例

- 検針データ処理システムの開発

電力小売事業に参集した電力会社が、2004年に東芝と共にスマートメータから送信される検針データを処理するシステムの開発を始めた。

当初、データベースにはRDBが採用され、送付されたデータから託送料金の計算を実行する仕組みを構築した。

2015年4月からは、一般家庭など低圧契約者に対するスマートメータの設置をスタートさせた。

設置された機器の数は数千台から一気に数百万台に増え、さらに今後は域内のすべての利用者への設置を計画している。

処理対象が数百倍に増加したことにより、従来と同じ時間内で処理を終わらせるためにはRDBに代わるデータベースの導入が必要となった。

データベースの活用事例

- 検針データ処理システムの開発

検討されたのは東芝が開発した「GridDB」である。

GridDBとは、ビッグデータ向けNoSQL型インメモリデータベースであり、ペタバイト級のビッグデータでも高速に登録と検索が行える。

数百万台ものスマートメーターから30分ごとに継続してデータが送信されてくるため、処理に要求されるパフォーマンスと信頼性には、従来の数十倍もの高い性能を要求される。

このことからRDBをチューニングでは必要な性能に到達できないと判断しGridDBを採用した。

当初は、十分な性能を発揮することができなかったが、検証と改善を進めることにより750万件の検針データを3分44秒で取り込み、1分32秒で処理を終わらせるという処理性能を示せることができた。

データベースの活用事例

- **グローバル遠隔通信サービス**

神戸製鋼所の汎用コンプレッサー「Emeraude ALE」に、東芝のIoTサービス基盤が採用され、グローバル遠隔通信を可能にする「コンプレッサM2Mクラウドサービス」を展開することとなった。

神戸製鋼所が提供するサービスでは、日本を含む世界各地で稼働する汎用コンプレッサーのデータを収集し蓄積することで、装置の稼働状況を一括管理できるよう「見える化」を進める。

神戸製鋼所の保守・サービス部門やエンドユーザーに、回線やクラウドサービスをワンストップで提供する。

これによりリアルタイムでの状態監視を可能としている。

また、東芝は今後の展開として、ビッグデータ解析による故障予知や予防保全などのサービスを計画している。

データベースの活用事例

- **建物エネルギー管理システム**

  東芝は2015年より建物エネルギー管理システム（BEMS）サービスを提供している。

  BEMSは建物のエネルギー需要を監視・制御するシステムである。
  住宅用または商業用ビルに備わる空調管理や照明、防犯などさまざまなものを監視・制御することができる。

  東芝が提供しているサービスでは、GridDBを使用して数百からの建物から2TBを超えるデータを収集・格納している。毎秒1,000件のレコードが処理されているが、問題を発生させずに必要なパフォーマンスを提供している。

データベースの活用事例

- **車両管理システム**

DENSO International Americaは、次世代の車両管理システム構築にGridDBを採用することとなった。

このシステムで取り扱われるデータは、車両に搭載されているさまざまなセンサーのデータになり、これらは車両のコンピュータで収集し一時解析される。解析結果はLTE通信でクラウドサーバーへ送られ、クラウドサーバーは受け取った内容をストリーミング表示し、オフラインでの高度な分析を実施する。

このシステムでは次の技術的な条件がある。

- 広大な地域で活動する無数の車両からデータがランダムにサーバに送付され、サーバー側は漏らすことなくデータベースへ書き込まなければならない。
- 取得したデータのストリーミング表示を行うことからリアルタイムに近い速度が求められており、データベースからの読み出しも高い性能が求められている。
- 管理対象は車両となるため、常に稼働し続ける可用性が必要となる。

GridDBが持つ機能により、これらの条件を満たすことができるとしている。

データベースの活用事例

- **オンラインゲームシステム**

  オンラインゲームを提供しているスクウェア・エニックスでは、当初のオンラインゲームのデータをRDBへ格納し処理していた。しかし、ゲーム規模の拡大とともに取り扱われるデータ量も拡大されるにつれてRDBでは対応できなくなってきた。

  ゲームプレイヤーに最適なゲーム体験を提供するために、スクウェア・エニックスはRDBからMongoDBへと切り替えることにした。

  MongoDBとはオープンソースのドキュメント指向型NoSQLで、開発元は10gen社（現MongoDB Inc.）である。

  この切り替えにより、大量に発生するデータの運用を格段に容易にし5年間のゼロダウンタイムを実現している。オンラインゲーム「Tomb Raider」では1日に0.5TBのデータが発生し、MongoDBでなければ対応が不可能であったと言及している。

データベースの活用事例

- **ECサイト向けサービス**

ECサイト向け事業者に対して、訪問者の購買行動を分析し販促サービスを提供する会社がデータストアとしてMongoDBを採用した。

ECサイトに訪問した一人一人の動作（マウスの動き、スマートフォンのジェスチャー等）を詳細なデータにしてデータベースへの格納を考えていた。しかし、取り扱われるデータ量は膨大な量となりRDBMSでは安定した運用を行うことができないことが判明した。

そこで、
◆ スキーマレスにデータが扱えること
◆ RDBMSと遜色なく柔軟なクエリが組めること
◆ 高速な読み書きとサーバーの増加を容易に行えること
からMongoDBを採用した。

これによりサーバー1台で1,000以上のアクセスが1秒間に集中しても耐えることができ、負荷に耐えられない場合は、サーバーを増やすだけでよくなった。
また、サーバーを増やした際に1台は障害対応用に、1台はバッチ集計になど柔軟な構成による運用を行うことができた。