

Data & BI

Big Data & AI

Data Flow & Automation

Data Infra & Security

Microsoft Azure Databricks SAP 마이그레이션 전략

데이터에 가치를 더하여 고객의 성장에 공헌합니다.

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

1. Why Azure Databricks
2. Why Migrate from SAP
3. Azure Databricks - SAP 마이그레이션 전략
4. Azure Databricks 단계별 프로세스
5. 『Azure Databricks Migration for SAP』 PoC 소개
6. 엠클라우드브리지 소개

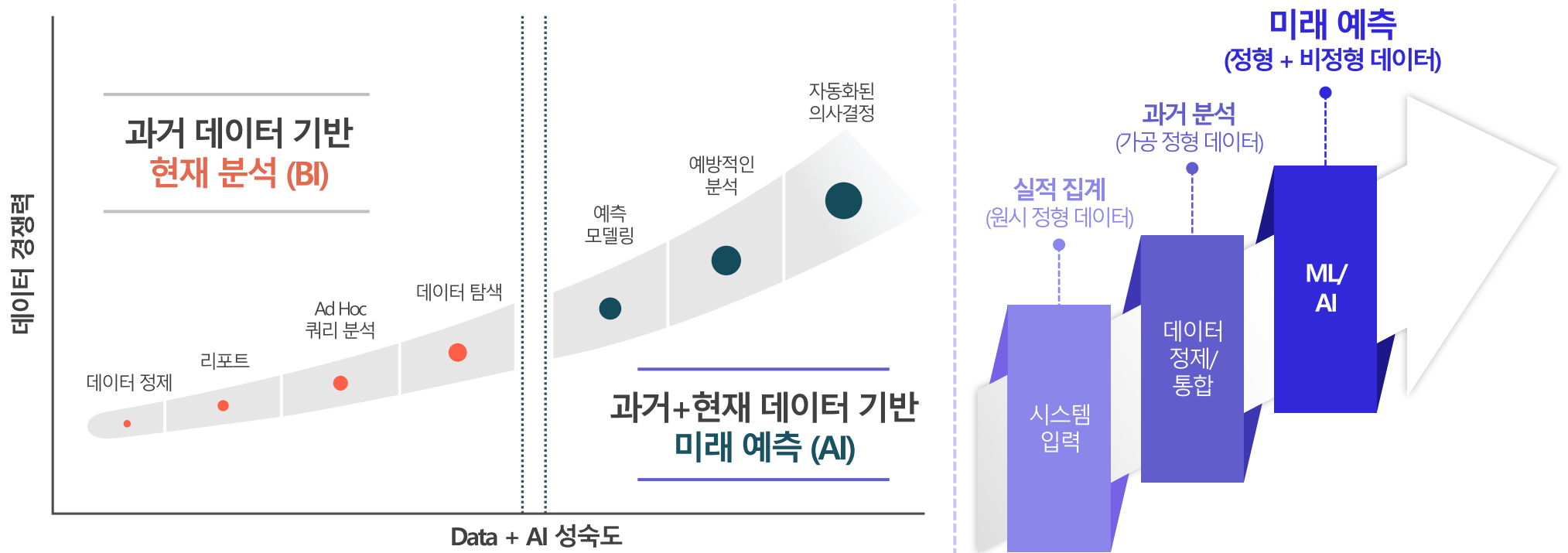


1. Why Azure Databricks

Digital Transformation 흐름에 따른 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 필수 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



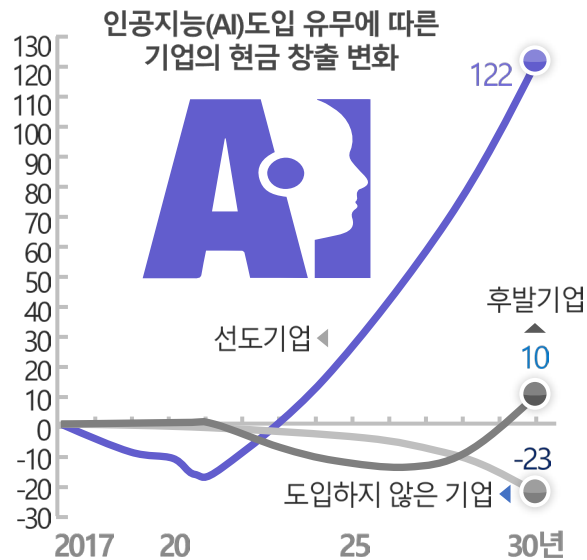
대량의 자연어 데이터를 이용하여 미래 예측이 가능한
빅데이터 플랫폼

빅데이터 플랫폼의 차별화된 경쟁력 중 AI/ML은 미래 예측을 위한 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 AI/ML 도입이 미래 기업의 경쟁력 결정

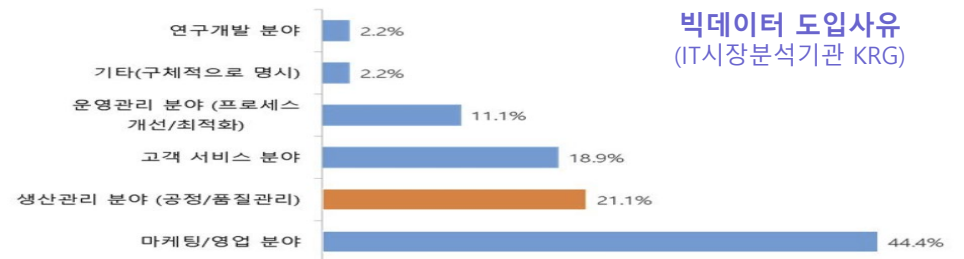
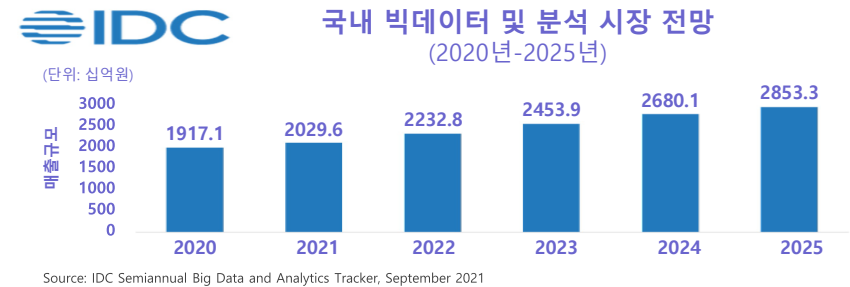
**AI 성공적 도입한 선두기업
10%만 살아남을 것**

(출처: 경영 컨설팅 기업 맥킨지, 단위: %)



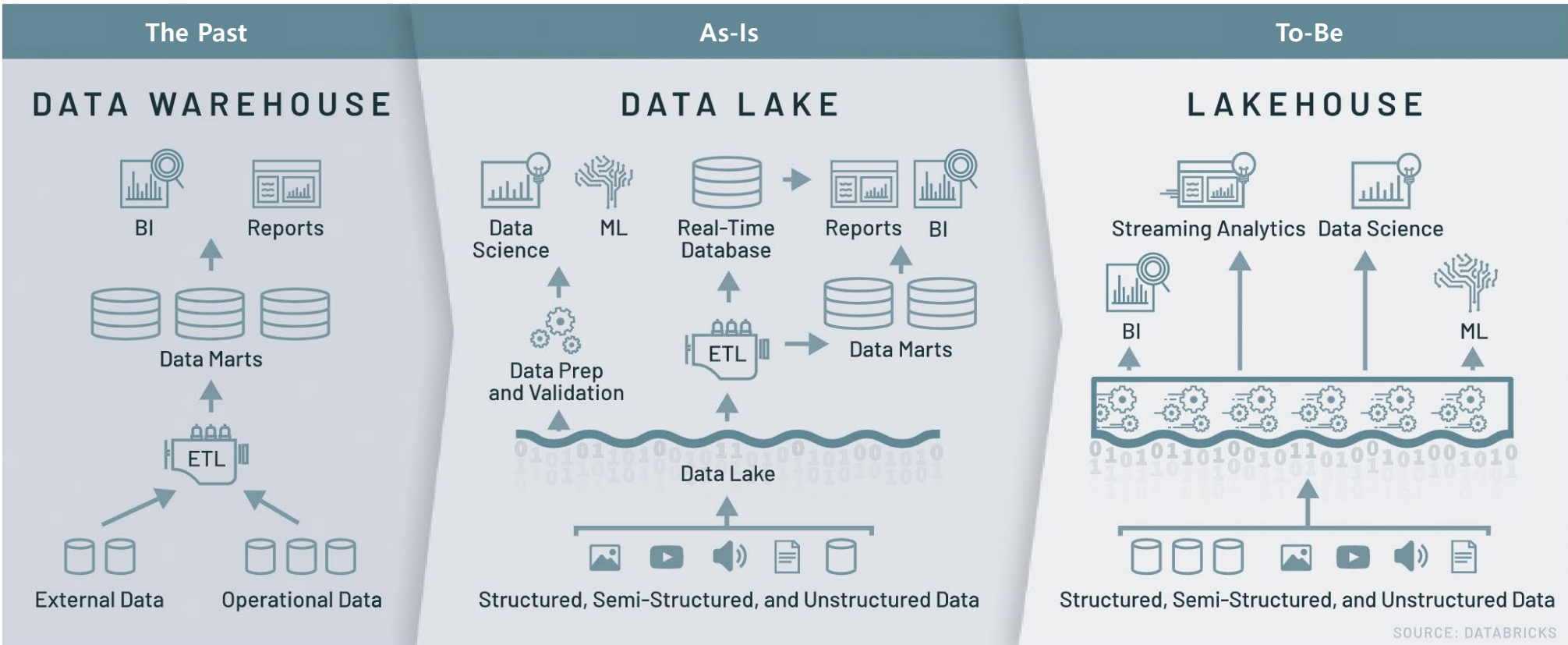
**국내 기업들 지속적인 투자 확대 중
(미래 경쟁력 제고와 생존의 기본 조건 인식)**

출처: IDC & IT시장분석기관 KRG



빅데이터 플랫폼의 차별화된 경쟁력 중 Lakehouse는 데이터 레이크와 데이터 웨어하우스의 최고의 요소를 결합한 새로운 개방형 아키텍처로, 저렴하고 신뢰성 높은 스토리지를 사용할 수 있게 된 현대사회에서 데이터 웨어하우스를 재설계해야 한다면 이러한 이점을 누릴 수 있습니다.

Lakehouse = Data Warehouse + Data Lake



데이터브릭스(Databricks)는 기업의 클라우드 통합 데이터 관리 플랫폼으로 전세계 7,000개 이상의 글로벌 기업과 Fortune 선정 500대 기업의 40% 이상 데이터 엔지니어링 및 데이터 분석, AI/ML을 활용하여 데이터 통합 및 관리하고 협업하는 클라우드 Data & AI 전문 플랫폼입니다.

기업의 클라우드 통합 데이터 관리 플랫폼



데이터, 분석, AI 워크로드를 통합하는
클라우드 데이터 플랫폼

Customers

7000+
across the globe

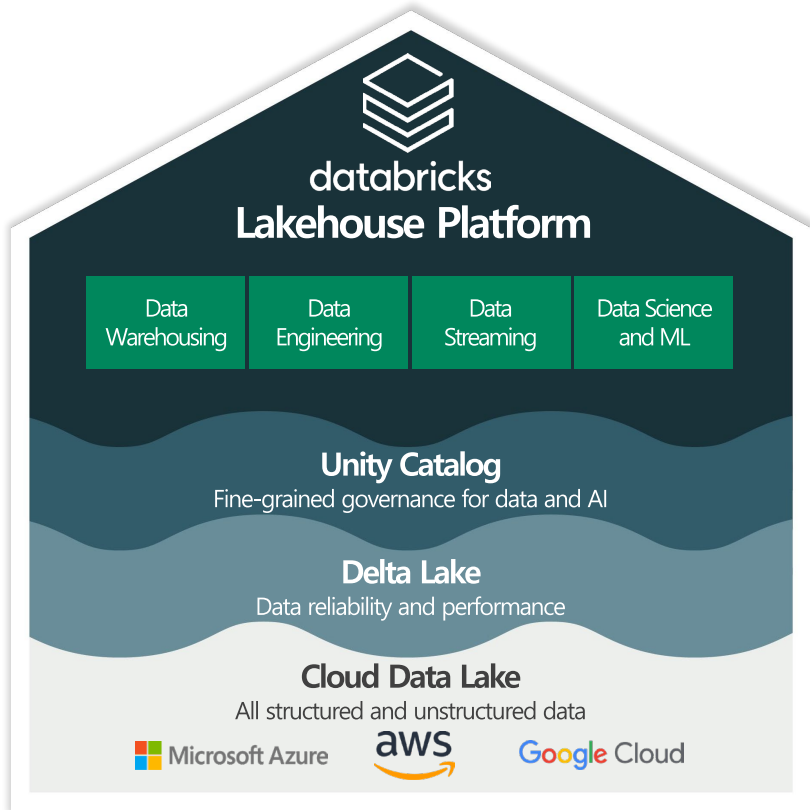


Original creators of:



데이터브릭스의 3가지 특징점은 데이터 저장 및 분석, AI/ML 등이 통합된 플랫폼에서 개발 되었으며, 오픈 소스, 국제 표준 형식으로 개발되어 편리하게 타시스템과 연결할 수 있고 데이터 분석가, 데이터 엔지니어, ML 엔지니어 등이 함께 데이터와 모델을 공유하며 협업해서 관리할 수 있습니다.

데이터브릭스 (Databricks) 3가지 특징점 요약



- ① 개방성(Open) & 통합성(Unify)**

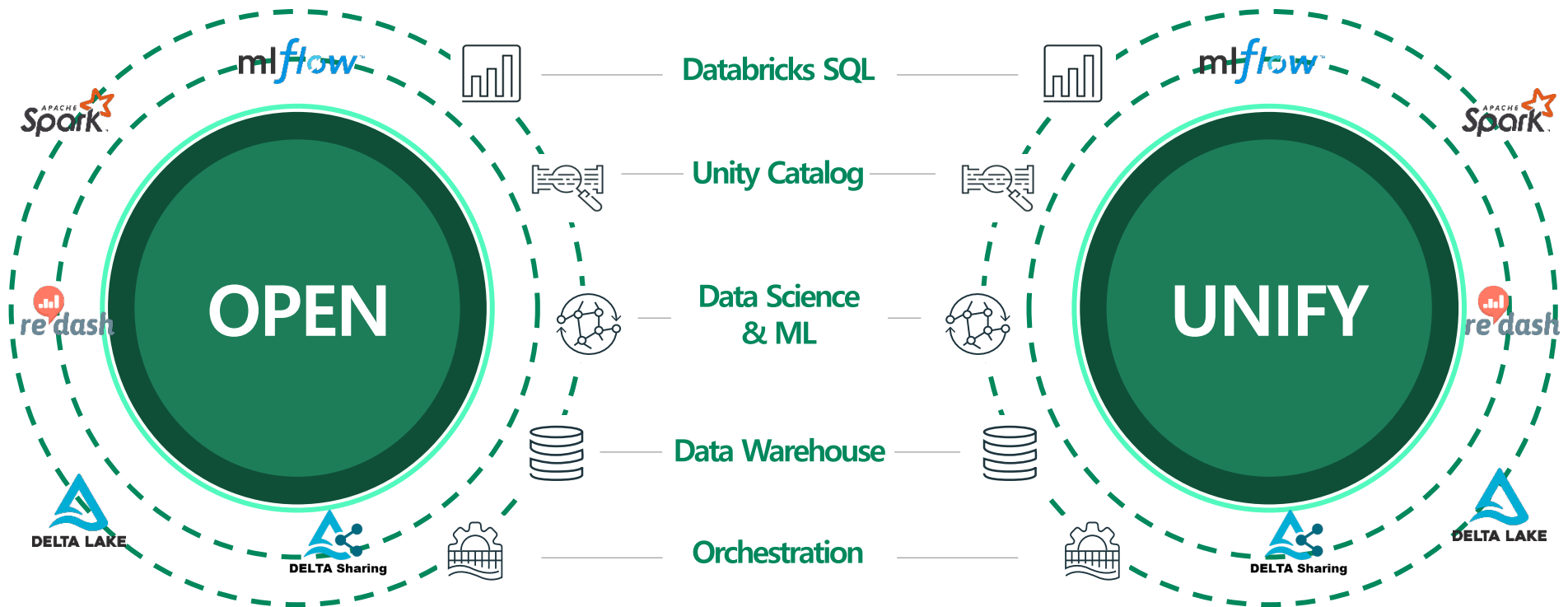
오픈 소스, 표준 형식으로 개발 되어 많은 종류의 시스템 및 개발 플랫폼과 연결 가능
- ② AI/ML 효율화 & 단순화**

단일 플랫폼 내에서 구성된 데이터를 통해 머신러닝과 AI를 지향함에 따라 데이터 관리의 효율화를 달성
- ③ 영역간 협업 강화**

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

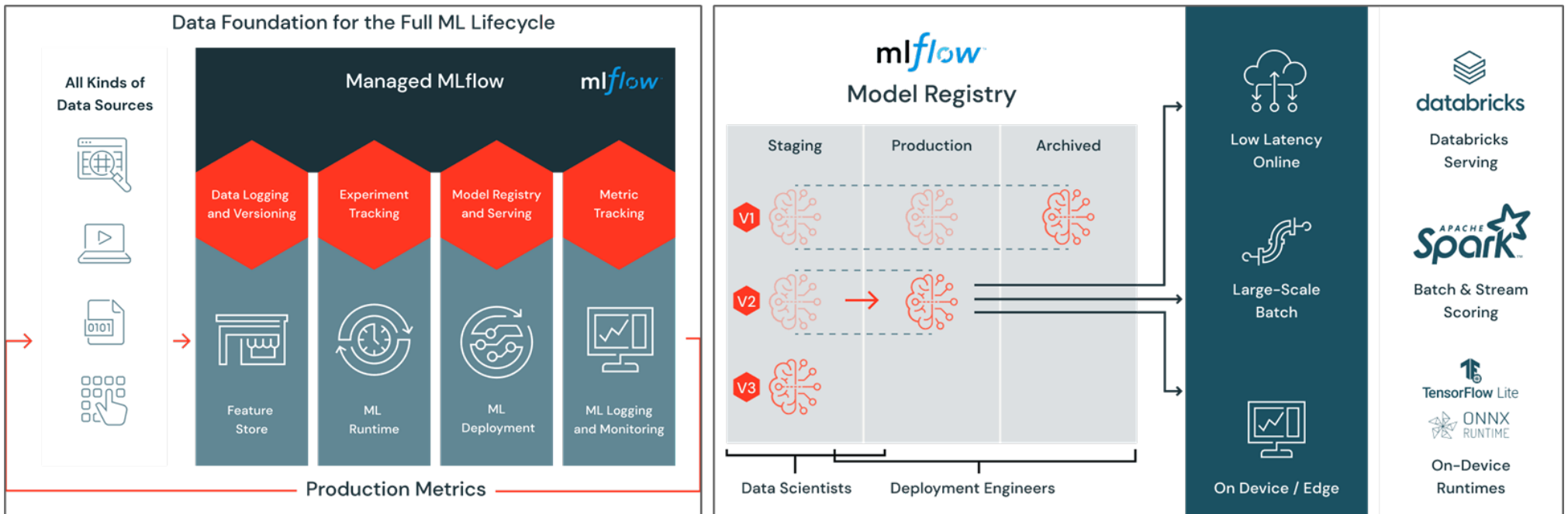
데이터브릭스는 개방성(Open) & 통합성(Unify)의 데이터 플랫폼으로서 외부의 많은 오픈 소스, 데이터 등과 데이터브릭스 플랫폼의 Databricks SQL, Serverless, Unity Catalog, Streaming, Restructured ML, Orchestration, Delta Table 등의 핵심 기술을 이용하여 효율적인 데이터 관리를 합니다.

