

# 대용량 센서 데이터 스트림 처리를 위한 프레임워크 설계

최현화, 원종호, 이훈순, 채미옥, 박재홍, 정원일, 김병섭, 이명철, 이미영,  
한국전자통신연구원 디지털 홈 연구단 인터넷 서버 그룹  
e-mail : {hyunwha, jhwon, hunsoon, mio, jaehong, wncung, powerkim, mclee, mylee}@etri.re.kr

## Design of the Framework for Processing Large Sensor Data Stream

Hyun-Hwa Choi, Jong-Ho Won, Hun-Soon Lee, Mi-Ok Chae,  
Jae-Hong Park, Warnill Chung, Byung-Seob Kim, Myung-Cheol Lee, Mi-Young Lee  
Internet Server Technology Group, Digital Home Research Division, ETRI

### 요 약

센서 데이터 스트림 처리 기술은 유비쿼터스 컴퓨팅의 핵심 기술로 그 중요성이 날로 증대되어, 현재 산업계는 물론 학계에서도 그 역량을 이에 집중하고 있다. 본 논문에서는, 센서 데이터 스트림에 대한 특징 및 이를 처리하기 위한 산업계와 학계의 동향을 알아본다. 그리고, 대용량 센서 데이터 스트림을 처리하기 위해 요구되어지는 기능을 정리하고, 이를 기반으로 센서 데이터 스트림 처리 시스템의 구조를 설계한다

### 1. 서론

하드웨어 발전으로 인한 마이크로 컨트롤러의 소형화와 무선 통신 기술의 발전은 경량의 센서 노드들이 무선 통신으로 연결되는 센서 네트워크를 가능하게 하고 있다. 이러한 센서 네트워크는 환경 감시, 가정 또는 사무실의 자동화 및 방범, 건강 모니터링, 재무 분석, 물류/유통 시스템, 웹 로그와 같은 다양한 유비쿼터스 응용 서비스를 제공한다.

차세대 정보 통신 비전으로 간주되는 유비쿼터스 컴퓨팅은 그 핵심 기술로, 센서 기술, 센서 네트워킹 기술 및 센서 데이터 스트림 처리 기술 등이 대두되고 있다. 특히, 센서 데이터 스트림 처리 기술은 다양한 센서(RFID, GPS, Temperatures, Vital signal)의 센싱 데이터들을 가공, 처리하여 응용 서비스에 제공하는 기술로 그 중요성이 날로 증대되고 있으며, 산업계는 물론 학계에서도 이에 그 역량을 집중하고 있는 실정이다. 센싱 데이터는 연속적이며(continuous), 무제한(unbounded)으로 생성되고, 생성된 데이터의 양이 시간에 따라 변하는(time-varying) 특징을 가진다.

그러므로, 유비쿼터스 컴퓨팅 환경에서 끊임없이

생성되는 센서 데이터 스트림 처리를 통한 고부가 가치의 응용 서비스를 쉽고 편리하게 제공하기 위해서, 기존의 데이터베이스 관리 시스템(Database Management System, DBMS)과는 다른 센서 데이터 스트림 처리 시스템이 필수적으로 요구된다. 이러한 센서 데이터 스트림 처리를 위한 시스템의 특징을 간단히 요약하면 다음과 같다.

- 대량의 실시간 센서 데이터 스트림의 입력은 데이터 업데이트(update)를 빈번히 발생시키며, 모든 데이터를 데이터베이스 관리 시스템이 저장하는 데는 한계가 있다. 즉, 센서 데이터 스트림 처리 시스템은 온라인으로 데이터를 처리하거나 요약 정보만을 저장할 수 있어야 한다.
- 실시간 정보의 반영을 원하는 대부분의 응용 서비스를 위하여, 센서 데이터 스트림은 실시간 처리가 요구된다.
- 계속적으로 들어오는 센서 데이터 스트림의 처리에 따른 변화의 추이를 모니터링할 수 있도록 영속적인 질의가 이뤄져야 한다.

- 다양한 사용자 요구를 반영하기 위하여, 다수의 질의가 한꺼번에 수행되어야 한다.

본 논문에서는, 다양한 종류의 스마트 객체<sup>1</sup>를 센싱함으로써 들어오는 센서 데이터 스트림의 효율적인 처리 및 다양한 조건들로 구성되는 이벤트<sup>2</sup>를 정의하여 특정 서비스를 실행할 수 있도록 함으로써, 다양한 유비쿼터스 환경 구축을 용이하게 할 수 있는 대용량 센서 데이터 스트림 처리 프레임워크, UbiCore 를 제안한다.

