

지금은 빅데이터 시대! 데이터 품질 관리로 경쟁 우위를 선점하라

빅데이터 시대가 도래하면서 기업이 개별적으로 데이터를 수집해 분석하는 것보다 훨씬 더 많은 데이터를 편리하게 이용할 수 있게 됐다. 고품질의 데이터는 기업의 비즈니스 전략을 수립하거나 정보를 기반으로 의사 결정하는 경우 필수 요소로 자리 잡았다.

많은 기업들이 데이터 품질 향상의 필요성을 충분히 알고 있고, 데이터 활용으로 얻는 이점에 대해서도 이해하고 있다. 그러나 어디서부터 시작해야 하는지 모르는 경우가 많다. 어떤 데이터를 좋은 데이터라고 정의하며, IT 인프라에 필요한 시간과 투자는 어떻게 만들어 낼지, 데이터가 끊임 없이 생성되고 있는 상황에서 비즈니스 운영에 영향을 미치지 않고 변화를 어떻게 구현할 것인지. 이와 같이 몇 가지 문제가 있음에도 불구하고 변하지 않는 사실이 하나 있다. 바로 '적절한 데이터 정제'를 거친 고품질 데이터의 전제조건은 '정확한 분석'이라는 사실이다.

1) 데이터 정제(Data Hygiene): 데이터의 정확성을 유지하기 위한 원칙 및 실천 사항



다음은 데이터 정제를 위한 5가지 핵심 요소와 데이터 품질 관리 시스템이 갖춰야 할 필수 요건, 그리고 각광받고 있는 데이터 품질 관리 솔루션인 Hitachi Content Intelligence에 대해 구체적으로 알아본다.



데이터 정제를 위한 5가지 핵심 요소

1 프로파일링으로 데이터 가치를 결정하라

데이터 품질이란 '가장 필요한 시점에, 신뢰할 수 있는 형태로 이용 가능한 올바른 데이터가 확보되어 있다'는 것을 의미한다. 데이터 프로파일링은 현재 보유하고 있는 데이터가 품질 요건에 부합하는지를 판단하는 과정으로, 데이터가 완전성과 정확성을 갖췄는지 여부를 파악하는 초기 단계다. 이를 통해 기존의 데이터셋을 점검하고, 데이터의 향후 사용 방식과 사용 목적을 정의하는 기준을 만들 수 있다.

2 데이터 소유권을 명확히 하라

모든 데이터가 동일한 조건에서 동일하게 생성되는 것은 아니다. 잠깐 사용 후 바로 폐기되는 데이터가 있는 반면, 비즈니스 운영에 중요한 데이터도 존재한다. 기업 데이터의 대부분은 일정 기간 동안 보관이 필요한 레거시 데이터들이다. 따라서 가능한 한 빠르게 데이터에 대한 소유권과 책임을 명확히 하는 것이 중요하다. 소유권의 부재는 곧 통제권 부재를 의미하기 때문이다.

3 데이터 정제는 단계가 아닌 여정임을 명시하라

데이터는 비즈니스 실행 방식을 바꿔주는 톨이다. 따라서 데이터 품질 향상은 일회성 프로젝트가 아니라 끊임없이 지속되어야 한다. 따라서 데이터 관리자를 지정해 데이터 소스 및 프로세스와 관련된 결과를 문서화하고, 보관해야 한다. 이렇게 하면 데이터 품질에 대한 접근방식이 지속적으로 개선되고, 이해관계자들과 더불어 데이터 품질 결과물을 감독하고 커뮤니케이션하는 방안을 찾을 수 있다.

4 'P-해킹(P-hacking)'을 경계하라

P-해킹은 원하는 결과를 얻기 위해 정확성이 떨어지는 데이터를 사용하거나 의도적으로 데이터를 조작하는 것을 의미한다. 충분한 시간을 갖고 데이터의 출처, 다른 데이터 소스와의 연관성, 비즈니스 관련성 등을 파악하면 의사결정을 내리거나 실행에 옮길 때 더 높은 정확성을 확보할 수 있다.

5 데이터 투자를 통해 조직을 선도하라

데이터는 기업의 가장 중요한 전략 자산이므로, 성과를 극대화하려면 데이터에도 충분한 투자가 이뤄져야 한다. 데이터 신선도 유지와 적절한 수준의 정제를 가능하게 해주는 데이터 품질 관리시스템은 이제 선택이 아닌 필수다. 이러한 시스템은 데이터 품질, 신뢰성, 완전성을 검토하고 평가함으로써 적절한 시점에 필요한 데이터가 문제를 해결할 수 있도록 지원한다.