개방성 (Open) & 통합성 (Unify) 데이터 플랫폼



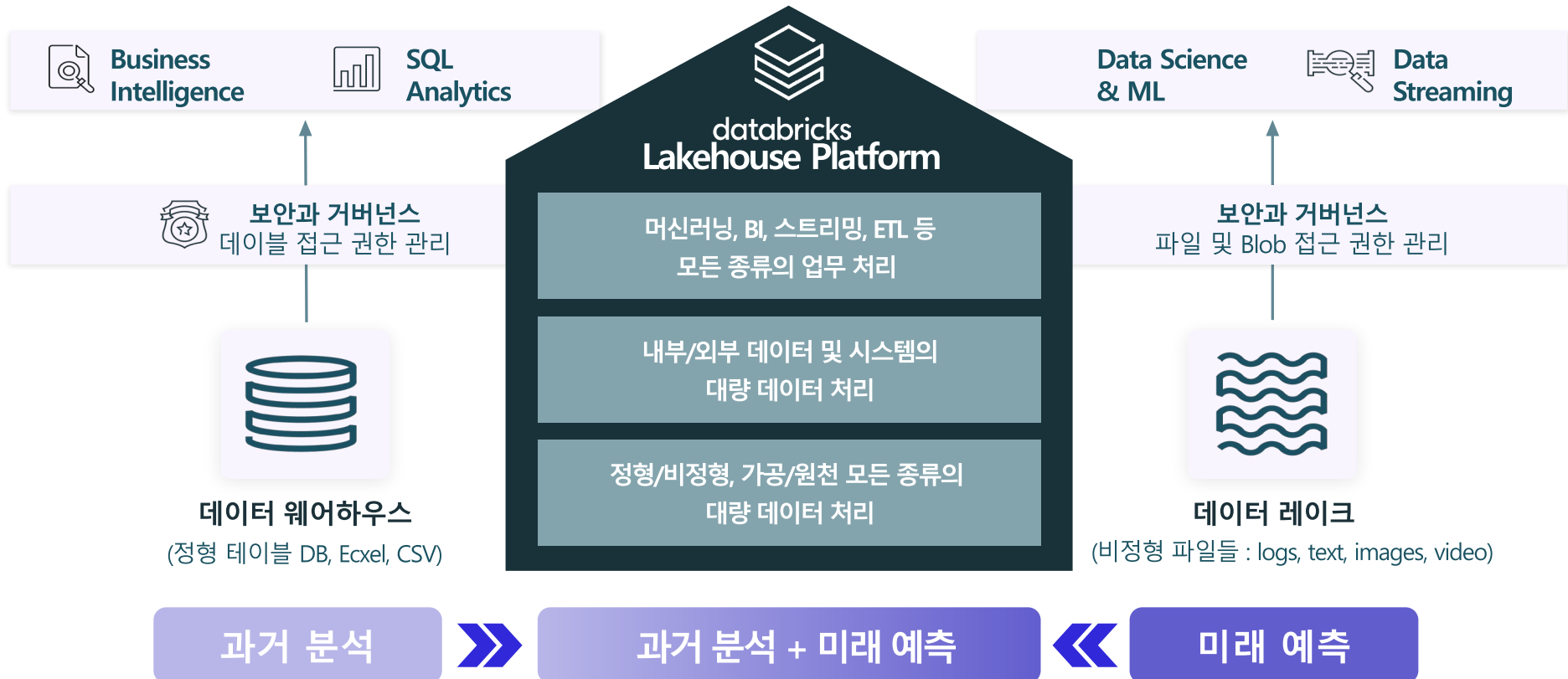
데이터브릭스는 AI/ML 효율화 & 단순화 데이터 플랫폼으로서 레이크하우스 기반으로 구축된 모든 규모의 모든 유형의 데이터에 액세스하고 탐색할 수 있습니다. 또한 대규모 언어 모델을 사용하여 대규모로 짧은 시간에 ML 모델을 배포할 수 있는 장점을 가지고 있습니다.

AI/ML 효율화 & 단순화 데이터 플랫폼



데이터브릭스는 영역간 협업이 강화된 데이터 플랫폼으로서 단일 플랫폼 내에서 모든 종류의 데이터를 ETL 하며 머신러닝, BI, 스트리밍, 분석 등 다양한 업무 처리가 가능한 단일 보안, 거버넌스 모델로 데이터를 통합하고 관리하는 빅데이터 플랫폼을 구성 합니다.

영역간 협업이 강화된 데이터 플랫폼



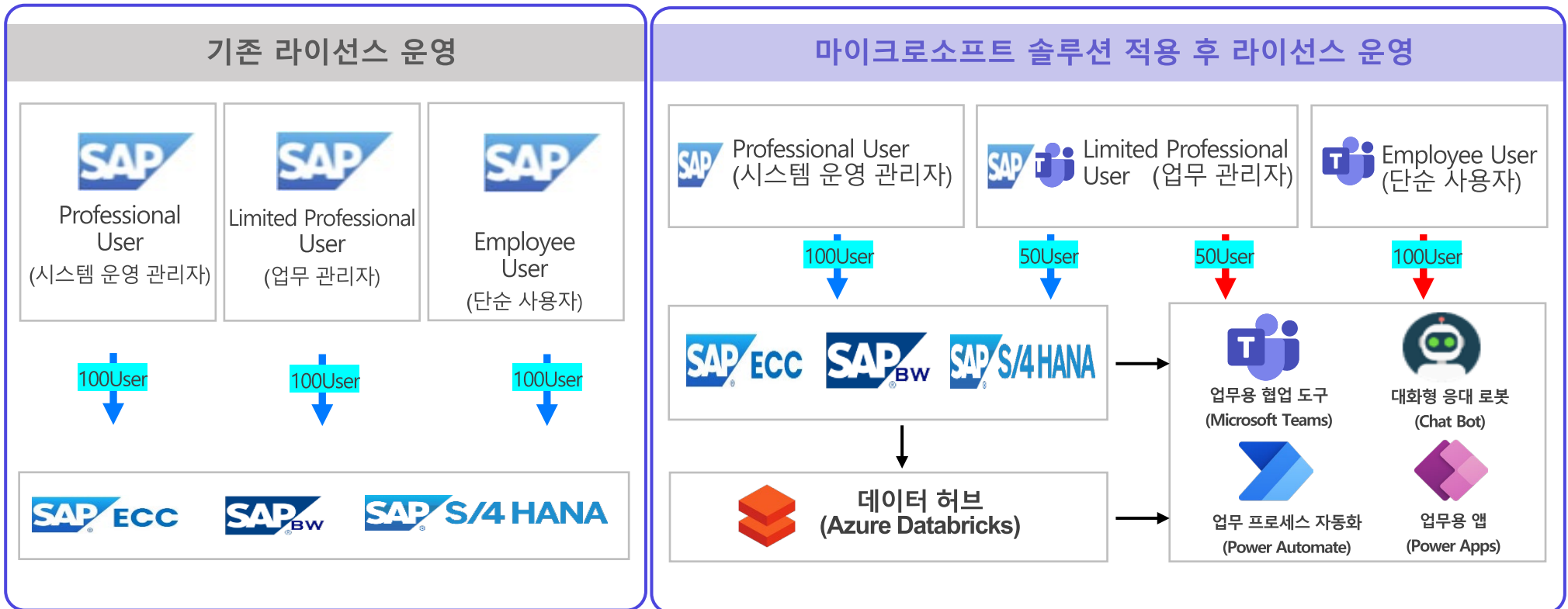


2. Why Migrate from SAP

Microsoft 솔루션은 사용자 라이선스의 경제적 합리성을 추구하여 SAP 운영자/관리자 및 IT개발자를 제외한 일반 사용자는 제공된 Data Set을 M365나 BI를 통해 업무를 수행하게 되므로 업무 생산성을 높이며 SAP 라이선스 미사용으로 인한 비용 효율화를 만들 수 있습니다.

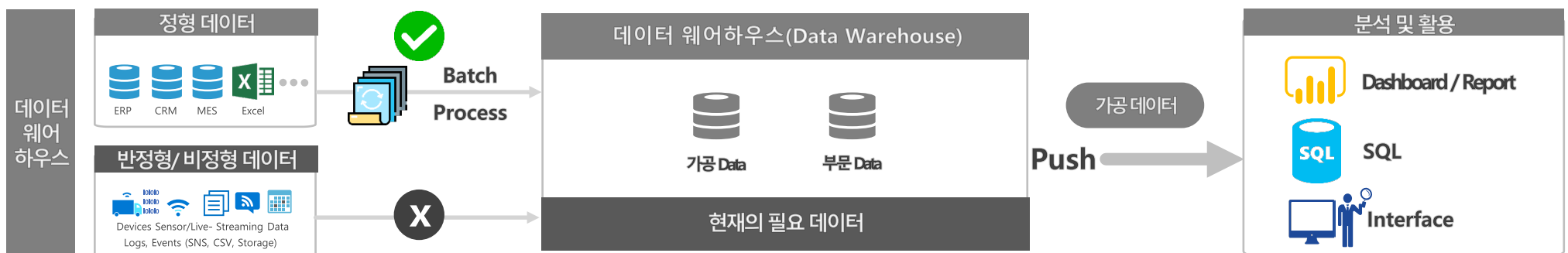
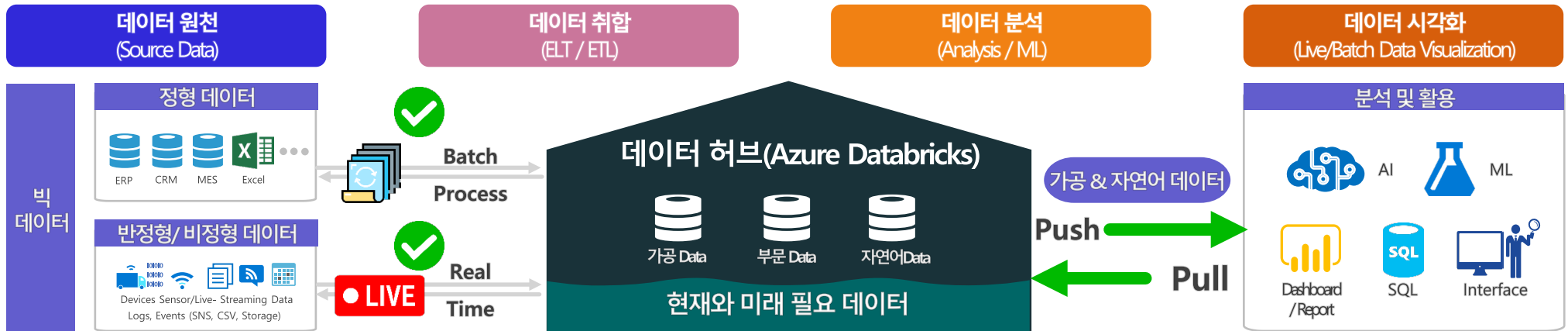
사용자 라이선스 경제적 합리성

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



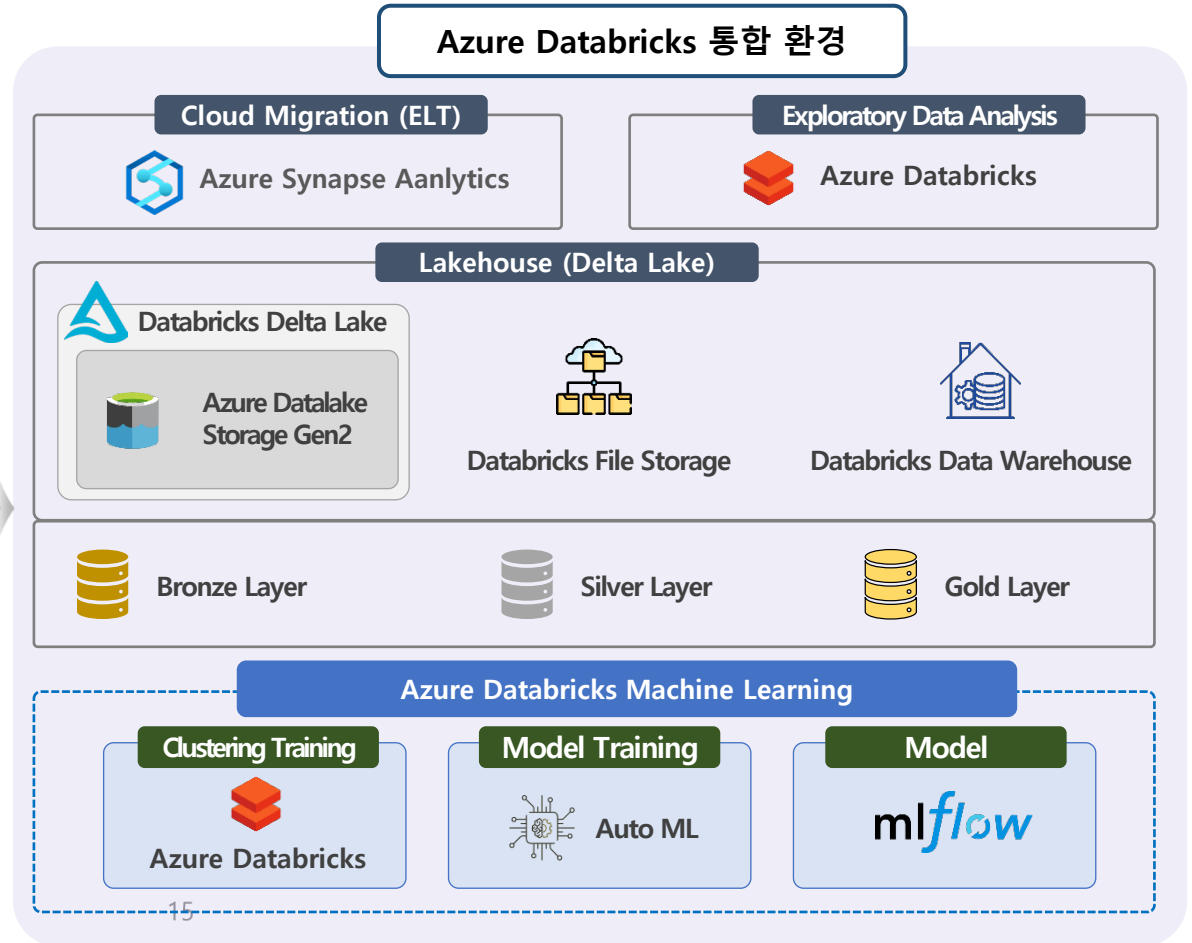
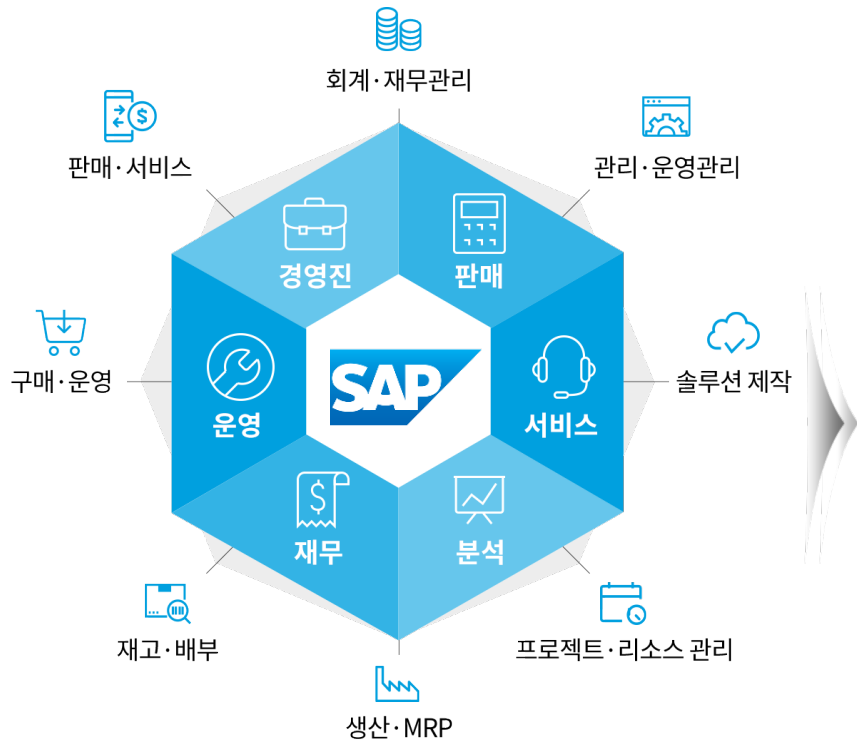
Azure Databricks 레이크하우스 기반 SAP 데이터 허브(SAP Data Hub) 구축으로 기업의 실시간 경영 상황이 반영된 데이터 분석 및 AI/ML 예측 분석 기반 경영 환경 구축을 위해 SAP 데이터 및 내외부의 정형, 비정형 데이터를 단일 플랫폼에서 분석할 수 있는 레이크하우스를 적용합니다.

레이크하우스 기반 SAP 데이터 허브(SAP Data Hub) 구축



Azure Databricks AI/ML을 통해 SAP 데이터 가치 향상에 기여할 수 있습니다. 이는 운영계와 분석계(정보계) 시스템을 철저하게 분리하여 정보계의 통합 환경을 구축하여 이 환경에서의 머신러닝에 최적화된 데이터를 산출하여 SAP 데이터 가치를 향상시킬 수 있는 환경을 만듭니다.

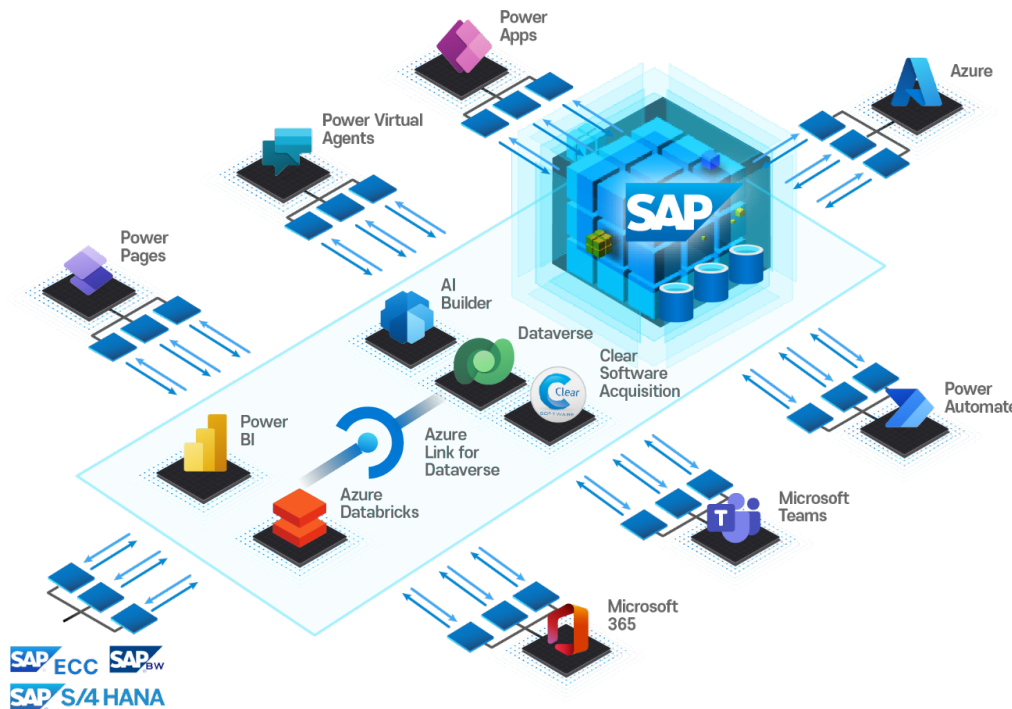
Azure Databricks AI/ML을 통한 SAP 데이터 가치 향상



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.

