

클라우드 기반 센서 데이터 관리 시스템 설계 및 구현

박경욱* · 김경욱** · 반경진** · 김응곤***

Design and Implementation of Cloud-based Sensor Data Management System

Kyoung-wook Park* · Kyong-og Kim** · Kyeong-jin Ban** · Eung-kon Kim***

요약

최근 대규모 센서 네트워크의 구축이 증가하면서 대규모의 센서 데이터를 효율적으로 관리하는 시스템이 요구되고 있다. 본 논문에서는 저비용, 높은 확장성 그리고 고 효율성을 지닌 클라우드 기반의 센서 데이터 관리 시스템을 제안한다. 제안된 시스템에서는 센서 데이터는 클라우드 게이트웨이를 통해 클라우드로 전송되며 이때 이상상황 검출과 이벤트 처리가 수행된다. 클라우드로 전송된 센서 데이터는 분산 컬럼 지향 데이터 베이스인 하둡 HBase에 저장되며 맵리듀스 모델 기반의 질의처리 모듈을 통해 병렬 처리된다. 처리된 결과는 REST 기반의 웹서비스를 통해 제공되므로 다양한 플랫폼의 응용프로그램과 연동이 가능하다.

ABSTRACT

Recently, the efficient management system for large-scale sensor data has been required due to the increasing deployment of large-scale sensor networks. In this paper, we propose a cloud-based sensor data management system with low cost, high scalability, and efficiency. Sensor data in sensor networks are transmitted to the cloud through a cloud-gateway. At this point, outlier detection and event processing is performed. Transmitted sensor data are stored in the Hadoop HBase, distributed column-oriented database, and processed in parallel by query processing module designed as the MapReduce model. The proposed system can be work with the application of a variety of platforms, because processed results are provided through REST-based web service.

키워드

sensor data management, cloud computing, hadoop, hbase, mapreduce

1. 서론

최근 무선 통신기기 및 센서 기술의 발달로 대규모의 USN(Ubiquitous Sensor Network)의 구축이 가능해지면서 여러 응용분야에서 이를 활용하는 방안이 연구되고 있다[1][2][3][4]. 예를 들어 도시 환경 모니터링 시스템을 구축하기 위해서는 도로, 하천, 공원, 폐기물 처리장과 같은 여러 장소에서 온도, 습도, 풍

속, 이산화탄소 농도, 소음 등과 같은 다양한 환경 데이터를 수집하는 대규모의 센서 네트워크를 구축해야 하며 수집된 대규모의 센서 데이터를 관리하는 센서 데이터 관리 시스템을 구축해야한다. 기존의 연구는 단일 서버 또는 그리드(Grid)로 구축된 다수의 서버에 분산 데이터베이스 시스템을 이용하여 센서 데이터를 관리하므로 시스템 확장이 용이하지 않으며 시스템 구축 및 관리 비용이 높다.

* 전남대학교 문화콘텐츠학부(kwpark@alex.chonnam.ac.kr)

** 순천대학교 컴퓨터학과(lisakim800@gmail.com)

** 순천대학교 컴퓨터학과(multiwave@sunchon.ac.kr)

*** 교신저자 : 순천대학교 컴퓨터학과(kek@sunchon.ac.kr)

접수일자 : 2010. 10. 20

심사(수정)일자 : 2010. 11. 22

게재확정일자 : 2010. 12. 10

최근 대규모의 데이터를 저장하고 관리하는 방안으로 클라우드 컴퓨팅(Cloud-Computing)이 각광받고 있다. 클라우드 컴퓨팅이란 인터넷 기술을 활용하여 '가상화된 IT 자원을 서비스'로 제공하는 컴퓨팅으로 사용자는 IT 자원(소프트웨어, 스토리지, 서버, 네트워크)을 필요한 만큼 빌려서 사용하고, 사용한 만큼 비용을 지불하는 컴퓨팅을 말한다[5][6][7]. 클라우드 컴퓨팅을 활용하여 관리하면 크게 두 가지의 장점을 지니게 된다. 첫째는 확장성이 높다. 클라우드를 통해 사용자는 필요한 만큼의 저장 공간과 데이터 처리를 위한 컴퓨팅 능력을 실시간으로 제공받을 수 있다. 둘째는 비용 부분이다. 사용한 만큼의 자원을 지불하면 되므로 구축을 위한 비용이나 서버 운영을 위한 비용을 절감할 수 있다. 이러한 장점으로 클라우드를 이용한 응용에 대한 연구가 활발히 진행 중이며 특히 구글의 맵리듀스(MapReduce) 병렬처리 모델이 발표된 이후 과학계산 분야에까지 그 영역을 넓히고 있다 [8][9][10].

