

クラウドコンピューティングにおける高性能ファイルシステムの検討

小林 賢司[†] 建部 修見^{††}[†] 筑波大学第三学群情報学類 ^{††} 筑波大学大学院システム情報工学研究科

1 はじめに

近年、インターネットを介してオンデマンドなサービスを提供するクラウドコンピューティングの需要が増えてきている。クラウドコンピューティングの1つの形態である IaaS は、多数のサーバ群やストレージ群を提供し、ユーザはその上で仮想計算機を動作させるモデルをとる。多数の仮想計算機も起動出来る IaaS は、大規模 PC クラスタを用いる科学技術分野にも期待される。そのようなデータインテンシブコンピューティングでは、スケーラビリティに優れ高速なディスク I/O バンド幅を実現するファイルシステムが必要とされる。

しかし、IaaS の特性を考慮したファイルシステムを、サービスとしてどのように提供すべきであるかはまだ明らかになってはいない。

よって本研究は、容易に仮想計算機台数をスケールできる IaaS 上で、仮想計算機間で動作する高性能な共有ファイルシステムについて検討する。

2 IaaS に適したファイルシステム

仮想計算機台数をスケールしただけディスク I/O バンド幅もスケールするようなファイルシステムが望ましい。本研究では、IaaS として現状最も商用的に利用されている Amazon EC2[1] のインスタンス (仮想計算機) 上で、以下の3つの共有ファイルシステムを動作させ、比較、検討する。

- S3FS[3]: 高信頼かつ高トラフィックに耐えうるパブリックなクラウドストレージサービスである Amazon S3[2] のディレクトリを、FUSE を用いてマウントするファイルシステムである。
- PVFS[4]: ファイル内のデータを、複数のファイルサーバのディスクにストライピングするクラスタファイルシステムである。ファイルアクセスの均一化、並列化が望める。メタデータサーバとファイルサーバで構成される。

- Gfarm[5]: 広域にも対応した分散ファイルシステムである。ファイル単位で分散させるため、複数ファイルへの同時アクセスに対する性能の向上が見込まれる。また複製ファイルを作成することにより、集中アクセスによる性能の低下も防ぐことができる。

3 システムの構成

Amazon EC2 のインスタンスはシャットダウンすると、インスタンス上のすべてのデータが消滅してしまうという特徴がある。そのため永続的なデータ保存場所を各ファイルシステムで提供しなければならない。Amazon Web Services では、永続的なデータ保存場所として Amazon S3 の他に Amazon EBS がある。Amazon EBS は、パブリックなアクセスが可能な Amazon S3 と異なり、外部デバイスとして Amazon EC2 のインスタンスに直接マウントして利用する。これを踏まえた上で、各ファイルシステムを用いたシステムの構成について述べる。

S3FS は、データ保存場所が Amazon S3 なので、ストレージ側の対応は必要がない。すべてのインスタンスが S3FS クライアントとなり、Amazon S3 をマウントすることになる。

PVFS は、特定数のインスタンスをメタデータサーバ、ファイルサーバとする (図 1)。ファイルデータの保存先に EBS ボリュームを用いる。

Gfarm は、メタデータサーバとなる1つのインスタンスと、クライアントでもありファイルサーバでもある複数のインスタンスから構成されるようにする (図 1)。PVFS と同様に、EBS ボリュームを用いる。

4 性能評価と考察

4.1 ディスク I/O 性能評価

各ファイルシステムにおいて、クライアントとなる複数インスタンスから、ファイルの同時読み書きによる性能評価を行った。ファイルは、クライアント毎に異なるファイルである。EC2 は仮想マシン仕様を選択できるが、ここでは Xeon2.3GHz 相当の仮想 CPU 1 コア、メモリが 1.7GB の仕様を用いる。ファイルサイズはキャッシュの影響を考慮し、メモリの約 3 倍の容量である 5GB とする。PVFS のファイルサーバ数は 16 とする。クライアント数を 1~128 の範囲で変化させ、評価を行った。

Study of High Performance File System on Cloud Computing

[†] Kenji Kobayashi

^{††} Osamu Tatebe

College of Information Sciences, Third Cluster of Colleges, University of Tsukuba ([†])
Graduate School of Systems and Informaion Engineering, University of Tsukuba (^{††})

