

오프라인 마켓에 적용 가능한 빅데이터 분석 시스템 구축 방안에 관한 연구

이후영*, 박구락**, 김동현***

공주대학교 멀티미디어공학과*, 공주대학교 컴퓨터공학부**, 우송대학교 IT융합학부***

A Study on Possible Construction of Big Data Analysis System Applied to the Offline Market

Hoo-Young Lee*, Koo-Rack Park**, Dong-Hyun Kim***

Dept. of Multimedia, Kongju National University*

Dept. of Computer Science & Engineering, Kongju National University**

Dept. of IT Convergence, Woosong University***

요 약 빅데이터는 현재 기업 경쟁력의 주요 자산으로 여겨지고 있고 향후에 그 영향력은 더욱 확대될 것으로 전망된다. 그 중요성을 인식한 기업들은 이미 빅데이터를 제품 개발과 마케팅에 적극적으로 활용하고 있으며 정치, 스포츠 등 사회 전반에 걸쳐 적용분야는 점점 늘어나고 있다. 그러나 시스템 구축에 따른 노하우 부족과 고비용은 빅데이터 시스템 도입에 여전히 큰 장애가 되고 있다. 본 논문에서는 중소기업 오프라인 마켓의 POS 판매 데이터를 빅데이터 시스템 중 오픈소스인 하둡(Hadoop) 및 하이브(Hive)를 기반으로 하는 빅데이터 시스템 구현을 목표로 한다. 이러한 융복합을 통해 단순히 손익분석과 재고관리 등에 집중되었던 기존 판매 시스템을 보완하여 고객의 소비 패턴과 선호도 조사, 수요에 대한 사전 예측이 가능하도록 하는 경영자의 합리적인 의사결정에 기초자료로 활용할 수 있을 것으로 기대된다.

주제어 : 빅데이터, 하둡, 하이브, 맵리듀스, 판매시점정보관리, 오프라인 마켓, 융복합

Abstract Big Data is now seen as a major asset in the company's competitiveness, its influence in the future is expected to grow. Companies that recognize the importance are already actively engaged with Big Data in product development and marketing, which are increasingly applied across sectors of society, including politics, sports. However, lack of knowledge of the system implementation and high costs are still a big obstacles to the introduction of Big Data and systems. It is an objective in this study to build a Big Data system, which is based on open source Hadoop and Hive among Big Data systems, utilizing POS sales data of small and medium-sized offline markets. This approach of convergence is expected to improve existing sales systems that have been simply focusing on profit and loss analysis. It will also be able to use it as the basis for the decisions of the executive to enable prediction of the consumption patterns of customer preference and demand in advance.

Key Words : Big Data, Hadoop, Hive, MapReduce, POS, Offline Market, Convergence

Received 6 July 2016, Revised 15 August 2016
Accepted 20 September 2016, Published 28 September 2016
Corresponding Author: Koo-Rack Park
(Kongju National University)
Email: ecgrpark@kongju.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

현대사회로 들어서면서 사회적으로 많은 변화가 일어나고 있으며, 특히 소비자들의 소비 형태가 바뀌어 가고 있는 실정이다. 이러한 소비 패턴의 변화에는 IT 기술의 발달이 많은 영향을 미치고 있으며, 특히 스마트 기술의 발전은 온라인 쇼핑몰의 발전과 확산을 더욱 가속화시키고 있고[1], 기존의 구매 행태인 구전 중심에서 구매하고자 하는 제품에 대한 많은 정보를 비교 및 입수한 후 구매하는 경향이 일반화되고 있다[2]. 그러나 소비자들은 많은 정보와 대안을 중심으로 논리적인 방법에 의해 인지적 구매만 하는 것이 아니라 개인차에 의한 습관적인 구매결정이 이루어지는 것이 일반적인 상황이다[3].

또한, 소량의 필요 품목을 중심으로 하여 주거지 근처의 근거리에서 구매하는 경향이 커짐에 따라 편의점 및 슈퍼마켓 등으로 진화하고 있다. 이러한 소매점을 효율적으로 관리하기 위해서 적정한 상품의 재고에 대한 관리가 필요하고[4], 재고 관리를 위하여 프랜차이즈형 판매점의 경우, 본사에서 판매시점관리(POS: Point of Sales)를 통하여 상품의 재고 파악을 실시간으로 처리하고 있고, 고객들이 선호하는 제품을 분석할 수 있어 효율적으로 매장을 관리할 수 있다[5].

