

국제물류 통합 플랫폼 서비스를 위한 데이터웨어하우스 스키마 설계⁺

이선영, 이종연
충북대학교 컴퓨터교육과
e-mail:{elesun97, jongyun}@chungbuk.ac.kr

Design of Data Warehouse Schema for International Logistics Platform Services

Sun Young Lee, Jong Yun Lee
Dept of Computer Education, Chungbuk National University

요 약

국제물류 EPCglobal Network 프레임워크에서 대용량 운송 데이터베이스의 누적과 경영 전략의 의사결정에 필요한 지식 탐사를 위해 데이터 웨어하우스 시스템 도입이 요구된다. 따라서 본 연구에서는 국제물류 통합 플랫폼을 지원하는 데이터 웨어하우스를 설계하기 위해 현재 데이터베이스 시스템을 분석하고 다차원 데이터 모델인 사실 별자리 스키마 구조를 이용한다. 또한 설계된 데이터웨어하우스를 통해 의사결정을 지원할 수 있는 지식탐사를 위한 질의들의 예를 보여준다.

1. 서론

데이터 웨어하우스는 다중 소스(데이터베이스)로부터 하나의 통합 스키마 구조로 모여진 정보들의 저장소이며, 보통 한 장소에 위치한다. W.H. Inmon에 따르면 “데이터웨어하우스란 경영 의사결정 과정을 지원하기 위한 주제 지향적, 통합적, 시계열적, 비휘발성 등의 특성을 갖는 데이터 집합”으로 정의하였다[1]. 데이터베이스 시스템의 발달과 함께 기업의 실세계로부터 많은 생산 및 생산기술 데이터베이스, 영업 및 영업지원 데이터베이스, 설계 데이터베이스 등 많은 데이터가 누적되어 왔다. 누적된 데이터베이스로부터 시간 변화에 따른 집계 데이터 축적과 정보(information) 또는 지식 탐사(knowledge discovery)라는 새로운 시도가 시작되었다. 전자를 누적된 대용량 데이터베이스로부터 시간 이력에 따라 집계 데이터(aggregate data)를 축적하는 데이터 웨어하우스라 하고, 후자를 축적된 데이터 웨어하우스로부터 새로운 지식을 탐사하는 데이터 마이닝(data mining)이라 한다[2]. 데이터 웨어하우스는 다양한 이질형 DBMS의 데이터베이스와 일반 파일 형태에서 데이터 전처리과정을 통해 구축된다.

1.1 연구동기

데이터웨어하우스는 현재 상태의 데이터뿐만 아니라 과거의 데이터도 시간과 함께 유지하므로, 누적된 통합 데이터를 분석하여 필요한 정보를 추출하는데 효과적이며 의사결정뿐만 아니라 운영업무도 지원할 수 있다. 국제물류 EPCglobal Network 프레임워크에서도 물류 흐름의 전과

정에 이르는 대용량 운송 데이터베이스의 누적과 더불어 경영 전략의 의사결정에 필요한 지식 탐사를 위한 데이터 웨어하우스 시스템 도입이 요구된다.

데이터 웨어하우스를 위한 가장 일반적인 데이터 모델은 다차원 데이터 모델이다. 기존의 ER 모델링은 주로 OLTP 시스템에서 빈번한 데이터의 변화에 대한 전체 데이터의 무결성(integrity)에 중점을 두고 있다. 그러나 다차원 모델링은 사용자의 다양한 요구사항을 충족시키는 복잡한 질의를 신속하게 처리하는데 목적을 두고 있다[3, 4, 5]. 차원 모형은 사용자 중심으로 중요한 데이터를 구별하여 하나의 테이블에 모아 놓기 때문에 개체-관계 모델과는 달리 스키마에서 테이블간의 중요도가 다르다. 따라서 데이터 웨어하우스 설계는 온라인 데이터 분석을 실현하기 위한 더욱 정확하고 주제 중심적(subject-oriented) 스키마 구조를 요구한다[2].

