

# Data and AI Virtual Forum Korea

AI여정에서 필수적으로 고려해야 할 기술기반

—

이지은 상무

Technical Sales Leader  
IBM Korea Cloud & Cognitive Software

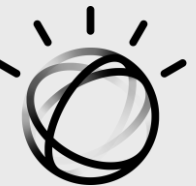
# Q1. AI 프로젝트의 성패, 데이터에 달려있다?

막상 AI 프로젝트를 하려고 하면, 가지고 있는 데이터는 많아도 실제로 필요한 데이터와 활용할 만한 데이터가 부족한 경우가 많습니다.

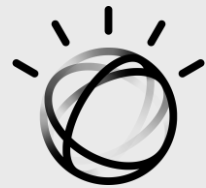
AI 프로젝트가 데이터의 제약없이 성공하려면 기업 내부에 가용한 데이터와 함께 외부의 다양한 정형, 비정형 데이터를 탐색하고 활용할 수 있어야 합니다. 상품 개발에 AI를 적용한 사례의 경우, 내부 데이터 뿐 아니라 외부기관의 시장분석 데이터, 소셜 데이터 등으로 필요한 데이터의 범위를 확장하였고, 이에 따라 AI 분석에 고객성향을 표현하는 다양한 속성을 추가하여 시장의 트렌드를 민감하게 상품 개발에 반영할 수 있었습니다.

또한 AI 프로젝트 결과가 비즈니스에 지속적으로 활용되려면, 이런 데이터의 범위와 깊이를 확장하는 과정이 시스템화 되어야 합니다. 사일로로 존재하는 내부 데이터와 외부로부터 필요한 데이터를 제공하기 위하여 1회성 수작업이 아닌 데이터 획득 및 저장 그리고 필요한 속성을 만드는 과정이 프로세스화 되어 자동화 되어 매달, 매주, 매일, 등 필요한 주기로 이루어지도록 해야 AI 분석은 물론 AI 적용 및 효과 모니터링까지 원활하게 이루어집니다.

AI를 위한 다양하고 복잡한 데이터 환경을 대응하기 위하여 물리적인 데이터 통합 뿐 아니라 데이터 양, 필요한 주기등을 고려하여 필요한 경우 데이터 가상화 기술을 함께 고려하여 데이터에 대한 접근성을 높여 다양한 데이터를 AI 프로젝트에 적극활용 합니다.