데이터 품질 관리 시스템의 4가지 필수 요소

올바른 데이터 품질 관리 시스템을 선택하는 것은 쉬운 일이 아니다. 어떤 시스템을 고려하든 가장 중요한 요소는 확장성, 유연성, 그리고 조직에 영향을 미칠 수 있는 새로운 데이터 유형과 소스를 지속적으로 학습하고 발전시켜 나가는 연속성이다. 다음의 4가지 사항은 비즈니스 목표가 무엇이든 간에 생산성과 데이터 품질 향상이라는 성과를 얻으려면 반드시 고려해야 하는 필수 요소들이다.

1 조직 내 모든 데이터를 종합하라

데이터는 조직 내 여러 곳에 동시에 존재한다. 성공적인 이니셔티브는 기업 내 데이터를 전체적으로 조망할 수 있으나 그 성패가 달려 있으며, 이를 가능하게 해주는 핵심이 바로 데이터의 품질이다. 따라서 데이터 품질 관리 시스템은 데이터 소스의 종류에 상관 없이 모든 데이터 유형의 통합을 지원해야 한다.

2 데이터를 늘 준비된 상태로 프로파일링하라

데이터 완전성과 정확성을 평가할 수 없다면 원하는 목적에 맞게 데이터를 정의하고 연계하는 작업이 어려워질 수 있다. 예를 들어, 데이터 완전성의 전제조건은 특정 데이터가 최신 버전이고 제품에 대한 상세 정보를 담고 있다는 사실을 보장하는 것이다. 이와 달리 데이터 정확성에서는 데이터가 담고 있는 가치가 더 중요하다. 따라서 데이터 프로파일링이 제대로 수행되어야 파일명을 통해 알 수 있는 데이터의 내용보다 더 많은 것을 파악할 수 있다.

3 데이터 품질을 보장하라

데이터의 완전성과 정확성이 확보된다면, 이를 통해 사용자는 품질이 보장된 데이터를 가치 있고 연관성 있는 비즈니스 정보로 전환할 수 있다. 데이터는 본래 의미가 훼손되지 않으면서 정화, 표준화, 확대 기술을 거쳐 사용자의 목적에 부합하는 더 유용한 정보로 재탄생하는 것이다.

4 데이터 시큐리티를 보장하라

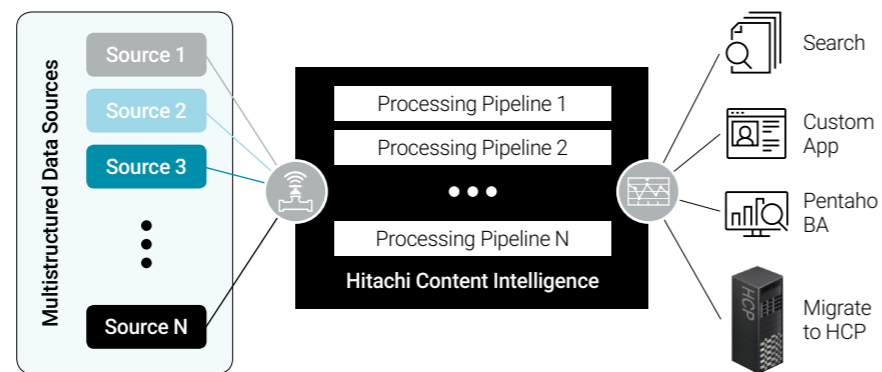
데이터 품질 관리 시스템의 핵심은 사용자가 원하는 시점에, 가장 정확한 최고 품질의 데이터를 사용할 수 있도록 지원하는 것이다. 특히 문제가 있는 데이터에 액세스하려는 사용자의 역할을 검증할 수 있도록 기업 보안 서비스와의 통합을 지원하고, 사용자의 데이터 액세스가 더 세분화된 단위로 관리, 통제돼야 하는 상황까지도 지원해야 한다.

데이터 품질 관리 솔루션 Hitachi Content Intelligence

이제는 방대하고 깊이 있는 데이터에서 신속하게 통찰력을 제공하는 솔루션이 기업에 반드시 필요하다. HCI는 데이터의 위치나 형태와는 상관 없이 기업 내 모든 데이터를 연계 및 통합할 수 있도록 지원한다. 아래 그림에서 볼 수 있듯이 HCI는 데이터가 다양한 목적으로 사용되도록 데이터를 준비시킨다.