1.2 기여도

따라서 본 연구는 국제물류 EPCglobal Network에서 지능형 국제물류 통합 플랫폼 서비스를 위한 데이터웨어하우스를 설계한다. 본 논문의 연구 내용은 다음과 같다. 첫째, 새로운 데이터 웨어하우스 스키마 설계를 위해 현재 국제물류 통합 데이터베이스 스키마를 분석한다. 둘째, 데이터베이스 분석을 기반으로 EPCglobal Network를 위한 데이터 웨어하우스 스키마를 설계한다.

본 논문의 구성은 다음과 같다. 2장에서는 데이터웨어하우스에 대하여 살펴보고, 3장에서는 현재 국제물류 통합 플랫폼의 데이터베이스 시스템을 분석한다. 4장에서 데이터웨어하우스 스키마 구조를 제시하고, 5장에서 지식 탐사

⁺본 논문은 2008년도 지식경제부 성장동력기술개발 사업의 일환으로 수행되었음.

를 위한 몇 가지 질의를 보여준다. 마지막으로 6장에서 결론을 맺는다.

2. 데이터 웨어하우스 스키마 구조

데이터 웨어하우스는 다차원 형태로 데이터를 일반화하고 응축한 형태로 기업의 전략적 의사결정에 필요한 데이터를 저장하며, 의사결정 지원 모델의 물리적 구현 형태로 지원되는 데이터 저장소이다. 데이터 웨어하우스를 위한 가장 일반적인 데이터 모델은 다차원 데이터 모델이다. 다차원 모델링은 사용자의 다양한 요구사항을 충족시키는 복잡한 질의를 신속하게 처리하는데 목적을 두고 있다[3, 4, 5]. 차원 모형은 사용자 중심으로 중요한 데이터를 구별하여 하나의 테이블에 모아 놓기 때문에 개체-관계 모델과는 달리 스키마에서 테이블간의 중요도가 다르다. 사용자의 관점에서 중요한 데이터를 포함하는 테이블을 스키마의 중심에 놓고, 이들을 분석하기 위한 관점을 나타내는 테이블을 주변에 배치한다. 다차원 데이터 모델의 종류는 별형 스키마(star schema), 눈송이형 스키마(snowflake schema), 사실 별자리형 스키마(fact constellation schema)가 있다. 다음은 이 세 가지 스키마 구조를 검토한다.

2.1 별형 스키마

별형 스키마는 다차원 데이터 웨어하우스 스키마 구조를 표현하는 가장 일반적인 데이터 모델로서 중앙에 사실 테이블(fact table)을 위치시키고, 그 주변에 사실 테이블의 속성으로 참조되는 차원 테이블(dimension table)로 구성된다[6]. 차원이란 조직이 데이터 레코드를 운용하는 이유의 대상이 되는 측면을 의미하고, 사실(Fact)이란 숫자로 표현되는 값(measure)을 의미한다. 사실 테이블은 사실 데이터를 표현하는 것으로 필요한 사실(측정치)과 관심 있는 차원 속성들, 그에 따른 부가적인 속성들로 이루어진다. 차원 테이블은 사실 테이블의 각 차원 속성에 대한 상세 정보이며, 사용자가 데이터를 분석하고자 하는 주요 분석 요인을 충족시키기 위한 정보를 유지 한다. 사실 테이블과 차원 테이블은 스키마 다이어그램에서 외부키(foreign key)를 통해 서로 연결된다. 예로, 수출입 데이터베이스에 대한 별형 스키마는 그림 1과 같다. 수출입량은 5 차원 “product type, subscribe, time, to_location, from_location”에 의해 표현된다. 스키마 구조는 수출입 데이터베이스의 5 차원 각각의 키 값과 측정치 “수출입량”을 포함하는 사실 테이블과 주변에 각 차원의 상세 정보를 나타내는 차원 테이블들로 구성된다. 사실 테이블은 차원 테이블의 키 값을 속성으로 하고, 이와 관계되는 수출입량의 측정치를 속성으로 표현된다.