Microsoft & SAP 통합 솔루션을 통한 비즈니스 혁신

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





Microsoft Teams



Power Apps



Power Automate



Power BI



Power Virtual



Power Pages



databricks



OpenAI



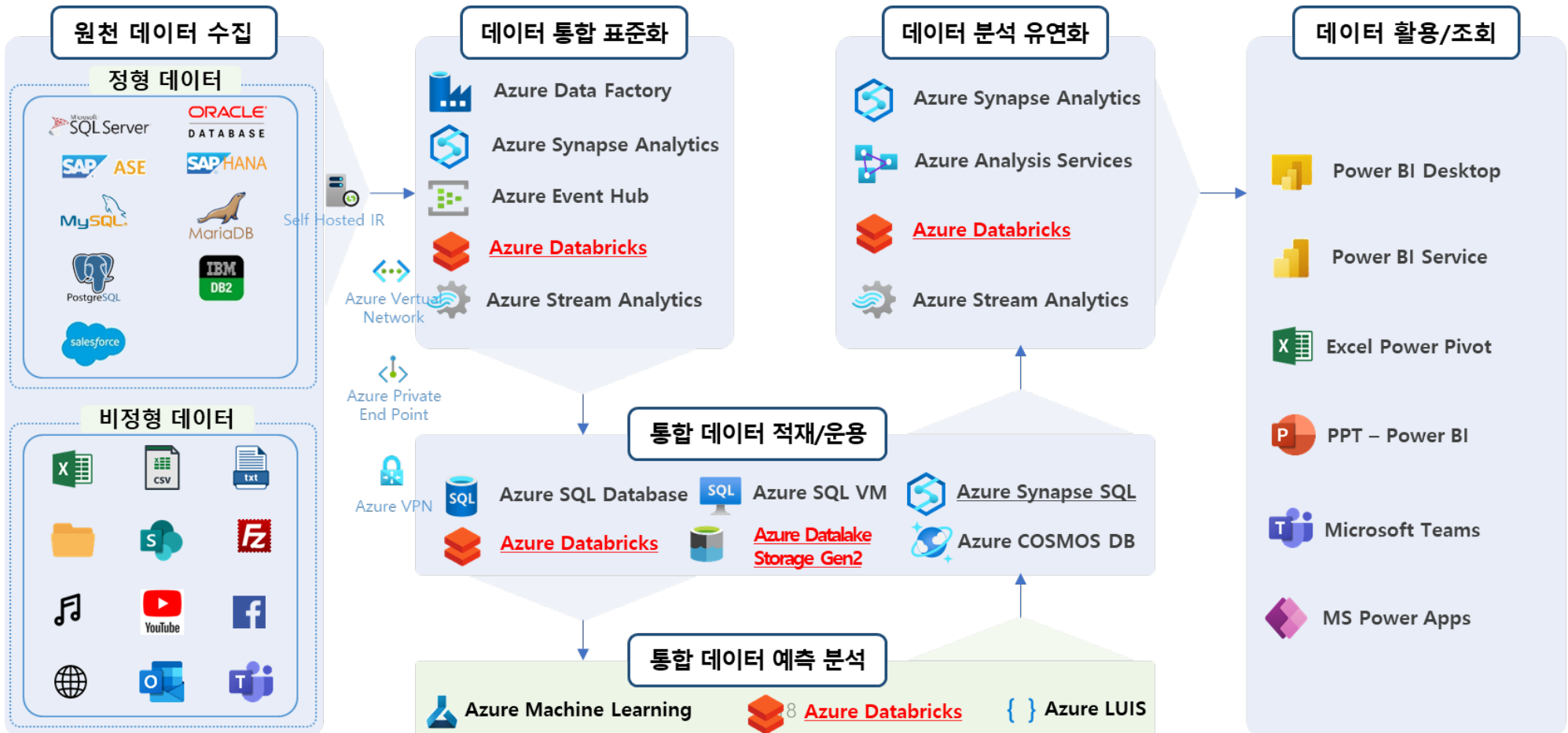
Microsoft



3. Azure Databricks - SAP 마이그레이션 전략

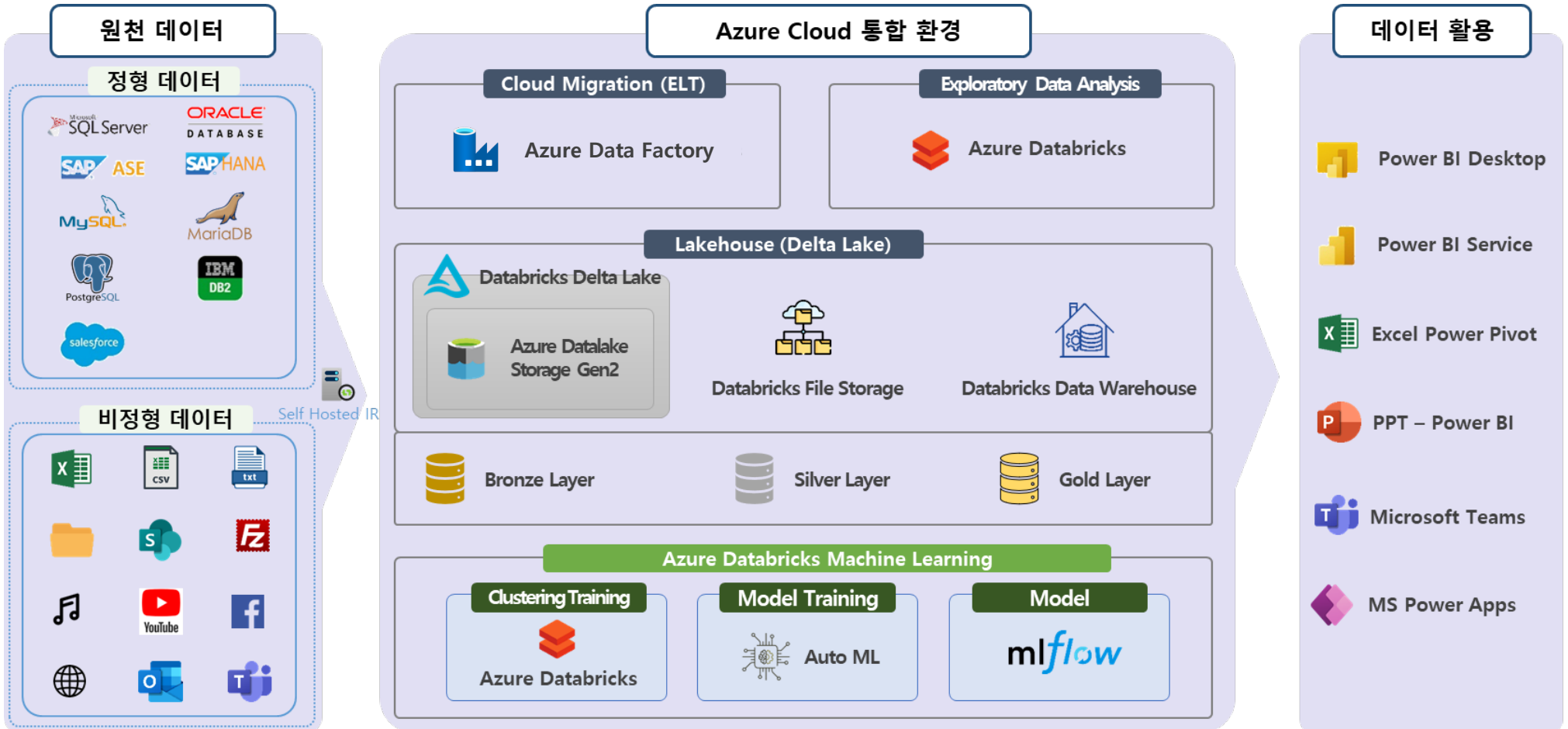
Powerful Azure Service Architecture는 원천 데이터로부터 Azure Cloud로 Data를 Migration 또는 수집시 Azure에서 Service중인 여러 가지 Solution을 선택하여 DW 형태로 저장한 다음 통계분석과 ML기법과 모델을 활용해 예측 데이터를 산출을 통해 BI Report로 시각화하는 일련의 프로세스로 진행됩니다

SAP · Databricks 데이터 마이그레이션 제안 아키텍처



SAP · Databricks 데이터 마이그레이션 제안 아키텍처는 개방성과 통합성을 보유한 Lakehouse를 기반으로 데이터 허브를 구축하여 SAP 데이터를 중심으로 기업에 연관된 모든 정보를 배치 또는 실시간 마이그레이션하여 통합 관리함으로써 데이터 분석(BI) 및 AI/ML 사용 및 확장성을 최대화시킵니다.

SAP · Databricks 데이터 마이그레이션 제안 아키텍처



SAP 마이그레이션의 단계별 상세 업무는 데이터 마이그레이션(Data Migration) 을 시작으로 ETL 및 코드 마이그레이션(Code Migration)을 순차적으로 진행하여, 기업의 시스템 환경에 따라 각 시스템 기반 마이그레이션과 수작업 마이그레이션을 아래와 같은 단계별 상세 업무 내용을 토대로 제공합니다.

SAP · Databricks 마이그레이션의 단계별 상세 업무 진행

Track 1 : 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

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

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

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)

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

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

Track 3 : 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

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

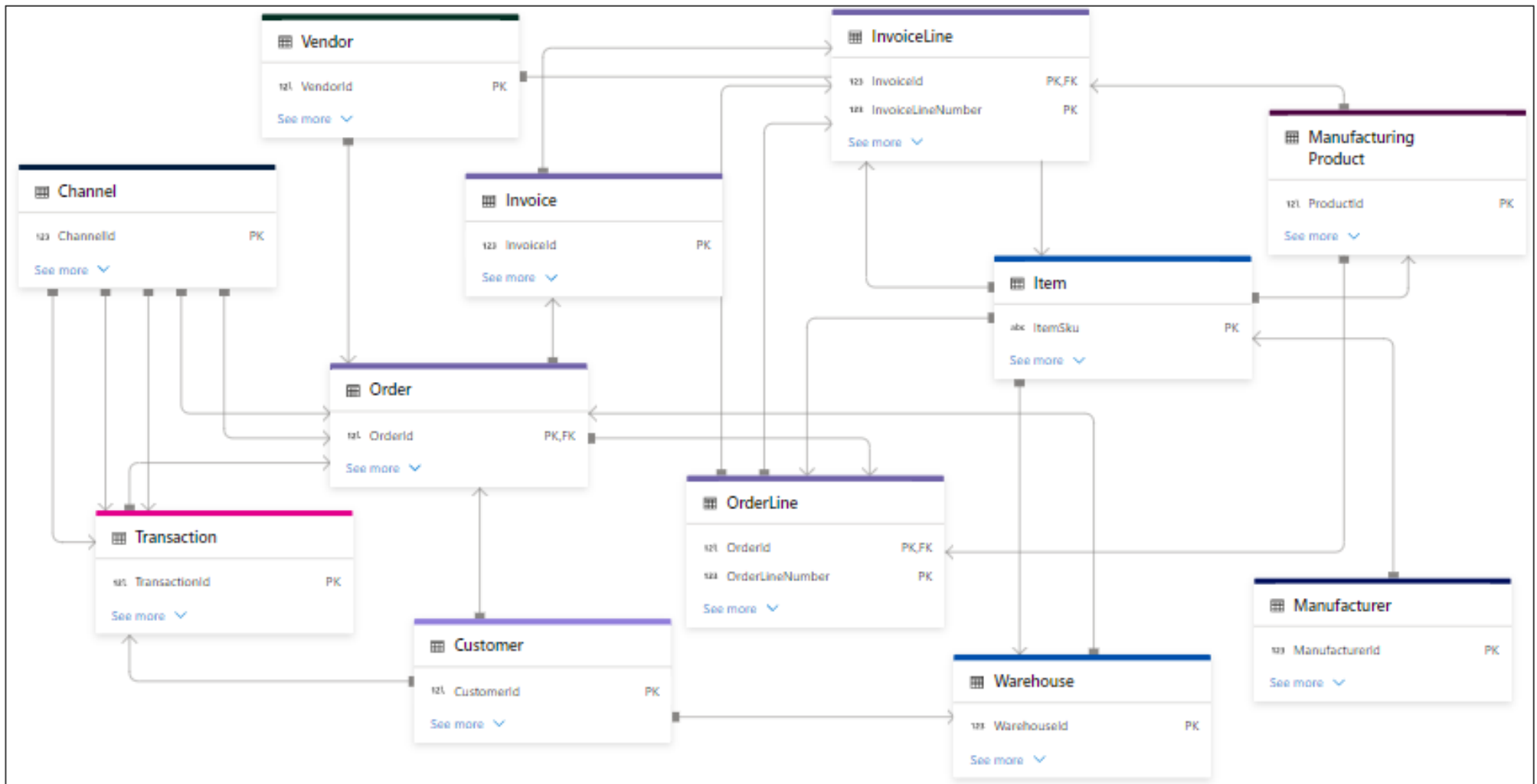
Manual Migration

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

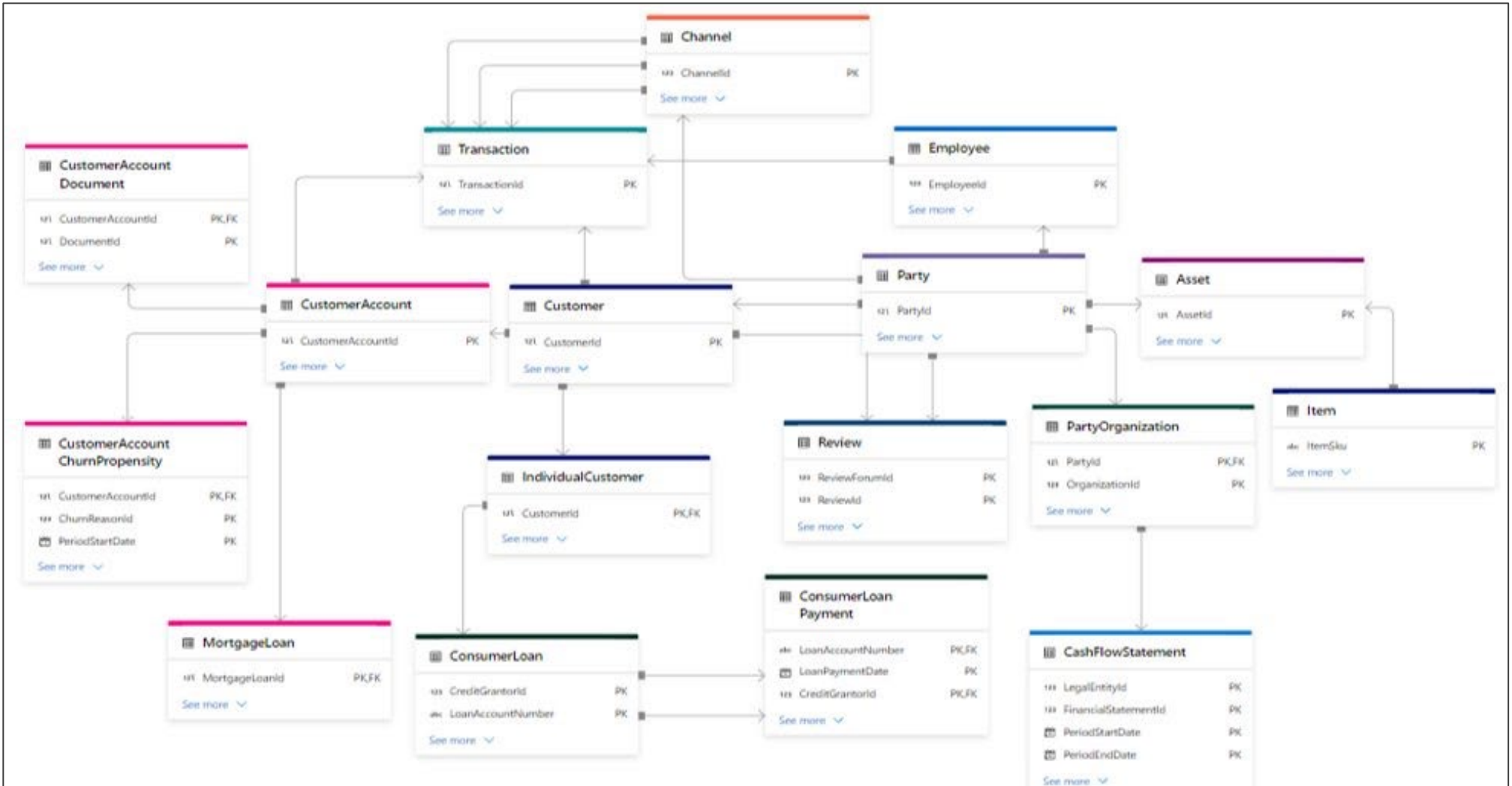
BladeBridge Migration

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

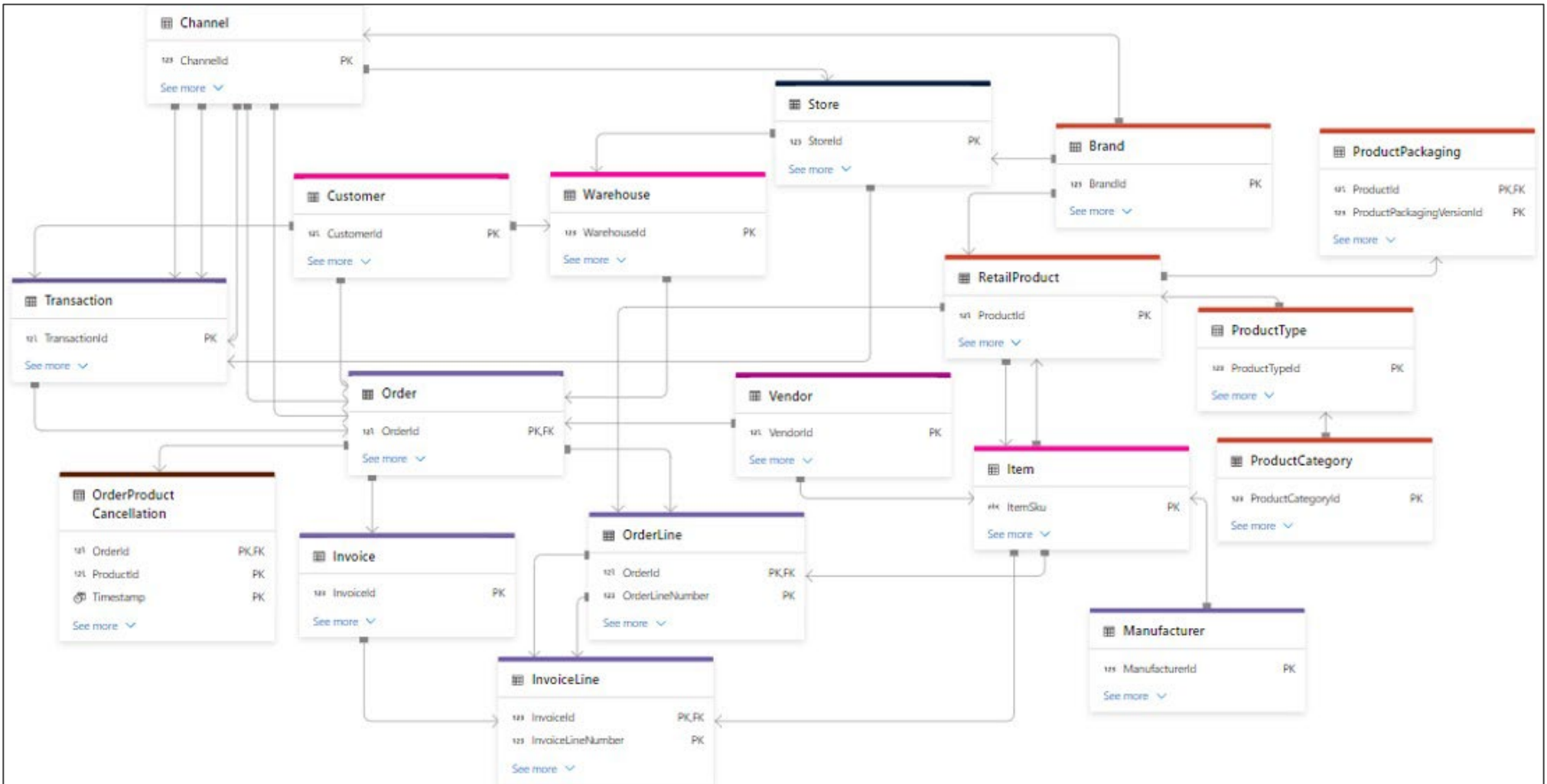
표준 데이터 모델 - 제조업



표준 데이터 모델 - 금융업

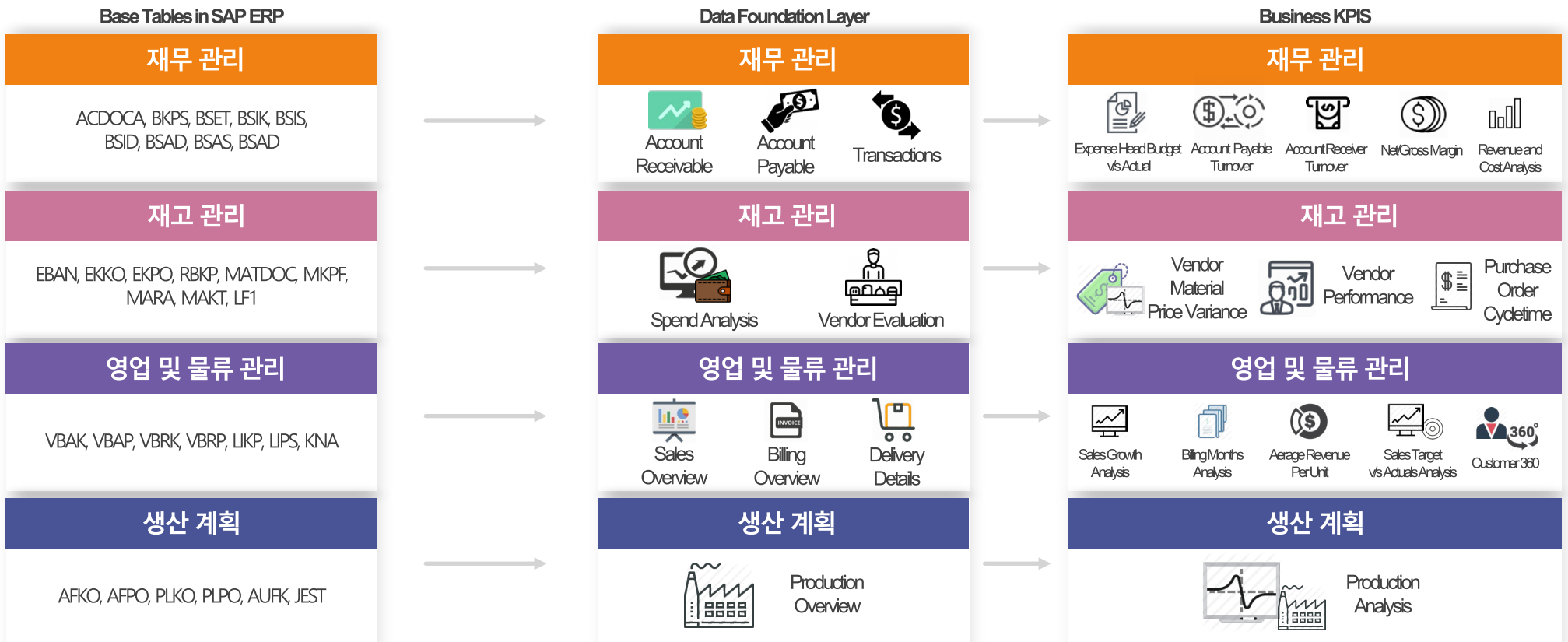


표준 데이터 모델 - 소비재산업



Azure Databricks로 마이그레이션 된 SAP 데이터는 데이터브릭스 레이크하우스 플랫폼을 기반으로 재무 관리(Finance & Controlling), 재고 관리(Inventory Management), 영업 및 물류(Sales & Distribution) 및 생산 계획(Production Planning) 등의 SAP 기반 전업무 영역 표준 KPI 분석 보고서를 제공합니다.

