4 데이터베이스의 디자인 개념

많은 시간과 예산을 들여 데이터베이스를 구축하고 의사결정이나 마케팅에 데이터를 사용했는데도 별 재미를 못 봤다는 경우를 들여다 보면, 애초에 그 데이터베이스의 디자인에 대한 개념이나 철학이 제대로 세워져 있지 않았던 경우가 많다.

무작정 건물 짓지 않듯, DB도 디자인 철학 있어야

데이터를 만지는데 무슨 디자인 철학까지 나오냐는 질문이 나올 법도 하지만, 데이터베이스와 같이 추상적 개념이 아닌 건물과 비교하면 이해가 쉬울 것이다.

건물을 지을 때에는 공사에 들어가기 전에 그 디자인(설계)에 대한 개념이 먼저 정립되어야 함은 의심의 여지가 없다.

그 건물은 주거공간인가, 사업장인가, 아니면 공장인가. 주거공간이라면 개인 가정을 위한 것인가, 아니면 여러 명이 공동으로 사용할 기숙사 같은 곳인가. 그 건물이 거주자가 단지 잠만 자는 곳인가, 아니면 다른 여가활동도 할 장소인가. 그렇다면 그 여가활동은 어떤 종류일 것인가. 설계와 건설에 관한 예산은 얼마나 될 것이며 또 그 건물을 운영하는 비용은 매월 얼마 정도로 책정되어 있나.

만약에 어떤 사람이 이런 질문에 대한 대답도 없이 무작정 건물을 짓기 시작한다면 그 사람은 돈이 썩어 나게 많거나 아니면 정신이 약간 이상한 사람이라고 봐도 무방할 것이다. 하지만 데이터를 오랫동안 마케팅에 사용해 온 미국에서도 많은 데이터베이스나 데이터 관련 프로젝트들이 이런 질문에 대한 대답 없이 무작정 시작되는 경우가 많다. 건물을 이루는 벽돌과 기타 자재들을 단순히 쌓기만 한 것이 아니듯, 데이터베이스에도 디자인 철학이 반드시 있어야 한다.

많은 회사들이 있는 데이터들을 그냥 한 곳에 모아 놓기만 하면 데이터베이스가 되는 양 사업을 추진하는데, 그것은 마치 감독 없는 영화나 설계사 없는 공사와 다를 바가 없다. 요즘 미국에서는 그런 디자인 철학을 가지고 모든 데이터와 분석에 관련된 사업이나 프로젝트를 총괄하는 CDO, 즉 Chief Data Officer라는 중역급 직책이 많이 생기고 있는데, 그들이 여기서 예로 든 영화 감독과 같은 존재이다.

CDO(최고데이터책임자, Chief Data Officer)의 역할

그런 CDO들의 역할이 기존의 IT계통 직책과 과연 어떤 점에서 다르며, 또 그들이 데이터를 만지는 사람이냐,

아니면 분석에 통달을 한 사람들이냐에 관한 질문들이 미국에서도 많이 나오고 있다. 거기에 대한 나의 생각은 "둘 다 아니다"라는 것이다.

CDO란 분명히 비즈니스를 먼저 생각해야 하는 직책이다. Analyst와 요즘 유행하는 Data Scientist라는 말의 차이도 그들이 수학이나 기술적인 요소들을 먼저 생각하느냐, 아니면 비즈니스에 더 중점을 두느냐로 볼 수도 있겠다. 기술적인 요소만 중요시하고 궁극적인 사업 목적을 등한시 하는 전문가는 Data Scientist가 아니라 Data Plumber, 즉 막힌 데나 뚫어주는 상, 하수도 배관공 같은 존재라는 극단적인 표현을 하는 이들도 있다.

