



様々なストレージおよびネットワークテクノロジーにおける SAP IQ 16 Multiplex のパフォーマンス評価

2014年4月

目次

まとめ	3
この調査の動機	3
ソリューション概要	4
SAP IQ 16 Multiplex	4
BAY STORAGE テクノロジー — BAYSTOR	5
ストレージ システム テクノロジーの比較	7
ベンチマーク プラットフォーム.....	9
測定	10
テスト シナリオ	10
ワークロード.....	11
結果	12
ストレージ デバイス	12
ストレージ ネットワーク 帯域幅	14
BAYSTOR アクセラレーション	15
結論	17
リソース.....	17
謝辞	18
APPENDIX A — システム構成	19
APPENDIX B — テスト結果の分析	19
ロード フェーズ.....	19
クエリー フェーズ	20

まとめ

SAP IQ は、卓越したスケーラビリティとパフォーマンスを発揮するカラム型の分析用データベースです。Simplex (シングル サーバー) の構成でも、Multiplex (サーバーのクラスター) の構成でも利用が可能です。IQ Multiplex は「シェアードエプリシング」と呼ばれる超大規模並行処理アーキテクチャーで、クラスター内の全てのサーバーが同時にアクセスする共有データベースを持つのが特長です。共有のメインデータベースと共有のテンポラリーストレージを使用するため、一般的なストレージベンダーが提供している高コストの大規模共有ストレージの使用を避けたいという要望が多くのお客様より寄せられました。BayStor は、ネットワーク接続したストレージ (Network Attached Storage : NAS) のソリューションを、ローカル接続したストレージ並みのパフォーマンスで提供するコスト効果の高い魅力的なソリューションでした。

このホワイトペーパーでは、DQP (Distributed Query Processing : 分散クエリー処理) が有効になっている時に、I/O が集中する並列データベースのロードとクエリーストリーム処理を行う IQ Multiplex のパフォーマンスに関して、様々なストレージテクノロジーと関連するストレージネットワーク帯域幅の選択の意義について展開します。調査対象のストレージテクノロジーには、SSD アレイ (フラッシュ) やいくつかのエンタープライズクラスのハードドライブアレイを含みます。

この文書では、また、BayStor コントローラーの独自のキャッシング機能により IQ Multiplex のパフォーマンスがどう加速するのかについても説明します。

このホワイトペーパーで示される結果は、I/O が集中するロードとクエリーのワークロードをベースにしています。この調査の結果、IQ は、I/O のパフォーマンスをストレスシェアするよう構成されているワークロードであっても、ハイレベルな CPU 使用効率を発揮すると結論づけました。そのため、同様なワークロードにおける IQ Multiplex の合理的なディプロイメント戦略とは、超大規模なデータセットをホストしながらも、広い帯域幅を提供できるエンタープライズクラスのハードドライブテクノロジーを使用することで、共有ストレージシステムのコストを効果的に抑えつつ、多くの CPU リソースを割り当てるといったことになります。

この調査の動機

2013年秋、Bay Storage Technology と SAP は、IQ に関連する 2 つの利用ケースに関して、低レイテンシー/高帯域幅のストレージを使用する共同ホワイトペーパー¹を発行しました。このホワイトペーパーの結論の 1 つが、Bay Storage の NAS は、ストレージディスクアグリゲーションのメリットを提供しつつ、ローカル接続のストレージ並みのパフォーマンスを発揮するという点でした。

このホワイトペーパーが出たことで、IQ の分散クエリー処理機能のメリットを発揮する大規模な IQ Multiplex 分析ディプロイメントに向いているコスト効果の高いストレージの有効性を評価してほしいというリクエストがお客様より寄せられました。この評価は、それぞれ 100 TB から 1 PB のデータセットのディプロイメントを視野に 10 TB のデータセットで行われました。

お客様からリクエストされたもの以外にも、このホワイトペーパーの調査には、いくつかの動機がありました。

- 以前のホワイトペーパーで使用された BayStor リリース 2 から追加された機能を含む BayStor リリース 3 の効果の調査です。特に、新しい BayStor マルチホスト機能では、高速パフォーマンスの BayStor コントローラーを IQ Multiplex で使用できるため、ネットワーク接続の共有ストレージでメインストアとシェアード Temp のホスティングがサポート可能になっていました。
- 以前の調査では、CPU の使用率が 100% に達することが時々あったため、制約はストレージのパフォーマンスのみなのか調べるため、ベンチマークテストの実施の要望がありました。
- 今回は、最初の調査で使用された DRAM やフラッシュテクノロジーに加えて、様々なエンタープライズハードドライブのパフォーマンス評価をする機会がありました。

この調査のテクニカル面での目的をまとめると以下のとおりです。

- IQ Multiplex の I/O ワークロードの特徴を理解する
- IQ Multiplex のパフォーマンスがストレージデバイスの違いによってどう変化するのか理解する – フラッシュ、エンタープライズハードドライブ、RAM ディスク
- 共有ストレージの IQ Multiplex ワークロードをサポートするために必要なネットワーク帯域幅を理解する

ソリューション概要

SAP IQ 16 Multiplex

SAP IQ は、高速パフォーマンスを発揮する分析データベースエンジンです。世界初の商用カラムストレージエンジンであり、オンライントランザクション処理よりも複雑なクエリーと分析、および大幅なデータ圧縮にフォーカスした最初の製品です。特許取得済のカラムインデックス付けサブシステムを有し、特にアドホックなクエリー処理に優れています。標準 ODBC や JDBC インターフェースをサポートしたオープンなプラットフォームで、豊富な SQL 言語を提供し、一般的なハードウェアおよび様々なオペレーティングシステムで稼働します。また、データベースはディスクに保持する一方、大きなキャッシュを持つことで、できる限り多くのデータをインメモリで展開します。SAP IQ データベースは、以下の3つのストアセットで構築されています。メタデータを保持するカタログストア、ユーザーデータを保持するメインストア、中間結果を保持するテンポラリストアです。「シェアードテンポラリストア」は、CPU 集中型のクエリーがフラグメントに分割され、サーバーで並列処理される、分散クエリー処理（DQP : Distributed Query Processing）をサポートするために使用されています。また、このシェアードテンポラリストアは、分散クエリー処理中に中間結果をサーバー間にパスするために使用されています。ネットワークの相互接続の速度が十分であれば、IQ は、シェアードディスクのかわりにネットワークを利用して DQP を行うことができます。それぞれのストアは、「DBSpace」と呼ばれる論理的データコンテナのセットとして実装されます。IQ は「Simplex」と呼ばれる単一のサーバー構成でも稼働できる一方、スケールアウト処理には、「Multiplex」と呼ばれるグリッドディプロイメントオプションが用意されています。Multiplex は、「シェアードエブリシング」構成で大規模な並列処理を行います。ユーザーのクエリーは、Multiplex クラスタ内の全ノードに分配され、コンカレントクエリーがそのシステム内に入出力されると、IQ サーバーは動的に CPU とメモリーの使用を再調整します。Multiplex では、1つのノードがコーディネーターになり、書き込み可能なトランザクションのトランザクション ID を管理します。他のノードは、リーダー（クエリーの実行のみ）にもリーダー/ライター（クエリーの実行またはトランザクションのアップデート）にもなることができます。同じサーバー環境内で、ワークロードを他から分離する必要があるマルチテナントのサポートが必要な場合には、ノードを「論理サーバー」と呼ばれるエラスティックなセットにグループ化することもできます。SAP IQ のシェアードエブリシングアーキテクチャーは、「データアフィニティ」機能で強化されており、このデータアフィニティ機能により、クエリーエンジンは、ワークをインテリジェントにパーティション化し、関連データがすでにキャッシュにあるノードに分散クエリー処理を割り当てます。キャッシュは、ワームのまま保たれるため、シェアードストアのロードが低減します。

