

Why Databricks Lakehouse on Azure

데이터에 가치를 더하여 고객의 성장에 공헌합니다.

Specialized Consulting Firm in **Data & AI** Cloud System



1. Why Big Data Platform ?
2. Why Databricks Lakehouse ?
3. SAP Data Migration to Databricks
4. Power Platform for SAP
5. 적용 및 고객 지원 서비스
6. 엠클라우드브리지 소개

1. Why Big Data Platform ?

1. Why Big Data Platform ?

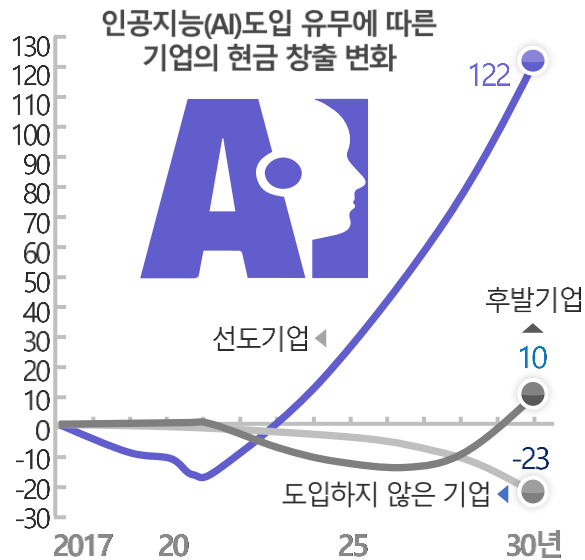
데이터 분석 기반 AI 도입이 미래 기업의 경쟁력 결정

빅데이터 플랫폼(Big Data Platform)의 도입은 데이터 분석 기반 경영 환경 구축을 통한 기업의 미래 경쟁력을 결정하는 필수 요소이며, 이를 적용함으로써 기업은 데이터 분석 기반 경영 의사 결정에 요구되는 내/외부의 정형/반정형/비정형 형태의 모든 데이터 통합을 확보 할 수 있습니다.

데이터 분석 기반 AI 도입이 미래 기업의 경쟁력 결정

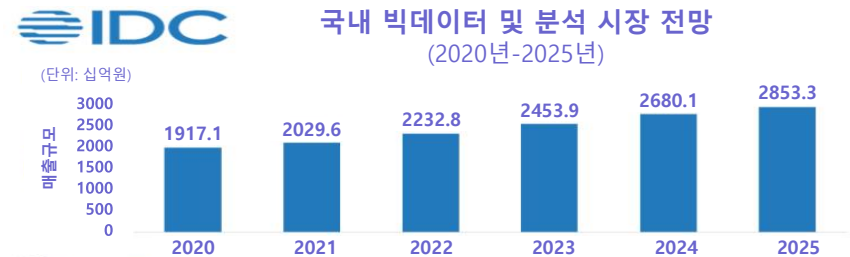
**AI 성공적 도입한 선두기업
10%만 살아남을 것**

(출처: 경영 컨설팅 기업 맥킨지, 단위: %)

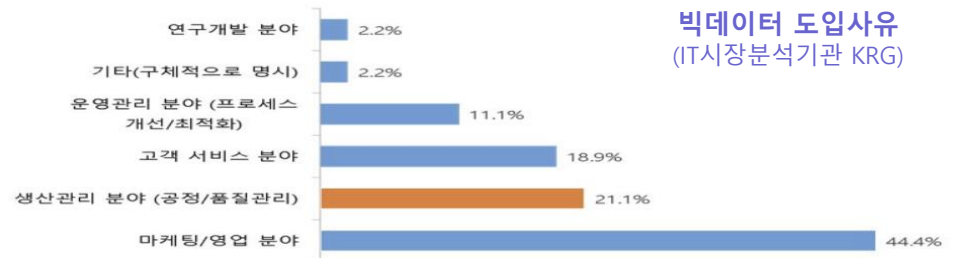


**국내 기업들 지속적인 투자 확대 중
(미래 경쟁력 제고와 생존의 기본 조건 인식)**

출처: IDC & IT시장분석기관 KRG



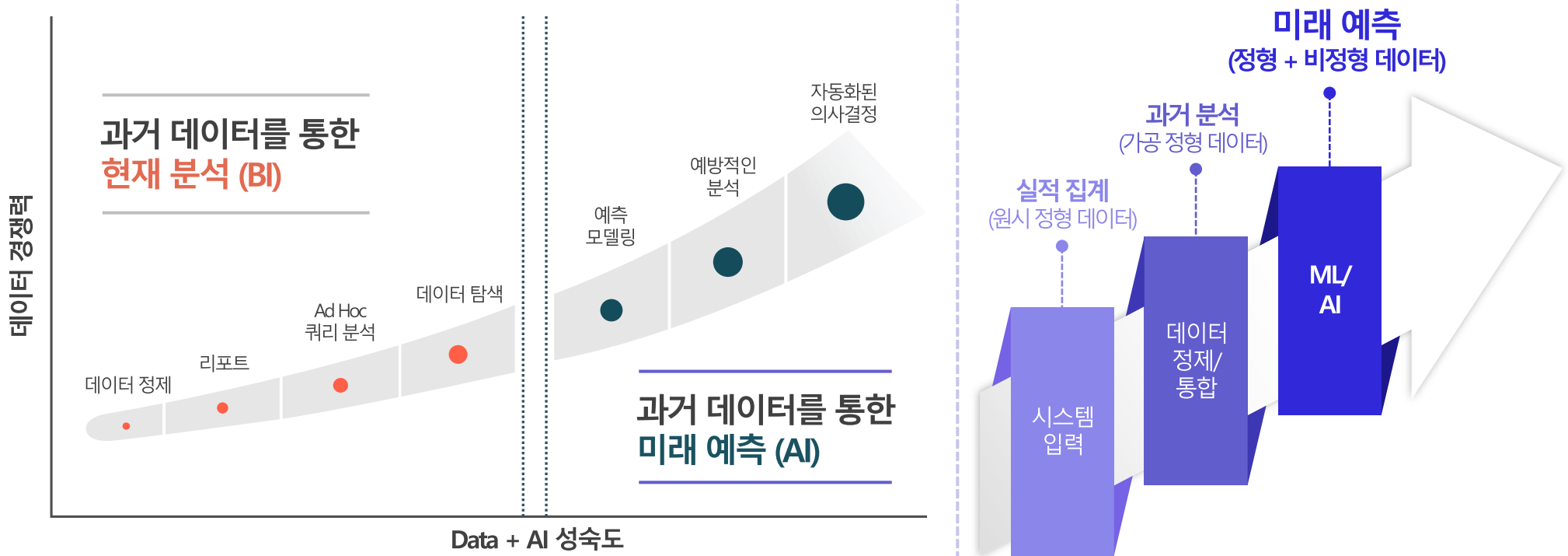
Source: IDC Semiannual Big Data and Analytics Tracker, September 2021



빅데이터 플랫폼(Big Data Platform)의 구축은 기업의 데이터 분석 환경이 과거의 경영 실적 분석에서 실시간 경영 상황이 반영된 미래 예측 및 대응 방안 마련 등으로 변화 하고 있으므로, 기업 활동에 관련된 과거 및 현재 정보를 실시간 반영 할 수 있는 데이터 분석 환경에 필수적 요소입니다.

미래 예측 시스템의 기반이 되는 빅데이터 플랫폼(Big Data Platform) 필요성

데이터 성숙에 따른 활용



빅데이터 플랫폼(Big Data Platform)을 통하여 미래 예측을 하기 위해서는 기업 내부의 가공된 정형데이터 확보를 넘어, 기업 내부/외부에 존재하는 정형/반정형/비정형 및 원천 데이터를 저장하여 미래 예측에 활용 할 수 있도록 모든 데이터의 관리 기반을 마련하는 것이 필수 요소입니다.

데이터 분석 기반 경영 환경을 위한 빅데이터 플랫폼(Bigdata Platform) 요건



대량의 자연어 데이터를 이용하여 미래 예측이 가능한
빅데이터 플랫폼

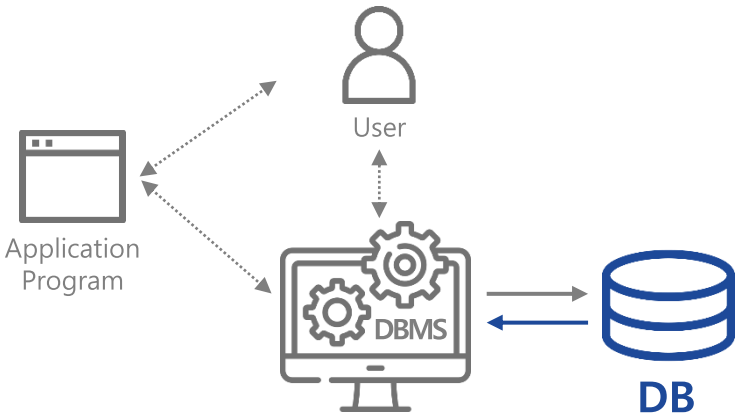
1. Why Big Data Platform ?

모든 데이터 형태가 처리 가능한 빅데이터 시스템 필요

빅데이터 플랫폼(Big Data Platform)은 기존의 데이터 베이스화가 가능한 정형화된 회사 내부 경영정보 관리에서 미래 예측을 위한 데이터 분석 기반 경영을 위하여는 정형 데이터 및 Log, IoT, SNS 등 정형화 되어 있지 않은 모든 데이터를 취합, 정제, 통합하는 데이터 시스템이 필수적입니다.

① 정형/비정형 등 모든 데이터 형태가 처리 가능한 빅데이터 시스템 필요

과거 - 정형 데이터



- 데이터 베이스화 하기에 가능한 정형 정보
- ERP 및 MIS 용으로 가공이 가능한 정보
- 항상 동일한 형태로 발생하는 정보

미래 - 정형/비정형 데이터



- 분석 및 예측을 위한 모든 형태의 데이터
- 로그, IoT, SNS 등 필요한 모든 형태의 데이터
- 업무 상황에 따라 가변적으로 발생하는 모든 데이터

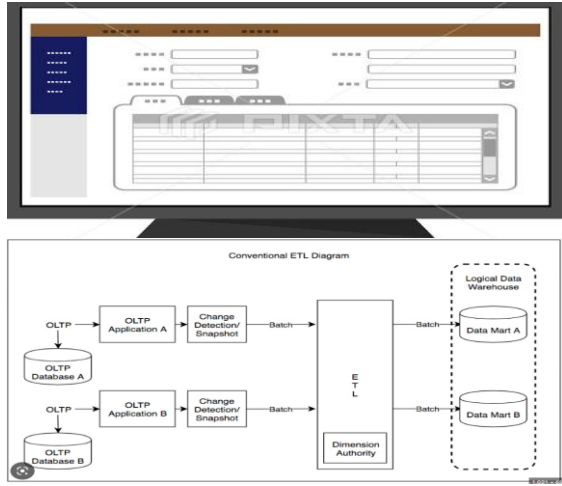
1. Why Big Data Platform ?

모든 데이터 소스가 처리 가능한 빅데이터 시스템 필요

빅데이터 플랫폼(Big Data Platform)은 기존 기업이 내부의 제한된 정보를 가공하여 분석에 사용하게 됨으로써 정보 분석 범위가 제한 되었으나, 데이터 분석 기반 경영 예측 시스템을 위하여는 가공 되지 않은 정보를 저장하고 필요에 따라 사용하기 위한 데이터 분석 시스템 도입은 필수적입니다.

② 가공/원천 등 모든 데이터 소스가 처리 가능한 빅데이터 시스템 필요

과거 - 가공된 소량 데이터



- 정보시스템을 통해 입력되며 가공된 데이터
- 필터링 되거나 클렌징 된 데이터
- ETL 작업을 통해 D/W or Data Mart로 가공된 데이터

미래 - 가공/원천 대량 데이터



- 분석 및 예측을 위한 대용량의 원시 데이터
- 오류 및 Risk 방지를 위한 전체 데이터
- 대용량의 실시간 발생 데이터

1. Why Big Data Platform ?

모든 위치의 데이터가 처리 가능한 빅데이터 시스템 필요

빅데이터 플랫폼(Big Data Platform)은 기존에 기업이 주요 내부 정보만을 제한적으로 저장하게 됨으로써 현재와 미래 판단에 필요한 연관된 내/외부 정보를 활용에 한계가 있었으나, 데이터 분석 기반 경영 예측 시스템의 도입을 위하여는 기업/내외 부의 데이터 관리 시스템은 필수적입니다.

③ 회사 내부/외부 등 모든 위치의 데이터가 처리 가능한 빅데이터 시스템 필요

과거 - 회사 내부 데이터




- 회사 내부 시스템에 입력된 데이터
- 입력 시 정보시스템에 필터링 된 데이터
- 실적 중심의 일차 가공 데이터

미래 - 회사 내부/외부 데이터



- 회사 정보가 발생된 모든 데이터
- 회사 내외부에서 발생한 가공되지 않은 데이터
- Risk 예측 및 미래 분석 등 활용성 높은 데이터

2. Why Databricks Lakehouse ?

Databricks Lakehouse는 Databricks SQL, Serverless, Unity Catalog, Streaming, ML 등의 핵심 기술을 기반으로 데이터 엔지니어링(Data Engineering), 데이터 분석(Data Analytics) 그리고 데이터 사이언스(AI/ML)를 단일 데이터 플랫폼에서 제공하여, 비용, 가버넌스 그리고 데이터 활용의 최선의 제품입니다.

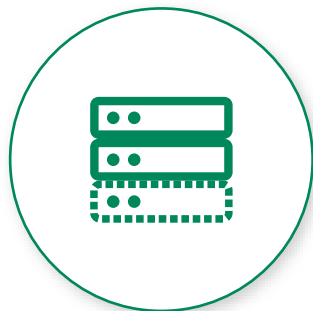
One Data Platform for Data Engineer , AI

Databricks's mission is to accelerate innovation for its customers by unifying Data Science, Engineering and Business.

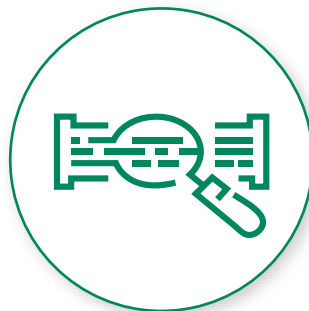
Databricks SQL



Serverless



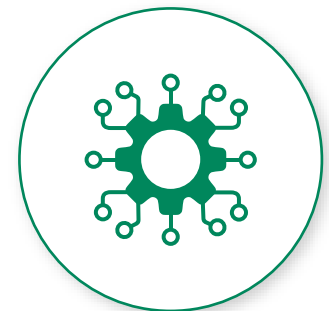
Unity Catalog



Streaming

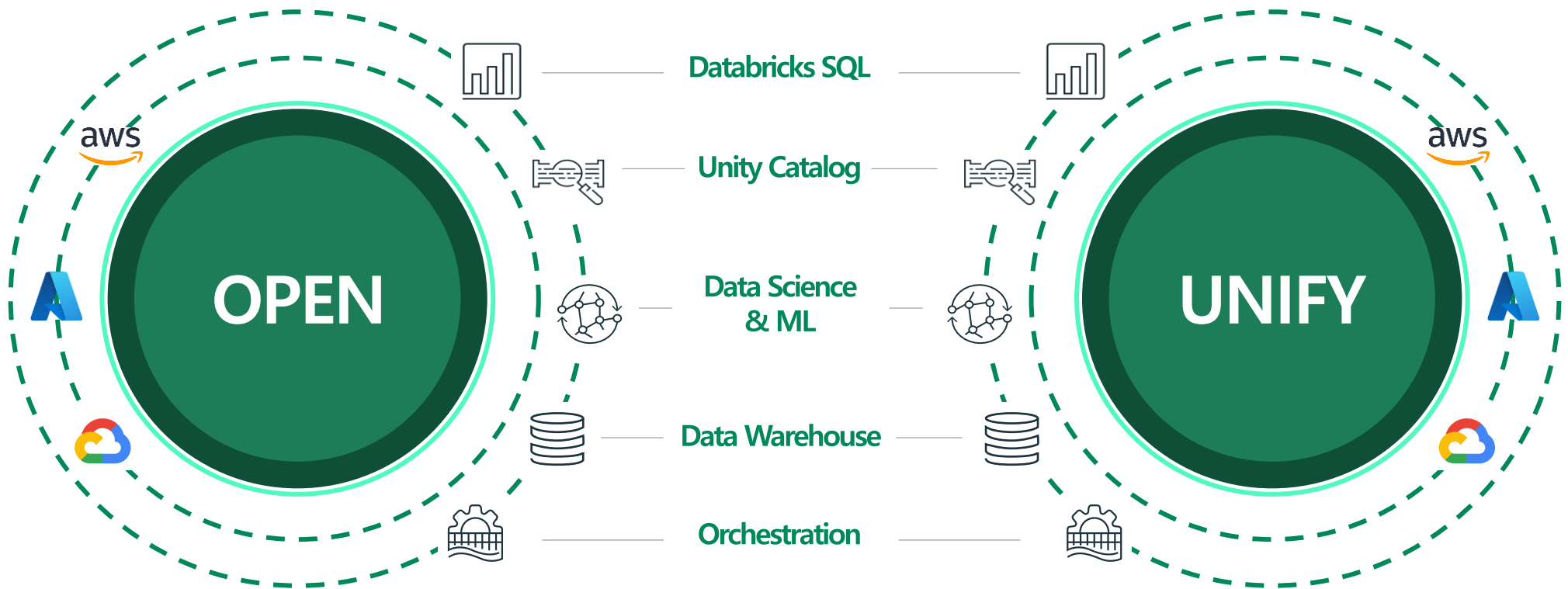


Restructured ML



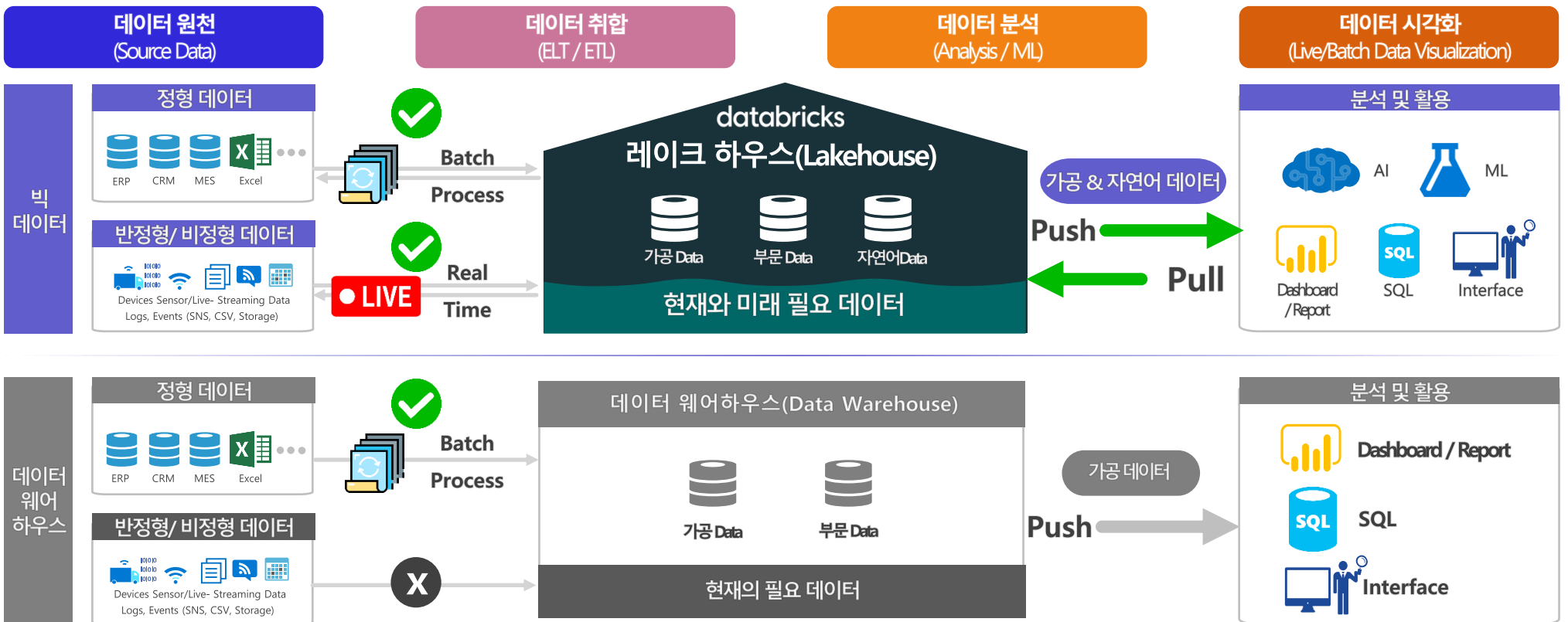
Databricks Lakehouse는 마이크로소프트(Microsoft), 아마존(AWS) 및 구글(Google) 등의 주요 3사 클라우드 플랫폼 서비스를 제공하고, 카탈로그, 머신러닝, 스트리밍, 델타 레이크 등의 외부 데이터 시스템과 시너지 효과를 내기 위하여, 개방성(Open)과 통합성(Unify)을 기반의 빅데이터 플랫폼입니다.

