

# 부가적인 비즈니스 기회 창출을 위한 '승부수'

데이터를 보다 효과적으로 운영하기 위해 다양한 시도가 계속되고 있다. 그 핵심은 바로 일관되고 통합된 데이터 베이스를 확보하는 것이다. 이는 영업이나 마케팅을 비롯해, 일반적인 관리에 이르기까지 모든 조직의 업무를 보다 능률적이고 효과적으로 수행하기 위한 것이다. 특히 회사간의 M&A가 빈번하게 이뤄지고, 유사한 프로그램들이 지속적으로 설치되고 있으며, 관련 업체들끼리 다양한 협력관계를 체결하게 되면서 이러한 요구사항은 더욱 중요시되고 있다. 데이터에 진정한 가치를 부여하기 위한 하나의 요건으로 받아들여지고 있는 통합 데이터베이스 구축의 필요성과 방법을 검토해보자. <편집자>

**당**

신의 영업 책임자가 이름, 주소, 전화번호가 포함된 영업 자료를 당신에게 주었다고 하자. 각 영업 담당자들은 이러한 자료를 활용해 영업을 보다 효과적으로 전개할 수 있는 기회를 갖게 될지도 모른다. 대부분의 경우 이와 같은 자료가 고객의 보다 깊은 내면을 파악하는 데에 도움이 될 것이라고 생각하기 때문에 고객 데이터베이스를 구축하는 데에 다양한 노력을 기울이고 있다.

지금 각 영업 담당자들의 연봉과 그들이 유사한 제품들을 판매하기 위해 소비한 비용, 그리고 그 제품을 구입한 사람들의 성향이 정리돼 있는 리스트가 있다고 상상해보자. 아마도 이런 의문이 생길 것이다. 현재 우리의 영업 담당자들은 어떤 수준의 영업 리스트를 가지고 있을까?

이러한 리스트는 결과적으로 향상된 영업자료를 확보함으로써 우수한 실적을 기록할 수 있도록 해주는 하나의 예이다. 사용자들은 원본 리스트에 추가적인 개인정보와 통계적인 정보(Demographic Information)를 부가함으로써 더욱 개선된 리스트를 확보할 수 있게 된다.

조직의 데이터가 가지고 있는 가치는 기본적인 엔터티 셋과

데이터가 가치를 지닐 수 있도록 하나의 통합된 시각을 제공, 추가적인 정보를 축적함으로써 엄청나게 확대될 수 있다. 많은 이들은 고객들과의 관계를 더욱 효과적으로 유지하기 위해 데이터를 확장하는 데에 온갖 노력을 기울이고 있다.

이는 영업의 능력을 극대화할 수 있도록 고객 프로파일(프레임워크 역할을 담당하는)을 생성하는 데에 초점을 두고 있는 것이다. 마치 고객 서비스를 위해 고성능의 장비를 도입한 것처럼 말이다.

지금부터 정보의 형태와 관련된 몇가지 사례를 제시할 것이다. 여기서 언급하고 있는 사례들을 면밀히 검토해보면 자신의 데이터를 보다 강력하게 개선할 수 있는 방안을 모색할 수 있을 것이다.

## 지리적 정보(Geographic Information)

우선 사용자는 주소를 표준화함으로써 고객에 대한 기록의 가치를 향상시킬 수 있다. 이러한 과정을 통해 사용자는 USPS(United States Postal Standard)와 같이 이미 오랜 기간 동안 정의돼 왔던 방식을 자신에게 적합하도록 수정할 수 있다. 결국 조직의 정보를 더욱 간결하고 분명하게 정리하



는 것이 가능해진다.

주소를 표준화할 때 사용자는 지리적인 정보(해당 지역의 코드나 위도/경도를 나타내는 등 구체적인 수치를 이용해)를 추가해야 한다.

### 인구통계학적인 정보(Demographic Information)

사용자는 고객의 연령, 결혼 여부, 성별, 수입, 국적 등 인구학적인 정보를 추가함으로써 고객 정보를 향상시킬 수 있다.

