

하둡기반 빅데이터 시스템을 이용한 스마트그리드 전력데이터 분석

Analyzing Smart Grid Energy Data using Hadoop Based Big Data System

조영탁* · 이원진** · 이인규*** · 온병원† · 최중인§

(YoungTak-Cho · WonJin Lee · Ingyu Lee · Byung-Won On · Jung-In Choi)

Abstract - With the increasing popularity of Smart Grid infrastructure, it is much easier to collect energy usage data using AMI (Advanced Measuring Instrument) from residential housing, buildings and factories. Several researches have been done to improve an energy efficiency by analyzing the collected energy usage data. However, it is not easy to store and analyze the energy data using a traditional relational database management system since the data size grows exponentially with an increasing popularity of Smart grid infrastructure. In this paper, we are proposing a Hadoop based Big data system to store and analyze energy usage data. Based on our limited experiments, Hadoop based energy data analysis is three times faster than that of a relational database management system based approach with the current system.

Key Words : Smart grid, Big data system, Power data analysis, Hadoop, Energy efficiency

1. 서론

스마트그리드란 에너지 생산 및 분배 시스템에 정보통신 기술을 접목하여 전력 수요와 공급의 양방향 통신이 가능하도록 하여 새로운 부가가치를 창출하는 차세대 지능형 전력망 인프라로 핵심 기술은 지능형 전력량계 시스템 (AMI: Advanced Measuring Instrument)의 양방향 통신망을 이용하여 전기 등의 에너지 사용에 대한 검침과 사용정보를 수집하고 관련 정보를 안내하며 다양한 요금제 적용 등의 서비스를 제공하는 것이다[1-3]. 기존의 스마트그리드 시스템은 지능형 전력량계를 이용하여 수집된 데이터를 관계형 데이터베이스로 저장하여 분석하였으나, 스마트그리드 보급 확산과 함께 데이터 량이 기하학적으로 증가함에 따라서 관계형 데이터베이스로 저장하고 분석하는 것이 어렵게 되었다[4]. 이러한 어려움을 극복하고자 본 논문에서는 하둡 기반의 빅데이터 시스템을 이용하여 전력데이터를 저장 및 분석하는 방안을 제안하고 이를 이용하여 스마트그리드 보급사업으로 수집된 전력 데이터를 분석하고자 한다.

현재 스마트그리드 확산사업으로 4개 지역단지 내의

1,880여개소의 지능형 전력량계로 부터 5분 단위로 전력 사용 데이터를 수신하고 있으며, 시범사업으로 현재까지 수집된 데이터는 대략 1억 개의 레코드로 데이터 크기가 4TB에 달한다. 따라서, 관계형 데이터베이스로 분석 연산을 수행하기에는 부담스러운 규모의 데이터이다. 추후 계획된 스마트그리드 확산사업에 따라 6개 지역으로 확장할 경우, 현재의 추세라면 데이터 크기가 28TB로 증가하고, 12개 지역으로 확장될 경우 데이터는 57TB가 훨씬 넘어서 관계형 데이터베이스로 저장하거나 분석하는 것이 현실적으로 어려워진다. 결과적으로, 하둡과 같은 빅데이터 저장 방법과 맵리듀스를 이용한 병렬 데이터 분석 시스템이 전력데이터를 저장하고 분석하는데 필요하다[4, 5]. 본 논문은 이렇게 수집된 에너지 사용 데이터를 하둡을 이용하여 분석하는 방안을 제시하고 그 분석 결과를 제시하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 스마트그리드에서 지능형 전력량계를 이용하여 에너지 사용 데이터를 수집하고 저장하여 분석하기 위한 빅데이터 시스템에 관하여 기술하고 3장에 이와 같이 수집되어 하둡시스템에 저장된 전력데이터를 하둡의 기능을 이용하여 분석하는 방법과 이를 스마트그리드 보급 사업으로 수집된 전력 데이터에 적용하여 분석한 결과를 기술한다. 결론 및 향후 계획은 4장에 이어진다.

2. 전력데이터 저장 및 분석을 위한 빅데이터 시스템

그림 1은 스마트그리드 전력데이터를 저장 및 분석하기 위한 서버 시스템의 구성을 보여준다. 현재 구축된 시스템은 빌딩 및 공장으로부터 지능형 전력량계 시스템을 이용하여 전력데이터를 수집하는 기능을 담당하는 수집 서버그

† Corresponding Author : Department of Statistics and Computer Science, Kunsan National University, Korea
E-mail : bwon@kunsan.ac.kr

* Department of Information and Computer Engineering, Ajou University, Korea

** Department of Computer Science, Kyonggi University, Korea

*** Sorrell College of Business, Troy University, USA

§ Smart Grid Research Center, Advanced Institutes of Convergence Technology, Korea

접수일자 : 2015년 2월 2일

수정일자 : 2015년 4월 14일

최종완료 : 2015년 5월 19일

롭과 이렇게 수집된 전력 데이터를 자동으로 빅데이터 시스템으로 저장 및 분석하기 위한 하둡 기반의 빅데이터 서버 그룹 그리고 외부 사용자에게 전력데이터 및 관련 서비스를 제공하기 위한 웹서비스 서버그룹으로 구성되어 있다. 본 논문에서는 데이터 수집 서버그룹과 빅데이터 서버그룹에 관하여 기술한다.

