



## OLAP 테크놀로지의 이해

광운대학교 조재희

### 1. OLAP 개요

OLAP은 “최종 사용자가 대규모 기업 데이터에 직접 접근해 능동적으로 정보를 분석하는 과정”이라고 정의할 수 있다. OLAP에 관한 정의 중 FASMI (Fast Analysis of Shared Multidimensional Information)라는 정의도 익히 알려져 있는데, 이것은 글자 그대로 “공유되는 다차원 정보를 신속히 분석할 수 있게 지원하는 전산 환경”을 의미한다[8]. OLAP 시스템의 목적은 사용자가 기업의 전반적인 상황을 이해할 수 있게 하고, 의사결정을 지원하는데 있다.

OLAP이란 용어는 1993년 E.F. Codd에 의해 처음 사용되었다. 관계형 데이터베이스의 아버지라고 할 수 있는 Codd는 자신이 1970년대 말에 만든 관계형 모델은 OLTP에는 적합하지만, OLAP에는 적합하지 못하므로 새로운 형태의 모델이 필요하다고 주장하였다[3]. 이 주장은 그 동안 방대한 양의 데이터를 축적해 놓은 기업들에게 상당한 호응을 얻었으며, 이 축적된 데이터를 어떻게 효과적으로 활용할 것인가로 관심을 돌리게 되는 계기를 마련했다. 이러한 경향은 기업간 경쟁이 치열해지고 세계화되면서 기업 데이터가 매우 중요한 전략적 자산이며, 이의 효과적인 활용은 기업 경쟁 우위를 향상시킬 수 있다는 인식이 확산되면서 더욱 가속화 되었다[표 1 참조].

OLAP은 90년대 중반부터 데이터웨어하우스라는 개념과 결합되어 기업의 전사적 아키텍처 상에 자리 잡게 되었다. 이로 인해 OLAP의 핵심 개념인 다차원 정보 분석이 협업 사용자들에게 깊이 인식되는 계기가 되었고, 데이터웨어하우스 데이터를 분석하는 근본적인 방법론으로 자리매김하게 되었다. 최근에는 CRM 프로젝트와 SCM 프로젝트 등이 활발하게 진행되면서 OLAP과 기타 비즈니스 인텔리전스 솔루션에 대한 업계의 관심이 증가되고 있는 추세이다.

한편 OLAP은 Balanced Scorecard, Business Performance Management, Risk Assessment, Early Warning System 등과 같은 비즈니스 애플리케이션에 반드시 포함되어야 하는 기능으로 인식되고 있다.

표 1 OLTP와 OLAP 비교

	OLTP	OLAP
정의	거래처리 프로세싱	분석처리 프로세싱
목적	비즈니스 운영	비지니스 선도
성격	단위업무 처리	기능업무 처리
저장내용	갱신된 현재값	요약, 종합자료
갱신내용	단순한 필드값 갱신	전체적인 계산/ 요약
사용시 자료량	적다	많다
데이터 구조	복잡	단순
사용법	단순 반복적	비정형적, 분석적
응답시간	즉시	분석시간 소요

### 2. OLAP 연구

경영 정보 분야에서의 OLAP 연구는 이론적이기보다는 실무 지향적인 연구이다. 이 연구는 80년대 초에 나온 의사결정 지원 시스템(DSS)의 확산 실패를 거울 삼아 “새로운 전산/정보 기술로 정보를 분석하여 기업 내에 배분하려는 전산화 노력”, 또는 “분석 프로세싱을 조직 내에 활성화시키기 위한 노력”이라고 해석할 수 있다.

#### 2.1 다차원 모델과 다차원 DB

OLAP 관련 연구는 가장 핵심적인 기능이라고 할 수 있는 다차원 모델 개발에 관한 연구로부터 출발했

다. 다차원 모델은 1983년 미국 컴쉐어사의 Crandall 박사가 최초로 만들었는데, 그는 데이터, 업무 구조, 그리고 업무 규칙을 함께 저장하는 아이디어를 ‘시스템 W’라는 메인프레임 기반 소프트웨어로 구현하였다[4].

또한 이 분야에 큰 축적을 남긴 연구는 아버소프트 웨어사가 1994년에 연구 개발한 에스베이스라는 다차원 데이터베이스이다[2]. 이 회사는, 지금은 하이페리온 사에 합병 되었지만, 당시로서는 유일한 클라이언트 서버용 다차원 데이터베이스를 출시하고 있었다.