업무 영역별 분석/예측 표준 데이터 모델 및 Report 제공

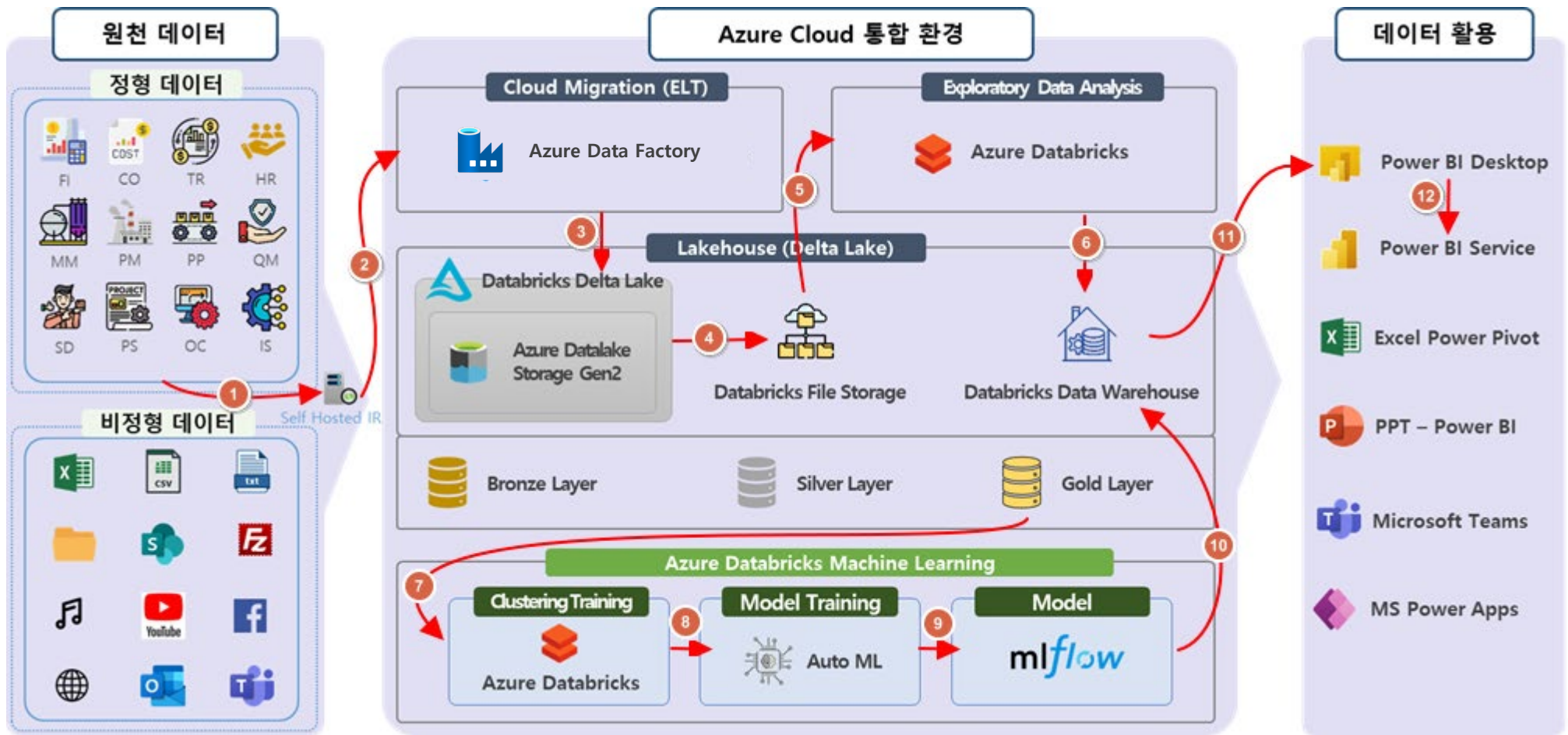




4. Azure Databricks 단계별 프로세스

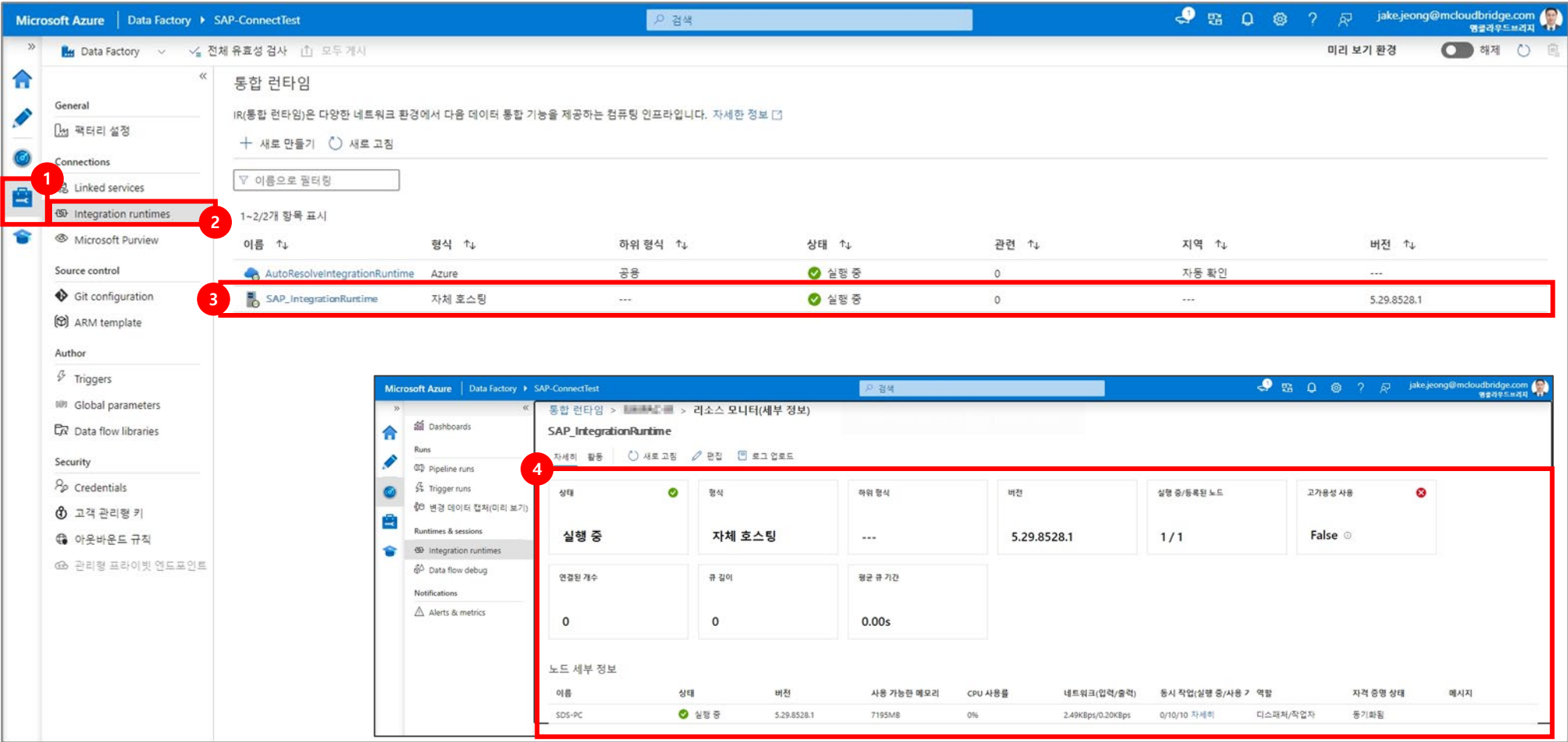
SAP · Databricks 데이터 마이그레이션 제안 아키텍처는 개방성과 통합성을 보유한 Lakehouse를 기반으로 데이터 허브를 구축하여 SAP 데이터를 중심으로 기업에 연관된 모든 정보를 배치 또는 실시간 마이그레이션하여 통합 관리함으로써 데이터 분석(BI) 및 AI/ML 사용 및 확장성을 최대화시킵니다.

SAP · Databricks 데이터 마이그레이션 제안 아키텍처



Self Hosted Integration Runtime은 서로 다른 네트워크 환경에서 데이터 통합 기능을 제공하기 위해 Azure Data Factory와 Synapse 파이프라인에서 사용하는 컴퓨팅 인프라입니다. 온-프레미스 네트워크 또는 Azure 가상 네트워크의 컴퓨팅 리소스에 대해 변환 작업을 실행할 수 있습니다.

안전한 데이터 반출, Self Hosted Integration Runtime



The screenshot shows the Microsoft Azure Data Factory console. The left-hand navigation pane has several items, with 'Integration runtimes' highlighted by a red box and a circled '1'. In the main content area, a table lists integration runtimes. The 'SAP_IntegrationRuntime' row is highlighted with a red box and a circled '2'. A circled '3' points to the 'SAP_IntegrationRuntime' row in the table. Below the table, a detailed view of the 'SAP_IntegrationRuntime' resource is shown, with a red box and a circled '4' highlighting the resource details.

이름	형식	하위 형식	상태	관련	지역	버전
AutoResolveIntegrationRuntime	Azure	공용	실행 중	0	자동 확인	---
SAP_IntegrationRuntime	자체 호스팅	---	실행 중	0	---	5.29.8528.1

이름	상태	버전	사용 가능한 메모리	CPU 사용률	네트워크(입력/출력)	동시 작업(실행 중/사용 ? 역할)	자격 증명 상태	메시지
SDS-PC	실행 중	5.29.8528.1	7195MB	0%	2.49KBps/0.20KBps	0/10/10 자체	디스커져/직업자	동기화됨

Azure Data Factory에서 (공식) SAP Connector를 사용하면 서로 SNC(Secure Network Communications)를 사용하여 SAP 사시스템에 연결을 지원합니다. 특히 Non-SAP 환경에서는 MS SQL, IBM 및 Oracle과 같은 기본 ANYDB(비 SAP) 데이터베이스에 연결하는 일반 커넥터를 사용할 수 있습니다.





SAP – Microsoft (공식 협약) SAP Connector

The screenshot displays the Microsoft Azure Data Factory interface for SAP-ConnectTest. The left sidebar shows the navigation menu with 'Integration runtimes' selected (1). The main area is divided into three panels:

- 연결된 서비스 (Connected services):** Shows a list of services including ADLS_Storage, Azure_OLTP_ADWS, AzureDatabricks, and SapEcc1. A red box (2) highlights the '+ 새로 만들기' (New) button. A red box (3) highlights the '+ 새로 만들기' button in the header.
- 연결된 서비스 편집 (Edit connected service):** Shows the configuration for 'SAP ECC' (5). Fields include: 이름 (SapEcc1), 설명 (SAP Connector를 활용한 연결), 통합 런타임을 통해 연결 (SIMPAC-IR), URL (http://vhcalnplci:8000/sap/bc/ping?sap-client=001), 사용자 이름 (DEVELOPER), and 암호 (Azure Key Vault).
- 새 연결된 서비스 (New connected service):** Shows a grid of service options (4). The 'SAP BW' option is selected.

Azure Data Factory에서 (공식) SAP Connector를 사용하면 서로 SNC(Secure Network Communications)를 사용하여 SAP 시스템에 연결을 지원합니다. 특히 Non-SAP 환경에서는 MS SQL, IBM 및 Oracle과 같은 기본 ANYDB(비 SAP) 데이터베이스에 연결하는 일반 커넥터를 사용할 수 있습니다.

SAP – Microsoft (공식 협약) SAP Connector

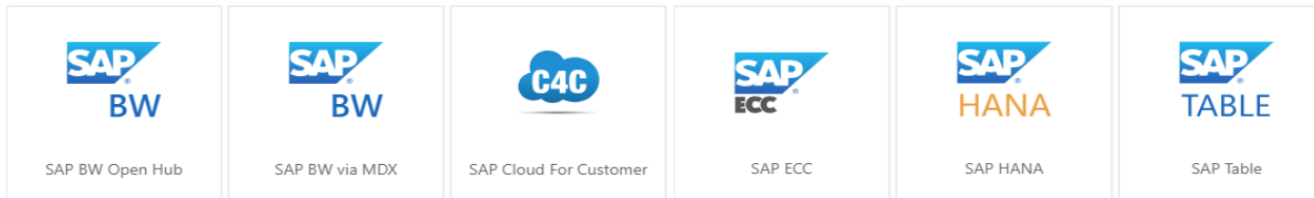
커넥터 이름	지원되는 SAP 시스템 버전	지원되는 SAP 개체	성능 및 확장성
SAP BW 커넥터 	<ul style="list-style-type: none"> SAP BW 버전 7.01 이상, 온프리미어 또는 클라우드 	<ul style="list-style-type: none"> OHD(Open Hub Destination) 로컬 테이블 	<ul style="list-style-type: none"> 병렬 처리 및 데이터 압축 기능을 SAP BW 시스템에서 제공하므로 대규모 데이터 추출 시에도 높은 성능 보임 SAP BW 시스템 내에 이미 최적화된 데이터 모델 등을 활용하여 데이터 추출 및 처리 속도가 빠름
SAP BW via MDX 커넥터 	<ul style="list-style-type: none"> SAP BW 버전 7.x, 클라우드 내 온프리미어(예: Azure) 	<ul style="list-style-type: none"> 정보 큐브 및 쿼리 큐브(BEX 쿼리 포함) 	<ul style="list-style-type: none"> 구성 가능한 데이터 파티셔닝을 기반으로 하는 기본 제공 병렬 로드 옵션을 제공 MDX 쿼리를 사용하여 데이터를 추출하기 때문에 다른 SAP 시스템에서 데이터를 추출하는 것보다 데이터 크기가 크고 복잡한 경우에 더 높은 성능을 발휘
SAP ECC 커넥터 	<ul style="list-style-type: none"> SAP ECC 버전 7.0 이상 SAP .Net Connector 필요 	<ul style="list-style-type: none"> SAP OData 서비스에 의해 노출된 엔티티 	<ul style="list-style-type: none"> 구성 가능한 데이터 파티셔닝을 기반으로 하는 기본 제공 병렬 로드 옵션을 제공 SAP 시스템과 직접 통신하며 데이터를 추출하기 때문에 높은 성능을 보임
SAP HANA 커넥터 	<ul style="list-style-type: none"> SAP ECC 또는 Business Suite 버전 7.01 이상의 기타 애플리케이션, 온프리미어 또는 클라우드 SAP HANA 클라이언트 라이브러리 필요 	<ul style="list-style-type: none"> SAP 투명 테이블, 풀링된 테이블, 클러스터 테이블 및 보기 	<ul style="list-style-type: none"> SAP HANA의 인메모리 데이터 처리 엔진을 활용하여 대규모 데이터에 대해 빠른 처리가 가능 데이터 모델링 및 쿼리 최적화 기능을 제공하여 데이터 검색, 처리를 빠르게 가능
SAP Table 커넥터 	<ul style="list-style-type: none"> 모든 SAP HANA 버전, 온프리미어 또는 클라우드 	<ul style="list-style-type: none"> HANA 정보 모델(분석/계산 보기) 	<ul style="list-style-type: none"> 구성 가능한 데이터 파티셔닝을 기반으로 하는 기본 제공 병렬 로드 옵션 제공 SAP 시스템에서 직접 데이터를 추출하므로 정합성이 보장되며, SAP 시스템의 데이터 캐시를 활용

Azure Data Factory에서 제공하는 (공식) Connector는 SAP 환경에서든 Non-SAP 환경에서든 일정 조건을 만족하면 Connector를 통하여 SAP Application 또는 DB에 연결할 수 있고 MS SQL, IBM 및 Oracle과 같은 기본 ANYDB(비 SAP) 데이터베이스에 연결하는 일반 커넥터를 사용할 수 있습니다.