최근 대규모의 센서 데이터를 저장 및 처리하는 방안으로 클라우드 컴퓨팅을 활용하는 연구가 진행되고 있다. [11]에서는 SOA를 통한 인터넷과 센서 네트워크의 통합을 제안하였고 [12]에서는 센서 네트워크와 클라우드를 이용한 웹기반의 응용 프로그램과의 결합에 대한 프레임워크를 제시하였다. [13]에서는 클라우드 서버를 통한 u-의료 시스템을 제시하였다. 기존의 연구들은 센서 네트워크와 클라우드와의 결합에 치중하고 있어 대규모 센서 데이터의 저장 방법이나 병렬 처리에 대한 연구가 부족하다. 또한, 빠른 응답을 요구하는 이상상황 검출(outlier detection)과 이벤트 처리(event processing)가 클라우드 내에서 이루어 지므로 응답시간이 늦어지는 단점을 지닌다.

본 논문에서는 대규모의 센서 네트워크에서 수집된 센서 데이터를 분산 저장하고 병렬 처리를 수행하는 센서 데이터 관리 시스템을 제안한다. 제안된 시스템은 클라우드 게이트웨이를 통해 센서 네트워크에서 수집된 센서 데이터를 클라우드로 전송한다. 이때 클라우드 게이트웨이는 이기종의 센서들로부터 수집된 데이터를 표준화된 메시지로 바꿔 클라우드로 전송하며 이때 규칙(rule)기반의 이상상황 검출과 이벤트 처리를 수행함으로써 빠른 응답처리가 가능하다.

클라우드로 전송된 센서 데이터는 분산 컬럼 지향

(column-oriented) 데이터베이스인 하둡(Hadoop) HBase에 저장된다. 분산 컬럼 지향 데이터베이스는 여러 노드에 분산 및 중복 저장하여 확장성이 좋고 맵리듀스 병렬 처리가 용이하다[14][15]. 본 논문에서는 효율적인 검색을 위한 데이터 스키마를 제시하고 맵리듀스 기반의 센서 데이터 처리 모듈을 설계하고 구현한다. 사용자는 REST(Representational State Transfer) 기반[16]의 웹 인터페이스를 통해 질의를 전송하고 이에 대한 결과를 제공 받도록 설계하여 다양한 플랫폼의 응용프로그램과 연동이 가능하다.

본 논문의 구성은 다음과 같다. 2절에서는 관련 연구로써, 하둡 HBase와 맵리듀스 병렬처리 모델에 대해 살펴본다. 3절에서는 본 논문에서 제안하는 클라우드 기반 센서 데이터 관리 시스템에 대해 기술하고 4절에서는 결론 및 향후연구 방향에 대해서 기술한다.

II. 관련연구

본 절에서는 본 논문에서 제안하는 센서 데이터 관리 시스템을 위한 클라우드 기술들에 대해 기술한다. 먼저 센서 데이터 저장을 위한 분산 컬럼 지향 데이터베이스 하둡(Hadoop) HBase에 대해 기술하고 센서 데이터의 병렬 처리를 위한 맵리듀스 모델에 대해 기술한다.

2.1 HBase

HBase는 구글의 BigTable을 모델로하여 하둡 HDFS (Hadoop Distributed File System)에 구현한 분산 컬럼 지향 데이터베이스이다[14][15]. HBase는 대규모 데이터 셋에 실시간으로 랜덤 액세스를 제공하며 최근 버전에서는 범위 검색을 지원한다. 또한 맵리듀스를 이용한 일괄처리 방식의 계산과 랜덤 액세스가 가능한 포인트 쿼리(point query) 방식 모두를 지원한다.

HBase에서 데이터는 테이블에 저장된다. 테이블은 로우(row)와 컬럼(column)으로 구성되며 테이블 셀(cell)은 로우와 컬럼의 교차 지점으로 원시 바이트 배열 형태로 데이터를 저장한다.