관리를 위하여 대부분의 매장에서는 판매시점관리를 사용하고 있으며, 이를 기반으로 상품의 제조 및 생산 단계에서 바코드 등을 활용하여 상품관리의 효율성을 증대시키고 있다[6]. 판매시점관리 시스템의 효율성을 높이기 위하여 무선 인터넷 환경을 기반으로 모바일 판매시점관리 시스템 개발에 관한 연구[7]가 이루어지고 있으나, 지금까지 대부분의 연구는 시스템의 편리성에 국한되어 실제 데이터 분석에 관련된 연구는 미미한 상태이다. 또한, 일반적인 판매시점관리 시스템에서 판매와 관련된 데이터는 재고 및 고객관리, 손익분석 등 제한적으로만 사용되고 있어 상품을 구매하는 고객의 구매 패턴 파악에는 거의 사용이 되지 않고 있기에[8], 최근 이슈가 되고 있는 빅데이터 분석을 통하여 구매 패턴을 분석하고자 한다. 빅데이터는 다양한 종류의 데이터를 생성, 수집, 분석 및 표현하면서 개인화된 현재 사회 구성원 마다 맞춤형 정보를 제공할 수 있다[9]. 이러한 빅데이터 동향은 구체적인 활용 산업과 응용 분야에 대한 사례와 벤치마킹 이슈가 주류를 이루고 있어, 빅데이터 분석 기반의 의사결정,

문제해결, 사업 기회 창출이 비즈니스의 중대한 화두가 되고 있다[10,11]. 또한 대규모 데이터 처리와 다양한 형태의 데이터 처리 기술이 진화함에 따라 사회문제 진단 및 현안 해결 도구로서 빅데이터의 잠재력이 주목받고 있으며[12], 최근 들어 데이터 분석 도구를 이용한 사회연결망분석에 관한 연구가 진행되고 있다[13].

이에, 본 논문에서는 중소규모 오프라인 마켓의 판매시점관리 시스템의 정보가 저장되어 있는 데이터베이스에 자료수집 API를 통하여 구매정보 데이터를 텍스트 형태로 추출한 다음 하이브(Hive)를 사용하여 하둡분산파일시스템(HDFS : Hadoop Distributed File System)에 적재하고 이를 HiveQL를 통하여 판매데이터를 분석할 수 있는 빅데이터 분석시스템 구축 방안을 제안한다.

2. 관련 연구

2.1 빅데이터의 활용

데이터를 수집하고 수집된 데이터는 전처리 과정으로 적절한 형태의 데이터 가공 및 필터링을 한다. 가공 데이터의 저장을 체계적으로 관리하여, 유용한 데이터는 정보처리 분석 과정에서 데이터의 시각화 및 가시화를 통하여 활용 가능하도록 한다[14]. 다음의 [Fig. 1]은 수집부터 활용까지의 단계를 나타낸 것이다.



[Fig. 1] Big Data Collection / Utilization Stage

빅데이터의 수집은 내부 시스템을 통해서 상당 부분 획득할 수 있다. 이미 기업이나 기관의 내부 시스템에는 업무 수행 과정을 통하여, RDBMS에 정형화된 많은 데이터가 축적되어 있다. 이밖에도 다양한 형식의 비정형 데이터(텍스트, 이미지, 영상 등)가 있을 수 있다. 그러나 빅데이터 시스템은 이러한 내부 데이터 외에도 SNS, 포털 사이트, 신문기사, 블로그 등 다양한 정보가 수집대상이며 이런 데이터 역시 빅데이터로 활용 될 수 있다.

이렇게 수집된 데이터들은 분산파일시스템에 저장되

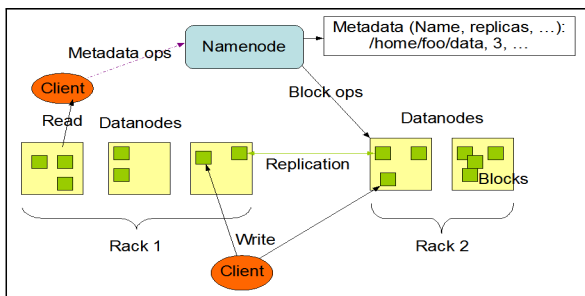
어 여러 저장소에 분리하여 저장되며, 빅데이터의 대용량 데이터를 처리하기 위한 개발된 오픈소스 기술인 하둡은 이에 대표적인 기술이다. 하둡분산파일시스템에 저장된 파일은 맵리듀스를 통한 분석 과정을 거쳐 의미 있는 데이터로 가공된다. 이렇게 가공된 데이터는 자연어 처리, 기계학습, 패턴인식, 데이터마이닝 통계 등의 다양한 분석기법을 통하여 분석되고, 데이터 시각화에 의해 금융, 마케팅, 공공분야 등 사회 전반에 걸쳐 사용된다.

