

데이터 웨어하우징에서의 보안성 연구방향

Security Direction on The Data Warehousing

심 감 식*

요 약

데이터 웨어하우스 이면의 있는 아이디어는 다양한 이질형 데이터베이스에 있는 데이터를 접근하는 것이 거주장 스텝다는 것이다. 이질적인 환경에서 질의를 처리하기 위해 몇몇 처리 모듈들이 서로 협력할 필요가 있다. 그러므로 다양한 데이터 원천(source)들에서 본질적인 데이터를 함께 가져다 놓는 곳이 데이터 웨어하우스이다. 이런 방법에서 사용자들은 웨어하우스만을 질의한다.

데이터 웨어하우스 개발에서는 부가적인 보안 사항을 초래한다. 예를 들면, 다양한 데이터 탐사 도구를 이용함으로써 정보를 연역할 수 있는가? 데이터 웨어하우스를 위한 적당한 감사 프로시유어는 무엇인가? 본 연구에서는 데이터 웨어하우스에서의 보안 문제들을 알아본다.

1. 개요

데이터 웨어하우징은 데이터 관리에서 가장 인기있는 주제 중의 하나이다. 몇몇 기관들은 자기 자신의 웨어하우스를 구축하고 있으며, 상용 DBMS 판매사들도 웨어하우징 제품을 판매하고 있다. 더욱이, 어떤 회사는 데이터 웨어하우스 개발에만 전념하고 있다. 그러면 데이터 웨어하우스란 무엇인가? 이 이면의 아이디어는 이질형 데이터베이스의 데이터를 접근하는 것이 종종 거주장스럽다(cumbersome)는 것이다. 몇 개의 처리 모듈들이 이질형 환경에서 질의 처리를 위해 서로 협력할 필요가 있다. 그래서, 이질형 데이터베이스에 있는 기

본적인 데이터를 데이터 웨어하우스에다 가져다 놓는다. 이런 방법으로 사용자는 웨어하우스만을 질의할 필요가 있다.

Inmon^[INMON93]이 설명한 것처럼, 데이터 웨어하우스는 주제 중심(subject-oriented)적이다. 그들의 설계는 그들을 활용하는 응용에 매우 종속된다. 그들은 다양한 이질형 데이터 원천(source)들을 통합한다. 그들은 지속적(persistent)이다. 즉, 웨어하우스는 데이터베이스와 매우 유사하다. 그들은 시간에 따라 변화한다. 그것은 웨어하우스를 구축한 데이터 원천들이 갱신될 때는, 그 변화가 웨어하우스에 반영되어야 하기 때문이다. 기본적으로 데이터 웨어하우스는 기업체나 조직체의 의사결정 지원 기능을 한다. 예를 들면, 데이터 원천들은 비가공(raw) 데이터를 가질 수 있는 반면에,

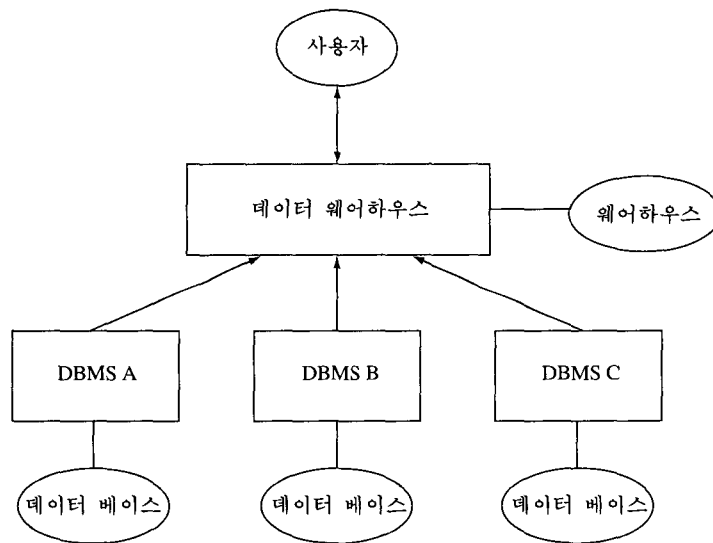
* 진주 산업대학교 조교수

데이터 웨어하우스에는 상관(correlated) 데이터, 요약 보고(summary reports), 그리고 비가공 데이터에 적용될 집계 함수(aggregate functions)가 있을 수 있다.