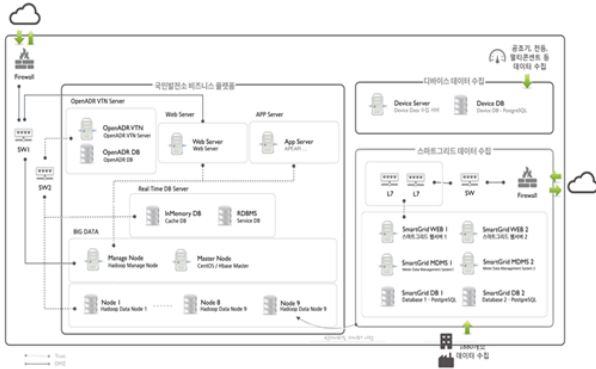


그림 1 전력데이터를 빅데이터로 저장하는 시스템
Fig. 1 Big data system architecture

2.1 데이터 수집 서버 그룹



그림 2 전력 사용 대쉬보드
Fig. 2 Energy usage dashboard

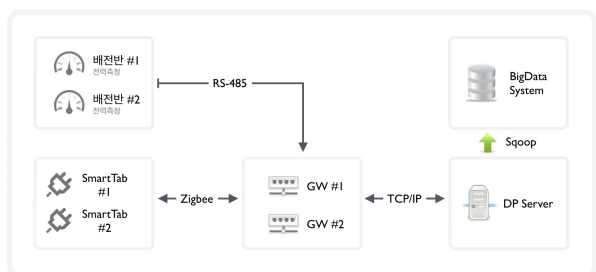


그림 3 사용전력 데이터 수집서버
Fig. 3 Energy usage data collection system

전력 데이터를 빌딩 및 공장으로부터 수집하고 관리하는 수집 서버그룹에는 전력데이터 및 수요 자원에 대한 웹서비

스 제공을 위한 웹 서버와 수집된 전력 및 수요자원 데이터 관리를 위한 데이터베이스 서버, 스마트미터 기기 및 수요자원 관리를 위한 수요자원 관리 서버로 구성되어 있다. 수집된 자료 및 시스템 보안을 위하여 수집 서버에는 자체 방화벽이 설치되어 외부 망으로부터의 접근을 제어하고, 서버의 가용성을 높이기 위하여 모든 서버는 이중화하였다. 또한, 부하를 분산하기 위하여 로드 밸런싱 스위치를 사용하여 특정 서버에 부하가 치중되는 것을 방지하였다.

공장 및 빌딩에서 스마트 미터를 이용하여 수집한 전력 사용데이터를 인터넷을 통하여 수집 서버에 보내면 수집 서버 그룹의 데이터베이스 서버가 관계형 데이터베이스인 Postgres를 이용하여 전력 사용데이터를 시간대별로 저장한다. 이렇게 저장된 데이터를 이용하여 추후 기준사용량(CBL: Customer Baseline Load)과 현재 전력 사용량을 그림 2에서처럼 사용자에게 대시보드로 전시한다. 시간대별로 저장한 관계형 데이터베이스의 경우, 개소 및 단지 단위로 기준사용량(CBL)을 계산하면 연산 속도가 느려서 온라인으로 사용자에게 서비스하기에는 적절하지 않다. 따라서 시간대별로 순차적으로 저장하는 관계형 데이터베이스 시스템보다는 개체별로 데이터를 저장하는 하둡의 HBASE와 같은 컬럼 중심의 비정형 데이터베이스 시스템이 사용자 서비스에 적절하다. 따라서, 데이터 분석을 위하여 관계형 데이터베이스로 저장된 전력 데이터를 하둡 기반의 저장장치로 이동하는 것이 필요하다. 자세한 방안은 다음 절에서 자세히 기술한다.

그림 3은 특정 공장 및 빌딩내부에서 전력사용데이터가 수집되어 빅데이터 시스템으로 저장되는 과정을 보여준다. 전력사용 데이터는 서브미터링 기기의 시리얼 통신(RS-485) 또는 스마트 콘센트의 지그비(Zigbee) 통신을 통하여 게이트웨이에 수집되고, 이렇게 게이트웨이에 수집된 전력 사용데이터는 랜선을 통하여 데이터(Data Portal) 서버에 저장된다. 이렇게 로컬 데이터 서버에 수집된 전력데이터는 인터넷을 통하여 수집서버로 전송된다[12]. 현재 스마트그리드 보급 사업으로 1880여 개소에서 전력 데이터 자료를 수집하여 Postgres 관계형데이터 베이스에 적재하고 이를 하둡의 스쿰과 플룸 기능을 이용하여 주기적으로 HBASE로 적재하여 데이터 분석에 사용한다.