개방성(Open) & 통합성(Unify) 데이터 플랫폼



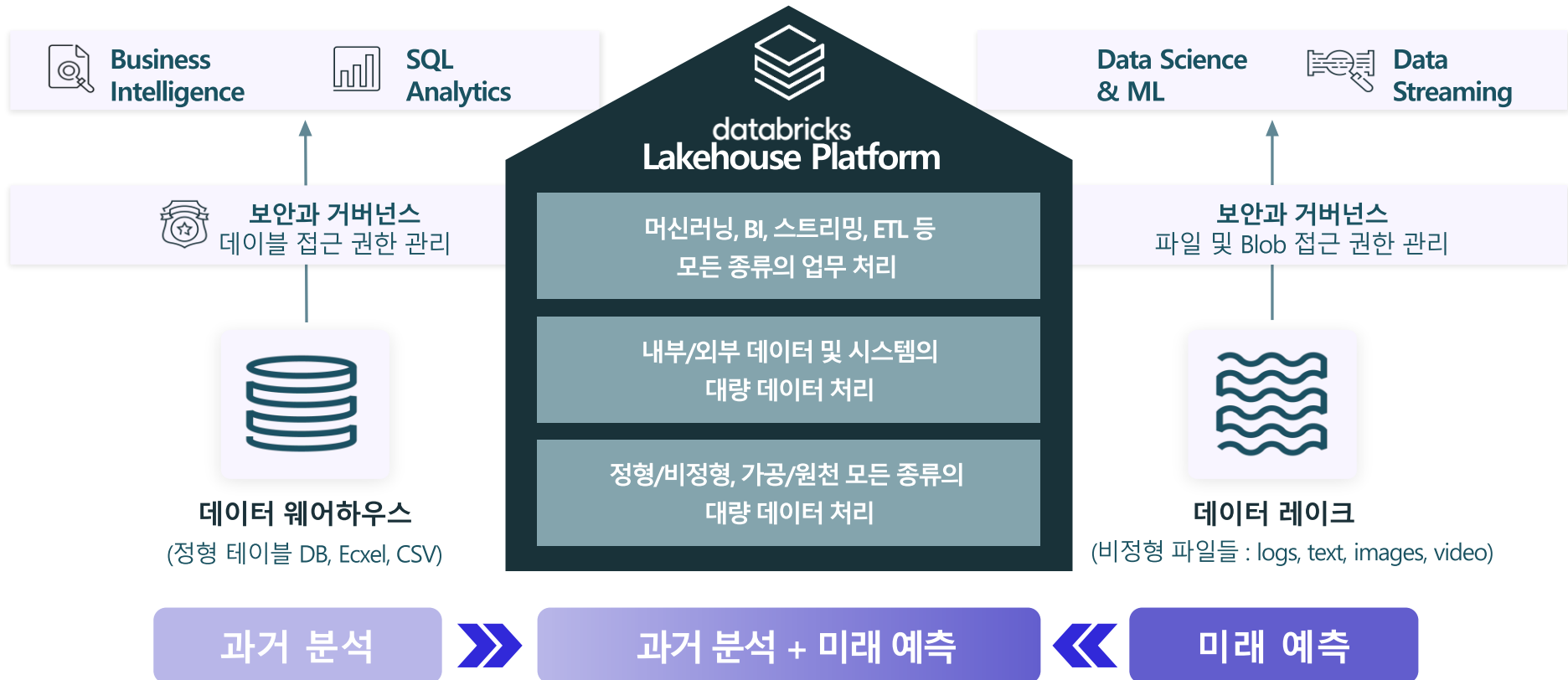
Databricks Lakehouse은 기업 내/외부의 분산된 정형, 반정형 및 비정형 데이터를 통합하는 빅데이터 플랫폼(Big Data Platform)으로써 가공 및 자연어 데이터의 실시간 수집 및 분석과 활용을 가능하게 하여 기존 데이터 웨어하우스(Data Warehouse)가 가진 데이터 분석 환경의 한계 보완합니다.

Databricks Lakehouse vs. Data Warehouse



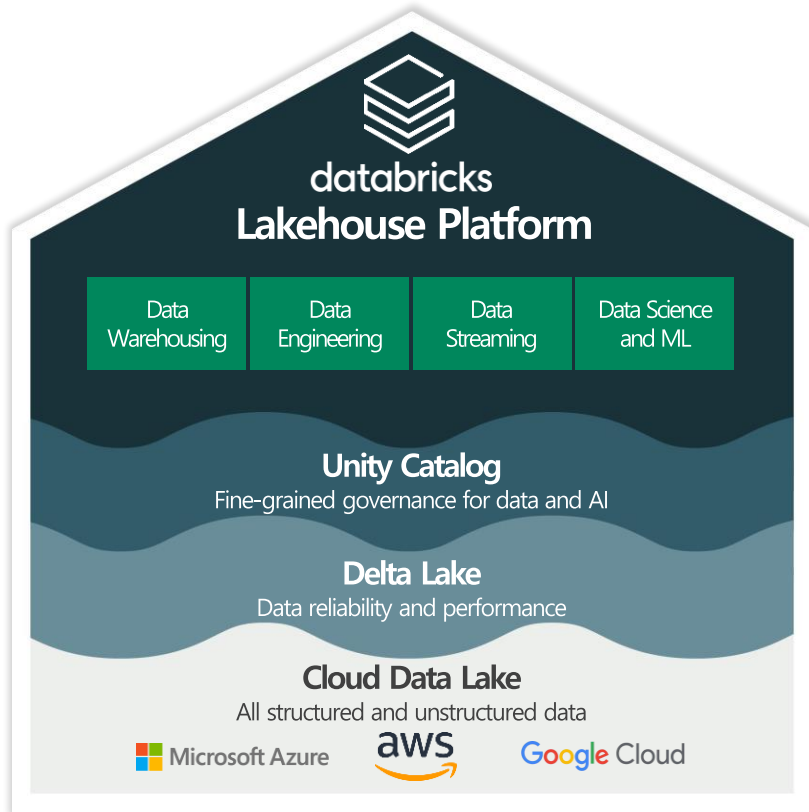
Databricks Lakehouse를 통해 기업은 단일 플랫폼 내에서 정형/비정형, 가공/원천 등 모든 종류의 데이터를 활용하여 ETL, 머신러닝, BI, 스트리밍 등 다양한 업무 처리가 가능해져 단일 보안, 거버넌스 모델로 데이터를 통합하여 관리하는 빅 데이터 플랫폼(Bigdata Platform)을 구성 할 수 있습니다.

빅데이터 관리 및 AI/ML 분석 기능이 제공되는 통합 플랫폼



Databricks Lakehouse는 데이터 저장 및 분석, AI/ML 등이 하나의 통합된 플랫폼에서 개발 되었으며, 오픈 소스, 국제 표준 형식으로 개발되어 편리하게 타 시스템과 연결할 수 있고 데이터 애널리스트, 엔지니어, 사이언티스트 등이 함께 데이터와 모델을 공유하며 협업해서 관리할 수 있습니다.

Databricks Lakehouse 업무적 특징점 소개



- ① 단순성 (Simple)**

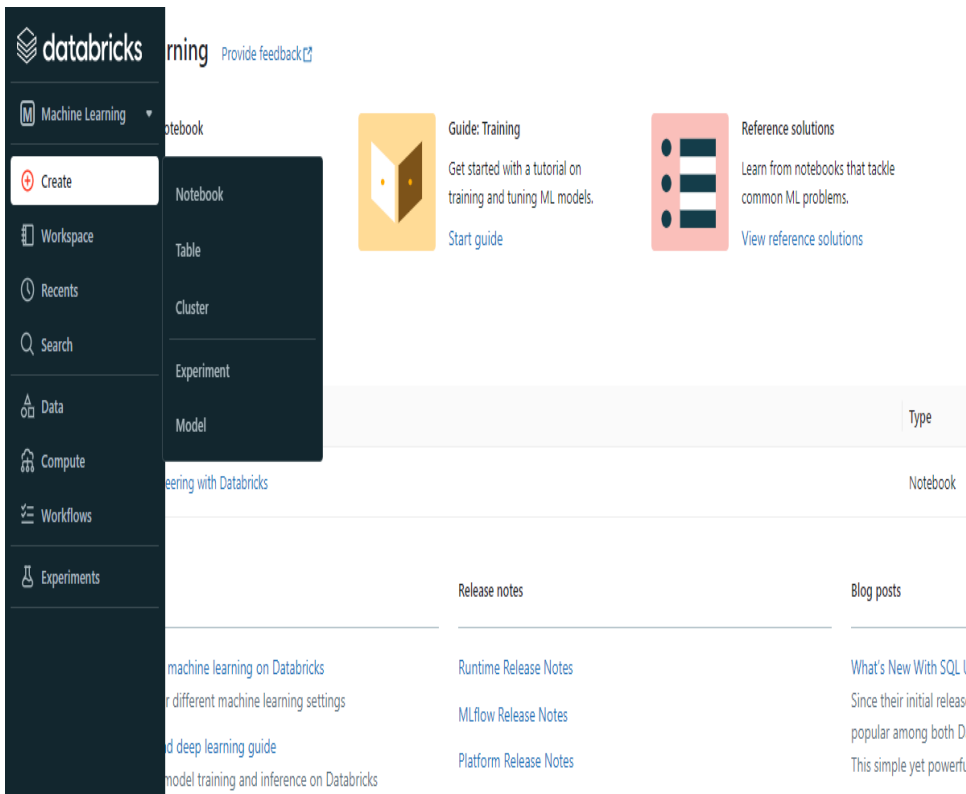
하나의 플랫폼에서 데이터 엔지니어링 및 ML, 분석 등 모든 종류의 업무 처리
- ② 개방성(Open)**

오픈 소스, 표준 형식으로 개발 되어 많은 종류의 시스템 및 개발 플랫폼과 연결 가능
- ③ 협업 (Collaborative)**

데이터 애널리스트, 데이터 엔지니어, 데이터 사이언티스트 등이 하나의 플랫폼에서 업무 처리

Databricks Lakehouse는 단일 플랫폼 내에서 보안 및 시스템 통합 관리와 데이터 생성 및 관리, 데이터 거버넌스 관리, 클라우드 설정(Cloud Configuration), ETL 작업 및 모니터링 등 모든 업무 처리가 가능함으로써 개발 및 관리를 단순화 하여 업무 효율성을 극대화할 수 있도록 지원합니다.

단일 플랫폼 내에서 모든 업무 처리가 가능한 단순성

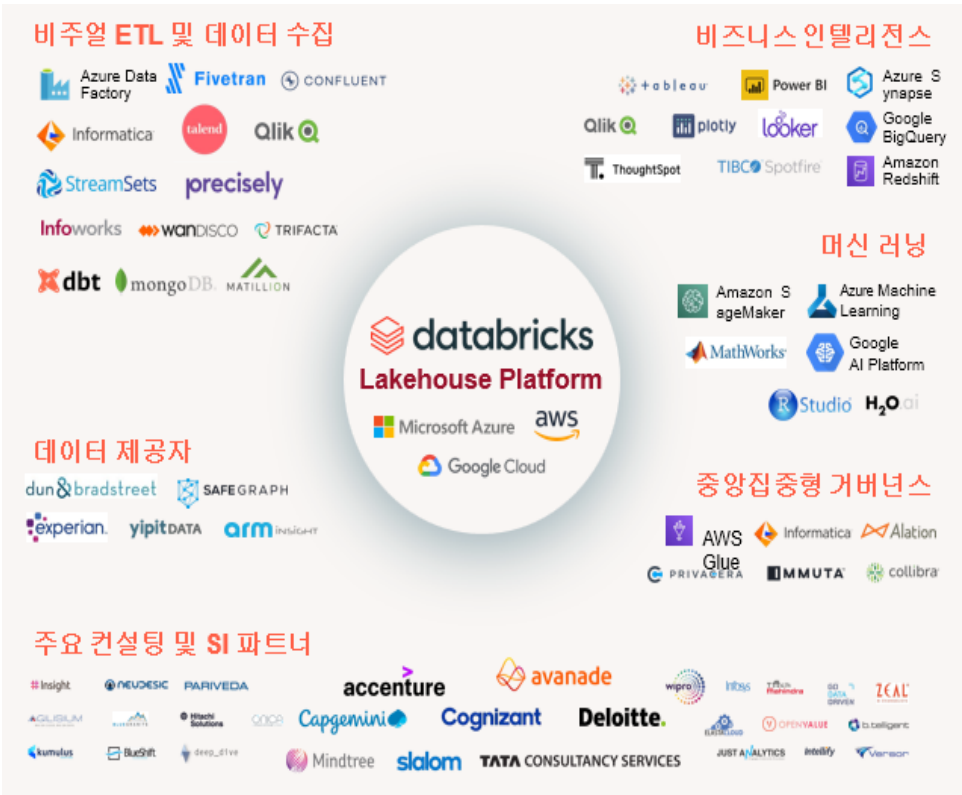


① 단순성 (Simple)

- 단일 플랫폼 내에서 보안 및 시스템 통합 관리
- 단일 플랫폼 내에서 데이터 생성 및 관리
- 단일 플랫폼 내에서 데이터 거버넌스
- 단일 플랫폼 내에서 Cloud Configuration 관리
- 단일 플랫폼 내에서 ETL 작업 및 모니터링

Databricks Lakehouse는 오픈소스 형식으로 개발되어 편리한 인터페이스가 가능하며 관련 정보를 데이터 수집과 BI, 머신러닝 등 여러 시스템 (450개 데이터 영역 파트너)과 호환하여 사용할 수 있고 안정성과 개방성을 확보하여 사용자는 필요한 데이터를 유연하게 개발하며 관리 할 수 있습니다.

인터페이스 및 ETL 작업 등 편리한 시스템 개방성



비주얼 ETL 및 데이터 수집

Azure Data Factory, Fivetran, CONFLUENT, Informatica, talend, Qlik, StreamSets, precisely, Infoworks, WANDISCO, TRIFACTA, dbt, mongoDB, MATILLION

비즈니스인텔리전스

Tableau, Power BI, Azure Synapse, Qlik, plotly, Looker, Google BigQuery, ThoughtSpot, TIBCO Spotfire, Amazon Redshift

머신 러닝

Amazon SageMaker, Azure Machine Learning, MathWorks, Google AI Platform, Studio, H2O.ai

데이터 제공자


dun & Bradstreet, SAFEGRAPH, Experian, YipitData, ARM INSIGHT

중양집중형 거버넌스

AWS Glue, Informatica, Alation, PRIVACERA, IMMUTA, collibra

주요 컨설팅 및 SI 파트너

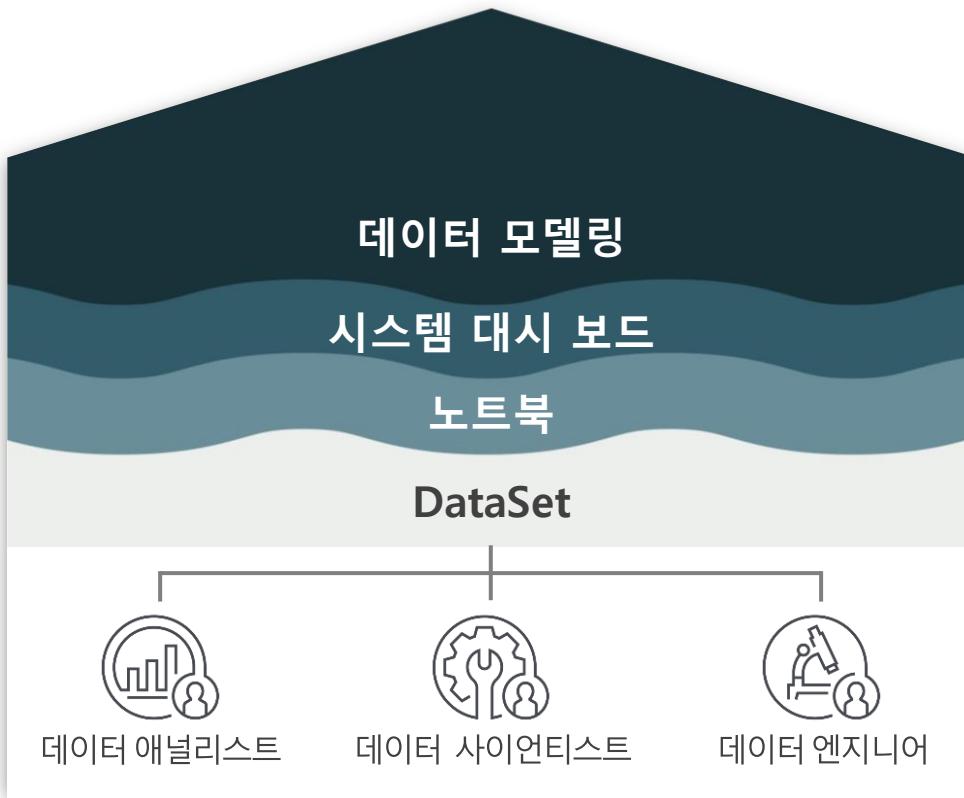
Insight, NEUESIC, PARIVEDA, accenture, avanade, wipro, Infosys, TCS, SAP, ZCAL, ACULUM, Hitachi Solutions, Capgemini, Cognizant, Deloitte, kumulus, BluShift, deep_drive, Mindtree, slalom, TATA CONSULTANCY SERVICES, JUST ANALYTICS, Intely, Vericon


 Databricks Lakehouse Platform
 Microsoft Azure, AWS, Google Cloud

- ### ② 개방성(Open)
- 오픈소스 형식으로 편리한 인터페이스
 - 많은 종류의 ETL 및 데이터 수집 기능
 - 3사 Cloud Service 에서 사용 가능
 - BI 및 머신러닝에서 여러 시스템과 호환
 - 450개 이상의 데이터 영역 파트너

Databricks Lakehouse는 데이터 애널리스트, 데이터 사이언티스트, 데이터 엔지니어가 하나의 플랫폼에서 협업하며 ETL/BI/ML 작업과 데이터 모델링 공유 및 대시보드/노트북을 이용한 정보 공유와 이력 관리 작업 및 편리한 데이터셋(Dataset) 생성과 스케줄 관리를 수행할 수 있도록 지원합니다.