### 사이코그래픽 정보(Psychographic Information)

사이코그래픽(시장을 분류할 때 사용되는 소비자의 생활양식 측정 기술) 정보는 각 개인들의 생활패턴이 어떤 방식으로 이뤄지는지를 정의함으로써 얻을 수 있는 추가적인 정보이다. 이러한 방식의 분석을 수행하기 위해서는 제품이나 현재 사용하는 브랜드, 브랜드에 대한 선호도, 조합 멤버십, 여가 활동 등을 그 기준으로 활용한다.



데이터의 가치 증대를 위해 효과적으로 활용될 수 있는 방법은 '진보된' 데이터베이스 내에 하나의 특정 데이터베이스보다 더욱 많은 정보를 저장할 수 있도록 조합하는 것이다. 이러한 방식은 대부분 데이터베이스 조인의 형태로 이뤄진다.

데이터의 가치를 향상시키는 능력은 SQL 문에 대한 숙련도로 좌우될 수 있는 것이 아니다. 오히려 정보를 통합하는 데에 필요한 질적인 분석 능력이 더욱 중요하다. 물론 데이터 통합 능력과 우수한 기술을 모두 가지고 있다면 더욱 영향력 있는 통합 데이터를 생성할 수 있을 것이다.

### 통합은 왜 필요한가?

우리는 왜 데이터를 통합해야 하는 것일까? 이는 깊이 있게 생각해볼 가치가 있는 문제이다. 지금부터 서술하는 몇개의

시나리오는 데이터 통합이 필요한 이유에 대해 명확한 답변을 제공할 것이다.

### 회사 합병(Corporate Mergers)

2개의 회사가 서로 합병됐을 경우 그들의 고객 데이터베이스는 반드시 통합돼야 한다. 고객 레코드가 반드시 통합돼야 하는 이유는 우선적으로 영업 활동을 전개함에 있어 고객들이 잠재적인 혼란에 빠지지 않도록 하기 위함이다.

예를 들어, 특정 날짜에 여러 명의 영업 담당자가 계속해서 한명의 고객을 방문했다고 하자. 고객은 그 날을 매우 고통(?)스러운 경험으로 기억할 것이며, 결국 체계적이지 못한 회사의 활동에 부정적인 평가를 내리게 될 것이다.

재미있는 것은 합병으로 인해 새롭게 설립되는 회사는 원래의 회사들이 가지고 있던 고객 데이터 셋 위에 비즈니스 인텔리전스 정보를 추가적으로 결합시킬 수 있게 된다는 사실이다.

### 관련 프로그램(Affinity Programs)

관련 프로그램은 이미 기존부터 포함돼 있던 고객층을 겨냥해서 커스터마이징이 이뤄진 제품, 즉 고유 시장층을 겨냥해 준비된 제품을 판매하기 위한 것으로 이해할 수 있다. 이는 특정한 형태나 영역과 관련해 요구를 가지고 있는 고객들을 위한 것이다.

관련 프로그램의 예로는 대학의 학생들을 위한 신용카드를 들 수 있다. 대부분 이러한 카드에는 그 학교의 중앙도서관 그림이 새겨져 있으며, 일정 수준의 금액이 매년 대학에 기부되는 프로그램을 운영하게 된다.

유사한 제품을 만들었을 때 조직들은 주요 고객 리스트를 통합함으로써 관련 프로그램이 효과적으로 운영될 수 있도록 해주는 작업에 착수한다. 마치 은행들이 고객의 계좌에 평균 5,000달러 이상의 잔고가 유지되도록 유도하는 것처럼 말이다. 또한 아이비리그(Ivy League) 학교의 졸업생 주소록처럼



통합된 데이터베이스를 구축하기도 한다.

### 공동 마케팅(Cooperative Marketing)

많은 회사들은 다른 회사와의 공동 마케팅 캠페인을 통해 상호협력 관계를 구축, 다양한 시너지 효과를 이끌어내고 있다. 이러한 캠페인을 실시하는 이들은 주로 고객이 앞서 구매한 어떤 상품을 토대로 또 다른 제품을 구매하도록 유도하는 데에 주안점을 둔다. 이전의 경험이 지속적으로 영향을 미칠 수 있다는 데에 주목하고 지속적인 계획을 수립하는 것이다.