# 가상화 기술

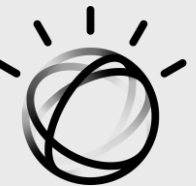


## Q2. 어떻게 신뢰할 수 있는 데이터를 확보할 수 있나?

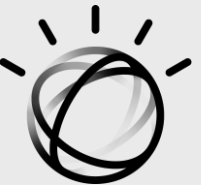
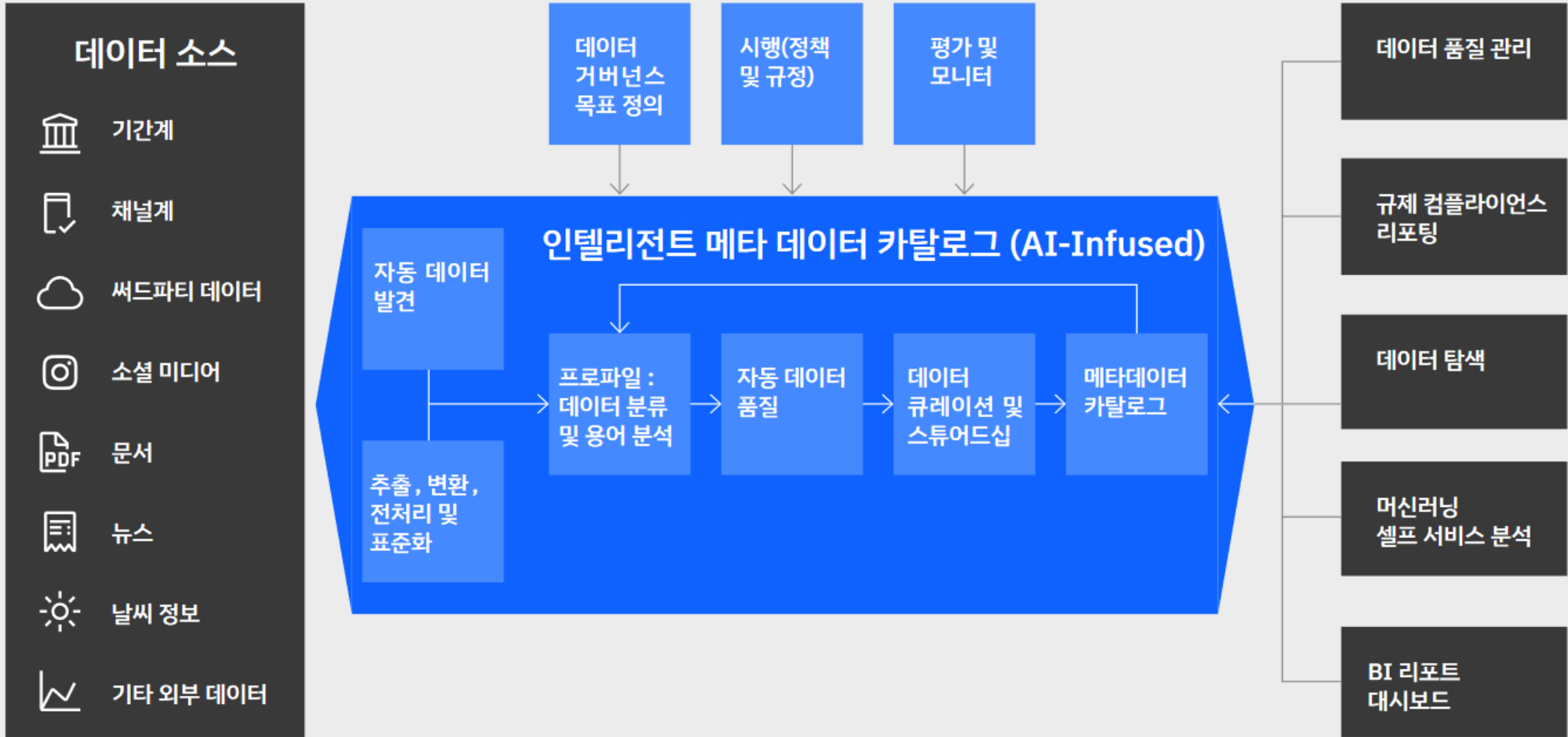
AI 프로젝트에서는 여러 AI use case 에 공통으로 필요한 데이터를 정의해야 하기도 하고 , 일회성이 아니라 지속적으로 제공되어야 하기때문에 , 데이터를 각각 사일로가 아닌 전사 관점으로 정리하여 신뢰할 수 있는 데이터를 투명한 프로세스로 제공할 수 있어야 합니다.

신뢰할 수 있는 데이터를 제공하고 유지하기 위하여 데이터 카탈로그를 생성하고 관리하는 것이 중요 합니다. 데이터 카탈로그에는 모든 데이터에 대한 위치와 DB,테이블, 컬럼 정보와 같은 기술메타뿐 아니라, 데이터의 비즈니스 의미를 모르면 분석 결과를 활용할 수 없으므로 비즈니스 용어를 정리하여 실제 데이터와 매핑하여 보여줄수 있어야 합니다. 그래야 AI 결과를 신뢰하고 잘 활용할 수 있습니다. 또한 비즈니스에 의미 없는 데이터가 원천적으로 들어오지 못하도록 특정 컬럼에 들어올 수 있는 데이터에 대한 규칙과 정책을 카탈로그에 정의하고 계속 들어오는 데이터에 대하여 적용하여 데이터 품질을 지속적으로 관리가능하게 됩니다. 많은 고객분들이 데이터 거버넌스에서 비즈니스 용어 관리의 중요성은 알고 있으나, 비즈니스 용어 정리가 잘 안되어 있거나 , 정리되어 있다고는 하나 계속 변화하는 데이터 환경에 대응하기 위한 노력이 너무 많이 들어지속적인 관리의 어려움을 얘기 하기도 합니다.

자동 메타 데이터 생성과 관리 기술이 점점 중요해지는데 , 최근에는 ML/DL 기술을 접목하여 비즈니스 용어를 자동 추천하고, 채택여부가 다시 학습되어, 추천의 정확도를 계속 높여 카탈로그에 지속적으로 반영할 수 있습니다.



# 데이터 카탈로그



# Q3. Talent, Trust, Data라는 도전과제를 극복하기 위해 갖추어야 하는 기술 기반은?

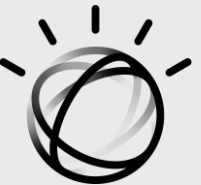
사용자 관리, 보안, 모니터링, 필요한 분석 환경에 대한 프로비저닝과 같은 공통 서비스와 함께 데이터 액세스부터 데이터 거버넌스, AI 분석 과 활용 까지 AI 라이프 사이클을 공통 플랫폼기반으로 유기적으로 지원해야되는거죠. 플랫폼위에 가용한 모든 데이터 에 대해 그 유형과 위치, 저장 방식에 상관없이 접근가능하도록하는 장치가 필요하고, 데이터에 대한 전처리 과정을 자동화하고, 가용한 데이터에 대해 검색이 가능하도록 데이터 카탈로그를 관리하게 되면, 다양한 AI 유즈케이스 가 필요로하는 데이터 액세스에 대한 노력과 비용이 감소하게 됩니다.