2.2 눈송이형 스키마

별형 스키마의 변형인 눈송이형 스키마는 차원 테이블을 재 정규화(normalization)함으로써 테이블들이 마치 눈꽃(snowflake) 모양이 되도록 만든 스키마이다.

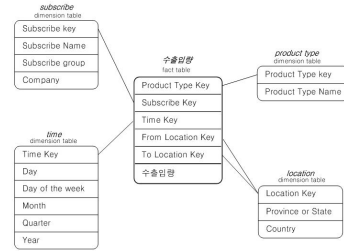


그림 1. 별형 스키마

결국 이전의 별형 스키마에서 차원 테이블을 정규화하여 데이터 중복을 제거한 형태가 된다. 이러한 눈송이형 스키마는 유지관리와 저장 공간 절약이 가능한 반면, 스키마가 복잡해져 스키마 브라우징 효과성이 떨어지므로, 데이터웨어하우스 설계 시 별형 스키마처럼 대중적이지 못한 단점이 있다. 데이터웨어하우스에서 차원 테이블의 크기는 문제가 되지 않고 많은 수의 조인이 문제가 되기 때문에 눈송이형 스키마를 사용하면 별형 스키마보다 응답 시간이 늦어진다.

2.3 사실 별자리형 스키마

복잡한 애플리케이션 시스템은 차원 테이블을 공유하는 다수의 사실 테이블을 요구할 수 있다. 이러한 경우 별형 스키마 집합으로 볼 수 있으며, 은하수 스키마(galaxy schema) 또는 사실 별자리형 스키마(fact constellation schema)라 부른다. 일반적으로 데이터 웨어하우스 설계 시 여러 개의 사실 테이블을 포함할 수 있으므로 사실 별자리형 스키마 구조를 갖는다고 할 수 있다. 예로, 영업(sales)과 운송(shipping) 데이터 웨어하우스에 대한 사실 별자리형 스키마는 영업 실적에 대한 사실과 운송 내역에 대한 사실의 두 개의 테이블로 표현된다. 본 논문에서도 Lead time, 수출입량, 처리 건수에 대한 사실 테이블을 포함하는 사실 별자리형 스키마를 설계한다.

3. 국제물류 플랫폼의 기존 데이터베이스 시스템

지능형 국제물류 플랫폼의 효율적인 서비스를 위한 데이터 웨어하우스를 구축하기 전에 현재 데이터베이스 시스템 현황을 살펴보고, 본 연구의 목표인 국제물류 플랫폼 데이터웨어하우스 구축 모델을 제시한다. RFID 기반 국제물류 통합 플랫폼은 국제물류(수입 및 수출)구간의 화물 흐름에 대한 RFID기반의 Visibility와 Traceability 서비스를 개발하여 국가전자무역 인프라와 연동함으로써 실시간 Visibility를 제공하고, 플랫폼에 축적된 RFID기반의 화물 정보 및 B2G 문서 정보를 제조(무역) 및 물류기업 등 참여 주체의 ERP(Enterprise Resource Planing) 시스템에 제공하여 생산/물류 정보관리를 위한 RFID 기반 글로벌 SCM(Supply Chain Management) 환경을 구축한다. EPCglobal Network의 시스템은 표1과 같다. 본 연구에서는 데이터웨어하우스 설계를 위해 국제 물류 통합 플랫폼

의 시스템들 중 Visibility 시스템과 RFID 기반 국제물류 정보 통합 관리 시스템 및 플랫폼 내의 데이터베이스를 분석하였다.