또한 기업의 데이터 보관 요구사항에 부합하기 위해 기업 내 모든 데이터가 유연성과 확장성을 갖춘 엔터프라이즈에 최적화 된 데이터 구조로 중앙화될 수 있도록 지원한다. 모든 데이터를 한 지점에서 효과적으로 관리, 통제, 결집 및 분석함으로써 기업은 운영 효율성 향상, 고객 경험 또는 새로운 비즈니스 기회를 확대할 수 있다.

목적에 부합하는 데이터를 지원하는 HCI



데이터 품질은 궁극적으로 균형에 관한 것이다. 100%의 완전성과 정확성을 보장하는 데이터를 확보하는 것이 어렵다는 건 기정 사실이다. 따라서 둘 중 하나를 선택하는 것이 아니라 적절한 균형을 확보하고자 한다면, 올바른 데이터 품질 관리 솔루션인 HCI가 정답일 것이다.

출처 Why Data Quality is Essential for Content Analytics, www.HitachiVantara.com, 2018년 2월

Hitachi Content Intelligence 산업별 적용 사례



정부
업무 효율성 및 만족도 향상,
데이터 보안 강화

민원인의 정보 요청에 신속히 대응하고, 전체 시스템의 데이터를 대상으로 조사를 완벽하고 빠르게 수행한다. 수십억 개에 달하는 정부 기록 데이터에 대한 키워드 검색이 가능하며, 전자화된 모든 기밀 문서는 공유가 금지된다.



컴플라이언스
국제 규제 준수 절차 간소화 및
데이터 무결성 보호

국제 교역 관련 데이터가 무역 규제와 요구사항을 준수하는지 꼼꼼히 검증한다. 개별 액세스와 조직 내 데이터 검색을 통해 법적 규제와 관련된 리스크 민감도를 단순화해 안전한 상태를 유지한다. 또한 정책을 기반으로 콘텐츠의 수정 또는 삭제를 방지한다.



법률 회사
효과적이고 철저한 문서 검색,
법률 담당자 업무 능력 향상

법률 담당 직원들이 사건별 자료에 액세스할 수 있으므로 사건과 관련 있는 모든



유통
데이터 주도적인 서비스 제공으로
고객 경험 향상

모든 고객 정보를 즉각 찾아낼 수 있어 개인 정보를 유추할 수 있는 데이터 삭제가 가능하며, 소매업체가 고객의 구매 데이터를 추적, 고객 구매 성향과 소비 습관을 파악할 수 있다.



통신
고객의 개인 정보 보호와
규제 준수 실행, 고객 만족도 향상

고객 지원, 문제 해결과 관련된 데이터 액세스 속도가 빨라져 통신 업체들의 경쟁력이 강화되었다. 뿐만 아니라 콘텐츠를 공유하기 전에 선택적으로 검열과 수정이 가능해져 리스크를 줄일 수 있다.



출판 및 언론
개인 정보 보호와 규제 준수 보장 지원

정보 폐기 권리를 강화하기 위해 모든 개인정보를 즉시 찾아내 파기한다. 규제 준수와 데이터 품질이 향상돼 모든 콘텐츠에 대해 간편한 검색과 감사가 가능하다.

문서의 상세 내용을 파악할 수 있다. 또한 신상품, 특허, 상표 등과 관련 있는 모든 지적재산권 문서를 바로 찾아낼 수 있다.



금융 서비스
고객과 금융사 간 커뮤니케이션 증가,
고객 이해도 향상

문서와 기록을 통해 지불 카드 정보를 즉각 찾아낼 수 있어 지불 카드와 관련된 보안 수준을 적절히 유지 가능하다. 더 나은 의사결정을 위해 다중 소스 검색을 지원하므로 고객은 자신의 금융 기록, 온라인 검색, 분석에 액세스할 수 있다.



의료 및 보험
규제 준수 지원,
환자 의료 클레임 이해도 향상

공유 콘텐츠에서 건강 관련 정보를 찾아내 삭제하며, 클레임이 발생하기 전에 사기 가능성을 파악할 수 있도록 콘텐츠를 분석한다. 전자의료기록 시스템 간 데이터를 공유하고, 환자에 대한 완벽한 의료 기록을 기반으로 트렌드를 체크한다.

출처 How Hitachi Content Intelligence Makes Key Industries Smarter, www.HitachiVantara.com, 2018년 3월