본문의 구성은 다음과 같다. 2 장에서는 RFID 데이터 처리를 위한 미들웨어와 현재 진행 중인 대표적인 데이터 스트림 프로젝트를 살펴본다. 3 장에서는 대량의 센서 데이터 스트림의 효율적인 처리를 위한 센서 데이터 스트림 처리 프레임워크를 설계하고, 4 장에서 결론을 맺는다.

## 2. 관련 연구

### 2.1 RFID 처리 미들웨어

EPCglobal 예서는 RFID 기술을 기반으로 유비쿼터스 컴퓨팅 서비스 환경에서 응용 서비스를 개발하기 위하여, EPC (Electronic Product Code), 전파 식별 (Radio Frequency Identification, RFID) 태그, Reader, SAVANT, ONS (Object Name Server), EPC 정보 서버 (Information Server)로 구성된 EPC Network Architecture 를 제시하고 그에 따른 표준을 제공하고 있다 [1]. 현재, 국내의 많은 업체들이 EPCglobal 에서 제시하는 표준을 따르는 tag, reader 와 같은 장치를 제작하고 있으며, Savant 에 해당하는 RFID 관련 미들웨어 제품들도 속속들이 출시되고 있다. 대표적으로 SUN 사의 Savant, CapTech 사의 TagsWare, GlobeRanger 사의 iMotion, ConnecTerra 사의 RFTagWare 등이 있다. SUN 사의 자바 플랫폼 기반의 RFID 이벤트 관리기(Event Manager)는 크게 Reader 와의 상호작용을 위한 Adapter, 유용한 데이터 추출 및 비즈니스 로직을 포함하는 Filter, 정보 로깅을 위한 Logger, 그리고 기업 응용을 위한 Enterprise Gateway 로 구성되며 네트워크 상의 오류 수용 및 소프트웨어 서비스의 동적 재배치와 같은 시스템의 융통성 있는 구성 능력을 제공한다 [2]. CapTech 사의 TagsWare 는 특정 기능을 가지는 Link 들의 조합을 통하여 센서 데이터 스트림의 처리 및 사용자의 질의를 구현할 수 있는 추상화된 인터페이스를 제공하는 것이 특징이다 [3]. GlobeRanger 의 iMotion 은 Microsoft .NET 프레임워크를 기반으로 데이터 정화(data smoothing), 필터링(filtering), 집합화(aggregation), 큐잉(queueing), 오류 처리(error handling) 및 스케줄링(scheduling)을 위한 이벤트 필터링을 제공하고 데이터의 교환 및 통합에서는 웹 서비스(Web Services) 기술을 이용한다 [3]. ConnecTerra 사의

RFTagWare 는 RFID 어플리케이션 개발 및 배치와 같은 인프라스트럭처의 관리에 주안점을 두고 있다 [5]. 이들 미들웨어는 모두 센서 데이터의 실시간 처리 및 응용 서비스 연동을 지원한다.

### 2.2 데이터 스트림 관리 시스템

여러 선도 연구 기관에서도 센서 데이터 스트림 처리에 적합한 시스템 (Data Stream Management System, DSMS)을 독자적으로 설계하고, 프로토타입(prototype)을 개발하고 있다. 일반적인 DSMS 의 구조는 그림 1 과 같다. DSMS 는 데이터 스트림의 양을 조절하기 위한 입력 모니터(Input Monitor), 윈도우 질의 처리를 위한 작업 저장소(working storage), 근사값 데이터를 저장하기 위한 요약 저장소(summary storage), 그리고 메타 데이터를 저장하는 정적 저장소(static storage)를 유지한다. DSMS 는 수행되는 연속 질의를 유지 및 관리하기 위한 질의 저장소(Query Repository), 질의 계획 수립 및 최적화, 실제 질의를 수행하는 질의 처리기(Query Processor), 연산 결과가 들어가는 출력 버퍼(Output Buffer)로 이뤄진다.