데이터를 담당하는 모든 작업자가 단일 플랫폼에서 협업

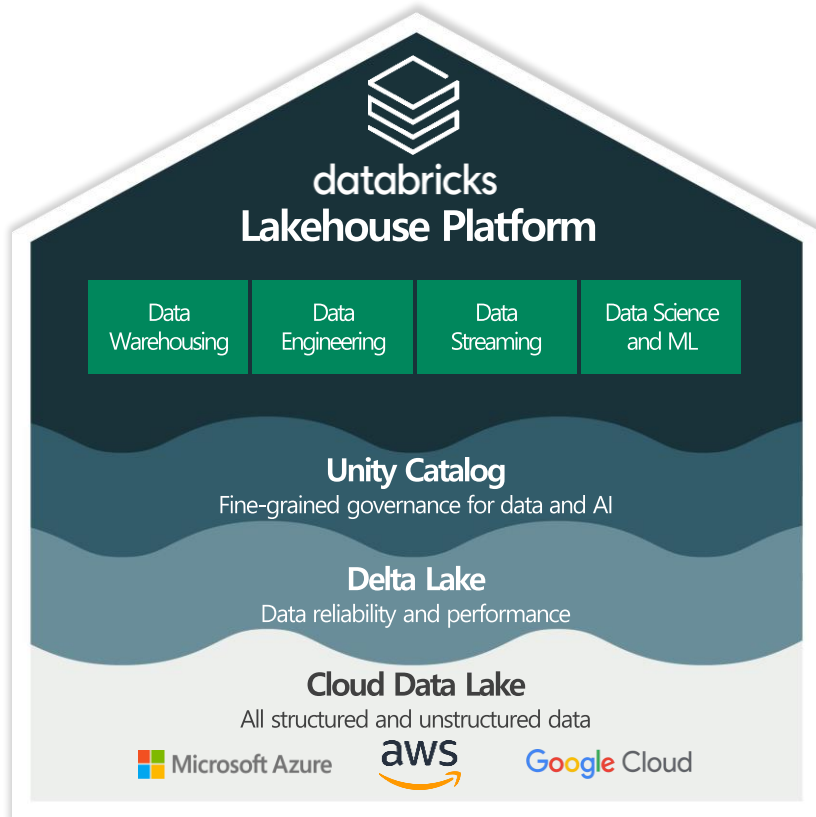


③ 협업 (Collaborative)

- 하나의 플랫폼에서 ETL, BI, ML 작업
- 머신러닝을 통한 데이터 모델링 공유
- 모든 사용자가 대시보드를 통한 정보 공유
- 노트북을 통한 프로그램 공유 및 이력 관리
- 편리한 Dataset 생성 및 스케줄 관리

Databricks Lakehouse는 프레임워크 역할을 하는 델타 레이크(Delta Lake)를 포함한 7가지 기능을 통하여 강력한 데이터 거버넌스 구축 및 데이터 통합 관리, 공유, 머신 러닝, 워크로드 구축 및 자동화 등을 제공함으로써 시스템을 단순하고 효율적으로 운영하여 더 빠르게 혁신할 수 있습니다.

Databricks Lakehouse 기능 소개



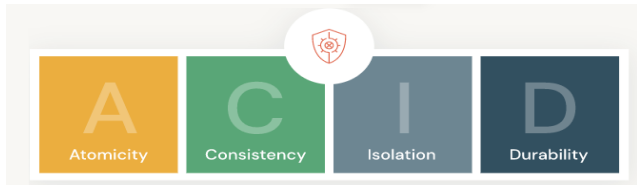
- ① **Delta Lake**
레이크하우스의 프레임워크
- ② **Unity Catalog**
통합 데이터 검색 및 보안, 권한 관리
- ③ **Data Engineering**
ETL 작업과 워크플로 오케스트레이션 기능
- ④ **Delta Sharing**
데이터를 중앙에서 관리하고 공유
- ⑤ **Machine Learning**
데이터 모델링부터 실험, 프로덕션을 관리
- ⑥ **Data Science**
안정적인 데이터 관리 및 분석을 지원
- ⑦ **Data Streaming**
스트리밍 데이터 워크로드 구축, 관리를 자동화

Databricks 델타 레이크 (Delta Lake)는 레이크하우스의 프레임워크 기능을 수행 할 수 있도록 모든 기능을 하나의 품으로 대체하여 데이터의 인사이트가 확대 되었으며, 공유데이터의 중앙 관리, 초고속 성능, 자동화 데이터 엔지니어링 등으로 데이터 관리 효율성을 높일 수 있도록 지원합니다.

① 레이크하우스의 프레임워크 델타 레이크 (Delta Lake)

데이터 인사이트 확대

Delta Lake는 ACID 트랜잭션과 스키마 적용을 지원하여 전사적으로 믿을 수 있는 데이터 인사이트를 확장하고, 데이터 레이크에서 직접 분석 및 여타 데이터 프로젝트를 실행하면서도 인사이트 도출 시간을 최대 50배나 단축할 수 있습니다.



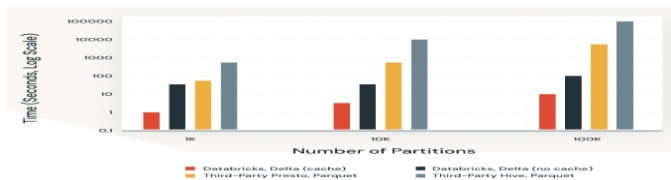
공유 데이터의 중앙 관리

데이터 위치에 구애받지 않고 다른 조직과 간편하게 데이터를 공유할 수 있으며 Unity Catalog와 기본적으로 통합되기 때문에 전사적으로 공유 데이터를 중앙에서 관리하고 감사할 수 있습니다



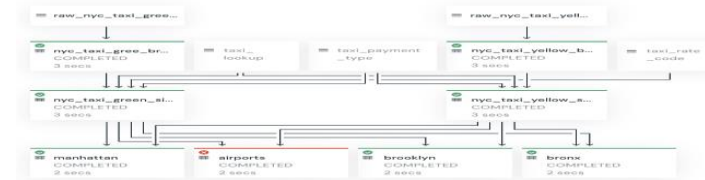
초고속 성능

Delta Lake는 Apache Spark™ 기반으로, 뛰어난 확장성과 속도를 제공하며 인덱싱과 같은 성능 기능을 염두에 두고 최적화하였기 때문에 최대 48%까지 ETL 워크로드 실행 속도가 빨라지는 효과를 만들 수 있습니다.



자동화 데이터 엔지니어링

Delta Lake 테이블은 선언적 파이프라인 개발, 데이터 안정성 개선 및 클라우드 규모 프로덕션 작업을 통해 레이크하우스의 기초를 구축하도록 도와 ETL 개발과 관리를 간소화 하여 줍니다.



Databricks 유니티 카탈로그 (Unity Catalog)는 데이터의 모든 담당자가 플랫폼에서 검색 및 보안, 권한 관리 작업을 진행하며, 데이터 자동 튜닝 및 최적화, 실시간 자동 리니지 작업, 모든 플랫폼으로 데이터 공유 등을 중앙에서 통합 관리함으로써 안정적으로 시스템이 관리됩니다.

② 통합 데이터 검색 및 보안, 권한 관리가 가능한 유니티 카탈로그 (Unity Catalog)

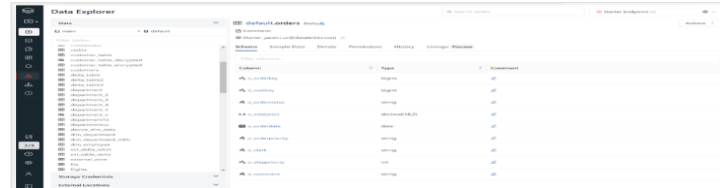
모든 데이터 담당자가 통합 데이터 관리

데이터 애널리스트, 데이터 엔지니어, 데이터 사이언티스트를 위한 통합 데이터 검색이 가능하며, 기본적으로 보안 기능과 사용자의 액세스 권한에 따라 검색 결과가 제한되고 개인정보 보호를 고려하여 보안 계층이 관리 됩니다.



자동 튜닝 및 데이터 최적화 작업

Unity Catalog는 메타데이터를 제공하고 테이블을 자동 튜닝하는 향상된 쿼리 성능을 제공하며 비동기식 자동 데이터 치밀화로 파일 용량을 최적화하고, 백그라운드에서 입출력(I/O) 지연을 자동으로 단축합니다.



실시간 자동 데이터 리니지

SQL, Python, Scala 및 R 형식의 모든 워크로드에서 실시간 자동 데이터 리니지를 사용하여 레이크 하우스에서의 데이터 흐름을 완전하게 확인할 수 있으며, 디버깅 및 권한 인식, 권한 제한. 다른 카탈로그와의 리니지 통합도 가능합니다.



모든 컴퓨팅 플랫폼으로 데이터 공유

Delta Lake와 Apache Parquet 형식의 기존 데이터를 모든 컴퓨팅 Platform으로 간편하게 할 수 있으며, Power BI, Tableau, Spark, pandas 및 Java와 기본적으로 통합되기 때문에 데이터를 받는 쪽에서 원하는 도구로 바로 공유된 데이터를 사용할 수 있습니다.



Databricks 데이터 엔지니어링 (Data Engineering)은 데이터 수집 간소화 및 자동 ETL 처리가 가능하며 중앙 집중형 모니터링 기능, 실시간 모니터링 및 시각화 기능, 데이터 거버넌스의 세분화 및 안전한 공유 등을 통하여 데이터 엔지니어링 작업의 효율성을 높일 수 있습니다.

③ ETL 작업과 워크플로 오케스트레이션 기능의 데이터 엔지니어링 (Data Engineering)

데이터 수집 간소화 및 자동 ETL 처리

Auto Loader를 활용하여 파일을 쉽게 가져올 수 있으며, Delta Live Table의 간단한 선언적 방식으로 ETL 및 ML 파이프라인을 구축하며 인프라 관리나 작업 오케스트레이션, 오류 처리, 복구와 같은 운영 복잡성과 성능을 최적화 자동화 합니다



중앙 집중형 모니터링 제공

Delta Live Tables, Jobs for SQL, Spark, 노트북, dbt, ML 모델 등을 포함한 전체 수명 주기에 대해 다양한 워크로드를 오케스트레이션 하며 최종 사용자에게 간단하게 심층적 중앙 집중형 모니터링을 제공합니다.



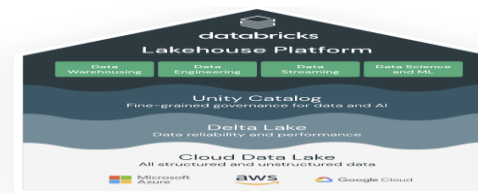
실시간 모니터링 및 시각화

모든 데이터와 AI 수명 주기에 대한 가시성을 제공하므로, 실시간으로 프로덕션 워크플로 상태를 확인하며 데이터 품질을 관리할 수 있고, 과거의 트렌드를 파악할 수 있으며 모든 각도에서 성능과 데이터 품질, 안정성 지표를 모니터링하고 시각화 할 수 있습니다.



데이터 거버넌스 세분화 및 안전한 공유

Unity Catalog를 결합하면 모든 데이터와 AI 자산에 세분화된 거버넌스를 제공할 수 있을 뿐만 아니라, 모든 클라우드에서 일관적인 데이터 탐색, 액세스, 공유 모델을 적용하여 거버넌스 방식을 단순화 하며 다른 조직과 간단하고 안전하게 데이터를 공유할 수 있습니다.




Databricks 델타 쉐어링 (Delta Sharing)은 모든 공유 데이터를 중앙에서 관리하고, 복제 없이 라이브 데이터를 공유하며, 오픈 된시스템 인터페이스를 사용하며, 개인 정보를 안전하게 보호하는 데이터 클린 룸 환경을 통해서 데이터 공유의 안정성 및 효율성을 높일 수 있도록 지원합니다.

④ 데이터를 중앙에서 관리하고 공유시키는 델타 쉐어링 (Delta Sharing)

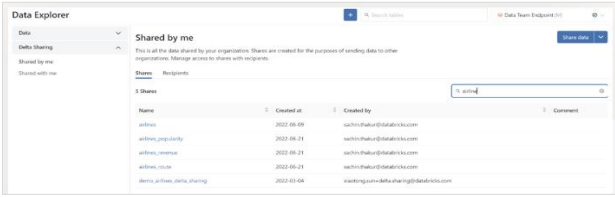
공유 데이터의 중앙 관리

Unity Catalog와 기본적으로 통합되어 전사적으로 공유 데이터를 중앙에서 관리하고 감사할 수 있으며 공급자, 협력업체와 신뢰를 바탕으로 데이터 자산을 공유할 수 있어 비즈니스 조율에 도움이 되고, 동시에 보안과 규정 준수 요구 사항에도 부합할 수 있습니다.



복제 없이 라이브 데이터 공유

데이터 플랫폼, 클라우드 또는 회사내의 분리된 시스템 내에 라이브 데이터를 다른 시스템에 복제하거나 복사하지 않고 공유합니다.



오픈 된 시스템 인터페이스

Databricks Platform이 아니더라도 어디서나 데이터 세트, 머신 러닝 모델, 대시보드, 노트북 등의 데이터 제품을 간편하게 발견, 평가하고 액세스할 수 있습니다.



개인 정보를 안전하게 보호하는 데이터 클린룸

어떤 클라우드에서나 고객 및 파트너와 협업 데이터 복제 없이 자신의 데이터 레이크에서 안전하게 데이터를 공유하며 사전 정의된 템플릿, 노트북, 대시보드를 사용한 공통적인 사용 사례를 통해 협업 대상을 안내하고 더욱 빠르게 인사이트를 얻을 수 있습니다.

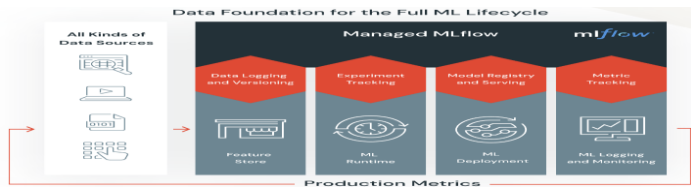


Databricks 머신 러닝 (Machine Learning)은 모든 유형의 데이터를 액세스 하고 탐색하며 실험 추적과 거버넌스 자동화 기능을 제공 하고, 모든 모델 수명 주기를 관리 하며, 대규모 데이터 머신 러닝 모델 배포 등을 통해 효율적으로 데이터 모델을 생성, 관리, 배포 할 수 있도록 지원합니다.

⑤ 데이터 모델링부터 실험, 프로덕션을 관리하는 머신 러닝 (Machine Learning)

모든 유형의 데이터를 액세스 하고 탐색

Databricks ML은 오픈 레이크 하우스 토대를 기반으로 빌드하여, 규모와 관계없이 어떤 유형의 데이터든 액세스, 탐색하고 준비 하도록 지원할 수 있으며 셸프 서비스 방식으로 여러가지 특징을 프로덕션 파이프라인으로 전환할 수 있습니다.



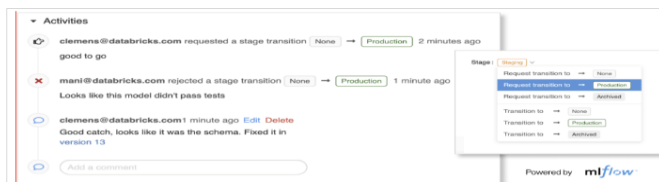
실험 추적과 거버넌스 자동화

이전 실행을 신속하게 확인하고 결과를 비교하며 필요한 경우 과거 결과를 재현할 수도 있으며 프로덕션할 모델의 가장 좋은 버전을 확인했다면, 이를 모델 레지스트리에 등록하면 배포 수명 주기에 따라 핸드오프를 간소화할 수 있습니다.



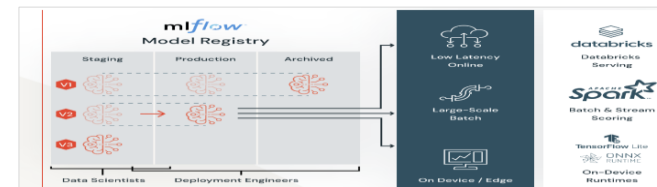
모든 모델 수명 주기 관리

모델은 실험, 스테이징, 프로덕션과 보관 등 다양한 단계를 거치면서 버전을 관리하고 이동할 수 있으며, 수명 주기 관리는 역할 기반 액세스 관리에 따라 승인 및 거버넌스 워크플로와 통합되어 관리되며 메모와 이메일 알림으로 정보를 제공합니다.



대규모 데이터 머신 러닝 모델 배포

서버 관리나 확장에 대한 제약을 걱정할 필요 없이 클릭 한 번으로 모델이 배포 가능하며 Databricks를 사용하면 엔터프라이즈급 가용성으로 어디에나 REST API 엔드포인트로 모델을 배포할 수 있습니다.



Databricks 데이터 사이언스 (Data Science)는 데이터 사이언스 워크플로 전체에 걸친 협업이 가능하며, 데이터 사이언스 중심 구성과 로우 코드 및 시각적 도구로 데이터 탐색이 가능하며 중앙 집중형 검색, 분석이 되므로 데이터 사이언스 작업이 효율적으로 진행될 수 있도록 지원합니다.

⑥ 안정적인 데이터 관리 및 분석을 지원하는 데이터 사이언스 (Data Science)

데이터 사이언스 워크플로 전체에 걸친 협업

Python, R, Scala 및 SQL로 협업 방식을 통해 코드를 쓰고, 대화형 시각화를 활용해 데이터를 탐색하며 공동 작성, 메모 작성, 자동 버전 관리, Git 통합과 역할 기반 액세스 관리 등이 제공되므로 신뢰를 기반으로 안전하게 코드를 공유할 수 있습니다.



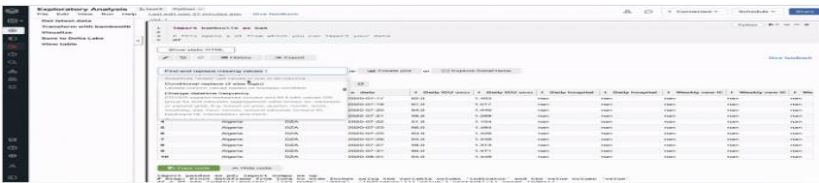
데이터 사이언스 중심 구성

노트북 컴퓨터의 데이터 용량에 제한을 받거나, 이용할 수 있는 컴퓨팅 파워로 행동 범위를 한정 짓지 않고 로컬 환경을 클라우드에 신속하게 마이그레이션하고, 노트북을 개인용 컴퓨팅, 자동 관리형 클러스터에 연결하여 사용가능 합니다.



로우 코드 및 시각적 도구로 데이터 탐색

기본 제공되는 네이티브 시각 도구를 사용해 데이터를 준비, 변환, 분석하면 전문지식 수준이 각기 다른 팀원이 모두 간편하게 데이터를 다룰 수 있으며 데이터 변환과 시각화를 마치면 배경에서 실행되는 코드를 생성할 수 있습니다.



중앙 집중형 검색, 분석

모든 데이터를 Delta Lake 내 한 곳에 정리하고 카탈로그화 하여 중앙집중형 데이터 스토어를 통해 기업 조직 내 어디에서나 검색 가능하게 할 수 있으며 데이터가 유입되면 자동 품질 검사를 실시해 데이터가 기대치에 부합하고 분석할 준비를 마쳤는지 확인합니다.



Databricks 데이터 스트리밍 (Data Streaming)은 스트리밍 데이터의 수집, 변환 및 ML, AI 작업에 적용이 가능하며 클러스터 활용 최적화 및 자동화된 운영 툴링 기능과 배치 및 스트림 처리를 위한 통합 API 기능을 통하여 데이터 스트리밍 관리의 최적화를 만들 수 있도록 지원합니다.

⑦ 스트리밍 데이터 워크로드 구축, 관리를 자동화 하는 데이터 스트리밍 (Data Streaming)

스트리밍 데이터 수집 및 변환

Delta Live Tables을 사용하여 스트리밍 데이터 파이프라인에 대한 데이터 입력 및 ETL을 단순화하며 데이터 엔지니어링에 대한 간단한 선언적 접근법을 활용하여 SQL 및 Python 등과 같이 이미 알고 있는 언어와 도구를 팀에 지원합니다.



스트리밍 데이터를 사용하여 ML, AI 작업

스트리밍 데이터를 사용하면 분석 및 AI의 정확도와 실행 가능성을 즉시 개선할 수 있습니다. 비즈니스는 스트리밍 데이터 파이프라인의 다운스트림 효과로 실시간 인사이트를 활용할 수 있습니다.



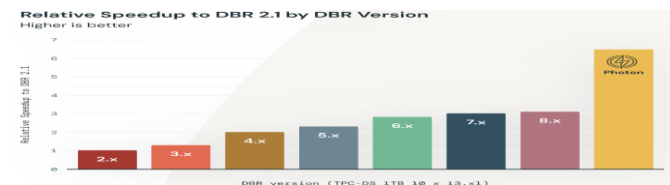
클러스터 활용 최적화 및 자동화된 운영 툴링

Enhanced Autoscaling는 각각의 고유한 워크로드에 컴퓨팅 리소스를 자동으로 할당함으로써 클러스터 활용을 최적화합니다. 이들 기능을 자동 데이터 품질 테스트 및 예외 관리와 함께 사용하면 운영 툴링을 구축 및 관리하는 데 사용하는 시간을 줄이고 데이터에서 가치를 창출하는 데 집중할 수 있습니다.