2.2 전력데이터 분석을 위한 빅데이터 서버 그룹

수집된 수요 전력 데이터는 분석을 위하여 그림 4에서처럼 빅데이터 시스템으로 전송되어 저장되는데, 빅데이터 시스템은 서버 그룹의 호스트와 네트워크를 관리하기 위한 매니저 노드, 하둡 기능 제공을 위한 마스터 노드, 데이터 저장 및 분석을 위한 10개의 데이터 노드로 이루어진다. 보안을 강화하기 위하여 매니저 노드와 마스터 노드만 외부 연결 스위치에 접속되어 서비스를 제공하고 데이터 노드는 내부 스위치에 연결되어 외부에서의 접속은 불가능하게 구성하였다. 각 노드는 CentOS 기반으로 빅데이터 서비스를 위한 하둡 에코시스템이 설치되어 있다. 마스터 노드는 하둡의 네임노드와 HBASE의 마스터로 동작하도록 설정하였고, 수요 전력 데이터는 데이터 노드에 중복 저장되어 동시에 접근하여 맵리듀스를 이용한 병렬처리가 가능하고 가용성을 증대

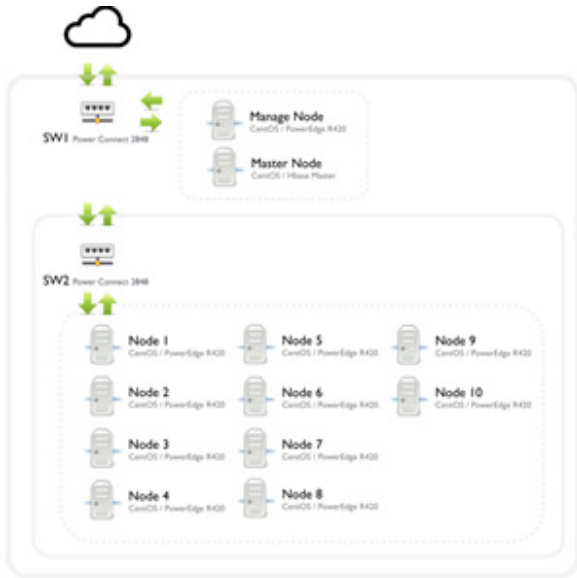


그림 4 빅데이터 시스템 구성
Fig. 4 Big Data System Group

하였다. 데이터 노드 1부터 데이터 노드 10까지는 하둡의 데이터 노드는 HBASE의 지역서버로 설정하여 데이터가 분산 저장되고 병렬처리를 지원하여 데이터를 병렬로 분석하는 공간으로 사용하였다. 특히, 노드 10의 경우는 스탠바이 네임노드로 설정하여 마스터 네임노드에 오류가 발생하였을 경우에 자동적으로 네임노드의 역할을 수행할 수 있도록 구성하여 서버의 가용성을 증가시켰다[6, 13].

2.3 하둡에코 시스템 구성

빅데이터 서버시스템의 노드에는 데이터 병렬 분석을 위하여 그림 5와 같은 하둡 에코 시스템을 설치하였는데, 스쿱(Scoop)과 플룸(Flume)은 데이터를 수집하고 저장하는 기능으로 수집 서버에서 전력 사용 데이터를 수집하여 빅데이터 시스템의 HBASE로 탑재할 때 사용한다. 데이터베이스로 정형화된 전력사용 데이터는 스쿱을 이용하여 일괄적으로 HBASE로 변환하여 저장하고 센서 데이터 등의 비정형 파일은 플룸을 사용하여 주기적으로 HBASE에 저장하도록 구성

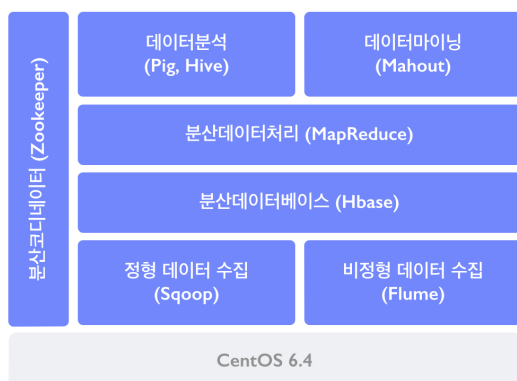


그림 5 하둡 에코 시스템
Fig. 5 Hadoop echo system

하였다[6, 7]. 이렇게 저장된 HBASE 형태의 데이터는 피그(Pig)와 하이프(Hive)등의 하둡 기능을 이용하여 별도의 프로그래밍 작업이 없이 손쉽게 분석할 수 있도록 하였다[8, 9]. 이 경우 내용량의 전력사용 데이터를 처리하기 위하여 하둡의 맵리듀스 분산처리 방식을 이용하고, 단순 통계 분석이 아닌 데이터마이닝과 기계학습 등 고급 분석이 요구되는 경우에 마헛(Mahout)에서 제공하는 알고리즘을 활용하였다[6].

