MapReduce と Hadoop

金 裕可里，木田 直人
Yukari KIN，Naoto KIDA

1　はじめに

近年，Web ページの検索や，地図の検索を行うために Google のサービスが頻繁に利用されており，必要不可欠なものとなっている.Web 検索は，莫大なデータを扱うのにも関わらず，瞬時に検索結果を表示させることを実現している．この事実の裏側では，想像以上に膨大な計算や多数のコンピュータの働きがある．

Google はその超巨大なコンピュータネットワークを用いて，膨大なデータ処理を分散化している．そのため，大量のデータを瞬時に処理することが可能になっている．その Google の分散処理システムは MapReduce[1] と呼ばれており，Google の検索技術を支えるコア技術である．

本稿では，MapReduce の原理と，MapReduce を実装し，誰でも簡単に大規模分散処理を実行できることを目的としたオープンソースソフトウェアの Hadoop[2] について述べる．

2　MapReduce

2.1　MapReduce の概要

MapReduce とは，多数のマシンで効率的にデータ処理を行う目的で Google によって考案されたフレームワークである．開発者は分散処理の難しい部分を MapRduce に任せることで，少ない労力で大規模な処理を実行できるようになる．

次に MapReduce の処理の流れについて述べる．MapReduce は，Fig. 1 にあるように，Map 処理，シャッフル，Reduce 処理の 3 つの手順から構成されている．

1. Map 処理入力データ（キーと値のペア）を受け取り，任意の形式に変換することで，必要な情報を抽出する．全ての Map 処理は並列実行することができる．
2. シャッフル
   Map によって作られたデータを整理し，データを任意の順に並べ替える．
3. Reduce 処理
   データをまとめて最終的に手に入れたい結果を作り上げるプロセスで，データ全体についての整理された処理結果を得る.

MapReduce のデータ処理をより理解するため，次節に具体例を述べる．

2.2　MapReduce 処理の具体例（転置インデックスの作成）

転置インデックスとは，単語とその単語がある Web ページをリストとしたものである．検索エンジンを実現

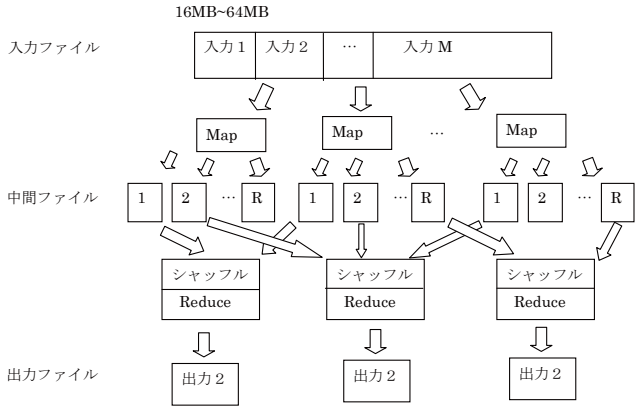

Fig.1　MapReduce の流れ（参考文献 [3] より参照）

するにあたって，まず必要なものとなる．

2.2.1　Map

まず，Fig. 2 に示すように，入力ファイル（Google が持つ全 Web ページ）を複数のファイルに分割する．分割は一般的に 16〜64MB ごとに行われる．入力ファイルが仮に 1MB だとすると，分割されたファイルは数万に及ぶ．次に，これらのファイルが，手の空いている PC に対して順に分配される．各 PC はファイルからデータ（キー＝ Web ページのアドレス，値＝各 Web ページの全テキストデータ）を次々と読み込み，開発者が用意した Map を呼び出す．

Map は新たなデータ（キー＝テキストデータを分割してできた単語，値＝その単語がある Web ページのアドレス）を出力する．PC はしばらくこれをメモリ上に蓄えるが，定期的に中間ファイルに保存する．中間ファイルは式 (1) の分割関数と呼ばれる関数に従って，あらかじめ指定した数（R）のファイルに分割される．

$$hash(\text{キー}) mod R \tag{1}$$

2.2.2　シャッフル

Fig. 3 に示すように，Map で中間ファイルが生成されると，ネットワークを経由して，Reduce 処理をする PC にその場所が伝えられ，シャッフルが始まる．

シャッフルでは，中間ファイルに書き込まれたキー（テキストデータを分割した単語）に従って，全てのデータが整理される．キーはシャッフルによって辞書順にちいさいものから順に選ばれるので Reduce の出力は必ずキーの順にソートされていると言える．

