

ベンチマークレポート

ーデータグリッド編ー

平成 22 年 9 月

グリッド協議会 先端金融テクノロジー研究会 ベンチマーク WG

《 目 次 》

1.	まえがき	1
2.	背景と目的	1
3.	経緯	1
4.	ベンチマーク環境	2
5.	データグリッドベンチマーク評価シナリオ	3
5.1	データグリッドベンチマークの概要	3
5.2	評価シナリオ	4
5.2.1	評価シナリオ0:事前テスト.....	4
5.2.2	評価シナリオ1:スケーラビリティ.....	5
5.2.3	評価シナリオ2:スループット.....	5
5.2.4	評価シナリオ3:マーケットデータフィード.....	6
6.	考察	9
7.	実施者	9

1. まえがき

本書は、グリッド協議会金融分科会ベンチマーク WG で実施した金融グリッドベンチマークのデータグリッドに関する評価の内容および結果に関する報告書である。

2. 背景と目的

欧米の金融機関では早くからグリッドコンピューティングが導入されていたが、それに比較して日本の金融機関での導入はまだ多くはない。国内の金融機関による導入促進と、ベンダーやインテグレータによるソリューション構築の支援が課題である。

そこで、金融業務に対してグリッドを適用する際に、アプリケーションの粒度やデータの配分・配置などを検討する材料となることを目的として、典型的な業務をモデル化したベンチマークプログラムを作成し、実環境での測定を行う。ただし、今回の測定結果は利用した環境上で得られた数値であり、他の環境にて測定した場合には異なる数値が得られるので注意されたい。

3. 経緯

グリッドの適用が容易と考えられる業務アプリケーションの一つであるリスク分析のモデル化から始めた。リスク分析業務のアプリケーションでは、市場リスク、信用リスクの他様々なリスクを、市場のデータや自社のデータを基に、モンテカルロ法に従って多数回の数値評価から統計的な処理によってリスクの指標を算出する。図 3-1 に示すように、一つのジョブがいくつかの DB から情報を読み取り、それぞれにデータを割り振り、複数のタスクを生成する。タスクは、タスク毎に異なるデータと、タスクに共通なデータを読み取り、割り当てられた計算処理を実行して、その結果を戻す。最終的な統計処理はジョブが最後に実施するが、実行時間のほとんどは、タスクの処理に使われる。

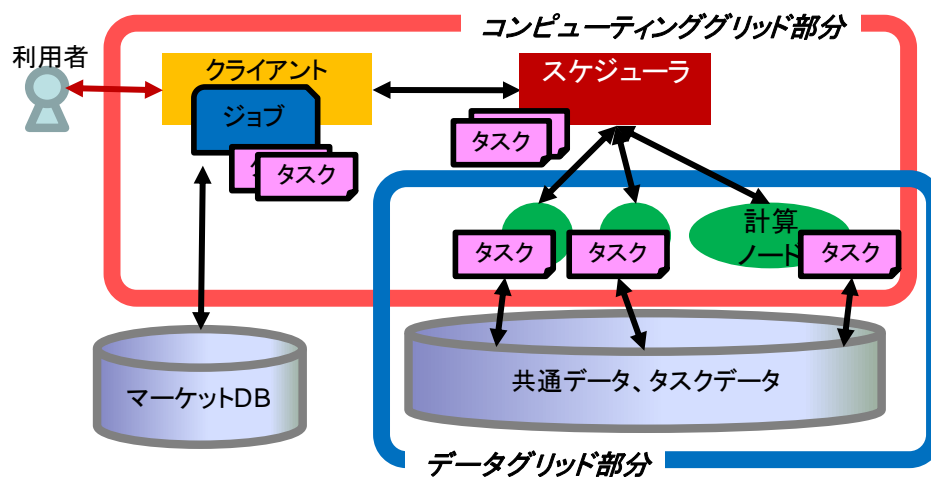


図 3-1 リスク分析の構成とベンチマークの範囲

リスク分析を、スケジューラがタスクを分散するコンピューティンググリッド部分と、タスクがデータにアクセスするデータグリッド部分とに分離し、それぞれにエッセンスを評価することにした。本書では、データグリッドの部分

について評価した結果について記述する。コンピューティンググリッド部分に関するベンチマークの結果については、ベンチマークレポート-コンピューティンググリッド編-を参照されたい。

