

2010 年度 自主課題研究概要レポート

Hadoop/HBase を用いた分散データベースサービスの構築・運用

情報システムコース 3 年 名列:265 竹下 淳（共同研究者:清水 裕亮）

1. 目的

本自主課題研究では、Google の基盤技術ⁱを参考に作成された、Hadoop/HBase を利用し分散データベースを構築し、それを用いた検索システムの構築・運用を目的とする。

2. Hadoop/HBase とは

Hadoop は、複数のマシンをネットワーク上で束ねて 1 つの空間としてデータを保持する分散ファイルシステム HDFSⁱⁱと分散処理環境フレームワーク HadoopMapReduce から構成されている。これらと、分散データベースⁱⁱⁱである HBase を用いることで複数のマシンにスケールアウト^{iv}して大規模データに対する処理を効率的に行えるようになっている。

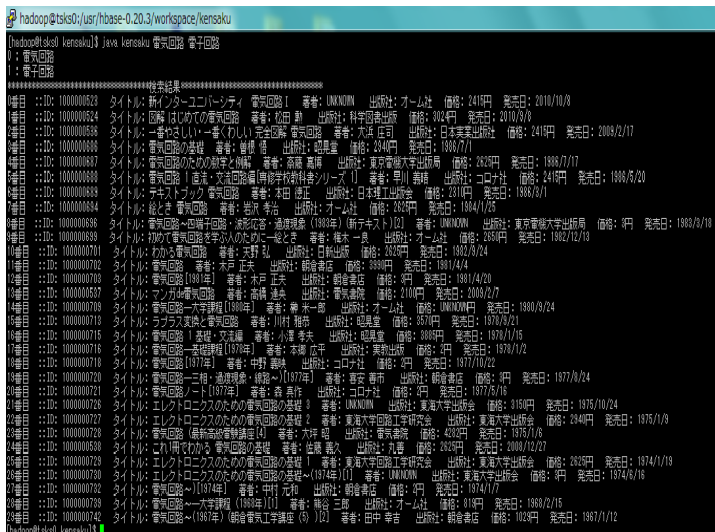
3. 開発したシステム

作成する検索システムは、書籍データを扱っている。検索は、書籍のタイトルに対して検索用インデックスを作成、HBase で保持し、クライアントが入力した検索ワードとインデックスがヒットする書籍データを検索結果として画面へ表示する。複数の検索ワードがある場合、ワードとのヒット数が多い書籍から順に表示を行う。

役割分担は、検索システムなどのプログラム作成・書籍データ収集を竹下が、HBase へのデータ格納を清水が担当している。なお、プログラム作成は JDK-1.6.0_21 を用いた Java で行っている。システム開発の過程の、検索インデックスを作成する部分で Hadoop の MapReduce を利用している。

図 1 に検索システムの実行結果を示す。初めに“電子回路”，“電気回路”を検索ワードとして入力しており、検索結果で検索ワードを含む書籍のデータを取得することに成功してい