SAP or Non-SAP 환경 – Microsoft (공식 협약) Connector

Access All Your Data

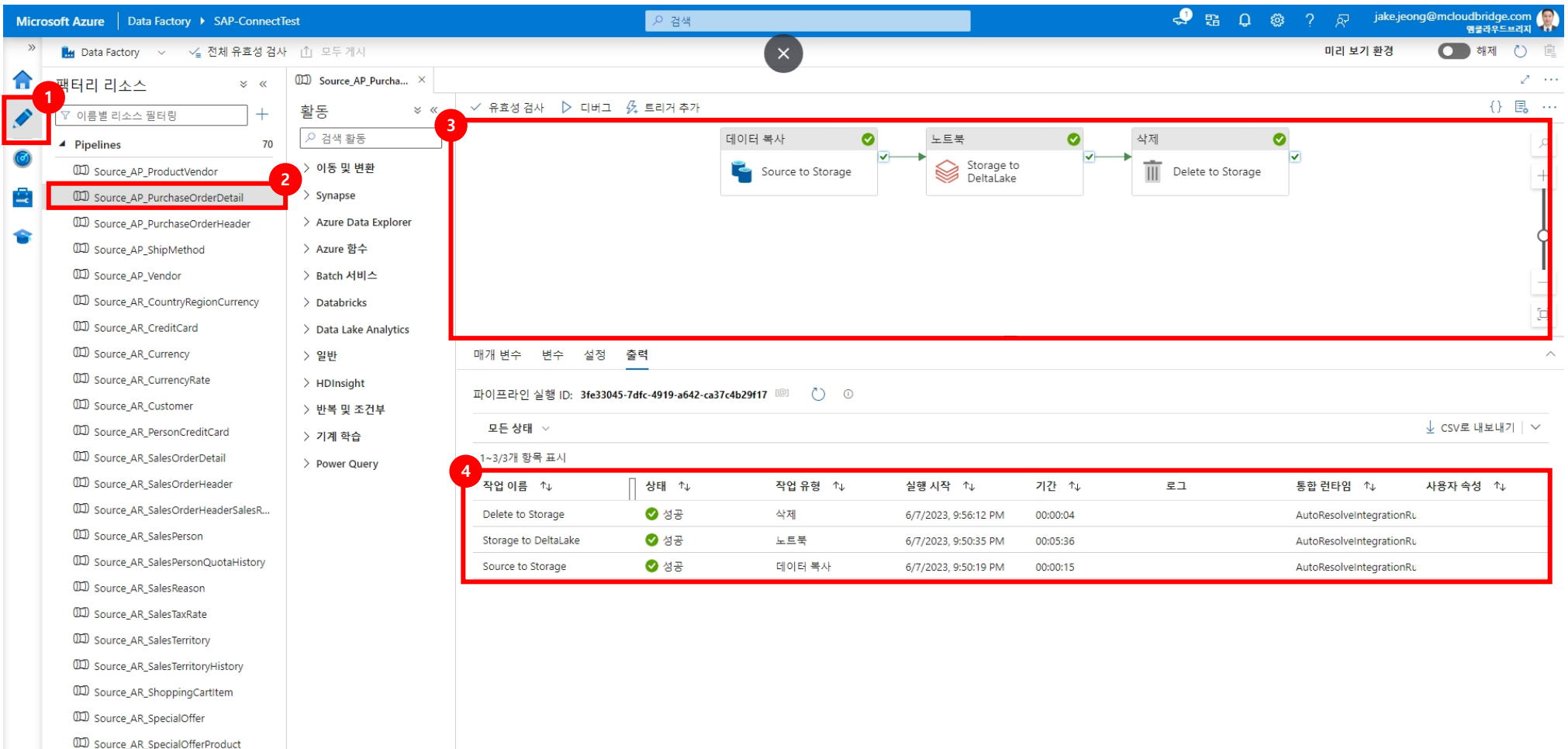
Single tool to enable data ingestion from SAP as well as other various sources, and data transformation via built-in Data Flow, integration with Databricks/HDInsight/etc.



Azure	Database & DW		File Storage	File Formats	NoSQL	Services & Apps		Generic
Blob Storage	Amazon Redshift	Phoenix	Amazon S3	Avro	Cassandra	Amazon MWS	PayPal	HTTP
Cosmos DB – SQL API	DB2	PostgreSQL	File System	Binary	Couchbase	CDS for Apps	QuickBooks	OData
Cosmos DB – MongoDB API	Drill	Presto	FTP	Common Data Model	MongoDB	Concur	Salesforce	ODBC
ADLS Gen1	Google BigQuery	SAP BW Open Hub	Google Cloud Storage	Delimited Text		Dynamics 365	SF Service Cloud	REST
ADLS Gen2	Greenplum	SAP BW MDX	HDFS	Excel		Dynamics AX	SF Marketing Cloud	
Data Explorer	HBase	SAP HANA	SFTP	JSON		Dynamics CRM	SAP C4C	
Database for MariaDB	Hive	SAP Table		ORC		Google AdWords	SAP ECC	
Database for MySQL	Impala	Snowflake		Parquet		HubSpot	ServiceNow	
Database for PostgreSQL	Informix	Spark				Jira	SharePoint List	
File Storage	MariaDB	SQL Server				Magento	Shopify	
SQL Database	Microsoft Access	Sybase				Marketo	Square	
SQL Managed Instance	MySQL	Teradata				Office 365	Web Table	
Synapse Analytics	Netezza	Vertica				Oracle Eloqua	Xero	
Search Index	Oracle					Oracle Responsys	Zoho	
Table Storage						Oracle Service Cloud		

Azure Data Factory에서 ELT 파이프라인 개발은 도구상자 중에서도 “데이터 복사” 도구를 사용하여 SAP 데이터를 Azure Storage로 ELT 합니다. 이때, Parquet 형태의 파일로 가져옵니다. 도구상자 중 “Databricks Notebook”을 사용하여 ADLS에 저장된 데이터를 Delta Lake로 다시 저장합니다.

Azure Data Factory – SAP Cloud ELT



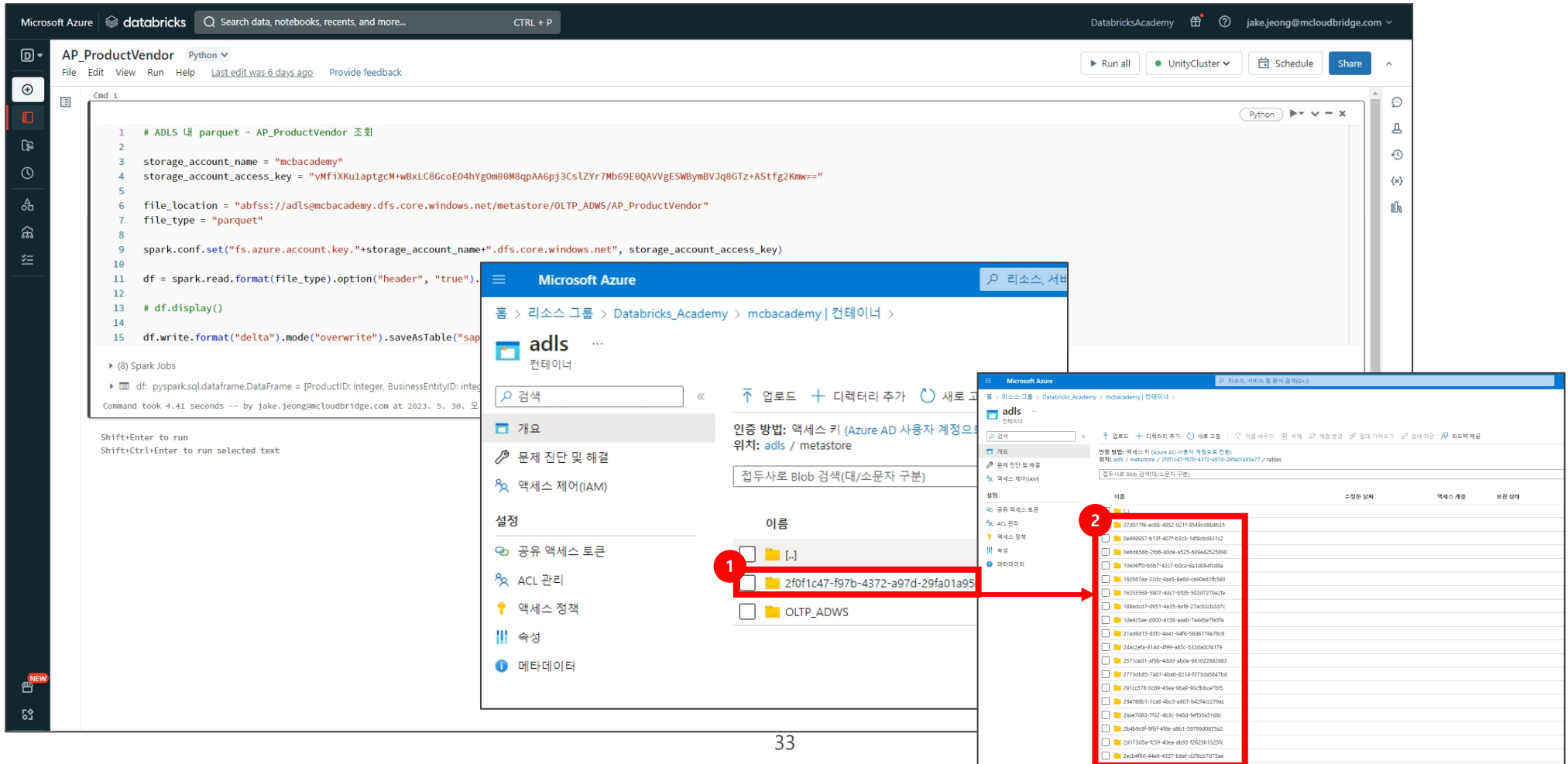
The screenshot displays the Microsoft Azure Data Factory interface for a pipeline named 'Source_AP_Purcha...'. The pipeline consists of three sequential activities: '데이터 복사' (Source to Storage), '노트북' (Storage to DeltaLake), and '삭제' (Delete to Storage). All activities show a green checkmark indicating successful completion.

Below the pipeline view, the execution history table is shown, detailing the status and execution time for each activity.

작업 이름	상태	작업 유형	실행 시작	기간	로그	통합 런타임	사용자 속성
Delete to Storage	성공	삭제	6/7/2023, 9:56:12 PM	00:00:04		AutoResolveIntegrationRu	
Storage to DeltaLake	성공	노트북	6/7/2023, 9:50:35 PM	00:05:36		AutoResolveIntegrationRu	
Source to Storage	성공	데이터 복사	6/7/2023, 9:50:19 PM	00:00:15		AutoResolveIntegrationRu	

Azure Databricks의 Notebook을 활용하여 ADLS에 저장된 데이터를 Delta Lake로 다시 저장합니다. Delta Lake로 저장이 되어야 Azure Databricks 내에서 Storage 파일을 Lakehouse로 호출하여 사용할 수 있는 환경 구성이 완료됩니다

Azure Databricks Notebook – Delta Lake 저장



The screenshot displays the Databricks Notebook interface with the following code:

```

1 # ADLS 내 parquet - AP_ProductVendor 조회
2
3 storage_account_name = "mcbacademy"
4 storage_account_access_key = "vMfiXKuIaptgcM+rWbXLC8GcoE04hYg0m0M8opAA6pj3CsLZYr7M69E0QAVVgESWbYmBVJq0GTz+ASTfg2Kmw=="
5
6 file_location = "abfss://adls@mcbacademy.dfs.core.windows.net/metastore/OLTP_ADWS/AP_ProductVendor"
7 file_type = "parquet"
8
9 spark.conf.set("fs.azure.account.key."+storage_account_name+".dfs.core.windows.net", storage_account_access_key)
10
11 df = spark.read.format(file_type).option("header", "true").load(file_location)
12
13 # df.display()
14
15 df.write.format("delta").mode("overwrite").saveAsTable("spark_catalog.adls.product_vendor")
  
```

The notebook output shows:

```

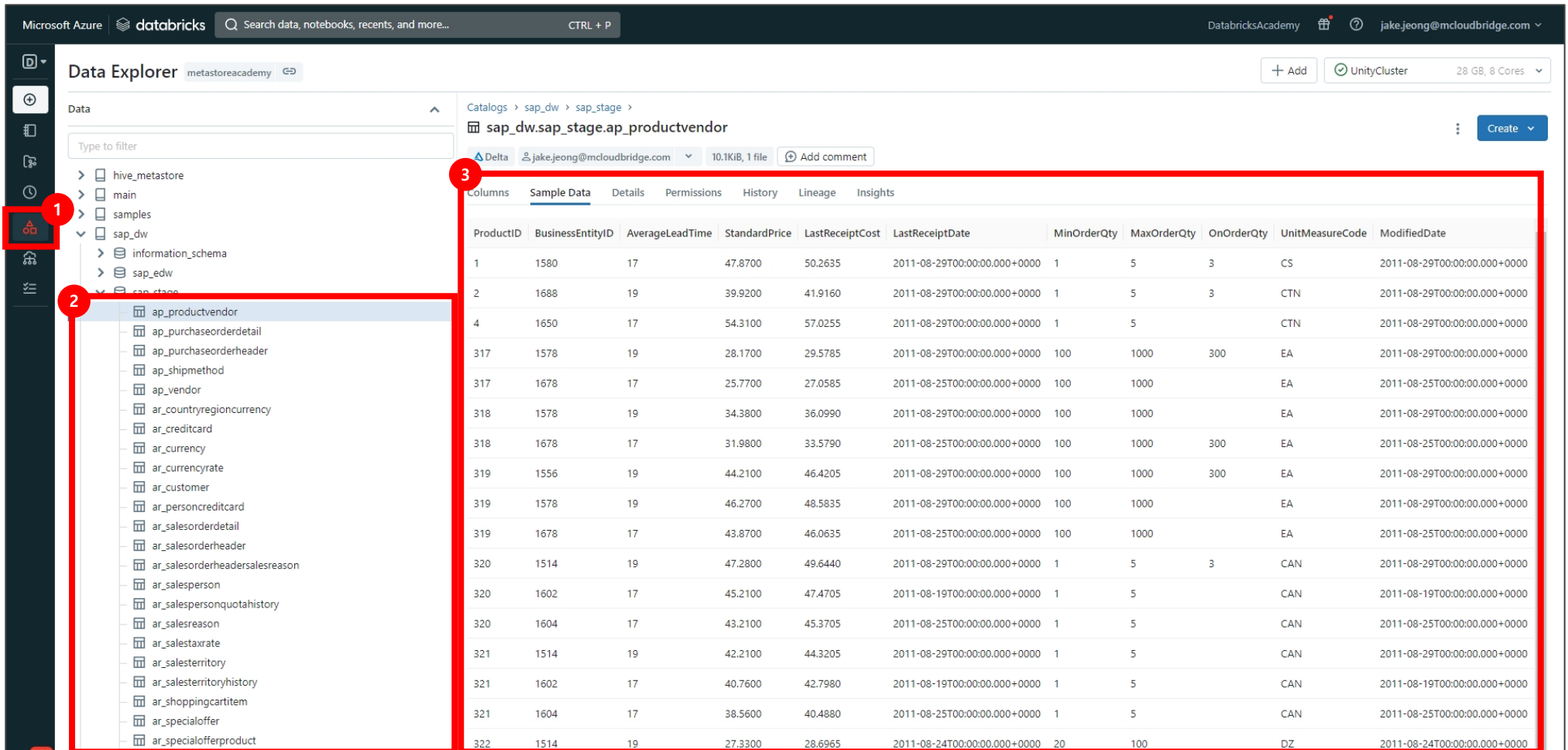
(8) Spark Jobs
(8) pyspark.sql.dataframe.DataFrame = [ProductID: integer, BusinessEntityID: integer]
Command took 4.41 seconds -- by jake.jeong@mcloudbridge.com at 2023. 5. 30. 오후 1:50:00
  
```

Two overlaid screenshots from the Microsoft Azure portal illustrate the configuration steps:

- Left Overlay:** Shows the 'adls' container details in the 'Databricks_Academy' resource group. The '인증 방법' (Authentication method) is set to '액세스 키 (Azure AD 사용자 계정으로 인증)' (Access key (authentication with Azure AD user account)). The '위치' (Location) is 'adls / metastore'. A red circle '1' highlights the '이름' (Name) field, which contains '2f01c47-f97b-4372-a97d-29fa01a95...'.
- Right Overlay:** Shows the 'adls' container's '설정' (Settings) page. A red circle '2' highlights the '공유 액세스 토큰' (Shared access token) field, which contains a long alphanumeric string.

Azure Databricks Delta Lake는 Databricks Lakehouse 플랫폼에 데이터 및 테이블을 저장하기 위한 기반을 제공하는 최적화된 스토리지 계층입니다. 특히 File Storage 형태의 데이터를 DB Table 형태로 조회할 수 있으며 이는 ETL을 진행한 데이터이기 때문에 추가적인 데이터 전처리와 모델링이 필요합니다.

Azure Databricks – Delta Lake 저장



The screenshot shows the Azure Databricks Data Explorer interface. The left sidebar (Data Explorer) shows a tree view of catalogs and schemas. A red box labeled '1' highlights the 'Data Explorer' icon. A red box labeled '2' highlights the path: 'hive_metastore' > 'main' > 'samples' > 'sap_dw' > 'sap_stage' > 'ap_productvendor'. A red box labeled '3' highlights the table details and sample data view.

ProductID	BusinessEntityID	AverageLeadTime	StandardPrice	LastReceiptCost	LastReceiptDate	MinOrderQty	MaxOrderQty	OnOrderQty	UnitMeasureCode	ModifiedDate
1	1580	17	47.8700	50.2635	2011-08-29T00:00:00.000+0000	1	5	3	CS	2011-08-29T00:00:00.000+0000
2	1688	19	39.9200	41.9160	2011-08-29T00:00:00.000+0000	1	5	3	CTN	2011-08-29T00:00:00.000+0000
4	1650	17	54.3100	57.0255	2011-08-29T00:00:00.000+0000	1	5		CTN	2011-08-29T00:00:00.000+0000
317	1578	19	28.1700	29.5785	2011-08-29T00:00:00.000+0000	100	1000	300	EA	2011-08-29T00:00:00.000+0000
317	1678	17	25.7700	27.0585	2011-08-25T00:00:00.000+0000	100	1000		EA	2011-08-25T00:00:00.000+0000
318	1578	19	34.3800	36.0990	2011-08-29T00:00:00.000+0000	100	1000		EA	2011-08-29T00:00:00.000+0000
318	1678	17	31.9800	33.5790	2011-08-25T00:00:00.000+0000	100	1000	300	EA	2011-08-25T00:00:00.000+0000
319	1556	19	44.2100	46.4205	2011-08-29T00:00:00.000+0000	100	1000	300	EA	2011-08-29T00:00:00.000+0000
319	1578	19	46.2700	48.5835	2011-08-29T00:00:00.000+0000	100	1000		EA	2011-08-29T00:00:00.000+0000
319	1678	17	43.8700	46.0635	2011-08-25T00:00:00.000+0000	100	1000		EA	2011-08-25T00:00:00.000+0000
320	1514	19	47.2800	49.6440	2011-08-29T00:00:00.000+0000	1	5	3	CAN	2011-08-29T00:00:00.000+0000
320	1602	17	45.2100	47.4705	2011-08-19T00:00:00.000+0000	1	5		CAN	2011-08-19T00:00:00.000+0000
320	1604	17	43.2100	45.3705	2011-08-25T00:00:00.000+0000	1	5		CAN	2011-08-25T00:00:00.000+0000
321	1514	19	42.2100	44.3205	2011-08-29T00:00:00.000+0000	1	5		CAN	2011-08-29T00:00:00.000+0000
321	1602	17	40.7600	42.7980	2011-08-19T00:00:00.000+0000	1	5		CAN	2011-08-19T00:00:00.000+0000
321	1604	17	38.5600	40.4880	2011-08-25T00:00:00.000+0000	1	5		CAN	2011-08-25T00:00:00.000+0000
322	1514	19	27.3300	28.6965	2011-08-24T00:00:00.000+0000	20	100		DZ	2011-08-24T00:00:00.000+0000

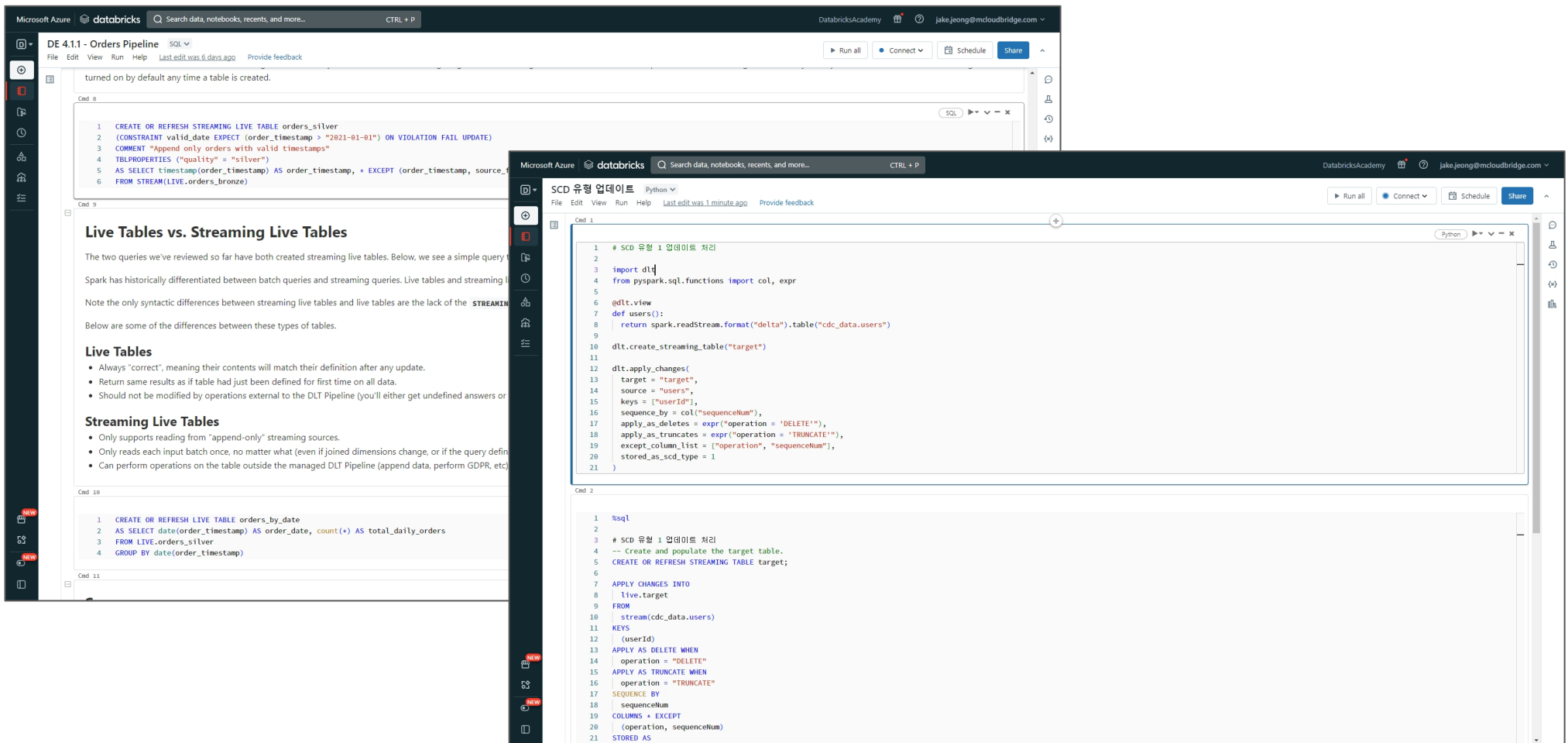
Azure Databricks Notebook에서는 Notebook 셀에서 Python 및 SQL 코드의 서식을 빠르고 쉽게 지정할 수 있는 도구를 제공합니다. Notebook 버전의 기록을 유지 관리하므로 Notebook의 이전 스냅샷을 보고 복원할 수 있으며 개발 언어를 선택하여 언어 혼합 개발이 가능합니다.