RFID 기반 국제 물류 플랫폼	
	국제 물류 Visibility 시스템
	인터페이스 연계 기술
	국제 물류 정보 변환 시스템
	RFID 기반 국제물류 정보 통합관리 시스템
국제물류 다기능/지능형 미들웨어	
	ALE 기반 대용량/실시간 RFID 이벤트 관리 시스템
	국제물류 EPC 네트워크 정보 서비스 시스템
	다기능/지능형 미들웨어 관리 모니터링 시스템
기업 및 관세사용 솔루션	

표 1 EPCglobal Network의 시스템

국제 물류 Visibility 시스템은 국제물류 흐름상의 전체 제조 흐름의 가시화와 추적을 제공하며, 이를 위해 EPC Network와 기업의 Legacy System, 기간망 시스템이 상호 연계되어 서비스를 제공한다. 국제물류 Visibility 시스템에서 제공하는 통계 서비스는 기간/Location별 RFID 발생이력 추이와 이력등록 실적 및 Lead Time 분석, 기간/BizStep별 이력등록 실적, e-Seal 이력등록 실적, 기간/문서별 등록 실적이 있다. RFID 기반 국제물류 정보 통합 관리 시스템은 각 거점에서 발생하는 다양하고 방대한 국제물류정보를 RFID 기반의 통합 데이터베이스로 구축하여 관리하는 시스템이다. 물류거점단계에서 발생하는 RFID 물류정보를 변화·축적·가공·연계·제공 관리한다. 국제물류 EPC 정보 서비스 시스템은 RFID 화물 정보 이벤트에 대한 표준 인터페이스를 제공하고 국제물류 EPCIS 이벤트 정보 캡처/쿼리 및 저장 관리를 지원하는 EPC 정보 서비스 시스템이다. 데이터베이스(DB)들은 국제 물류 플랫폼 통합 DB, 다기능/지능형 미들웨어 DB, 솔루션 DB의 세 부분으로 나누어 볼 수 있으며, 참여 업체들의 Legacy DB도 포함된다. 국제물류 통합 DB는 물류거점단계에서 발생하는 RFID 물류정보를 변화·축적·가공·연계·제공 관리한다. RFID 기반 데이터들이 들어오면 해당하는 물류 정보를 솔루션 DB에서 가져와 사용한다. 현재 시스템은 국제물류 통합 플랫폼의 국제물류 흐름에서 생기는 정보들을 이용해 통합 관리하고 있으며 미들웨어에 EPCIS 데이터 정보가 저장된다. 거점별 RFID 실시간 정보들이 미들웨어에 전달되어 저장되고 이것들은 각 솔루션 데이터베이스를 통해 국제물류 정보 통합 DB에 저장된다.

4. 데이터 웨어하우스 스키마 설계

EPCglobal Network를 위한 데이터 웨어하우스는 일반 데이터 웨어하우스와는 달리 최고 경영자 관점이 아닌 각 거점별 사용자(관리자) 관점에서의 필요한 ‘인전 지역의 보세 창고의 리드타임(Lead time)은 얼마인가?’, ‘지역별

리드타임은 얼마인가?’, ‘가장 짧은 리드타임을 갖는 지역은 어디인가?’와 같은 질문에 정보를 제공할 수 있어야 한다. 차원모형을 설계하는 기법은 일정하게 정해진 것은 아니지만 개발하고자 하는 시스템의 목적에 맞게 본 논문에서는 사실 별자리 차원모형(Fact constellation schema)을 이용하여 설계 한다.