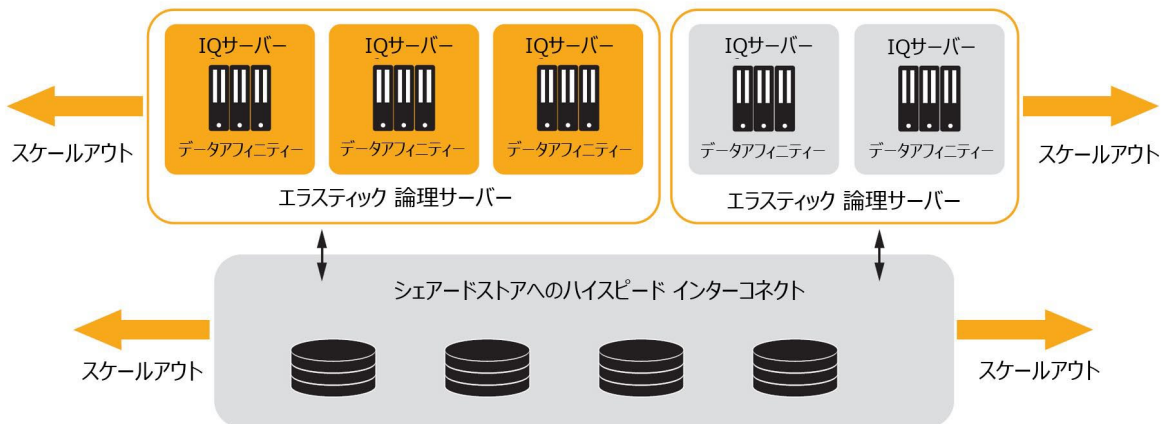


図 1. SAP IQ Multiplex アーキテクチャー

SAP IQ 16 には、RLV (Row Level Versioned: 行レベルバージョン管理) と呼ばれる書き込みが最適化された、インメモリーのデルタストアも実装されています。このRLV ストアでは、データベースへの高速ローディングをサポートしているとともに、以下のメリットがあります。

- オペレーショナルデータの継続的な分析
- 高速なコンカレントデータ変更
- 大容量のメモリとコアのフットプリントの利用

BAY STORAGE テクノロジー — BAYSTOR

BayStor は、ネットワーク接続された DRAM や x86 の一般的なサーバー上の階層化されたストレージにデプロイするための、ソフトウェア・イネーブルドの、革新的なテクノロジーで、超高速パフォーマンスストアとして DRAM を直接使用したり、標準的なストレージデバイスのアクセラレーションキャッシュとして使用することができます。BayStor は、DRAM へのキャッシュの有無にかかわらず、あらゆるストレージデバイスのネットワーク接続に使用することができます。BayStor は、Linux システムからは標準的なブロックデバイスとして見えるため、I/O が集中するアプリケーションでも修正なしにアクセラレーションすることが可能です。BayStor のリリース 3 では、この IQ Multiplex の調査で使用している機能がいくつか追加されています。

- 直接の、マルチホストのサポート。複数の IQ およびアプリケーションサーバーで高速パフォーマンスの共有ストレージが実現できます。
- オプションであるストレージデバイスの DRAM キャッシュ。ローカルデータへのアクセスにメリットがあります。
- オプションである Early Write Acknowledgement。書き込み Acknowledgement レイテンシーにセンシティブなワークロード処理を加速します。この機能は、トランザクションのワークロードにメリットのある機能です。

以下の図は、BayStor のアーキテクチャーの全体像と、帯域幅と IOPS (Input/Output Per Second) でキャパシティをスケールできる機能を表したものです。

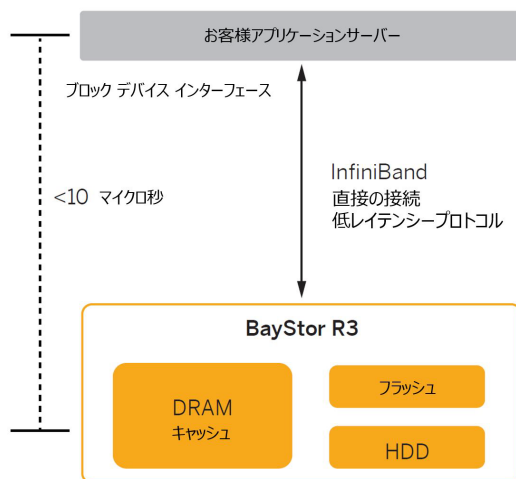


図 2. BayStor アプライアンス アーキテクチャー

BayStor は、IQ Multiplex に対してコスト効果の高い共有ストレージプラットフォームであるだけでなく、R3 の追加機能により、この調査においてさらなるメリットがありました。

- BayStor は、フラッシュ PCIe カード、フラッシュ SSD、ハードドライブ、DRAM など幅広い標準ストレージデバイスをサポートしています。そのため、この調査で使用する様々なフラッシュ、エンタープライズハードドライブ、DRAM デバイスの共通ベンチマークプラットフォームとして、同じ BayStor を使用することができます。
- BayStor は、56 Gbps InfiniBand をサポートしているため、十分すぎるストレージ ネットワーク ファブリック ヘッドルームを使用することで帯域幅による制約を除外でき、またどのストレージネットワーク帯域が必要なかを理解することができます。
- BayStor で実現するネットワークの超低レイテンシーにより、ネットワークレイテンシーの影響を除外することができるため、従来のストレージネットワークソリューションにとっては何百マイクロ秒から数ミリ秒ものメリットがあります。この調査において、重要なパフォーマンス要因である SAN ファブリックのレイテンシーは、無視することができます。

