

비즈니스 성공을 돕는 통찰력의 핵심 ‘데이터 분석 아키텍처’를 수립하는 세 가지 방법

데이터가 넘쳐나는 시대다. 기업, 정부는 물론이고 개인도 스마트폰을 몇 번 터치하기만 하면 방문했던 장소와 소비 지출 내역 등의 흔적뿐 아니라 누구와 함께 있었는지도 금세 알 수 있다. 기업이 취합하는 데이터는 이보다 훨씬 더 방대하다. 이처럼 방대한 데이터를 통해 기업은 얼마나 제대로 된 통찰력을 얻고 있는지 반문하지 않을 수 없다.

수년 전, 하버드 비즈니스 리뷰(Harvard Business Review) 발표 자료에 따르면 기업이 비정형 데이터를 실제 활용하는 비중은 1%도 채 되지 않았다. 다행히 최근 조사에서는 전체 데이터의 2/3 정도로 활용 수치가 높아졌지만 이 중에서 통찰력을 얻을 수 있는 분석 데이터는 아직도 많지 않은 현실이다.

데이터를 제대로 활용하면 제품과 서비스의 품질을 향상할 수 있고, 비용과 리소스를 줄여주며, 궁극적으로 세상까지 변화시킬 수 있다. 그런데 이처럼 강력한 힘을 가진 데이터를 충분히 활용하는 기업은 많지 않다. 방대한 데이터와 통찰력 사이의 간극을 해소할 방법은 무엇일까. 지금부터 비즈니스 성공을 돕기 위한 데이터 통찰력의 핵심이 될 세 가지 방안을 살펴본다.



하나. 데이터 접근성의 우선순위를 설정하라

데이터는 필요로 하는 사람이 원하는 시점에 정확히 이용할 수 있어야 한다. 그러나 현실은 그렇지 못한 경우가 대부분이다. 여기에는 몇 가지 이유가 있다.

적절한 데이터 접근을 방해하는 요인

기업의 각 부서는 목적에 맞춰 데이터를 운영하고, 그에 따라 서로 다른 클라우드를 선택한다. 많은 데이터가 사일로화 되는 이유다. 여기에 부서별로 거버넌스가 다르기 때문에, 이에 맞춰 데이터 저장과 태깅을 할 수밖에 없다. 따라서, 분석가와 과학자들은 표준화된 뷰는 커녕 전체적인 뷰를 얻기조차 쉽지 않다.

데이터 아키텍처가 갈수록 복잡해지는 원인도 있다. 데이터 파이프라인의 개별 컴포넌트는 상당한 수준의 구성을 요구한다. 이 때문에 사일로화된 부서들의 특화된 튜닝과 기술이 필요하다. 엔터프라이즈 데이터 아키텍처는 대부분 최신 데이터 솔루션과 기존 솔루션이 공존하기 마련이다. 하지만 이 두 솔루션은 에지투코어(edge-to-core) 멀티 클라우드 파이프라인에서 함께 동작하도록 설계돼 있지 않다.

마지막으로, 사일로화되고 상호성이 결여된 아키텍처로 인해 각 사업부에는 암호화된 데이터들이 존재한다. 암호 해독과 상호 액세스에는 많은 시간이 소요되고 복잡하며, 별도의 코딩 작업이 필요하다.

지능화된 데이터 아키텍처로 문제 해결

이러한 문제를 해결하고 싶다면 다음의 세 가지 방법으로 접근할 것을 제안한다.

첫째, 사용자의 요구사항이 반영된 간결한 지능형 엔드투엔드 데이터 아키텍처를 구현해야 한다.

둘째, 검색, 수집, 카탈로그화, 태깅, 저장, 보안, 거버넌스, 액세스 등 데이터 라이프사이클 전반에 대해 “이렇게 하면 데이터를 분석에 쉽게 활용할 수 있을까?” 끊임없이 질문하라. 가격이 저렴한 이질적인 오픈소스 툴로 데이터 아키텍처를 구축할 수 있다는 말에 현혹될 수 있다. 그러나 상호작용할 수 없는 시스템 유지와 암호 해독에 필요한 리소스 비용을 고려한다면, 실제 비용 절감 효과는 아주 적다. 가장 저렴한 아키텍처를 찾을 게 아니라, 정확한 데이터를 신속하고 효율적으로, 안전하게 분석가에게 제공한다는 목표로 시작해야 한다.

셋째, 특정 사업부의 필요에 맞춰 기존의 컴포넌트를 평가하지 말고 “이 데이터가 현재 어떤 문제를 해결하고 있으며, 잠재적으로 회사의 어떤 문제를 해결할 수 있나?”를 질문하라. 액세스할 데이터의 상호의존성을 파악해 코딩과 액세스 제약 조건을 제거해야 한다. 이렇게 하면 분석가가 필요할 때 암호 해독을 하지 않고 언제든지 데이터에 간편하게 액세스할 수 있을 것이다.