다차원 DB는 “DSS 기능이 내장된 데이터베이스” 혹은 “분석 프로세싱을 위하여 재구축된 데이터베이스”로 정의된다. 이 DB는 다양한 측면을 고려해야 하는 경영 환경을 그대로 반영할 수 있게 디자인되어 있으며, 다양적으로 업무를 구조화하는 틀을 제공하고 있다[9]. 다양한 데이터는 물론 업무 구조와 규칙도 함께 내장되어 있으며, 사용자 뷰를 제공하기 위해서 별도의 코딩이 필요없이 애플리케이션이 제공된다. 다년간의 비즈니스 컨설팅 경험에서 필요성이 인식되어 개발·보급되었으며, 논리적인 구조는 다차원적인 입방체로 흔히 설명되지만, 디스크에 저장할 때의 물리적인 구조는 일종의 다차원 배열 구조이다.

다차원 DB는 다소 생소한 용어를 사용하는데, 그것들은 ‘관점/차원(Viewpoint)’, ‘관점에 속하는 항목들(Viewpoint Members)’, ‘변수(Variables)’, 그리고 ‘기간(Periods)’이다. 기업의 업무를 다차원 모델로 분석하면 대체로 사업부, 생산 지역, 제품, 산업, 시장 구분, 판매 경로 등이 차원으로 설정되며, 주요 변수로는 실적과 계획 대비 매출액, 비용, 이익 등과 같은 계정 과목들, 수송량, 재고량, 인원, 성장을 등이 사용된다.

다차원 DB는 데이터간의 관계를 입체화, 또는 시각화 해줌으로써 데이터셋 안에 담긴 패턴에 대한 이해를 돋는다. 또한, 문제의 출처를 파악할 수 있도록 심층 분석 기능 (Drill-down)과 What-if 분석 기능, Goal-seeking 기능 등이 제공된다. 따라서 사용자는 다양한 각도로 문제를 분석할 수 있게 된다.

## 2.2 Codd의 OLAP 기능에 관한 연구

1970년대 말에 12가지 관계형 DB 원칙을 발표했던 Codd 박사는 1993년에 12가지 OLAP 법칙을 발표했다[3]. 그는 관계형 DB가 유연성 있는 데이터 분

석에는 적합치 않다고 인정하며, 최종 사용자들의 주 업무인 정보 분석 프로세스를 위해서 필요한 기능을 선정하였다. 무엇보다도 다차원 뷰를 제공하여야 하며, 클라이언트/서버 환경에서 작동되어야 하며, 직관적인 데이터 조작 기능, 복수 사용자의 동시접속 등을 갖추어야 한다는 것이 핵심이다.

Codd 박사의 이 연구는 특정 소프트웨어 회사의 연구비를 받고 수행되었기 때문에 발표논문 내용에 관해 많은 논란이 있었다. 그러나 OLAP 도입에 관심을 가지고 있는 기업은 이 기능들을 OLAP 서버 개발 툴 선정시 일종의 제품 평가 가이드라인으로 사용하고 있다. 가트너 그룹이나 Pendse 등이 추가적으로 OLAP에 필요한 기능을 선정하였으며, 표 2와 같이 항목을 나눠 분류되기도 한다.

표 2 OLAP 제품 평가기준

기본요소	다차원 관점 제공 직관적인 데이터 조작 정보 접근성 OLAP 데이터의 사전연산 다양한 수준의 OLAP 분석 모델 클라이언트/서버 아키텍쳐 투명성 다중사용자 지원
	비정규화된 데이터 처리 OLAP 결과의 분리저장 널 값의 추출 널 값의 처리
특수요소	질의 유연성 질의 응답성능의 일관성 물리적 레벨의 자동 수정
질의요소	차원간 동질성 제한없는 차원 및 레벨 제한없는 차원간 조작
차원컨트롤	

## 3. OLAP과 다차원 모델

### 3.1 다차원 모델링

기업 정보 시스템에서 데이터베이스의 제1차적인 역할은 각종 업무에서 발생하는 데이터의 저장고 역할이다. 기업 본원 업무인 구매, 생산, 유통, 그리고 판매에서 발생하는 데이터는 물론이고, 지원 업무인 마

케팅, 기획, 전산, 그리고 회계에서 발생하는 데이터의 규모 또한 대단하다. 뿐만 아니라 인터넷이 판매 및 마케팅 채널로 등장함에 따라 새로운 형태의 데이터를 담고 있는 웹 관련 데이터베이스들이 출현하고 있다.

