



데이터 레이크에 관한 5가지 오해

기업이 데이터의 가치와 중요성 그리고 새로운 기회와 수익 모델을 창출할 수 있는 가능성을 인식함에 따라, 모든 데이터를 관리하고 인사이트를 이끌어내는 ‘최고의’ 접근 방식을 제공하는 기술이 넘쳐나고 있습니다.

이러한 이유로 데이터 레이크는 많은 관심을 받고 있는 참조 아키텍처 중 하나입니다. 하지만 데이터 레이크에 대한 여러 가지 오해로 인해 기업과 IT 관리자는 옵션을 조사하는 과정에서 소중한 시간을 허비하게 됩니다.

1.

데이터 레이크는 단일 클라우드에만 구축 가능하다? → 거짓

데이터 레이크는 단 하나의 장소나 클라우드, 또는 온프레미스에만 국한되지 않습니다. 하이브리드 옵션을 사용하여 여러 클라우드에 걸쳐 데이터 레이크를 구축할 수 있습니다. 데이터 레이크는 기술에 의존하지 않는 참조 아키텍처입니다. 조직에서 데이터 중심 운영을 통해 다양한 데이터 유형의 대규모 액세스를 지원하고 셀프 서비스 분석을 통해 사용자를 지원하기 위해 사용하는 접근 방식입니다.

2.

Hadoop만이 유일한 데이터 레이크이다? → 거짓

데이터 레이크라는 용어는 흔히 Hadoop 또는 Hadoop 지향 객체 스토리지와 관련되어 있지만, Hadoop을 통합하지 않고도 데이터 레이크를 효과적으로 개발 및 사용할 수 있습니다. 예를 들어, 효과적인 데이터 레이크는 서로 다른 관계형 데이터베이스 관리 시스템을 기반으로 할 수 있습니다. 데이터 레이크는 다양한 기술을 통합하여 인사이트를 제공하는 시스템을 구축하고, 이를 통해 데이터 과학자가 비즈니스 니즈를 해결하는 데 필요한 데이터를 민첩하게 탐색할 수 있게 해줍니다.

3.

데이터 레이크를 사용하면 모든 데이터를 덤프할 수 있으며, 거버넌스가 필요 없다? → 거짓

소프트웨어와 하드웨어는 데이터 레이크 솔루션의 핵심 요소이지만, 데이터 카탈로그, 데이터 품질, 데이터 거버넌스, 관리 프로세스 또한 이에 못지 않게 중요합니다.

일부 데이터 웨어하우스가 엄청난 양의 데이터가 갇혀 있는 거대한 블랙홀이 된 것처럼, 우수한 거버넌스 정책이 적용되지 않으면 데이터 레이크는 데이터 늪으로 전락할 수 있습니다.

4.

데이터 레이크의 성공 여부를 판단하는 기준은 액세스다? → 거짓

데이터를 중앙에 모아두는 것은 진정한 분석 솔루션이 아닙니다. 목표는 의미 있는 비즈니스 인사이트를 제공하는 데이터 분석을 실행하고 새로운 수입원, 고객 유지 모델 또는 제품 확장 기회를 발굴하는 것입니다.

하지만 이런 데이터는 모든 사용자가 신뢰할 수 있고 관련성이 있으며 사용할 수 있어야 합니다. 데이터 레이크에는 비즈니스 용어와 관련된 지능형 메타데이터 카탈로그가 있어야만 해독하기 어려운 코딩 데이터를 맥락에 따라 이해할 수 있도록 해줍니다. 또한 정형 및 비정형 정보 자산 및 거버넌스 패브릭의 소스와 품질을 높여 정보 보호, 표준화, 효율적인 관리 및 신뢰성이 보장될 수 있도록 해줍니다.

5.

데이터 레이크는 데이터 웨어하우스의 대안이다? → 거짓

데이터 레이크는 다수의 엔터프라이즈 데이터 웨어하우스(EDW) 뿐 아니라 소셜미디어 또는 IoT 등 다른 데이터 소스를 통합할 수 있습니다. 이 모든 기능이 거버넌스를 내장할 수 있는 데이터 레이크에 통합되어 조직 전반에 걸쳐 사용자가 신뢰할 수 있는 데이터를 손쉽게 검색할 수 있게 해줍니다.

따라서 데이터 레이크는 데이터 과학자와 분석가가 데이터를 손쉽게 탐색하고, 혁신과 비즈니스 성장을 가속화할 수 있는 새로운 관점, 인사이트, 기회를 발견할 수 있도록 EDW 환경을 개선합니다.



멀티클라우드 데이터 레이크의 이점

잘못된 정보에 좌우돼서는 안 됩니다. 관리형 데이터 레이크는 기업 전반에 걸쳐 다양한 정형 및 비정형 데이터에 대한 액세스를 제공하는 동시에 어디에서나 신뢰도와 보안을 보장합니다.

관리형 데이터 레이크는 비즈니스 니즈에 최적화되었을 때 분석을 가속화하고 인사이트의 정확도를 개선할 수 있으며 그 이유는 다음과 같습니다.



데이터가 안전하고 신뢰할 수 있는
인프라에 보관되어 있음.



통제된 데이터 피드가 데이터 레이크를
신뢰할 수 있는 정보로 채우고, 정보
자산, 메타데이터, 비즈니스 컨텍스트를
문서화하여 실시간 데이터 흐름을
제공함.



데이터의 품질, 출처, 계보를 올바르게
파악할 수 있음.



데이터 과학자가 용어의 의미 때문에
어려움을 겪지 않고 즉시 작업에 착수할
수 있도록 데이터를 비즈니스 언어로
저장함.



데이터가 적절하게 분류, 보호, 관리됨.

오해는 이제 그만!

[IBM 데이터 관리 포트폴리오](#)는 고객이 데이터 레이크에 대한 일반적인 문제점과 오해를 피할 수 있도록 지원하며 성공적으로 관리형 데이터 레이크를 구현하는데 필수적인 단계를 안내합니다.

IBM 데이터 레이크 솔루션 알아보기

더 많은 데이터 소스로부터 얻은 다양한 데이터 유형을 활용하여 스마트한 결정을 내리세요.



eBook 읽기

관리형 데이터 레이크로 핵심 비즈니스 인사이트를 도출하는 기회를 창출하는 방법을 확인하세요.

