

# 빅데이터 기반의 수요자원 관리 시스템 개발에 관한 연구

## A Study on Demand-Side Resource Management Based on Big Data System

윤재원\* · 이인규<sup>†</sup> · 최중인\*  
 (Jae-weon Yoon · Ingyu Lee · Jung-In Choi)

**Abstract** - With the increasing interest of a demand side management using a Smart Grid infrastructure, the demand resources and energy usage data management becomes an important factor in energy industry. In addition, with the help of Advanced Measuring Infrastructure(AMI), energy usage data becomes a Big Data System. Therefore, it becomes difficult to store and manage the demand resources big data using a traditional relational database management system. Furthermore, not many researches have been done to analyze the big energy data collected using AMI. In this paper, we are proposing a Hadoop based Big Data system to manage the demand resources energy data and we will also show how the demand side management systems can be used to improve energy efficiency.

**Key Words** : Big data system, Demand response, Hadoop, Distributed database

### 1. 서론

에너지 패러다임이 공급위주의 정책에서 ICT를 이용한 정보 교환으로 필요한 시기에만 수요를 조정하는 수요관리 중심으로 변화되고 있다. 이러한 수요관리중심 체계는 수요자원 관리가 중요한 요인이다. 게다가 수요자원은 지속가능성 및 환경 규제에 대응하기 위한 친환경 에너지로 분류되어 더욱 더 중요성이 부각되고 있다[1,2,3]. 그러므로 수요자원 관리를 위한 다양한 에너지 관리 시스템이 개발되었다. 하지만, 수요자원 관련 데이터는 천문학적인 데이터 사이즈로 기존의 관계형 데이터베이스로 저장하고 분석하기에는 어려움이 있다[7,8,16]. 본 논문에서는 이러한 대용량의 전력정보 데이터를 기반으로 하는 수요자원을 확장성을 갖춘 하둠 기반의 빅데이터 시스템을 이용하여 관리하기 위한 방안을 제안하고자 한다.

그림 1은 스마트그리드 사업단에서 수집하고 있는 전력 데이터로 4개 지역 단지 내의 1,880개소로부터 수신하고 있는 전력데이터의 대략적인 규모를 보여주고 있다[2]. 각각의 서브미터링 기기에서 5분 단위로 데이터를 수신하여 연간 2억 개의 데이터가 저장된다. 한번에 2KB를 수신할 경우, 대략 400GB의 데이터가 수집되고 있다. 향후 추진될 IoT를 고려하면 한 곳에 150여개의 센서 노드가 설치된다고 가정하여 300억 개의 데이터가 저장되어야 하고, 한걸음 나아가 스마트그리드 보급사업과 확산사업을 고려하여 국가 전체로 보면, 연간 약 4000조 개의 데이터가 수신된다. IoT 적용 시에는 60경개의 데이터가 1년 동안 수신되고 600경개의 데이

터가 10년 동안에 수신되어 일반 데이터베이스로는 처리하기 곤란하게 된다. 이러한 데이터를 저장하고 분석하기 위한 수요자원 관리시스템을 위하여 본 논문에서는 하둠을 이용한 빅데이터 시스템을 제안한다.

본 논문은 다음과 같이 구성된다. 2장에서는 수요자원 관리를 위한 하둠 에코시스템 기반의 빅데이터 시스템에 관하여 기술하고, 3장에서는 이러한 빅데이터 시스템에 저장된 전력 데이터를 활용한 수요관리 서비스 및 플랫폼에 관하여 기술한다. 결론 및 향후 계획은 4장에 이어진다.

### 2. 빅데이터 시스템 설계

이 장에서는 수요자원 관리를 위하여 수용가별 전력 사용 데이터를 저장하고 분석하기 위한 관계형 데이터베이스 기반의 시스템을 소개하고 이의 문제점을 해결하기 위한 빅데이터 시스템과 전력 데이터를 수집하여 빅데이터로 구축하기 위한 과정을 자세하게 기술한다.