다수 사용자를 위한 협업 플랫폼을 통하여 AI 를 개개인의역량에 의존하기 보다는 AI 경험을 자산으로 관리하여 기업내 기술 역량 내재화로 이어지면 안정적인 AI 기술 공급이 가능하게 됩니다.

또한 AI 결과를 업무에 적용해보고 플랫폼 상에서 AI 결과의 정확도에 대한 주기적인 모니터링과 피드백을 하게 되면 신뢰할 수 있는 AI 를 지속적으로 실 업무에 적용할 수 있습니다. 최근 AI 기술은 오픈 소스 중심으로 다양 한 기술들이 제공되므로 플랫폼에서 각종 오픈소스를 지원해야 하겠죠. 다만 오픈소스를 개인용도가 아닌 기업업무에 사용해야 하므로, 안정적으로 안전하게 기업에서 활용하기 위한 사전 테스트와 기술 지원 체계가 갖춰진 플랫폼이 필요합니다.

특정 AI 기술 을 사용하기 위하여 특정 클라우드 플랫폼만을 사용해야 한다면 구축 후 다른 환경으로의 이동이 쉽지 않아 큰 제약이 되기 때문에 Private, On-Prem , Public 을 포함한 다양한 Multi Cloud 환경에서 운영할 수 있는 쿠버네티스기반의 컨테이너 환경과 같은 개방형 클라우드 기술을 고려해야 합니다.

IBM 은 기업이 AI 를 제대로 활용할 수 있도록 체계적인 접근방법으로 AI Ladder 를 제시해왔습니다. AI Ladder 는 데이터 액세스 부터 거버넌스, 분석 및 적용에 이르는 AI Full Lifecycle 을 구현하기 위한 단계별 접근법과 기술을 포함합니다 . 이를 실현하는 플랫폼으로 Cloud Pak for Data 를 제공하고 있습니다. 트랙세션에서 AI Ladder 개념과 Cloud Pak for Data 의 상세 기술에 대하여 만나보실수 있습니다.