그 위에서 데이터와 분석을 총괄하는 사람이 기술적인 요소에 대해 아예 무지할 경우에는 마치 영화감독이 영화에 관한 제반 기술을 모르는 것과 마찬가지로 위태한 경우라 할 수 있다. 반대로 과거에 CTO로 불렸건 현재에 CDO로 불리건 (혹은 Chief Analytics Officer이건) "기술적"인 타이틀을 가진 중역급 인사가 기술적인 면을 우선하여 비즈니스를 도외시한 의사결정을 하기 시작하면 그 끝이 좋을 수가 없다. 데이터와 그에 관련된 데이터베이스 구축, 또 그것을 이용한 분석과정을 크게 보자면 ① 매출을 올리기 위한 것, 그리고 ② 비용을 절감하여 수익성을 높이는 것, 이 두 가지의 목적을 위해 있는 것이기 때문이다.

굳이 순서를 따지자면 필자는 비즈니스의 목적을 최상위 개념으로 놓고, 그 다음에 그 비즈니스에 관련된 질문에 대한 대답을 주는 분석과정이 있으며, 모든 데이터와 데이터베이스는 그러한 고등적 분석을 위한 최적화된 모습으로 존재해야 한다고 생각한다.

하지만 대단히 불행하게도 현실은 그 정반대인 경우가 허다하다. 즉 경영자와 의사결정자들은 분석하는 사람들이 만들어 주는 자료에만 의존하는 거꾸로 된 종속적 관계에 있으며, 분석을 하는 Analyst나 Data Scientist들은 그들을 위해 최적화 되어 있지도 않은 데이터와 허구한날 씨름을 하며 데이터베이스의 기존 구조와 관련 툴 셋(Toolset)의 한계 내에서 일을 하고 있는 것이다. 참으로 안타까운 현실이 아닐 수 없다.

목적 · 기능 결정 없는 Wish-list 나열은 설계가 아니다

심지어는 전문가가 데이터베이스를 디자인하는 경우에도 그 데이터베이스가 해야 할 일들에 대한 명확한 정의가 없어

나중에는 예산도 초과하고 제대로 된 대답도 나오지 않는 구조를 만들기도 한다. 목적과 기능의 우선순위에 대한 결정이 없이 모든 사용자들의 Wish-list를 단순 나열하는 것은 설계라고 봐주기도 어렵다.

예를 들어 많은 사업계획서에서 다음과 같은 황당한 문구를 흔히 보곤 한다. "모든 마케팅과 회계에 관련된 데이터를, 새 구매자를 창출하는 행위이건 기존 CRM 프로그램의 일부이건 상관없이 국내와 국외에서 모든 사용 가능한 채널을 총 망라하여 수집 관리하되, 업데이트를 실시간으로 할 것". 이런 '야심 찬 요구사항'이 버젓이 등장하곤 하는데, 누가 어디서 어떤 프레젠테이션이나 책을 보고 와서는 이런 요구사항들을 한 문장에 몰아넣었는지 모르겠으나, 이런 식이라면 "하늘을 날 수 있는 2톤 트럭"을 만들겠다고 하는 게 더 현실적이라 할 수 있을 정도이다.

다시 비유를 들자면, 그런 계획을 세우기 전에 "왜 트럭이 날도록 해야 하지?"라는 기본적인 질문을 먼저 하는 것이 순서일 것이다. 만약 '실시간 업데이트'라는 옵션이 실제 영업에 반드시 필요하다고 하면 그것이 가능하도록 해야 할 것이다. 그러나 그 이전에, 실시간 업데이트가 중요할 정도로 분초를 다퉈 의사결정을 내리는 사람들이 조직 내에 몇이나 있으며, 데이터베이스를 관리하는 비용이 몇 배로 늘어나도 그런 옵션을 반드시 가지겠냐는 질문을 누군가는 반드시 해야 한다는 것이다. 다시 말하지만 데이터베이스도 건물을 지을 때와 같은 자세로 임해야 한다.

비싼 자재를 쓰고 공간을 시원시원하게 넓히면 좋다는 것을 모르는 사람은 아무도 없다. 설계사나 시공자가 그것을 몰라서가 아니라, 현실적으로 모든 건물을 그런 식으로 지을 수는 없기 때문인 것이다.

데이터베이스를 만들고 관리하는 툴 셋(Tool set)을 파는 회사들도 도움이 되지 않는 경우가 많다. 대개 과도한 기능은 그런 툴 셋에 딸려 있는 옵션인 경우가 많다. 그런 제품을 파는 사람들의 입장에서는 많은 옵션을 더하는 것이 그들의 수익성을 올리는 길이기 때문에 당연히 그럴 수밖에 없다.