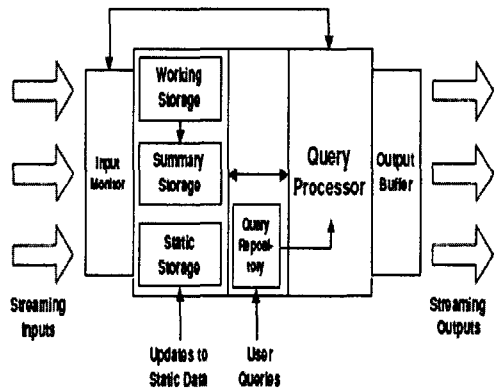


그림 1. 일반적인 DSMS 구조

Brown 대학, Brandeis 대학, MIT 대학이 공동으로 수행중인 Aurora 프로젝트는 QoS 기반 연산자 스케줄링, 부하 감쇄 방법, 풀 방식과 푸시 방식 데이터 처리를 함께 할 수 있는 복합 데이터 스토리지 최적화 등과 같은 실시간 데이터 이슈에 초점을 맞추고 있다. Aurora 프로토타입은 데이터 스트림의 한정된 영역에 동작할 수 있는 윈도우 기반 연산자를 포함하는 Filter, Map, Union, BSort, Aggregate, Join, Resample 의 연산자 집합을 제공하고 있다[6][7]. 스탠포드 대학에서 수행 중인 STEAM(STandford stream data Manager) 프로젝트는 SQL 을 확장하여 스트림 데이터에 관련된 연속 질의를 위한 CQL(Continuous Query Language)을 정의하였으며, Relation 과 스트림의 관계 및 그 변환을 위하여 Istream, Dstream, Rstream 과 같은 변환 연산자를 제공한다 [8][9]. 버클리(Berkeley) 대학에서는 공개 데이터베이스 시스템인 PostgreSQL 을 변형하여 연속 질의를

<sup>1</sup> Smart chip 등이 내장, 부착된 물건, 사람, 건물 등의 객체

<sup>2</sup> 특정 상황의 발생 여부

수행하는 TelegraphCQ 프로젝트를 수행 중에 있으며, Grouped filter 를 이용하는 SteM 과 튜플의 라우터 및 연산 스케줄러인 Eddy 를 통해 공유(sharing) 와 적응성(adaptivity)에 중점을 두고 있다 [10][11]. 또한 데이터 처리 프로젝트로 XML 문서 처리를 위한 NiagaraCQ [12], 인터넷 상의 변화 내용 모니터링을 위한 OpenCQ [13], 센서 네트워크용 분산 데이터 처리 시스템을 위한 COUGAR 등이 있다 [14][15].

### 3. 대용량 센서 데이터 스트림 처리 프레임워크 (UbiCore) 설계

#### 3.1 UbiCore 특징

우리가 개발하고자 하는 대용량 센서 데이터 스트림 처리 프레임워크인 UbiCore 의 목적은 다수의 각종 센서로부터 들어오는 대용량 센서 데이터 스트림을 처리하여, 유비쿼터스 환경의 다양한 응용에서 요구하는 정보를 실시간으로 제공해줌으로써, 서비스 제공자가 센싱되는 상황 및 이에 대응하는 서비스 개발에만 전념할 수 있도록 하는 프레임워크를 제공하고자 하는 것이다. 따라서, 대용량 센서 데이터 스트림 처리 프레임워크 설계 시, 다양한 센서와 서비스를 쉽게 연동할 수 있도록 하기 위해, 표준 프로토콜과 인터페이스를 따르도록 하며, 서비스가 운용되는 환경 및 발생하는 센서 데이터 스트림의 양의 다양성을 고려하도록 한다.

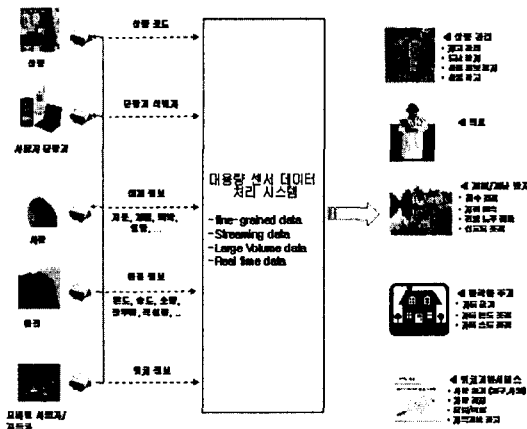


그림 2. UbiCore 운영 환경

대용량 센서 데이터 스트림 처리 프레임워크의 특징은 다음과 같다.

- 다양한 RFID 리더 및 각종 센서와 연동하여 센서 데이터를 입력받을 수 있으며, 새로운 센서 연동을 위한 확장이 용이하다.
- 다양한 XML 형식을 갖는 센서 데이터 스트림을 실시간으로 처리한다.

