HDFS 対応型スケールアウト NAS の性能評価

佐藤 充[†1]　神崎 浩貴[†2]　川田 翔士[†2]　杉山 俊[†3]　大塚 真吾[†1]

ストレージ製造の大手である EMC 社の Isilon が，Hadoop の分散ファイルシステムである HadoopDistributedFileSystem(HDFS)への対応を発表した．Hadoop は HDFS と分散処理フレームワークである MapReduce で構成される大規模データのための並列分散処理基盤であり，複数のコンピュータを連携させて運用する事で，テラバイトからペタバイト級の巨大ファイルも扱うことができる．しかし，詳しい性能や各種パラメータの傾向などは明らかになっていない．そこで，本研究では，ベンチマークを用いて通常の HDFS と HDFS 対応の Isilon を用いた場合での書込みと読込み性能の比較検証を行う．

Performance Evaluation of Scale out NAS for HDFS

MAKOTO SATO[†1]　HIROKI KANZAKI[†2]　SHOJI KAWATA[†2]
SHUN SUGIYAMA[†3]　SHINGO OTSUKA[†1]

1. はじめに

スケールアウト NAS は, 拡張性や管理性において従来のスケールアップ NAS にはない利点を持ち, 容量と性能をシームレスに拡張することができる．このため，小規模構成から必要に応じて徐々に拡張するという対処が可能になった. スケールアウト NAS はクラスタを単一のファイルシステムとして管理し，データの実際の物理的位置を意識する必要がない[1]．その中でも最大手である EMC 社の Isilon が，Hadoop の分散ファイルシステムである HadoopDistributedFileSystem(HDFS)をネイティブサポートする事を発表し，最新の OS(OneFS)では HDFS に対応している．

Hadoop は HDFS と分散処理フレームワークである MapReduce で構成される大規模データのための並列分散処理基盤であり，複数のコンピュータを連携させて運用する事で，テラバイトからペタバイト級の巨大ファイルも扱うことができる．HDFS は Namenode と呼ばれる唯一のノードと，Datanode のクラスタで構成されており，Datanode はブロックと呼ばれる固定長に区切られた HDFS 内データを管理している．

Namenode はメタデータと呼ばれるファイルの属性情報やファイルシステムの情報を管理している．Namenode は HDFS における単一障害点であり，ダウンした場合，ファイルシステムはオフラインとなるという問題があるが，Isilon では HDFS における単一障害点などの問題を解決している．

また，HDFS(Hadoop)環境では，Datanode の HDD 故障に対して，冗長性を持たせるレプリカ機能を保持しており，ユーザは利用状況に応じてレプリカ数を決める事が可能である．レプリカ数が多いほど冗長性は向上するが，HDD へのアクセスが多くなるため，処理能力は低下する．Hadoop の一般的な利用ではレプリカ数を 3 程度にしているが、HDFS 対応型の Isilon では Isilon 自身が冗長性を担保しているため, レプリカ数を 1 として処理を行う事が可能である．したがって，HDFS 性能を最大限に活かすことが可能である．

このように，Isilon は HDFS における問題点の一部を解決しつつ，高度な処理能力を保持しているが，詳しい性能や各種パラメータの傾向などは明らかになっていない．そこで, 本研究では, 実際に HDFS 対応の Isilon を用いて Hadoop 環境を構築し, ベンチマークを用いて通常の HDFS と HDFS 対応の Isilon を用いた場合での書込みと読込み性能の比較検証を行う．

2. 関連研究

本研究の関連研究としては，石井らによるマルチコア CPU 環境における仮想計算機を用いた Hadoop システムの評価[2]や櫻井による分散ファイルシステムの性能評価に関する研究[3]，百瀬らによる高遅延環境における分散ファイルシステム Hadoop の遠隔データアクセス特性の評価[4], 奥寺らによる MapReduce 環境におけるアドホックなクエリを対象とした Adaptive indexing 適用モデルの提案とその評価[5]などが存在する．