예를 들어 이메일 자동화 프로그램을 판다고 해서 들여다 보면, Meta-table도 자동으로 업데이트 해주며, 지저분한 데이터도 알아서 정리하고 가공해주고 (뭐가 틀린 데이터인지에 대한 정의는 누가 내려주나?), 분야가 다른 데이터도 저절로 연결해주며 (무슨 기준으로 또 어떤 Match-key로?), 타깃 모델도 알아서 만들어 수신자를 선택해 이메일도 자동으로 보내주고 (모델 타깃은 누가 정하나?), 주기적 캠페인 관리는 물론이고 (월요일이 좋은지 목요일이 더 효과적인지 어떻게 알고?), 반응 데이터까지 수집해서 (모든 채널을 통해?) 어느 형식이건 원하는 대로 리포트도 만들어 준다는 식이다.

그게 다 자동으로 이루어진다면 그것 참 신통방통한 프로그램이라고 할 수 있다. 그렇다면 그 소프트웨어가 장기적, 단기적 마케팅 계획까지 세워주냐는 질문을 해 볼만 하다. 그 도구는 과연 사람들의 행동에 대한 뉘앙스와 그들의 의도까지도 이해가 가능한가? 그렇다면 정말 그 컴퓨터는 공상과학영화에 등장하는 Hal(2001년 스페이스 오디세이)이나 Mr. Data 같은 이름을 붙여서 영화에 등장할 법하다.

옵션 잔뜩 붙은 관리 툴 셋의 함정

하지만 분명한 것은 툴 셋 자체에 막대한 투자를 하고 후회를 하지 않은 기업은 정말 찾아보기 어렵다는 것이다.

그 이유는 이렇다. 첫째, 인간대신 목적에 대한 생각까지 해주는 기계는 아직 없다. 둘째, 개개인 사용자들의 구체적 사용목적을 모르는 상태에서 소프트웨어나 툴 셋을 개발하려면 수많은 사용 용도를 미리 다 예상해서 프로그램을 짜야 하는데, 그렇게 되면 구체적 목적을 가진 프로그램을 짜는 솔루션에 비해 가격이 훨씬 비싸질 수 밖에 없다.

문제는, 그런 프로그램을 사용해서 얻게 되는 비용 효과나 수익 증대가 프로그램을 사기 위해 지불한 가격보다 적은 경우가 허다하다는 것이다. 경영진에서는 투자를 많이 했기 때문에 그 본전을 반드시 뽑아야 한다는 압력을 넣을 것인데, 매출이 두 배로 뛰고 마케팅 비용이 반으로 줄어도 감당하기 어려운 투자를 해 놓고 거기에 목표를 맞춘다는 것은 누가 봐도 무리한 주문이 되는 것이다.

게다가 이런 경우는 기성복을 맞춤 양복보다 몇 배나 더 비싸게 주고 산 형국이니, 지름길을 아는 전문가의 입장에서 보면 안타깝기 그지 없는 일이다. 불특정 다수의 불확실한 목적들을 충족시키기 위해 만든 기성 도구보다는, 목적이 분명한 데이터베이스와 툴 셋의 조합에 효과적으로 투자하는 것이 현명한 방법이다. 그리고 첫 단계로 Need Analysis, 즉 사용 필요 분석을 통해 그 목적을 분명하게 정립하는 것이 제대로 된 순서다.

사용 필요 분석으로 목적 분명하게 정립해야

이런 경우 전문가, 또는 데이터베이스나 분석(Analytics) 컨설턴트의 역할을, 다시 자동차와 비교해서 예를 들어보자.

어떤 사용자가 뜬금없이 "스포츠카만큼 빠른 트럭"을 원한다고 치자. 이런 주문을 받은 사람의 목적이 순전히 차를 만들어 파는 것이라면, 그 가격이 수백만 불에 육박하더라도 그런 차를 개발하여 만들어 주는 것이 파는 사람에겐 이익이다.

