

100 万イベント/秒のアクセスに耐える統合センサネットワークのための分散データベースシステム

末松 直樹[†] 宮崎 敏明[†]

会津大学大学院コンピュータ理工学研究科[†]

1. 序論

センサネットワークは物理環境の観測に広く普及し、その観測データの蓄積にはリレーショナルデータベース (RDB) などが用いられている。しかし、大規模センサネットワークにおいては単位時間あたりに生成されるデータ量が膨大となることから、従来型の RDB では全ての観測データを保存することは難しい。近年、大規模データを扱うための分散 DB が幾つか提案され、実用化されている。本稿では多数の独立したセンサネットワークを統合したシステムにおいて、分散 DB の一つである Cassandra[1] が有用であることを示し、実際に計 100 万イベント/秒のデータ蓄積が可能であることを示す。

2. 大規模統合センサネットワークにおける DB に対する要求項目

本稿では、日本各地に配置した複数の独立したセンサネットワークを、インターネットを介して統合した大規模統合センサネットワークシステムを想定する。本システムで用いられる DB では、センサ数が増えるほど、またセンシング間隔が短くなるほど、単位時間あたりの書き込み要求件数が増加する。以下に大規模統合センサネットワークの DB に対する要求項目を列挙する：

- 分断耐性：想定するセンサネットワークは地理的に広範囲をカバーするシステムであり、障害等によるネットワーク分断が懸念される。もしネットワークが分断されても、個々の DB は稼働し続ける必要がある。また、ネットワークの分断ではなく、システムの一部が故障した、または、故意に取り外された場合であっても、個々の DB は動作し続ける必要がある。これは「単一障害点を持たないこと」または「着脱性の確保」と同意である。
- 容量拡張性：ストレージ容量が足りなくなった場合でも、システムを止めずに容易に容量拡張できることが望ましい。

- アクセシビリティ：センサネットワークから生成される全てのデータを書き込むのに十分な性能が必要である。本稿では、ネットワーク接続された 100 万個のセンサが毎秒データを取得するとして、最大で計 100 万イベント/秒の書き込み要求を満たすことを目標とする。

3. 望ましいデータ分散戦略

現在広く普及している RDB は、基本的にマシン 1 台の単体システム構成であり、性能を上げるには、ハードウェアをより高性能なものに交換するスケールアップ方法となる。しかし、1 台のマシンで DB システムを構成することは前章で述べた単一障害点を持つという意味で望ましくない。よって、本稿では複数マシンを使った分散 DB システムを候補とする。まず、大規模統合センサネットワークにおいてどこにデータがストアされるべきかを論じてから、好適な分散 DB の構造を提案する。

分散 DB システムにおけるデータのストア戦略には、以下の 2 つが考えられる。

- 分散ストア：あるセンサネットワークのデータは DB システムを構成するマシン集合（以下 DB クラスタ）内のいずれかのマシンにランダムに分散して保存される（図 1. 分散ストア）。

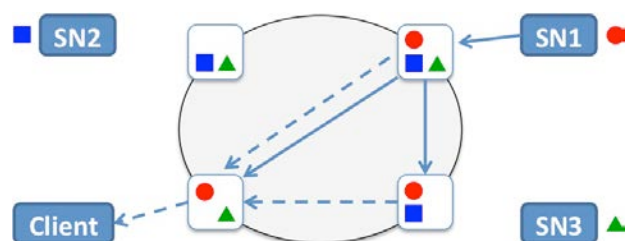


図 1. 分散ストア

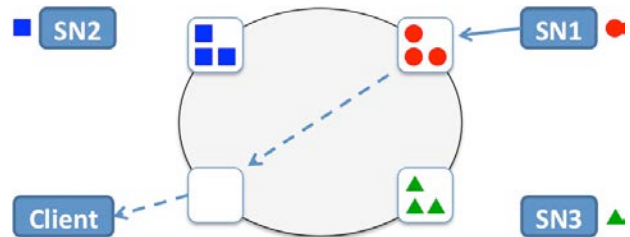
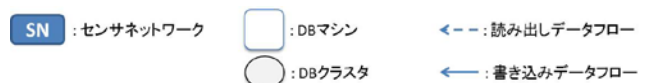


図 2. 近傍ストア



A Distributed Database System for an Integrated Sensor Network to Handle One Million Events-per-second Data Accesses

[†]Naoki Suematsu, [†]Toshiaki Miyazaki

[†]Graduate School of Computer Science and Engineering, The University of Aizu

- 近傍ストア：あるセンサネットワークのデータは近傍の固定 DB マシンに保存される(図 2. 近傍ストア)。

ネットワーク分断耐性・単一障害点・着脱性の確保の観点では近傍ストアが望ましい。近傍ストアでは、一部のセンサネットワークが、近傍の DB マシンと共に、システムから取り除かれても、他のセンサネットワークのデータはシステムから消失することはない。