†1 神奈川工科大学　情報工学科
†2 神奈川工科大学大学院　工学研究科
†3 図研ネットウエイブ株式会社

3. 並列分散処理

3.1 Hadoop

技術の進歩により，様々な局面で大量のデータが発生している．モバイル機器に搭載されている GPS やカメラ，赤外線測域センサによるユーザの行動履歴など，分析をすることで有益な情報が得られるといわれているビックデータは日々増大し続けている．

このような中で, Hadoop は大規模データを手軽に複数のマシンに分散して処理できる並列分散処理基盤として注目を集めている. また近年では, ビックデータだけではなく，企業における売り上げデータ処理や集計処理など，一定時間内に処理を終わらせたい場合にも用いられるケースが増えてきているため，今後がさらに期待される．

Hadoop は分散ファイルシステムである HDFS と分散処理を担う MapReduce，データベースの基盤となる hBase で構成されている．この中でも重要なのが分散処理を実行する MapReduce 処理で，データは Key と Value の組み合わせで管理されている．MapRecude 処理は大規模なデータを小さなデータに分割し必要な情報を抽出する Map 処理，同じ Key を持つ組み合わせを束ねる Shuffle 処理，それをまとめて結果を出力する Reduce 処理の 3 つからなる．

HDFS は，大規模データを分割して複数のディスクで管理するファイルシステムである．データを複数のディスクから平行して読込むことでスループットを向上させ，大規模データを効率よく処理させる工夫がされている．また，デフォルトでレプリケーション数が 3 に設定されているため，一部のマシンが故障しても，データの損失を防ぐことができるようになっている．

3.2 Isilon

Isilon はアイシロンシステムズにより開発されたスケールアウト NAS である．NAS コントローラが分離していないのが従来のスケールアウト NAS とは異なる点である．

ディスク装置には CPU，メモリ，ディスク，ネットワークといったハードウェアのほか，ファイルシステムなどのソフトウェアが搭載されている．これらの装置はノードと呼ばれ，最小で 3 ノードから最大で 144 ノードまで拡張することができる．ノードの追加は容易であり，ノードが追加されると自動的にデータを再配置し，最適化する．

従来の NAS には拡張性に欠点が存在するが, ディスクの容量が増加してきたことにより，大容量のデータまでカバーすることができるようになった．しかし，単一では限界が存在するため，ボリュームが増加し，管理する負荷が大きくなってしまう．

スケールアウト NAS では，このような問題を解決し，日々増大するデータに対応するため，拡張することができるように設計されている．

4. ベンチマークを用いた性能評価

本研究では，実際に HDFS 対応の Isilon を用いて Hadoop 環境を構築し, ベンチマークを用いて通常の HDFS と HDFS 対応の Isilon を用いた場合での書込みと読込み性能の比較検証を行った．

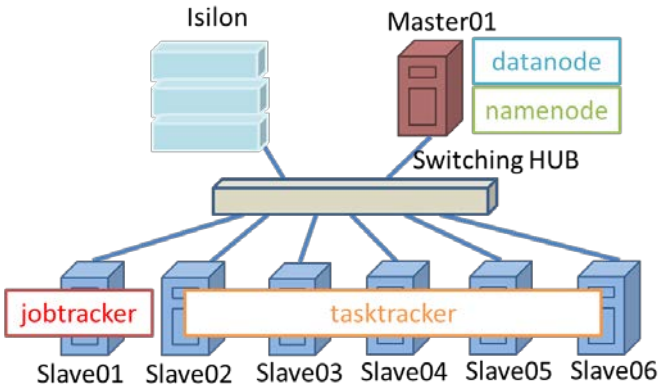

Figure 1 構成図

4.1 実験環境

本実験に用いたハードウェアは, 6 台の計算ノード, Isilon, Isilon と同等の性能を持つ HP 製サーバ(以下 HP サーバ)である. 各計算ノードは CPU:Intel corei5-2500K, メモリ:8GB, HDD:1TB の SATA(7,200rpm)×1 本を搭載している．Isilon は CPU:Nehalem Quad Core，メモリ:6GB，HDD:500GB の SATA (7,200rpm)×12 本を搭載している．また，HP サーバは，CPU:Xeon E5607 2.26GHz，メモリ:6GB，HDD:1TB の SATA (7,200rpm)×8 本を搭載している．OS については，計算ノードと HP サーバは CentOS(6.2)，Isilon は OneFS である．