한편 데이터베이스는 데이터 수집 및 저장 역할 뿐만 아니라 정보 분석을 위해서 재사용되고 있다. 이것이 데이터베이스의 제2차적인 역할인데, 정보 분석용 데이터베이스 즉 데이터웨어하우스가 그 대표적인 예이다. 데이터웨어하우스의 기본 사상은 애써 모아둔 데이터베이스를 이제는 경영 전략 및 마케팅 전략에 활용해 보자는 것이다. 이를 위해서는 다차원성을 내포하고 있는 데이터베이스를 그야말로 다차원적으로 분석할 수 있는 형태로 변환시켜야 하는데, 이것이 바로 다차원 모델링이다.

데이터 분석 업무 자체를 구조화하기 위한 모델링 방법으로 2가지 방법이 존재한다. 그것은 다차원 모델링(Multi-dimensional modeling)과 차원 모델링(Dimensional modeling)인데, 전자는 다차원 데이터베이스를 기반으로 하며, 큐브 방식(하이퍼 큐브와 멀티 큐브)으로 표현된다[5,7]. 큐브(입방체)는 축과 좌표로 구성되는데, 이러한 축은 차원으로, 좌표는 차원 항목으로 이해할 수 있다. 차원은 사용자가 정보를 분석하고자 하는 하나의 관점을 나타내며, 데이터에 대한 사용자의 인식을 바탕으로 하여 유사한 일련의 항목들이 차원을 구성한다. 후자는 관계형 DB를 기반으로 하는 모델링 방법으로써 일반적으로 1개의 사실 테이블과 복수의 차원 테이블들이 만들어지며, ‘스타 스키마’와 ‘스노우플레이크 스키마’가 주요 모델링 방법으로 사용된다[6,10].

### 3.2 다차원 모델 구성 요소

OLAP의 핵심은 다차원 정보 분석인데, OLAP 시스템은 다차원 정보 분석을 지원하기 위해 다차원 모델을 구축한다. 다차원 모델은 OLAP 툴 벤더들에 따라 비즈니스 영역(Business Area), 비즈니스 개요(Business Abstract), 유니버스(Universe), DSS 시스템 등과 같은 다양한 용어로 표현한다. 아직까지 다차원 모델의 표준이 정립되어 있지 않기 때문에, 되도록 모든 OLAP 기술의 근간을 이루는 개념들을 중심으로 알아본다.

다차원 모델을 구성하는 가장 기본적인 요소는 차원이며, 차원은 계층 구조와 차원 항목을 가진다. 또

차원 항목에는 다양한 비즈니스 룰을 반영한 관계식이 설정될 수 있다. 각 요소들간의 관계를 정리해 보면 그림 1과 같다.

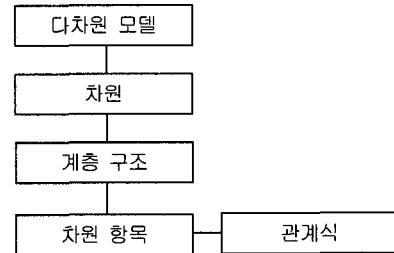


그림 1 다차원 모델의 구성 요소

#### 3.2.1 차원과 차원 항목

차원은 다차원 모델의 기본구조로, 큐브의 구조적 속성(Structural Attribute)을 나타낸다. 차원을 어떻게 정의할 것인가 하는 문제는 OLAP 제품들에 따라 조금씩 차이를 보이고 있다. 차원이 각 제품에 따라 물리적으로 구현되는 방식은 많은 차이가 있다. 하지만 논리적으로 볼 때 차원은 사용자의 관점에서 다른 항목과 독립적인 일련의 모든 항목들의 집합으로 볼 수 있다. 이러한 차원은 전적으로 사용자의 분석 목적에 적합하게 생성되며 사용자의 관점에서 정의된다.

다차원 모델은 일반적으로 큐브(Cube)로 많이 표현되는데, 이 경우 차원(Dimension)은 큐브를 구성하는 축(Axis)으로 생각할 수 있다. 예를 들어 변수, 매장, 제품이라는 3개 차원으로 구성된 큐브를 좀 더 시각적으로 표현해 보면 그림 2와 같이 될 것이다.