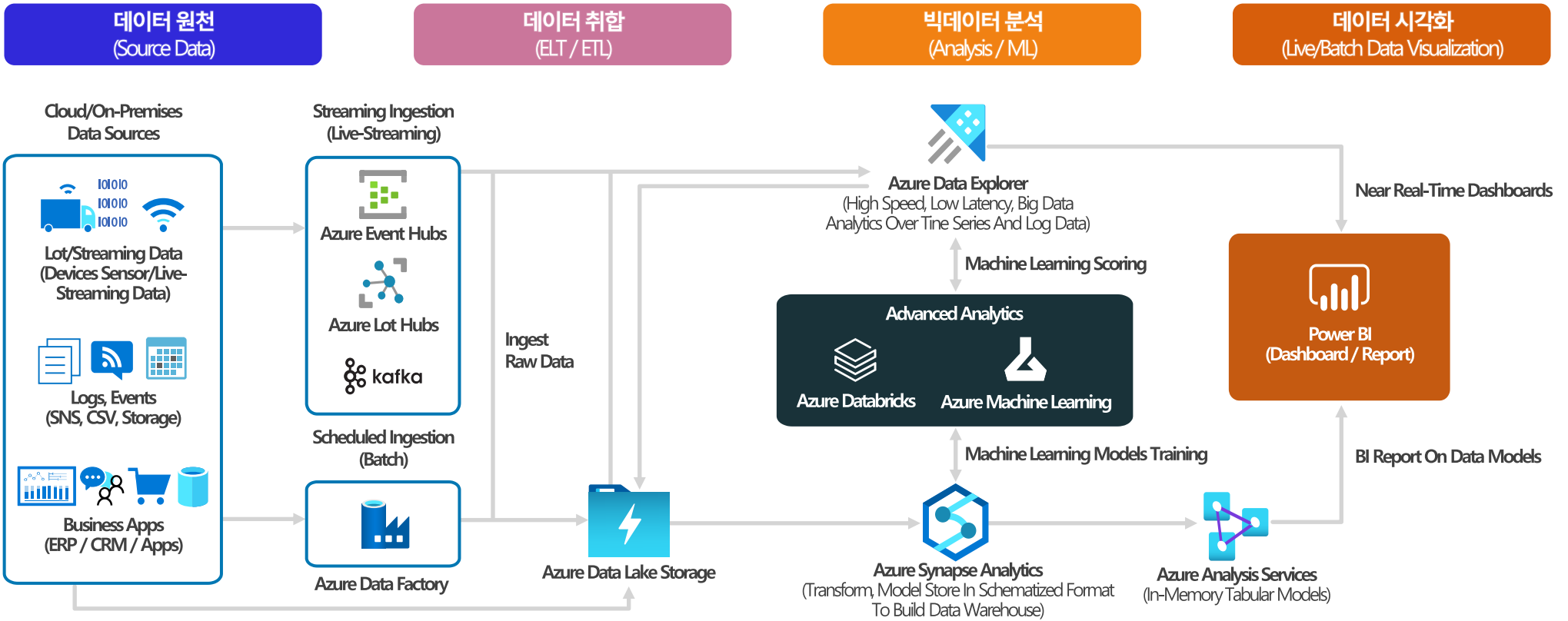
배치 및 스트림 처리를 위한 통합 API를 제공

배치 및 스트림 처리를 위한 통합 API를 제공하며 Spark 워크로드는 Apache Spark API와 호환되는 차세대 레이크하우스 엔진인 Photon으로 가속화되며, 수천 개의 노드로 자동 확장되고 최고의 비용당 성능비를 제공합니다.



Databricks on Microsoft Azure 빅데이터 플랫폼은 Azure 상에서 정형 및 비정형의 대량 데이터의 취합(ELT) 및 가공, 저장(Storage), 분석 (Analysis/ML)을 통한 데이터 모델링(Data Modeling)에서 데이터 시각화(BI) 보고서까지 End-to-end 빅데이터 플랫폼 (Big Data Platform) 환경을 제공합니다.

Databricks on Microsoft Azure 빅데이터 플랫폼(Big Data Platform)



Databricks Lakehouse는 전세계 7,000개 이상의 글로벌 기업과 Fortune 선정 500대 기업의 40% 이상이 데이터 엔지니어링 및 데이터 분석, ML, AI 활용을 위해 사용하고 있는 데이터 분석, 보안, 공유, 거버넌스 관리, AI 워크로드를 통합하여 관리하고 협업하는 클라우드 Data & AI 전문 플랫폼입니다.

글로벌 기업의 통합 클라우드 관리 데이터 플랫폼



데이터, 분석, AI 워크로드를 통합하는
클라우드 데이터 플랫폼

Customers

7000+
across the globe

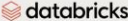


Original creators of:



Databricks Lakehouse는 시장 전문 기관 가트너(Gartner)의 데이터베이스 성능과 생산성에 관련된 조사에서 Magic Quadrant DBMS(Database Management System) 부문과 DSML(Data Science and Machine Learning Platforms) 부문에서 경쟁 솔루션 중 선두기업으로 선정 및 우수성을 인정 받았습니다.

Databricks Lakehouse 시장 전문 기관 평가

 **databricks**

2021 Gartner Reports

Databricks named a Leader in 2021 Gartner® Magic Quadrant™ in both DBMS and DSML

Proving leadership across the data and AI lifecycle

Gartner

Databricks is the only cloud-native vendor named a Leader in both Gartner Magic Quadrants:

- Cloud Database Management Systems
- Data Science and Machine Learning Platforms

Databricks is proud to announce that Gartner has named us a Leader in both the 2021 Magic Quadrant for Cloud Database Management Systems and the 2021 Magic Quadrant for Data Science and Machine Learning Platforms.



2021 Gartner® Magic Quadrant™ for Cloud Database Management Systems



2021 Gartner® Magic Quadrant™ for Data Science and Machine Learning Platforms

Databricks Lakehouse는 현재 전세계의 유통, 제조, 금융, 생명 과학, 통신/게임 등 각 산업 분야에서 정형/비정형 및 가공/원천, 회사 내부/외부에서 발생하고 있는 빅데이터를 이용하여 현상 분석 및 발생 가능한 리스크 방지, 미래 예측 분석 등의 작업을 통하여 기업의 가치를 높이고 있습니다.

Databricks Lakehouse 도입 및 활용 사례



유통

- 소비자 이탈 예측
- 세분화된 수요 예측
- 고객 평생 가치 예측
- 고객 세분화 관리
- 실시간 POS 정보 분석
- 판매 예측 & 광고 속성



제조

- 예측 유지 관리 (IoT)
- ESG 성과 분석
- 고객 360도 뷰 정보 생성
- 안전 재고 관리
- 최적화된 오더 피킹
- 세분화된 수요 예측



금융

- 자금 세탁 방지
- 고객 세분화 관리
- 실시간 금융 사기 방지
- 평판 위험도 관리
- 고객 평생 가치 관리
- 규제 보고 관리



생명 과학

- 약물 부작용 감지
- 병리학 이미지 분석
- 게놈 전체 연관 연구
- 메디케어 위험 조정
- 건강의 사회적 요인 관리
- 코호트 구축 지식 관리



통신/게임

- 게임에서의 독성 검출
- 비디오 체감 품질 관리
- 생존 분석 및 평생 가치
- DNS를 통한 위협 탐지
- 고객 세분화 관리
- 세분화된 수요 예측

Databricks Lakehouse는 현재 전세계의 유통, 제조, 금융, 생명 과학, 통신/게임 등 각 산업 분야에서 정형/비정형 및 가공/원천, 회사 내부/외부에서 발생하고 있는 빅데이터를 이용하여 현상 분석 및 발생 가능한 리스크 방지, 미래 예측 분석 등의 작업을 통하여 기업의 가치를 높이고 있습니다.

Databricks Lakehouse 글로벌 고객사

Healthcare & Life Science



Retail & CPG



Media & Entertainment



Financial Services



Public Sector



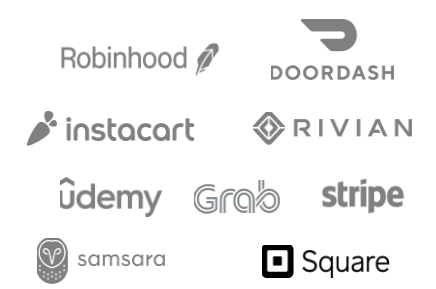
Manufacturing & Logistics



Energy & Utilities



Digital Native



3. SAP Data Migration to Databricks

SAP Data Migration to Databricks 프로그램은 온프레미스(On-Premise)고객 및 아마존, 구글 등의 Azure 클라우드 환경의 SAP 고객사를 위한 프로그램으로, SAP ERP 데이터 및 내·외부 업무 데이터 통합 관리 및 분석을 위한 단일 데이터 플랫폼과 현재와 미래의 분석 환경을 제공합니다.

SAP 고객을 위한 Databricks 마이그레이션 프로그램

Scenario 1



Net New Customer

A Customer who has not adopted a public cloud and prefer to continue SAP on-prem or Private Cloud for some more time

Scenario 2



Customer with SAP on Compete Cloud

A customer with SAP Infra on compete cloud like AWS/GCP

Scenario 3



Customer with SAP on Azure

A Customer with SAP infra already on Azure

SAP Extend and Innovate Scenarios

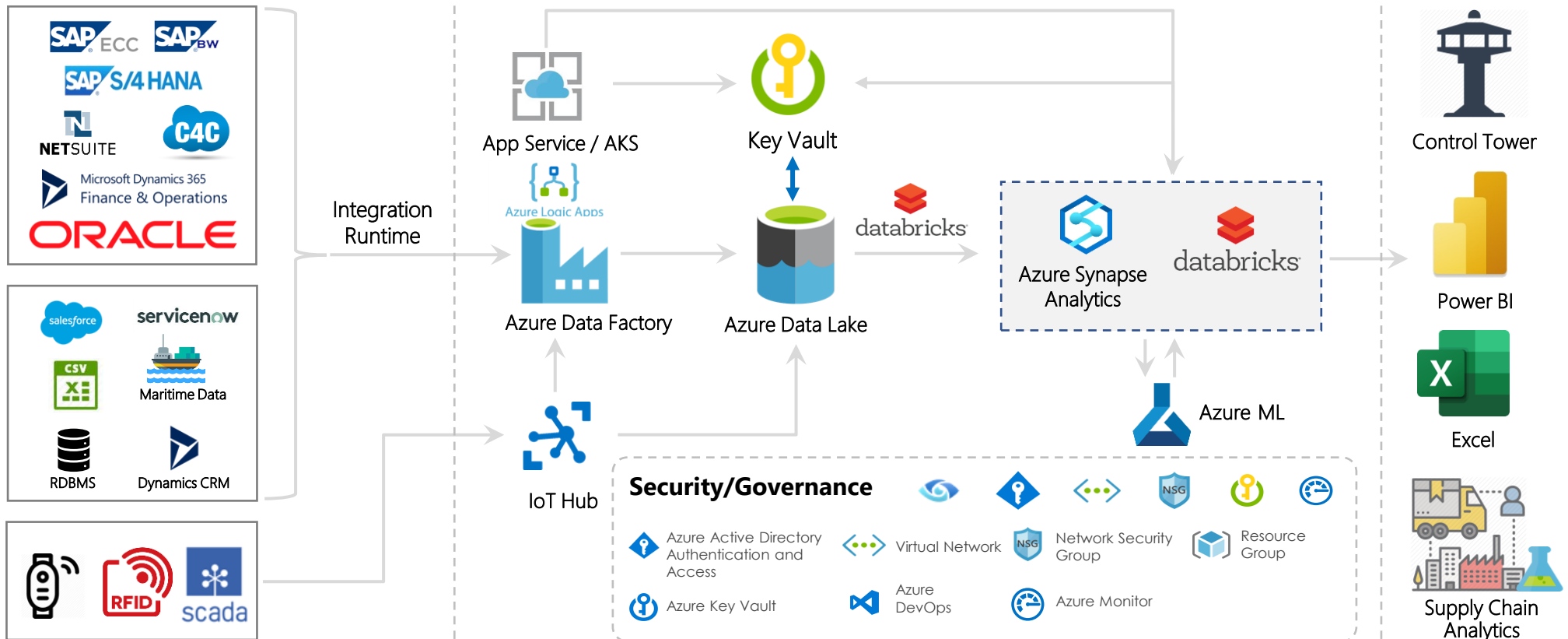
- Innovation beyond Infra
- Complete Digital Transformation
- Drive Customer Stickiness on Azure

- Initiate Azure Cloud Presence
- Engage with Microsoft ecosystem
- Gives the customer a Multi-Cloud setup
- Driving an eventual Migration to Azure Infra

- Boost Azure Consumption with of Azure services with Extend and Innovate solution and solution accelerators

SAP Data Migration to Databricks 프로그램은 개방성(Open)과 통합성(Unify) 을 보유한 Databricks Lakehouse를 기반으로, SAP ERP 및 중
 심으로 기업에 연관된 기업 모든 정보를 배치 또는 실시간으로 마이그레이션 하여 통합 관리 함으로 데이터 분석(BI) 및 AI/ML 사용성을 최대화 시킵니다.

통합 데이터 플랫폼 Databricks 아키텍처(Architecture)



SAP Data Migration to Databricks 프로그램은 코드 마이그레이션(Code Migration)을 시작으로 ETL 및 데이터마이그레이션(Data Migration)을 순차적으로 진행하여, 이는 기업의 시스템 환경에 따라 각 시스템을 통한 마이그레이션과 수작업(Manual) 마이그레이션이 진행 됩니다

마이그레이션의 단계별 상세 업무 진행

Track 1 : Code Migration

Migrate all logics, functions, Stored Procedures as follows :

- Extract Teradata scripts as .sql files
- Assess using BladeBridge Analyzer
- Categorize Jobs as per complexity
- Pilot Migration to Databricks PySpark using BladeBridge Converter
- Documentation

Deliverables

- Teradata script Analysis report
- Bill of Materials for Databricks

Options Available

- a) **Option 1:** Manual Migration
- b) **Option 2:** Migration using BladeBridge (Recommended)

Track 2 : ETL

Migrate all ETL/ETL jobs as follows :

- Scope ETL jobs
- Functional and Technical Assessment(Manual) – ETL Jobs
- Categorize Jobs as per complexity
- Pilot Migration to Azure Data Factory Jobs(Manual)
- Documentation

Options Available

Manual Migration (Recommended)

- a) **Option 1:** Migration to PySpark Jobs on Databricks
- b) **Option 2:** Migration to Azure Data Factory

Track 3 : Data Migration

Migrate data to Azure :

- Sources: Data sources plugged in Teradata
- Migrating subset of data from source systems to Azure
- Create copy pipelines for data migration in Azure Data Factory
- Execute data transfer
- Testing and Validation
- Document

Options Available

- a) **Option 1:** Connect all required source systems to Azure using ADF and ingest data
- b) **Option 2:** Migration a subset of all required table data from above sources

Manual Migration

- Hard to analyse pattern
- Understand Use-cases and data models from scratch

BladeBridge Migration

- Automate code conversion.
- Easier Analysis
- Reader file available
- Writer file available

PySpark Jobs on Databricks

- Code based approach
- Install drivers to use spark connectors
- Bit complex to leverage other Azure tools

Using Azure Data Factory

- GUI Based approach
- Native connectors to multiple source systems
- Easy to integrate with other Azure Tools

Direct connection using ADF

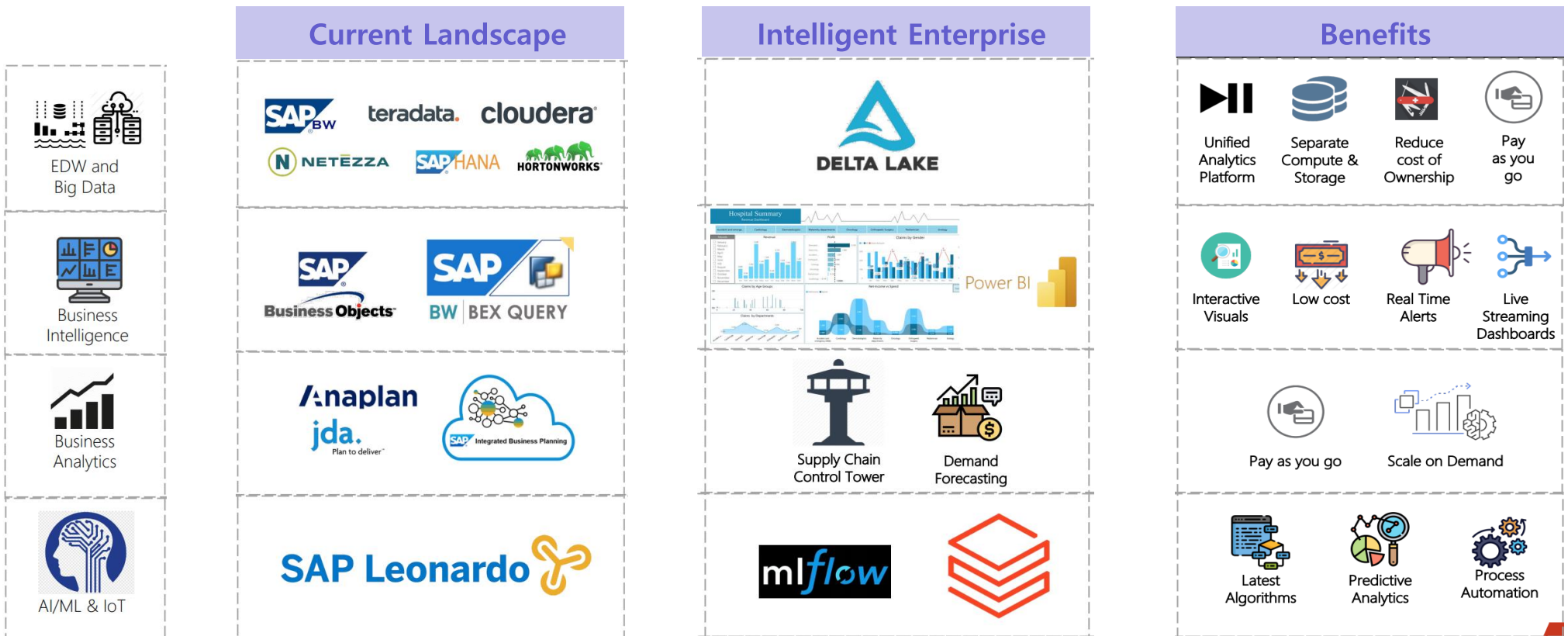
- For pilot, approvals might delay connectivity

Migrate subset of source Data to Azure

- Feasible approach for pilot.
- Cons : Initial connectivity in the converted jobs must be changed again

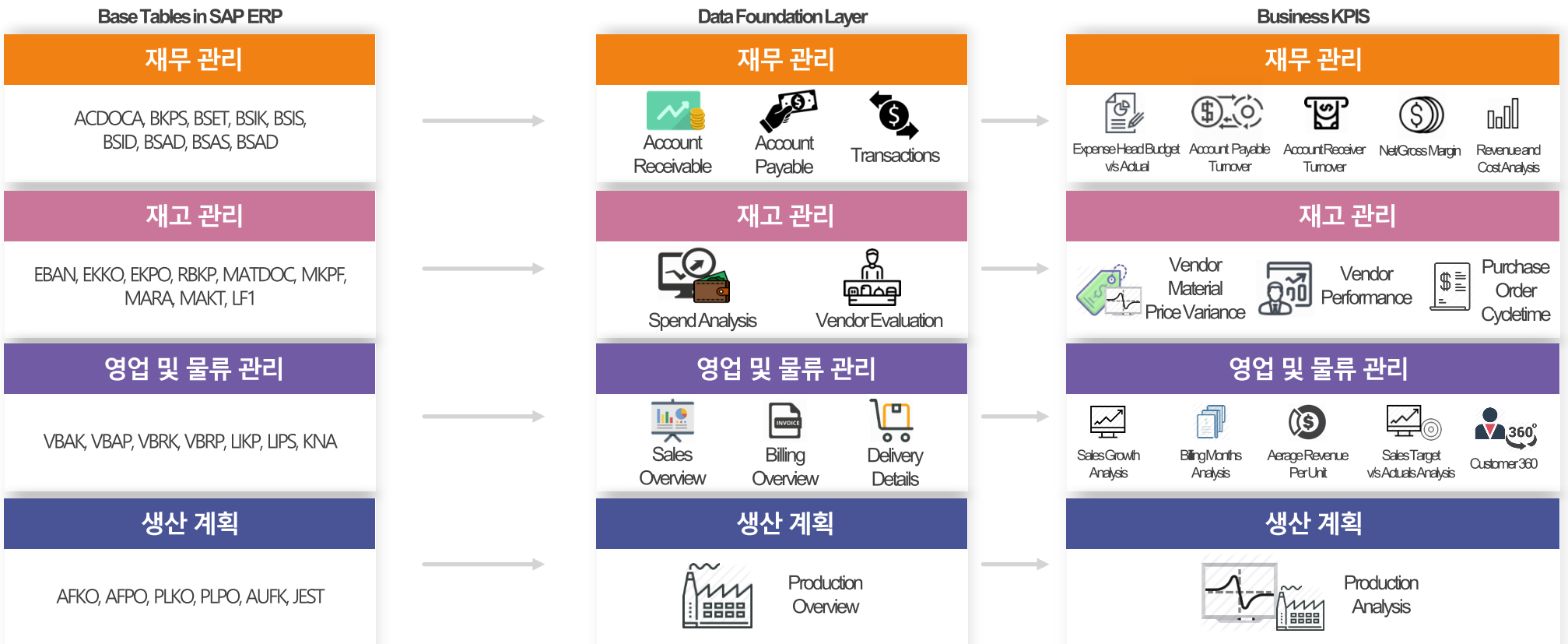
SAP Data Migration to Databricks 프로그램은 분리된 과거 데이터 분석 중심(Business Analysis) 기업 정보 활용 한계를 뛰어 넘어, 효율적 비용으로 업무 예측(Predictive Analytic),실시간 위험 알림(Real Time Alert) 등의 현재와 미래의 예측이 가능한 통합 단일 업무 플랫폼을 도입을 지원합니다.