표 1 전력 데이터 특성

Table 1 Data Characteristics

지역	4개 지역 (산업유통단지/테크노파크)
AMI 갯수	1,880
레코드 숫자	3억개 이상의 순시전력데이터 1억개 이상의 수요전력데이터
데이터 수집 기간	1년 6개월 (2012.12 ~ 2014.6)
데이터 갱신	5분 (순시전력 XML 포맷) 15분 (수요전력)

3. 전력데이터 분석

스마트그리드 보급 사업으로 공장 및 빌딩에서 수집한 전력 사용 데이터는 표 1에 요약되어 있는 것처럼 4개 지역에서 (산업유통단지 3곳, 테크노 파크 1곳) 1880개의 AMI를 이용하여 18개월간 수집한 데이터로 15분 단위의 1억개 이상의 전력사용 레코드가 관계형 데이터베이스로 적재되어 있다. 데이터는 5분마다 한번 씩 수집되고 있으며 Postgres 관계형 데이터베이스 테이블에 타임스탬프, AMI 위치, 전력사용량 등의 정보와 같이 XML 포맷으로 저장되어 있다. 이 중에 수요 전력 데이터는 15분 단위로 계산되어 별도의 테이블에 저장된다.

3.1 전력데이터 특성

그림 6은 2014년 1월 한달 간의 전력 사용 데이터를 보여 준다. 공장과 빌딩의 근무일인 월요일부터 금요일까지는 전력 사용량이 피크를 이루고 휴무일인 토요일과 일요일에는 전력 사용량이 급격히 줄어드는 반복적인 패턴을 보여주고 있다. 따라서, 수요전력 예측 모델 작성 시에는 근무일과 휴무일을 분리하여 두 개의 모델을 작성하는 것이 정확한 예

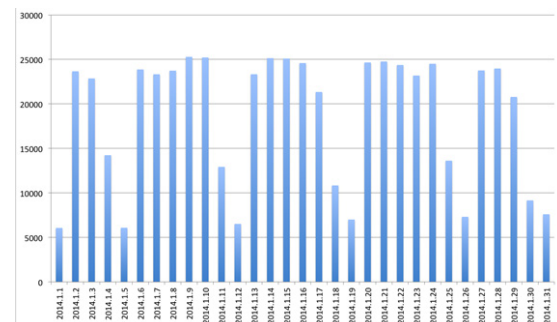


그림 6 월간 전력 사용 데이터
Fig. 6 Energy usage data for one month

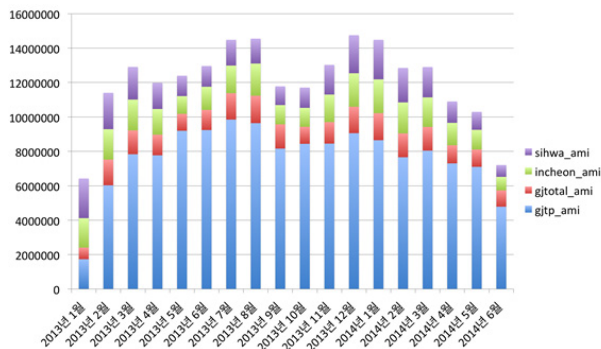


그림 7 월별 전력사용
Fig. 7 Monthly power usage

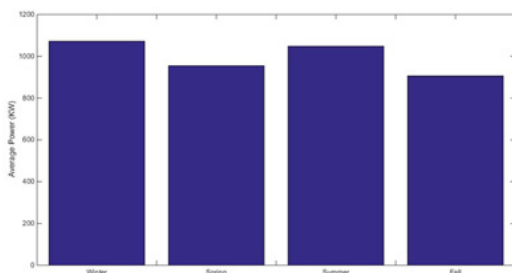


그림 8 계절별 평균 전력 사용량 (봄: 3~5, 여름: 6~8, 가을: 9~11, 겨울: 12~2)
Fig. 8 Average power usage by season

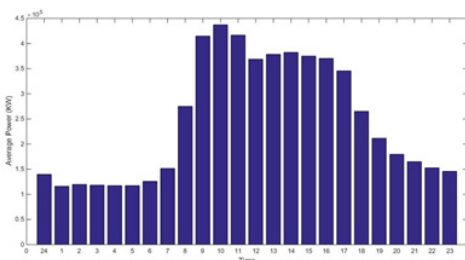


그림 9 시간대별 평균 전력 사용량
Fig. 9 Average power usage by time

측에 유리하다. 수요전력 예측 모델에 관한 세부 사항은 다음 장에 자세히 기술한다.