また、データグリッドを活用することが期待されるマーケットデータフィードについても評価を行った。マーケットデータフィードとは、東証などの取引所が出力する売買情報、情報ベンダーが複数の国内/海外取引所などから取得した売買情報を統合して出力する情報の総称である。主に TCP/IP でリアルタイム出力され、1メッセージが数百バイト程度であり、1秒間に数百から時に数千、万メッセージの出力となる場合がある。マーケットデータフィードを受信処理するシステムは、連続的に発生するメッセージの受信イベントに合わせて、遅延なく処理することが求められる。

4. ベンチマーク環境

ベンチマークを実行する環境として、新日鉄ソリューションズ株式会社のグリッド・ユーティリティ・コンピューティングの検証施設 NS Solutions Grid/Utility Computing Center (NSGUC)を使用した。実験を実施した期間は2回。一度目は2008年10月から11月の4週間コンピューティンググリッドのベンチマーク測定と同時期に行った。コンピューティンググリッドのベンチマーク測定の後半に余裕ができたため、並行してデータグリッドの事前テストを実施した。この期間では、IBM および HP のブレードサーバから、最大400超コアのサーバを使用した。また、2009年7月6日からの1週間と7月27日からの1週間の合計2週間では、データグリッドのベンチマーク測定の本番を実施した。一週目は、IBM のブレードサーバ14ノード2筐体を使用し、二週目は、IBM のブレードサーバ14ノードを3筐体使用した。

また、今回のデータグリッドベンチマークでは、以下のソリューションを対象とした。評価結果は、用いたソリューション毎に述べる。なお、Coherence の測定においては日本オラクルとは関わりがなく、協議会に参加するユーザ企業の所有するライセンスを利用し、本WGが独自に評価を行ったものである。

表 4-1 実装ソリューション

ソフトウェア名	ベンダー	実装言語
Coherence	日本オラクル	Java
WXS: WebSphere eXtreme Scale	IBM	Java
Caché	インターシステムズ	Java

5. データグリッドベンチマーク評価シナリオ

5.1 データグリッドベンチマークの概要

図 3-1 に示すような業務を実行する際に、データグリッドとしては、図 5-1 に示すように大きく分けて三通りの構成に分けることができる。一つ目は、タスクの処理を、データを保持するサーバ上で行い、同一サーバのデータのみを参照するように割り当てるものである。この場合が最も高い性能を引き出すことが可能であるが、どのようにデータを割り当てるかは、アプリケーションの内容にも依存する。実装ソリューションによっては、タスクをデータの存在するサーバに自動で割り当てるようにするものがある。二つ目は、タスクの処理を、データを保持するサーバ上で行うが、異なるサーバのデータも参照する場合である。この場合、別のサーバにアクセスすることから、性能は一つ目の場合に比べて落ちることになる。これら二つの構成は、タスクとデータ間の関連に強く依存するため、全てのデータグリッド実装ソリューションで実現できるわけではない。三つ目は、データは全てデータグリッド専用サーバで保持し、CPU 処理専用サーバからアクセスする場合である。常にネットワークを介してデータにアクセスすることになり、データをどこに置くかにはあまり依存せずに実装することができる。また、どのデータグリッド実装ソリューションでも、この構成をとることは可能である。そこで、データグリッドベンチマークでは、三つ目の場合での性能を評価するものとする。