MapReduce 処理は 1 台の JobTracker と 5 台の TaskTracker で行い，HDFS 処理は NameNode と DataNode ともに Isilon が行う．また，比較対象として，NameNode と DataNode を HP サーバで行う手法でも実験を行った．

デフォルトでは, Hadoop の各ノードの Map タスク数は 2, Reduce タスク数は 1 に設定してあるが，本研究では Map タスク数を 2，Reduce タスク数を 2 に設定して実験を行った．また，レプリカ数は通常の HDFS では 1，Isilon による処理では Isilon を 3 台使用するため, Isilon のノード数とレプリカ数は 3 になっている．

4.2 実験内容

今回の実験では基本的な性能を調べるため, Hadoop 付属のベンチマークである Teragen，Grep，Terasort を使用し，読込みと書込みの性能比較を行った．実験ではブロックサイズのパラメータを変化させ，それによる処理速度の傾向の比較を行った．計測は各 3 回ずつ行い，平均値を本実験の結果とする．それぞれのベンチマークで行われる処理は

以下の通りである．

(1) Teragen
1 レコード 100 バイトの文字列を指定された数だけ生成するプログラム．書込み性能を測るときに用いられる．

(2) Grep
入力されたデータの中に出現する指定された文字列の回数をカウントするプログラム．読込み性能を測るときに用いられる．

(3) Terasort
入力されたデータをソートして出力するプログラム．書込み性能と読込み性能を測るときに用いられる．

また，Grep と Terasort においては，処理を行うためにデータを用意する必要があるが，今回は Teragen で生成したランダムな文字列データを用いて実験を行った．

今回の実験で扱うデータのサイズは 20GB，40GB，60GB の 3 種類とし，ブロックサイズは 32MB，64MB，256MB，512MB，1GB の 5 種類でパラメータを変化させ，それぞれの処理を実行した．なお，Terasort 処理に関しては，32MB，64MB，256MB の 3 種類で比較を行った．

ブロックサイズの設定に関して，Isilon では最小 4KB から最大で 1GB とすることができる．デフォルトのブロックサイズは 64MB となっているため，今回の実験ではデフォルト値である 64MB 前後と中間値の 512MB，最大値の 1GB の値を設定して実験を行った．

4.3 実験結果

通常の HDFS による実験結果を図 2 から 4 に示す．また，HDFS 対応の Isilon を用いて実験を行った結果を図 5 から 7 に示す．縦軸は処理にかかった時間（秒），横軸はデータサイズを表している．

結果から，通常の HDFS による Teragen 処理では，ブロックサイズが大きいほど処理時間が短くなっていることがわかる．特にデータサイズ 60GB での 32MB，64MB と 256MB，512MB，1GB の差は約 50 秒もあった．Isilon による Teragen 処理では，ブロックサイズの違いによる処理時間の差は見られなかった．しかし，通常の HDFS によるものと処理にかかった時間を比較すると，通常の HDFS ではデータサイズ 20GB で 200 秒ほどかかっているのに対し，Isilon を用いた場合では半分の 100 秒で処理が完了していた．

次に Grep 処理の結果より，通常の HDFS による処理では，Teragen と同様にブロックサイズの大きいほうが処理にかかる時間が少ないことがわかる．また，Teragen とは違い，32MB と 64MB との差もデータサイズが大きくなるにつれて広がっていく結果となった．Isilon による Grep 処理では，通常の HDFS による処理と傾向は似ており，ブロックサイズの値は大きいほうが処理時間が短くなったが，最大値の 1GB よりも 256MB のほうが若干早く処理が終わっていた．