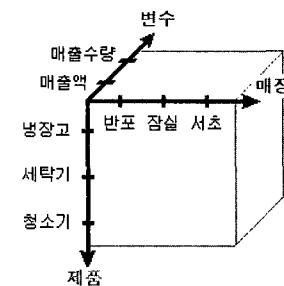


그림 2 차원과 차원 항목

차원이 큐브를 구성하는 축이라고 할 때, 각 축의 좌표에 해당하는 것이 차원 항목(Member 혹은 Element)이라 할 수 있다. 즉 차원은 차원 항목들로

구성된다. 그림 2에 표현된 큐브를 살펴보면 제품 차원은 냉장고, 세탁기, 청소기 항목으로, 매장 차원은 반포, 잠실, 서초 항목으로 구성되어 있다. 또 변수 차원의 항목은 매출액과 매출 수량이다.

차원과 차원 항목은 일차적으로 사용자의 입장에서 분석 목적에 적합하게 설계된다. 차원에 따라서는 차원 항목의 수가 매우 많을 수 있다. 예를 들어, 제품 차원의 경우 수천 개의 항목을 가질 수도 있는데, 창고형 할인매장에서 취급하는 수많은 상품이나 삼성전자에서 생산하는 수많은 가전 제품들을 생각해 보면 쉽게 이해가 갈 것이다.

### 3.2.2 큐브와 셀

각 차원을 구성하는 항목들의 조합에 의해 만들어지는 공간을 셀(Cell)이라 하며, 바로 여기에 데이터 값이 저장된다. 그림 3은 그림 2에 표현된 3차원 큐브에 존재할 수 있는 셀을 시작적으로 나타내고 있다.

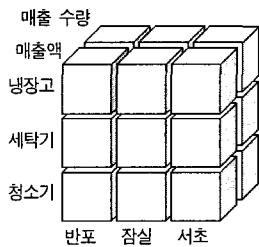


그림 3 큐브와 셀

셀은 큐브를 구성하는 차원들이 가진 항목들의 조합 수만큼 존재한다. 예를 들어, 그림 3의 3차원 큐브는 변수 차원 항목이 2개, 매장 차원 항목이 3개, 제품 차원 항목이 3개이므로  $18(2 \times 3 \times 3)$ 개의 셀이 존재한다. 각 셀은 차원 항목들의 조합에 의해 참조될 수 있다. 여기서 셀은 논리적으로 존재하는 가상의 공간이며, 데이터의 물리적 존재 여부와는 별개의 개념이다.

차원과 차원 항목은 다차원 모델을 구성하는 가장 기본적인 요소라고 할 수 있다. 한편 차원과 차원 항목은 계층 구조와 관계식을 가질 수 있다. 이와 같은 요소로 인해 다차원 모델은 다양한 비지니스 로직을 모델링할 수 있으며 다양한 사용자의 요구를 수용할 수 있다.

### 3.2.3 계층 구조

다차원 모델을 구성하는 대부분의 차원들은 일반적으로 계층 구조(Hierarchy 혹은 Hierarchical Structure)를 가진다. 계층 구조는 차원 항목들 간에 존재할 수 있다. 그림 4는 ‘기간’ 차원을 예로 사용하고 있는데, 이 차원을 구성하는 항목들 간에 존재할 수 있는 계층 구조를 시작적으로 나타내고 있다.

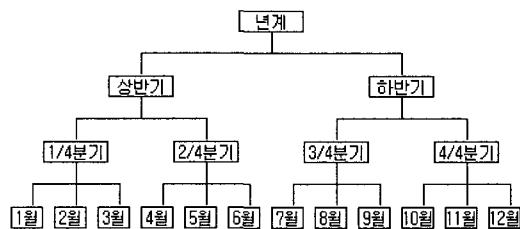


그림 4 차원 항목간 계층 구조

이 예에서 ‘기간’ 차원은 19개 항목으로 구성되어 있다. 이 계층 구조로 인해서 항목들 사이에는 상대적인 위치에 따라 다양한 관계가 설정되며, 이러한 관계는 관계식이나 다차원 질의에서 매우 유용하게 사용된다.