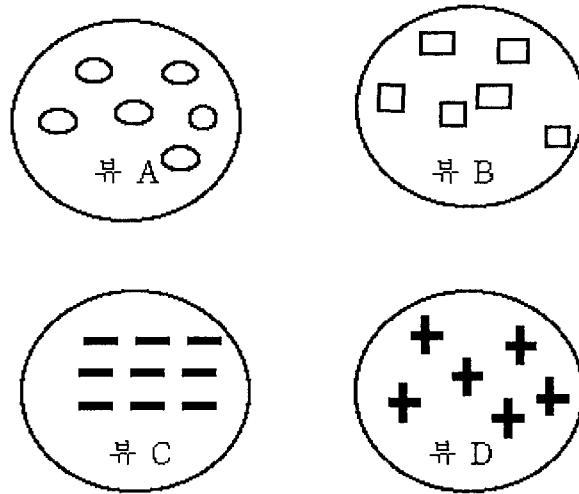
그림 1은 데이터 웨어하우스를 나타내고 있다. 데이터 원천은 DBMS A, B, 그리고 C에 의해 관리된다. 이들 데이터베이스에 있는 정보는 병합(merge)되어 웨어하우스에 더해진다. 정보를 병합하는 여러 가지 방법이 있다. 하나는 데이터베이스를 단순히 중복(replicate)시키는 것이다. 이것은 이질형 데이터베이스 접근에 대한 장점이 없다. 두 번째의 경우는 정보를 중복시키는 것이지만, 불일치(inconsistency)나 중복성(redundancy)은 없다. 세 번째 접근방법은 데이터베이스에 있는 정보의 부분집합을 선택하여 그것을 웨어하우스에 배치시킨다. 여기에는 몇 가지 논점이 있다. 부분집합을 어떻게 선택할 것인가? 그들을 무작위(random) 선택할 것인가 혹은 데이터를 선택하는 어떤 방법이

있는가? 예를 들면, 관계형 데이터베이스라고 가정 할 때, 한 릴레이션에 있는 다른 모든 행(row)을 택해서 이 행들을 웨어하우스에 저장한다. 세 번째 접근방법을 약간 변경시킨 네 번째 접근방법은 사용자들이 제시하는 질의 유형을 결정한 다음에 데이터를 분석하여 사용자가 필요한 데이터만을 저장시킨다. 이것을 온라인 트랜잭션 처리(OLTP)와 반대되는 온라인 분석 처리(OLAP: on-line analytical processing)라고 부른다.

데이터 웨어하우스에서 데이터는 응용마다 서로 다르게 볼(view) 수도 있다. 즉, 데이터는 다차원적(multidimensional)이다. 예를 들면, 급여지불(payroll) 부서는 어떤 형식(format)으로 된 데이터를 필요로 한 반면에, 프로젝트 부서는 다른 형식으로 된 데이터를 필요로 할 수도 있다. 웨어하우스는 그런 다차원 데이터를 지원해야 한다. 동일 데이터에 대한 다중 뷰(multiple view)는 그림 2에 나타나 있다.



[그림 1] 데이터 웨어하우스 예



[그림 2] 데이터의 다중 뷰(view)

웨어하우스를 형성하기 위해 데이터 원천들을 통합할 때, 응용을 분석하여 웨어하우스에 배치할 적절한 데이터를 선택하는 것이 과제이다. 단지 요약(summary)이나 평균(average)만을 데이터 웨어하우스에 저장할 때는 어떤 계산(computation)이 수행되어야 한다. 웨어하우스가 질의에 대한 모든 정보를 가지고 있는 것은 항상 아니다. 이 경우, 질의 실행을 완료하기 위해서 웨어하우스는 이질형 데이터 원천들로부터 데이터를 가져온다. 또다른 과제는 각각의 데이터베이스들이 갱신될 때 웨어하우스를 어떻게 할 것인가이다. 그 갱신을 웨어하우스에 전파(propagate)시키는 방법은 무엇인가? 보안을 어떻게 유지할 것인가? 이것들이 연구되고 있는 논점들이다.

2. 데이터 웨어하우스를 위한 기술

그림 3은 데이터 웨어하우스 기술을 나타낸다 [THUR97a]. 즉, 데이터 웨어하우스를 개발하려면 몇 가지 기술들이 통합되어야 한다. 이들에는 이질

형 데이터베이스 통합, 통계 데이터베이스, 데이터 모델링, 메타데이터 관리, 접근방법 및 색인, 질의 언어, 데이터베이스 관리(administration), 데이터베이스 보안, 분산 데이터베이스 관리, 그리고 고성능 데이터베이스 관리가 포함된다. 이 섹션에서는 데이터 웨어하우스 맥락으로 이들 기술을 간략히 조사하겠다.