最後に Terasort 処理の結果より，通常の HDFS による処理では，Teragen，Grep と同様に，ブロックサイズが大きいほうが処理の速度は速い傾向にあった．また，データのサイズが大きくなるにつれて徐々にその差は広がっていった．しかし，これとは対照的に，Isilon による Terasort 処理では，ブロックサイズが 32MB，64MB の結果がもっとも早く，256MB とは大きく差を開いた結果となった．

以上の実験結果から，通常の HDFS と Isilon では異なる傾向があることが分かった．通常の HDFS では，書込み処理と読込み処理の両方でブロックサイズのパラメータは実験で設定した値の中で最も大きい 1GB がもっとも性能が良く，ブロックサイズが小さくなるにつれて処理速度が遅くなる傾向にあった．Isilon に関しては，Teragen と Grep では通常の HDFS と同様にブロックサイズが大きいほうが性能が良い結果となったが，Terasort においてはこれとは対照的に，ブロックサイズが小さいほど処理速度が早い傾向にあることがわかった．

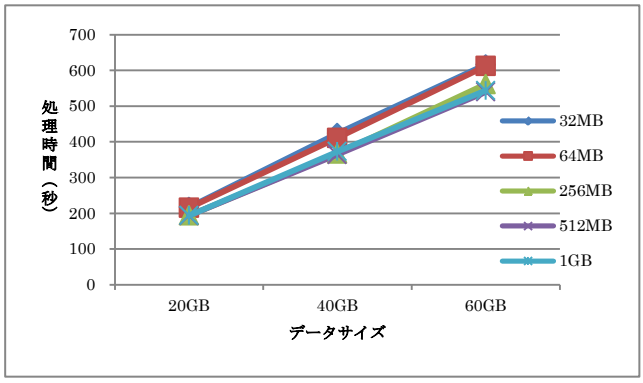

Figure 2 HDFS：Teragen

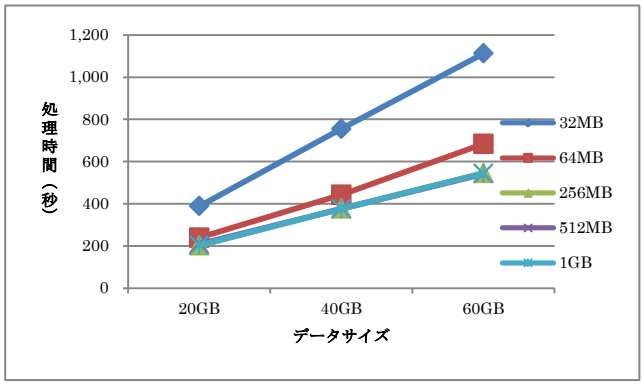

Figure 3 HDFS：Grep

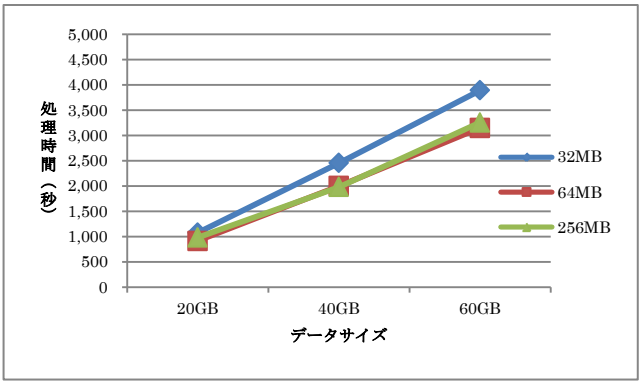

Figure 4 HDFS：Terasort

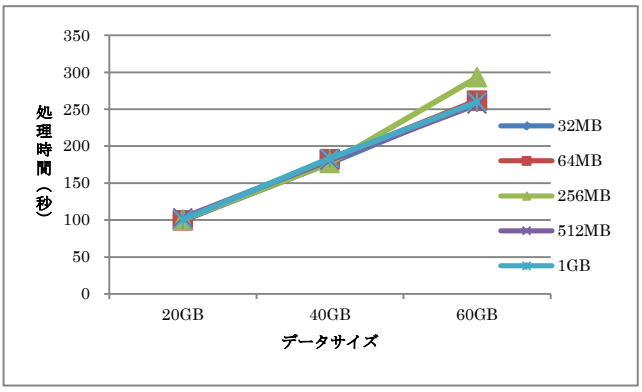

Figure 5 Isilon：Teragen

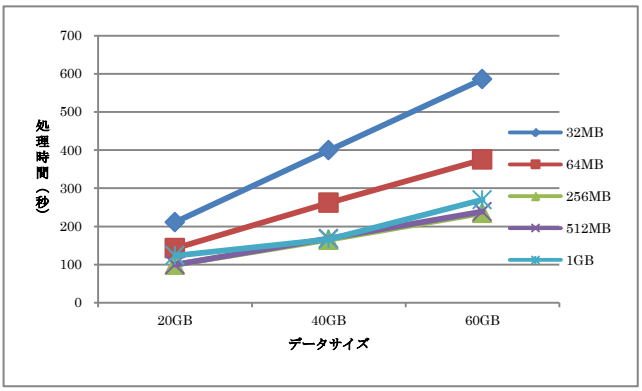

Figure 6 Isilon：Grep

