

# 並列データ処理実行時の IP-SAN 統合型 PC クラスタの動作特性解析

原 明日香 (指導教員:小口 正人)

## 1 はじめに

近年、情報システムにおいて処理される情報量が爆発的に増大しており、その中からユーザが必要とする情報を高速に取り出すことが求められている。そこで膨大なデータを処理するために、本研究ではバックエンドとフロントエンドのネットワークを統合した IP-SAN 統合型 PC クラスタを構築して利用した。ただしアプリケーション実行時にクラスタのノード間通信や I/O の実行がどのように振舞いシステム性能に影響を与えているのかなど詳しい解析は行われていない。そこで本研究では、相関関係抽出などの並列データ処理を実行し、システムのモニタを行い解析することによって、IP-SAN 統合型クラスタの詳しい振舞いを明らかにする。

## 2 IP-SAN 統合型 PC クラスタ

我々は図 1 のような SAN を使用した PC クラスタにおいて、IP-SAN を導入することで、図 2 のような Front-end と Back-end のネットワークを同じ IP ネットワークに統合した IP-SAN 統合型 PC クラスタの実現を考えている。ネットワークを統合することでネットワーク構築コストと管理コストの削減が期待出来るが、ネットワーク統合により性能がネットワークバウンドになる可能性が懸念される。

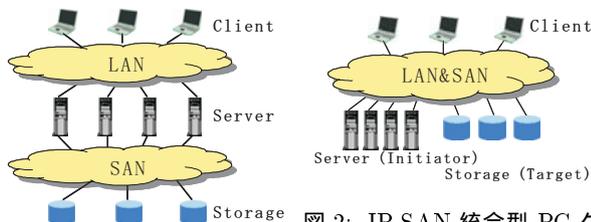


図 1: 通常の PC クラスタ

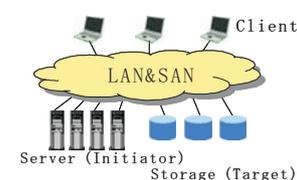


図 2: IP-SAN 統合型 PC クラスタ

## 3 並列データマイニング

相関関係抽出のアルゴリズムとしてよく知られているものに、Agrawal らによる Apriori アルゴリズムと Han らによる FP-growth アルゴリズムがある。Apriori は発見された頻出アイテムセットから候補アイテムセットを生成し、数え上げを行う。FP-growth は巨大なトランザクションデータベースをコンパクトに圧縮したデータ構造である FP-tree を利用することで、候補パターンを生成せずに頻出パターンを抽出する。

相関関係抽出で扱うデータはしばしば巨大であるため、データベースを分散し計算処理を並列化する必要がある。Apriori の並列相関関係抽出のアルゴリズムはいくつか提案されているが、本研究ではハッシュ関数を用いて Apriori を並列化する HPA (Hash Partitioned Apriori) を用いる。FP-growth の並列相関関係抽出のアルゴリズムは、既存研究において提案された PFP (Parallel FP-growth) を用いる。

## 4 実験概要

本研究では、アクセス性能が高い SAS ディスクをデータの読み書きを高速化する RAID0 構成とすることで、高性能なストレージを持つディスクを Target ストレージとした IP-SAN 統合型 PC クラスタを構築した。この際に、クラスタ構築管理ツールである Rocks を用いた。また、クラスタの通信状況をモニタリングするために Ganglia のインストールを同時に行った。表 1 にクラスタノードのスペックを示す。

表 1: Experimental setup : PCs

|             |   |
|-------------|---|
| OS          | Initiator : Linux 2.6.18<br>Target : Linux 2.6.18                     |
| CPU         | Initiator : Intel Xeon 3.6GHz<br>Target : quad-core Intel Xeon 1.6GHz |
| Main Memory | Initiator : 4GB<br>Target : 2GB                                       |
| HDD         | Initiator : 250GB SATA<br>Target : 73GB SAS × 2 (RAID 0)              |

## 5 実験結果と考察

### 5.1 I/O 性能測定 (1 対 1)

まず、構築したクラスタの基本性能を測定するために、ハードディスクベンチマークの bonnie++ を用いて、SATA ディスクと SAS ディスク (RAID0 構成) であるローカルデバイス、および SAS ディスク (RAID0 構成) を Target ストレージとして Initiator と Target を 1 対 1 接続させた iSCSI の計 3 種類のストレージについて Sequential read アクセスと Sequential write アクセスを測定した。

この結果から、ローカルディスクとして SAS ディスク (RAID0 構成) を用いた場合の性能が格段に良いことが分かった。また、iSCSI を用いた通信の方がローカルディスクとして SATA ディスクを用いた場合よりも性能が良いことが分かった。これらにより、今回高性能なディスクを IP-SAN 統合型 PC クラスタに導入できたということが確認できた。

### 5.2 並列データマイニング (1 対 1)

次に、上記実験と同じ環境で、HPA アルゴリズムと PFP アルゴリズムの並列化プログラムについて、アイテム数を 1000 とし、トランザクション数が 1M、2M、4M、8M のトランザクションデータを用い、最小サポート値を 0.7 % として実行した。

この結果から、HPA の場合どのクラスタにおいても性能はほとんど変わらず、PFP の場合ベンチマークで格段に性能が良かった SAS ディスクを用いたクラスタの性能がやや劣るといことが分かった。この時のモニタリングの結果から、ネットワークにはまだ十分余裕があり、IP-SAN のトラフィックを統合してもネットワークの帯域にあまり影響を与えないことが分

かった。また、今回使用した HPA のプログラムはディスクアクセス性能やメインメモリ容量の影響を受けにくく、PFP のプログラムはディスクアクセス性能の影響を受けにくい、メインメモリ容量の影響を受けやすいと考えられる。