Migration Service for SAP – Databricks for SAP Customers



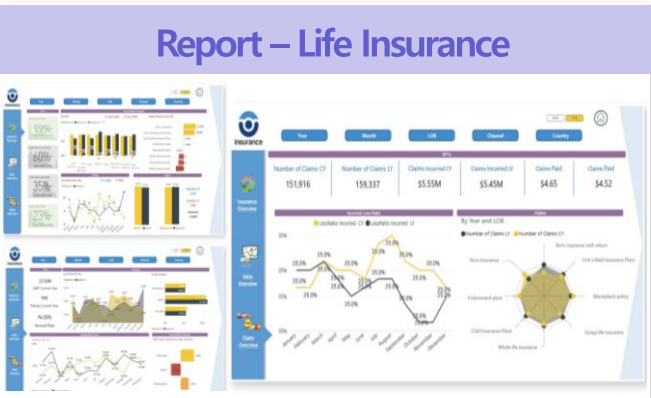
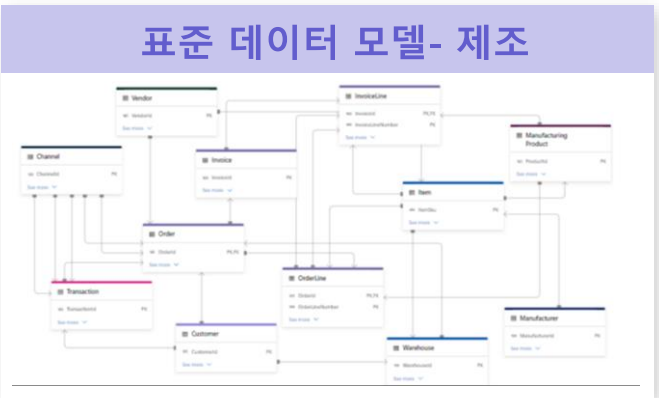
SAP Data Migration to Databricks 프로그램을 통해 적용된 Databricks Lakehouse는 재무 관리(Finance & Controlling), 재고 관리(Inventory Management), 영업 및 물류(Sales & Distribution) 및 생산 계획(Production Planning) 등의 전 업무 부분의 SAP 기반 P표준 KPI 데이터/보고서를 제공합니다.

전 SAP 업무 부분 데이터 모델과 보고서 제공



SAP Data Migration to Databricks 프로그램을 통해 적용된 Databricks Lakehouse는 제조(Manufacturing), 금융(Banking) 및 보험(Life Insurance) 등의 각 산업군의 분석(BI) 및 예측(Prediction)을 위한 표준 데이터 모델(Standard Data Model)과 표준 보고서(Analysis Report)를 제공합니다

산업별 분석/예측을 위한 위한 표준 데이터 모델과 표준 보고서(Report Snapshots)



SAP Data Migration to Databricks 프로그램을 위한 POC 프로그램은 Databricks Lakehouse를 기업에 적용함으로써 이의 실제 검증에 지원
 하는 프로그램으로 전체 5주에 걸쳐, 7-8개의 마스터 테이블과 1개의 트랜잭션 테이블의 마이그레이션과 1-2개의 KPI보고서를 개발 진행하게 됩니다.

SAP Data Migration to Databricks 프로그램을 위한 POC

Week 1



Establishing connectivity between SAP ECC or S/4HANA by setting up the environment.

Week 4



Developing ETL pipeline within Azure Data Factory for both historical and incremental loads.

- Migrating the logic from SAP BW/HANA to corresponding logic within Lakehouse on Azure Databricks.
- Develop an abstracted data model consisting of Fact and Dimension Views within Lakehouse

Week 5



Power BI reports on top of the data model developed within Lakehouse and the delta engine

Assumptions



Manual Migration

- 1 Transactional Table (Header and Item)
- 7 - 8 Master Tables



1 reporting tab within Power BI will be developed during this effort



1 - 2 KPIs will be developed
 As part of this effort within the data model layer

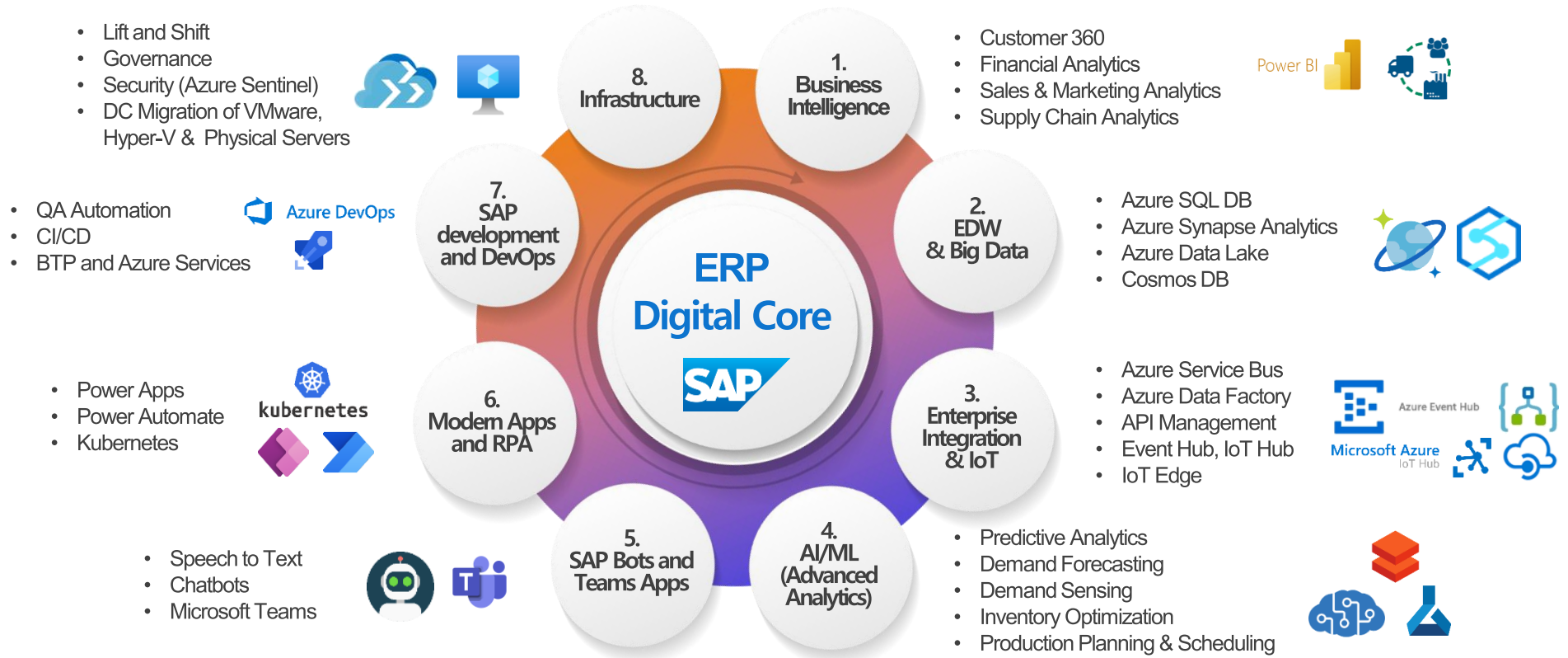


Transformation logic would consist of basic transformations such as lookups, basic conditional derivations and aggregations

4. Microsoft Power Platform for SAP

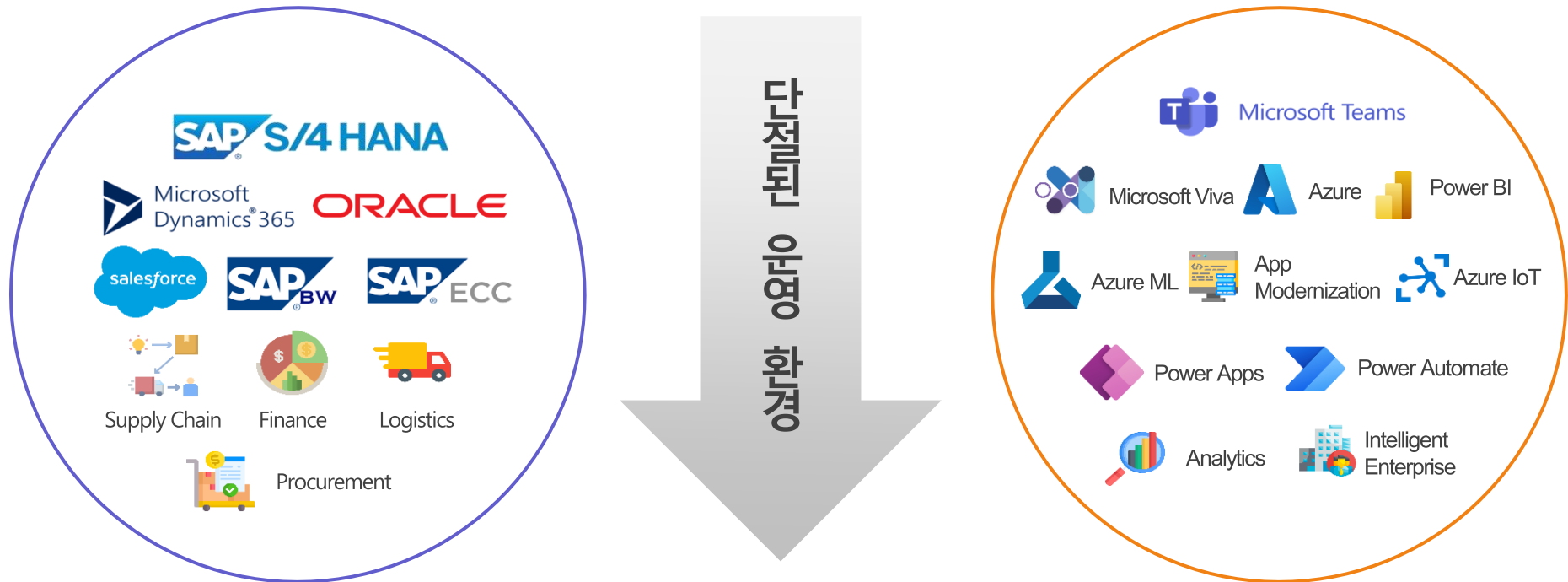
SAP ERP 사용자를 위한 마이크로소프트 솔루션은 Azure와 Power Platform의 8가지 혁신적인 기술을 활용하여 SAP 및 비 SAP 데이터의 통합과 AI/ML을 통한 예측, 시각화 분석을 수행하며 RPA를 이용한 자동화 및 최고의 보안 기능 활용 등으로 SAP 업무 혁신과 비용 효율화를 만들수 있습니다.

SAP 업무 혁신과 비용 효율화를 위한 마이크로소프트 8가지 솔루션



기존의 단절된 시스템 운영 환경은 기간 업무 시스템인 SAP ERP 시스템 및 비 ERP 시스템이 각자 고유한 운영 환경에서 개발 되고 운영됨으로써 비효율적 사용자 업무 환경과 데이터 활용의 제한, 라이선스 비용 부담 등의 문제가 발생하여 기업의 민첩성과 생산성 제고에 제한적 요소가 되고 있습니다.

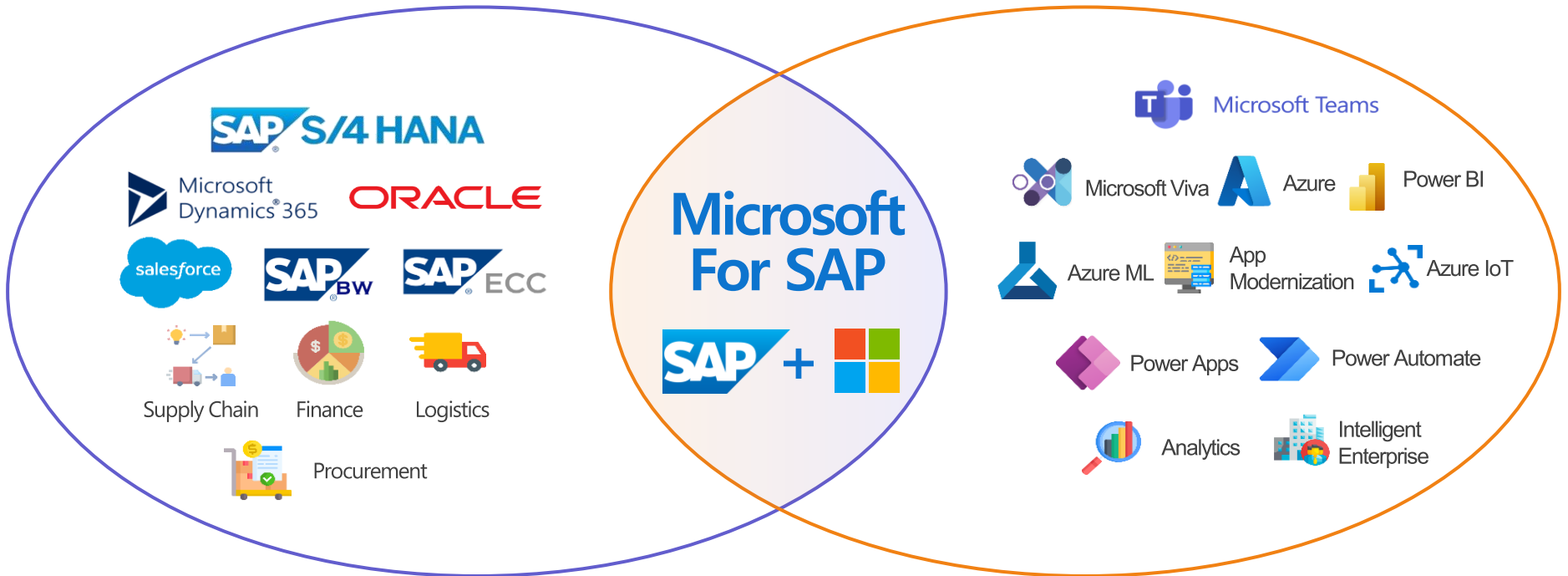
기존의 단절된 시스템 운영 환경의 문제점



기업의 민첩성과 생산성 제고에 제한적 요소

마이크로소프트 솔루션 기반하의 운영 환경은 사용자에게 익숙한 마이크로소프트 환경에서 SAP ERP를 비롯한 모든 업무가 통합되어 개발, 운영됨으로써 효율적 사용자 업무 환경과 데이터 활용, 라이선스 활용을 이룰 수 있고 이를 통해 기업은 SAP 업무 혁신과 비용 효율화를 이룰 수 있습니다.

마이크로소프트 솔루션 기반하의 업무 환경



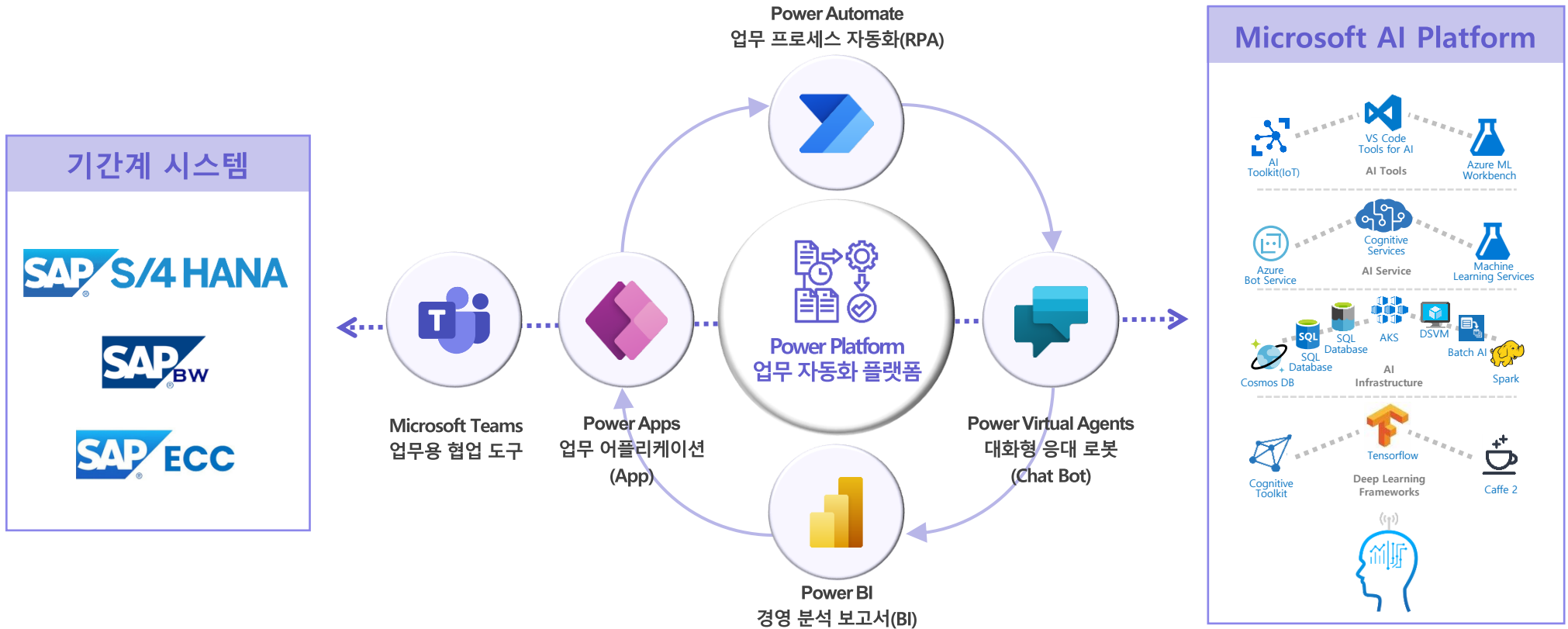
1. 효율적 사용자 업무 환경, 2. 효율적 데이터 활용, 3. 효율적 라이선스 활용

4. Microsoft Power Platform for SAP

1) 효율적 사용자 업무 환경

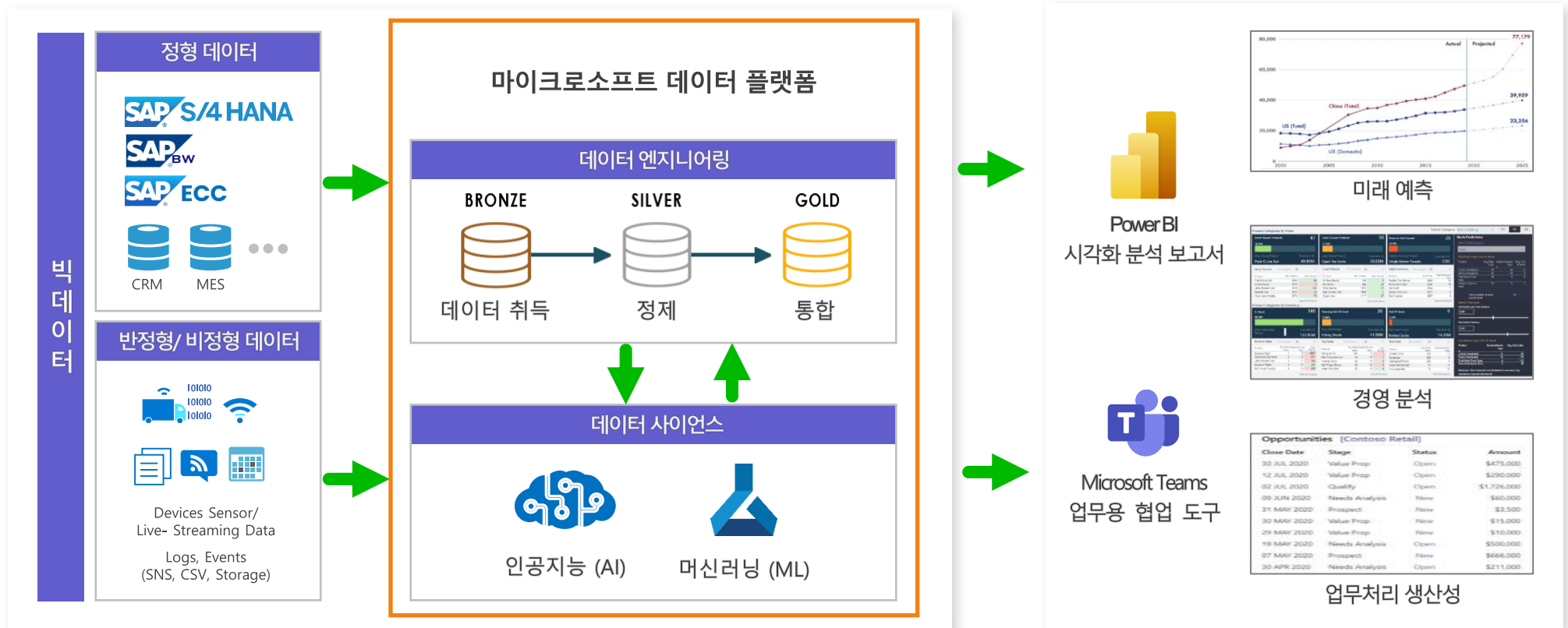
SAP ERP 사용자를 위한 마이크로소프트 솔루션은 Microsoft AI 기반 솔루션인 Power Platform의 업무 자동화(RPA) 및 업무 앱(App), 대화형 응대 로봇, 데이터 분석 서비스와 협업 도구인 Teams 등을 통하여 생산성 높은 효율적 업무 처리 환경이 구축 됨으로써 SAP 업무 혁신을 이룰 수 있습니다.