그림 1 스마트그리드에서 수집하고 전력데이터  
 Fig. 1 Energy usage data

<sup>†</sup> Corresponding Author : Advanced Institute of Convergence Technology, Seoul National University, Korea  
 E-mail : inlee@snu.ac.kr

\* Advanced Institute of Convergence Technology, Seoul National University, Korea

Received : June 5, 2014; Accepted : July 23, 2014

### 2.1 데이터베이스 기반 수요관리 시스템

그림 2는 스마트그리드 사업단에서 전력 데이터를 수집 관리하는 시스템의 구성을 보여준다. 외부 망으로부터의 접근을 제어하는 방화벽(Firewall) 설치되어 있고 웹서버와 데이터베이스 서버, 스마트 미터(AMI)를 관리하는 서버(MDMS)로 구성되어 있다. 서버의 가용성을 늘리기 위하여 서버를 중복화 하였고 로드 밸런싱 스위치를 사용하여 부하를 분기하였다.

스마트 미터가 설치된 공장 및 빌딩 1880개소에서 5분 단위로 보내는 전력 사용 관련 데이터는 관계형 데이터베이스인 Postgres를 이용하여 데이터베이스에 시간대별로 저장된다. 이렇게 저장된 데이터는 사용자가 요구 시에 기준사용량(CBL: Customer Baseline Load)을 실시간으로 계산하여 화면에 전시한다. 시간대별로 저장된 데이터는 단지 및 개소별 연산인 CBL에 매우 불합리적이다. 따라서, 시간대별로 순차적으로 저장하는 시스템보다는 개체별 저장을 유도하는 비정형데이터 베이스 시스템이 절실히 요구된다[12,13].

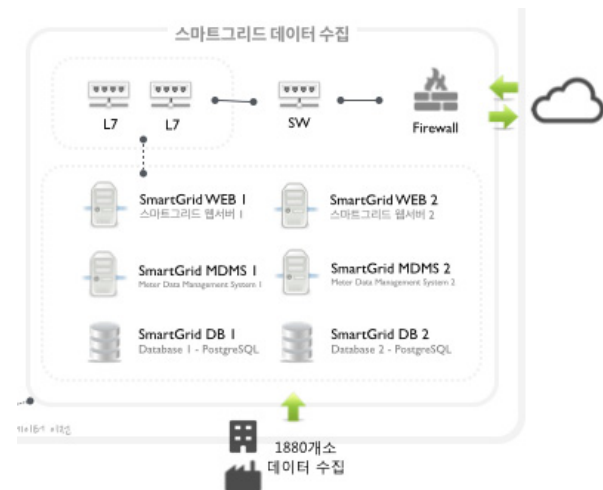


그림 2 전력데이터 수집 시스템  
Fig. 2 Energy usage data collection system

### 2.2 빅데이터 시스템

앞에서 언급한 것처럼 수요자원을 관리하기 위한 대용량의 전력 및 비전력 데이터는 천문학적인 데이터 크기와 데이터의 빠른 증가세로 인하여 기존의 관계형 데이터베이스로 저장하여 분석하기에는 어려움이 있다. 또한, IoT를 활용한 수요자원 관리 시스템에서 센서를 통해 수집되는 비정형 데이터의 증가 추이를 고려하면 시스템 확장이 용이한 하둠을 이용한 빅데이터 시스템이 적절하다.