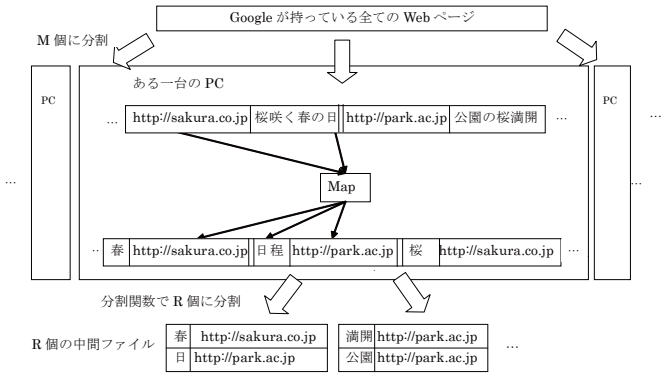

Fig.2　Map 処理の流れ（参考文献 3）より参照）

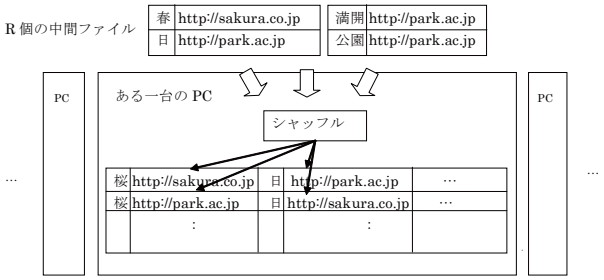

Fig.3　シャッフル処理の流れ（参考文献 3）より参照）

2.2.3　Reduce

Fig. 4 に示すように，Reduce は，シャッフルの終わったグループから順に始められる．各グループの中間ファイルには複数のキー（テキストデータを分割してできた単語）が書き込まれているので，同じキーを持つすべての値が集められて Reduce が呼び出される．

Reduce に渡される Reduce の出力はグループごとに一つのファイルとして Google の巨大分散ファイルシステムの GFS（Google File System）に書き込まれる．結果としてグループの数（R 個）の出力ファイルが生成される．

2.3　MapReduce が有効な処理

MapReduce が有効に働く処理に，以下のものがある．

- 検索エンジンの転置インデックス作成
- grep
- ソート
- 平均値と分散計算
- PageRank 計算
- PageRank の高いウェブページを検索
- ドキュメント内のリンクの収集
- ログ解析

これまでに述べた MapReduce は Google 独自の技術であり，誰もが使えるものではない．しかし，同様の技術を実現している Hadoop と呼ばれるオープンソースソフトウェア（OSS）がある．次章では分散処理を誰にでも簡単に行うことのできる Hadoop について述べる．

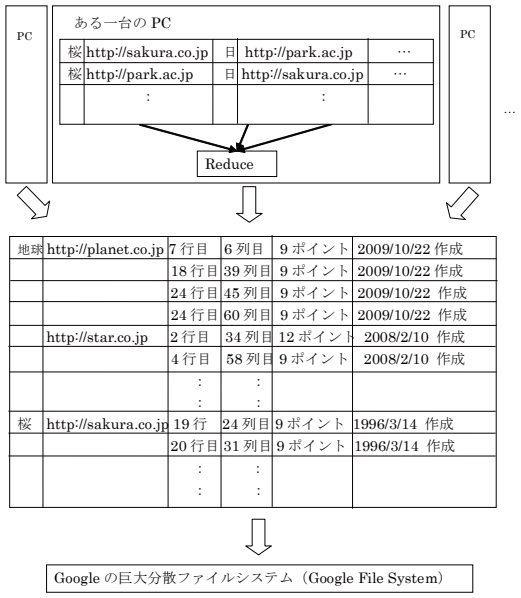

Fig.4　Reduce 処理の流れ（参考文献 3）より参照）

3　Hadoop

Hadoop とは, 大量のデータを手軽に複数のマシンに分散して処理できるオープンソースのプラットフォームである．Hadoop におけるデータ分散処理のベースとなっているのは前章で述べた MapReduce である．

Hadoop とは, 面倒な分散処理を, プログラマが「簡単に」扱えることを目的としたプラットフォームである．

3.1　Hadoop の構造

Hadoop は Google の基盤ソフトウェアである GFS(Google File System) と，Mapreduce のオープンソース実装である．Hadoop は HDFS(Hadoop Distributed File Systm)，Hadoop Mapreduce から構成されている．Fig. 5 にあるように，Google の基盤ソフトウェアに対応させると, 前者は, GFS, 後者は Mapreduce に対応する．分散データベース Big Table も hBase と呼ばれるオープンソース実装によって実装された．