국제물류 EPCglobal Network의 데이터 웨어하우스에서 전체적인 Lead time, 비즈니스 프로세스 구간별 Lead time, 비즈니스 프로세스 구간별 event 수, 전체적인 수출입 실적, 연도별·월별 수출입 총액, 지역별 수출입 실적, 국가별 무역현황, 지역(통관)별 수출실적, 국가별 품목별 수출입실적, 성질별 수출실적, 품목별 국가별 수출입실적, 품목별 수출입업체와 같은 통계자료를 생성하기 위해서 앞 절에서 살펴본 국제물류 통합 데이터를 중심으로 데이터 스키마를 생성하였다. 즉, EPCglobal Network를 위한 데이터 웨어하우스에서 Lead time, 이벤트 수, 수출입 실적에 대한 정보를 제공하는 것에 중점을 두었다. 리드타임(Lead time)이란 일반적으로 최종 소비자에게 제품이 전달되기까지의 각 설정 범위간 소요시간으로, 본 연구에서는 lead time을 물류의 이동 및 처리 시간으로 정의한다. 예로 창고 내의 리드타임, 운송 구간별 리드타임 등으로 정의 될 수 있다. 이벤트 수는 각 거점별로 물류의 이동에 따른 이벤트를 나타낸다. 이 두 측정값은 비즈니스 프로세스 전체 또는 구간별, 지역별, 구독자별, 품목별, 시간별로 정보를 얻고자 한다. 지역 차원은 국가, 도(주), 시 등을 나타내고 BizLocation은 물류 프로세스의 거점을 나타낸다. Subscribe 차원은 구독자로서 EPCIS를 가지고 있는 사용자이다.

Lead time, 이벤트 수, 수출입량에 대한 세 개의 사실 테이블(Fact table)을 갖는 사실 별자리형 스키마는 그림 2와 같다. 첫 번째 lead time의 측정값을 갖는 사실 테이블에는 시간(time) 차원, 제품(Product) 차원, 지역(Location) 차원, 구독자(Subscribe) 차원, BizLocation 차원과 연결고리를 가지는 기본 코드를 포함한다. 두 번째 event count의 측정값을 갖는 사실 테이블에는 시간 차원, 제품 유형(Product Type) 차원, 지역 차원, 구독자 차원, BizLocation 차원과 연결고리를 가지는 기본 코드를 포함한다. 세 번째 수출입량 사실 테이블은 시간 차원, 제품 유형 차원, 지역 차원(도착지, 선착지), 구독자 차원과 연결고리를 가지는 기본 코드를 포함한다. 차원테이블은 사용자가 분석하고자 하는 기준이 되는 것으로 기간(일, 주, 월, 분기, 년도), 제품(품명, 품종), 지역(국가, 도시, 항구), BizLocation(보세운송사, 선사, 포워드, 화주), 구독자 차원을 갖는다. 제품의 경우 lead time은 제품 및 제품 유형별로 분석이 필요하지만 이벤트 수와 수출입량의 경우는 제품 유형별로 분석이 필요하다. 따라서 시간, 제품, 제품 유형, 지역, BizLocation, 구독자의 차원을 기본적으로 설정하고 구독자의 부가적인 정보를 구하기 위해 회사(Company)테이블을 추가하였다.