그림 3은 수요자원 관리를 위하여 구축한 대용량의 전력 데이터를 저장하기 위한 빅데이터 시스템의 하드웨어 구성을 보여준다. 두 대의 Switch HUB가 각각 내부와 외부의 네트워크를 연결하는 역할을 한다. 외부 연결을 위한 스위치는 매니저 노드와 네임노드만 연결이 가능하게 구성하고 나머지 데이터 노드는 내부 연결을 위한 스위치에 연결하여 보안을 강화한다. 매니저 노드는 내부 네트워크에 연결된 노드

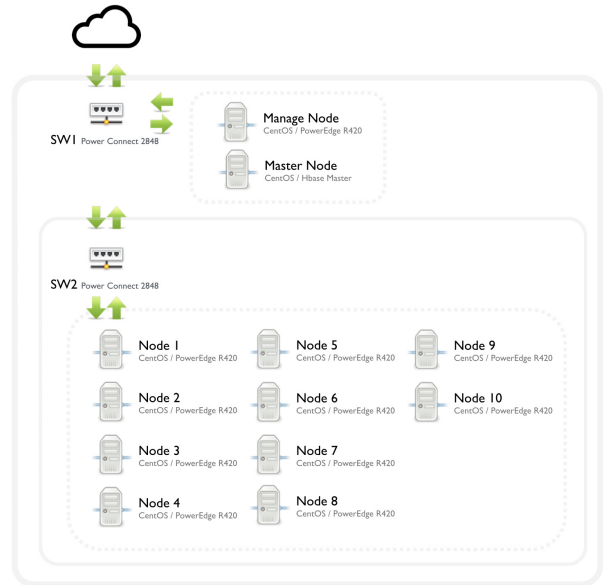


그림 3 수요관리 빅데이터 시스템  
Fig. 3 Demand resources management system

들의 시스템 관리를 담당하고 빅데이터 처리 시스템인 하둠을 관리하는 마스터 노드와 열 개의 데이터노드가 대상이다. 또한, 이를 모니터링하고 컨트롤하기 위한 모니터링 콘솔이 서버 랙에 설치되어 있다. 각 노드에는 CentOS (Release 6.4)와 하둠 (Ver 2.2.0) 에코시스템이 그림 4와 같이 설치되어 있다. 각 노드는 인텔 제온 프로세서(E5-2430 CPU)에 8TB의 하드디스크를 가지고 있다. 매니저 노드는 16GB의 메인메모리가 장착되어 있으며, 부하가 큰 마스터 노드는 32GB의 메인 메모리를 가지고 있으며, 그 외의 데이터 노드는 24GB RAM 메모리로 구성되어 있다. 자세한 노드별 소프트웨어 제원은 테이블1에 자세하게 기술되어 있다.

마스터 노드는 하둠의 네임노드와 HBase의 HBase 마스터로 동작하도록 설정하여 빅데이터로 저장된 수요자원에 접근하여 맵리듀스로 분석 가능하도록 해준다. 노드 1부터 노드 10은 하둠의 데이터노드와 HBase의 지역서버(Region Server)로 설정하여 실제로 데이터가 저장되며 병렬로 데이터를 분석되는 공간으로 사용된다. 노드 10의 경우는 스탠바이 네임노드(Standby NameNode)로 설정하여 주 네임노드에 오류가 발생하였을 때 자동적으로 네임노드의 역할을

표 1 노드별 소프트웨어 설치내역

Table 1 Software specification

장비	설치내역	비고
Master Node	- CentOS 6.4 - Hadoop 2.2.0 - HBase 0.98.0 - Sqoop 1.4.4	- Hadoop의 Name Node 와 HBase의 Hbase Master
Name Node	- CentOS 6.4 - Hadoop 2.2.0 - HBase 0.98.0 - Sqoop 1.4.4	
Data Node 1 ~ 9	- CentOS 6.4 - Hadoop 2.2.0 - HBase 0.98.0 - Sqoop 1.4.4	- Hadoop의 Data Node 와 HBase의 Region Server
Data Node 10	- CentOS 6.4 - Hadoop 2.2.0 - HBase 0.98.0 - Sqoop 1.4.4 - Grails 2.2.2 - MySQL-Server 5.1.73 - NginX 1.6.0	- Standby Name Node - OpenADR2.0 VTN Server

수행할 수 있도록 구성하였다. 추가적으로 노드 10에는 EnerNOC의 VTN(Virtual Top Node) 서버 구성을 위하여 Grails 프레임워크에 MySQL 데이터베이스 시스템과 NginX 웹서버가 설치되어 있다.