사례 01 자동화된 카탈로그 프로세스 도입 → 복잡한 데이터 검색 문제 해결

강력한 거버넌스와 데이터 관리를 통해 수많은 데이터를 빠르게 취합할 수 있는 아키텍처를 구축한 기업이라도 분석가에게 신속하고 간편하게 데이터를 제공하지 못할 수 있다. 히타치 벤타라는 최근 한 금융 서비스 기업의 프로젝트를 진행하면서 통찰력을 지연시키는 요인을 확인할 수 있었다. 이 회사는 거버넌스와 데이터 관리 프로세스를 세심하게 설계한 후 하루에 수백만 건의 보안 데이터 파일을 성공적으로 처리하게 되었다. 그러나 프로세스가 복잡해 데이터 가용성의 지연을 초래했고, 데이터가 추상화되어 있어 최종 사용자가 검색하기 쉽지 않았다.

히타치 벤타라는 비즈니스 성장을 위한 데이터 활용 계획이 무엇이었는지 고객에게 다시 질문을 시작했다. 이 회사는 API 기반의 자동화된 카탈로그 프로세스를 도입했다. 그 결과, 데이터 가용성에 소요되는 시간이 줄었고, 사용자에게 친숙한 셀프서비스 ‘마켓플레이스’ 데이터 카탈로그 UI를 도입해 복잡한 데이터 검색 문제도 해결할 수 있었다.

사례 02 사일로화된 데이터를 데이터 레이크로 통합 → 데이터 중복 문제 해결

많은 기업이 사일로화되고 중복된 데이터를 갖고 있다. 분석가들이 신뢰할 수 있고, 가치 있는 데이터를 확보하는 데 어려움을 겪는 이유다. 한 글로벌 기업의 경우, 효율성은 떨어지는데 비용은 고가인 데이터 아키텍처 탓에 분석팀이 어려움을 겪고 있었다. 이 회사에는 25개 사업부가 각각 관리하는 25개의 RDB(관계형 데이터베이스) 사일로가 존재했다. 히타치 벤타라는 데이터 전환 프로젝트를 통해 회사의 모든 데이터를 카탈로그화된 데이터 레이크로 통합하고 데이터 중복 부분을 제거하도록 지원했다. 그 결과 데이터를 일일이 찾아다닐 필요가 없게 되었고, 신속한 분석 프로젝트를 수행할 수 있었다. 모든 데이터가 통합되면서 데이터 저장 공간도 4PB에서 수백TB 정도로 줄었다. 비용 절감 효과까지 얻게 된 것이다.

다시 한번 강조하지만 늘 이렇게 질문하라. “데이터의 활용 목적이 무엇인가, 누가 데이터를 이용하고 있는가, 사용자가 원하는 시점에 안전하게 데이터에 액세스할 수 있는가?”



“데이터가 필요한 사람이 원하는 시점에 정확히 데이터를 이용할 수 있어야 한다. 그러나 현실은 그렇지 못한 경우가 대부분이다.”

“늘 스스로 질문하라. 데이터의 활용 목적이 무엇인가, 누가 데이터를 이용하고 있나, 사용자가 원하는 시점에 안전하게 데이터에 액세스할 수 있는가?”

둘. 거버넌스를 위해 자동화 활용을 극대화하라

기업이 취합하는 모든 데이터의 검토와 분류, 태깅 등을 수작업으로 한다는 건 불가능하다. 분석가가 활용할 정도의 수준이 되려면 몇 개월이 걸릴 수도 있고, 정작 그때가 되면 데이터는 이미 쓸모 없게 돼버린다. 민감한 데이터 처리도 문제다. 데이터의 개인정보보호 기준을 충족시키지 못하면 리스크가 그만큼 크다.

오래된 데이터 거버넌스 프로그램과 달리 최신 솔루션들은 AI와 머신러닝을 이용해 보안 정책과 데이터 개인정보보호 기준을 준수하면서도 신속하게 데이터를 분류해 카테고리화한다. AI 프로그램은 사용자 피드백에 대한 학습 과정을 거쳐 진화하므로 훨씬 더 효과적이다. 또 AI를 적용하면, 데이터 카테고리화 과정에 사람의 편견이 개입되지 않아 데이터셋의 유용성도 향상된다.

사례 03 루마다 데이터 카탈로그 도입 → 데이터 분류와 액세스 속도 향상

히타치 밴타라의 자동화 기술을 기반으로 민감한 데이터 관리 방식을 대폭 향상한 고객이 있다. 이 회사는 수작업으로 인해 데이터 분류에만 엄청난 시간을 쏟아 붓고 있었다. 이를 해결하기 위해 루마다 데이터 카탈로그가 지원하는 AI 기반 데이터 카탈로그 프로세스가 도입되었다. 솔루션 도입 후, 데이터 액세스가 대폭 빨라졌고, 며칠씩 소요되던 프로비저닝 시간도 1~2시간으로 단축돼 성능과 정확도 역시 크게 향상되었다.