월별 전력사용량은 그림 7에서 보여주듯이 겨울철인 12월과 1월, 여름철인 7월과 8월에는 다른 시기와 비교하여 다소 높은 전력 사용량을 보인다. 이는 외부 온도에 의한 냉난방 운용의 결과이다. 동일한 맥락으로 그림 8에서 계절별 사용량을 보면, 여름과 겨울철에는 냉난방을 위해서 봄과 가을보다 높은 전력 사용량을 보인다. 따라서, 보다 정확한 수요전력 예측을 위하여 계절 및 외부 온도 데이터를 고려하여야 한다[10, 11]. 시간대별로는 그림 9에서 보여주듯이 출근 시간인 9시 부터 전력 사용량이 서서히 증가하면서 12시경에 전력 사용량이 피크를 이루고 오후에 들어서면서 전력 사용량은 서서히 감소하여 퇴근 시간인 6시 이후에 전력 사용량이 급격히 감소하는 반복적인 패턴을 보인다.

그림 10은 산업별 전력 사용량을 보여주고 있다. 산업 유

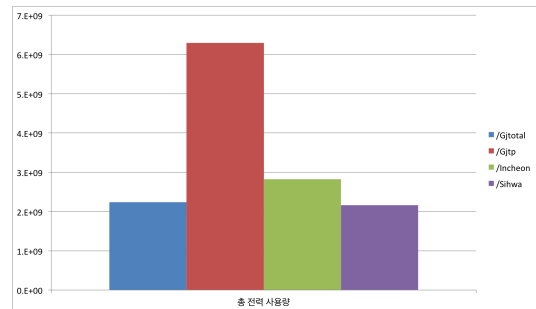


그림 10 산업별 전력 사용량
Fig. 10 Power usage by industry

통 단지 지역들은 서로 비슷한 전력사용량을 보여주고 있으나 생산 시설인 산업용 로봇이 즐비한 테크노 공장단지에서는 산업 유통 단지보다 상대적으로 많은 에너지를 사용하고 있다. 그림 11은 산업별 일단위 전력사용 특징을 보여주는 그래프로, 산업유통단지인 상단 세 개 그래프는 주말에 전력 사용량이 현저히 줄어드는 특징을 보여주는 반면에 산업용 로봇이 많은 테크노파크인 네 번째 그래프는 주말에도 비슷한 전력사용량을 보여준다. 이러한 산업별 특이 패턴은 에너지 효율화를 위한 계획 수립에 유용하게 사용될 것이다.

3.2 관계형 데이터베이스와 하둡의 성능 비교

AMI별 전력 데이터는 5분마다 XML 형태로 저장되고 15분마다 수요 전력 데이터를 XML 데이터로부터 추출하여 표 2와 같은 테이블 스키마로 저장된다. 표 2에 보이는 것처럼 수요전력 데이터는 타임스탬프, 위치정보, 전력관련 데이터가 함께 저장된다. AMI위치는 게이트웨이명파 URL로 식별되고, 해당 시간의 수요전력과 요금에 함께 저장된다. 데이터가 누적되면서 대용량으로 기존의 관계형 데이터베이스에서 SQL(Structured Query Language)을 이용하여 분석하기에는 어려움이 있다. 표 3은 관계형 데이터베이스로 저장한 데이터를 SQL로 처리하는 경우와 HBASE로 적재한 후 하둡의 하이브(Hive)를 이용한 맵리듀스 병렬처리로 연산하는 경우의 기본적인 연산 수행 시 성능 비교를 보여준다. 표 3에서 보여주는 것처럼 관계형 데이터베이스 시스템에서 SQL을 사용하는 경우 대부분의 연산이 하둡을 사용했을 경우보다 3~5배 이상 소요되기 때문에 사용자에게 온라인으로 서비스하기에는 적절하지 않다. 하지만 하둡의 맵리듀스 기능을 이용한 하이브의 경우에는 확장성이 우수하여 데이터의 크기가 증가하여도 맵퍼와 리듀서를 최적화하여 안정적인 속도로 서비스하는 것이 가능함을 알 수 있다.