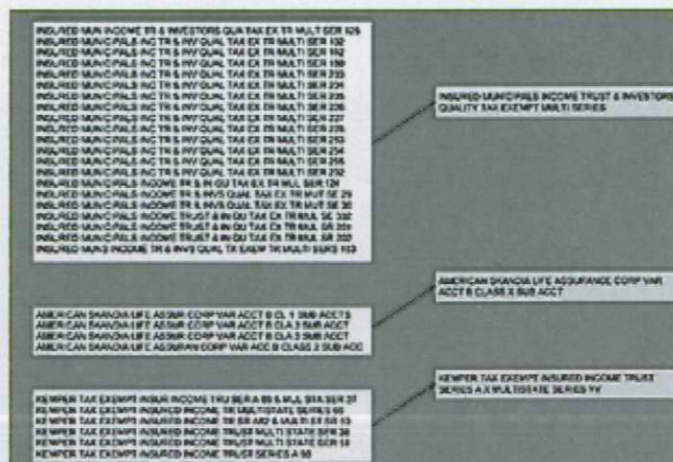
장거리 전화 공급자와 항공사가 하나의 프로그램을 만들어 고객들이 장거리 전화를 신청했을 때 마일리지 보너스를 제공하는 것은 공동 마케팅 캠페인의 좋은 예가 될 수 있다.

### 데이터 클리닝(Data Cleansing)

대규모 고객 데이터베이스에서는 이사를 하거나 이름을 변경하는 고객의 수도 실로 엄청날 것이다. 또 결혼을 하거나 이혼을 하는 경우도 빈번하게 발생한다. 이러한 변화는 모두 고객 데이터베이스의 정확도를 떨어뜨리는 결과를 가져온다.

더욱이 데이터를 등록하는 것은 바로 사람이기 때문에 이들은 언제라도 실수를 범할 가능성을 지니고 있다. 이 시나리오는 개별적인 형태로 또는 하나의 사내 엔티티로 구성하기 위해 가지고 있는 레코드들이 중복될 수 있음을 의미한다.

〈그림〉을 보면 중복된 많은 내용들을 자동으로 인지할 수 있도록 하는 것이 그리 수월하지 않음을 알 수 있다(이는 U.S.



▲ 하나의 특정 개념을 다양한 방법으로 표현하는 것이 가능하다.

Securities and Exchange Commission's Edgar 데이터베이스에서 추출한 실제 데이터를 포함하고 있다). 이렇게 중복된 내용을 제거하고 데이터베이스를 깨끗하게 정리할 수 있는 하나의 방법은 중복된 데이터를 서로 링크 시키는 것이다.

또 하나의 프로세스(Householding이라고 일컬어진다)에는 데이터베이스 레코드의 속성에 기반을 둔 개개의 단위 그룹들을 통합하는 과정이 포함된다. 독자적으로 구축된 남편과 부인의 레코드를 가족이라는 하나로 그룹으로 통합하는 것은 그 가장 간단한 하나의 예가 될 수 있다.

### 의학 진단(Medical Diagnosis)

건강을 위해 치료를 결심하는 것과 마찬가지로 데이터에 대한 진단을 실시함으로써 데이터 통합을 추진한다. 한 부모의 히스토리와 프로파일에 부합되는 내용을 검토하면서 또 다른 부모의 히스토리를 대응시켜 일치하는 사항이 있는지 찾는다.

### 사기 검출(Fraud Detection)

건강보험 사기와 같은 영역에서는 부정적인 행동들을 분석, 이를 토대로 유사한 패턴을 찾아 데이터 셋을 구성한다.

### 데이터 통합과 레코드 연계

어떤 2개의 레코드들이 선택된 키의 셋에 기반을 두고 조합될 경우, 우리는 그 2개의 레코드가 링크됐다고 표현한다. 일

반적으로 레코드 연계는 오직 선택된 키들이 정확하게 일치했을 때에만 이행된다. 하지만 다음과 같은 이유 때문에 간단한 것처럼 보이는 레코드 연계도 한계를 가지고 있다.

- 레코드 연계는 부정확하다: 2개의 레코드들이 동일한 엔티티를 의미한다고 할지라도 어떤 속성 값의 근소한 차이는 이들 레코드가 서로 연계되지 못하도록 할 것이다.

- 정보 소스는 동시성이 없다: 상대적으로 데이터베이스의 고객 정보는 정적인 상태로 머무르게 마련이다. 하나의 데이터베이스에 있는 데이터가 또 다른 데이터베이스 내에 있는 정보와 완전히 동기화 된다는 것은 거의



불가능하다. 아니 분명한 차이를 가진다는 것이 더욱 적절할 것이다.