계층 구조는 차원을 구성하는 항목들을 그룹핑하는 기본적인 수단을 제공한다. 어떤 기업이 취급하는 제품 수는 많은 경우 수천 가지에 달할 수 있는데, 이러한 상황에서 사용자는 보다 효과적으로 정보를 조회하고 분석하기 위해 제품들을 특정 기준에 따라 그룹화 하는 것이 일반적이다. 예를 들어, 전체 제품을 가전용품, AV기기, 사무통신기기와 같이 커다란 부문으로 분류하고 가전용품을 다시 냉장고, 세탁기, 청소기 등의 제품 군으로 분류함으로써 보다 효과적으로 정보를 조회할 수 있다. 일반적으로 계층 구조는 데이터가 집계(Aggregation, 혹은 Consolidation)되는 기본적인 경로를 나타내며 분석 목적에 따라 복수의 계층 구조가 존재할 수 있다.

### 3.2.4 관계식

다차원 모델에서 항목들 사이에는 계층 구조와 함께 다양한 관계식이 설정된다. 관계식은 일반적으로 계층 구조만으로 표현하기 힘든 항목들 사이의 관계를 정의하며, 이러한 관계에 의해 입력된 데이터를 바탕으로 새로운 데이터가 생성된다.

표 3 승하차 패턴 분석용 다차원 모델의 차원, 차원 항목, 그리고 관계식

차원	차원 항목수	차원 항목	비즈니스 로직		
측정지표	5	승차건수, 하차건수			
		합계	승차건수 + 하차건수		
		차이	승차건수 - 하차건수		
		승하차율	승차건수/하차건수 x 100		
일	366	1월1일, 1월2일, 1월3일, …, 1월 31일	합계		
		:			
		12월1일, 12월2일, 12월3일, …, 12월31일			
시간대	27	01시, 02시, 03시, …, 11시, 12시	오전계	합계	
		13시, 14시, 15시, …, 23시, 00시	오후계		
승차권 구분	3	MS (승차권)	합계		
		RF (카드식)			
지하철역	154	방화, 개화산, 김포공항, 송정, …, 고덕, 마천	5호선계	합계	
		봉화산, 화랑대, 태등입구, …, 불광, 독바위	6호선계		
		장암, 도봉산, 수락산, 마들, …, 천왕, 온수	7호선계		
		모란, 수진, 신흥, 단대오거리, …, 천호, 암사	8호선계		

그림 2에서 보았던 매출 모델의 경우, 변수 차원은 매출액과 매출 수량으로 구성되어 있다. 그런데 사용자는 분석 과정에서 매출액과 매출 수량을 이용하여 새로운 항목을 만들어 볼 수 있다. 즉, 사용자는 매출액을 매출 수량으로 나누어 평균 매출 가격을 구해 볼 수 있다. 이 경우, ‘매출액/매출 수량’을 관계식이라 한다.

이러한 관계식에 의해 다양한 비지니스 룰이 표현될 수 있다. 일반적으로 대부분의 관계식은 변수, 기간, 그리고 유형과 같은 차원에 집중된다. OLAP 제품들은 관계식을 효과적으로 구성할 수 있도록 다양한 함수를 제공하고 있다. 하지만 제품에 따라 지원되는 함수의 수와 종류에는 차이가 있으며, 일반적으로 다차원 데이터베이스 제품들이 관계형 OLAP 제품에 비해 보다 다양한 함수를 제공한다.

하나의 차원 항목이 가지는 관계식은 특별한 조건이 없는 경우, 나머지 차원 항목들의 모든 조합에 대해 적용된다. 이처럼 다차원 모델은 항목간에 관계식을 설정함으로써 다차원 모델 상에 존재하는 방대한 셀에 대해 일일이 관계식을 부여하는 번거로운 작업을 거치지 않고서도 무수히 많은 셀의 값을 계산해 낼 수 있다. 항목간의 관계식은 방대한 분량의 데이터를 효과적으로 계산해 낼 수 있도록 한다.

## 4. OLAP 사례 : 지하철 승하차 패턴 분석

### 4.1 프로젝트 배경과 다차원 모델링 결과물

도시철도공사는 수입원은 승차권 판매가 주종을 이뤘으나, 지하철역사 내 점포 임대료와 광고제재료도 점차 증가하고 있다. 그러나 현재의 역별 임대료나 게재료는 역의 특성에 관계없이 일률적으로 책정되어 있다. 따라서 이 요금을 상향 책정하기 위해서는 무엇보다 먼저 역별 인구이동 패턴을 다각적이고 상세하게 파악해야 한다고 판단했다. 이것은 요금 인상시 현재의 입주 업체와 광고주를 설득하기 위해서도 필요하고, 장차 새로운 입주 업체나 광고주를 유치하는데 필요한 자료임에 틀림없다.