Azure Databricks – Notebook을 활용한 다양한 언어로 Data Cleansing

The screenshot displays the Azure Databricks Notebook interface. On the left, a code cell (Cmd 3) contains Python code for reading IoT Hub data into a stream and writing it to Delta. The main workspace shows a notebook titled 'DE 2.4 - Cleaning Data' with several code cells. Cmd 15 shows a Python script that deduplicates users based on their first touch timestamp. Cmd 16 shows a confirmation message: 'Let's confirm that we have the expected count of remaining records'. Cmd 17 shows a SQL query to count distinct users. Cmd 18 shows another Python script for dropping duplicates. Cmd 19 shows a 'Validate Datasets' section. On the right, a command (Cmd 4) displays the output of the IoT stream as a DataFrame. Below this, a line chart (Cmd 5) visualizes the data, plotting humidity, temperature, windspeed, and rpm against timestamp. The chart shows a clear upward trend in temperature and humidity over time, while windspeed and rpm remain relatively stable.

Azure Databricks Delta Live Tables은 관리하기 쉬운 쿼리로 크거나 복잡한 쿼리가 있을 경우는 뷰로, 여러 다운스트림 쿼리가 테이블을 사용하고 View가 쿼리될 때마다 다시 계산됨을 방지하고자 할 경우는 **구조화된 뷰**로, 지속적으로 또는 점진적으로 증가하는 데이터는 **스트리밍 테이블**로 CDC와 SCD를 지원합니다.

Azure Databricks – Delta Live Table



The screenshot displays the Azure Databricks workspace interface. The left pane shows a document titled "Live Tables vs. Streaming Live Tables" with the following content:

Live Tables vs. Streaming Live Tables

The two queries we've reviewed so far have both created streaming live tables. Below, we see a simple query that...

Spark has historically differentiated between batch queries and streaming queries. Live tables and streaming live tables...

Note the only syntactic differences between streaming live tables and live tables are the lack of the **STREAMING** keyword...

Below are some of the differences between these types of tables.

Live Tables

- Always "correct", meaning their contents will match their definition after any update.
- Return same results as if table had just been defined for first time on all data.
- Should not be modified by operations external to the DLT Pipeline (you'll either get undefined answers or errors).

Streaming Live Tables

- Only supports reading from "append-only" streaming sources.
- Only reads each input batch once, no matter what (even if joined dimensions change, or if the query definition changes).
- Can perform operations on the table outside the managed DLT Pipeline (append data, perform GDPR, etc).

The right pane shows a Python notebook titled "SCD 유형 1 업데이트 처리" (SCD Type 1 Update Processing) with the following code:

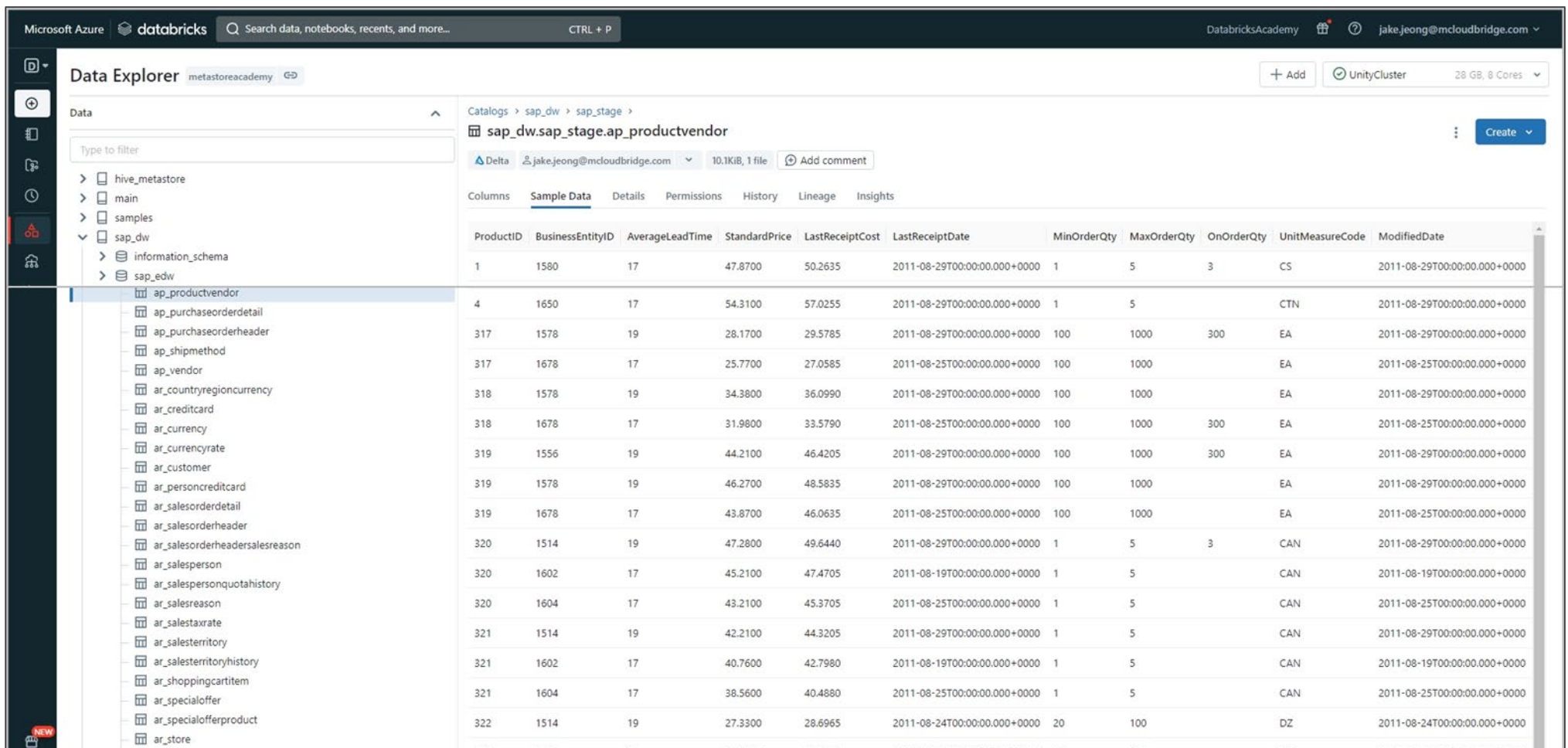
```
1 # SCD 유형 1 업데이트 처리
2
3 import dtl
4 from pyspark.sql.functions import col, expr
5
6 @dtl.view
7 def users():
8     return spark.readStream.format("delta").table("cdc_data.users")
9
10 dtl.create_streaming_table("target")
11
12 dtl.apply_changes(
13     target = "target",
14     source = "users",
15     keys = ["userId"],
16     sequence_by = col("sequenceNum"),
17     apply_as_delete = expr("operation = 'DELETE'"),
18     apply_as_truncate = expr("operation = 'TRUNCATE'"),
19     except_column_list = ["operation", "sequenceNum"],
20     stored_as_scd_type = 1
21 )
```

The bottom pane shows a SQL notebook with the following code:

```
1 %sql
2
3 # SCD 유형 1 업데이트 처리
4 -- Create and populate the target table.
5 CREATE OR REFRESH STREAMING TABLE target;
6
7 APPLY CHANGES INTO
8   live.target
9 FROM
10   stream(cdc_data.users)
11 KEYS
12   (userId)
13 APPLY AS DELETE WHEN
14   operation = "DELETE"
15 APPLY AS TRUNCATE WHEN
16   operation = "TRUNCATE"
17 SEQUENCE BY
18   sequenceNum
19 COLUMNS * EXCEPT
20   (operation, sequenceNum)
21 STORED AS
```

Azure Databricks Data Warehousing은 데이터베이스 스키마, 테이블 및 뷰와 같은 친숙한 관계로 클라우드 개체 스토리지의 Delta Lake와 함께 저장된 데이터를 구성합니다. Databricks Lakehouse는 엔터프라이즈 데이터 웨어하우스의 ACID 트랜잭션 및 데이터 거버넌스와 데이터 레이크의 유연성 및 비용 효율성을 결합합니다.

Azure Databricks – Data Warehousing

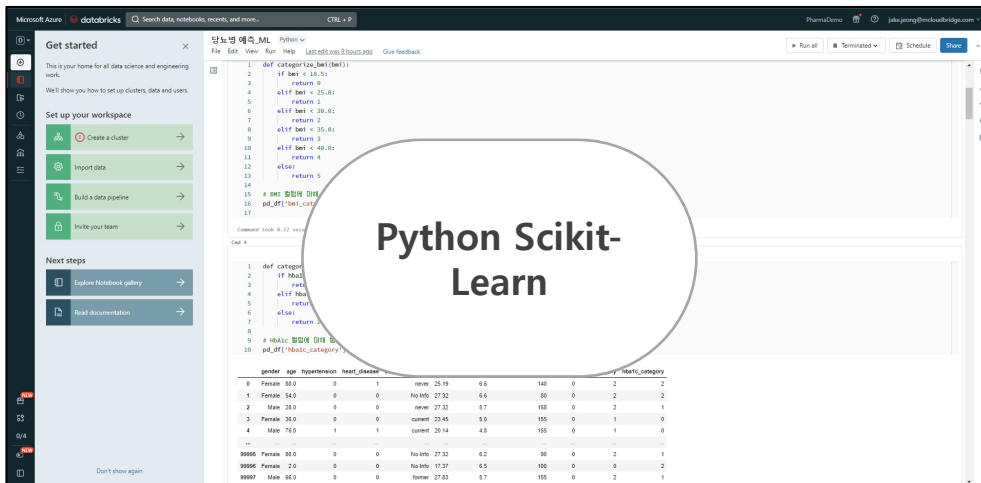


The screenshot displays the Azure Databricks Data Explorer interface. The left sidebar shows a tree view of the data catalog, with 'sap_dw' expanded to show 'sap_edw'. The main pane shows the table 'sap_dw.sap_stage.ap_productvendor' with the following columns: ProductID, BusinessEntityID, AverageLeadTime, StandardPrice, LastReceiptCost, LastReceiptDate, MinOrderQty, MaxOrderQty, OnOrderQty, UnitMeasureCode, and ModifiedDate. The table contains 12 rows of data.

ProductID	BusinessEntityID	AverageLeadTime	StandardPrice	LastReceiptCost	LastReceiptDate	MinOrderQty	MaxOrderQty	OnOrderQty	UnitMeasureCode	ModifiedDate
1	1580	17	47.8700	50.2635	2011-08-29T00:00:00.000+0000	1	5	3	CS	2011-08-29T00:00:00.000+0000
4	1650	17	54.3100	57.0255	2011-08-29T00:00:00.000+0000	1	5		CTN	2011-08-29T00:00:00.000+0000
317	1578	19	28.1700	29.5785	2011-08-29T00:00:00.000+0000	100	1000	300	EA	2011-08-29T00:00:00.000+0000
317	1678	17	25.7700	27.0585	2011-08-25T00:00:00.000+0000	100	1000		EA	2011-08-25T00:00:00.000+0000
318	1578	19	34.3800	36.0990	2011-08-29T00:00:00.000+0000	100	1000		EA	2011-08-29T00:00:00.000+0000
318	1678	17	31.9800	33.5790	2011-08-25T00:00:00.000+0000	100	1000	300	EA	2011-08-25T00:00:00.000+0000
319	1556	19	44.2100	46.4205	2011-08-29T00:00:00.000+0000	100	1000	300	EA	2011-08-29T00:00:00.000+0000
319	1578	19	46.2700	48.5835	2011-08-29T00:00:00.000+0000	100	1000		EA	2011-08-29T00:00:00.000+0000
319	1678	17	43.8700	46.0635	2011-08-25T00:00:00.000+0000	100	1000		EA	2011-08-25T00:00:00.000+0000
320	1514	19	47.2800	49.6440	2011-08-29T00:00:00.000+0000	1	5	3	CAN	2011-08-29T00:00:00.000+0000
320	1602	17	45.2100	47.4705	2011-08-19T00:00:00.000+0000	1	5		CAN	2011-08-19T00:00:00.000+0000
320	1604	17	43.2100	45.3705	2011-08-25T00:00:00.000+0000	1	5		CAN	2011-08-25T00:00:00.000+0000
321	1514	19	42.2100	44.3205	2011-08-29T00:00:00.000+0000	1	5		CAN	2011-08-29T00:00:00.000+0000
321	1602	17	40.7600	42.7980	2011-08-19T00:00:00.000+0000	1	5		CAN	2011-08-19T00:00:00.000+0000
321	1604	17	38.5600	40.4880	2011-08-25T00:00:00.000+0000	1	5		CAN	2011-08-25T00:00:00.000+0000
322	1514	19	27.3300	28.6965	2011-08-24T00:00:00.000+0000	20	100		DZ	2011-08-24T00:00:00.000+0000

Azure Databricks ML Library를 통해 모델을 학습을 지원하고 있습니다. 주로 scikit-learn은 Python 라이브러리 중 단일 노드 기계 학습에 널리 사용되며 Apache Spark MLlib는 분류, 회귀, 클러스터링, 공동 작업 필터링, 차원 감소, 일반적인 학습 알고리즘 및 유틸리티로 구성된 Apache Spark 라이브러리입니다.

Azure Databricks – Machine Learning Library

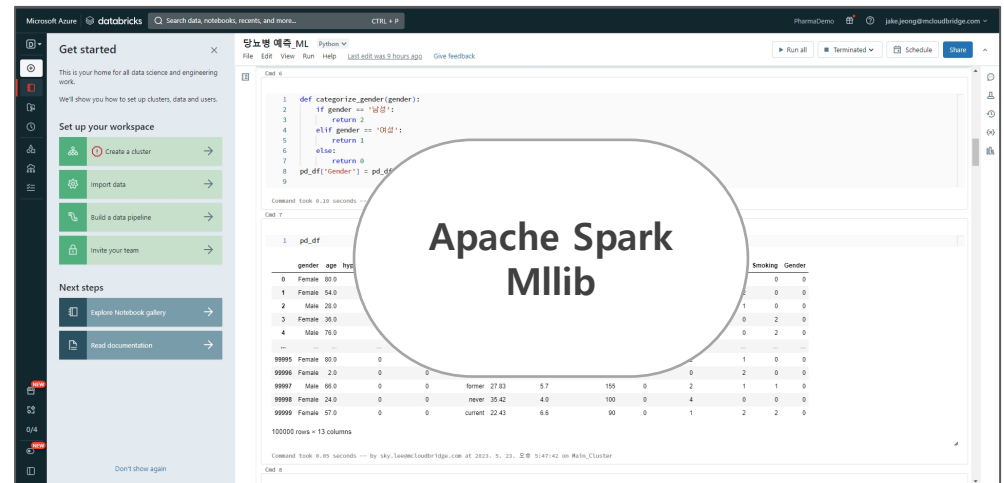


Python Scikit-Learn

```

1 def categorize_bmi(bmi):
2     if bmi < 18.5:
3         return 0
4     elif bmi < 25.0:
5         return 1
6     elif bmi < 30.0:
7         return 2
8     elif bmi < 35.0:
9         return 3
10    elif bmi < 40.0:
11        return 4
12    else:
13        return 5
14
15 # BMI 범주에 따라 BMI
16 pd_df['bmi_category'] = pd_df['bmi'].apply(categorize_bmi)
  
```

gender	age	height_cm	weight_kg	heart_rate_bpm	cholesterol_mg_dl	glucose_fasting_mg_dl	hypertension	diabetes	smoking	bmi_category
Female	30.0	160	55	70	120	90	0	0	never	2
Female	54.0	160	66	80	120	90	0	0	never	2
Male	28.0	170	57	75	100	80	0	0	never	1
Female	38.0	165	60	75	100	80	0	0	never	1
Male	70.0	175	45	65	100	80	0	0	former	0

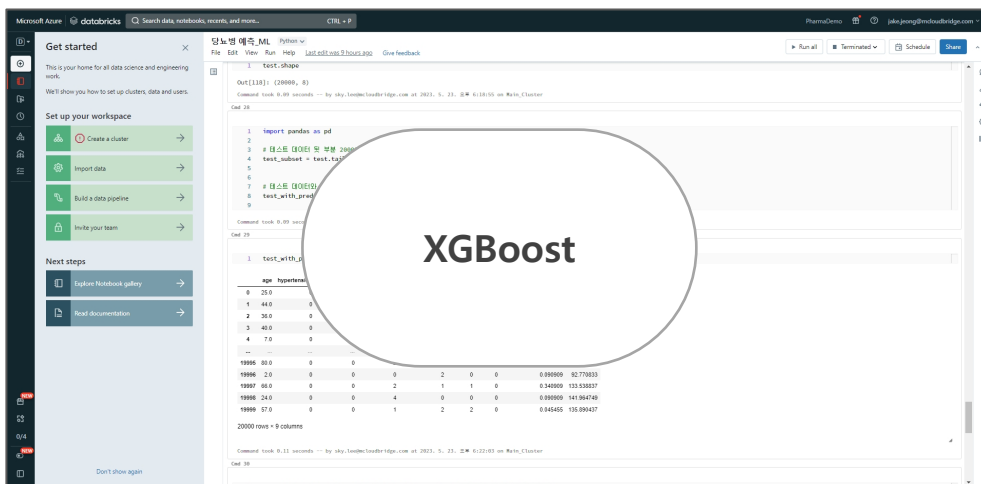


Apache Spark MLlib

```

1 def categorize_gender(gender):
2     if gender == '남성':
3         return 2
4     elif gender == '여성':
5         return 1
6     else:
7         return 0
8 pd_df['gender'] = pd_df['gender'].apply(categorize_gender)
  
```

gender	age	height_cm	weight_kg	heart_rate_bpm	cholesterol_mg_dl	glucose_fasting_mg_dl	hypertension	diabetes	smoking
Female	30.0	160	55	70	120	90	0	0	0
Female	54.0	160	66	80	120	90	0	0	0
Male	28.0	170	57	75	100	80	0	0	1
Female	38.0	165	60	75	100	80	0	0	2
Male	70.0	175	45	65	100	80	0	0	2

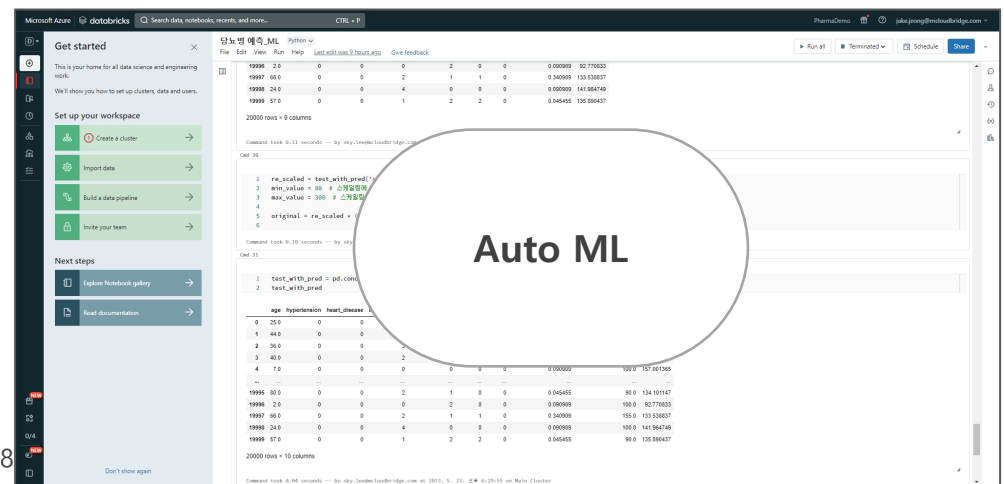


XGBoost

```

1 test_shape
2 # Import pandas as pd
3 # Import XGBoost as xgb
4 test_subset = test_train
5
6 # Import XGBoost
7 xgb = xgb.XGBClassifier()
8 test_with_xgb = xgb.fit(test_subset, test_subset['diabetes']).predict(test_subset)
  
```

age	hypertension	diabetes	smoking	area_under_curve	auc
30.0	0	0	0	0.899000	0.770353
28.0	0	0	0	0.940000	0.833037
38.0	0	0	0	0.930000	0.819478
70.0	0	0	0	0.845455	0.694247



Auto ML

```

1 test_with_xgb = xgb.fit(test_subset, test_subset['diabetes']).predict(test_subset)
2 test_with_lr = lr.fit(test_subset, test_subset['diabetes']).predict(test_subset)
  
```

age	hypertension	diabetes	smoking	area_under_curve	auc
30.0	0	0	0	0.899000	0.770353
28.0	0	0	0	0.940000	0.833037
38.0	0	0	0	0.930000	0.819478
70.0	0	0	0	0.845455	0.694247

Azure Databricks AutoML은 데이터 세트에 기계 학습을 자동 적용하는 기능으로 여러 모델을 자동으로 만들고, 조정하고, 평가하는 일련의 평가판을 수행하고 기록합니다. 모델 평가 후 이에 대한 결과를 표시하고 Python Notebook에 각 평가판 실행에 대한 소스 코드를 제공하고 있습니다.