실제로, 내가 워싱턴 D.C.에서 뉴욕시로 이사했을 때 온라인 전화 디렉토리의 내 리스트는 거의 1년이 지난 이후에나 업데이트됐다.

· 정보가 유실될 수 있다: 부정적인 측면에서 살펴볼 때 데이터 베이스 조인은 통합을 실시하는 동안 원래 포함하고 있던 중요한 정보 가운데 어떤 데이터 셋을 유실하게 될지도 모른다는 것을 의미한다.

일례로, 최근 내 딸아이 앞으로 어떤 편지가 전달됐다. 그 편지에는 원거리 전화 서비스를 신청했을 경우 비행기를 무료로 이용할 수 있도록 5,000 마일의 보너스 마일리지 제공 서비스 내용이 적혀 있었다.

사실 이 편지는 서비스 이용과 관련해 의사결정 권한을 가지고 있는 아내에게 전달돼야 했다. 하지만 얼마전 이사를 하면서 딸이 아내의 카드를 이용해 서비스를 신청하면서 자신의 의사결정권자로 표기했기 때문에 편지가 딸에게 보내진 것이었다.

연계 작업의 표준화가 제한적으로 이뤄질 때, 결국 그것은 동일한 업무처리를 위해 비능률적이고도 소요비용을 증대시키는 결과를 가져온다. 현재 업계에서 가장 일반적으로 실시되고 있는 마케팅 방법으로는 DM을 들 수 있다. 하지만 상당수의 메일링은 부정확한 정보를 기반으로 하고 있다. 심지어 동일인에게 여러 개의 메일이 발송되는 경우도 있으며, 주소지가 잘못돼 있는 경우도 빈번하게 발생한다.

현재의 고객들은 그들이 이전에 구입했던 제품들에 대해 정확히 기억하고 있지 않다. 하지만 담당자들은 이러한 사실을 깨닫지 못하고 어리석은 방식으로 잘못된 마케팅을 계속하고

있는 것이다. 체계적이지 못한 마케팅은 고객들에게 불신감을 심어줄 것이며, 결국 기업에게는 고객 이탈이라는 치명적인 결과를 선사할 것이다.

예를 들어, 어떤 서비스 회사가 고객에게 발생하는 변화(지리적인 이동 - 이사가 그 좋은 예이다)를 정확하게 파악하지 못한 상태라고 하자. 이 회사가 고객에게 적절한 수준의 서비스를 제공한다는 것은 기대하기 어려우며, 실질적인 관점에서 매달 사용자에게 요금을 청구하는 작업도 혼선을 빚게 될 것이다.

## 퍼지 매칭

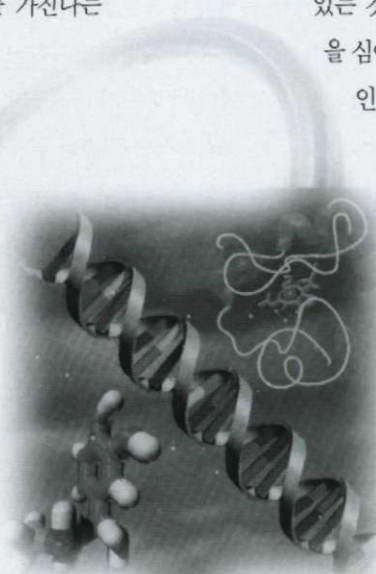
연계 작업을 제한적으로 표준화하기 위한 하나의 방법으로 퍼지 매칭(Fuzzy Matching)을 들 수 있다. 퍼지 매칭은 근사치(Approximate)와 거의 동일한 개념으로 이해할 수 있는데, 이 기술을 적절히 활용함으로써 연계 작업을 어느 정도 표준화하는 것이 가능하다.

표준화된 연계는 정확히 부합하는 어떤 2개의 값을 요구한다. 퍼지는 이러한 요구사항을 일부 완화시킴으로써 우리가 기억하기 어렵거나 파악하기 힘든 것들을 가장 근접한 수준에서 연계시킬 수 있도록 기회를 제공한다.

다음 예를 살펴보자. 나는 가끔 앞으로 개최될 데이터웨어우징 세미나에 초대받곤 한다. 최근에 받은 어떤 우편물에는 주소가 다음과 같이 적혀 있었다.