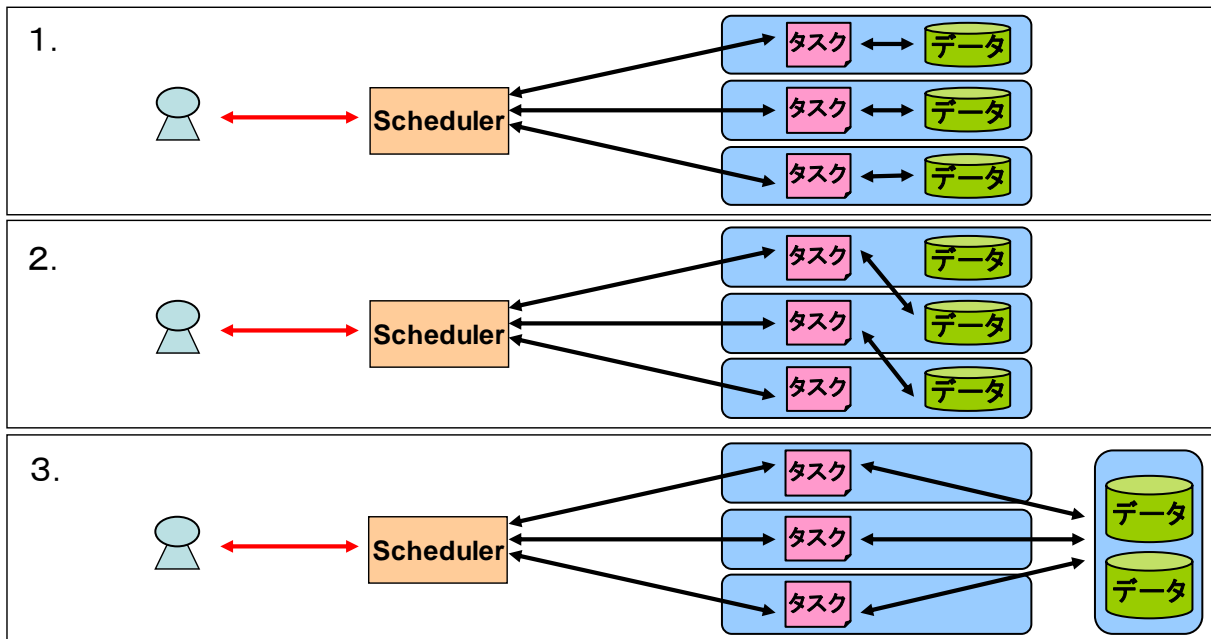


図 5-1 データグリッドの主な構成方法

データグリッドのベンチマークを行う環境は、基本的に図 5-2 に示す構成となる。データを保持するデータノードと、データノードにアクセスしてデータを読み書きするクライアントから成る。図 5-1 に示すタスクは、このクライアント上で実行される。この構成において、ベンチマークで可変なパラメータは、表 5-1 に示すように、データノード数、クライアント数、データオブジェクトサイズ、全データサイズの四つと考えられる。ただし、データグリッド上の全データサイズは、データアクセス性能にはほとんど影響しないため、本ベンチマークにおいては変化させなかった。

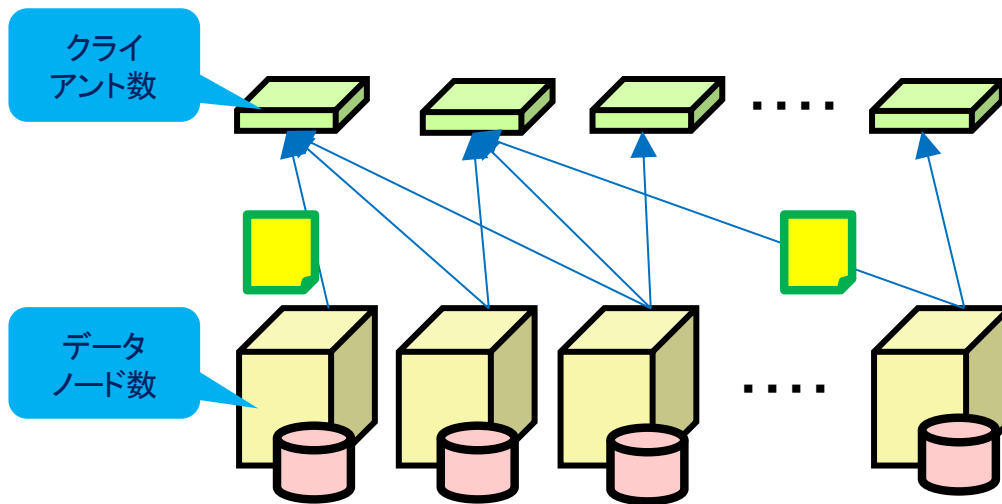


図 5-2 データグリッドベンチマークの構成

表 5-1 データグリッドベンチマークの可変パラメータ

パラメータ	内容	単位
データノード(コア)数	データグリッドを保持するノード(コア)数	個
クライアント数	データグリッドにアクセスするクライアント数	個
データサイズ	データグリッド上に存在するデータの総量	KB
オブジェクトサイズ	データグリッド上に存在するデータオブジェクトのサイズ	KB