네임노드가 데이터의 위치정보를 보관하고 있어 하둡 시스템에서 가장 중요한 요소 중의 하나이다. 따라서 하둡2.0은 액티브 네임노드와 스탠바이 네임노드가 동시에 존재하여 네임노드의 지속적인 서비스를 보장한다. 액티브 네임노드가 저널 노드에 로그를 기록하면, 저장된 로그데이터를 스탠바이 네임노드 (Standby NameNode) 가 접근할 수 있다. 또한 디렉토리 서비스를 위하여 액티브 네임노드 (Active NameNode) 는 각각의 데이터 노드에 데이터를 분산시켜 저장하고 저장이 완료된 데이터에 대한 메타데이터를 네임노드의 디스크에 저장한다. 하나의 데이터는 기본적으로 3개의 노드에 분산 저장되어 데이터의 안전성을 보장한다[12,13].

### 2.3 하둡에코 시스템

저장된 수요관리 데이터는 그림 4에서 보이는 하둡 시스템으로 분산처리가 가능하다[14,15,16]. 스쿱 (Squoop)과 플룸 (Flume)은 하둡에서 데이터를 수집 저장하는 기능을 제공한다. 데이터베이스로 정형화된 수요관리 데이터는 스쿱을 이용하여 일괄적으로 HBASE에 저장하고 센서 데이터등의 비정형 파일은 플룸을 사용하여 저장하게 된다. 이렇게 저장된 HBASE 형태의 데이터는 피그 (Pig)와 하이브(Hive)를 이용하여 분석할 수 있는데, 대용량의 수요관리 데이터를 처리하기 위하여는 하둡의 맵리듀스 분산처리 방식으로 접근 및 분석하여야 한다. 또한 데이터마이닝과 기계학습 등 고급 분석이 요구되는 경우에 마훗 (Mahout)에서 제공하는 기능을 활용한다.

### 2.4 전력데이터 수집 및 저장

그림 5는 전력 데이터가 수집되어 빅데이터 시스템에 저장되는 과정을 보여준다. 전력 데이터는 스마트 콘센트 또는 서브미터링의 지그비(Zigbee) 또는 시리얼 (RS-485) 통신을 통하여 게이트웨이에 수집되고, 이렇게 게이트웨이에 수집된 전력데이터는 랜선을 통하여 DP (Data Portal) 서버에 저장된다. 이렇게 수집된 전력 데이터를 추후 맵리듀스

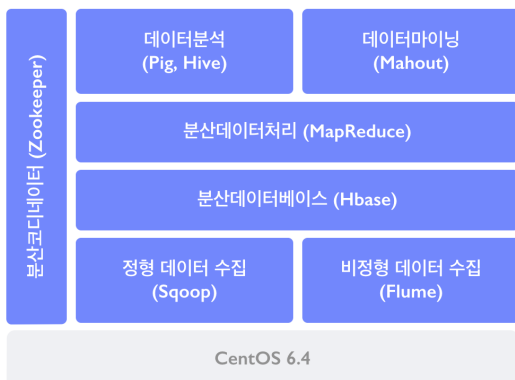


그림 4 하둡 에코 시스템  
Fig. 4 Hadoop echo system

를 이용한 빅데이터 분석을 위해 하둡시스템의 HBase로 스쿱(Squoop)을 사용하여 일괄 적재한다.