2.2 하둡

하둡은 더그 커팅이 2003년 발표된 구글의 논문을 벤치마킹하여 컴퓨터 클러스터를 통한 대용량 데이터 처리가 가능하도록 개발한 오픈 소스 프레임워크이며, 크게 하둡분산파일시스템과 맵리듀스로 구성되어있다.

2.2.1 하둡분산파일시스템

다음의 [Fig. 2]는 하둡분산파일시스템의 구조로서 파일의 메타정보를 가지고 있고 클라이언트에 의한 파일접근을 통제하는 마스터 서버인 네임노드와 이를 백업하는 보조 네임노드, 그리고 스토리지를 관리하는 많은 수의 데이터 노드간의 주/종 구조를 가지고 있다. 또한, 파일의 병렬처리를 위한 고가용성을 제공하고 데이터 결합에 효율적으로 대처 할 수 있도록 각 파일을 1개 또는 그 이상의 블록(기본 64M)으로 나누어 각 데이터노드에 분산 저장된다[15].

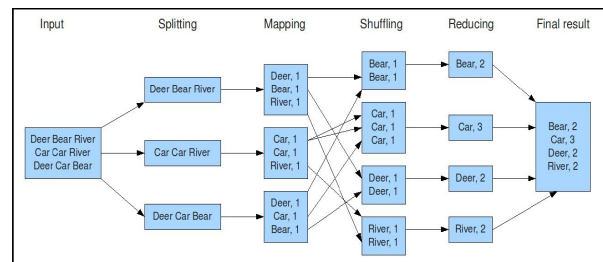


[Fig. 2] HDFS Architecture

2.2.2 맵리듀스

맵리듀스는 많은 양의 데이터를 독립된 작업으로 분산하여 병렬처리가 가능하도록 설계된 프로그래밍 모델이다. 맵리듀스는 전체적인 작업의 상태를 관리하는

JobTracker, 맵 또는 리듀스 작업을 실행하는 TaskTracker로 구성되어 있다. 맵핑의 단계는 입력된 전체 데이터를 나누어 개별적으로 정의된 함수인 맵퍼로 전달되어 중간 데이터 리스트를 생성하는 단계로, 최종 데이터를 생성하기 전의 상태이다. 또한 리듀싱은 중간 데이터 리스트를 이용하여, 리듀서로 정의된 함수를 통해 최종적인 데이터로 취합하는 단계라 할 수 있다[16]. 맵리듀스의 핵심과정은 입력 데이터를 <Key,Value> 의 한 쌍으로 묶은 후 같은 키를 가지는 데이터의 합을 구하는 과정이라고 할 수 있다. 다음의 [Fig. 3]은 하둡 프로젝트에서 기본으로 제공하는 맵리듀싱 프로그램 중 글자의 수를 카운트 하는 프로그램을 이용하여 사용자가 입력한 문자들을 맵리듀싱하는 과정을 나타낸 것이다[17]. 먼저 입력된 데이터는 개행 문자를 기준으로 분리하고, 자간 공백에 의해서 다시 분리된 후 맵핑 과정을 거쳐 기본적인 맵으로 이뤄진 데이터 세트를 완성하여, 셔플링과 리듀싱 과정을 거쳐 입력된 파일에 해당 단어가 몇 번이나 반복되는지 카운트를 하게 되고, 이를 취합하여 최종 파일을 출력하게 된다.



[Fig. 3] The Overall MapReduce Word Count Process

2.2 하이브

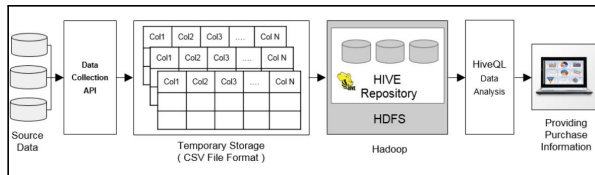
하둡에서 동작하는 아파치 하이브(Apache Hive)는 데이터웨어하우스 형태의 인프라 구조로서 데이터의 요약 및 질의, 분석 기능을 제공한다[18].