ストレージ システム テクノロジーの比較

このセクションでは、評価したストレージテクノロジーについて解説します。全てのストレージは、IQ が直接管理できるよう、個別の「raw」デバイスとして現れるようにしました。

各ノードのローカル Temp は、Samsung SM843 SSD をベアで使用しました。各ドライブは 240 GB のキャパシティーがあり、200 GB にオーバープロビジョニングしてガベージコレクション量を減らすことで、エンデュランスとパフォーマンスを強化しました。これらの 6 Gbps SATA ドライバーは、使用する前に事前調整しました。

以下は、BayStor がホストし、ネットワーク接続された、共有ストレージテクノロジーです。明記されていない限り：

- 各テクノロジーは、単一の BayStor コントローラーでホストしています。これは、ストレージファブリックへ単一の 56 Gbps InfiniBank リンクを使用しています。
- 各テクノロジーの評価は、16 のデバイス（シェアード Temp に 8、IQ メインデータベースに 8 引き当て）の構成で行いました。
- 個々のドライバーのフォームファクターは、2 ½" ドライブでした。

HDD ディスクアレイ

エンタープライズクラスのハードドライブテクノロジー 3 種を評価しました。

- Enterprise Performance 15K RPM、600 GB、6 Gbps SAS（備考：6 Gbps SAS コントローラーに接続）
- Enterprise Performance 10K RPM、1.2 TB、6 Gbps SAS
- Enterprise Capacity 7.2K RPM、1 TB、6 Gbps SAS

SSD アレイ

ドライバーは、256GB Samsung 840Pro SSD を使用しました。ガベージコレクション量を低減し、エンデュランスとパフォーマンスを強化するため 200 GB にオーバープロビジョニングしました。これらの 6 Gbps SATA ドライバーは、使用前に事前調整しています。

BAYSTOR RAM ディスク

あらゆるパフォーマンス特性において、BayStor 独自の RAM ディスクテクノロジーには実質的な優位性があります。4 KB ブロックにおいて 10 マイクロ秒以下のレイテンシー、優れた帯域幅と IOPS を実現します。この評価では、5 つの BayStor RAM ディスクをディプロイしましたが、それぞれのキャパシティーは 350 GB でした。IQ のメインストアに 3、シェアード Temp に 2 使用しました。5 つの BayStor コントローラーのそれぞれを 56 Gbps InfiniBand リンクでストレージファブリックに接続しました。複数のリンクによる追加の帯域幅は、このホワイトペーパーの結果の要因ではありません。

ストレージ パフォーマンス サマリー

各ストレージテクノロジーのパフォーマンスの詳細は以下のとおりです。

	4 KB ランダム						1 MB シーケンシャル	
	レイテンシー (マイクロ秒)		帯域幅 (MB/秒)		IOPS		帯域幅 (MB/秒)	
BayStor	読み込み	書き込み	読み込み	書き込み	読み込み	書き込み	読み込み	書き込み
SSD	230	520	280	110	70,000	30,000	500	500
15 K HDD	5,000	2,000	2	2	550	450	250	250
10 K HDD	7,000	3,000	2	2	400	350	200	200
7.2 K HDD	12,000	5,000	1	1	250	200	110	110
RAM ディスク	8	8	2,000	2,300	500,000	600,000	4,000	2,500

図 3. 様々なストレージテクノロジー毎のパフォーマンス

重要なポイントは、ネットワーク接続されたストレージであっても、BayStor によって、ローカル接続されたストレージと本質的に同じアプリケーションパフォーマンスが発揮できることです。また、BayStor のネットワーク接続モデルは、ストレージリソースを移動して再引き当てするためのリソースの共有と管理の柔軟性が必須となる仮想環境において、ローカル接続されたストレージのオペレーションに大きなメリットがあります。

ベンチマーク プラットフォーム

このベンチマーク プラットフォームをまとめると、以下の図のとおりです。

- CPU : IQ Multiplex は、3 つのノードで構成しました。それぞれのノードは 16 コアで、合計 48 コアでした。
- メモリー : それぞれのノードの DRAM は 384 GB ですが、IQ パラメーターは、メインデータストアの I/O ロードを増やすために Multiplex 全体でデータベースの約 10% のみキャッシュするように設定しました。これは、IQ Multiplex が大規模なデータベース（数百 TB 以上）をサポートする場合の典型的な設定です。
- ネットワーク : ストレージファブリックは、スイッチドかつノンブロッキングの 56 Gbps InfiniBand ネットワークで、BayStor の低レイテンシープロトコルを使用しています。

IQ のページサイズ、ストレージスピンドル数、その他のパラメーターは、全てのテストケースで一律にしました。システムまたはベンチマークの最適化やチューニングは行っていません。

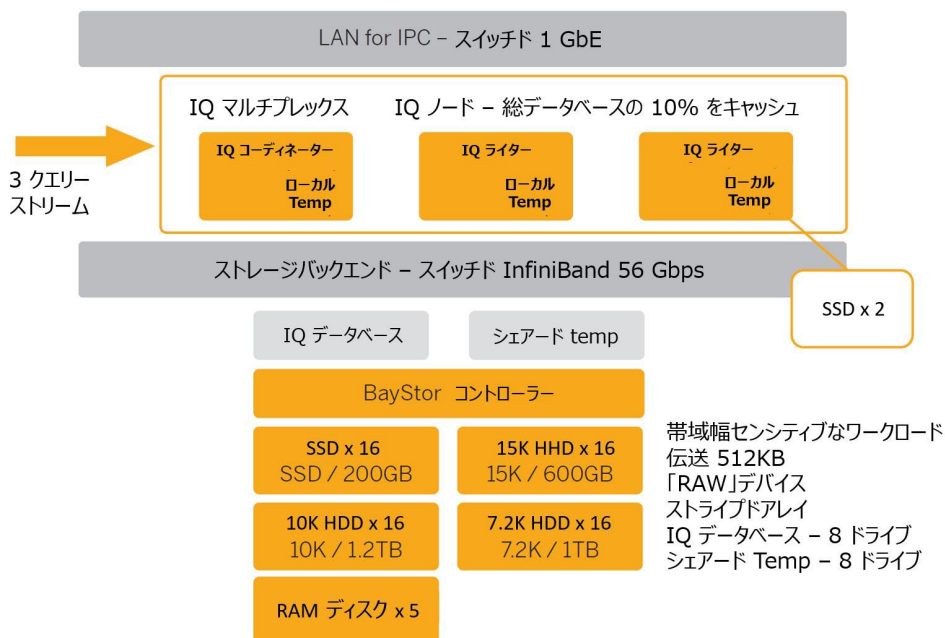


図 4. ベンチマークしたプラットフォーム

IQ Multiplex ノードのソフトウェアおよびハードウェアの設定の詳細については、Appendix A を参照してください。

測定

ロードとクエリーテストフェーズの経過時間測定に加えて、ベンチマークプラットフォーム上で、いくつかの測定を行いました。

- 各 IQ Multiplex ノード：各ノードのローカル Temp、シェアード Temp、メインストアへの I/O 帯域幅および CPU 使用率を測定しました。
- BayStor コントローラー：シェアード Temp とメインストアへの IQ Multiplex のアグリゲートを測定しました。