도시철도공사 CIO인 P본부장은 다차원 모델링 전문가인 K교수에게 이 공사가 운영하는 4개 노선 (총 149개 역)의 역별 승하차 패턴을 분석할 수 있는 OLAP 시스템 개발을 의뢰하기로 했다. 의뢰를 받은 K교수는 먼저 공사 내의 승하차 분석 업무 담당자와 대면하여 업무 분석을 진행하였고, 이 결과 그림 5와 같이 차원 5개를 가진 다차원 모델을 도출하였다.

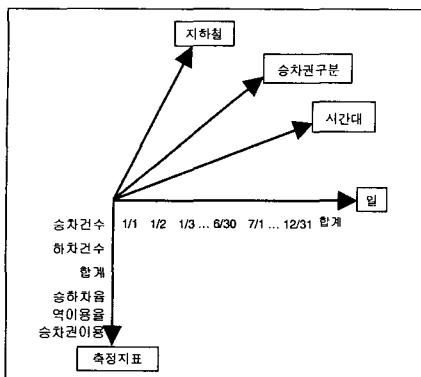


그림 5 지하철 승하차 패턴 분석을 위한 다차원 모델

이어서 주요 변수, 차원 항목, 그리고 관계식이 도출되었으며, 이 결과물은 표 3과 같다.

#### 4.2 생성 가능한 다차원 뷰와 다차원 모델 크기

표 4 네비게이션 가능한 다차원 뷰

	역	승차권	일	시간대	측정지표
역	X	-	-	-	-
승차권	O	X	-	-	-
일	O	O	X	-	-
시간대	O	O	O	X	-
측정지표	O	O	O	O	X

이렇게 다차원 분석 시스템을 구축하면 표 4와 같이 많은 양의 정보를 생성할 수 있다. 이 표에서 각 셀들은 보고서 뷰 그룹을 표시하며, “O” 표시는 의미 있는 정보를, “X” 표시는 의미없는 정보를 각각 나타낸다. 다차원 모델의 보고서 뷰 그룹 갯수는 “차원 갯수×(차원갯수-1)/2”인데, 따라서 이 5차원 모델

의 뷰 그룹은 10개이다. 그런데 여기서 주의할 점은 각 차원이 고유한 계층 구조를 가지고 있다는 점이다. 그러므로 각 셀은 한 개의 뷰를 의미하는 것이 아니라 매우 많은 조합의 뷰를 산출해 낼 수 있는 뷰 그룹인 것이다.

한편 모델의 크기는 셀의 수로 표현하는데, 셀의 수는 각 차원 항목의 개수를 곱함으로써 산출된다. 그런데 다차원 모델의 셀은 2가지 종류가 있다. 그것은 데이터 셀(data cell)과 결과 셀(result cell)이다. 데이터 셀에는 모델에 입력된 원시 데이터가 들어 있고, 결과 셀에는 모델을 계산할 때 모델의 계층 구조식 및 관계식에 의해 새롭게 생성되는 값들이 저장된다. 실제로 이 결과 셀 안에는 사용자가 원하는 정보들이 들어 있으며, 사용자는 주로 이 스페이스를 네비게이션하게 된다.

표 5에서 볼 수 있듯이 이 다차원 모델은 측정 지표 차원이 3개 항목, 일 차원이 366개 항목, 시간대 차원이 27개 항목, 승차권 구분 차원이 3개 항목, 그리고 지하철역 차원이 154개에 달함으로써 약 2,270만 개의 셀을 보유하고 있으며, 이 중 결과 셀이 77.1%를 차지하고 있다.

#### 4.3 샘플화면과 그 해석

이렇게 구축된 OLAP 시스템은 어떤 형태로 정보를 출력하는지 샘플화면을 사용하여 설명해 본다. 그림 6은 RF카드, 즉 신용카드 또는 선불카드로 지하철을 이용하는 추세가 전년도 대비 증가일로에 있음을 보여주고 있다. 모든 역에서 RF카드 이용율이 2001년도에는 30%대 이었으나, 2002년도에는 50%대 임을 볼 수 있다. 특히 RF카드 사용자는 보통권이나 정액권 사용자와는 달리 직장인 또는 경제활동에 종사하는 성인으로 분류되는데, 이 점은 입주 업체 또는 광고주에게 시사하는 바 크다.