그런데 중간에 전문가가 버티고 앉아 사용자가 왜 그런 차를 원하는 지를 묻기 시작하면 얘기가 달라진다. "원하는 것"과 "필요한 것"은 엄연히 다른 것이고, 이것 저것 사용 용도를 따져보니 그런 차는 극히 예외적인 경우를 제외하고는 거의 필요가 없고, 대형 밴을 한 40~50대 사서 효과적으로 운용하면 수백만 불의 반도 안 되는 예산만 써도 충분하겠다는 결론이 나올 수 있다.

이렇게 목적을 분명히 하고 거기에 맞는 기술과 도구를 조합하는 것이 데이터를 총괄하는 사람들의 책임인 것이다. 순전히 빠르고 큰 차가 좋아서 그런 데에 돈을 쓰는 것은 말도 안 되는 일이라고 생각을 하면서도, IT의 세계에서는 그런 말도 안 되는 일들이 자주 벌어지는 것이 현실이다.

데이터베이스를 만들 때에는 단순히 "있는 데이터를 일단 모조리 한곳에 몰아넣으면 미래에 어떤 사용자가 어떻게든 가치 있는 정보를 얻어내겠지"라는 막연한 바램으로 일을 시작해서는 안 된다. 또, 목적을 정했으면 그대로 추진 해야지 도중에 자꾸 다른 목적을 더하는 것도 공사를 지연시키고 예산을 낭비하게 되는 첩경이다. 그래서 제작에 책임을 진 사람이 어떤 기능이 전체적으로 더 우선인지 확실히 정해야 하는 것이고, 그 책임자는 회사 내 많은 사람들의 요구에 "no"라고 말할 수 있어야 한다.

그것은 단순히 기술적인 지식이 있다고 해서 할 수 있는 일은 아니다. 그렇기 때문에 이 글의 서두에서 CDO들은 비즈니스를 대변하는 사람이어야 한다고 말한 것이다. 윗사람에게건 동료에게건 다른 사용자에게건 단순히 듣기 좋은 말만 하는 것은 아무나 할 수 있다. 그러려면 전문가고 뭐고 다 필요 없는 일이지 않겠는가.

반면 무작정 예산을 줄이겠다고 해서 용도가 전혀 다른 데이터베이스를 우격다짐으로 다른 목

적으로 사용하는 것 역시 바람직하지 못하다. 다시 자동차의 예를 들자면, 그런 것은 2인승 스포츠카로 이삿짐을 나르는 형국인데, 물론 불가능하지는 않겠지만 누가 봐도 바람직한 방법은 아니다. 목적이 다른 데이터베이스를 혼동해서 사용하면 전체적 효율은 떨어질 수밖에 없고, 게다가 오류도 빈번해진다(부정확한 정보는 없느니만 못한 경우가 많다). 마케팅의 경우를 보자면, 많은 마케터들이 회계나 재무, 심지어는 재고관리를 위해 만들어진 데이터베이스를 가공 없이 그냥 쓰다가 많은 곤란을 겪는 것을 미국에서도 흔히 볼 수 있다. 데이터베이스는 다 비슷한 것이라고들 여긴 탓이다. 또한 여러 옵션들의 역할과 기능이 뭔지도 잘 모르고, 그런 상황에서 벗어나고자 하는 의지도 없는 경우가 많기 때문이다.

마케팅 전용 데이터베이스에서 가능한 질문들

지금까지 강조한, "목적이 분명한 데이터베이스"가 어떤 것인지를 설명하기 위해

이쯤에서 '마케팅 전용 데이터베이스'(marketing database)란 과연 어떤 것인지 고찰해 보는 것도 유익하겠다.

CRM(Customer Relationship Management)이란 개념이 한국에도 이미 많이 (때로는 악명 높게) 알려져 있기에 그것을 예로 들겠다. 정말로 CRM전용 데이터베이스가 잘 구축되어 활용되고 있다면 단 몇 분 안에 별 문제없이 다음 질문에 대한 대답이 나올 수 있을 것이다.

Q 일년 넘게 거래해 온 고객들의 지난 12개월 간 거래 당 평균 구매액수는 얼마인가?