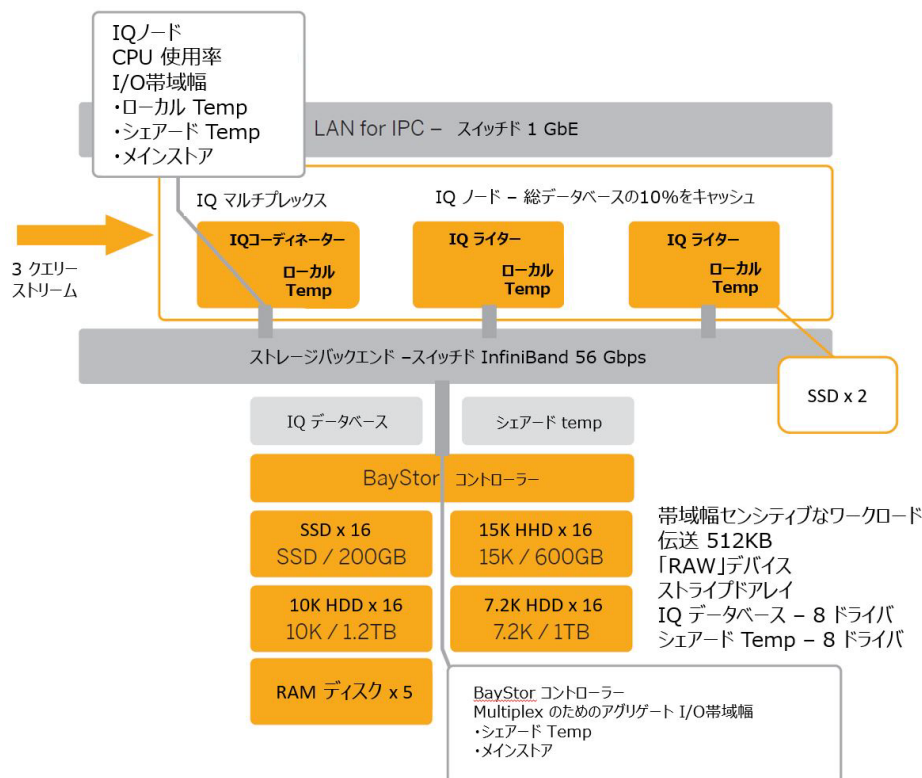


図 5. ベンチマークプラットフォーム測定

テスト シナリオ

この調査におけるベンチマークワークロードとプラットフォームを選択する一番の目的は、共有ストレージのパフォーマンスに負荷をかけるためのロード / クエリーワークロードを設定することでした。また、この調査の結果が、ソフトウェア、CPU、あるいはネットワークのパフォーマンスによる制約を受けた結果にならないようにする意図がありました。

テストケースは、2 段階のフェーズ --- IQ Multiplex データベースへ生データをロードし、その後 3 つのクエリーストリームを実行 --- で構成しました。また、テストケースは上記のストレージテクノロジーそれぞれで実行しました。さらに、フラッシュやハードドライブデバイスのテストは、「アクセラレーション」でも再度実行しました。BayStor コントローラーは、読み込みや書き込みのアクセスどちらもキャッシュし、書き込みでは acknowledgement が加速しました。

ワークロード

この調査に使用されたワークロードは、大規模顧客の評価により動機付けされ、TPC-DS（ビッグデータ基盤の性能検証）をベースにしました。実行時間を合理的なものにするため 'Scale 1000' TPC-DSデータセットを選択しました。この調査の目的は、ストレージテクノロジー間の比較を行うために I/O 集中型のワークロードを作成することであるため、ワークロードへの追加の最適化、インデックス作成、実行は行いませんでした。TPC-DS 結果における「ヒーロー」を求めることを追究したわけではありません。

ロード

この Scale 1000 TPC-DS ベンチマークには、918 GB の事前生成された生データがありました。23 のテーブルへのロードは、3 つの Multiplex ノードに対して並行に行われました。それぞれのテーブルにロードする生データのサイズは大幅に異なるため、ロード時間の長い棒は、これら 3 ノードの中の 1 ノードが、最大のテーブルをロードする時間を示します。コーディネーターノードは、小さいテーブルのうち 19 テーブルを、一方、2 つの書き込みノードのそれぞれは大きな 2 テーブルをロードしています。結果として、IQ データベースのサイズは 893 GB でした。

クエリー

2 つのクライテリアをベースに、99 の TPC-DS クエリーから 26 選択しました。

- クエリーは、CPU 集中型ではなく、I/O 集中型でした。
- クエリーは、[「ゴルディオックスの原理」](#) にもとづいていました — 短すぎず、長すぎないランタイムが存在し、短めのクエリーは I/O にそれほどストレスを与えませんでした、実行時間が長くなるクエリーが結果を支配しました。

26 のクエリーテンプレートの順列化されたストリームは、3 つの IQ Multiplex ノードのそれぞれに同時に投入され、26 のクエリーテンプレートそれぞれの特定の値のさらなるランダム化に、TPC-DS キットからのクエリー生成が使用されました。

シェアード Temp ストレージが使用されるように、DQP (Distributed Query Processing : 分散クエリー処理) を有効化しました。

結果

ストレージ デバイス

以下のグラフは、先に説明した各ストレージテクノロジーのロードとクエリーフェーズの経過時間で相対パフォーマンスを比較したものです。結果は、7200 RPM Enterprise Capacity ハードドライブアレイのパフォーマンスに正規化したものです。