- 센서 데이터에 대한 실시간 로깅을 지원한다.
- 센싱된 데이터를 기반으로 한 특정 이벤트를 확장 XQuery를 이용하여 정의할 수 있다.
- 연속적으로 들어오는 센서 데이터 스트림에 대한 다수 개의 연속 XQuery 처리를 지원한다.
- 정의된 이벤트 발생 시에 Web Services, JMS 를 연동할 수 있는 방법을 제공한다.
- 각종 센싱 데이터에 대한 사용자의 인증에 따른 보안을 제공하기 위해 센싱 데이터의 필드별 및 조건 기반의 fine-grained 접근 제어를 제공한다.
- 분산 환경 하에서의 처리 기능을 지원 한다.
- 관리가 용이하도록 센서 통합 관리 도구 등 GUI 기반의 자원 도구를 제공한다.

#### 3.2 UbiCore 블록 구조

대용량 센서 데이터 스트림 처리 프레임워크는 크게 각종 센서를 연동하고, 이들로부터 들어오는 대용량의 센서 데이터 스트림을 처리하는 데 필요한 이벤트 및 특정 이벤트 발생 시에 구동할 응용 서비스를 정의할 수 있어야 한다. 이러한 기능을 가지는 대용량 센서 데이터 스트림 처리 프레임워크의 블록 구조는 그림 3과 같다.

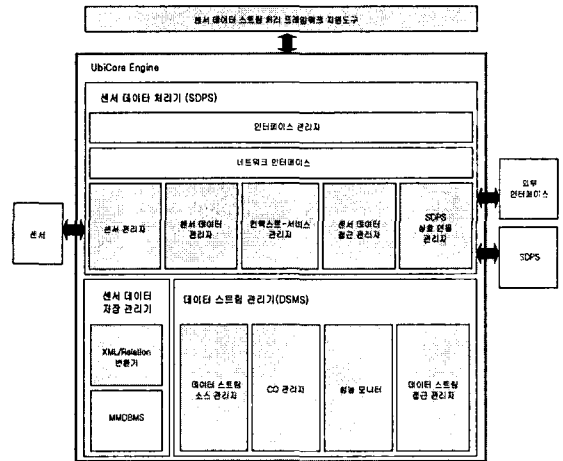


그림 3. UbiCore 블록 구조도

- UbiCore 지원 도구  
센서 데이터 스트림 처리 프레임워크 지원도구는 대용량 센서 데이터 스트림 처리 프레임워크에서 지원하는 주요 자원 및 기능인 센서, 센서 데이터, 컨텍스트-서비스의 운영 및 관리 모니터링의 기능을 지원하는 GUI 기반 도구이다.
- 센서 데이터 처리기  
센서 데이터 처리기는 대용량 센서 데이터 스트림의 처리 기능을 제공하는 시스템으로, 센서의 등록

및 연결 정보를 관리하는 센서 관리자, 센서 데이터 소스 관리, 센서 데이터 브라우징 및 센서 데이터 로 그를 관리하는 센서 데이터 관리자, 컨텍스트-서비스 정보를 관리하고, 서비스 메시지 생성, 실제 원격 서비스 호출을 담당하는 컨텍스트-서비스 관리자, 센서 데이터 접근 권한을 설정하는 센서 데이터 접근 관리자, 다른 센서 데이터 처리기와와 상호 연동을 담당하는 SDPS 상호 연동 관리자로 구성된다. 여기서 센서 데이터 소스는 연동된 센서들을 그룹화하여 데이터 스트림 관리기에 전달하는, 실제 데이터 스트림 관리기의 입력 데이터 소스를 말한다. 컨텍스트-서비스에서 컨텍스트는 영속 질의(Continuous Query, CQ)를 통해 표현되는 이벤트 정의를 말하며, 서비스는 실제 구동될 응용 서비스의 정의 및 전달할 메시지를 생성하는 데 있어 필요한 매핑 정보를 포함한다.

● 데이터 스트림 관리기

데이터 스트림 관리기는 센서 데이터 처리기로부터 들어오는 데이터 스트림에 대한 스키마 및 입력 쿼리를 담당하는 데이터 스트림 소스 관리자, 실제 영속 질의(CQ)를 파싱, 실행 계획 수립, 스케줄링 및 실행을 담당하는 CQ 관리자, 영속 질의의 실행기 성능 측정을 하여 스케줄링의 정보를 제공하기 위한 성능 모니터, 데이터 스트림에 대한 사용자의 접근 권한 및 인증 정보를 관리하는 데이터 스트림 접근 관리자를 포함한다.

● 센서 데이터 저장 관리기

센서 데이터 저장 관리기는 센서로부터 전달되는 센서 데이터 중에서 히스토리컬 질의(historical Query)를 지원하기 위해 영구히 저장할 필요성이 있는 데이터의 저장을 담당하며, 이는 XML/Relation 변환기 및 MMDBMS(Main Memory DataBase Management System)로 구성된다.