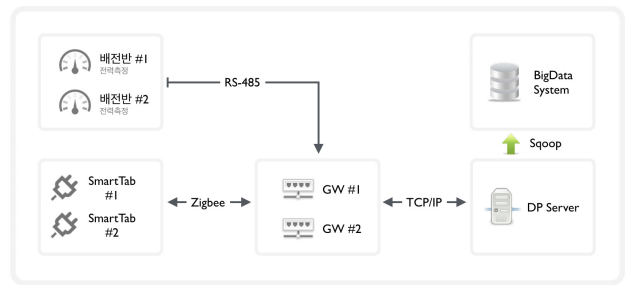


그림 5 전력데이터 수집과정  
Fig. 5 Energy usage data collecting process

## 3. 수요자원 관리 시스템 설계

이 장에서는 빅데이터 시스템을 이용한 수요자원 관리 시스템 및 서비스와 대표적인 응용프로그램인 수요반응 애플리케이션에 관하여 기술한다.

### 3.1 수요자원 관리 시스템

그림 6은 구축된 전력 정보 데이터 시스템을 이용한 수요관리 서비스 시스템 구성을 보여준다. 수요관리 서비스 시스템에는 커다랗게 세 가지 중요한 요소가 있다[4,5]. 첫째는 데이터를 수집하여 분석하고 관리하는 빅데이터 기술부분이고, 둘째는 수요 반응을 구현하는 OpenADR2.0 규격을 통한 VTN (Virtual Top Node) 서버와 VEN (Virtual End Node) 사이의 통신부분이다. 그리고 마지막으로 수집된 데이터를 제3의 업체(3RD Party)가 새로운 서비스를 개발하여 제공하기 위한 API 서비스로 볼 수 있다.

수용가로부터 실시간 저장되는 전력 데이터들은 시각화하고, 특정 수요관리 자원에 관한 세부내용을 분석하여 효율적인 수요관리를 할 수 있도록 도와준다. 또한, 통계와 기계학습 기능을 이용하여 각 장비의 전력 데이터 패턴을 분석하고 그에 따른 향후 전력소비를 예측하여 장비별 수요관리를 할 수 있도록 한다. 수용가의 전력 데이터 수집이 늘어갈수록 장비별 패턴분석의 오차범위가 현저히 줄어들어, 전반적인 수요관리 효율은 상당히 높아질 것이다.

그림 7은 현재 제작된 프로토타입 수요관리시스템의 대쉬보

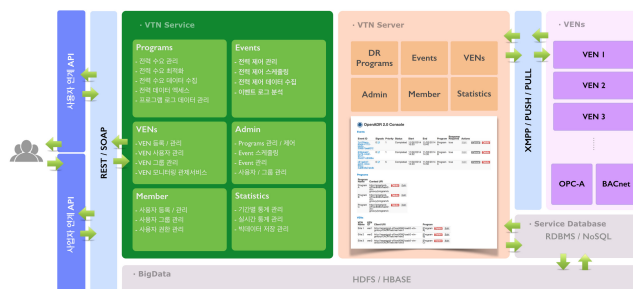


그림 6 수요자원 관리 시스템  
Fig. 6 Demand resources management system

드 화면을 보여준다. 각 수용가별 기준사용량 (CBL: Customer Baseline Load) 대비 현재 전력사용량을 비교하여 화면에 전시하고 수요반응 이벤트 발생 시에는 목표 절감량 대비 현재 생산량을 표시한다.



그림 7 수요자원 관리 시스템 대쉬보드  
Fig. 7 Demand management system dash board

### 3.2 수요반응 (DR: Demand Response) 시나리오

수요관리 시스템은 수용가 기본관리부터, 수용가내의 모든 장비들까지 등록하여 수요자원을 관리할 수 있도록 한다. 등록된 장비는 게이트웨이(Gateway)에서 취합되어, 해당 수요반응 클라이언트로 이동되고 수요반응서버로 전달되어 최종적으로 빅데이터 시스템으로 저장되어 관리된다. 수요반응서버에 등록된 장비는 상태정보를 모니터링하고 필요시 해당 장비에 직접적으로 이벤트(Event)를 발생시킬 수 있다. 건물과 해당 장비의 관리, 제어를 통해 적정수준의 수요자원을 관리 할 수 있도록 응용프로그램이 수요관리 시스템에서 제공된다.