또한, 하이브는 맵리듀스 프로그램 작성에 필요한 고수준의 인터페이스를 제공하고 있으며, 쿼리 친화적 언어인 HiveQL이라는 질의문을 통하여 맵리듀스를 구현할 수 있다. 일반적으로 질의 하나에 그 이상의 맵리듀스 잡(Job)으로 실행된다. 따라서 맵리듀스 병렬 프로그래밍에 익숙하지 않은 네트워크 분석가들도 HiveQL의 인터페이스를 통하여 편리하게 질의어를 작성하여 맵리듀스 잡을

실행시킬 수 있다[19]. 이러한 하이브는 UI, Driver, Compiler, Meta Store, Execution Engine 의 5가지 구성요소로 되어 있다[20].

3. 제안 시스템

본 논문에서 제안하는 전체적인 구성은 다음의 [Fig. 4]와 같다.



[Fig. 4] System Configuration

매장에 설치된 판매시점관리 시스템으로부터 발생되는 상품 판매와 관련된 데이터는 판매시스템과 연결된 각각의 데이터베이스에 저장되어 있다. 수집 API는 이러한 데이터베이스에 저장된 데이터를 CSV(Comma Separate Value) 형태의 파일로 변경하여 서버의 임시저장소에 저장하는 역할을 수행한다. 제안 시스템에서 사용된 수집 API의 중요 알고리즘은 [Fig. 5]와 같다.

```

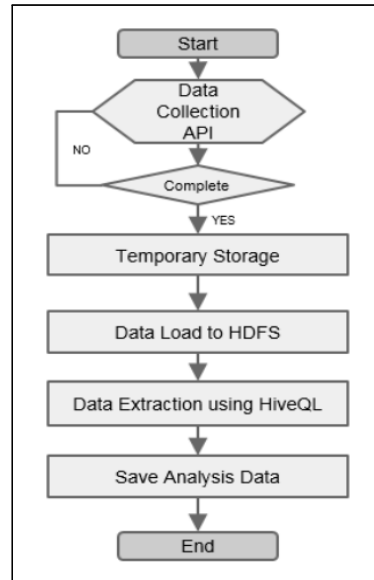
Data Collection API Class {
    Class.forName(Database Driver Loading);
    Connection connection Connection Object Create;
    try {
        connection = DriverManager.getConnection(Database Connect);
        PreparedStatement pstmt Object Create;
        String SQL Generation to be Extracted from the DB;
        pstmt.connection.prepareStatement(Query);
        ResultSet resultSet = pstmt.executeQuery();
        BufferedWriter out = new BufferedWriter(Generating an
Output Object File);
        while(Data Exists in the Result Set) {
            out.write(resultSet.getString(Data Extraction));
        }
        out After Closing the File Output Object;
    } catch() {
        Exception Handling
    } finally() {
        pstmt Object Close;
        connection Object Close;
    }
}
    
```

[Fig. 5] Collection API Class

제안 시스템의 데이터 수집 API는 각 시스템의 서버 환경과 데이터베이스 스키마(Schema)의 구조적 차이를 고려하여 다양한 환경에서도 API의 변경만으로도 분석 시스템을 사용 할 수 있는 장점을 가지고 있다. 수집 API 를 통해 생성된 임시파일은 하이브의 LOAD DATA 명령어를 통해 하둡분산파일시스템의 하이브 저장소에 테이블 형태로 저장되고 하이브에서 제공하는 질의어인 HiveQL을 사용해서 테이블 간의 조인을 통해 데이터를 추출 할 수 있고 이를 서버에 저장한다.

3.1 시스템 프로세스

다음의 [Fig. 6]은 제안 시스템의 프로세스로서, 크게 3단계로 구성되어 있다.