### 5.3 I/O 性能測定 (n 対 1)

これまでの実験においては、IP-SAN 統合型 PC クラスタの Initiator と Target を 1 対 1 接続していたが、iSCSI の使われ方を考えると、数台の Initiator が 1 台の Target にアクセスする形が一般的であると考えられる。Initiator と Target を 1 対 1、2 対 1、4 対 1、8 対 1 にそれぞれ接続することで、ネットワークと Target に高負荷をかけた環境を構築した。そのシステムにおいて、SAS ディスク (RAID0 構成) であるローカルデバイス、SAS ディスク (RAID0 構成) を Target ストレージとして接続させた iSCSI のストレージについて、ハードディスクベンチマーク bonnie++ を用いて、Sequential read アクセスと Sequential write アクセスを測定した。

その結果、Sequential read アクセスにおいてはどの接続方式においても性能はほとんど変わらず、Sequential write アクセスにおいては負荷を高くしていくごとに性能が少し低下した。このとき、ネットワーク帯域にはまだ余裕があり、CPU 負荷が高くないことから、ネットワークではなく Target のストレージ I/O がボトルネックであると考えられる。

### 5.4 並列データマイニング (n 対 1)

bonnie++ と同様に、Initiator と Target を 1 対 1、2 対 1、4 対 1、8 対 1 でそれぞれ接続することで、ネットワークに高負荷をかけた環境を構築した。そこで、HPA アルゴリズムと PFP アルゴリズムの並列化プログラムについて、アイテム数を 1000 とし、トランザクション数が 1M、2M、4M、8M のトランザクションデータを用い、最小サポート値を 0.7 % として実行した。プラットフォームとしては、先の bonnie++ 実験時と同じ環境を用いた。

その結果、どちらのプログラムにおいても、負荷の量を変えた場合においても実行時間はほとんど変わらなかった。このときネットワーク帯域にはまだ余裕があり、CPU 使用率が非常に高くなっていることから、ネットワークではなく Initiator の CPU がボトルネックであると考えられる。

### 5.5 複数プロセス (n 対 1)

次に、bonnie++ を実行させながら HPA アルゴリズムと PFP アルゴリズムの並列データマイニングプログラムについて、アイテム数を 1000 とし、トランザクション数が 1M、2M、4M、8M のトランザクションデータを用い、最小サポート値を 0.7 % として実行した。プラットフォームとしては、先の実験時と同じ環境を用いた。

その結果、図 3 のように高負荷を与えてトランザクションデータ量を増やしていくごとに実行時間が長くなった。このとき図 4 に示すようにネットワーク帯域にはまだ余裕があり、図 5 および図 6 から、Target における処理が軽いときは、Initiator の CPU 使用率が高く、Target における処理が重いときは、Target のストレージ負荷が高くなっていることから、Initiator の CPU または Target のストレージ I/O がボトルネックであ

ると考えられる。

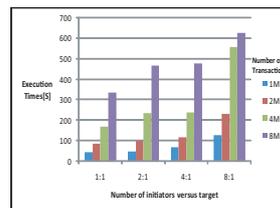


図 3: bonnie++ と HPA を実行させたときの HPA 実行時の実行時間

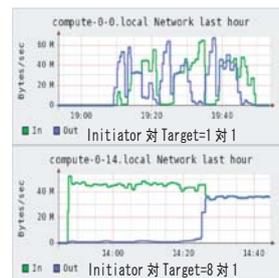


図 4: bonnie++ と HPA を実行させたときのネットワークトラフィック

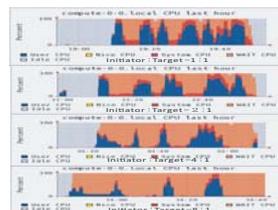


図 5: bonnie++ と HPA を実行させたときの Initiator の CPU 使用率

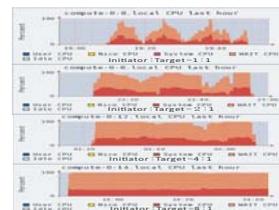


図 6: bonnie++ と HPA を実行させたときの Target の CPU 使用率

## 6 まとめと今後の予定

ネットワークと Target に高負荷をかけた IP-SAN 統合型 PC クラスタにおいて、並列データマイニングおよび複数プロセスを実行させたときの性能評価を行った。その結果いずれの場合においても当初懸念されたネットワークバウンドにはならず、IP-SAN 統合型 PC クラスタの柔軟なシステム構成を実現することが可能である。

今後は他のデータ処理アプリケーションを用いシステムの評価を行うことで、IP-SAN 統合型 PC クラスタの最適化などを行っていく。

### 参考文献

- [1] 原明日香、神坂紀久子、山口実靖、小口正人:”並列データマイニング実行時の IP-SAN 統合型 PC クラスタのネットワーク特性解析”, 分散、協調とモバイル (DICOMO2007) シンポジウム, 3H-2, pp.705-714 札幌, 2007 年 7 月
- [2] Asuka Hara, Kikuko Kamisaka, Saneyasu Yamaguchi, and Masato Oguchi:”Analyzing Characteristics of PC Cluster Consolidated with IP-SAN using Data-Intensive Applications”, In Proc. 20th IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS2008), No.631-042, Orland, Florida, November 2008
- [3] 原明日香、神坂紀久子、山口実靖、小口正人:”IP-SAN 統合型 PC クラスタにおける並列データ処理アプリケーション実行時の特性解析” コンピュータシステム研究会 (CPSY), 京都, 2008 年 12 月