이질형 데이터베이스 통합(heterogeneous database integration)은 데이터 웨어하우스에서 기본적인 구성요소이다. 웨어하우스를 구축하기 위해서는 여러 이질형 데이터 저장소(store)들에 있는 데이터가 통합되어야 하기 때문이다. 그러나, 주된 차이점이 있다. 종종 이질형 데이터베이스 통합에서는 데이터를 저장하기 위한 단일 저장소(single repository)가 없다. 그러나, 웨어하우스에서는 웨어하우스의 데이터를 저장하기 위한 단일 저장소가 보통 있고 이 저장소도 관리되어야 한다.

통계 데이터베이스(statistical database)는 합계(sum), 평균(average), 그리고 그밖의 집계(aggregate)와 같은 정보를 유지한다. 통계

데이터베이스에 대한 많은 논점이 있다. 예를 들면, 데이터베이스가 갱신될 때 요약 데이터를 어떻게 유지보수할 것인가? 개개의 데이터 항목들을 어떻게 보호할 것인가? 예를 들면, 평균 봉급은 U등급으로 한 반면에 개개의 봉급들은 S 등급으로 한다. 웨어하우스는 요약정보를 보유하고 있으므로, 통계 데이터베이스를 관리하는 기법들이 웨어하우스에서도 검토될 필요가 있다.

- * 데이터 모델링
- * 데이터베이스 보안
- * 이질형 데이터베이스 통합
- * 분산 데이터베이스
- * 접근 방법 및 색인
- * 질의 언어
- * 데이터베이스 관리(Administration)
- * 고성능 컴퓨팅
- * 메타데이터

[그림 3] 데이터 웨어하우스 기술

데이터 모델링(data modeling)은 데이터 웨어하우스 구축에 기본적인 일(task)이다. 그 데이터 모델은 이질형 데이터 원천들이 이용하고 있는 데이터 모델의 영향을 받을 것인가? 백지 상태에서(scratch) 새로운 데이터 모델을 개발해야 하는가? Inmon은 데이터 모델을 개발하는 몇 가지 단계를 제시했다^[INMO93]. 고 수준(higher level)에는 세 단계가 있다. 즉, 협력(corporate) 모델 개발, 기업(enterprise) 모델 개발, 그리고 웨어하우스 모델 개발이다. 중간 수준(middle level)에서는 각 주제(subject)에 대한 모델이다. 물리적 수준(physical level)에서는 키(key)와 같은 특성들을 포함한다. 어떤 사람들은 이것이 너무 긴

과정이므로 곧바로 웨어하우스 모델로 해야 한다고 주장한다. 데이터 웨어하우스 개발에 대한 많은 경험들이 발표되고 있기 때문에, 이런 논점은 해결될 것이다.

적절한 접근방법 및 색인(access method and index) 전략이 웨어하우스를 위해 개발되었다. 예를 들면, 웨어하우스는 구조적(structured)이어서 질의 처리가 용이하다. 다음과 같은 예제 질의가 있을 수 있다. 의사들은 1995년에 50K 이상 되는 빨간색 차를 몇 대 샀는가? 이 질의를 처리하기 위해서는 많은 릴레이션이 조인(join)되어야 한다. 실제 데이터를 조인하지 않고, 연관(associated) 데이터에 대한 비트맵(bit map)을 결합함으로써 결과를 얻을 수 있다. 즉, 웨어하우스는 비트맵 색인(여기서, 해답이 데이터베이스에서 긍정이면 비트 맵에 1이 있다)이라는 색인 전략을 활용한다. 즉, 차의 색깔이 빨간색이면, 연관된 비트 맵에는 1이 될 것이다. 이것은 간단한 예제이다. 현재의 연구는 더 복잡한 접근방법과 색인 전략 개발에 초점을 두고 있다.

웨어하우스에 대한 적절한 질의 언어를 개발하는 것이 논점이다. 이것은 활용될 데이터 모델에 의존될 것이다. 만일, 모델이 관계형이라면, SQL 기반 언어가 적당할 것이다. 만약 데이터 모델이 객체-지향이라면, ODMG 기반 언어가 적당할 것이다. 웨어하우스에서도 역시 비주얼(visual) 인터페이스가 필요할 것이다.