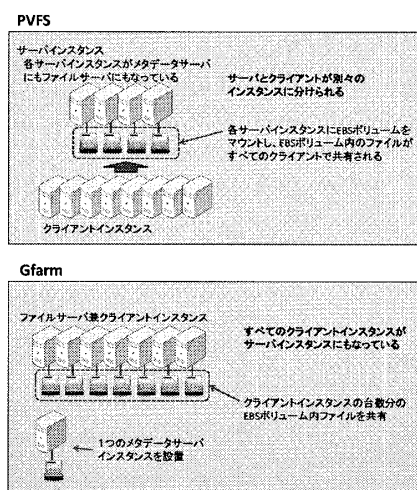


図 1: EC2 上での PVFS と Gfarm の構成

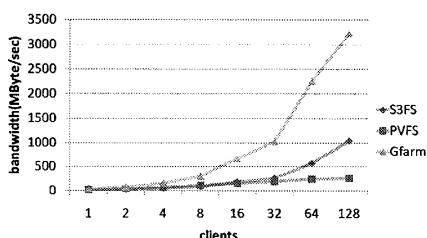


図 2: write 性能

クライアント数が 128 のときに総 I/O バンド幅は、書込時 S3FS で 1042MB/s, PVFS で 256MB/s, Gfarm で 3221MB/s (図 2) . 読込時 S3FS で 1380MB/s, PVFS で 174MB/s, Gfarm で 6689MB/s という性能となった (図 3) . S3FS と Gfarm では読み書きともに個々のバンド幅が低下することなく、総 I/O バンド幅がスケールされることが確認された。

4.2 考察

S3FS は、Amazon S3 に依存しており、その詳しい内部構成は明かされていないが、複数のデータセンターや RAID 構成によってスケールされていると思われる。しかし書き込み時に、クライアント数が 64 のときと 128 のときに、それぞれ 1 つのクライアントと 5 つのクライアントで I/O エラーが見受けられた。

PVFS は、クライアント数がサーバ数を上回ると、総 I/O バンド幅は多少性能向上してはいても個々のバンド幅は大きく低下している。分散するディスクやネットワークの数が限られるため、そこでボトルネックになっていると思われる。

Gfarm は、S3FS と同様にスケールしているが、S3FS よりも個々のバンド幅が大きいいため、大きく総 I/O バンド幅を向上させている。クライアントとなるインスタンスがファイルサーバとなっているため、ク

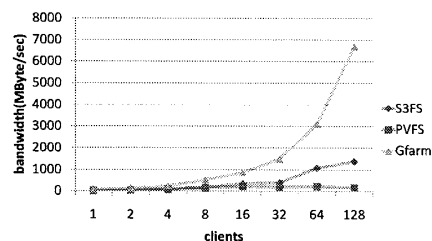


図 3: read 性能

ライアント数が 128 のとき、実質インスタンス 128 個分のディスクやネットワークの資源を共有できることになる。ファイルの局所性を考慮しており、書込はローカルの EBS ボリュームに、読込はネットワークに近いサーバを選択することも、性能向上の理由である。またサーバの台数が基本的に固定である PVFS と異なり、Gfarm は新たにファイルサーバをスケールすることが容易でもある。よって Amazon EC2 上では 3 つのファイルシステムの中で、性能面で Gfarm が最も適しているといえる。

5 おわりに

本稿では、Amazon EC2 を利用して IaaS 上での共有ファイルシステムのディスク I/O 性能の評価を行った。S3FS と Gfarm においてインスタンス数が大規模化してもスケールブルに向上する結果が得られ、特にファイルの局所性を考慮した Gfarm はより高性能であり、IaaS 上での有用性が示された。今後はこの結果を踏まえ Amazon EC2 に依存しないクラウド環境での評価、またその上で動作するアプリケーションでの評価を行っていく。

謝辞 本研究の一部は、文部科学省科学研究費補助金特定領域研究課題番号 21013005 および文部科学省次世代 IT 基盤構築のための研究開発「e-サイエンス実現のためのシステム統合・連携ソフトウェアの研究開発」、研究コミュニティ形成のための資源連携技術に関する研究 (データ共有技術に関する研究) による。

参考文献

- [1] Amazon Elastic Compute Cloud Developer Guide API Version 2008-12-01.
- [2] Amazon Simple Storage Service Developer Guide API Version 2006-03-01.
- [3] S3FS. <http://code.google.com/p/s3fs/>
- [4] PVFS. <http://www.pvfs.org/>.
- [5] 建部修見, 曾田哲之. 広域分散ファイルシステム Gfarm v2 の実装と評価. 情報処理学会研究報告, 2007-HPC-113, pp.7-12, 2007 年 12 月.