표 5 승하차 패턴 분석용 다차원 모델의 크기

	측정지표	일	시간대	승차권구분	지하철역	셀 수	비율
데이터셀	2	365	24	2	149	5,185,920	22.9%
결과셀	3	1	3	1	5	17,943,270	77.1%
전체	5	366	27	3	154	22,679,190	100.0%

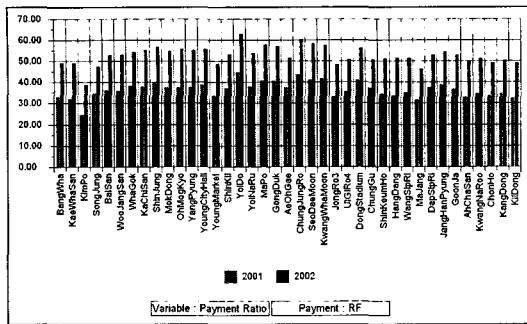


그림 6 RF카드 승차권 이용률

그림 7은 2001년도의 호선별 승하차 인원을 비교하고 있다. 도시철도공사는 5호선부터 8호선을 운영하고 있는데, 호선간 수송인구 격차가 큰 문제점을 안고 있다. 8호선의 경우는 분당선이므로 큰 문제 없으나, 6호선의 경우 수송분담율이 10% 정도 밖에 못 미치어 문제점으로 지적되고 있다. 일단 호선별 문제점을 파악한 후, 다른 차원에 대해서도 심층적으로 비교 분석해 봄으로써 원인을 밝혀내고, 대안을 마련하는 계기를 가질 수 있다. 이 그림에서 발견할 수 있는 또 하나의 패턴은 도시철도공사는 전반적으로 승차하는 인원 수가 하차하는 인원 수보다 많다는 점이다. 이것은 5호선~8호선 산하의 지하철 역은 상대적으로 도심 혹은 변화가에 위치해 있지 않아서 바로 하차하는 승객보다 환승하는 승객이 많기 때문이라고 설명할 수 있다.

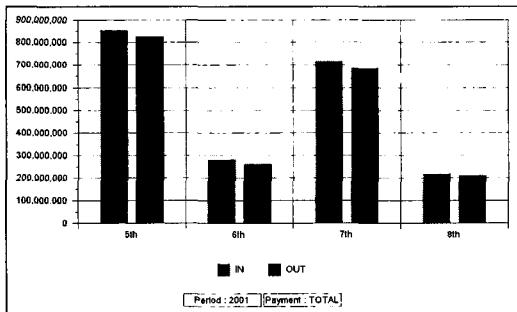


그림 7 호선별 승하차 인원 비교(2001년도)

그림 8과 그림 9는 시간대별 승하차 인구 비교를 위해 참조할 수 있는 차트이다. 그림 8은 하계역으로 써 베드타운인 이곳은 출근시간인 7~8시에 승차 인원이 많고, 반대로 저녁시간에 하차 인원이 많다. 이

에 반해서 도심지역인 종로3가는 8~9시에 하차 인원이 많고, 반대로 퇴근시간에 승차 인원이 많다. 현재 이 차트는 1년 전체를 집계한 상태를 나타내고 있지만, 사용자는 OLAP 시스템을 이용하여 특정요일 또는 특정일(연휴 다음날, 비온 날 등)을 드릴다운하여 조회해 볼 수도 있다. 즉, 다양한 변수들이 지하철 이용에 어떤 영향을 미치는지 파악할 수 있게 된다.