테이블 내의 하나의 로우는 로우 키와 컬럼으로 구

성되며 컬럼은 컬럼-패밀리(column family)로 그룹지어진다. 로우는 테이블의 주 키인 로우 키에 의해 정렬되며 기본적으로 바이트 순으로 정렬되며 모든 테이블 접근은 테이블의 로우 키를 통해서 이루어진다. 컬럼-패밀리는 하나 이상의 멤버들을 지닌다.

기존의 RDBMS는 대규모 확장성과 분산처리에 여러 가지 제약사항들을 가지고 있는 데에 반해 HBase는 노드만 추가하여 선형적으로 확장이 가능하다. 또한 맵리듀스를 이용하여 대규모 데이터 셋에 대한 병렬처리가 가능하여 빠른 데이터 처리가 가능하다.

2.2 맵리듀스

MPI와 같은 병렬 처리 기술은 고성능의 컴퓨팅을 요구하는 응용 분야에 초점을 맞추고 있어 처리할 데이터 양이 많은 경우엔 적용하기 어렵다. 또한 대규모 데이터 병렬 처리를 위해서는 데이터 증가에 따른 확장성 용이해야 하며 노드 간 데이터 이동에 따른 네트워크 트래픽이 최소화되도록 잡(job) 분산이 이루어져야한다. 맵리듀스는 이와 같은 사항을 고려한 구글에서 발표한 병렬처리 모델이다. 맵리듀스는 키-값(key-value) 기반의 데이터를 병렬 처리하는 모델로 여러 노드에 분할되어 저장된 입력 데이터에 대해 각각 맵(Map) 태스크를 병렬 수행하여 중간 결과들을 생성하고 이 결과들에 대해 리듀스(Reduce) 태스크를 수행하여 최종 결과를 산출하는 2단계로 구성된다.

이러한 맵리듀스 모델을 오픈 소스로 개발한 것이 아파치 그룹의 하둡 시스템이다. 하둡은 현재 IBM, Adobe, Powerset, 페이스북 등의 여러 회사에서 사용되고 있다.

III. 클라우드 기반 센서 데이터 관리 시스템

이 절에서는 본 논문에서 제안하는 클라우드 기반의 대규모 센서 데이터를 위한 관리 시스템에 대해 기술한다. 시스템의 전체적인 구성은 <그림 1>과 같다.

3.1 클라우드 게이트웨이

클라우드 게이트웨이 서버는 여러 센서 네트워크에서 발생하는 대량의 센서 데이터를 클라우드로 전송하는 역할을 수행한다. <그림 2>는 클라우드 게이트

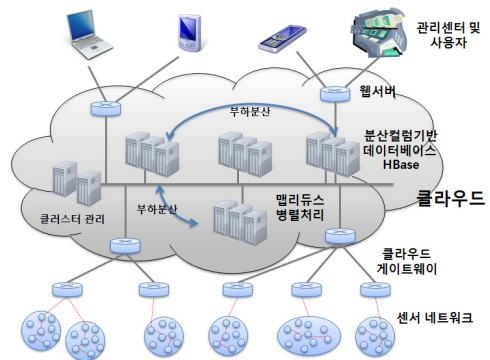


그림 1. 클라우드 기반 센서 데이터 관리 시스템 개념도

Fig. 1 Conceptual diagram of cloud-based sensor data management system

웨이 구성을 나타낸다. 클라우드 게이트웨이는 센서 네트워크 공통 인터페이스를 통해 이기종의 센서 데이터들을 표준화된 메시지(XML)로 변환하여 클라우드로 전송하고 일정 기간의 센서 데이터를 로컬 스토리지에 저장한다. 그리고 이상 상황 검출 및 지능형 이벤트 처리 모듈을 통해 응급 상황에 대한 제어신호 및 이벤트 메시지를 센서나 관리자에게 전송한다. 이때 시계열 질의와 같은 이전 시점의 센서 데이터가 필요하면 로컬 스토리지에 저장된 이전 시점의 센서 데이터를 활용함으로써 게이트웨이 내에서 이상상황 검출이나 지능형 이벤트 처리를 수행한다. 클라우드 서버는 데이터 처리능력은 높으나 빠른 응답을 보장하기는 힘들다. 따라서 클라우드 게이트웨이가 이러한 부분을 담당함으로써 빠른 응답 처리가 가능하다.