관계형 데이터베이스의 경우에는 연산 속도가 HBASE보다 느린 원인을 그림 12에서 보여주고 있다. 관계형 데이터베이스의 경우에는 행(row)을 중심으로 데이터가 저장되기 때문에 특정한 열(Column)에 대한 연산을 수행하려면 디스크에서 메모리로 여러 블록을 읽어 들여야 한다. 예로, 그림 12의 경우에 주어진 쿼리문을 수행하기 위하여 3개 블록을 메모리로 읽어 들여야 한다. 반면 HBASE와 같은 열기반 저장인 경우에는 동일한 연산 수행 시 디스크 입출력이 줄어든다. 그림 12의 경우 동일한 쿼리문을 수행하기 위하여

2개의 블록 만 메모리로 읽어 들이면 된다. 이와 같은 이유로 열에 대한 연산 수행 시 기존의 관계형 데이터베이스보다는 열기반의 저장 방법이 유리하다. HBASE의 경우 이러한 열기반으로 데이터를 저장하여 속도를 증대시키고 있다. 또한, HBASE의 경우 동일한 데이터를 세 곳의 별도의 위치에 저장하여 병렬로 연산이 가능하고 데이터 오류시에도 복구가 가능하도록 구성한다. 이러한 특징으로 하둡의 맵리듀스를 이용하면 시스템 성능을 최적화 할 수 있는 다수의 맵퍼와 리듀서가 자동으로 생성되어 병렬로 연산 처리하여 수행속도가 훨씬 빠르게 된다.

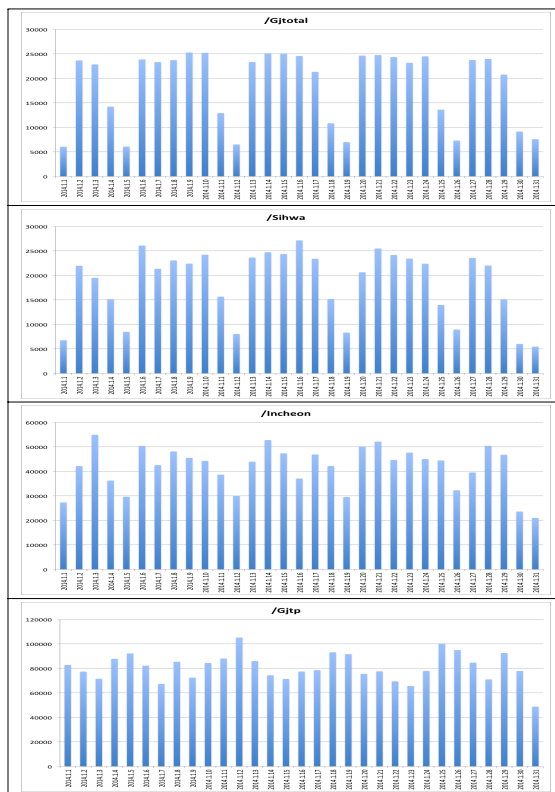


그림 11 산업별 월간 전력 사용량

Fig. 11 One month power usage by industry

표 2 수요전력 데이터 테이블

Table 2 Demand table

Field	Content	Type
mid	아이디	Serial
ttime	시간	Timestamp
sgname	게이트웨이명	Text
url	장치 URL	Text
demand	수요전력	Real
rdemand	무효전력	Real
pf	역률	Real
charge	요금	Real
whacc	누적전력사용량	Real

SELECT url, avg(demand) from demands group by url order by url;
→ url과 demand 값을 저장하고 있는 블록을 읽어서 연산 수행

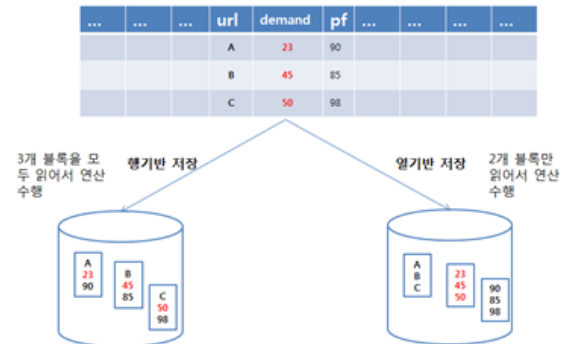


그림 12 행기반 대비 열기반 저장 공간

Fig. 12 Row based vs. Column based storage scheme

표 3 연산별 관계형 데이터베이스와 하둡 맵리듀스 성능 비교

Table 3 Performance comparison between RDBMS and Hadoop

연산	RDBMS (SQL)	Hadoop (Hive)
Count (record)	60 sec	20 sec
Count (unit)	79 sec	20 sec
Count (site)	64 sec	20 sec
Group (by unit)	77 sec	22 sec
Group (by site)	69 sec	20 sec
Group (by time)	65 sec	23 sec
Group (by hour)	93 sec	23 sec
Group (by month)	72 sec	25 sec
Group (by season)	21 sec	18 sec