[Fig. 6] System Process

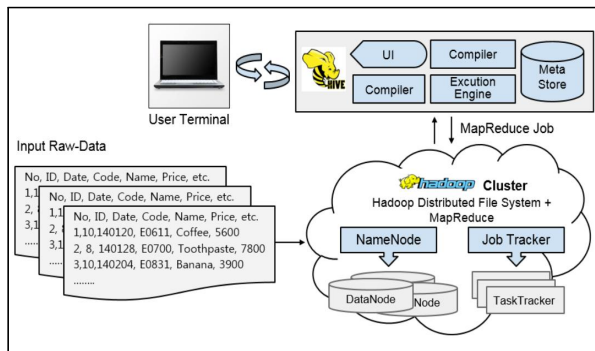
첫째, 데이터 수집 단계로서 중소 규모 매장 안에 설치된 다수의 판매관리시점 시스템은 데이터베이스와 연결되어 판매 데이터를 실시간으로 저장한다. 수집하여 저장된 데이터는 시스템에 따라 다양한 구조(Schema)를 가지고 있기 때문에 직접적으로 하둡분산파일시스템에 적재되지 않고, 구현한 수집 API를 통하여 소비자정보, 제품정보, 구매정보 등을 포함하는 CSV 형태의 데이터로 변환되어 수집 임시 데이터 저장소에 저장된다.

둘째, 하둡분산파일시스템에 데이터를 적재하는 단계로서, 임시저장소에 저장된 데이터를 하이브에서 제공하

는 LOAD DATA 명령을 통해 생성한 하둡분산파일시스템의 하이브 레포지토리 내에 있는 하이브 테이블에 적재하는 단계이다.

셋째, 데이터 분석 단계로서 하이브에서 제공하는 질의어인 HiveQL을 통해 각각의 테이블 사이에 조인을 이용하여 분석 데이터를 추출 한다. 추출된 데이터는 서버에 저장되기 때문에 같은 데이터를 대상으로 한 동일한 질의에 대하여 반복적인 분석 작업이 불필요하다.

하둡분산파일시스템에 저장된 자료는 맵리듀스 과정을 거치게 된다. 다음의 [Fig. 7]은 하이브의 처리과정을 나타낸 것으로 하이브는 HiveQL을 통해 맵리듀스의 사용 편의성을 높이기 위해 등장한 하둡의 서브 프로젝트이다.



[Fig. 7] HiveQL Query Process

4. 실험 및 고찰

본 논문에서 제안한 시스템의 실험 환경 구축을 위해 Mac OS version 10.11.3을 이용하여 서버환경을 구축하였다. 데이터베이스는 MySQL-5.1.41를 사용하였고, 빅데이터 분석을 위해 Hadoop 2.6.0 과 하이브 0.14.0을 이용하였다.

실험은 가상의 샘플 데이터를 사용하여 특정 고객의 물품구입 정보를 통해 선호 제품 목록을 추출하는 방법으로 다음과 같이 실험을 진행하였다.

첫째, 자바로 구현한 수집 API를 통해 고객의 정보가 임시저장소인 레포지토리에 CSV 형태의 파일로 저장된다. 하둡분산파일시스템을 실행한 후 하이브 콘솔 명령으로 고객 테이블, 상품 테이블, 구매정보 테이블을 각각 생성한다.

다음의 <Table 1>은 테이블 구조의 예이다.

<Table 1> Table Structure

Customer	Product	Purchase
customer_id	product_id	customer_id
customer_name	product_company	product_id
customer_gender	product_kind	purchase_code
customer_address	product_name	purchase_date
customer_phone	product_cost	purchase_cost

둘째, 각각의 테이블을 생성한 후 하이브 콘솔에 Load Data 명령을 실행하여 임시저장소의 파일을 하둡분산파일시스템의 하이브 테이블에 고객 데이터, 물품 데이터, 고객별 상품 구매 데이터를 다음의 [Fig. 8]과 같이 각각의 테이블에 입력한다.

```
hive> LOAD data local inpath '/User/bigdata/Documents/sample_customer_table.csv'
overwrite into table customer_table;
hive> LOAD data local inpath '/User/bigdata/Documents/sample_purchase_table.csv'
overwrite into table purchase_table;
hive> LOAD data local inpath '/User/bigdata/Documents/sample_product_table.csv'
overwrite into table product_table;
```

[Fig. 8] Sample Data Input

셋째, 샘플 데이터 셋의 입력이 완료된 후 다음의 [Fig. 9]과 같이 HiveQL을 통하여 주요 3개 테이블의 정보를 조인하여 특정 소비자의 물품구매 선호도, 지출금액, 매장일 방문 패턴 등을 분석 할 수 있다.

```
hive> select
    ct.customer_id, pt.product_kind, count(pt.product_kind) cnt
from
    customer_table ct
    Inner join purchase_table bt on ct.customer_id = bt.customer_id
    Inner join product_table pt on bt.product_id = pt.product_id
Group by ct.customer_id, pt.product_kind
having ct.customer_id = 'gildonghong';
```