4. 好適な DB の提案と評価

Apache Cassandra は、P2P(Peer to Peer)ネットワーク技術を基盤としたオープンソースの key-value 型分散 DB 管理システムである。Cassandra はデータをストアするマシンの決定方法をメソッドとして選択できる。あるメソッドを用いることで、前述した近傍ストアを実現できる。

以下に、近傍ストアの設定をした Cassandra クラスタを用い、計 100 万イベント/秒のデータ書き込みを試みた検証結果を示す。表 1 に示した仕様の物理マシン 30 台に XenServer[2]をインストールし、各マシンに表 2 に示す仮想マシンを 4 台ずつ起動させた。これら計 120 台の仮想マシンを用い、図 3 に示すように、60 台の仮想マシンを書き込みクライアントとし、別の 60 台の仮想マシンで前述した Cassandra システムを構成し、検証環境を構築した。次に、計 100 万イベント/秒以上となるよう、各書き込みクライアントから 16668 イベント/秒の同時書き込みを 6 時間 30 分行った(16668 イベント/秒×60 台≧100 万イベント/秒)。評価結果は下記の通りである。全ての仮想マシンにおいて 16668 イベント/秒の継続書き込みが可能な持続時間が存在するが、その持続時間は各仮想マシンで異なった。また、当該持続時間は、同一仮想マシンであっても試行毎に異なった。さらに、書き込み序盤はどのホストも 1 秒前後の速度で書き込めていたが、徐々に書き込み速度が低下した。ある試行において、4 秒以上の書き込み時間を記録した後 1 回でも 1 秒以内の書き込みができたホスト、すなわち、書き込み速度の回復があったホストの割合は 41/60 であった。一方、5 秒以上の書き込み時間を記録した後、書き込み速度の回復があったホストの割合は 22/60 であり、ホスト全体の半数以下となった。図 4 は、計 60 台の各ホストにおいて、5 秒以内の書き込み完了が可能な持続時間をプロットした分布図である。最も継続書き込み時間が短かったマシンでも 26 分間の継続書き込みが可能であったことが分かる。これは、本検証システム環境において、計 100 万イベント/秒の 5 秒以内の書き込みが少なくとも 26 分間可能であったことを示している。

表 1. 物理マシン 1 台の仕様

ハードウェア	
CPU	Intel® Xeon® CPU E3-1240 V2 @ 3.40GHz (4 コア 8 スレッド)
メモリ	8 GB
HDD	500 GB、7200rpm
ソフトウェア	
OS	XenServer 6.2.0

表 2. 仮想マシン環境

ハードウェア割り当て		ソフトウェア	
CPU	2	OS	CentOS 6.6
メモリ	1700 MB	JDK	1.7.0
HDD	100 GB	Cassandra	2.1.2

書き込みクライアント
を起動

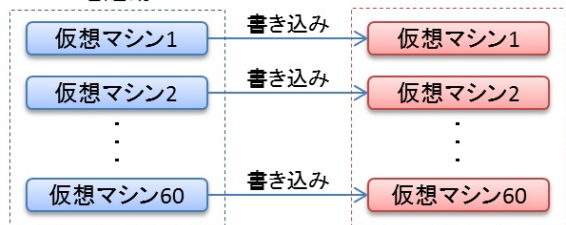


図 3. データ書き込み検証環境

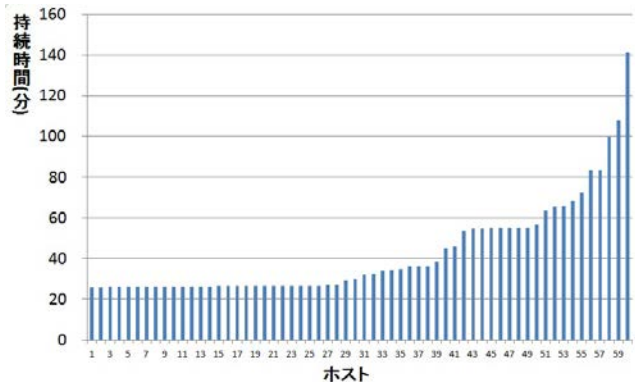


図 4. 各ホストにおける 5 秒以内の書き込み完了が可能な持続時間

5. 結論

本稿では、複数のセンサネットワークを統合して一つのシステムとした大規模統合センサネットワークに好適な分散 DB 構成の提案し、検証実験により計 100 万イベント/秒の書き込みができる見通しを得た。

謝辞 本研究の一部は、総務省戦略的情報通信研究開発推進制度 (SCOPE No. 121802001) の支援を受けて実施したものである。

参考文献

- [1] Apache Cassandra, <http://cassandra.apache.org/>
 [2] Citrix XenServer, <http://www.citrix.co.jp/>