1) 효율적 사용자 업무 환경



SAP ERP 사용자를 위한 마이크로소프트 솔루션은 SAP 데이터 및 비 SAP 모든 데이터를 통합하여 AI/ML 기능을 활용한 데이터 플랫폼을 구축한 뒤 이를 예측 및 분석 작업과 팀즈에서 직접적인 업무 처리를 통하여 효율적 데이터 활용이 가능해 짐으로써 SAP 업무 혁신을 이룰 수 있습니다.

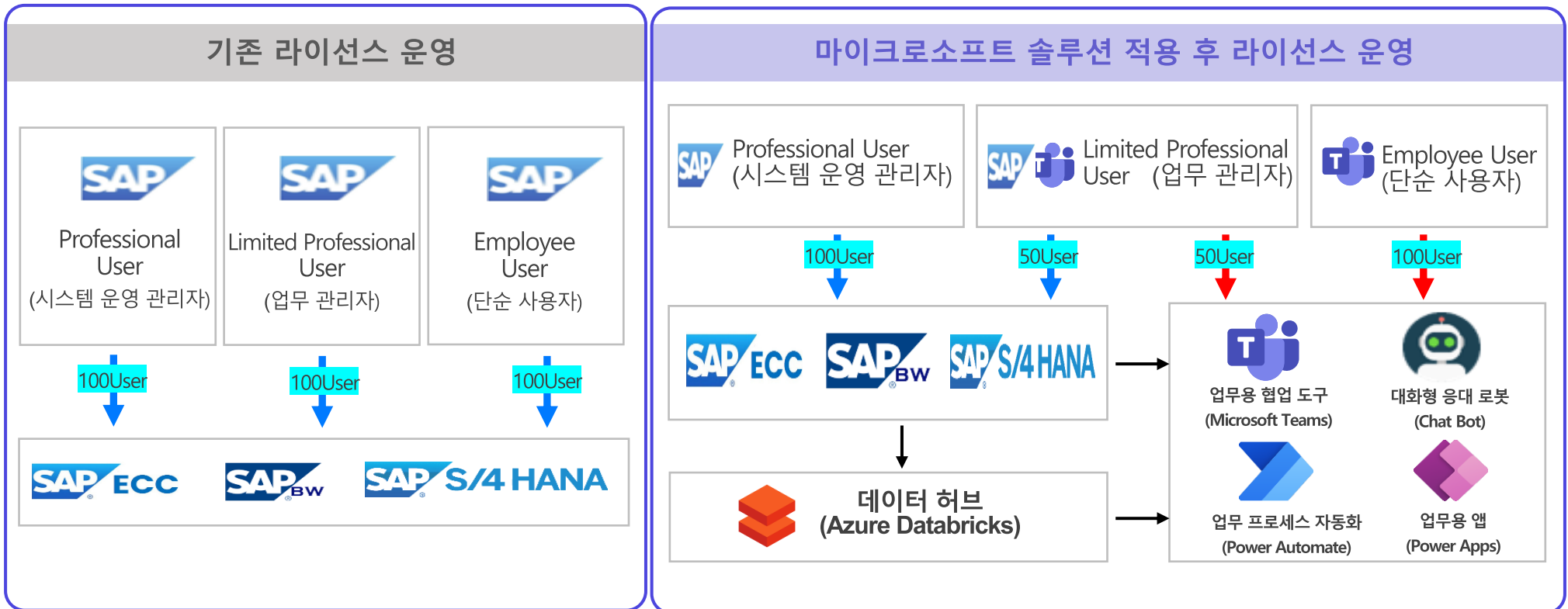
2) 효율적 데이터 활용



SAP ERP 사용자를 위한 마이크로소프트 솔루션은 SAP를 운영하는 관리자를 제외한 일반 사용자가 익숙한 업무 협업 솔루션인 Teams를 사용하여 ERP 업무를 수행하게 됨으로써 효율적인 업무 처리로 생산성을 높이는 효과와 함께 SAP 라이선스 미사용으로 인한 비용 효율화를 만들 수 있습니다.

3) 효율적 사용자 라이선스

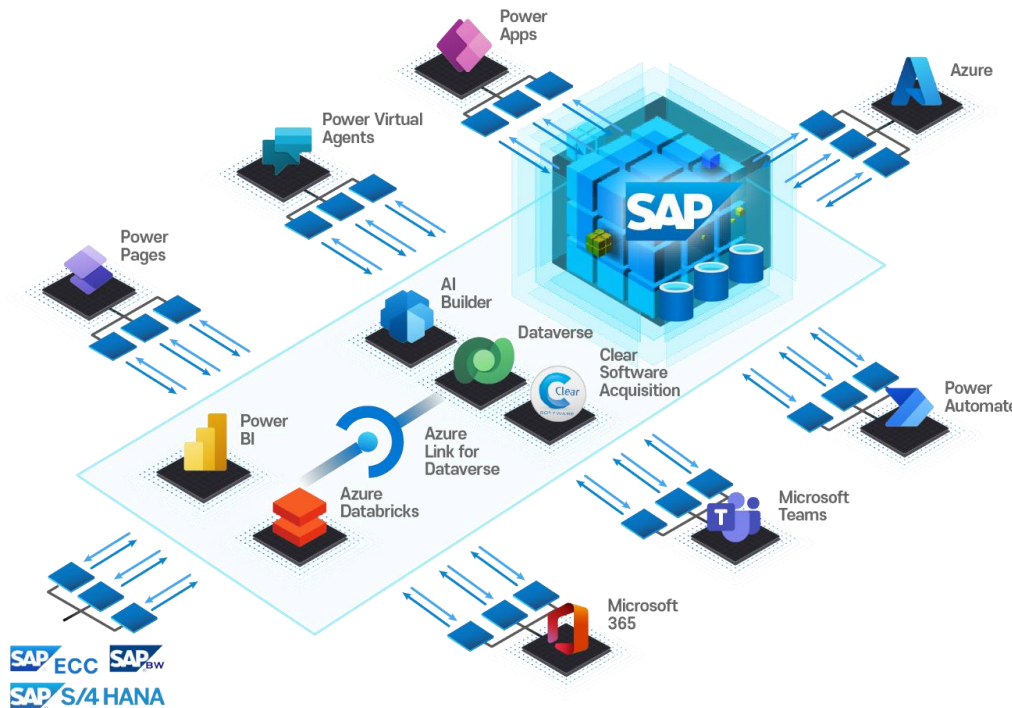
▶ 사용자 : 시스템 운영 관리자, 업무 관리자, 단순 사용자 각 100User 기준 ▶ 라이선스 : **SAP** / **Microsoft365**



SAP 활용을 위한 마이크로소프트 솔루션은 Azure 클라우드 기반의 로우코드 개발 플랫폼과 데이터브릭스를 이용하여 SAP 운영 시스템을 쉽게 구축하며, 이를 활용한 효율적 업무 환경 구성, 데이터 활용성 증대, 효율적 라이선스 활용 등을 통하여 기업의 비즈니스 혁신을 이룰 수 있습니다.

마이크로소프트 통합 솔루션을 통한 비즈니스 혁신

SAP on Azure innovation with Microsoft's world leading low code and Databricks













5. 적용 및 고객 지원 서비스

Microsoft Azure Databricks 기반 Teams 및 Power Platform 활용으로 업무 혁신을 희망하는 고객사를 대상으로 SAP 업무용 데이터를 아래 4가지 Pattern 중 한가지를 선정하여 업무에 적용 함으로써 비즈니스 민첩성을 높이고 업무 혁신과 비용 절감을 이룰 수 있는 서비스를 제공합니다.

마이크로소프트를 활용한 4가지 적용 Pattern

			
			
<p>SAP ERP등의 기업 업무에 필요한 업무자동화 (RPA)</p>	<p>Microsoft Teams 기반의 SAP ERP 협업 환경 구축</p>	<p>Microsoft 로우 코드(Low-Code) 기반의 SAP ERP UI 구축</p>	<p>SAP ERP 등의 업무 데이터를 Databricks를 통한 ML/AI 시스템구축</p>

Microsoft Azure Databricks 기반 Teams 및 Power Platform 활용으로 업무 혁신을 희망하는 고객사를 대상으로 SAP 업무용 데이터를 RPA를 통하여 자동화 함으로써 단순 반복적인 수작업의 자동화 및 수작업으로 인한 오류를 방지하여 업무 생성을 높일 수 있는 서비스를 제공합니다.

Pattern 1 : SAP ERP등의 기업 업무에 필요한 업무자동화 (RPA)

Scenario and Challenges

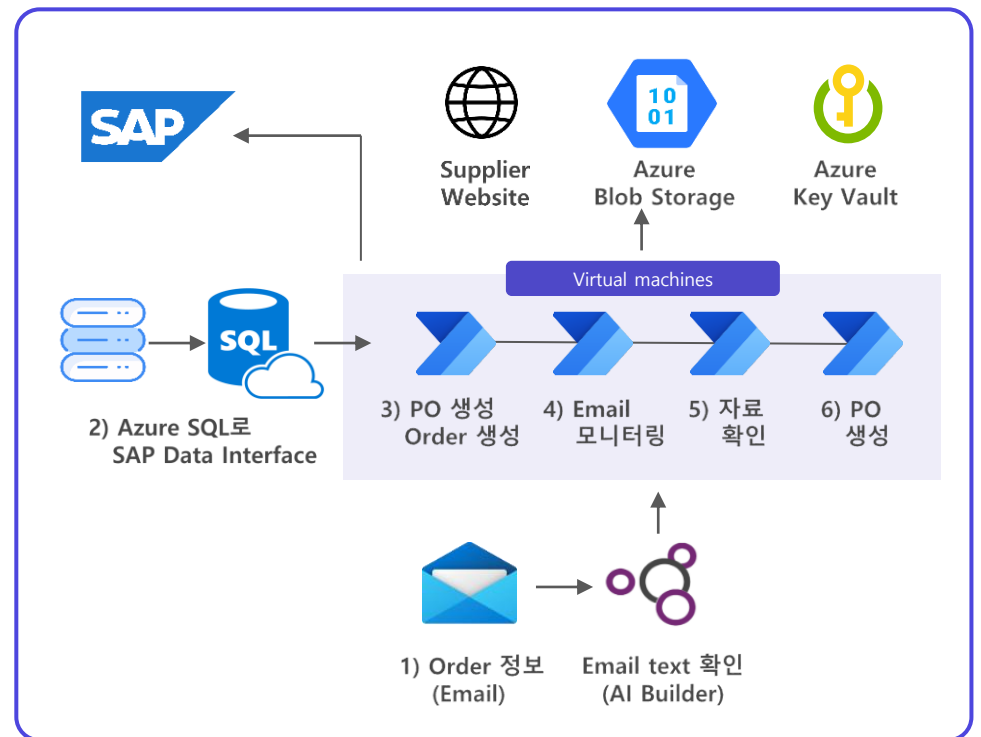
Scenario: 대부분의 경우 비즈니스 프로세스를 완료하는 데 필요한 정보가 SAP 시스템과 외부 시스템에 동시 존재하기 때문에 담당자는 업무 처리를 위해 여러 IT 시스템을 이동하며 작업해야 합니다.

Challenges:

- 이는 오류 가능성을 높이고 생산성을 감소시키며 궁극적으로 프로세스 민첩성을 감소 시킨다.
- SAP에 데이터를 읽고 입력하기 위해 사람의 행동에 의존하므로 잠재적인 비효율성과 오류가 발생합니다.

Business Outcomes:

- 더 적은 오류로 더 빠른 엔드 투 엔드 처리.
- 프로세스 전반에 걸쳐 더 나은 추적 기능으로 문제를 더 일찍 식별하고 수정



Microsoft Azure Databricks 기반 Teams 및 Power Platform 활용으로 업무 혁신을 희망하는 고객사를 대상으로 SAP 데이터를 익숙한 업무 환경인 Teams내의 Chatbot을 통하여 처리함으로써 업무 처리 신속성, 데이터 관리 효율성 등으로 업무 생성을 높일 수 있는 서비스를 제공합니다.

Pattern 2 : Microsoft Teams 기반의 SAP ERP 협업 환경 구축

Scenario and Challenges

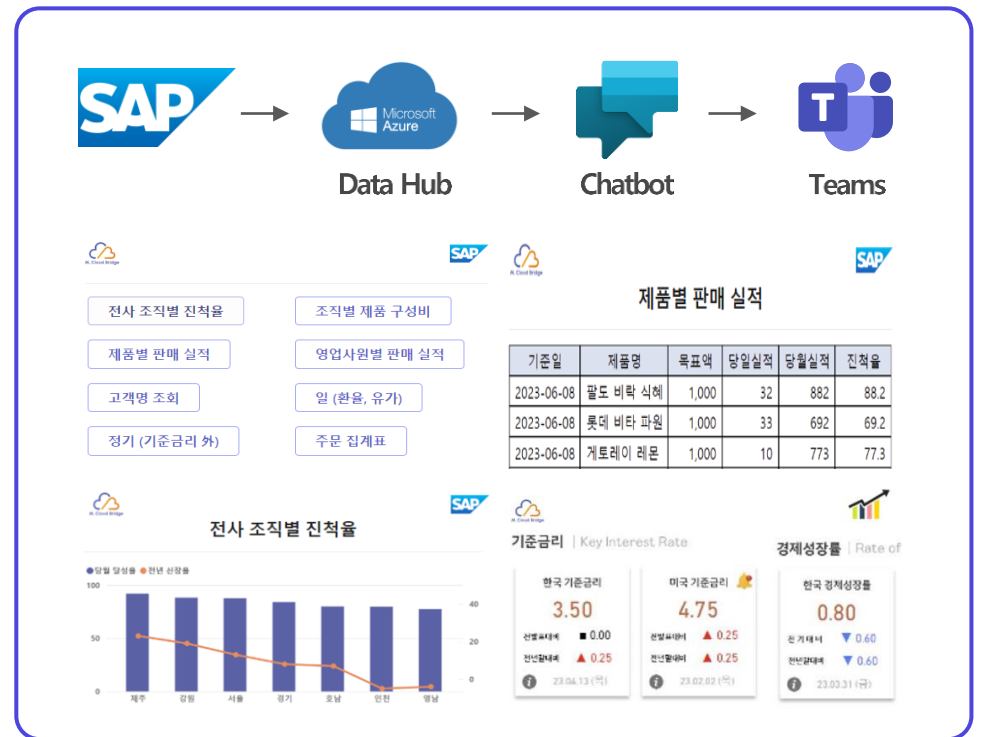
Scenario: 업무 담당자들은 협업 도구를 사용하여 다른 사람과 의사소통과 정보를 공유하며, 질문하고, 메모 하는 등 대부분의 시간을 익숙한 업무 환경에서 작업합니다.

Challenges:

- 사용자는 일반적으로 자신이 알고 있고 친숙한 프론트엔드 환경을 사용하는 것을 선호합니다.
- 이는 복잡한 ERP 화면 사용시 보다 생산성이 우수합니다.

Business Outcomes:

- SAP 데이터를 기반으로 한 협업 경험과 이미 알고 있고 매일 사용하는 툴의 프로세스를 통해 생산성 향상
- 수동 및 반복 작업을 자동화 하며 협업 환경에서 처리된 업무는 기록 되어 관리



Microsoft Azure Databricks 기반 Teams 및 Power Platform 활용으로 업무 혁신을 희망하는 고객사를 대상으로 SAP 데이터를 Power Apps를 통하여 처리함으로써 경제적인 개발 비용과 M365의 통합된 보안관리, 모바일 기기의 활용 등으로 업무 생성을 높일 수 있는 서비스를 제공합니다.

Pattern 3 : Microsoft 로우 코드(Low-Code) 기반의 SAP ERP UI 구축

Scenario and Challenges

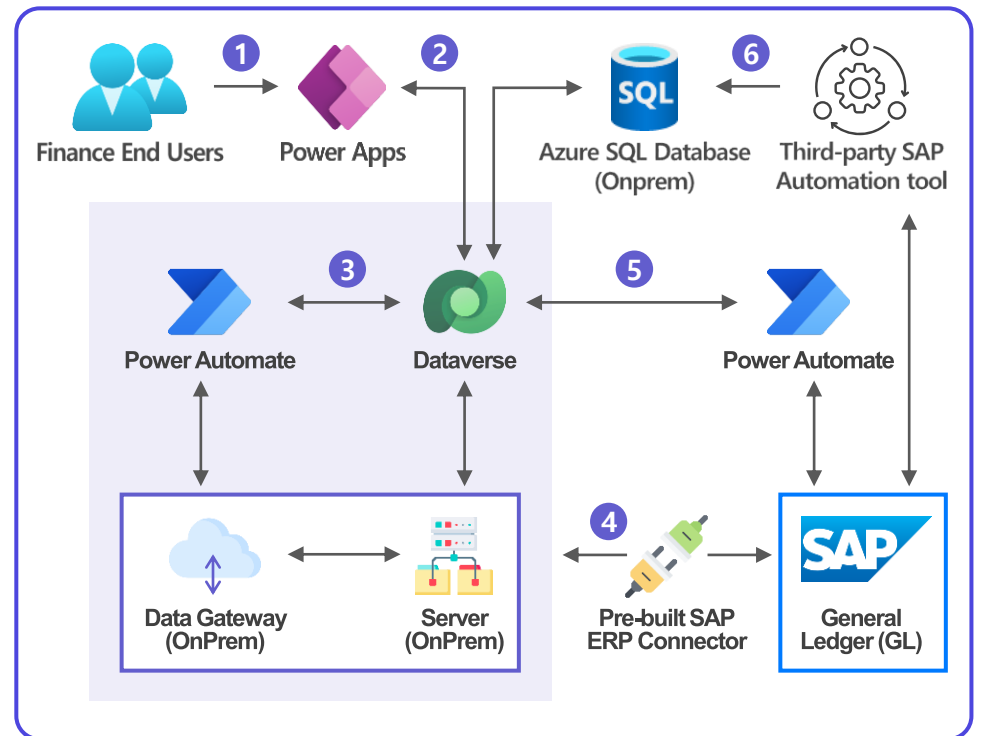
Scenario: 데이터는 SAP 시스템에서 관리 되지만 ERP의 복잡한 화면과 관리 프로세스와는 달리 신속하고 간단하게 입력이 필요한 비즈니스 프로세스가 있습니다.