이 사례에서 보는 것처럼 거버넌스 규칙을 자동화하는 기술은 이미 시장에 나와 있으니, 한번 시도해보는 것도 좋은 방법일 것이다. 거버넌스 정책을 준수하면서도 민감한 데이터를 충분히 효과적으로 파악해 태깅할 수 있다.

“오래된 데이터 거버넌스 프로그램과 달리 최신 옵션들은 AI와 머신러닝을 이용해 보안 정책과 데이터 개인정보보호 기준을 준수하면서 신속하게 데이터를 분류해 카테고리화한다.”

셋. 상품 개발 시점부터 분석을 함께 고려하라

데이터 접근성 우선순위 설정, 자동화 활용 극대화에 이은 성공적인 데이터 분석 아키텍처를 위한 마지막 방안은 ‘분석을 염두에 둔 설계’다. 상품과 솔루션 개발을 시작할 때부터 분석 인프라를 함께 설계한다면 데이터 분석의 수준을 한층 높일 수 있다.

분석의 궁극적인 목적은 기업의 전략적 목표를 지원하고, 이를 실행 가능한 것으로 만들기 위해서다. 그러나 대부분은 “제품만 제대로 작동하면 되지. 분석과 보고서는 나중에 추가해도 되잖아. 분석이 전략적 성공의 핵심도 아니지 않나.”라고 생각하기 쉽다. 이런 마인드라면 시간이 지날수록 비용이 증가하고, 통찰력을 얻는 데 걸리는 시간도 길어진다.

개발자들은 가능한 최소한의 입력만으로 비즈니스 프로세스를 가능하게 하는 제품을 만들고 싶어한다. 이런 방식은 데이터 분석팀이 비즈니스 프로세스를 지원하는 데이터에 접근할 수는 있지만, 이후 모니터링을 위한 데이터 디코딩과 추가 작업이 매우 어렵고 고급 데이터 분석과 활용이 뒷전이 될 수도 있다.

물론 신속하고 효과적인 해결책이 있다. 상품과 솔루션 개발을 시작할 때부터 분석 인프라도 함께 설계하는 것이다. 그리고 가능하다면, 상품과 비즈니스 프로세스의 설계 및 재설계, 또는 디지털 전환을 시작할 때 데이터 분석팀을 참여시켜야 한다.

사례 04 상품 출시와 분석 시스템 동시 구현 → 1년 만에 분석 성과

히타치 밴타라는 한 거대 CMET(커뮤니케이션, 미디어, 엔터테인먼트, 기술) 기업의 신상품 개발을 지원한 적이 있다. 히타치 밴타라는 최신 데이터와 분석 전문 그룹을 신속히 배치해 각 사업부의 상품 관리자 및 개발자와 협력할 수 있도록 했다. 각각의 팀은 사업부별 핵심 목표를 설계하면서 긴밀히 협력해 전략적 지도를 완성했다. 이를 통해 비즈니스 프로세스 데이터를 적절히 배치해 상품 출시 당일의 활동을 모니터링하고, 향상된 분석을 바탕으로 미래의 전략적 의사 결정을 지원하도록 했다.

결과는 대만족이었다. 상품이 출시되고 몇 주 동안 이 회사는 미처 예기치 못한 문제를 해결할 수 있었다. 높은 비용을 발생시켜 매출에도 영향을 미칠 수 있는 주요 이슈를 파악한 것이다. 이는 비즈니스 프로세스 파이프라인의 누수 지점을 제대로 찾아낸 덕분이다. 이뿐만이 아니다. 단순한 모니터링 활동에서 벗어나, 시간당 수백만 건의 기록에 맞먹는 활용 데이터셋에 고도의 분석 모델링까지 적용할 수 있었다. 많은 기업이 분석의 성과를 높이기 위해 10여 년 이상 고군분투하고 있는 것과 달리, 이 회사는 상품을 출시한 후 성과를 얻기까지 1년도 채 걸리지 않았다.

데이터 분석을 염두에 둔 프로세스 수립

상품과 서비스를 출시하는 시점부터 분석팀을 함께 운영할 수 있을 만큼 충분한 여력을 갖춘 기업은 많지 않다. 그러나 이런 상황이 관리자와 경영진이 초기 단계에서 분석 작업을 소홀히 해도 된다는 이유가 될 수는 없다.

현재 취합되고 있는 데이터를 어떻게 분석할지 고민해보지 않은 기업이라면 비즈니스 성숙 곡선을 향상시키는 과정에서 어려움이 끊이지 않을 것이다. 데이터 분석을 염두에 두어야 한다. 분석할 수 있는 형태로 데이터가 포맷되고, 기본적인 보고 수준에서 벗어나 예측 분석이 가능한 프로세스로 빠르게 옮겨갈 수 있다.

출처 : Modern Data Foundation for AI-Driven Results, www.hitachivantara.com, 2021년 12월