David Loshin  
Prod Manager  
Jones Smith Co  
750 X Ave  
New York, NY 10019

얼마 후 다시 하나의 우편물을 받았는데 그 우편물에 쓰여 있는 내용은 첫 번째 우편물과 비교할 때 많은 차이가 있었다. 아니 주소를 제외하고는 모두 다르게 적혀 있었다. 그 내용은 다음과 같다.





H. D. Loshin  
Product Manager  
Jones Smith & Co  
750 X Ave Fl 12  
New York, NY 10019

이들 2개의 레코드는 서로 전혀 연계가 이뤄지지 않았던 것이다. 퍼지 매칭 스키마는 이들 이름과 주소 레코드에 각각 유사한 필드를 할당한다(표 참조). 서로 상이하게 기록된 이 2개의 주소가 동일한 사람의 것이라는 사실을 직감적으로 알 수 있기 때문이다.

하지만 어떤 정보(흥미로운 데이터웨어하우징의 사이코그래픽 아이템)를 하나 추가한다고 가정해보자. 이 경우 2개의 레코드를 자동으로 연결해 서로 균형을 이루도록 해주는 프로세스를 생성할 수 있을지도 모른다. 한가지 흥미로운 점은 내 집의 주소가 동일하게 적혀있는 초대장을 받기도 했다는 사실이다.

퍼지 매칭은 그들의 데이터 확장 프로그램을 더욱 강력하게 해줌으로써 비즈니스 기회를 확대시키는 결과를 가져다준다. 동일한 엔티티(고객, 주소, 비즈니스 등)에 대한 설명이 서로 상이한 경우 이를 동일시하는 것은 지금도 가능하다. 대규모의 동일한 데이터 셋은 다양한 관점에서 이점을 제공한다.

앞서 언급한 예는 겨우 빙산의 일각에 불과하다. 이를 통해 우리가 추가적으로 얻을 수 있는 주요 기능에는 다음과 같은 것들이 있다.

- 보다 많은 레코드를 연계시킨다면 고객의 행동을 분석함으로써 엄청난 통찰력을 얻는 것이 가능해진다.
- 직접 마케팅 캠페인에서 2중 또는 3중으로 중복된 메일링을 제거할 수 있어, 결국 상당한 비용절감 효과를 가져올 수 있다.

- 퍼지 매칭을 통해 조직적인 부정행위에서 분명하게 드러나는 어떤 패턴을 감지하는 것이 가능하다.
- 데이터 셋은 가장 최근의 정보와 함께 업데이트될 수 있다.
- 조직에서는 밀접한 연계 결과(결혼, 새로운 집 장만, 아기 출산 등과 같이 생활패턴에 변화를 주는)로부터 중요한 고객 정보를 추출할 수 있게 된다.

### 비즈니스 톨 추가

다음 단계는 비즈니스 톨을 추가하는 보다 지능적인 수준의 통합이다. 앞서 살펴본 예에서 우리는 자신에게 적합한 프로세스에 비즈니스 톨을 조심스럽게 접목시켰다. 이러한 톨은 연계를 목적으로 하고 있는 데이터 레코드에 '흥미로운 데이터웨어하우징'이라는 특성을 추가한다.

동일한 건물의 동일한 회사에 'Loshin'이라는 성을 가진 2명의 사람이 존재할 가능성은 매우 희박하다. 이는 성이 'Smith'이고 이름이 'John'인 사람 2명이 뉴욕의 어느 고층 건물에서 함께 일하고 있다는 것과 마찬가지이다.

추가적으로 비즈니스 톨을 첨가하는 것은 데이터의 유용성을 증대시키는 결과를 가져온다.

비즈니스 톨을 추가함으로써 얻을 수 있는 또 하나의 이점으로는 레코드를 링크하지 않았을 때보다 더

욱 진보된 로직을 부과하는 것이 가능해진다는 사실을 꼽을 수 있다.

우리는 통합된 레코드에 임베딩된 콘텐츠를 기반으로 하는 통합연계 운영의 결과를 하나의 톨로 사용할 수 있다. 예를 들어, 만일

내 딸아이에게 어떤 제안을 했던 장거리 회사가 "최근





6개월 동안 구매했던 상품 가운데 어린이표는 이러한 혜택에서 제외됩니다"라는 룰을 적용했다면, 이번처럼 마케팅 캠페인으로 인해 곤란함을 겪는 실수를 피할 수 있었을 것이다.