3.2 센서 데이터 스키마

클라우드 내에서 센서 데이터들은 확장성 및 분산저장을 위해 기존의 RDBMS와는 다른 컬럼-기반의 데이터 저장 시스템을 활용한다. 테이블은 로우 키와 컬럼으로 구성되며 컬럼은 컬럼패밀리로 그룹지어진다. 테이블 내의 데이터는 로우 키로 정렬되어 있으며 데이터의 접근은 로우 키를 통해서 이루어진다. 따라서 효율적인 저장 및 빠른 검색을 위해서는 기존의 RDBMS와는 다른 스키마 설계가 필요하다. 주요 테이블들의 스키마는 <그림 3>과 같다.

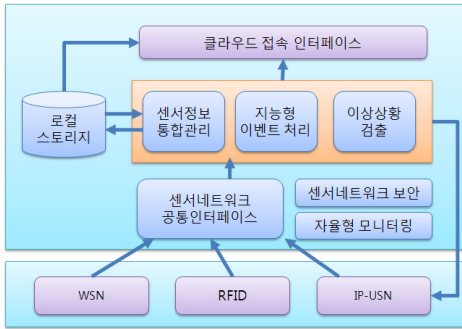


그림 2. 클라우드 게이트웨이
Fig. 2 Cloud gateway

sensor_id 테이블은 센서의 정보를 저장하는 테이블이다. 센서들은 사용자, 위치 및 센서 종류에 따라 특정 그룹에 속하게 된다. 센서 검색 및 그룹별 센서 검색이 용이하도록 <group_id>와 <sensor_id>를 조합하여 로우 키를 구성한다. 그리고 컬럼 패밀리에 센서의 정보를 저장하는 <sensor_info>와 센서의 위치 정보를 저장하는 <location>을 두어 센서에 정보를 저장한다.

sensor_type 테이블은 센서의 종류와 데이터의 종류 및 단위를 나타내는 <type_info>와 그리고 정상 데이터의 범위를 나타내는 <valid_range>를 컬럼 패밀리로 지닌다.

sensor_data 테이블은 수집된 센서 데이터를 저장하는 테이블로 최근 수집된 데이터 검색 및 범위 질의를 효율적으로 수행할 수 있도록 <sensor_id>와 역타임스탬프(reverse-timestamp) 값을 조합하여 로우 키를 구성한다

3.3 맵리듀스 센서 데이터 처리

HBase는 클라우드 내의 여러 노드들에 센서 데이터를 분산시켜 저장한다. 각 노드들의 HTable 저장된 센서 데이터는 <그림 4>와 같이 맵리듀스 모델에 의해 병렬처리된다. 센서 데이터는 클라우드 내의 각 노드들의 HTable에 분산 저장되며 테이블 매퍼(TableMapper)에 의해 각 노드에서 병렬로 사용자 질의를 수행한다. 각 노드의 테이블 매퍼를 통해 얻어진 키-값 리스트 결과들은 정렬 및 복제되어 테이블 리듀서(TableReducer)로 전송된다. 테이블 리듀서는 이들 결과들을 통합하여 최종 결과를 테이블에 저장한다.