데이터베이스 관리(database administration) 기법이 웨어하우스 관리를 위해 활용되어야 한다. 웨어하우스 관리자(administrator)가 있어야 하는가? 웨어하우스 관리자와 데이터 원천 관리자 사이의 관련성은 무엇인가? 웨어하우스는 얼마나 자주 감사(audit)되어야 하는가? 웨어하우스는 반드시 감사되어야 하는가? Inmon은 웨어하우스 감사에 대한 좋은 아이디어가 없는 이유를 제시했다^[INMO93]. 또다른 관리 논점은 갱신을 데이터베이스로 전파하는 것이

다. 대부분의 경우 데이터 원천 관리자들은 그들 데이터에 대한 트리거(trigger)를 강요하고 싶지 않을 것이다. 이런 경우 자동적으로 갱신을 전파시키기는 어려울 것이다.

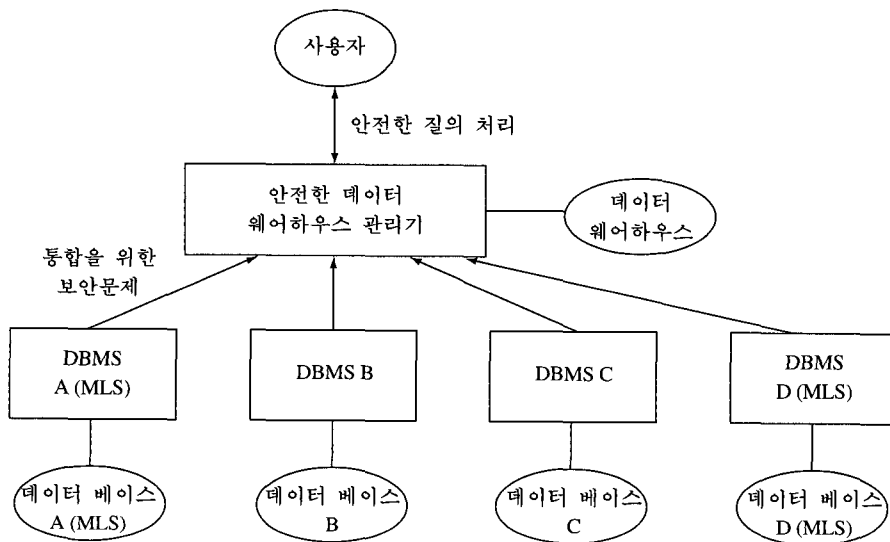
웨어하우스 보호는 주요 논점이다. 이질형 데이터베이스 시스템의 통합에서의 보안성 논점이 여기서 역시 검토될 필요가 있다. 또한, 통계 데이터베이스 보안 역시 중요한 역할을 한다. 웨어하우스 유지보수 시에도 역시 보안 제어(security control)가 시행되어야 한다. 이것은 질의, 메타데이터 관리, 그리고 웨어하우스 갱신에 영향을 줄 것이다. 뿐만 아니라, 다단계 보안(multilevel security)이 필요하다면, 부가적인 고려사항이 있다. 예를 들면, 웨어하우스에서 신뢰적인(trusted) 구성요소는 무엇인가?

병렬(parallel) 데이터베이스 관리를 포함한 고성능 컴퓨팅(high performance computing)은 데이터 웨어하우스에서 중요한 역할을 한다. 목적은 사용자가 복잡한 질의에 대하여 신속하게 해답을 얻는 것이다. 그러므로, 병렬 질

의 처리 전략이 웨어하우스에서도 인기가 되고 있다. 효율적인 질의 처리를 위해서 적절한 하드웨어와 소프트웨어가 필요하다.

메타데이터 관리(metadata management)는 데이터 웨어하우스에서 또다른 중요 기술이다. 문제는 메타데이터를 정의하는 것이다. 메타데이터는 데이터 원천들로부터 나올 수 있다. 메타데이터에는 데이터 원천들과 웨어하우스 사이의 사상(mapping)도 포함된다. 웨어하우스에 특정한 메타데이터도 역시 있다.

분산(distributed) 데이터베이스 기술은 데이터 웨어하우스에서도 역할을 한다. 웨어하우스를 중앙집중으로 할 것인가 혹은 분산시킬 것인가? 만일 분산 시킨다면, 분산 데이터베이스 관리에 대한 대부분의 기술들이 데이터 웨어하우스에 적용된다. 비 분산의 경우에는, 은행에서 처럼, 여러 지점(branch)에 대한 하나의 중앙 웨어하우스가 있다. 분산 웨어하우스 경우에는, 각 지점들이 자신의 지역(local) 웨어하우스를 가지며 웨어하우스들은 서로서로 통신을 한다고 가정할 수 있다.