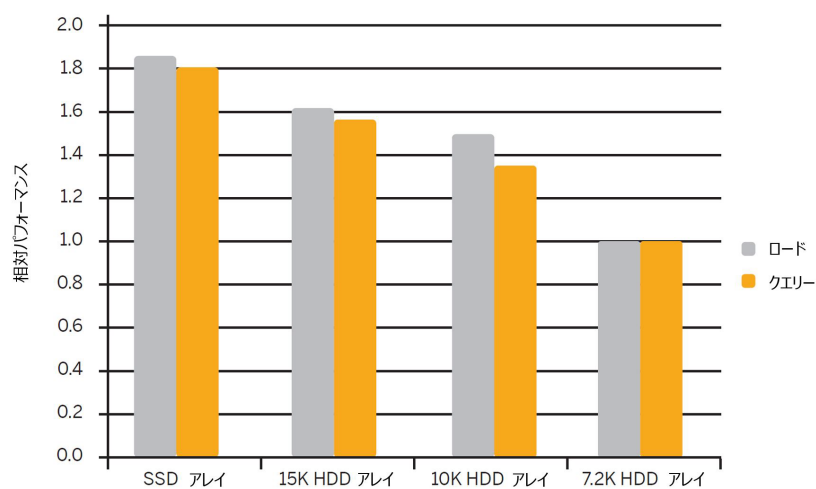


図 6. ストレージ テクノロジーのパフォーマンス比較

帯域幅を占有するこのワークロードでは、SSD/フラッシュアレイの相対パフォーマンスが、15K Enterprise Performance ハードドライブアレイよりわずかに良いことが観察されました。特に、非常に大きなデータベースをホスティングしている場合には、SSD アレイのプレミアムコストがパフォーマンスの面でメリットがあるようにはみられませんでした。

さらに、予想どおり、10 K と 7.2 K ハードドライブアレイへ移行すると、相対パフォーマンスはある程度の低下がみられました。しかしながら、これらのデバイスのコストとキャパシティのメリットは、大きなデータベースをホスティングする際には、とても魅力的です。

また、このワークロードにおけるこれらのテクノロジーの相対パフォーマンスの違いは、相対帯域幅には直接比例しませんでした。理由は、このワークロードの CPU 使用率について実験した結果から明らかです。以下のチャートは、コーディネーターノードのクエリーフェーズでの CPU 使用率を示したものです。シェアード Temp とメインストアをホスティングしている Enterprise 15K ハードドライブアレイであっても、IQ の CPU 使用率は、ハイレベルなものでした。

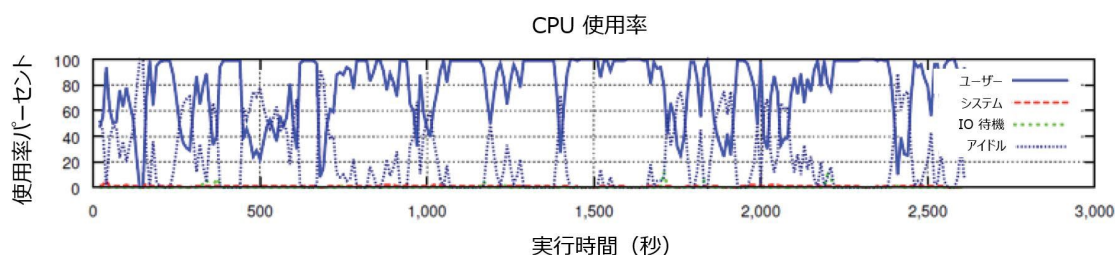


図 7. コーディネーターノードの CPU の使用率 - Enterprise 15K HDD アレイ

これをさらに展開するため、以下のグラフのように、この調査に BayStor RAM ディスクテクノロジーを含めました。BayStor RAM ディスクは、読み込み帯域幅 4 GB/秒、および 4 KB ランダムブロックアクセスレイテンシーが10 マイクロ秒以下で、パフォーマンスの観点では、理想的なストレージテクノロジーです。この非常に優れたパフォーマンスを利用しても、ロードフェーズの全体パフォーマンスは、他のテクノロジーの最善時の約 1.5 倍でした。クエリーフェーズのパフォーマンスは SSD アレイのものと同程度でした。

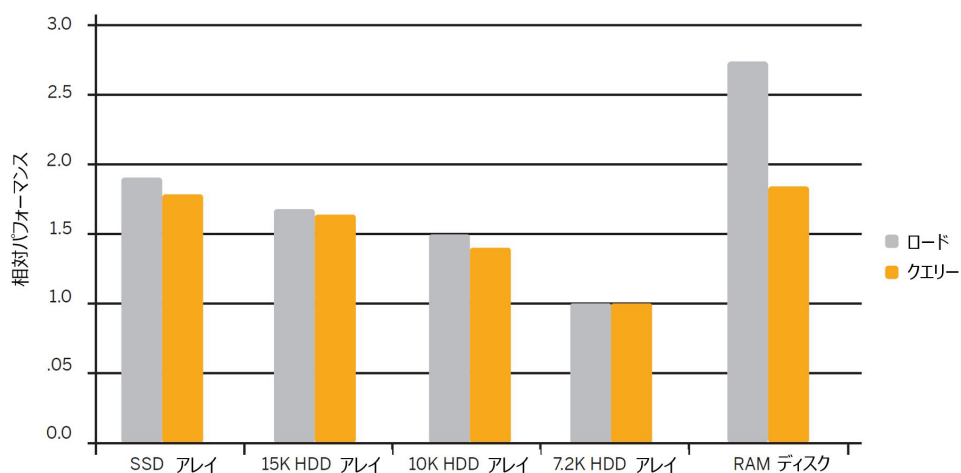


図 8. BayStor RAMディスクを含めたストレージテクノロジーのパフォーマンス比較

この結果の理由は、コーディネーターノードのロードおよびクエリーフェーズの CPU 使用率を示す以下のグラフからも明らかです。ロードフェーズでは、CPU 使用率は、75 % を前後します。高いものの、ボトルネックではありません。しかしながら、クエリーフェーズでは、かなり長時間にわたって CPU 使用率が 100% に達します。そのため、I/O が唯一のパフォーマンス制約になるワークロードとプラットフォームのベンチマークを構成したいという我々の希望に対し、IQ Multiplex は効果的に CPU リソースを使用し、パフォーマンスの制約を I/O から CPU にシフトさせているのがわかります。