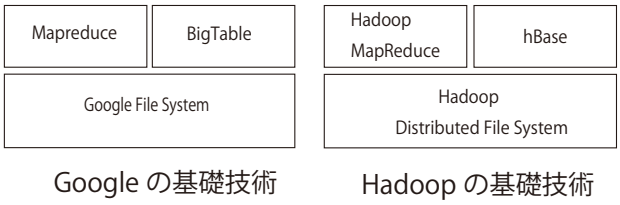

Fig.5　Google，Hadoop 基盤技術の対応関係（参考文献 2）より参照）

3.2　HDFS の概要

HDFS の構造は GFS の構造を踏襲している．GFS とは多数の PC 上に構築される分散ファイルシステムのことである．PC を追加するだけで，システムを停止する

ことなくディスク容量・読み込み性能・書き込み性能を拡大でき，スケーラビリティがあるといえる．また，大容量であり，数百 TB クラスの大規模なディスクの提供を可能にしている．さらに，障害を自動的に監視・検出・修復が可能である．

GFS の特徴としては同じファイルを異なるマシンに重複して持たせることで, 一部のマシンが故障してもファイルが失われないという点が挙げられる（冗長化）．Google では何万台ものサーバーが常時稼動しているので，1 日に多くのマシンが壊れる．それに耐えられるような耐故障性の高い分散ファイルシステムを持つことは巨大データを扱う上で必須である．

3.3 Hadoop MapReduce

Hadoop MapReduce の構造は Google MapReduce の構造を踏襲している．Hadoop のみがもつ機能として，ファイルの分散キャッシュ機能と先に延べた Hadoop Streaming がある．

3.3.1 Hadoop Streaming

Hadoop はすべて Java で記述されており, MapReduce 処理を書く場合も基本的には Java でプログラムを書くことが想定されている．ただし Hadoop Streaming という拡張パッケージを用いると，C/C++・Ruby・Python など任意の言語と標準入出力を用いて MapReduce 処理を書くことも出来る．

3.4 Hadoop の利点

Hadoop では Map 関数と Reduce 関数を実装する．Shuffle はすでに実装されているため，利用者が作るのは Map と Reduce だけでよい．Hadoop の仕組みにのっとって機能を実装さえすれば，自動で分散処理が行えるのだ．1 台では処理にかなり時間がかかるような大量のデータも,Hadoop によって, 複数マシンに分散させることで, 驚くべきスピードで処理を行うことができる. 例えば, 今まで 1 台でやっていた, あるログ集計処理を, Hadoop (マスタ 1 台, スレーブ 19 台) で行うようにしたところ, 従来は 6 時間 6 分 35 秒だったのが, Hadoop を使うと,5 分 34 秒になったという結果がでている．分散コンピューティングと聞くと，研究室レベルでの活用とか，Google や Yahoo!，IBM といった大企業しか使えないのではないかと思われがちだ．Hadoop の流儀にしたがって機能を実装すれば簡単に分散処理を実現できる．

4 まとめ

MapReduce と Hadoop は莫大になったデータをより効率的に，安価に，簡単に処理できるようにするために編み出された工夫である．MapReduce はプログラミングモデルの名前であり，Hadoop は MapReduce を活かして作られたオープンソースのプラットフォームである．これらの利点から，Yahoo!をはじめとする様々な企業が積極的に導入しており，改良も進んでいる．

5 今後の展望

米国 Amazon．com 傘下の Amazon Web Services（AWS）は 4 月 2 日,「Amazon Elastic MapReduce」のベータ版をリリースしたと発表した．Hadoop クラスタのセットアップはかなり複雑な作業であった．しかし，Amazon Elastic MapReduce は，Hadoop クラスタを大幅に利用しやすくすることを目指したサービスであるのだ．

同社によると，Elastic MapReduce を使えば,

- 手軽なポイント＆クリック操作で Hadoop ジョブを作成，実行，監視，制御できる．
- ハードウェアを大量購入することもなく，（ハードウェアを）ラックに収めてネットワークにつないで管理するといった作業も必要もない．
- リソースが足りなくなることや，組織内のほかのメンバーとの共有について心配することもない．監視もチューニングも不要
- システムやアプリケーション・ソフトウェアのアップグレードに時間をかける必要もない．
- 時間単位課金で実行できる．

という機能がある．

参考文献

1) Jeffrey Dean, Sanjay Ghemawat：MapReduce: Simplified Data Processing on Large Clusters, OSDI'04: Sixth Symposium on Operating System Design and Implementation，December，2004
2) エヌ・ティ・ティ レゾナント株式会社，株式会社 Preferred Infrastructure：Hadoop 調査報告書，2008
3) 西田圭介：Google を支える技術 巨大システムの内側の世界，技術評論社，2008