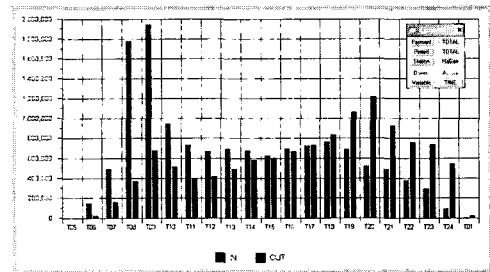


그림 8 하계역 시간대별 승하차 인구

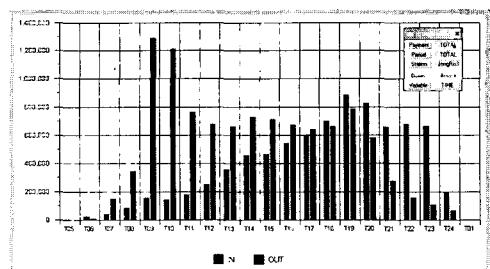


그림 9 종로3가역 시간대별 승하차 인구

이상으로 OLAP 시스템의 효과를 사례를 통해서 간접적으로 표현해 보았다. 실제로 이 시스템을 사용하게 되면 2,200만 개의 셀에 담긴 방대한 패턴 정보를 다차원적이고 심층적으로 분석할 수 있는데, 이렇게 되면 도시철도공사는 승객 기반에 대한 이해가 깊어짐과 동시에 각 역의 특성에 대한 이해가 높아지게 된다.

## 5. 결 론

다차원 모델로 디자인된 OLAP 시스템을 이용하면 업무 담당자는 이제까지 자기가 가지고 있었던 업무 지식을 훨씬 뛰어 넘는 패턴들을 발견할 수 있게 된다. 이렇게 되면 기업은 왜곡되거나 좁은 시야에서 벗어나, 보다 종합적이고 다차원적으로 시장과 고객을 보는 눈이 생긴다. 따라서 이러한 고객 기반에 대

한 이해를 바탕으로 보다 적극적이고 심도있는 전략을 펼칠 수 있게 된다.

오늘날의 경영 환경은 과거보다 더욱 경쟁적이고 역동적이다. 세계시장에서 굴지의 기업들과 경쟁을 벌이고 있는 국내 기업들은 하루가 다르게 증폭되는 변화의 물결을 체험하게 된다. 이렇게 급박한 시대일수록 기업의 사활은 그들의 정보 시스템의 상황 적응 능력과 정보를 분석/취합하는 능력에 따라 좌우된다. OLAP과 데이터마이닝을 필두로 한 비즈니스 인텔리전스 부문은 계속적으로 기업 내에 확산될 것이다. 기업이 경쟁에서 우위를 차지하는 능력은 종국적으로는 기업이 가지고 있는 비즈니스 인텔리전스 능력의 질과 효율성, 그리고 얼마만큼 조직 내에서 활성화되어 있느냐에 달려있다고 해도 과언이 아니다.

## 참고문헌

- [1] 조재희, 박성진, OLAP 테크놀로지, 시그마인사이트컴, 1999.
- [2] Arbor Software, "Multidimensional Analysis: Converting Corporate Data into Strategic Information," White Paper, 1994.
- [3] Codd, E.F., S.B. Codd and C.T. Salley, "Providing OLAP to User-Analysts: An IT Mandate," White Paper, Codd & Date Inc., 1993.
- [4] Crandall, Richard L., "Multi-Dimensionality in a Decision Support System," White Paper, 1983.
- [5] Finkelstein, Richard, "Understanding the Need for On-Line Analytical Servers," White Paper, Comshare, 1994.

[6] Kimball, Ralph, "A Dimensional Modeling Manifesto," DBMS Magazine, August 1997, <http://www.dbmsmag.com/9708d15.html>.

[7] McFadden, Fred R. and H.J. Watson, "The World of Data Warehousing: Issues and Opportunities," Journal of Data Warehousing, Vol.1, No.1, pp.61-71, 1996.

[8] Pense, Nigel, and R. Creeth, "The OLAP Report: Succeeding with On-Line Analytical Processing," White paper, 1995.

[9] Stamen, Jeffery P., "Structuring Databases for Analysis," IEEE Spectrum, pp.55-58, October 1993.

[10] Youness, Sakhr, Professional Data Warehousing with SQL Server 7.0 and OLAP Services, pp.67-71, Wrox Press, Birmingham (UK), 2000.

---

## 조 재 희



1983 연세대학교 경영대학(학사)  
1985 Miami University (Ohio) 경영학  
(석사)  
1991 University of Nebraska 경영학  
(박사)  
1992~1993 펜타컴퓨터 EIS/DSS사업부  
컨설턴트  
1993~1994 LG-EDS 기술연구소 컨설  
턴트

1998~2001 광운 정보통신 창업지원센터 센터장  
1994~현재 광운대학교 경영대학 경영정보학과 교수  
관심분야 : 디차원 모델링, Business Intelligence, 데이터자산의  
전략적 활용  
E-mail : simon@kw.ac.kr

---