수요반응서버는 그레일(Grails) 프레임워크에 그루비(Groovy) 언어로 구성되어 있으며, 가상머신(GVM)기반으로 실행되고 있다. 웹서버로는 NginX가 사용되고 있고, 데이터베이스는 관계형 데이터베이스인 MySQL DB로 구축되어 있다. 수요반응 클라이언트는 파이썬(Python)기반으로 구성되어, 수요반응서버를 폴링(Polling) 방식(혹은 xmpp)을 통하여, 주기적으로 호출하게 된다[4,9,11]. 수요반응서버와 수요반응클라이언트와의 이벤트 발생은 통신규격인 OpenADR2.0을 사용하게 되며, OpenADR2.0은 HTTP (HyperText Transfer Protocol) 기반의 XML (eXtensible Mark-up Language) 형식 데이터로 전송된다. 앞서 언급한대로 수요반응서버와 수요반응클라이언트간의 통신은 수요반응 클라이언트에서 수요반응서버 쪽으로 이벤트를 폴링(Polling) 방식으로 확인하게 되므로, 수요반응클라이언트에서 자신의 아이디를 담아 수요반응서버에 전달하고, 수요반응서버에서 해당 이벤트 등록 여부를 확인 후 이벤트 내용을 수요반응 클라이언트 쪽으로 전달하는 방식으로 이루어진다. 또한, 수요반응 클라이언트에서는 이벤트 데이터 이외에 추가적인 장비데이터를 수요반응서버 측으로 전송하여, 해당 수용가의

장비데이터를 수집하여 빅데이터로 저장한다.

### 3.3 시스템 서비스 API

수요관리 서비스 시스템은 수용가로부터 수집된 데이터를 제3의 업체가 해당 비즈니스에 손쉽게 활용할 수 있도록 그림 8과 같이 API를 제공한다. 보안상의 문제는 SSL 또는 TLS를 접속 시에 사용함으로 안전성을 높이고 인증된 사용자만 이용이 가능하도록 되어있다. 또한, 인증된 사용자는 자유롭게 수집된 데이터를 활용하여 연계 어플리케이션을 개발하고 배포할 수 있도록 구성되어 있다. API는 간단하고 명료하게 통신할 수 있는 RESTful (혹은 SOAP) 방식으로 구성하여 수용가에 대한 정보 및 통계, 수요관리 및 제어, 과금, 이벤트 등 수집된 모든 정보를 제3의 업체에 제공할 수 있도록 한다. 이로써 제공되는 데이터를 활용해 새로운 서비스를 창출하여 제공할 수 있도록 지원한다.

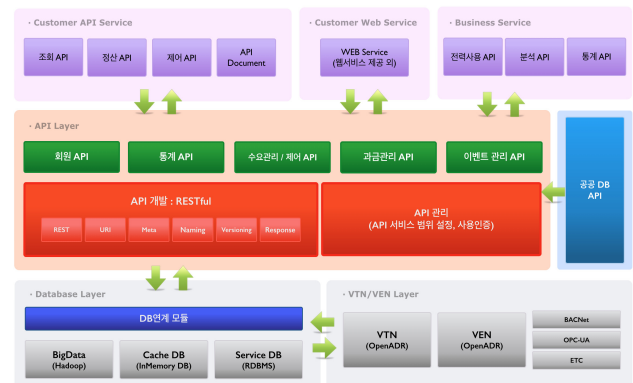


그림 8 수요자원 관리 시스템 서비스 API  
Fig. 8 Demand management API Services

## 4. 결 론

본 논문에서는 수요자원 관리를 위한 빅데이터 시스템 구성과 이를 이용한 수요자원 관리시스템의 서비스플랫폼을 제안하였다. 또한 수요관리 시스템의 어플리케이션으로 자동수요반응 시스템 구성에 대하여 기술하였으며, 제3의 개발자가 공개되는 데이터 접근 API 및 어플리케이션을 이용하여 서비스를 개발 및 제공할 수 있도록 하였다.