以下に述べる評価シナリオにおいて、マーケットデータフィールド以外では、これらのパラメータを変化させつつ性能を評価する。

5.2 評価シナリオ

以下、いくつかの評価シナリオを述べる。ただし、評価を行った実装ソリューションそれぞれが全ての評価シナリオを実行したわけではなく、ソリューション毎に評価したシナリオは異なる。評価を行ったシナリオは、ソリューション毎の報告を参照のこと。

5.2.1 評価シナリオ 0 : 事前テスト

ベンチマーク測定を 2008 年と 2009 年の二回に渡って実施したが、2008 年は、事前テストという位置づけで以下のパラメータに対して性能を測定した。

- 1) データノード数 8 固定、クライアント数 100 固定、オブジェクトサイズ (1KB, 100KB, 1M, 10M)
- 2) データノード数 (1, 4, 8)、クライアント数 100 固定、オブジェクトサイズ 1KB
- 3) データノード数 8 固定、クライアント数 (1, 10, 100)、オブジェクトサイズ 1KB

いずれの場合もアクセスは Read /Update /Write の 3 種類で、それぞれ 1000 回繰り返す。ここで Update は Read を行った後に Write を行う処理を指す。ソリューションによっては Update 専用のインタフェースが用意されているため、個別にテストを実施した。

5.2.2 評価シナリオ 1 : スケーラビリティ

この評価では、ベンチマークパラメータに対する各ソリューションのスケラビリティを調べる。ただし、オブジェクトのサイズは通常大きな値を取ることはなく、また、実行時間はほぼオブジェクトのサイズに比例することが分かっているため、ここではオブジェクトのサイズを 128 バイトに固定した。すなわち、評価シナリオ1では、データノード数を固定してクライアント数を変化させた場合、およびクライアント数を固定してデータノード数を変化させた場合の二通りを実施した。

クライアント数は 1,000 個まで増やして測定を行うこととした。ただし、1,000 台のサーバを用意して測定することは困難であるため、1 台の物理サーバ上にクライアントの役割を果たすタスクを複数立ち上げて行った。

データオブジェクトは、1データ 128 バイトであるが、データグリッド上の総量は 1GB となるだけオブジェクトを生成した。1GB のデータを、使用するデータノードに分担して保持する。また、データグリッドでは、全てのデータが Replication を1つ作成しているものとする。従って Replication を含めると全データグリッドの総量は 2GB となる。データグリッドではデータに冗長性を持たせるため、全てのデータについて複製をひとつ用意する。従って複製を含めたデータ総量は 2GB となる。また複製を作成しないケースについても、特定のパラメータセットについて測定し、両者の差を評価する。

クライアントがアクセスするデータオブジェクトはランダムに設定するものとする。1GB には 8 百万のオブジェクトが存在するため、複数のクライアントでアクセスするオブジェクトが重複する可能性は極めて低い。クライアントからのアクセスは Put および Get のみとする。

評価シナリオ1では、二通りの評価を行う。一つはデータノードを 2 および 8 個と固定し、クライアント数を 1, 10, 100, 1,000 と変化させ、クライアント数に対するスケラビリティを評価する。他方は、クライアント数を 100 および 1,000 と固定し、データノード数を 1,2,4,8 と変化させ、データノード数に対するスケラビリティを評価する。

5.2.3 評価シナリオ 2 : スループット

この評価では、データグリッドへのアクセス性能としてスループットを評価する。すなわち、データグリッドにアクセスするクライアントを増やして行き、単位時間あたりの処理量が頭打ちとなる様子から限界を測定する。データノード数が変われば、処理可能量も変わるので、条件毎に測定を行う。このスループットは、データノードが持つ、データアクセス能力を測るものであり、データノード数に比例して能力が高くなることが期待される。

データのオブジェクトサイズは、シナリオ1と同じく 128 バイトとする。クライアントからデータオブジェクトへのアクセスは、Put および Get のみとする。データグリッドのソリューションの中には複数の Put/Get 命令をまとめて発行することでオーバーヘッドを減らして高速化するものがある。しかし、ここでは、1オブジェクト毎にアクセスするものとする。