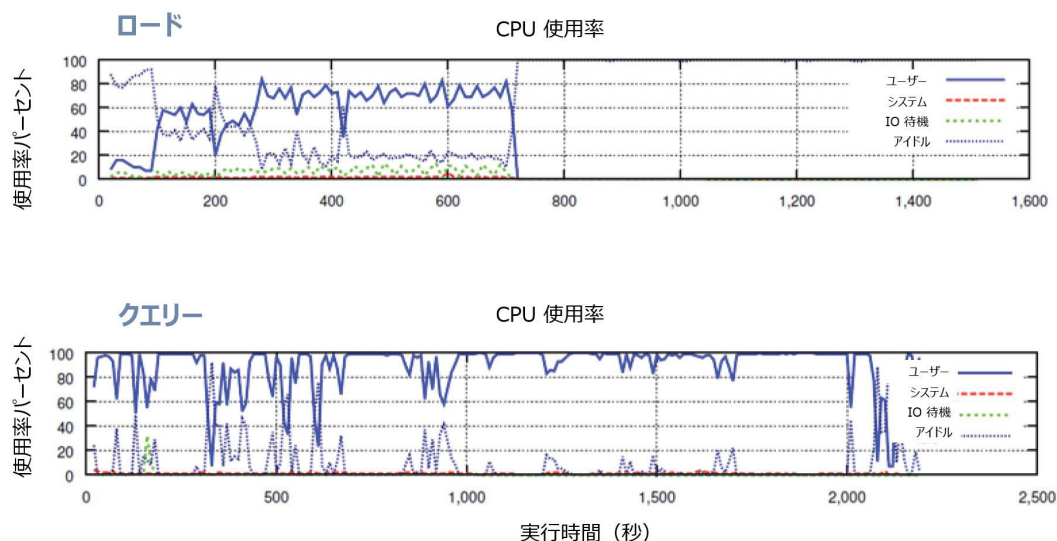


図 9. BayStor RAM ディスクデプロイ時の CPU 使用率

ストレージ ネットワーク 帯域幅

この調査の目的の1つは、ストレージネットワーク帯域幅の要件を理解することでした。下のグラフは、BayStor コントローラーで測定された IQ Multiplex の I/O 帯域幅の全体ロードを示しています。また、これには、シェアード Temp およびメインデータベースへのアクセスが含まれています。ストレージネットワークは 56 Gbps InfiniBand リンクとノンブロッキング スイッチでデプロイされてるため、ストレージネットワークに関連した制限に陥ることなく、帯域幅の要件を評価するのに十分なヘッドルームがありました。

この 3 ノード、48 コアの Multiplex においては、ロードとクエリーのフェーズの大半で、10 Gps リンクで十分であることがわかりました。しかしながら、アグリゲートされた帯域幅では、ピーク時に 10 Gbps を超えました。このホワイトペーパーで詳細は述べませんが、他の実行では、10 Gps リンクをデプロイした場合、全体として 10% のパフォーマンスの低下がみられました。

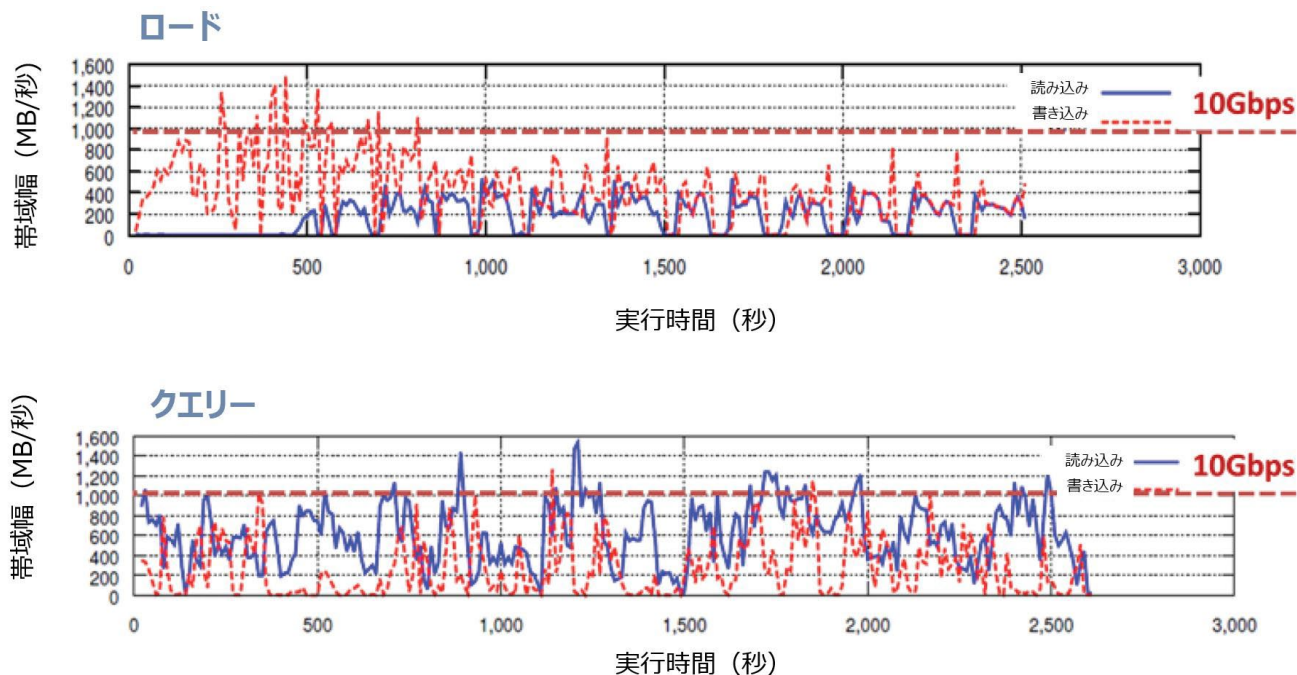


図 10. BayStor コントローラーで測定した IQ Multiplex ストレージネットワーク帯域幅

これらの結果は、この調査で使用された特定のワークロードとベンチマークプラットフォームをベースにしていることに注意してください。しかしながら、これらの結果により、1コアあたりの帯域幅 20 MB/秒が、I/O が集中するワークロードをホストする IQ Multiplex における合理的なガイドラインになると考えられます。

BAYSTOR アクセラレーション

BayStor コントローラーには、early write acknowledgement 機能に加え、DRAM ストレージにおけるストレージの大量のキャッシング機能がありますが、以下の図で示されるように、BayStor アクセラレーション機能を有効化した状態でも、このワークロードを実行してみました。BayStor コントローラーの DRAM ストアは、データベースの約 30% をキャッシュしました。コスト効果の高い、キャパシティにフォーカスしたテクノロジーを使用するほど相対パフォーマンスのメリットが増大することがわかります。これらの BayStor アクセラレーション機能は、コスト効果の高い、キャパシティにフォーカスしたストレージテクノロジーをデプロイしながら、パフォーマンスを加速させるのに役に立つ補足テクノロジーだと感じています。