A 거래 내역이 다 기록되어 있어도 개인별로 정리가 되어있지 않다면 이 질문에 대답하려면 시간이 오래 걸릴 것이다. 심지어는 아예 개인별 ID가 중복 없이 제대로 정립되어 있지 않은 경우도 많다.

Q 개인 별 마지막 거래일을 기준으로 지난 12개월 간 활동적인 고객과 비 활동적인 고객의 숫자는 얼마인가?

A 놀랄 만큼 많은 회사들은 자신의 고객 숫자를 정확히 파악하지 못하고 있다. "대충 한 백만 명이 좀 안 된다"라는 것은 실제로 아는 것이 아니다. 그 중에는 지난 2~3년간 아무 거래도 하지 않은 고객도 많을 수 있다. 그들을 아직도 고객이라고 부를 수 있을까?

Q 구매 채널 별로 각 고객 당 거래 사이의 평균 날짜의 수가 얼마인가?

A 이 질문은 쉽게 보일 수도 있지만 만약 데이터가 구매기록만 있고 고객 기준, 게다가 채널 별로 정리되어 있지 않으면 평균 날짜는 고사하고 채널 별 마지막 거래일을 알기도 어려울 것이다.

Q 모든 마케팅 채널을 망라하여 개개인 고객의 예상 가치를 충족시키는데 몇 번의 캠페

인과 접촉이 필요하였는가?

A 사실 이것은 정말 어려운 질문이다. 우선 고객의 예상가치는 구매 기록을 사용한 통계적 모델이 제대로 짜여 있어 그것이 액수로 표현되어 있어야 하며, 모든 마케팅 캠페인의 기록과 반응여부도 고객별로 구분 관리 되어야 한다. 하지만 많은 데이터베이스에는 고객의 구매 기록과 회사가 그들을 접촉한 기록이 연관 없이 따로 관리되고 있다. CRM에서 흔히 말하는 "Closed-loop marketing", 즉 순환고리가 연결된 마케팅이란 이런 기본적인 데서부터 시작된다. 그 고리가 끊어져 있으면 CRM이라고 불러 주기도 어색하다.

Q 고객을 접촉할 수단은 무엇이며, 접촉을 시도할 때 성공률은 얼마인가? 고객별로 그들이 선호하는 (혹은 기피하는) 채널을 알고 있는가?

A 우편, 이메일, 전화, 문자, 방문, 매스컴 등 정말 많은 채널이 있지만 그것이 다 같은 것이 아니며, 그 접촉 순서와 조합에 따라 결과는 달라진다. 이 모든 것도 고객별로 관리되어야 한다.

Q 고객을 접촉할 때 모든 고객에게 동일한 메시지를 보내는가, 아니면 그들이 선호하는 상품과 접촉 채널 별로 개인별 최적화(Customized and Personalized)가 가능한가?

A 데이터베이스를 만들어 관리하면서 모든 고객을 다 똑같이 대한다면 그런 노력 자체가 무의미하다. 첫째로 상품, 채널 별로 누구를 언제 접촉하고 누구를 접촉하지 말아야 할지, 둘째로는 분석을 통해 특정 고객이나 비 고객을 상대하기로 정했으면 무슨 Offer를 어떤 형태의 메시지를 가지고 어떤 채널로 접촉할 것인지를 확실하게 하는 것이 데이터베이스 마케팅의 두 가지 기본이다.

Q 어떤 상품이 Gateway Product, 즉 고객들이 처음으로 사용을 시도하는 제품이며, 그것이 다른 상품의 구매와 어떻게 연결이 되어 있는가?

A 이것은 상품에 관한 질문으로 보일지도 모르지만 모든 상품이 카테고리와 개인 고객별로 동시에 정리되어 있지 않으면 이 질문에 대답하는 데 굉장히 오래 걸릴 수 있다. 많은 데이터베이스에는 상품의 SKU 번호나 제품설명이 카테고리 없이 단순 나열되어 있는데, 그런 데이터는 비정형 데이터와 다를 바가 없고, 이런 질문에 대답하려면 장시간의 노력이 필요하게 되거나 아예 대답을 못하게 될 수도 있다.