また、シナリオ1と同様、全てのデータが **Replication** を1つ作成しているものとする。

評価シナリオ2では、データノード数を 2, 4, 8 台に固定し、クライアント数を増加させた際の応答からスループットを評価する。

5.2.4 評価シナリオ3：マーケットデータフィード

この評価シナリオでは、ベンチマークとしての性能評価を目指すのではなく、特定のシステム構成が期待する性能を発揮できるものかどうかを検証する。東証における株情報のデータフィードでは、将来的に秒1万件もの情報送信が行われると予想されている。このような情報送信に対して、どのような構成のデータグリッドを配することにより対応可能であるかを確認するためのシナリオである。

データグリッドには、約 3,000 銘柄（実際には 3,322 銘柄）の株データを保持する。各データは、**key**、**value**、**dummy data** から構成される。**key** は 9 バイトの文字列データで株の銘柄、**value** は 4 バイトの整数で株価の値、**dummy data** は 124 バイトの **byte array** でその他のデータ、をイメージしている。

データの更新には、とある日の朝 9 時前後における実際のデータを使用した。30 分間で約 27 万件のデータが配信されている。模擬的にデータの配信密度を高めるため、実際よりも 10 倍速および 20 倍速で配信を実施する。データノードと別のサーバをフィーダとして設定し、このデータに基づいて指定時刻になったらデータノード上の値を更新する。完全に指定された時刻に値を更新することは困難なため実際には遅延が生じることとなる。

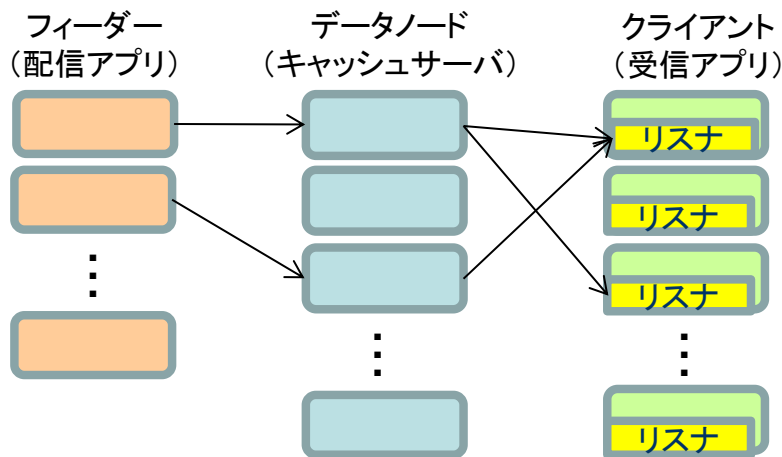


図 5-3 シナリオ3におけるシステムの概略構成図

データの更新は、データグリッドの機能によりリスナに通知される。リスナは 300 用意し、それぞれが 10 銘柄ずつ監視している。配信からリスナが受信するまでの遅延を確認し、使用するデータノードの台数との関係を明らかにする。

図 5-4 に、今回用いたフィードデータの更新頻度を示す。10 秒毎に更新された件数をプロットしている。9 時に大きなピークがあるが、それ以前にも、何度か取引が発生している。