3.3 맵리듀스를 이용한 데이터 분석 응용

전력관리 시스템에서 기준사용량 (CBL) 계산은 전력수요 관리 및 요금계산, 수요반응 등 부가서비스 제공을 위한 중요한 핵심 연산중에 하나이다. 기준 사용량 계산 방법에 관한 많은 연구가 이루어져 있고 그 계산 방법이 복잡하지만, 본 논문에서는 기준사용량 계산을 단순화하여 동일 시간대의 10일간 평균 전력 사용 값으로 정의하였다. 즉, 동일한 타임태그를 갖는 레코드의 10일간 전력사용량 데이터를 평균한 값을 기준사용량으로 정의한다. 이러한 CBL 계산은 컬럼 기반의 대표적인 연산으로 앞에서 언급했듯이 관계형 데이터베이스로 계산하는 것은 비효율적이다. 실제로 기준사용량을 계산하기 위한 관계형 데이터베이스와 하둡의 하이브 간에 연산 속도를 비교하면 표 4에서 보이는 것처럼 관계형 데이터베이스로 저장된 데이터를 이용하여 기준사용량을 계산하는 연산은 하둡의 하이브를 사용한 경우보다 3 배 정도 느리다.

에너지의 원활한 수요 공급을 위하여 수요전력에 관한 정

표 4 기준사용량 계산 연산

Table 4 CBL computation cost

기준사용량 연산	RDBMS (SQL)	Hadoop (Hive)
유닛 (단일 AMI)	60 sec	20 sec
단지 (동일 sname)	79 sec	20 sec
전체	64 sec	20 sec

확한 예측이 필요하다. 이를 위하여 하둡의 에코시스템 중 마하트(mahout)에 포함된 선형회귀분석을 이용하였다. 근무일과 휴무일에 전력사용 패턴이 현저하게 다르므로 하나의 회귀모델을 사용하여 수요 전력을 예측하기 보다는, 근무일(weekday)과 휴무일(weekend)을 분리하여 두 가지 다른 선형회귀모델로 예측하였다. 그림 13에 2014년 1월의 근무일 회귀모델에 의한 예측값과 실제값과 휴무일 모델에 의한 값을 보여주고 있다. 근무일(weekday)인 경우에는 예측 패턴이 실제 전력사용 값과 비교적 정확하게 매치되고 있다. 하지만, 상대적으로 휴무일(weekend)인 경우에는 피크전력 예측에서 근무일보다 오차가 크게 발생한다. 이러한 현상은 휴무일(weekend)인 경우 회귀분석 값이 트레이닝 데이터에 지나치게 의존하는 현상이 발생하고 다양한 휴무일의 특성으로 (i.e., 주말 또는 주초에 휴일이 발생하여 주말이 길어지는 경우, 주중에 휴일이 발생하는 경우 등) 장기간의 다양한 데이터가 필요하나 현재의 데이터 규모로는 휴무일에 수요전력을 정확하게 예측하기에는 부족하다. 좀 더 정교한 회귀분석 모델을 사용하거나 오랜 기간의 데이터를 수집하면 이러한 현상은 해소될 것으로 예측된다.

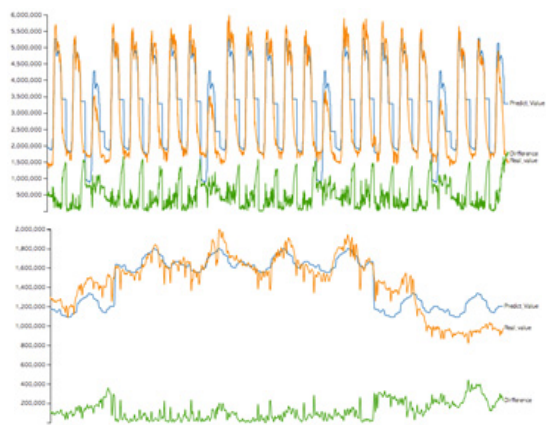


그림 13 한달간 (2014년 1월) 수요전력 예측: 근무일(상단) 휴무일 (하단); 실측값(붉은색), 예측값(파란색), 에러(녹색)

Fig. 13 Demand predict for a month (2014. 1): weekday(top) and weekend(bottom); real value(red), predicted value(blue), error(green)

4. 결 론

본 논문에서는 하둡 기반의 빅데이터 시스템을 이용한 전력 사용 데이터를 저장 및 분석 하는 방안을 제시하였으며,

스마트그리드 확산 사업으로 수집된 전력 사용 데이터를 활용하여 전력 데이터의 특성을 분석하였다. 분석에 따르면 전력 사용 데이터는 대용량으로 관계형 데이터베이스를 이용하여 분석하는 것 보다 하둡을 이용한 빅데이터 시스템으로 구축한 후 병렬처리를 지원하는 맵리듀스를 이용한 분석 방안이 3~5배 정도 빠르다. 스마트그리드 확산사업으로 데이터의 용량이 기하학적으로 증가할 것을 고려하면 관계형 데이터베이스를 이용한 데이터 분석이 어려울 것으로 예측된다.