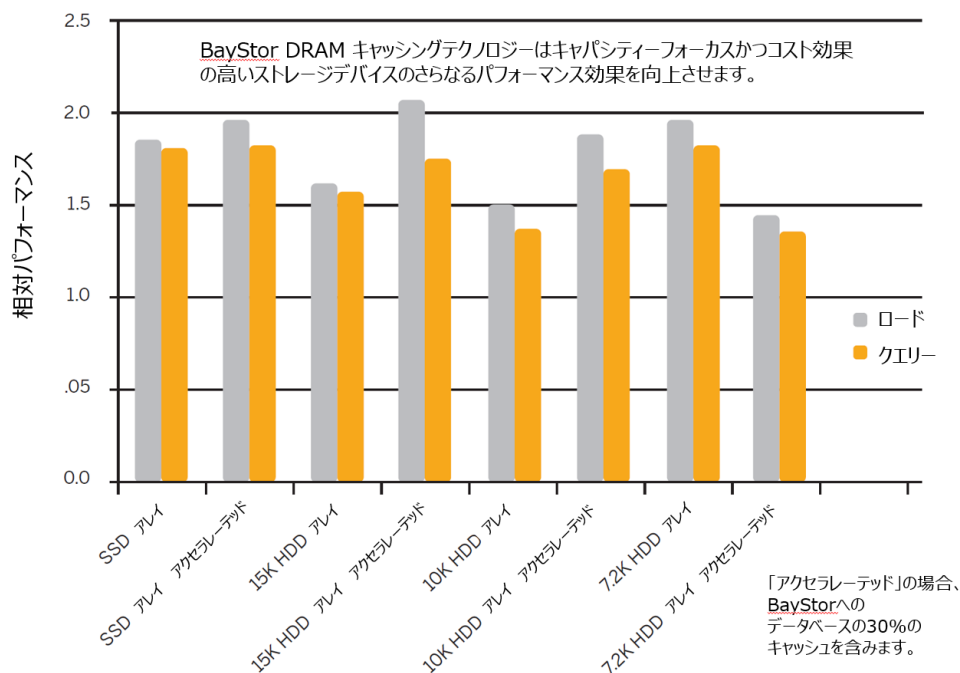


図 11. BayStor アクセラレーションを使用したパフォーマンス比較

このホワイトペーパーには含まれていない、追加の調査では、パフォーマンスのメリットの大半は、キャッシュからもたらせることが示されました。Early Write Acknowledgement は、この帯域幅占有的なワークロードではわずかなメリットしかありませんでした。IQ Simplex における RLV やトランザクションデータベースなど、データベースの更新が多いワークロードの場合、Early Write Acknowledgement がより適切であると予想しています。

結論

このワークロードにもとづき、IQ は、共有 I/O パフォーマンスにストレスをかけるよう構成されたワークロードであっても、高レベルの CPU 使用率を効果的に実現できると結論づけました。そのため、このホワイトペーパーで使用されたワークロードに近い特性のワークロードにおける IQ Multiplex の適正なデプロイメント戦略は、多くの CPU リソースを引き当てるとともに、エンタープライズクラスのハードドライブテクノロジーを使用して、シェアードストレージシステムのコストを効果的に抑えることで、非常に大規模なデータセットをホスティングする場合でも、圧倒的なパフォーマンスを実現することができるということになります。

特に、

- BayStor の NAS のパフォーマンスは、ローカルで接続されたストレージ同様のパフォーマンスを発揮します。
- Enterprise Performance ハードドライブは、帯域集中型のワークロードにおいて非常に効果的であり、実質より魅力的な価格帯でフラッシュソリューションに近いパフォーマンスを発揮します。
- Enterprise Capacity ハードドライブは、Enterprise Performance ハードドライブと比較すると低パフォーマンスであるものの、パフォーマンスとコスト比較の観点では、特に、非常に大規模なデータベースでは、とても魅力的です。

キャパシティの要件を考慮することなく、I/O が集中するワークロードにおける IQ Multiplex のストレージネットワーク帯域幅のデプロイメントガイドラインを 1 コアあたり 20 MB/秒と提案することができます。これには、ストレージネットワークとともにストレージアレイの総キャパシティを含みます。

このワークロードでは、SSD / フラッシュなどより高価なストレージテクノロジーを使用するケースでは弱いことがわかりました。しかしながら、以前のホワイトペーパー²より、フラッシュは、データベースのアップデートに RLV が使用されている IQ Simplex においては、非常に大きなパフォーマンスのメリットがあることがわかっています。

リソース

SAP/Bay Storage Technology 共同ホワイトペーパー

- “Accelerating SAP Sybase IQ Performance with Bay Storage High-Performance and Low-Cost Data Access Technology”, September 2013, <http://scn.sap.com/docs/DOC-47983>.

SAP IQ:

- SAP web サイト: <http://www.sap.com>
- SAP IQ: <https://www.sap.com/japan/products/sybase-iq-big-data-management.html>
- SAP IQ のマニュアル: https://help.sap.com/docs/SAP_IQ?locale=en-US
- SAP IQ のコミュニティ: <https://community.sap.com/search/?by=updated&ct=all&mt=01200314690800002024&q=SAP%20IQ>

Bay ストレージ:

- Bay Storage Technology web サイト: <http://www.baystoragetechnology.com/>

Seagate:

- Seagate web サイト: <http://www.seagate.com/>
- Seagate Enterprise ハードドライブ: <http://www.seagate.com/internal-hard-drives/enterprise-hard-drives/>

謝辞

このテクノロジーの評価は、Bay Storage Technology によって行われました。この調査の中で使用されたハードドライブテクノロジーは、Seagate により提供されました。このホワイトペーパーの SAP IQ の説明にかかわるコンテンツや、追加のレビューおよびコメントは SAP が行いました。

²“Accelerating SAP Sybase IQ Performance with Bay Storage High-Performance and Low-Cost Data Access Technology”, September 2013, <http://scn.sap.com/docs/DOC-47983>.