Azure Databricks – Machine Learning AutoML

Create AutoML experiment with feature store tables

```
from pyspark.sql import *
from pyspark.sql.functions import current_timestamp
from pyspark.sql.types import IntegerType
import math
from datetime import timedelta
import mlflow.pyfunc

def rounded_unix_timestamp(dt, num_minutes=15):
    """
    Ceilings datetime dt to interval num_minutes, then returns the unix timestamp.
    """
    nsecs = dt.minute * 60 + dt.second + dt.microsecond * 1e-6
    delta = math.ceil(nsecs / (60 * num_minutes)) * (60 * num_minutes) - nsecs
    return int((dt + timedelta(seconds=delta)).timestamp())

rounded_unix_timestamp_udf = udf(rounded_unix_timestamp, IntegerType())

def rounded_taxi_data(taxi_data_df):
    # Round the taxi data timestamp to 15 and 30 minute intervals so we can join with the pickup and dropoff features
    # respectively.
    taxi_data_df = (
        taxi_data_df.withColumn(
            "rounded_pickup_datetime",
            rounded_unix_timestamp_udf(taxi_data_df["tpep_pickup_datetime"], lit(15)),
        )
        .withColumn(
            "rounded_dropoff_datetime",
            rounded_unix_timestamp_udf(taxi_data_df["tpep_dropoff_datetime"], lit(30)),
        )
        .drop("tpep_pickup_datetime")
        .drop("tpep_dropoff_datetime")
    )
    taxi_data_df.createOrReplaceTempView("taxi_data")
    return taxi_data_df

taxi_data = rounded_taxi_data(raw_data)
display(taxi_data)
```

Table

	trip_distance	fare_amount	pickup_zip	dropoff_zip	rounded_pickup_datetime	rounded_dropoff_datetime
1	4.94	19	10282	10171	1455469200	1455471000
2	0.28	3.5	10110	10110	1454611500	1454612400

Azure Databricks MLFlow는 엔드투엔드 기계 학습 수명 주기를 관리하기 위한 오픈 소스 플랫폼으로 ML 모델 학습 실행 및 ML 프로젝트를 추적하고 보호하기 위한 통합 환경을 제공합니다. 각 실험을 통해 실행을 시각화, 검색 및 비교할 수 있으며 다른 도구에서 분석을 위해 실행 아티팩트 또는 메타데이터를 다운로드할 수 있습니다.

Azure Databricks – MLFlow

Forecast power output with the production model

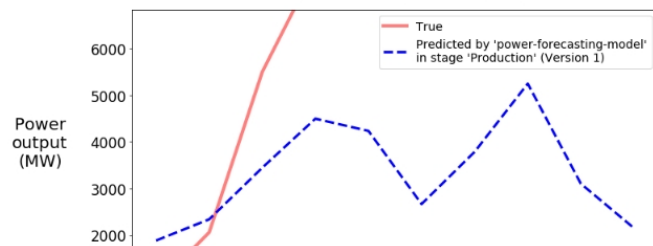
In this section, the production model is used to evaluate weather forecast data for the wind farm. The `forecast_power()` application loads the latest version of the forecasting model from the specified stage and uses it to forecast power production over the next five days.

```
def plot(model_name, model_stage, model_version, power_predictions, past_power_output):
    import pandas as pd
    import matplotlib.dates as mdates
    from matplotlib import pyplot as plt
    index = power_predictions.index
    fig = plt.figure(figsize=(11, 7))
    ax = fig.add_subplot(111)
    ax.set_xlabel("Date", size=20, labelpad=20)
    ax.set_ylabel("Power\output\n(MW)", size=20, labelpad=20, rotation=0)
    ax.tick_params(axis='both', which='major', labelsize=17)
    ax.xaxis.set_major_formatter(mdates.DateFormatter('%m/%d'))
    ax.plot(index[:len(past_power_output)], past_power_output, label="True", color="red", alpha=0.5, linewidth=4)
    ax.plot(index, power_predictions.squeeze(), "--", label="Predicted by 'Ns'\nin stage 'Ns' (Version %d)" % (model_name, model_stage, model_version), color="blue", linewidth=3)
    ax.set_ylim(ymin=0, ymax=max(3500, int(max(power_predictions.values) * 1.3)))
    ax.legend(fontsize=14)
    plt.title("Wind farm power output and projections", size=24, pad=20)
    plt.tight_layout()
    display(plt.show())

def forecast_power(model_name, model_stage):
    from mlflow.tracking.client import MLflowClient
    client = MLflowClient()
    model_version = client.get_latest_versions(model_name, stages=[model_stage])[0].version
    model_uri = "models://{model_name}/{model_stage}".format(model_name=model_name, model_stage=model_stage)
    model = mlflow.pyfunc.load_model(model_uri)
    weather_data, past_power_output = get_weather_and_forecast()
    power_predictions = pd.DataFrame(model.predict(weather_data))
    power_predictions.index = pd.to_datetime(weather_data.index)
    print(power_predictions)
    plot(model_name, model_stage, int(model_version), power_predictions, past_power_output)
```

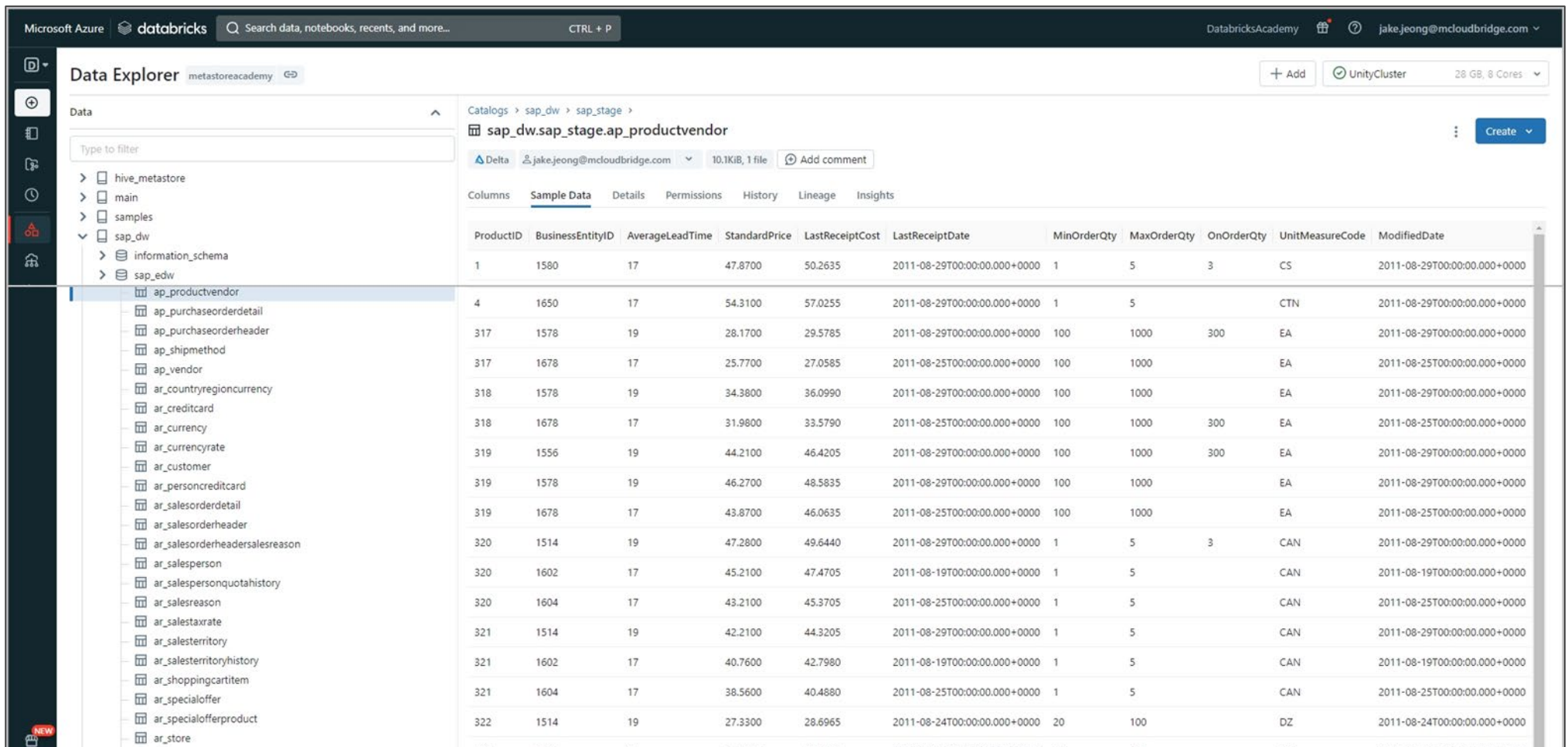
forecast_power(model_name, "Production")

Wind farm power output and projections



Azure Databricks Data Warehousing은 ML 모델을 학습시킨 결과와 예측값에 대한 결과를 Row 데이터로 Data Warehouse 내에 데이터로
 서 함께 적재하여 관리할 수 있습니다. Data Warehouse에서 관리되는 데이터는 전사 표준화 데이터를 지향하며 Self 분석 환경을 제공합니다.

Azure Databricks – Data Warehousing

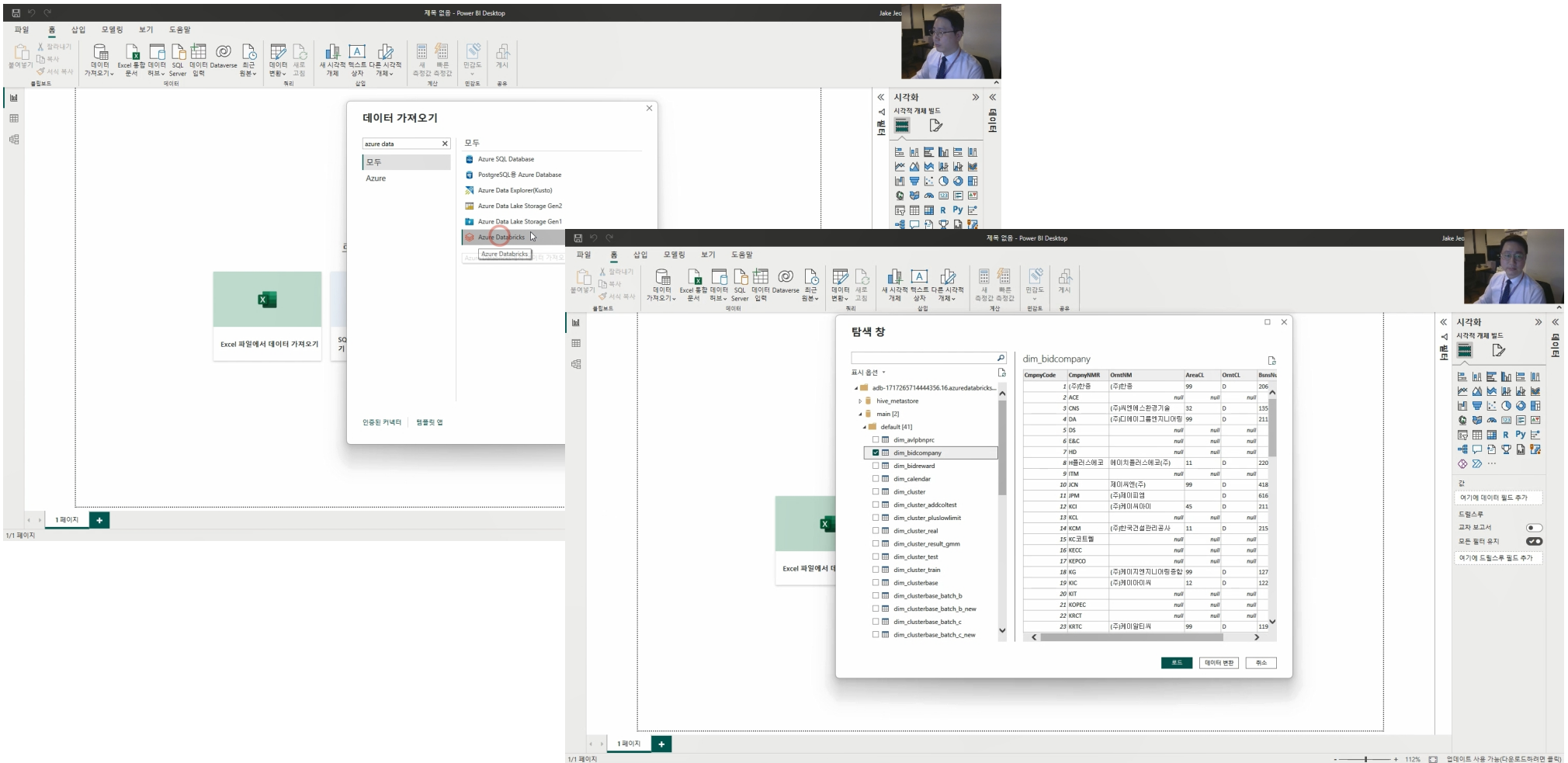


The screenshot displays the Azure Databricks Data Explorer interface. The left sidebar shows a tree view of the data catalog with 'sap_dw' selected. The main pane shows a table named 'sap_dw.sap_stage.ap_productvendor' with the following columns: ProductID, BusinessEntityID, AverageLeadTime, StandardPrice, LastReceiptCost, LastReceiptDate, MinOrderQty, MaxOrderQty, OnOrderQty, UnitMeasureCode, and ModifiedDate. The table contains 12 rows of data.

ProductID	BusinessEntityID	AverageLeadTime	StandardPrice	LastReceiptCost	LastReceiptDate	MinOrderQty	MaxOrderQty	OnOrderQty	UnitMeasureCode	ModifiedDate
1	1580	17	47.8700	50.2635	2011-08-29T00:00:00.000+0000	1	5	3	CS	2011-08-29T00:00:00.000+0000
4	1650	17	54.3100	57.0255	2011-08-29T00:00:00.000+0000	1	5		CTN	2011-08-29T00:00:00.000+0000
317	1578	19	28.1700	29.5785	2011-08-29T00:00:00.000+0000	100	1000	300	EA	2011-08-29T00:00:00.000+0000
317	1678	17	25.7700	27.0585	2011-08-25T00:00:00.000+0000	100	1000		EA	2011-08-25T00:00:00.000+0000
318	1578	19	34.3800	36.0990	2011-08-29T00:00:00.000+0000	100	1000		EA	2011-08-29T00:00:00.000+0000
318	1678	17	31.9800	33.5790	2011-08-25T00:00:00.000+0000	100	1000	300	EA	2011-08-25T00:00:00.000+0000
319	1556	19	44.2100	46.4205	2011-08-29T00:00:00.000+0000	100	1000	300	EA	2011-08-29T00:00:00.000+0000
319	1578	19	46.2700	48.5835	2011-08-29T00:00:00.000+0000	100	1000		EA	2011-08-29T00:00:00.000+0000
319	1678	17	43.8700	46.0635	2011-08-25T00:00:00.000+0000	100	1000		EA	2011-08-25T00:00:00.000+0000
320	1514	19	47.2800	49.6440	2011-08-29T00:00:00.000+0000	1	5	3	CAN	2011-08-29T00:00:00.000+0000
320	1602	17	45.2100	47.4705	2011-08-19T00:00:00.000+0000	1	5		CAN	2011-08-19T00:00:00.000+0000
320	1604	17	43.2100	45.3705	2011-08-25T00:00:00.000+0000	1	5		CAN	2011-08-25T00:00:00.000+0000
321	1514	19	42.2100	44.3205	2011-08-29T00:00:00.000+0000	1	5		CAN	2011-08-29T00:00:00.000+0000
321	1602	17	40.7600	42.7980	2011-08-19T00:00:00.000+0000	1	5		CAN	2011-08-19T00:00:00.000+0000
321	1604	17	38.5600	40.4880	2011-08-25T00:00:00.000+0000	1	5		CAN	2011-08-25T00:00:00.000+0000
322	1514	19	27.3300	28.6965	2011-08-24T00:00:00.000+0000	20	100		DZ	2011-08-24T00:00:00.000+0000

Microsoft Power BI는 셀프 서비스 비즈니스 인텔리전스 기능을 사용하여 대화형 시각화를 제공하는 비즈니스 분석 서비스로, Azure Databricks를 데이터 원본으로 사용하는 경우 데이터 과학자 및 데이터 엔지니어를 넘어 Azure Databricks 성능 및 기술의 이점을 모든 비즈니스 사용자에게 가져올 수 있습니다.