Q 기본적인 구매 기록, 즉 RFM(Recency, Frequency, Monetary) 데이터가 예측적 모델에 쉽게 사용될 수 있는가?

A 모든 데이터의 종류를 망라해 실제 구매기록에 기초한 예측이 가장 정확하다. 만약에 Analyst들이 그들의 시간의 대부분을 통계가 아닌 데이터 가공에 쓰고 있다면 이 질문에 대한 대답은 분명한 "아니오"이며, 그런 구매 기록이 간단한 필터링과 리포트에만 쓰이고 있다면 "아직 갈 길이 멀었다"가 답이다.

일단 몇 가지 예만 든 것이지만, 여기서 보다시피 이 중엔 기술적인 질문이 하나도 없다. 그 데이터베이스가 관계형 DB(Relational Database)이건 하둡(Hadoop)에 자리잡고 있건 또 무슨 소프트웨어를 사용해서 만들어졌건 그런 것들은 본질이 아닌 것이다. 아무리 효율적인 구조로 데이터를 많이 저장하고 또 단편적인 정보를 빨리 나열할 수 있다 하여도, 이러한 비즈니

스가 원하는 대답을 ① **빠른 시간 안에,** ② **정확하고,** ③ **매번 일관성 있게 제공**해주지 못한다면 다만 임시적이라도 (데이터마트의 개념에 입각해) 그 데이터베이스의 구조를 바꿔야 하는 것이다. 그런데 그 바꾸는 과정이 불가능하지는 않지만 전문 프로그래머가 수십 페이지에 이르는 플로우차트를 따라가며 며칠을 밤새고 일해야 가능한 것이라면 그 데이터베이스는 마케팅에 최적화되어 있다라고 말할 수가 없다. 많은 자동차는 4개의 바퀴로 굴러가지만 그 중에는 세단도 있고, SUV도 있고, 버스도 있고, 트럭도 있는 것이다. 과거에 자동차를 구경도 해보지 못한 사람들이나 차는 다 그게 그거다라고 말하듯이, 데이터베이스도 마찬가지다.

디자인 철학 어긋나면 '데이터 풍요 속의 빈곤'

다시 마케팅의 기본으로 돌아가면, 마케팅은 사람을 상대하는 것이기 때문에 모든 정보는 '구매자 중심'으로 재구성되어야 한다.

아무리 구매기록을 철저하게 관리하고 있어도 그것은 '구매기록'이지 '구매자에 대한 정보'가 아니다. 사람을 상대하는 데에 쓰는 데이터베이스라면 모든 기록이 '개개인 구매자를 묘사하고 있는 형태'를 가지고 있어야 한다.

그런데, 관계형 DB라는 것 자체가 사람이 아닌 구매기록(Transaction Data), 채널, 상품, 혹은 그것들을 관리하는 부서들 중심이다. 게다가 빅 데이터 시대에 접어들면서 비정형 데이터(Unstructured Data), 즉 자유형으로 생긴 데이터도 엄청나게 많아졌는데, 많은 현대식 빅 데이터는 방대한 저장과 빠른 속도의 검색을 주목적으로 튜닝이 되어 있어서, 단 한 단계만 더 깊이 들어가 고객이나 질문자 중심으로 구조로 바꾸려 해도 그게 쉽지가 않은 것이다.

따라서, 앞으로는 숫자형(Numeric) 데이터와 비숫자형(Character or Categorical) 데이터를 고객 중심의 변수(Variable)로 효과적으로 전환하고, 또 그것을 통계적 예측 모델에 잘 활용하는 방법들을 구체적으로 다루게 될 것이다.

"풍요 속의 빈곤"이라는 표현이 어울리게도 많은 의사결정자들은 정말로 많은 데이터를 눈앞에 놓고도 정작 그들의 질문에 대한 대답을 제대로 얻지 못하는 경우가 많은데, 그것은 앞에서 말한 대로 데이터베이스의 디자인 철학 자체가 어긋나 있어서 일어나는 일이다. 막상 트럭이 필요한데 눈 앞에 스포츠카만 즐비하다면, 낭패도 그런 낭패는 없듯이 말이다.