APPENDIX A — システム構成

IQ Multiplex は、3 つのノードで構成 — コーディネータ 1 に、ライター 2。全ノードが SAP IQ 16 SP2。各 Multiplex ノードは同じソフトウェアを使用かつ同じハードウェア設定。

- Linux Centos 6.3
- デュアルソケット Intel E5-2670 プロセッサ (Sandy Bridge)
 - 16 cores
 - 2.6 GHz
 - 384 GB メモリー
- ローカルストレージに LSI SAS/SATA コントローラー
 - ローカル Temp は Samsung SM843 SSD (240 GB。200 GB をオーバープロビジョン) x 2。個別の「ロー」デバイスとして設定
- ネットワークインターフェース
 - 1 GbE – LAN および IQ ノード間通信
 - 56 Gbps InfiniBand – ストレージネットワーク

全ての SAP IQ サーバーノードの設定ファイルは以下のとおり：

```
-gd all
-gl all
-gm 32
-gp 4096

-iqmc 35000
-iqtc 35000
-iqlm 10000
```

APPENDIX B — テスト結果の分析

このホワイトペーパーでまとめられた全体結果に加えて、この測定では、このベンチマークプラットフォーム上で実行されるワークロードの CPU 特性と I/O 特性を掘り下げることができました。この Appendix で示す結果は、Enterprise Performance 15 K ハードドライブアレイで実行したワークロードにもとづいています。

ロードフェーズ

各 IQ Multiplex ノードは、このワークロードで使用されている 23 テーブルのいくつかを担当します。以下のグラフは、そのうち 1 つの書き込みノードの I/O ロードと CPU 使用率を示しています。このノードは、最大のテーブルのロードを担当しており、ロードフェーズで測定した経過時間の長いボールはこのノードのものでした。

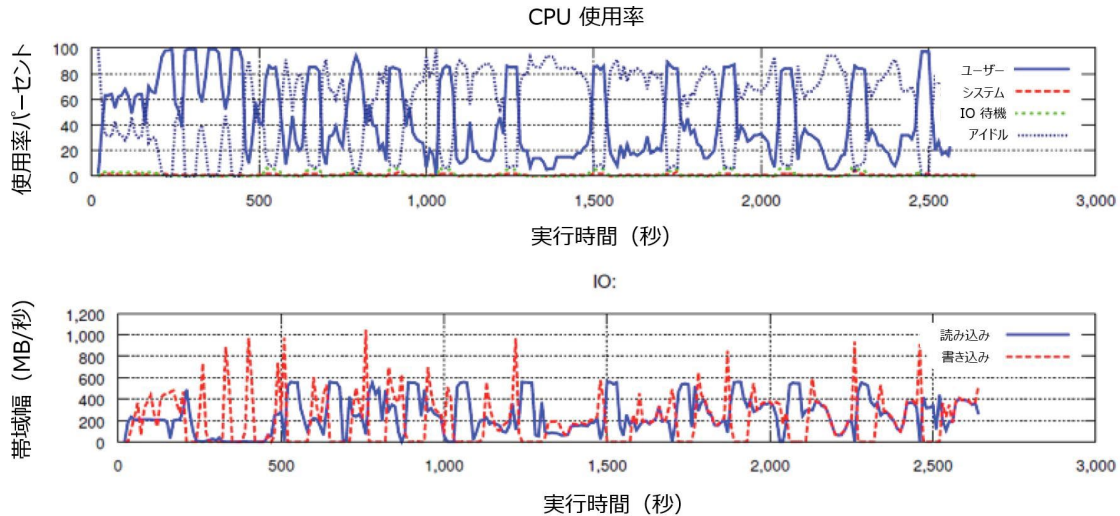


図 12. 最大のテーブルをロード中の書き込みノード

CPU の使用は非常にバイモーダルであり、この 16 コアのノードにわたって低レベルの使用と交互に高レベルで使用されていることが観察されました。

また、非常に興味深いことに、ロードフェーズの I/O 特性は書き込みに限定されたものではありませんでした。上記のグラフが示すように、ロードの最初の段階の経過後、ほぼ同等量の読み込みと書き込みがあります。

クエリー フェーズ

以下のグラフは、クエリーストリームと DQP (Distributed Query Processing : 分散クエリー処理) Multiplex 実行中に、3 つある IQ ノードのうちの 1 つでみられる I/O のアクティビティを示しています。各 IQ ストアで I/O アクティビティがあります。

- ローカル Temp - 少量の急激な上昇がみられるアクティビティ。本質的には、全てのアクティビティが書き込み
- シェアード Temp - 書き込みと読み込みがほぼ同じ一定量のアクティビティ
- メインストア - 予想どおり、このピュアなクエリーワークロードでは、読み込みのアクティビティのみ

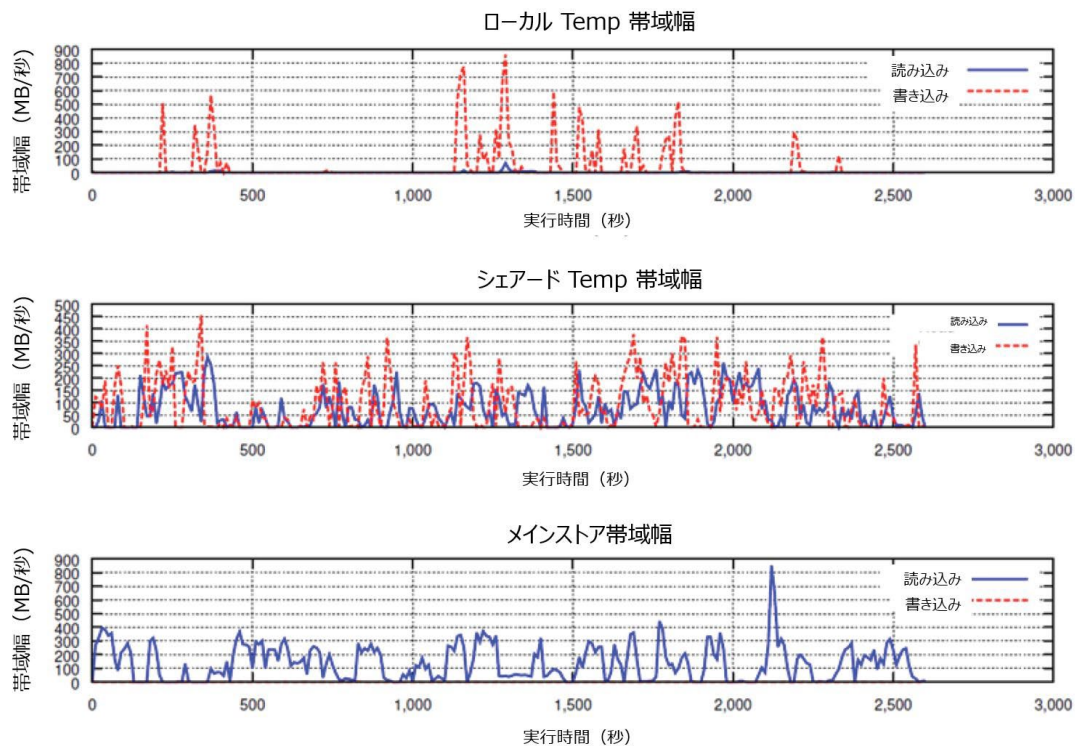


図 13. クエリストリームを実行中のコーディネーターノード