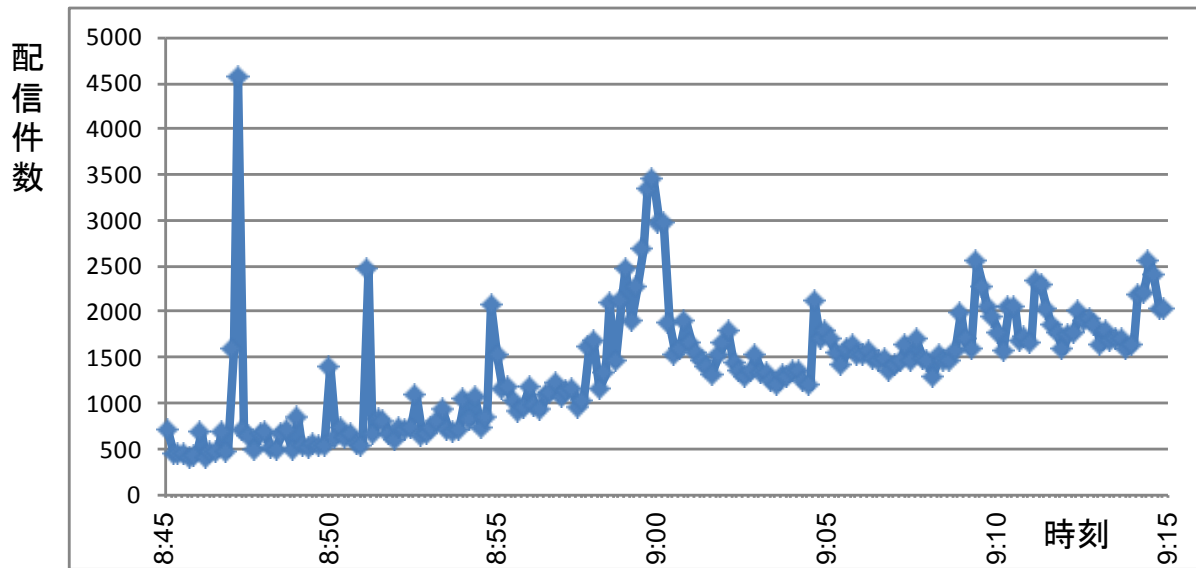


図 5-4 時刻毎の配信件数(10 秒毎)

10 秒毎に見た場合に配信件数が最も多いのは 8 時 47 分 10 秒の約 4,500 件、次が 9 時の約 3,500 件である。配信間隔を 10 分の 1、および 20 分の 1 にした場合、配信の密度は 4,500 件/秒および 9,000 件/秒となる。10 秒の間で配信密度のバラツキがあるため、瞬間的には 1 万件/秒を超える配信も発生する可能性がある。また、配信件数の多い銘柄の順に配信件数をプロットしてみたのが、図 5-5 である。この 30 分間で、最も多い配信のあった銘柄は 1 つで 1,680 件の配信が行われた。また、604 の銘柄については、この期間中配信は存在しなかった。

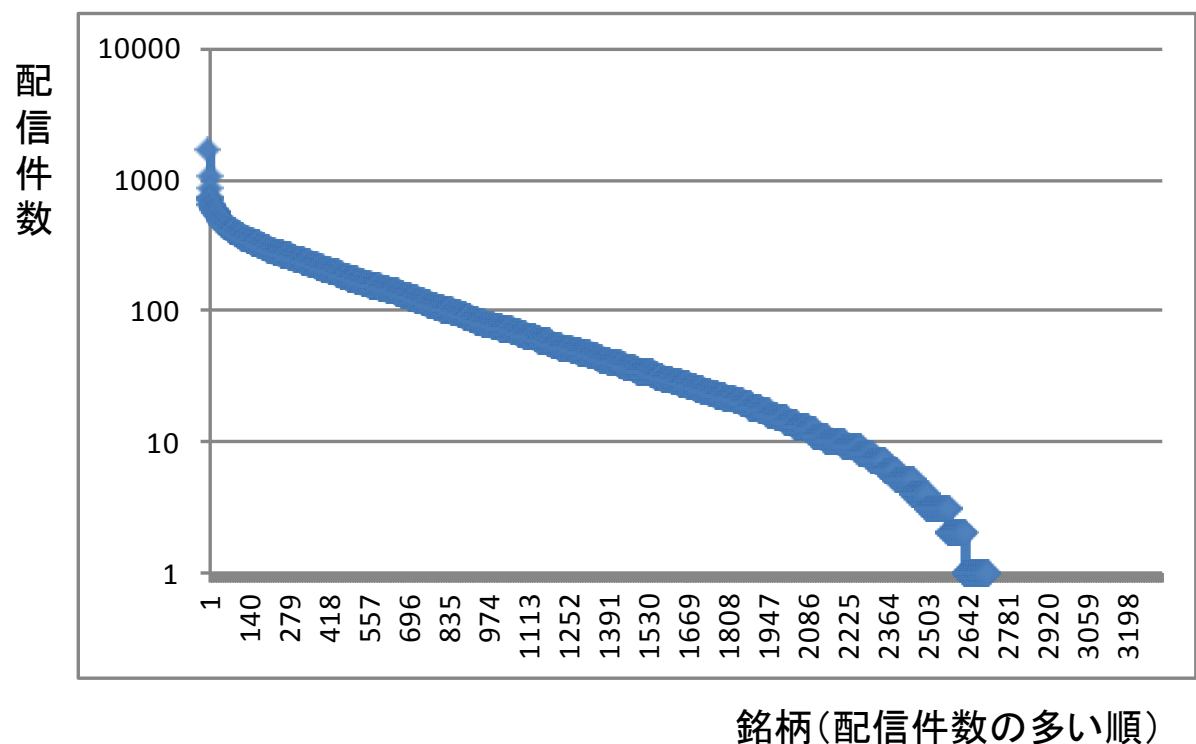


図 5-5 銘柄毎の配信件数(配信件数の多い順に並び替えた)