[그림 4] 데이터 웨어하우스에서 보안

3. 데이터 웨어하우스에서 보안성

이 섹션에서는 [THUR96, THUR97b]에 나와 있는 데이터 웨어하우스에서의 보안 문제들을 논의할 것이다. 웨어하우스의 형성을 위한 데이터베이스들의 통합과 웨어하우스로부터 정보 검색에서 보안문제는 주된 고려사항이다. 이것은 그림 4에서 설명하고 있다. 이 예제에서 네 개의 DBMS 중에서 두 개는 다단계 보안(MLS) DBMS이다.

이질형 데이터베이스 통합에서의 보안성 연구는 웨어하우스 구성에서의 보안성을 성취하는데 중요한 기여를 할 것이다. 예를 들면, 웨어하우스를 구성하기 위해 많은 이질형 데이터베이스들을 통합할 때, 우리는 많은 보안 정책들을 취급해야만 할 것이다. 여기서 주된 문제는 불일치 정책들을 처리하는 것이다. 우리는 다양한 충돌(conflicts)을 해결할 필요가 있으며 또한, 웨어하우스를 위한 적당한 보안 정책을 만들 필요가 있다. 연합(federated) 데이터베이스 관리에서의 보안성에 대한 연구가 [BLAU95]에서 보고되었다. 우리는 웨어하우스에서의 보안 정책을 개발함에 있어서 그런 연구를 검토할 필요가 있다. 다른 문제는 다음과 같은 사항에서의 보안 영향(impact)을 포함한다. ① 웨어하우스에 대한 데이터 모델, ② 개개의 데이터베이스에서 나온 갱신으로부터 웨어하우스에 대한 적절한 갱신 요청을 생성하는 것, ③ 웨어하우스에 대한 메타데이터의 개발, 그리고 ④ 개개의 데이터 원천과 웨어하우스 간의 사상과 변환에서의 보안성을 유지하는 것 등이다. 역시, 개개의 데이터베이스 관리 시스템들 중에서 몇 개가 그림 4에 나타난 것처럼 다단계 보안이라면, 웨어하우스에 대한 신뢰 컴퓨팅 베이스(trusted computing base)와 같은 어떤 부가적인 보안 관련 사항이 있어야 한다.

웨어하우스 구성에 있어서 많은 다른 중요

한 보안 고려사항들이 있다. 이것은 다음과 같은 사실에 기인된다. 즉, 이질형 데이터베이스를 통합할 때는 데이터 저장소(repository)의 개발을 가정하지 않는 반면에, 웨어하우스의 경우에는 보통 물리적인 데이터 저장소가 있다. 보안에 관계된 예제는 다음과 같은 것이다. 웨어하우스 데이터베이스는 요약 정보를 줄 수 있다. 이런 요약 정보는 이질형 데이터베이스들에 있는 데이터에서 종종 유도된다. 웨어하우스에 있는 요약 정보에서 이질형 데이터베이스들에 있는 기밀 정보를 연역하지 못하게 하는 것은 중요하다. 그러므로, 통계 데이터베이스 보안 뿐만 아니라 추론 및 집산화(aggregation) 문제에서의 발전은 웨어하우스 보안에 중요한 역할을 할 것이다.

이전에 논의된 것은 이질형 데이터 원천으로부터 웨어하우스를 구성함에 있어서의 보안 문제에 초점을 두었다. 웨어하우스가 운영되는 동안, 보안성은 지속적으로 유지되어야 한다. 예를 들면, 웨어하우스에서의 행동(action)들은 감사될 필요가 있다. 의문은 기존의 데이터베이스 감사 기법이 웨어하우스에서도 사용가능한 것인가이다. 다른 문제는 다음과 같은 것이다. 특정한 웨어하우스 관리자나 웨어하우스 보안 요원이 있어야 하는가? 웨어하우스 관리자/보안 요원과, 웨어하우스를 개발하는데 사용된 이질형 데이터베이스의 관리자/보안 요원 사이의 관련성은 무엇인가? 적절한 질의 수정 기법이 웨어하우스에서 개발 가능한가? 질의 의존적인 웨어하우스를 구성할 때, 웨어하우스에서 시행될 접근 제어 규칙들을 고려해야 하는가? 접근 방법이나 색인 전략에서의 보안 영향은 무엇인가? 웨어하우스에 대한 보호 메카니즘으로서 뷰가 사용될 수 있는가? 이들 문제에 대한 연구가 진행되고 있다. 예를 들면, 스탠퍼드 대학의 데이터 웨어하우스 프로젝트 [ZHT95]에서는 웨어하우스에서 구체화(materialized) 뷰의 기법을 연구하며 또한, 데