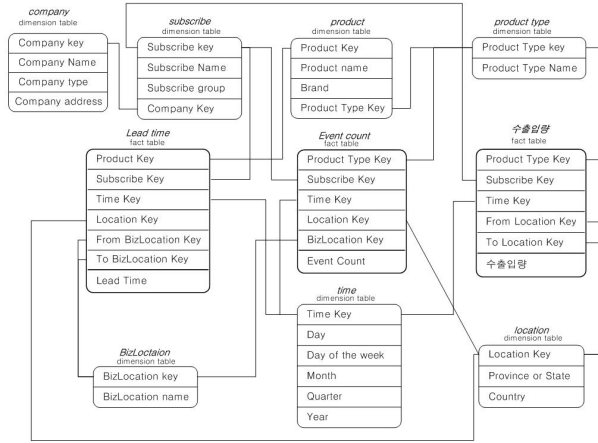


그림 2. EPCglobal Network를 위한 사실 별자리형 데이터 웨어하우스의 스키마 설계

차원의 속성을 결정하는 것은 검색한 결과를 사용자에게 전달하는 방법의 편의성을 증진시키는 가장 큰 요소가 된다. 각 차원의 속성은 그림 2의 각 차원테이블에 나타났다. 지역 차원의 지역키는 수출입량 사실테이블과의 연결에서 출발지와 도착지의 정보를 나타낼 수 있도록 From Location Key와 To Location Key로 연결된다. 리드 타임 사실 테이블에서 BizLocation별 리드타임 뿐만 아니라 구간별 리드 타임값을 분석하기 위해 From BizLocation key와 To BizLocation key를 BizLocation 차원의 연결고리로 두고 있다.

제품의 생산에서 배송, 소비자까지 물류 흐름의 전 과정에 이르는 대용량 운송 데이터베이스의 누적과 더불어 경영 전략의 의사결정에 필요한 지식 탐사를 위한 데이터 웨어하우스 시스템이 요구된다. 따라서 본 연구에서는 국제 물류 통합 플랫폼의 현재 데이터베이스 시스템을 분석하고 의사 결정을 지원하는 지식탐사에 알맞은 데이터웨어하우스 구축을 위한 스키마 구조를 설계하였다. 또한 설계된 데이터웨어하우스를 통해 의사결정을 지원할 수 있는 질의들의 예를 보여주었다. 국제물류 EPCglobal Network의 설계된 데이터 웨어하우스를 중심으로 lead time 조회, event 수 조회, 수출입량 조회가 가능할 것이다.

5. 의사결정 질의의 SQL 표현

설계된 데이터 웨어하우스를 바탕으로 의사결정을 위한 필요한 지식을 얻고자 할 때 사용할 수 있는 SQL 문장을 살펴본다. 구축된 국제물류 EPCglobal Network의 주요 지식 탐사내용은 설계된 데이터 웨어하우스를 중심으로 전체적인 lead time 조회, 비즈니스 프로세스 구간별 lead time 조회, 비즈니스 프로세스 구간별 event 수 조회, 전체적인 수출입 실적 조회 등이 가능하다.

(1) 월별 BizLocation별 event 수를 구하는 SQL문장

```
select month, BizLocation_name, sum(event count)
from event_count_fact_table a1, time_dim_table a2,
BizLocation_dim_table a3,
where a1.time_key=a2.time_key
and a1.BizLocation_key=a3.BizLocation_key
group by a2.month, a3.BizLocation_name;
```

(2) Location별 lead time을 구하는 SQL문장

```
select location, sum(Lead time)
from Leadtime_fact_table a1, location_dim_table a2
where a1.location_key = a2.location_key
group by a2.location
```

6. 결 론

현재 국제물류 EPCglobal Network 프레임워크에서도

참고문헌

- [1] W. H. Inmon, *Building the Data Warehouse*, JohnWiley & Sons, 1996.
- [2] Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, second edition, Morgan Kaufmann, 2006.
- [3] 장동인, *(실무자를 위한)데이터 웨어하우스*, 대청 미디어, 1999.
- [4] 김락성, “데이터 웨어하우스를 이용한 기업 의사결정지원시스템의 설계”, *産業經營研究*, Vol.25 No.1 2002.
- [5] P. Tan, M. Steinbach and V. Kumar, *Data Mining*, Addison Wesley, 2006.
- [6] M. Levene and G. Loizou, “Why is the snowflake schema a good data warehouse design?” *Information Systems* 28 (3), pp. 225-240, 2003.
- [7] 이현창, 김경창, 지원철, “지역 스키마 통합에 의한 효율적 데이터 웨어하우스 설계”, *한국전문가시스템학회*, pp. 215-222, Vol.2 No.1 1997.
- [8] 강현철 외 6인, *데이터 마이닝 방법론*, 자유아카데미, 2006.
- [9] S. Naouali and R. Missaoui, “Flexible query answering in data cubes,” In *DaWaK*, pages 221-232, 2005.