これらのデータを配信するフィーダについては、余裕を持って 8 台のサーバを用いることとした。本シナリオで検証したいのは、データグリッドの更新とリスナへのメッセージ送信の部分であるため、データの配信部分がボトルネックにならないことを懸念しての処置である。なお、配信するデータをフィーダに分配する方法は、配信件数の多い順にサイクリックに割り当てる。ただし、割り当て量をなるべく均等にするため、奇数回目と偶数回目では、逆並びとする。具体的には、表 5-2 の通りである。

表 5-2 配信データのフィーダへの割り当て方法

配信件数の多い株価の順	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	...
分担するフィーダの番号	1	2	3	4	5	6	7	8	8	7	6	5	4	3	2	1	1	2	3	...

その結果各フィーダが配信する件数は以下の通りとなる。配信件数のバラツキは 3%程度であり、よく分配されて、配信の性能上問題はないと思われる。

表 5-3 フィーダ毎の配信件数

フィーダの番号	1	2	3	4	5	6	7	8
配信件数	35240	35340	34432	34302	34259	34017	33784	34438

一方、リスナは、実際の業務での状況を模して、300 用意する。それぞれが 10 件の銘柄を持ち受け、値の更新（配信）があった場合にデータグリッドからメッセージが飛ぶようにする。各リスナが待ち受ける銘柄は、実際の業務を模して決定した。具体的には、各リスナはまず配信頻度の多い順に 25 位までの銘柄から 2 つ選択する。ちょうど 300 通りの選択方法がある。次いで順位の 26 から 825 までの 800 銘柄から 100 おきに 8 銘柄選ぶ。このように選択された 10 銘柄をそれぞれのリスナが担当し、データグリッド上に配信（更新）された場合に、クライアント上のリスナに通知する。

6. 考察

各ソリューションの評価結果は、個別の冊子を参照のこと。

今回の評価では、共通のシナリオを設定したが、ソリューションの有する機能が異なるため、一つの尺度で比較することは難しい。しかし、いずれのソリューションの場合でも、規模に制約はあるものの、使用した台数に応じた性能が得られていることが確認できた。また、性能のボトルネックはプロセッサよりもネットワークで生じやすいので注意が必要である。

7. 実施者

本ベンチマークの実施にあたって、以下の方々にご協力頂いた。

氏名	所属	分担
阿部 憲幸	元日本オラクル株式会社	仕様策定
池上 努	産業技術総合研究所	仕様策定、測定結果分析、報告書作成
池田 佳弘	みずほ証券株式会社	仕様策定
伊藤 智	産業技術総合研究所	仕様策定、測定結果分析、報告書作成
伊藤 秀和	三菱UFJ証券株式会社	仕様策定、測定結果分析、報告書作成
伊藤 宏樹	新日鉄ソリューションズ株式会社	測定環境用意
大嶋 裕子	産業技術総合研究所	事務局
沖原 久憲	QUICK 株式会社	仕様策定、実験データ提供、報告書作成
神社 純一郎	新日鉄ソリューションズ株式会社	測定環境提供
北田 顕啓	プラットフォーム コンピューティング株式会社	仕様策定
白坂 純一	野村証券株式会社	仕様策定
谷口 肇	三菱UFJ証券株式会社	仕様策定
田村 圭	株式会社大和総研	仕様策定
田村 俊明	GigaSpaces Technologies	仕様策定
中台 慎二	日本電気株式会社	仕様策定
中田 秀基	産業技術総合研究所	仕様策定
野村 俊朗	野村証券株式会社	仕様策定
堀之内 道仁	元データシナプス株式会社	仕様策定
根岸 史季	日本アイ・ビー・エム株式会社	仕様策定
広瀬 哲也	日本電気株式会社	仕様策定
堀田 稔	インターシステムズジャパン 株式会社	仕様策定、環境構築・測定、測定結果分析、報告書作成
山本 学	日本アイ・ビー・エム株式会社	仕様策定、環境構築・測定、測定結果分析、報告書作成