[Fig. 9] HiveQL Table Join and Group

다음의 [Fig. 10]은 구매 물품의 선호도 분석의 데이터 추출 결과로서 고객이 선호하는 구매 물품을 카운트 하는 과정이다.

```

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2016-07-25 23:31:24,928 Stage-2 map = 0%, reduce = 0%
2016-07-25 23:31:32,381 Stage-2 map = 100%, reduce = 0%
2016-07-25 23:31:40,800 Stage-2 map = 100%, reduce = 100%
Ended Job = job_1469454979716_0006
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 HDFS Read: 2234 HDFS Write: 223 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
gildonghong Processed food 13
gildonghong Biscuit 4
gildonghong agriculture products 7
gildonghong Noodles 3
gildonghong Fish cake 3
gildonghong Plastic Container 2
gildonghong Milk 2
gildonghong Meat 9
gildonghong Drink 4
gildonghong Toothpaste 2
Time taken: 35.74 seconds, Fetched: 10 row(s)
    
```

[Fig. 10] Data Extraction Result

판매시점관리 시스템으로부터 데이터베이스로 입력된 판매 데이터를 기반으로 오픈소스인 하둡과 하이브를 사용하여 특정 고객의 구매목록을 분석하는 시스템을 실험하였다. 수집 API를 이용하여 이기종 시스템 간에 데이터 확보가 가능하도록 설계하여 활용도를 높였고, SQL(Structured Query Language)과 유사한 형태인 HiveQL을 지원하는 하이브를 설치하여 데이터 분석의 편의성을 향상시켰다. 본 제안 시스템을 통하여 판매시점관리 시스템의 판매 데이터를 대상으로 다양한 분석이 가능하며, 분석 대상 고객의 소비 패턴 및, 물품 구매 선 호도 조사에 이용될 수 있고, 또한 경영자의 입장에서는 및 고객 수요에 대한 사전 예측 등의 분석 자료로 사용할 수 있어 합리적 의사결정에 중요한 지표가 될 수 있을 것으로 기대된다.

5. 결론

빅데이터는 일시적인 트렌드 이거나 혹은 실생활과는 멀리 떨어져 일부 분야에서만 사용하는 특별한 기술이 아니다. 빅데이터는 이미 정치, 문화, 경제, 스포츠 등 일상 전반에 걸쳐 적용되고 있는 기술이며, 향후 그 파급영역은 더욱더 늘어날 것으로 예상하고 있다. 그러나 많은 장점을 가지고 있지만, 빅데이터 시스템 구축과 분석의 어려움 등으로 일부 대기업을 중심으로 금융과 전자통신 분야에 집중되어 사용되고 있는 실정이다. 이에 본 논문에서는 중소규모 오프라인 매장에서 발생하는 대량의 구매 데이터를 바탕으로 오픈소스인 하둡과 분석 툴인 하이브를 이용하여 빅데이터 처리 시스템 구축 방안

을 제안하였다. 이를 통해 기존의 POS 시스템이 손익분석, 재고관리에 집중되었던 문제를 개선하여 대규모 데이터의 처리와 분석이 가능하며, 이를 통해 경영자의 합리적 의사결정에 자료를 확보하여 중소 마켓이 대형 마켓에 대한 경쟁력을 갖출 수 있을 것이다. 향후 연구에서는 더 많은 판매데이터 수집을 위해 각각의 다른 데이터베이스시스템 간에 이 데이터 수집이 가능하도록 하는 데이터 수집 API 연구와 더불어 다양한 분석 기법 및 시각화에 대한 연구가 계속되어야 할 것이다.