# AI 도전 과제

## 데이터

AI의 핵심이지만 복잡성이  
발전을 저해함

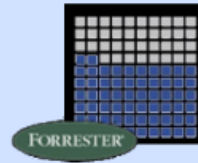


60%

데이터 품질 관리에 어려움을  
겪습니다

## 인재

AI 인력 수요는 증가하나  
인재 부족

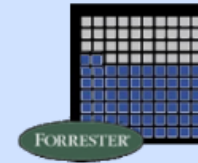


62%

인재 확보 및 스킬 구축에  
어려움을 겪습니다

## 신뢰

AI 시스템 및 프로세스에  
대한 회의적인 시각 존재



62%

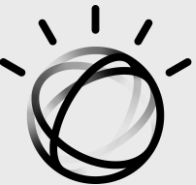
AI 프로덕션 준비를 위한  
접근방법이 필요합니다

Stuck in  
Experimentation

51%

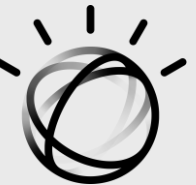
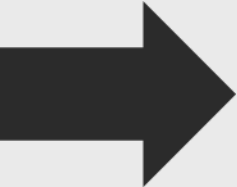
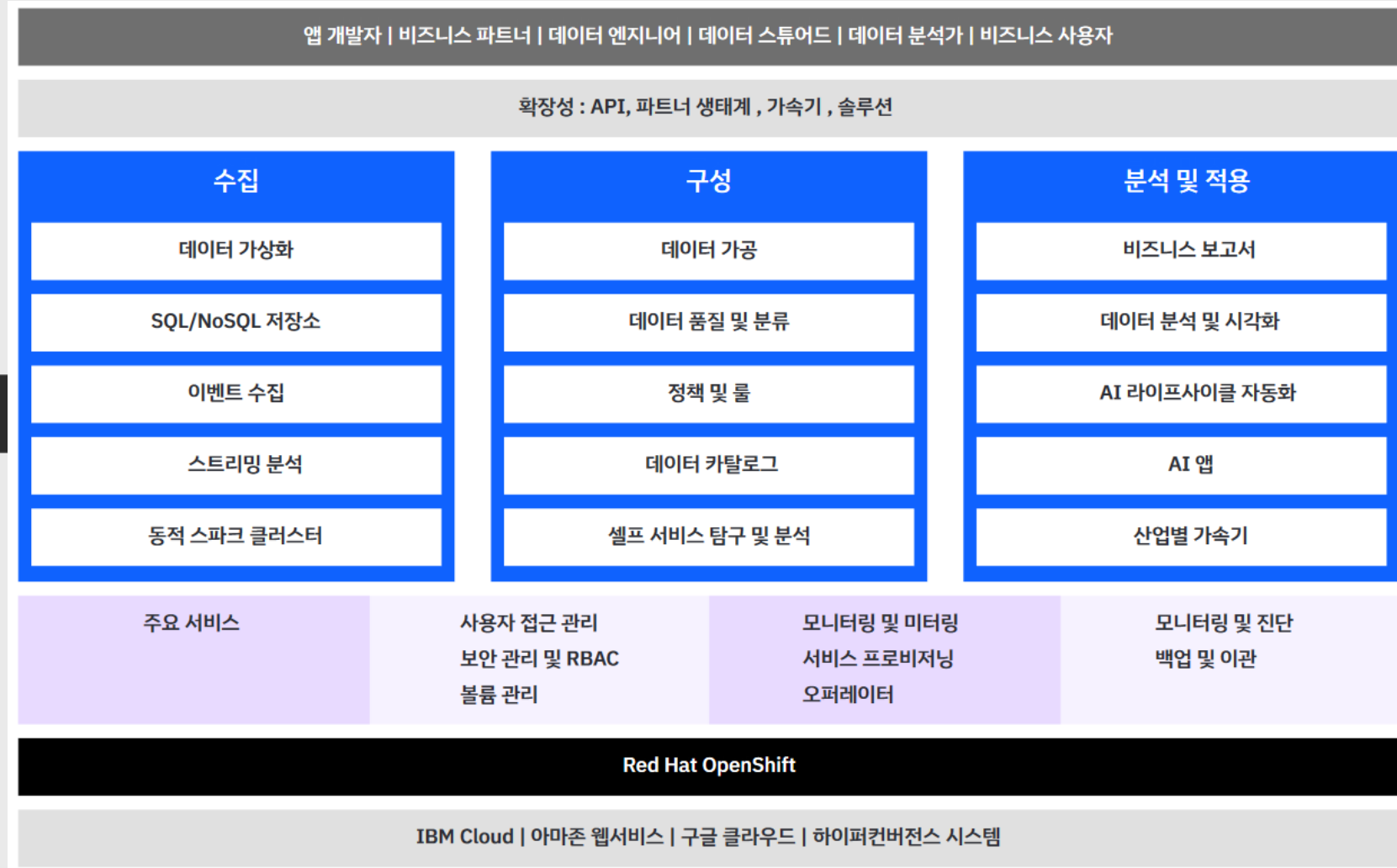
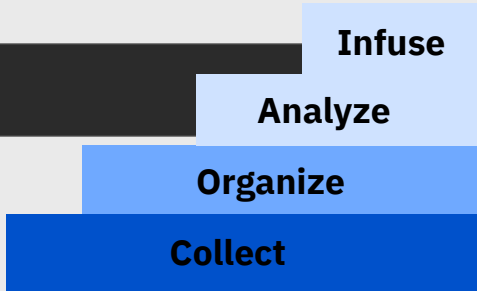
AI의 운영, 유지 및 확장에  
어려움을 겪습니다

출처 : 2019 Forrester 보고서 , Challenges That Hold Firms Back From Achieving AI Aspirations



# 클라우드 기반의 현대화된 데이터-AI 플랫폼

## The AI Ladder





# 기업의 디지털 혁신을 지원하기 위한 IBM의 AI 기술 진화 방향은?

#1. AI 기술을 어디서든 활용할 수 있도록 AI Everywhere를 지향합니다. 있는 곳이라면 Public , On-prem, Private 환경 어디서든 AI 기술이 활용될 수 있도록 하는 것을 목표로 합니다.

#2. 오픈소스 기반 AI 플랫폼을 지원하는 것으로 지속적으로 AI 모델을 개발하고 개발된 AI 서비스를 실제 업무 운영에서 실행하고 제대로 작동하고 있는지 지속해서 모니터링 하는 것입니다.

#3. AI 기술의 신뢰성 강화를 위해 Trust AI에 힘을 쏟을 것입니다. 설명가능한 AI 기술등을 통하여 기업들이 AI 를 실 업무에 적극 활용할 수 있도록 합니다.