이 규칙들은 전형적으로 'if-then' 문장의 형태를 지니게 된다. 즉, 어떤 조건이 이행됐을 때에 비로소 그에 대한 결과가 적용될 수 있도록 하는 것이다. 2개 이상의 레코드가 통합된 프로세스 내에 연결됐을 때에는 그 레코드들의 콘텐츠가 참인지 검토한다. 사람들은 장거리 전화와 비행기 마일리지가지고 있는 룰에 다음과 같이 정형화된 룰을 계속해서 추가하길 원하고 있다.

1. 각각의 비행 레코드에 대해 사용자가 어떤 사용권한을 가지고 있는지(예를 들어 의사결정권자인지 아닌지)를 명시함으로써 레코드에 의한 서비스를 분명하게 정의할 수 있다. 다시 말해, 티켓을 구매할 때 신용카드를 사용하는 것이 일반화되면서 아이들도 자신이 가정의 '의사결정권자'라고 제멋대로 표기하는 것이 가능해진 것이다.

2. 만일 티켓을 구입한 사람이 의사결정권자가 아니라면, 장거리 전화 서비스 업체는 그 사람에게 특정 마케팅 캠페인에 대한 메일을 보내지 않았을 것이다.

또 다른 옵션은 비즈니스 룰을 사용하는 캠페인을 부분적으로 수정하는 것이다.

동일한 가정에 비행기를 이용할 사람이 한명 이상 거주한다고 가정하자. 이 경우 의사결정권자가 아닌 사람이 새로운 장거리 서비스를 선택했다면 그 가정에 있는 다른 모든 사람들도 보너스 마일리지 혜택을 적용 받을 수 있게 규칙을 수정하는 것이다. 사용자는 데이터를 통합함으로써 다양한 기회를 모색할 수 있게 되는데, 이는 비즈니스 룰의 결과를 더욱 효과적으로 만들어준다.

고객 데이터베이스가 통합 됐을 때 마치 '의사결정권자'를 추측할 수 있게 되는 것처럼 보다 개선된 분석 데이터를 파생시킬 수 있다.

공동 마케팅 프로그램을 통해 얻을 수 있는 결과는 더욱 능률적이다. 이를 통해 응답 비율을 확대시킬 수 있기 때문이다.

## Resources

- 데이터플럭스(DataFlux)  
[www.dataflux.com](http://www.dataflux.com)
- 퍼스트로직(Firstlogic)  
[www.firstlogic.com](http://www.firstlogic.com)
- 오라클(Oracle)  
[www.oracle.com](http://www.oracle.com)
- QM소프트(QMSoft)  
[www.qmsoft.com](http://www.qmsoft.com)
- SSA  
[www.searchsoftware.com](http://www.searchsoftware.com)
- 트릴리엄 소프트웨어(Trillium Software)  
[www.trilliumsoft.com](http://www.trilliumsoft.com)
- 배리티(Vality)  
[www.vality.com](http://www.vality.com)

사용자는 마케팅 프로그램에서 티켓 계층이 응답하는 경향을 분석함으로써 특성을 부여하고, 독특한 규칙을 추론해 관련 프로그램을 향상시킬 수 있다.

나아가 비즈니스 룰을 추가함으로써 정보의 중복을 피하는 것과 마찬가지로 데이터를 깨끗하게 운영할 수 있는 능력을 확보하게 된다.

### 통합을 향해

다중 데이터베이스로부터 데이터를 통합한다는 것은 전혀 새로운 아이디어가 아니다. 하지만 기술과 시스템의 성능이 빠르게 발전하면서 지능적인 기술을 채택, 색다른 각도에서 어떤 결과를 전망해보는 것이 가능하도록 변화되고 있다.

다양한 프로세스와 데이터 개선을 통해 가치를 추가하고자 새로운 방법론에 초점이 맞춰지고 있다. 이러한 방법론은 보다 많은 회사들이 마케팅이나 영업 프로그램, 고객과의 관계를 증대시키는 등의 작업을 더욱 유효하게 전개할 수 있도록 해줄 것이다. 