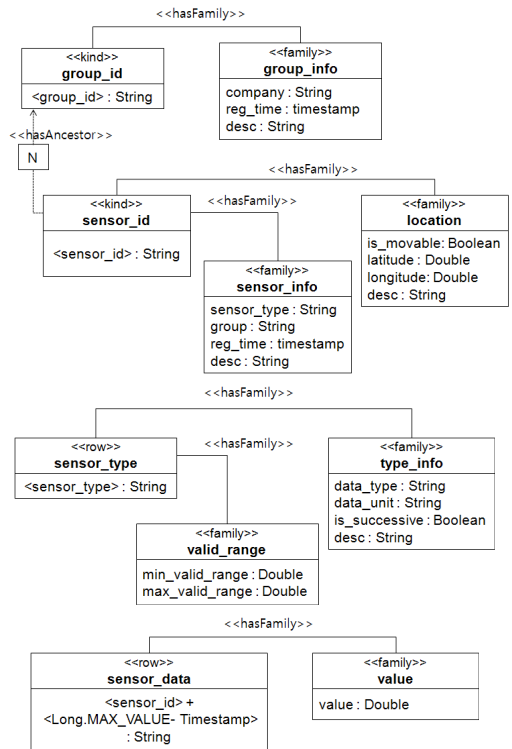


그림 3. 센서 데이터 스키마
Fig. 3 Sensor Data Schema

3.4 시스템 인터페이스

본 논문에서 제안된 시스템은 다양한 플랫폼의 응용 프로그램과 연동이 가능하도록 REST 기반의 웹 인터페이스를 제공한다. <그림 5>는 센서 데이터 관제 클라이언트를 보여준다. 센서ID, 그룹ID 및 센서 종류, 시각, 값의 크기 등의 다양한 조건으로 센서 데이터를 검색할 수 있도록 인터페이스를 구성하였다. 질의를 전송받은 서버는 맵리듀스를 수행해 센서 데이터를 병렬 처리한 후 그 결과를 XML 문서 형태로 전송한다. 웹 클라이언트는 전송받은 데이터를 자바 스크립트와 자바 FX를 이용하여 웹 브라우저에 출력한다. 센서 목록과 이들에 대한 정보는 테이블 형태로 출력되며 센서를 선택하면 해당 센서의 상세 정보와 검색 시점의 센서 데이터 값들을 차트 형태로 시각화하여 출력된다.

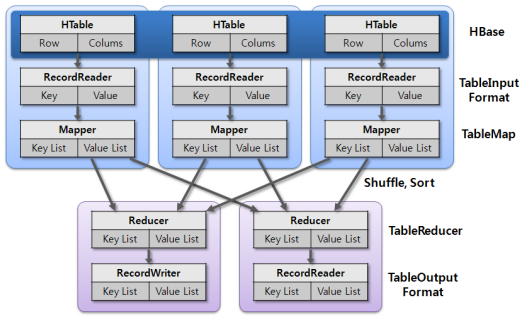


그림 4. 맵리듀스 수행과정
Fig. 4 MapReduce Execution

IV. 결론

센서 네트워크는 구축되는 순간부터 지속적으로 데이터가 발생하므로 센서 데이터를 관리하는 서버는 높은 확장성과 데이터처리 능력을 지녀야한다.

본 논문에서는 이러한 요구사항들을 적은 비용으로 만족시키는 클라우드 기반 센서 데이터 관리 시스템을 제시하였다. 제안된 시스템은 대규모의 센서 데이터를 클라우드에 분산 저장하기 위해 컬럼 지향 데이터베이스인 하둡 HBase를 이용하였다. 효율적인 검색 및 관리를 위한 데이터 스키마를 제안하고 맵리듀스 모델을 이용한 센서 데이터 병렬처리 모듈을 구현하였다. 또한 REST 기반의 웹 인터페이스를 통해 질의를 전송하고 이에 대한 결과를 제공 받도록 설계

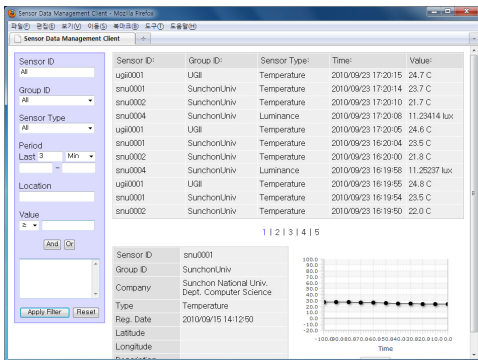


그림 5. 센서 데이터 관리 클라이언트
Fig. 5 Screenshot of the Sensor Data Management Client

하여 다양한 플랫폼의 응용프로그램과 연동이 가능하다.

감사의 글

본 지식재산권은 지식경제부 및 정보통신산업진흥원의 지원을 받아 수행된 연구결과임.
(10-기반. 산업원천기술개발사업)