현재의 빅데이터 시스템은 하둡시스템의 장애 감시 및 복구는 가능하나 전체 시스템 고장시의 대책이 준비되어 있지 않다. 향후 분산 시스템 모니터링 및 복구 기능을 추가하여 시스템 가용성을 높일 예정이다. 또한 전력데이터 이외의 다양한 센서로 부터 수집되는 데이터를 저장하는 기능과 다양한 데이터 포맷을 상호호용 가능 하도록 전환하는 ETL 기능 등이 추가될 예정이다.

### 감사의 글

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

## References

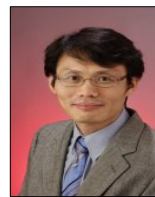
- [1] Smartgrid website, <http://www.smartgrid.or.kr>, Korean Smart Grid Institute, 2013.
- [2] 2013 Smart Grid Extension Plans, Korean Smart Grid Institute, June 2013.
- [3] Demand side management using ICT and technology, ministry of trade, industry and energy, August, 2013.
- [4] OpenADR Alliance website, <http://www.openadr.org>
- [5] EnerNOC website, <http://www.enernoc.com>
- [6] Demand Reductions from the Application of Advanced Metering Infrastructure, Pricing Programs, and Customer-Based Systems - Initial Results, SmartGrid.gov, 2013.
- [7] Ultra Large-Scale Power System Control Architecture, SmartGrid.gov, October 2012.
- [8] Demand response architecture and load management algorithms for energy-efficient power grid: a survey, Y. Law, T. Alpcan V. Lee, A. Lo, M. Palaniswami, Seventh International Conference on Knowledge, Information and Creativity Support Systems, 2012.
- [9] OpenADR Open Source Toolkit: Developing Open Source Software for the Smart Grid, Charles McParland, IEEE Power and Energy Society General Meeting, July 2011.
- [10] Training material: PJM 101, PJM 2013.
- [11] Implementing of OpenADR2.0a Profile for Demand Response in Smart Grid, KERI.
- [12] Hadoop The Definitive Guide, 3rd Ed. W. Tom, O'Reilly Media, July 2009.
- [13] Hbase: The Definitive Guide, G. Lars, September 2011.
- [14] Programming Hive, C. Edward, O'Reilly Media, November 2012.
- [15] MapReduce: Simplified Data Processing on Large Clusters, J. Dean and S. Ghemawat, Communications of the ACM, 2008.
- [16] Big data understanding and applications, B. On, KETEP, December, 2012.

## 저 자 소 개



### 윤재원 (尹 哉 元)

2008년 건국대에서 석사학위를 받은 후, 웹프로그래머로 근무하였으며, 2013년 부터는 차세대융합기술원 스마트그리드 연구센터에서 ICT를 활용한 비즈니스 플랫폼 개발 과제에 참여 중이며, 연구 관심 분야는 IoT를 활용한 에너지 효율화.



### 이인규 (李 仁 奎)

2007년 펜실베이니아 주립대에서 박사학위를 받은 후, 미국 알라바마주 트로이 주립대에서 교수로 재직하였고, 2013년 부터는 차세대융합기술원 스마트그리드 연구센터에서 비즈니스 플랫폼 및 빅데이터 마이닝에 관하여 연구를 수행 중.



### 최중인 (崔 重 仁)

1987년 메사추세츠 대학교에서 박사학위를 받은 후, 가천대학교에서 교수로 재직 한 후, 2013년부터 차세대융합기술원 스마트그리드 연구센터에 센터장으로 근무 중이며, 주요 연구 관심 분야는 스마트그리드 및 트랜잭티브 에너지.