3.3 UbiCore의 주요 기능

대용량 센서 데이터 스트림 처리 프레임워크에서는 EPC 표준을 지원하는 리더는 물론 온도, 습도 등을 측정하는 각종 센서를 연동할 수 있도록 하기 위하여 다양한 센서 데이터 형식을 관리할 수 있는 PML core 스펙을 따르며, 불필요한 정보 및 사용자의 요구에 따른 데이터 제거를 수용하기 위하여 데이터의 필터링 기능을 지원한다. 컨텍스트를 표현하는데 있어서는 크게 영속 질의(Continuous Query)와 히스토리컬 질의(Historical Query)를 지원한다. 질의의 형태는 XQuery(XML Query language) 기반의 스트림 데이터 질의 언어를 사용한다. 영속 질의 표현을 위하여 시간 또는 데이터의 수량에 의존하는 슬라이딩 윈도우 개념을 포함하여, 질의의 수행 시점 및 질의의 수명을 묘사할 수 있도록 한다. 질의 표현에 있어서 기본적으로 EQ, NE, LT, LE, GT, GE 를 포함한 비교 연산, 덧셈(addition), 뺄셈(subtraction), 곱셈(multiplication), 나눗셈(division), 몫(modulus)의 수식 연산, AND, OR 를 포함한 논리 연산, 포함 여부(contains), 비교(like), 부분 문

자열(substring)의 문자열 연산, AVG, COUNT, MAX, MIN, SUM 의 집합 함수(aggregation function) 연산 및 합집합(set), 교집합(intersection), 차집합(difference)의 집합(set) 연산을 제공한다. 또한 XML 구조 연산, AFTER 의 키워드를 이용한 시간 순서 연산, 기존 DBMS 의 Selection, Projection, Join, Order-by 및 Group-by 연산을 지원하며, 사용자 정의 함수를 지정할 수 있는 통로를 제공한다.

4. 결론

본 논문은 최근 이슈화 되고 있는 센서 데이터 처리에 관련해, 대용량의 센서 데이터 스트림을 효율적으로 처리하여 유비쿼터스 컴퓨팅 환경하의 응용 서비스에 실시간 정보를 제공하기 위한 프레임워크를 설계하였다. 대용량 센서 데이터 스트림 처리 프레임워크는 실시간 데이터 스트림 처리 및 사용자가 관심 있는 상황, 즉 이벤트를 정의할 수 있도록 지원하기 위하여 영속 질의와 히스토리컬 질의를 지원한다. 또한, 다양한 센서 및 응용 서비스를 연동할 수 있도록 하였으며, 대량의 센서 데이터의 효율적인 처리를 위하여 분산 환경을 고려하여 설계하였다.

참고문헌

[1] <http://www.epcglobalinc.org/>  
 [2] <http://www.sun.com/rfid>  
 [3] <http://www.captchventures.com>  
 [4] <http://www.globeranger.com>  
 [5] <http://www.connecterra.com>  
 [6] <http://www.cs.brown.edu/research/aurora>  
 [7] D. Abadi D. Carney, U. Cetintemel, M. Cherniack, C. Convey, S. Lee, M. Stonebraker, N. Tatbul, S.Zdonik. Aurora: A New Model and Architecture for Data Stream Management. In VLDB Journal, August 2003  
 [8] <http://www-db.stanford.edu/stream>  
 [9] D. Carney, U. Cetintemel, M. Cherniack, C. Convey, S. Lee, G. Seidman, M. Stonebraker, N. Tatbul, and S. Zdonik. Monitoring Streams: A New Class of Data Management Applications. In proceedings of the 28th International Conference on Very Large Data Bases (VLDB'02), Hong Kong, China, August 2002  
 [10] <http://telegraph.cs.berkeley.edu/techover.html>  
 [11] Calton Pu and Ling Liu. "Update Monitoring: The CQ project". (invited paper) in: The 2nd International Conference on Worldwide Computing and Its Applications - WWCA'98, Tsukuba, Japan, Lecture Notes in Computer Science 1998, pp396-411.  
 [12] <http://www.s.wise.edu/niagara>  
 [13] Calton Pu and Ling Liu. "Update Monitoring: The CQ project". (invited paper) in: The 2nd International Conference on Worldwide Computing and Its Applications - WWCA'98, Tsukuba, Japan, Lecture Notes in Computer Science 1998, pp396-411.  
 [14] <http://www.cs.cornell.edu/boom/2003sp/ProjectArch/CougarSM/index.php>  
 [15] Johannes Gerhke, "COUGAR DESIGN AND IMPLEMENTATION"