참고 문헌

- [1] K. Aberer, G. Alonso, and D. Kossmann, "Data Management for a Smart Earth - The Swiss NCCR-MICS initiative", SIGMOD Record, Vol. 35, No. 4, pp. 40-45, 2006.
- [2] K. Aberer, M. Hauswirth, and A. Salehi, "Infrastructure for Data Processing in Large-Scale Interconnected Sensor Networks", 2007 International Conference on Mobile Data Management, pp. 198-205, May 2007.
- [3] C. Jardak, J. Riihijärvi, and P. Mähönen, "Extremely Large-scale Sensing Applications for Planetary WSNs", In Proceedings of the 2nd ACM international Workshop on Hot Topics in Planet-Scale Measurement, pp. 1-6, June 2010.
- [4] K. Aberer, G. Alonso, and D. Kossmann, "Data Management for a Smart Earth", SIGMOD Record, Vol. 35, No. 4, pp. 40-45, 2006.
- [5] B. Hayes, "Cloud Computing", Communications of The ACM, Vol. 51, No. 7, pp. 9-11, July 2008.
- [6] I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud computing and Grid Computing 360-Degree Compared", Grid Computing Environment Workshop(GCE'08), pp. 1-10, Nov. 2008.
- [7] R. Buyya, C. S. Yeo, and S. Venugopal, "Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities", Proceedings of the 2008 10th IEEE International Conference on High Performance Computing and Communications, pp. 5-13, 2008.
- [8] J. Dean and S. Ghemawat, "Mapreduce: Simplified Data Processing on Large Clusters", In OSDI'04: 6th Symposium on Operating

System Design and Implementation, December 2004.

- [9] H. Liu and D. Orban, "GridBatch: Cloud Computing for Large-Scale Data-Intensive Batch Applications", Proc. 8th IEEE International Symposium on Cluster Computing and the Grid (CCGRID'08), pp. 295 - 305, May 2008.
- [10] C. Vecchiola, S. Pandey, and R. Buyya, "High-Performance Cloud Computing: A View of Scientific Applications", 2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks, pp. 4-16, 2009.
- [11] V. Rajesh, J. M. Gnanasekar, R. S. Ponmagal, and P. Anbalagan, "Integration of Wireless Sensor Network with Cloud", 2010 International Conference on Recent Trends in Information, Telecommunication and Computing, pp. 321-323, 2010.
- [12] M. Hassan, B. Song, and E. Huh, "A Framework of Sensor-cloud Integration Opportunities and Challenges", Conference On Ubiquitous Information Management And Communication, pp. 618-626, 2009.
- [13] X. Lee, S. Lee, P. T. True, L. T. Vinh, A. M. Khattak, M. Han, D. V. Hung, M. M. Hassan, M. Kim, K. Koo, Y. Lee, and E. Huh, "Secured WSN-integrated cloud computing for u-life care", In Proceedings of the 7th IEEE Conference on Consumer Communications and Networking Conference, pp. 702-703, Jan. 2010.
- [14] Hadoop : <http://hadoop.apache.org>
- [15] HBase : <http://hadoop.apache.org/hbase>
- [16] Roy Thomas Fielding, "Architectural Styles and the design of Network-based Software Architectures", PhD. thesis, University of California, Irvine, 2000.

저자 소개



박경욱(Kyoung-wook Park)

1996년 8월 : 순천대학교 전자계산학과 (이학사)

1999년 8월 : 전남대학교 전산통계학과 (이학석사)

2004년 8월 : 전남대학교 전산학과 (이학박사)

※ 관심분야 : 병렬 및 분산처리, 그래프 이론, 알고리즘



김경옥(Kyong-og Kim)

2005년 2월 : 한려산업대학교 사회복지학과 졸업(문학사)

2008년 2월 : 순천대학교 컴퓨터과학과 (이학석사)

2010년 2월~현재: 순천대학교 컴퓨터과학과 박사과정

※ 관심분야 : 영상처리, 컴퓨터 그래픽스, USN, RFID, 증강현실



반경진(Kyeong-jin Ban)

2003년 2월 : 순천대학교 컴퓨터과학과 (이학사)

2005년 2월 : 순천대학교 컴퓨터과학과 (이학석사)

2007년 8월 : 순천대학교 컴퓨터과학과 박사수료

※ 관심분야 : 컴퓨터 그래픽스, RFID, USN



김응곤(Eung-kon Kim)

1980년 2월 : 조선대학교 전자공학과 (공학사)

1986년 2월 : 한양대학교 컴퓨터공학과 (공학석사)

1992년 2월 : 조선대학교 컴퓨터공학과 (공학박사)

1993년 3월 ~ 현재 : 순천대학교 컴퓨터과학과 교수

※ 관심분야 : 영상처리, 컴퓨터 그래픽스, 멀티미디어, HCI