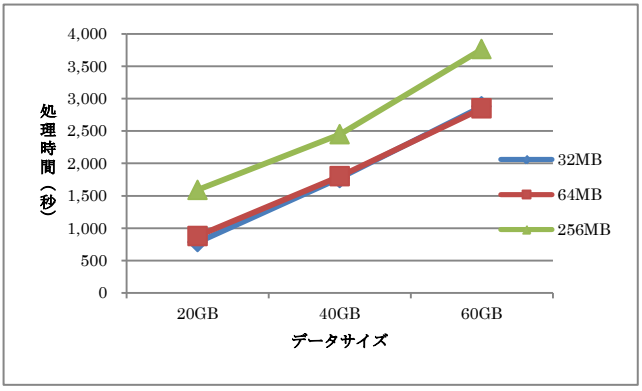

Figure 7 Isilon：Terasort

5. おわりに

本研究では，HDFS 対応の Isilon を用いて Hadoop 環境を構築し，ベンチマークを用いて通常の HDFS と HDFS 対応の Isilon を用いた場合での書込みと読込み性能の比較検証を行った．今後はブロックサイズだけではなく，使用するノードの数やディスクの数といった様々なパラメータの値を変更した場合の処理速度がどのように変化するか傾向を調査したい．

参考文献

1) スケールアウト NAS「アイシロン」のすべて：<http://ascii.jp/elem/000/000/730/730401/>
2) 石井朝葉，金鎔煥，中村純哉，大下福仁，角川裕次，増澤利光，マルチコア CPU 環境における仮想計算機を用いた Hadoop システムの評価，研究報告ハイパフォーマンスコンピューティング（HPC）2012-HPC-136(20),，pp. 1-7, 2012.
3) 櫻井雅志，分散ファイルシステムの性能評価に関する研究，2011
4) 百瀬明日香，小口正人，高遅延環境における分散ファイルシステム Hadoop の遠隔データアクセス特性の評価，電子情報通信学会 DE 研 &PRMU 研 (パターン認識・メディア理解研) 共催，6: 19-24，2011.
5) 奥寺昇平，横山大作，中野美由紀，喜連川優， MapReduce 環境におけるアドホックなクエリを対象とした, Adaptive indexing 適用モデルの提案とその評価，2012.
6) EMC ISILON スケールアウト NAS の高可用性とデータ保護，2013.
7) EMC Isilon スケールアウト NAS による Hadoop ストレージ環境の構築，2012.