Challenges:

- SAP에서 개발시 ERP의 표준화된 기능을 사용해야 하므로 비즈니스 민첩성을 감소 시킨다.
- 사용자 요구 사항에 맞춰서 ERP 시스템 개발시에는 많은 비용과 시간이 소요된다.

Business Outcomes:

- 개발 비용 및 기간 단축
- 사용자 요구 수준에 부합한 프로세스 구축
- 모바일 장치 활용성 강화 (예 : 카메라, 이미지 인식, GPS 등)



Microsoft Azure Databricks 기반 Teams 및 Power Platform 활용으로 업무 혁신을 희망하는 고객사를 대상으로 정형 및 비정형이 포함된 SAP 및 내외부 비 SAP 데이터를 취합 및 정제하여 성능과 안정성이 보장된 데이터 허브로 통합한 뒤 ML/AI 예측 및 시각화 분석 서비스를 제공합니다.

Pattern 4 : SAP ERP 등의 업무 데이터를 Databricks를 통한 ML/AI 시스템구축

Scenario and Challenges

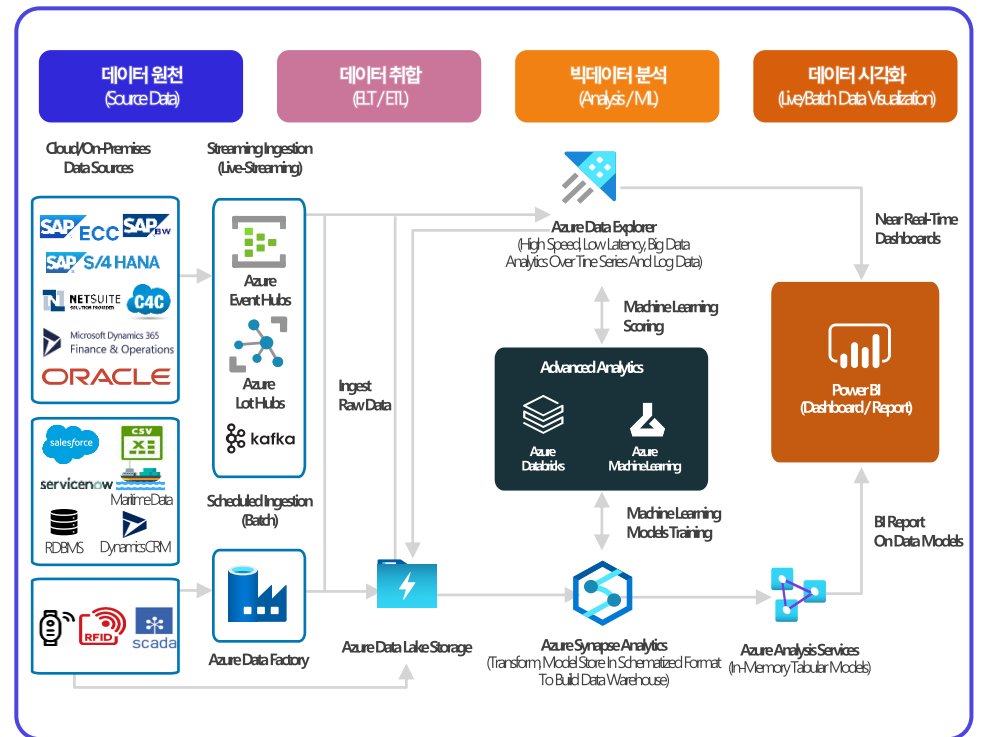
Scenario: SAP 데이터 및 회사 내/외부 정형/비정형 대량 데이터를 통합하여 미래예측에 활용하고 이를 시각화 하여 상세 분석을 해야 하는 비즈니스 프로세스가 있습니다.

Challenges:

- 데이터가 분산 관리 되어 있고 대량의 데이터는 취합 및 정제에 많은 시간과 비용이 발생한다.
- 전문적인 빅데이터 관리 솔루션 부재로 ML/AI를 활용한 예측 업무를 수행 할 수 없음

Business Outcomes:

- 단일 빅데이터 플랫폼을 활용한 데이터 허브를 구축함으로써 데이터 안정성, 신속성, 거버넌스 확보
- ML/AI를 활용한 미래 예측 환경 구축
- 높은 가시성을 지닌 시각화 시스템을 통한 분석력 강화



Databricks Lakehouse 적용 및 고객 지원 서비스는 Databricks Lakehouse 도입 및 적용을 검토하는 고객사를 대상으로 요구 사항을 분석하여 고객사의 요구 사항에 적합한 지원을 전문 컨설턴트와 엔지니어의 개발 컨설팅 프로젝트, 파일럿 프로젝트, 교육 및 기술지원 서비스를 제공합니다.

개발 / 컨설팅 프로젝트

- * 업무 시스템 /프로세스분석
- * 개발 / 컨설팅 수행
- * 교육 및 유지보수

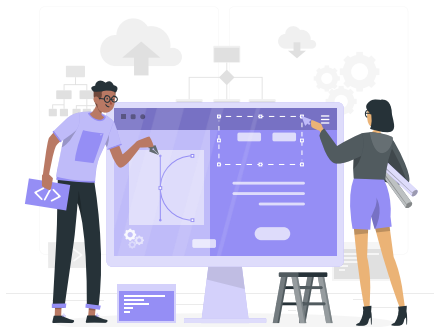
개발 / 컨설팅 범위에 대한 계약 및 수행
상주 / 비상주 프로젝트 수행



파일럿 프로젝트

- * 샘플 데이터 구조 분석
- * 프로토타입 개발
- * 생성 데이터 검증

개발 / 컨설팅 범위에 대한 계약 및 수행
비상주 프로젝트 수행



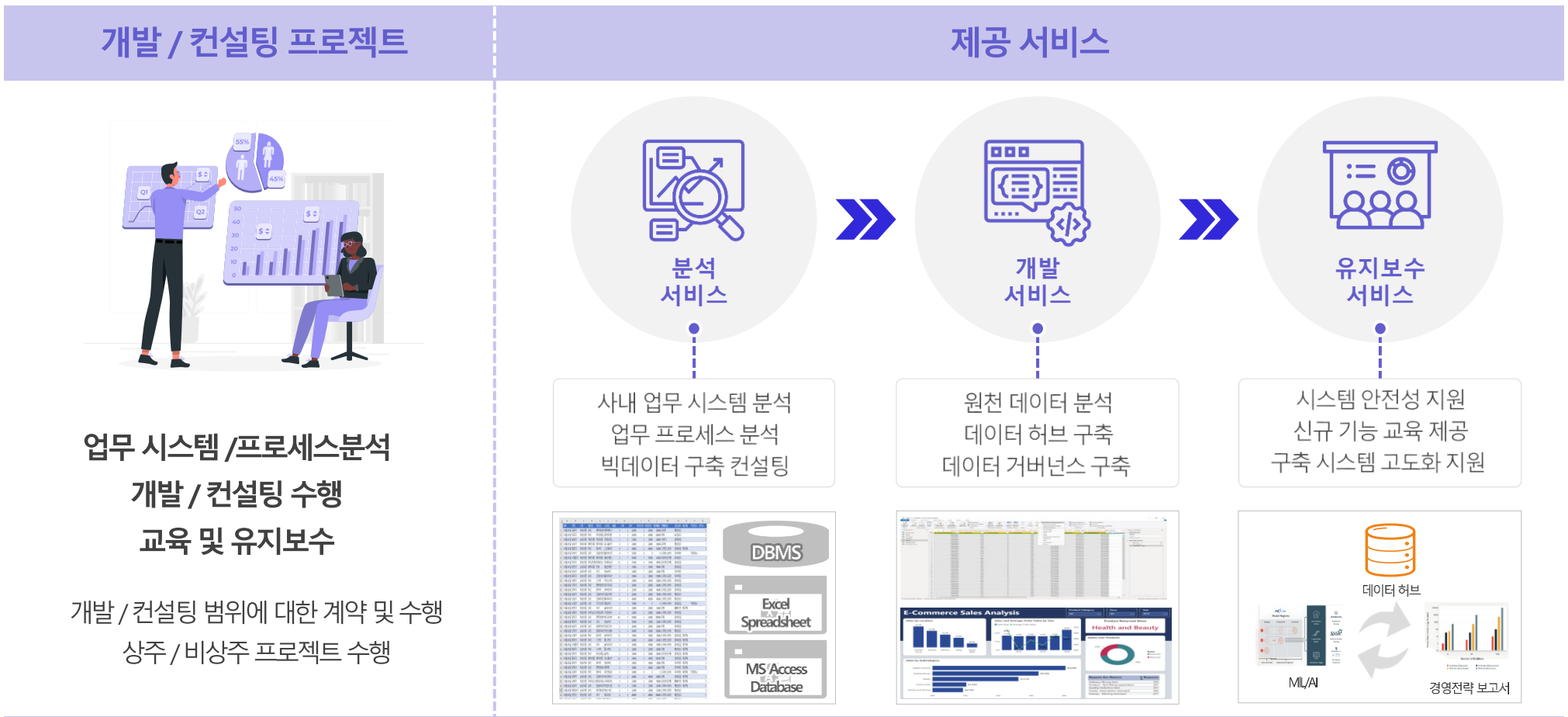
교육 및 기술지원 서비스

- * 교육 지원
- * 개발 지원
- * 기술지원

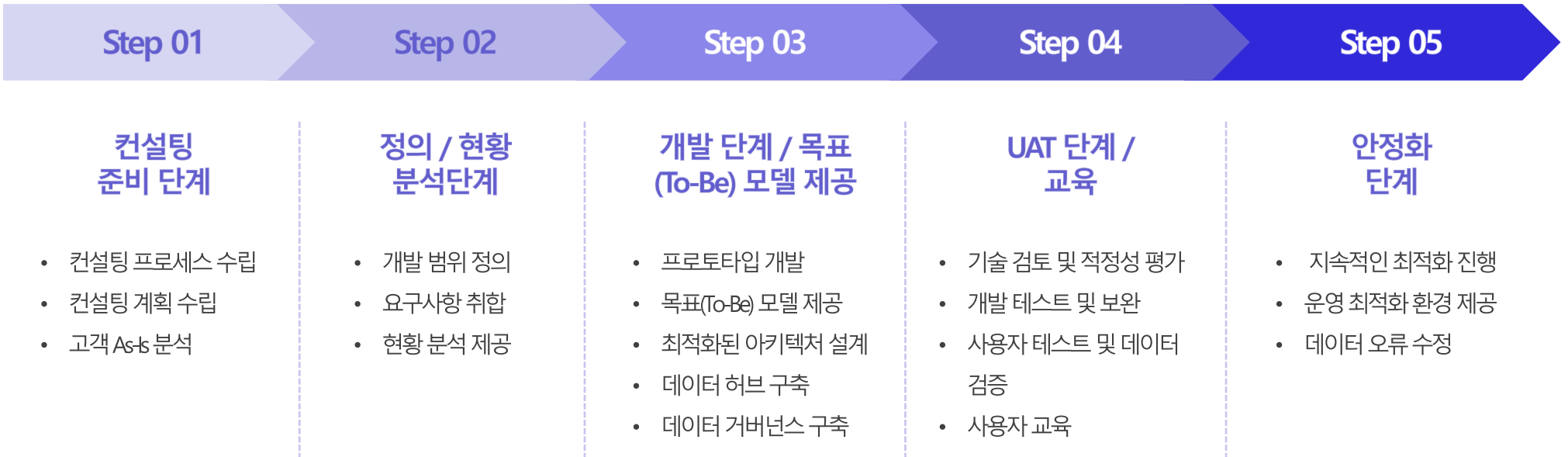
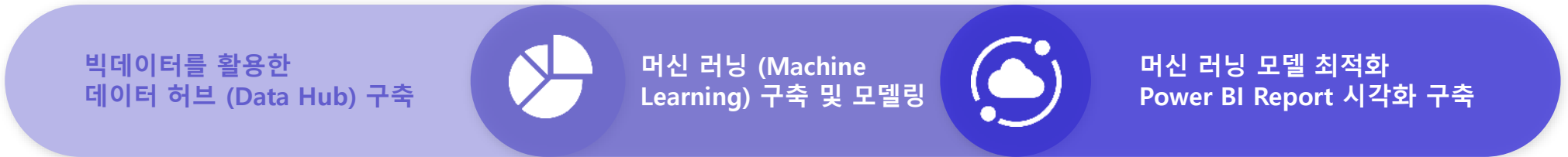
표준 200시간 선 계약 후 실사용



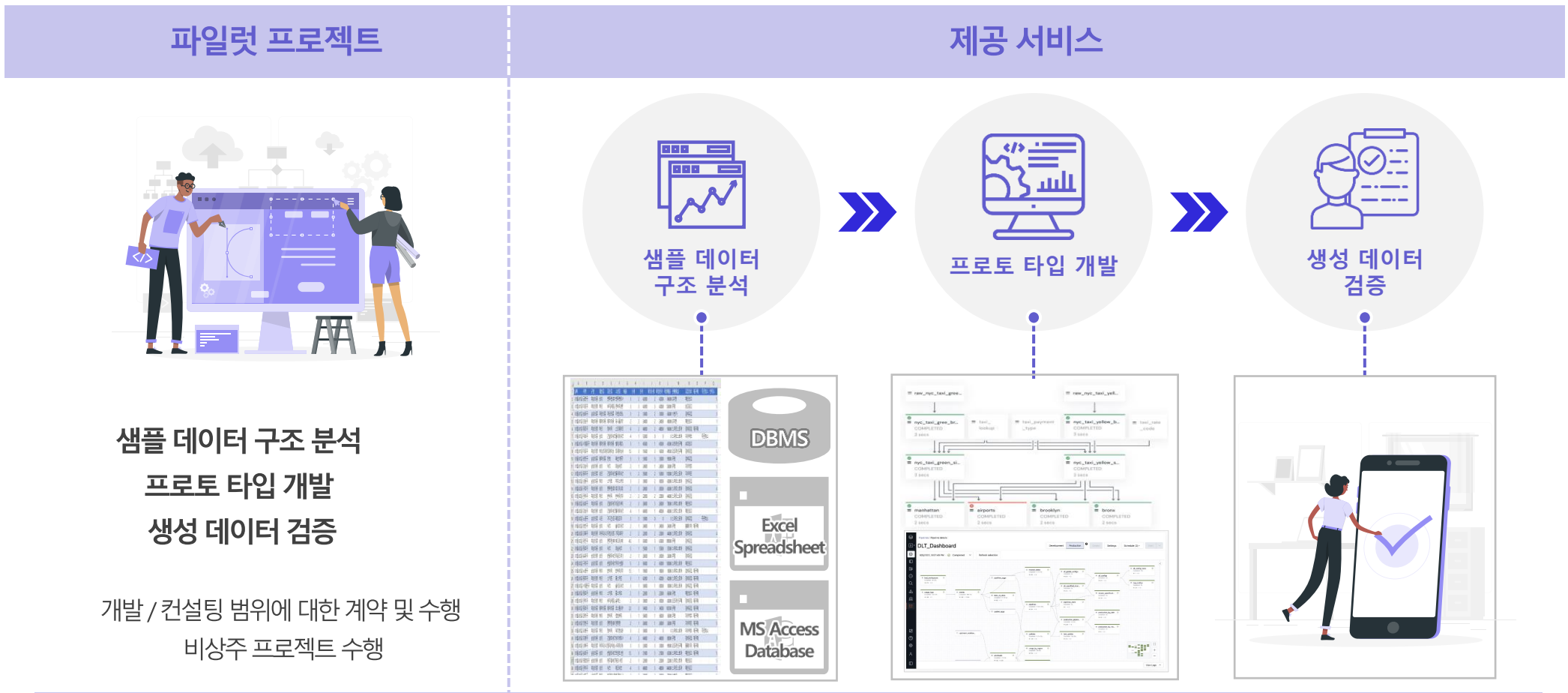
Databricks Lakehouse 개발/컨설팅 프로젝트는 데이터 통합시스템 및 AI/ML, 경영전략 보고서(B) 개발을 Databricks Lakehouse를 통해 구축된 데이터 허브를 이용하고자 하는 고객 대상으로 전문 컨설턴트와 엔지니어를 통한 분석, 개발, 유지보수 서비스를 3단계로 제공하는 서비스입니다.



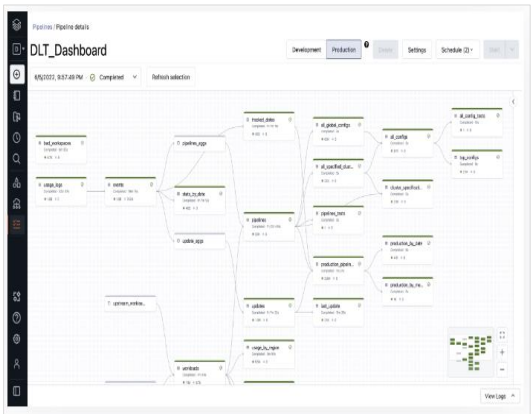
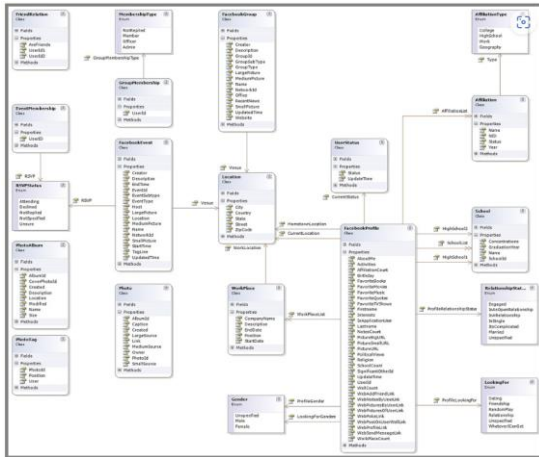
Databricks Lakehouse 개발/컨설팅 프로젝트는 기업의 요구 사항을 파악 후 요건 정의 및 분석 과정을 수행하며 개발 과정에서는 프로토타입 개발, 목표(To-Be) 모델 개발, 개발 적정성 평가, 사용자 교육 등의 과정을 수행 후 안정화 및 최적화 단계를 지원하는 전체 개발 서비스를 제공합니다.



Databricks Lakehouse 파일럿 프로젝트는 솔루션 도입 전, 본 프로젝트에 앞서 실제 사용중인 주요 데이터를 프로토타입 형태로 개발하여 시스템 적합성과 성능을 검증할 수 있도록 샘플 데이터 구조 분석, 데이터브릭스 테이블 생성 및 마이그레이션, 데이터 검증을 3단계로 제공하는 파일럿 서비스입니다.