る。なお、本来は HBase アクセス時のログが実行結果と同時に排出されるが、図 1 では割愛している。



ID	検索結果
1000000523	タイトル: 新インテュスコノミシティ 電気回路 I 著者: UNKNON 出版社: オーム社 価格: 241円 発売日: 2010/10/8
1000000524	タイトル: 図解 はじめての電気回路 著者: 丸岡 剛 出版社: 科学図書出版 価格: 3024円 発売日: 2010/9/9
1000000538	タイトル: 一番やさしい 一番くわしい 完全図解 電気回路 著者: 大浜 正司 出版社: 日本実業出版社 価格: 2419円 発売日: 2009/2/17
1000000608	タイトル: 電気回路の基礎 著者: 曾根 信 出版社: 朝倉書店 価格: 2340円 発売日: 1989/7/1
1000000637	タイトル: 電気回路のための数学と物理 著者: 斎藤 義典 出版社: 東京電機大学出版局 価格: 2820円 発売日: 1988/7/17
1000000690	タイトル: 電気回路 I 直交・交流回路編 鳥羽学校新書シリーズ I 著者: 早川 義晴 出版社: コロナ社 価格: 2419円 発売日: 1989/5/20
1000000693	タイトル: テキストブック 電気回路 著者: 本田 信正 出版社: 日本理工出版会 価格: 2310円 発売日: 1988/3/1
1000000694	タイトル: 総とき 電気回路 著者: 岩沢 幸治 出版社: オーム社 価格: 245円 発売日: 1984/1/25
1000000695	タイトル: 電気回路～応用電子回路～演習問題 (1989年) (新テキスト) ② 著者: UNKNON 出版社: 東京電機大学出版局 価格: 9円 発売日: 1989/9/18
1000000699	タイトル: 初めて電気回路を学ぶ人のために 総とき 著者: 佐々木 一 出版社: オーム社 価格: 2350円 発売日: 1982/12/19
1000000701	タイトル: わかる電気回路 著者: 矢野 弘 出版社: 日刊工業新聞 価格: 2020円 発売日: 1982/9/24
1000000702	タイトル: 電気回路 著者: 木戸 正夫 出版社: 朝倉書店 価格: 399円 発売日: 1981/4/4
1000000709	タイトル: 電気回路 (1981年) 著者: 木戸 正夫 出版社: 朝倉書店 価格: 9円 発売日: 1981/4/29
1000000637	タイトル: マンガ版電気回路 著者: 斎藤 義典 出版社: オーム社 価格: 2100円 発売日: 2009/2/7
1000000709	タイトル: 電気回路―大卒課程 (1989年) ① 著者: 藤 孝一 出版社: オーム社 価格: UNKNON円 発売日: 1980/9/24
1000000719	タイトル: ラプラス変換と電気回路 著者: 川村 雅市 出版社: 朝倉書店 価格: 3570円 発売日: 1979/9/21
1000000715	タイトル: 電気回路 I 基礎・交流編 著者: 小澤 孝夫 出版社: 朝倉書店 価格: 398円 発売日: 1979/7/15
1000000716	タイトル: 電気回路―基礎課程 (1979年) 著者: 藤 孝一 出版社: 朝倉書店 価格: 9円 発売日: 1979/7/12
1000000719	タイトル: 電気回路 (1979年) 著者: 中野 義典 出版社: コロナ社 価格: 9円 発売日: 1977/10/22
1000000720	タイトル: 電気回路―基礎・交流編 ① (1979年) 著者: 藤 孝一 出版社: 朝倉書店 価格: 9円 発売日: 1977/9/24
1000000721	タイトル: 電気回路ノート (1979年) 著者: 森 真作 出版社: コロナ社 価格: 9円 発売日: 1977/5/16
1000000726	タイトル: エレクトロニクスのための電気回路の基礎 ① 著者: UNKNON 出版社: 東海大学出版会 価格: 3150円 発売日: 1975/10/24
1000000727	タイトル: エレクトロニクスのための電気回路の基礎 ② 著者: 東海大学回路工学研究会 出版社: 東海大学出版会 価格: 2840円 発売日: 1975/7/9
1000000728	タイトル: 電気回路 (橋本高次郎著) ④ 著者: 大坪 昭 出版社: 朝倉書店 価格: 429円 発売日: 1975/7/6
1000000638	タイトル: これ1冊でわかる 電気回路の基礎 著者: 佐藤 寛久 出版社: 丸善 価格: 2820円 発売日: 2000/12/27
1000000729	タイトル: エレクトロニクスのための電気回路の基礎 I 著者: 東海大学回路工学研究会 出版社: 東海大学出版会 価格: 2830円 発売日: 1974/1/19
1000000730	タイトル: エレクトロニクスのための電気回路の基礎～(1974年) ① 著者: UNKNON 出版社: 東海大学出版会 価格: 9円 発売日: 1974/6/16
1000000732	タイトル: 電気回路―I (1974年) 著者: 中野 義典 出版社: 朝倉書店 価格: 9円 発売日: 1974/1/7
1000000739	タイトル: 電気回路―大卒課程 (1985年) ① ② 著者: 藤 孝一 出版社: オーム社 価格: 815円 発売日: 1983/2/15
1000000742	タイトル: 電気回路～(1987年) (朝倉電気工学講座 (9) ①) ② 著者: 田中 幸吉 出版社: 朝倉書店 価格: 1020円 発売日: 1987/7/12

図 1. 検索システム実行結果

4. まとめ・反省

初期構想では、検索システムを Web ページ上で利用できるサービス運用を目指していたが、役割分担の時点でうまく作業の分担が出来ていなかったため負担が一方に偏り、検索システムの動作確認までに終わってしまったことが悔やまれる。しかし、そんな中でも、コンピュータシステムを運用するにあたり、負荷分散の面で重要となるだろうスケールアウトの考え方を Hadoop/HBase を利用することで学べたことが非常に有意義だった。

研究全体を通してチームによる共同開発を行うという場面で、自分達はまだ経験不足であることを改めて実感し、今回の研究で培った経験を今後のシステム設計・開発などの場面に活かし、このような活動に積極的に参加していきたいと思う。

ⁱ Hadoop の基盤技術の HDFS・HBase・HadoopMapReduce は Google の基盤技術 GFS・BigTable・MapReduce を参考に作成されている

ⁱⁱ Hadoop Distributed File System の略

ⁱⁱⁱ 格納するデータを複数のマシンに分散しているデータベース

^{iv} サーバ処理能力向上の手法の 1 つであり、サーバの台数を増やし負荷分散させることでサーバ能力を向上させる ⇔ スケールアップ