현재의 시스템은 그래픽 사용자 인터페이스가 제공되지 않아서 일반 사용자가 전력 데이터를 분석하거나 예측하기 위하여 명령어 기반의 도구를 사용하여야 한다. 향후, 웹 기반의 사용자 인터페이스를 제공하여 일반 사용자가 손쉽게 전력 사용량을 예측할 수 있는 서비스를 제공할 계획이다. 또한, 최근에 인메모리 연산을 이용하여 연산 성능을 향상시키기 위한 시스템이 개발되어 하둡 빅데이터 시스템 환경에서 사용되고 있다. 향후, 이와 같은 인메모리 기능을 활용하여 분석 성능을 향상 시킬 계획이다. 마지막으로 오프라인으로 분석하는 기능 이외에 온라인상에서 스트리밍으로 저장되는 전력 사용 데이터를 처리 및 분석하기 위한 시스템으로 확장할 계획이다.

감사의 글

본 연구는 2013년도 산업통산자원부의 재원으로 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. 20132010101800)

References

- [1] Smartgrid website, <http://www.smartgrid.or.kr>, Korea smartgrid association, 2013.
- [2] Demand Reductions from the Application of Advanced Metering Infrastructure, Pricing Programs, and Customer-Based Systems - Initial Results, SmartGrid.gov, December, 2012.
- [3] Demand side management using ICT and technologies, Ministry of Trade, Industry and Energy, August, 2013.
- [4] Big data understanding and applications, B. On, KETEP, December, 2012.
- [5] Ultra Large-Scale Power System Control Architecture, SmartGrid.gov, October 2012.
- [6] Hadoop The Definitive Guide, 3rd Ed. W. Tom, O'Reilly Media, July 2009.
- [7] Hbase: The Definitive Guide, G. Lars, O'Reilly, August 2011.
- [8] Programming Hive, C. Edward, O'Reilly Media, November 2012.
- [9] MapReduce: Simplified Data Processing on Large Clusters, J. Dean and S. Ghemawat, Communications of the ACM, Vol. 51, No. 1, pp. 107-113, 2008.
- [10] A Big Data Management System for Energy

Consumption Prediction Models, W. Lee, B. On, I. Lee and J. Choi, Proceedings of International Conference on Digital Information Management (ICDIM), Bangkok, Thailand, pp. 156-161, October, 2014.

- [11] Simple one-pass algorithm for penalized linear regression with cross-validation on MapReduce, e-Print arXiv:1307.0048 (2013)
- [12] Development of Sensor Based Energy Management System, D. Um, J. Choi and I. Lee, Journal of the Korean Institute of Illuminating and Electrical Installation Engineers, Vol. 28, No. 10, pp. 69-74, 2014.
- [13] A Study on Demand-Side Resource Management Based on Big Data System, J. Yoon, I. Lee and J. Choi, The Transactions of the Korean Institute of Electrical Engineers, Vol. 63, No. 8. pp. 1116:1120, 2014.



온 병 원 (溫 炳 原)

2007년 미국 The Pennsylvania State U. 컴퓨터공학과 졸업(박사). 2008년~2009년 캐나다 U. of British Columbia 박사후 연구원. 2010년~2011년 U. of Illinois at Urbana-Champaign ADSC연구소 선임연구원. 2011년~2014년 차세대융합기술연구원 공공데이터연구센터 센터장. 2014년~현재 군산대학교 통계컴퓨터과학과 조교수.
E-mail : bwon@kunsan.ac.kr.



최 중 인 (崔 重 仁)

1987년 MIT 제어계측 졸업(박사). 1987년~2014년 가천대학교 에너지IT학과 교수. 2013년~현재 차세대융합기술연구원 스마트그리드연구센터 센터장.
E-mail : jichoi@snu.ac.kr

저 자 소 개



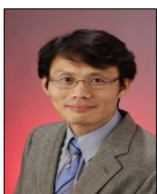
조 영 탁 (曹 永 卓)

2014~현재 아주대학교 정보컴퓨터공학과 학사과정 재학 중.
E-mail : nintyning@gmail.com



이 원 진 (李 原 振)

2014년 경기대학교 컴퓨터과학과 졸업.
2013~2014년 차세대융합기술연구원 공공데이터연구센터 연구보조원.
E-mail : ckask12@kyonggi.ac.kr



이 인 규 (李 仁 奎)

2007년 미국 The Pennsylvania State U. 컴퓨터공학과 졸업(박사). 2007년~2013년 미국 Troy U. 경영대학 조교수. 2013년~2014년 차세대융합기술연구원 스마트그리드연구센터 책임연구원. 2015년~현재 미국 Troy U. 경영대학 부교수.
E-mail : inlee@snu.ac.kr