Step 1 - 현황 분석 및 지표데이터 선정



Step 2 - 데이터 생성 환경 구성



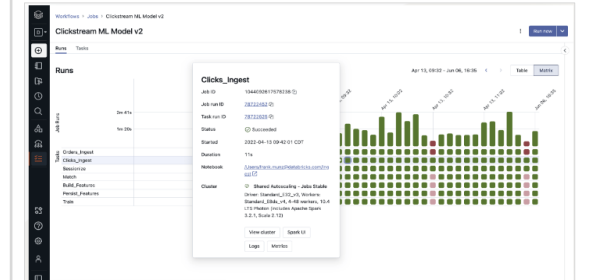
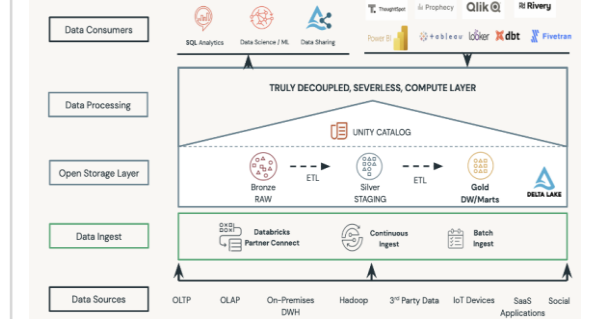
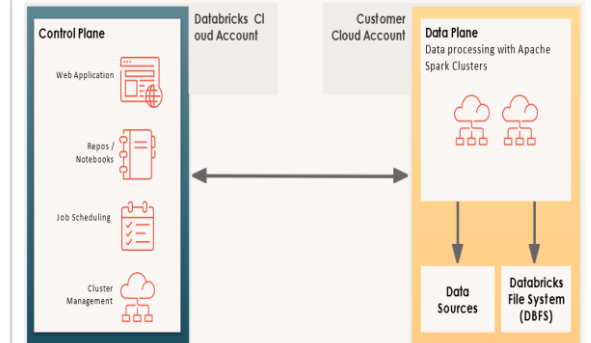
Create Cluster
 New Cluster
 Cluster Name: (Workers: 0.9 GB Memory, 3 Cores, 0 DDB, 11 executors, 4.0 GB Memory, 0.08 Cores, 1 DDB)

Databricks Runtime Version: (4.3 (includes Apache Spark 2.3.1, Scala 2.11))
 Python Version: (2)

Instance Profile: (Free 6GB Memory: As a Community Edition user, your cluster will automatically terminate after an idle period of two hours. For more configuration options, please upgrade your Databricks subscription.)

Availability Zone: (us-west-2c)

Step 3 - 프로토타입 개발



Databricks Lakehouse 교육 및 기술지원 서비스는 기업 내 자체 도입 및 활용으로 계약 시간 내 기술지원이 필요한 고객사를 대상으로 기술지원 서비스를 제공하며, 지속적 활용을 위한 기본 및 정규 교육을 자체 보유하고 있는 커리큘럼을 통해 체계적인 정기 교육으로 지원하고 있습니다.

교육 및 기술지원 서비스	제공 서비스					
 <p>교육 지원 개발 지원 기술지원</p> <p>협정 시간(표준 200시간) 선 계약 후 실사용</p>	서비스 대상	서비스 내용	담당	단 위	위치/지원 방법	시간 차감
	 고객 교육	기본 교육 수강(4시간)	전임 강사	강의당/ 1인 기준	교육센터	0시간 <small>(계약기간내 무제한)</small>
		정규 교육 수강(2일)				9시간 차감
	 요구사항 분석 개발 지원	요구 사항 분석 지원	컨설턴트	시간당		
		시스템 개발 지원				
	 라이브 및 안정화 지원	라이브 지원	컨설턴트	시간당		
		라이브 안정화 지원				
	 기술 지원 서비스	기술 문의 지원	컨설턴트	시간당		
		장애 지원				

빅데이터 체험 교육 커리큘럼

Databricks 빅데이터 분석 과정		Databricks Lakehouse 구축 과정				
시간	상세 커리큘럼	일차	상세 커리큘럼			
(10)	등록		유니티 카탈로그 사용 방법 실습			
(10)	엠클라우드브리지 및 제공 서비스 소개	1일차	Azure Databricks 개요, 엔지니어링 델타 테이블 생성 및 데이터 수집 실습			
(230)	Databricks를 이용한 데이터 엔지니어링	2일차	데이터 처리/통합	- 데이터 레이크 확장을 위한 Delta Lake	Spark 오토 로더 이용 데이터 수집 실습	
				- 데이터 파이프라인을 위한 프레임워크 Delta Live Tables		
				- 데이터 거버넌스 위한 Unity Catalog		
	- 워크플로우 오케스트레이션					
	Databricks를 이용한 데이터 분석			- 데이터 카탈로그 서비스 Unity Catalog		델타 라이브 테이블 실습
				- Spark Notebook을 이용한 데이터 처리와 분석		
Power BI를 이용한 데이터 시각화	- 머신 러닝과 Mllow	머신 러닝 기초, Mllow 실습				
	- Databricks 데이터 불러오기		데이터 관리, 제어 및 시각화 실습			
	- 데이터 모델링					
	-	3일차	머신 러닝 및 데이터 시각화			

빅데이터 교육 일정 및 인원

MS x Databricks	Databricks	정규 교육
월 1회/4시간	월 1회/4시간	월 1회/4일
최대 40명	최대 40명	최대 40명
		

교육 장소 : 위워크 삼성2호점 컨퍼런스룸



6. 엠클라우드브리지 소개

엠클라우드브리지는 디지털 전환시대의 핵심인 클라우드 기반 데이터 분석 및 시각화(Data & BI), 빅 데이터 및 AI (Big Data & AI), 데이터 관리(Data Flow & Automation), 이를 위한 데이터 인프라 및 보안(Data Infra & Security) 컨설팅 서비스 전문 회사로써, 고객 환경에 적합한 서비스 제공을 위해 라이선스 공급, 적용&구축, 유지보수 그리고 교육 서비스를 통합 지원합니다.



회사명
엠클라우드브리지(주)



대표이사
이 혁재

+ 설립일 Establishment	2022년 09월 19일
+ 자본금 Capital	16억
+ 관계사 Affiliated company	다이나믹러닝 인브레인 비주얼데이터 M공작소
+ 업종 Type of business	클라우드 서비스 전문 및 교육 사업
+ 주소 Address	서울특별시 강남구 테헤란로 242 아이타워역삼 7층
+ 연락처 Tel Fax	Tel. 02. 552.9700 Fax. 02. 552. 9799
+ 홈페이지 Homepage	www.mcloudbridge.com

엠클라우드브리지는 메이븐코리아의 마이크로소프트 클라우드 Data & AI 조직에서 분리해 지난 25년간 마이크로소프트(MS) 라이선스 총판 및 리셀러로 활동해온 비주얼 데이터와 합병한 기업으로, “데이터에 가치를 더해 고객의 성장에 공헌한다”는 목표로 설립 되어 Microsoft, Google, Databricks, Snowflake 그리고 Datadog 의 Data & AI 서비스 제공과 'AhnLab', 'Kaspersky'과 보안 솔루션을 제공하고 있습니다.

2013	2015	2022	2023~Present
<p>메이븐코리아 설립</p> <p>Microsoft의 골드 파트너 인증 자격을 갖춘 파트너로써 전문 컨설턴트의 컨설팅 서비스로 Microsoft Dynamics AX, CRM 및 클라우드 솔루션의 구현 서비스를 제공합니다.</p>	<p>마이크로소프트 다이내믹스 프레지던트 클럽 가입</p> <p>메이븐코리아는 뛰어난 비즈니스 성과와 고객에게 가치 있는 비즈니스 솔루션 제공 및 헌신을 인정 받아 Microsoft President's Club에 초대 되었습니다.</p>	<p>MCloudBridge 설립 및 비주얼데이터 M&A 완료</p> <p>엠클라우드브리지는 2022년 설립되어 지난 25년간 마이크로소프트 총판 및 리셀러 파트너로 활동해 온 비주얼데이터와 M&A를 완료했습니다.</p>	<p>클라우드 기반 데이터 및 AI 전문 컨설팅 회사</p> <p>Microsoft, Google, Databricks, Snowflake 그리고 Datadog 인증 자격을 갖춘 전문 컨설팅 및 지원 그룹을 보유하고 고객에게 클라우드 데이터 분석 및 적용 서비스를 제공하는 Data & AI 전문 컨설팅 회사입니다.</p>

엠클라우드브리지는 고객 환경에 적합한 클라우드 데이터 서비스 제공을 위해 빅데이터 및 인프라 기술 협력사, 개발센터, 각 지역 거점 오피스에서 유기적으로 협업하고 있으며, 자체 보유하고 있는 클라우드 스쿨을 통한 고객 맞춤 교육 서비스를 제공하고 있습니다.

Big Data 협력

I | N | B | R | E | I | N

엠클라우드브리지 교육 센터



Data Infra/Security 협력

VisualData

본사

서울 강남구 테헤란로 242
아이타워역삼 7층

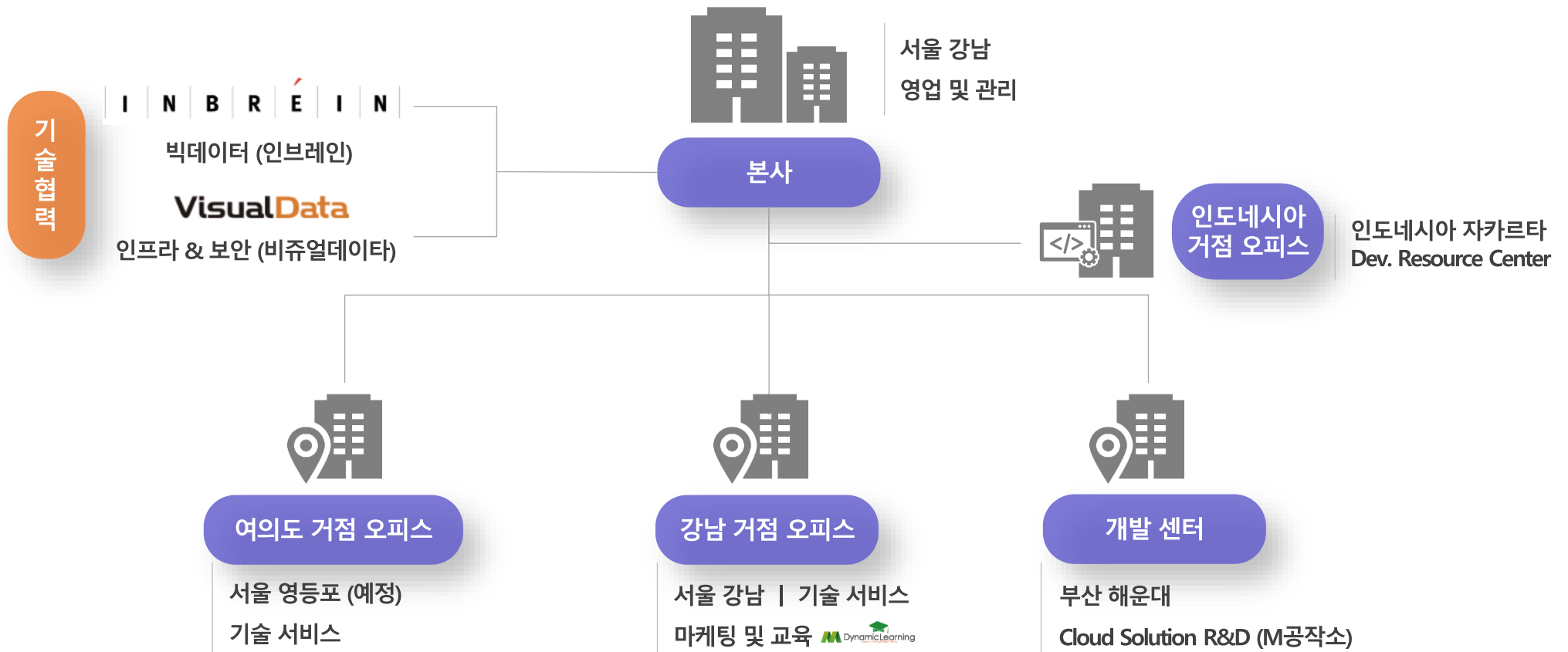
개발 센터

부산 | 해운대구

거점 오피스

서울 | 강남구
인도네시아 | 자카르타
서울 | 여의도 (2023년 4월 오픈 예정)

엠클라우드브리지는 고객 환경에 적합한 클라우드 데이터 서비스 제공을 위해 빅데이터 및 인프라 기술 협력사, 개발센터, 각 지역 거점 오피스에서 유기적으로 협업하고 있으며, 자체 보유하고 있는 교육센터를 통한 고객 맞춤 교육 서비스를 제공하고 있습니다.

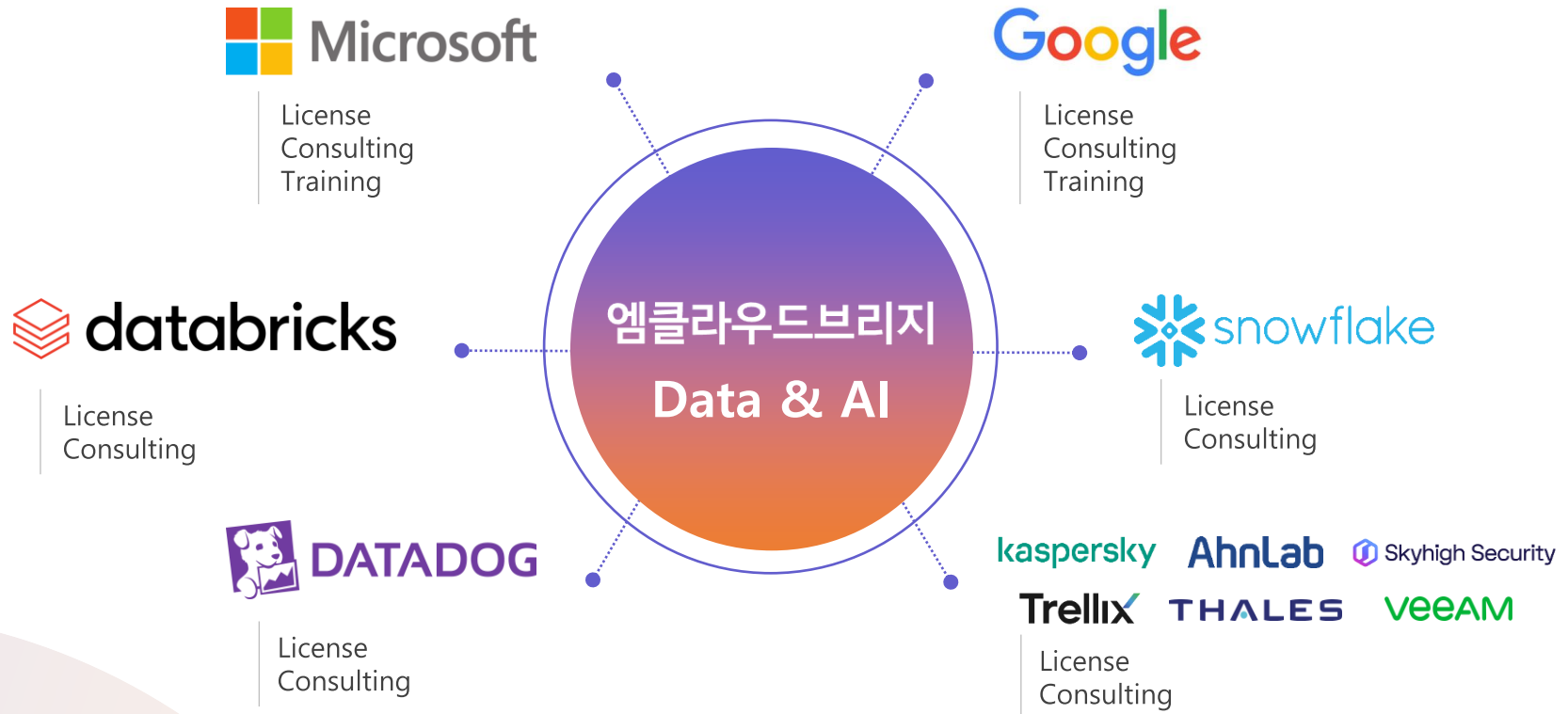


엠클라우드브리지는 디지털 전환 시대, 고객 환경에 적합한 클라우드 데이터 분석 환경 제공을 위해 조직된 각 분야 평균 10년 이상의 전문가 그룹으로서, Data 서비스 전문 컨설팅 기업 도약이라는 목표를 가지고 투자 합병을 통한 조직의 실행력 및 전문성을 갖추어 나아가고 있습니다.



엠클라우드브리지는 Data 서비스 전문 컨설팅 기업 도약이라는 비전으로 기존 서버 및 보안(IaaS & PaaS) 중심의 클라우드 서비스 시장에서 Data 플랫폼, 분석 서비스 제공 차별화 전략을 기반 Big Data & AI, Data Infra & Security 서비스 역량 강화 투자를 전개할 계획입니다.









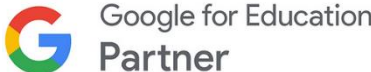










“데이터에 가치를 더하여 고객의 성장에 공헌합니다”

클라우드 기반 Data & AI 전문 컨설팅 기업

Specialized Consulting Firm in Data & AI Cloud System







엠클라우드브리지는 Data & AI 전문 컨설팅 서비스 기업으로써 고객 환경에 적합한 클라우드 데이터 분석 환경 및 서비스 제공을 위해 Microsoft, Google, Databricks 등 인증 자격을 갖춘 전문 컨설팅 및 지원조직을 갖추고 라이선스 공급, 적용 & 구축, 유지보수 그리고 교육 서비스를 통합 지원합니다.

보유 완료 및 취득 자격 (Data & AI)

Microsoft Partner		Google Partner		Databricks Partner	
					
				Snowflake Partner	
					
				Datadog Partner	
					

엠클라우드브리지는 Data Infra & Security 전문 컨설팅 서비스 기업으로써 고객 환경에 적합한 온프레미스 및 클라우드 데이터 보안 환경 제공을 위해 Kaspersky, AhnLab, Skyhigh, Trellix, Thales, Veeam 인증 자격을 갖춘 전문 조직을 통한 라이선스 공급, 적용 & 구축, 유지보수, 교육 서비스를 통합 지원합니다.

보유 완료 및 취득 자격(Data Infra & Security)

<p>Kaspersky</p>	<p>Ahnlab</p>	<p>Skyhigh Security</p>
<p>Platinum Partner</p>	<p>Developing Partner</p>	<p>Gold Partner</p>
<p>  Kaspersky Endpoint Security for Business </p>	<p> AhnLab V3 Endpoint Security </p>	<p>  Security Service Edge </p>
<p>Trellix</p>	<p>Thales</p>	<p>Veeam</p>
<p>Gold Partner</p>	<p>Developing Partner</p>	<p>Silver Value-Added Reseller</p>
<p>  Endpoint Security  Data Security </p>	<p>  CipherTrust Manager </p>	<p>  Backup & Replication™ </p>

<p>Data & BI</p> <p>데이터 분석 & 시각화</p>	<p>Big Data & AI</p> <p>빅데이터 및 AI</p>	<p>Data Flow & Automation</p> <p>데이터 관리 및 자동화</p>	<p>Data Infra & Security</p> <p>데이터 인프라 및 보안</p>
 <p>중외제약</p> <p>THE HYUNDAI</p>  <p>영보화학주식회사 YOUNGBO Chemical Co., Ltd.</p>  <p>BAIKSAN</p>  <p>(주)한국문화진흥</p>  <p>SIMPAC</p> <p>FERRERO</p> <p>KoMiCo</p>	 <p>LOTTE INEOS</p>  <p>영보화학주식회사 YOUNGBO Chemical Co., Ltd.</p>  <p>BAIKSAN</p>  <p>(주)한국문화진흥</p> <p>'TORAY' 도레이첨단소재</p>  <p>SONGWON</p>	<p>THE HYUNDAI</p>  <p>LOTTE INEOS</p>  <p>영보화학주식회사 YOUNGBO Chemical Co., Ltd.</p>  <p>EDWARDS</p> <p>TOSHIBA</p>	 <p>MINISTOP</p>  <p>CJ BIOSCIENCE</p> <p>DOHWA</p> <p>FERRERO</p>  <p>LT 메탈</p>  <p>EA</p>  <p>대인제지</p> <p>CERAGEM</p> <p>OTOS®</p>  <p>GC 녹십자</p>  <p>SK 하이닉스</p>  <p>SK 하이스택</p>  <p>ILJINRAD</p>  <p>TESS</p>  <p>BAIKSAN</p>  <p>MEGABOX LIFE THEATER™</p>  <p>phognix HOTELS & RESORTS</p>  <p>JTBC</p> <p>The JoongAng</p>

Data & BI

Big Data & AI

Data Flow & Automation

Data Infra & Security

Thank You

T. 02.552.9700

E. info@mcloudbridge.com

H. www.mcloudbridge.com

데이터에 가치를 더하여 고객의 성장에 공헌합니다.

Specialized Consulting Firm in **Data & AI** Cloud System