REFERENCES

- [1] Q. Y. Hao, S. J. Lee, K. R. Lee, "The Acceptance of Customer Reviews in Taobao", Journal of the Korea Convergence Society, Vol. 6, No. 4, pp. 205-212, 2015.
- [2] S. H. Park, I. H. Lee, H. Y. Ahn, "Analysis of the Retail Channels Resulting from Changing Consumption Trends", Hana Institute of Finance, No. 13, 2009.
- [3] G. D. Seo, J. E. Lee, "A study on the Effect of Consumer Lifestyle on Brand Attitude, Brand Attachment influence upon Brand Loyalty", Journal of Digital Convergence, Vol. 14, No. 4, pp. 185-192, 2016.
- [4] K. W. Ko, D. C. Kim, "The Analyses of the Operational Efficiency and Efficiency Factors of Retail Stores Using DEA Model," Korean Management Science Review, Vol. 31, No. 4, pp. 135-150, 2014.
- [5] K. C. Ahn, et al., "POS Data Analysis System based on Association Rule Analysis", Korea Society of Industrial Information Systems, Vol. 17, No. 5, pp. 9-17, 2012.
- [6] James Dyson, "Paperless Receipt Solution (PRS) System", James Dyson Foundation, 2015.
- [7] O. B. Kwon, H. C. Shin, "Mobile Point-of-Sales System", Journal of Information and Security, Vol 7, No. 3, pp. 87-93, 2007.
- [8] K. C. Ahn, et al., "POS Data Analysis System based on Association Rule Analysis", Journal of the Korea Industrial Information System Society, Vol. 17, No. 5, pp. 9-17, 2012.
- [9] Y. S. Jeong, Y. T. Kim, G. C. Park,, "Subnet Selection

Scheme based on probability to enhance process speed of Big Data”, Journal of Digital Convergence, Vol. 13, No. 9. pp. 201-208, 2015.

- [10] S. H. Namn, K. S. Noh, “A Study on the Effective Approaches to Big Data Planning”, Journal of Digital Convergence, Vol. 13, No. 1, pp. 227, 235, 2015.
- [11] K. S. Noh, S. T. Park, K. H. Park, “Convergence Study on Big Data Competency Reference Model”, Journal of Digital Convergence, Vol. 13, No. 3, pp. 55-63, 2015.
- [12] B. C. Kim, “A study on Utilization of Big Data Based on the Personal Information Protection Act”, Journal of Digital Convergence, Vol. 13, No. 12, pp. 87-92, 2014.
- [13] K. C. Choi, J. A. Yoo, “A reviews on the social network analysis using R”, Journal of the Korea Convergence Society, Vol. 6, No. 1, pp. 77-83, 2015.
- [14] G. S. Hang, “Big Data Platform Strategy: Big Data is Changing Business Platform Future Revolution”, Electronic Times, pp. 83-97, 101-105, 193-203, 2013.
- [15] <http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html>
- [16] J. H. Kwak, et al., “Large-scale Data Analysis based on Hadoop for Astroinformatics”, Journal of KIISE, Vol. 17, No. 11, pp. 587-591, 2011.
- [17] http://hci.stanford.edu/courses/cs448g/a2/files/map_reduce_tutorial.pdf
- [18] Venner, Jason, “Pro Hadoop”, Apress, 2009.
- [19] Y. H. Lee, Y. S. Lee, “Yet Another BGP Archive Forensic Analysis Tool Using Hadoop and Hive”, Journal of KIISE, Vol. 42, No. 4, pp. 541-549, 2015.
- [20] Apache Hive: <https://cwiki.apache.org/confluence/display/hive/Design>

이 후 영(Lee, Hoo Young)



- 2002년 2월 : 우송대학교 컴퓨터과 학과 (공학사)
- 2015년 3월 ~ 현재 : 공주대학교 대학원 멀티미디어공학과
- 관심분야 : 빅데이터, 하둡, 하이브, 정보보안
- E-Mail : hooyoung.paul.lee@gmail.com

박 구 락(Park, Koo Rack)



- 1986년 2월 : 중앙대학교 전기공학과 (공학사)
- 1988년 2월 : 숭실대학교 전자계산학과 (공학석사)
- 2000년 2월 : 경기대학교 전자계산학과 (이학박사)
- 1991년 4월 ~ 현재 : 공주대학교 컴퓨터공학부 교수
- 관심분야 : 경영정보, 정보통신, 전자상거래
- E-Mail : ecgrpark@kongju.ac.kr

김 동 현(Kim, Dong Hyun)



- 1986년 2월 : 중앙대학교 전기공학과 (공학사)
- 2005년 2월 : 공주대학교 컴퓨터멀티미디어공학과 (공학석사)
- 2010년 2월 : 공주대학교 컴퓨터공학과 (공학박사)
- 2016년 2월 ~ 현재 : 우송대학교 IT융합학부 겸임교수
- 관심분야 : 빅데이터, 영상처리, 지리정보, 시물레이션
- E-Mail : dhkim977@naver.com