이터 원천이 갱신될 때 뷰를 유지하는 기법도 연구하고 있다. 웨어하우스 환경에서 뷰를 통한 보안 실행은 더 연구가 필요하다.

데이터 웨어하우스를 관리하는 것은 어떤 문제점이 있다. 예를 들면, 웨어하우스를 구성할 때, 많은 이질형 데이터 원천들을 통합하는 것을 고려해 보자. 데이터베이스들이 갱신될 때, 웨어하우스가 자동적으로 갱신되는 방법으로 실행될 트리거(trigger) 메카니즘이 필요할 것이다. 그러나 많은 경우에 있어서 자치성 때문에, 각각의 데이터베이스 관리자들은 그들 데이터에 대해서 트리거가 실행되지 않기를 바랄 것이다. 이것이 웨어하우스를 갱신하는데 주된 문제점이다. 감사에 대해서도, 웨어하우스를 감사하는 것이 항상 좋은 아이디어가 아니다고 Inmon^[INM093]은 서술하고 있다. 예를 들면, 웨어하우스가 감사되어야만 한다면, 어떤 관련되지 않는 데이터가 웨어하우스에 저장될 수도 있다.

여기서 논의된 많은 문제들이 보여주는 것은 보면, 데이터 웨어하우스에서의 보안은 데이터베이스 관리 시스템, 통계 데이터베이스 그리고 이질형 데이터베이스의 통합 등에서의 보안성에 대한 결함이다. 웨어하우스를 안전하게 하는 해결책이 개발되기 전에 더 많은 연구를 해서 웨어하우스에 특정한 보안 문제들을 결정할 필요가 있다.

4. 결 론

이 연구에서는 데이터 웨어하우징에 대한 총괄을 알아보고, 몇몇 보안 논점들을 서술하였다. 데이터 웨어하우징에서 웨어하우스를 형성하기 위한 데이터베이스 통합에서의 보안성, 웨어하우스 질의에서의 보안성을 논의하였다.

이 연구에서 논의된 데이터 웨어하우징에서의 보안 논점들은 매우 초보적인 조사이다. 그러나, 연구에서는 운영 데이터 웨어하우스에서

설명되어야만 하는 몇몇 보안 해결사항을 확인하였다. 이 연구에서는 문제점을 단지 확인만 하였다. 의미있는 해결책이 개발되기 전에 많은 연구가 필요하다.

참 고 문 헌

- [BLAU95] B. Blaustein, C. McCollum, A. Rosenthal, K. Smith, and L. Notargiacomo, "Autonomy and Confidentiality : Secure Federated Data Management", Second International Conference on Next Generation Information Technologies and Systems, Naharia, Israel, June 1995.
- [INMO93] W. H. Inmon, "Building the Data Warehouse", John Wiley and Sons, 1993.
- [THUR96] B. Thuraisingham, "Data Warehousing, Data Mining, and Security", Published in the Proceedings of the 10th IFIP Working Conference in Database Security 1996; also published in the Proceedings of the 19th National Information Systems Security Conference, 1996.
- [THUR97a] B. Thuraisingham, Data Management Systems Evolution and Interoperation, CRC Press, 1997.
- [THUR97b] B. Thuraisingham, "Security Issues for Data Warehousing and Data Mining", Database Security Volume X : Status and Prospects, Chapman & Hall, 1997.
- [ZHUG95] Y. Zhuge, H. Garcia-Molina, J. Hammer, and J. Widom, "View Maintenance in a Warehousing

Environment", Proceedings of the
ACM SIGMOD Conference, San
Jose, CA, May 1995.

□ 著者紹介



심 감 식

1985년 2월 전남대학교 계산통계학과(학사)

1987년 2월 전남대학교 대학원 계산통계학과(석사)

1993년 8월 전남대학교 대학원 전산통계학과(박사)

1993년 11월 ~ 현재 진주산업대학교 교양과정부 조교수

※ 주관심분야 : database security, data warehousing and mining