Azure Databricks – Power BI Connection

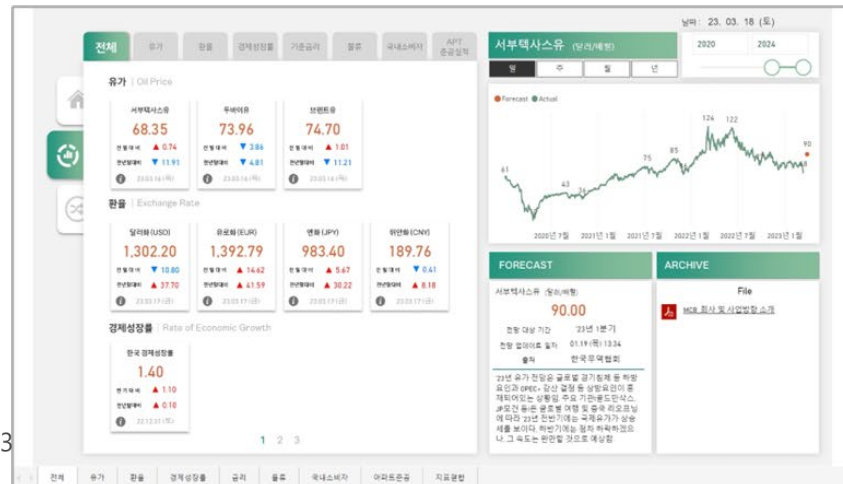
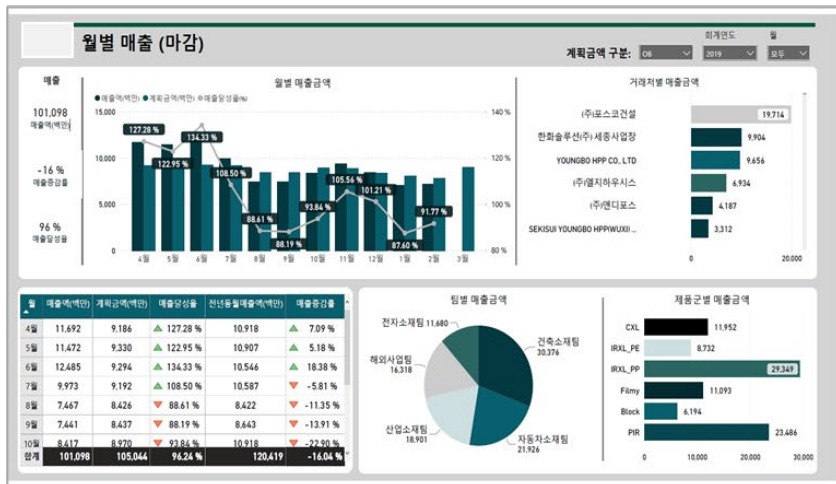
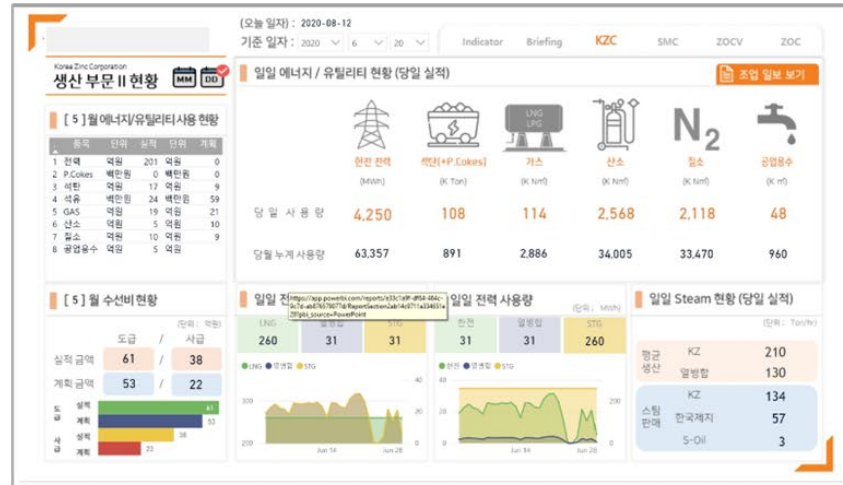
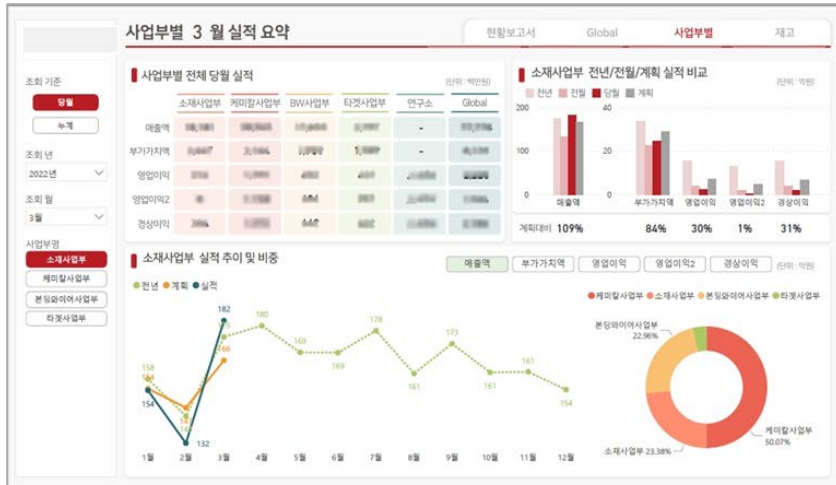


The screenshot shows the Power BI Desktop interface. The 'Get Data' dialog box is open, listing various data sources. 'Azure Databricks' is selected. Below it, a 'Table View' window displays a list of tables from the 'dim_bidcompany' dataset. The table view shows columns: CompanyCode, CompanyName, OrntNM, AreaCL, OrntCL, and Bonus.

CompanyCode	CompanyName	OrntNM	AreaCL	OrntCL	Bonus
1	(주)안동	(주)안동	99	D	206
2	ACE		null	null	null
3	CNS	(주)씨앤에스환경기술	32	D	135
4	DA	(주)대에이그룹엔지니어링	99	D	211
5	DS		null	null	null
6	EB&C		null	null	null
7	HD		null	null	null
8	H플러스에코	에이치플러스에코(주)	11	D	220
9	ITM		null	null	null
10	JCN	제이씨엔(주)	99	D	418
11	JPM	(주)제이피엠		D	618
12	KCI	(주)케이씨아이	45	D	211
13	KCL		null	null	null
14	KCM	(주)한국건설관리공사	11	D	215
15	KC&K	케이씨&케이		null	null
16	KECC		null	null	null
17	KECO		null	null	null
18	KG	(주)케이지엔지니어링중합	99	D	127
19	KIC	(주)케이아이씨	12	D	122
20	KIT		null	null	null
21	KOPEC		null	null	null
22	KRCT		null	null	null
23	KRTC	(주)케이알티씨	99	D	119

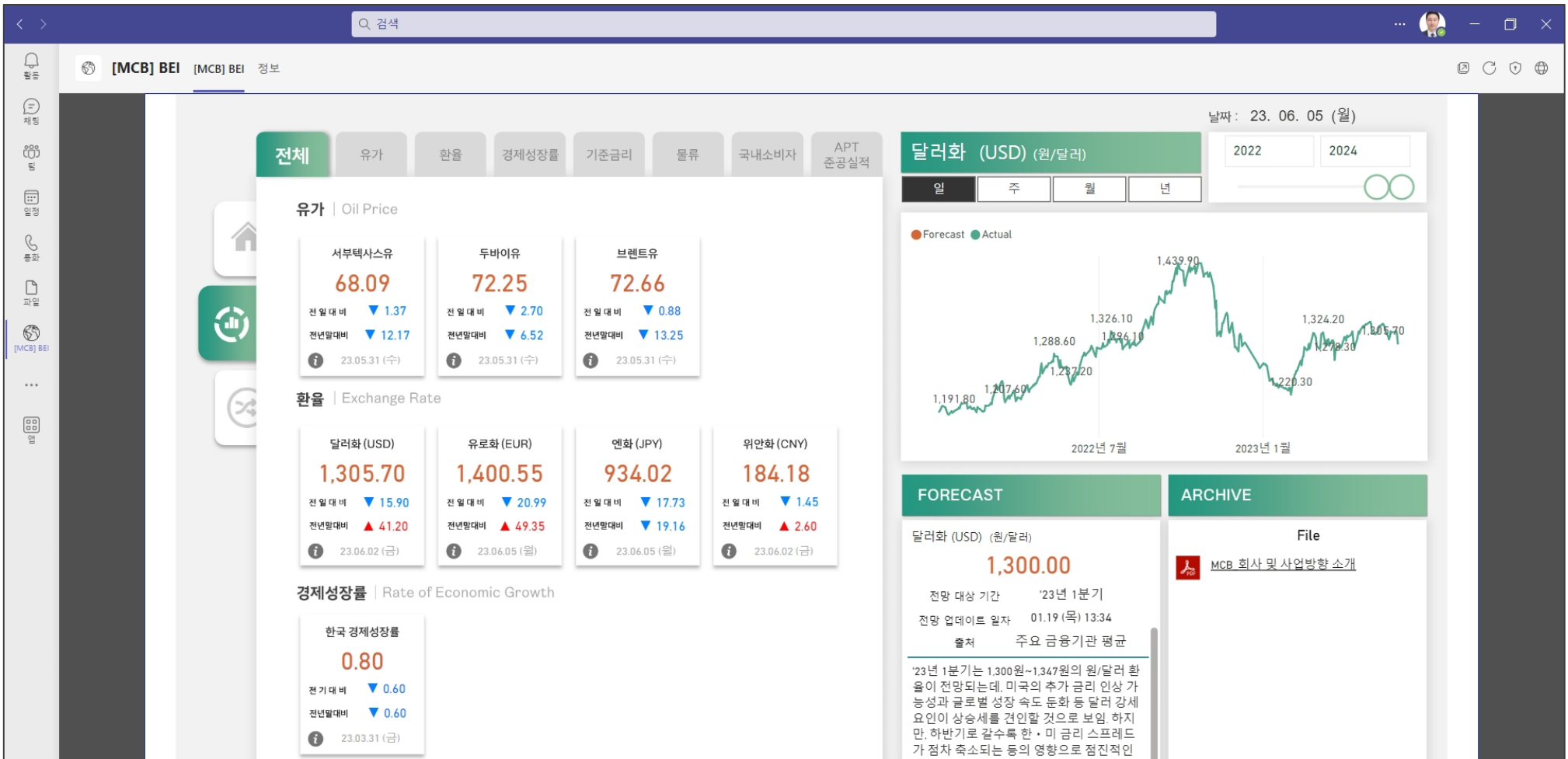
Microsoft Power BI Web Publish를 사용하여 보고서 및 대시보드에 대한 액세스 권한을 사용자에게 부여할 수 있으며 보고서를 공유할 경우 공유받은 사용자들은 이를 보고 조작할 수 있지만 편집할 수는 없습니다. 받는 사람은 보고서 및 대시보드에 표시되는 것과 동일한 데이터를 볼 수 있습니다.

Microsoft Power BI Web Portal Publish



Microsoft Power BI Web Publish 진행 후 Teams의 Power BI 앱에는 웹 브라우저에서 Power BI 서비스를 볼 때 표시되지 않는 기능이 있습니다.
 Teams의 Power BI 홈페이지에서 Teams에서 봤던 모든 Power BI 탭을 볼 수 있습니다. Teams의 검색 환경을 사용하여 최근 보고서, 대시보드, 앱을 찾을 수 있습니다.

Microsoft Power BI – Teams Collaboration



The screenshot displays a Microsoft Power BI dashboard within a Teams application. The interface includes a search bar at the top, navigation icons on the left, and a main content area with several data cards and a chart.

Navigation and Context:

- Search bar: 검색
- Page title: [MCB] BEI [MCB] BEI 정보
- Date: 날짜: 23. 06. 05 (월)

Data Cards:

유가 | Oil Price

상품명	현재 가격	전일 대비	전년말 대비	기준일
서부텍사스유	68.09	▼ 1.37	▼ 12.17	23.05.31 (수)
두바이유	72.25	▼ 2.70	▼ 6.52	23.05.31 (수)
브렌트유	72.66	▼ 0.88	▼ 13.25	23.05.31 (수)

환율 | Exchange Rate

통화	현재 환율	전일 대비	전년말 대비	기준일
달러화 (USD)	1,305.70	▼ 15.90	▲ 41.20	23.06.02 (금)
유로화 (EUR)	1,400.55	▼ 20.99	▲ 49.35	23.06.05 (월)
엔화 (JPY)	934.02	▼ 17.73	▼ 19.16	23.06.05 (월)
위안화 (CNY)	184.18	▼ 1.45	▲ 2.60	23.06.02 (금)

경제성장률 | Rate of Economic Growth

지표	현재 성장률	전기 대비	전년말 대비	기준일
한국 경제성장률	0.80	▼ 0.60	▼ 0.60	23.03.31 (금)

달러화 (USD) (원/달러) Chart:

The chart shows the exchange rate of USD against the Korean Won from July 2022 to January 2023. It compares actual data (green line) with a forecast (orange line).

날짜	실제 (Actual)	예측 (Forecast)
2022년 7월	1,191.80	1,207.60
2022년 8월	1,237.20	1,288.60
2022년 9월	1,296.10	1,326.10
2022년 10월	1,439.90	1,439.90
2022년 11월	1,220.30	1,220.30
2022년 12월	1,279.30	1,279.30
2023년 1월	1,324.20	1,324.20
2023년 2월	1,305.70	1,305.70

FORECAST: 달러화 (USD) (원/달러) 1,300.00

전망 대상 기간: '23년 1분기
 전망 업데이트 일자: 01.19 (목) 13:34
 출처: 주요 금융기관 평균

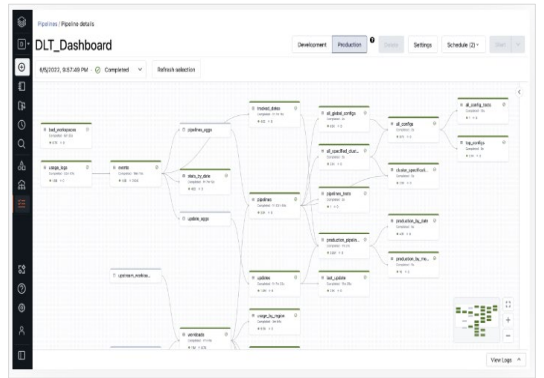
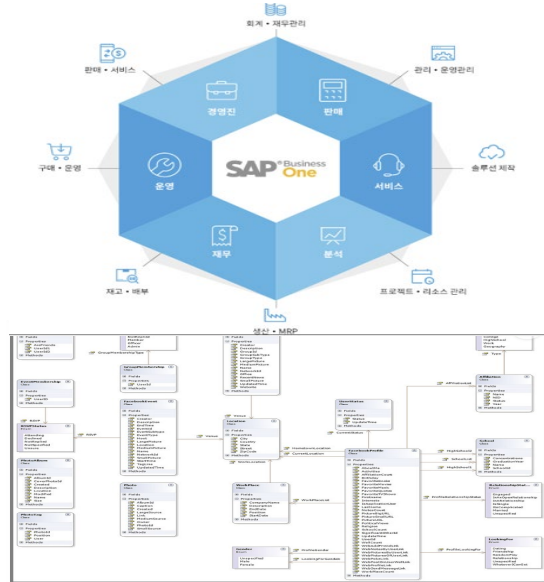
ARCHIVE: File: MCB 회사 및 사업방향 소개

Text in Forecast Card:
 '23년 1분기는 1,300원~1,347원의 원/달러 환율이 전망되는데, 미국의 추가 금리 인상 가능성과 글로벌 성장 속도 둔화 등 달러 강세 요인이 상승세를 견인할 것으로 보임. 하지만, 하반기로 갈수록 한·미 금리 스프레드가 점차 축소되는 등의 영향으로 점진적인...

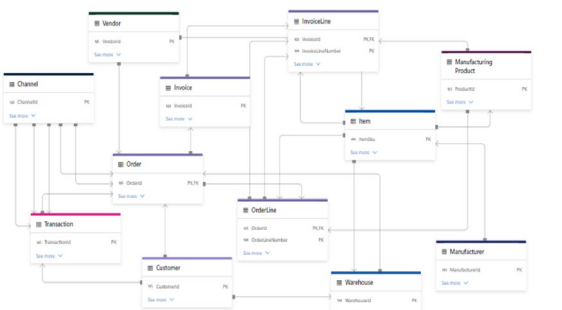
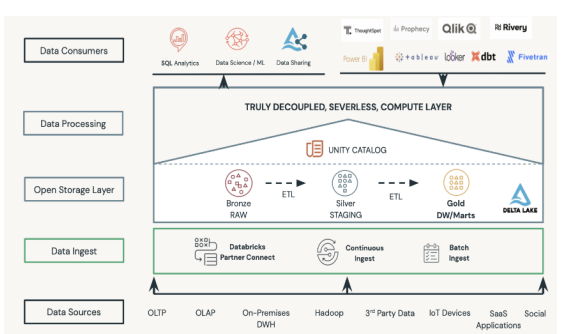


5. 『Azure Databricks Migration for SAP』 PoC 소개

Step 1 - 현황 및 요구사항 분석



Step 2 - 데이터 허브 구축



The screenshot shows the 'Create Cluster' form in Databricks. The form includes fields for Cluster Name, Databricks Runtime Version (4.8), Python Version (2), and Availability Zone (us-west-2c). The 'Create Cluster' button is highlighted in red.

Step 3 - 표준 보고서 수정 개발



SAP 마이그레이션의 단계별 상세 업무 중 PoC 진행 범위는 데이터 마이그레이션(Data Migration) 을 시작으로 ETL로 순차적으로 진행하
 여, PoC 성격상 단기간 내에 최적의 성과를 산출할 수 있도록 데이터 마이그레이션에 중점을 두고 진행하고자 합니다.

SAP · Databricks 마이그레이션의 단계별 상세 업무 진행

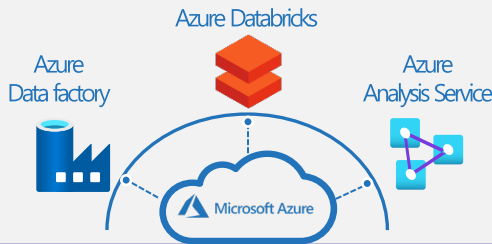
Track 1 : Data Migration		Track 2 : ETL		Track 3 : Code Migration	
Migrate data to Azure : <ul style="list-style-type: none"> • Sources: Data sources plugged in Teradata • Migrating subset of data from source systems to Azure • Create copy pipelines for data migration in Azure Data Factory • Execute data transfer • Testing and Validation • Document Options Available <ol style="list-style-type: none"> Option 1: Connect all required source systems to Azure using ADF and ingest data Option 2: Migration a subset of all required table data from above sources 		Migrate all ETL/ETL jobs as follows : <ul style="list-style-type: none"> • 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 <p>Manual Migration (Recommended)</p> <ol style="list-style-type: none"> Option 1: Migration to PySpark Jobs on Databricks Option 2: Migration to Azure Data Factory 		Migrate all logics, functions, Stored Procedures as follows : <ul style="list-style-type: none"> • 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 <ul style="list-style-type: none"> • Teradata script Analysis report • Bill of Materials for Databricks Options Available <ol style="list-style-type: none"> Option 1: Manual Migration Option 2: Migration using BladeBridge (Recommended) 	
Direct connection using ADF <ul style="list-style-type: none"> • For pilot, approvals might delay connectivity 	Migrate subset of source Data to Azure <ul style="list-style-type: none"> • Feasible approach for pilot. • Cons : Initial connectivity in the converted jobs must be changed again 	PySpark Jobs on Databricks <ul style="list-style-type: none"> • Code based approach • Install drivers to use spark connectors • Bit complex to leverage other Azure tools 	Using Azure Data Factory <ul style="list-style-type: none"> • GUI Based approach • Native connectors to multiple source systems • Easy to integrate with other Azure Tools 	Manual Migration <ul style="list-style-type: none"> • Hard to analyse pattern • Understand Use-cases and data models from scratch 	BladeBridge Migration <ul style="list-style-type: none"> • Automate code conversion. • Easier Analysis • Reader file available • Writer file available

본 PoC 프로그램은 SAP 데이터의 AI & 머신러닝 분석 플랫폼 구축 및 데이터 활용을 희망하는 고객사를 대상으로 본 프로젝트에 앞서 클라우드 데이터 허브 구축 환경 구성, SAP 데이터 마이그레이션 및 분석 플랫폼 구축, 운영 관리 교육을 제공하여 시스템 적합성과 솔루션 검증을 제공합니다.

PoC 프로그램 혜택

총 1,600만원 혜택

**클라우드 데이터
 인프라 환경 구성**
 (Azure Platform + Databricks)



기업 빅데이터 통합 분석 환경 효율 증명

Azure 클라우드 데이터 환경 제공

300만원

**SAP 데이터 마이그레이션 및
 데이터 허브 분석 플랫폼 구축**
 (Databricks Lakehouse)



기업 빅데이터 분석 플랫폼 효율 증명

마이그레이션 및 데이터 허브 구축 제공

1,000만원

**데이터 허브
 사용 및 운영 관리 교육**
 (Azure / Databricks)



사용 및 운영 효율 증명

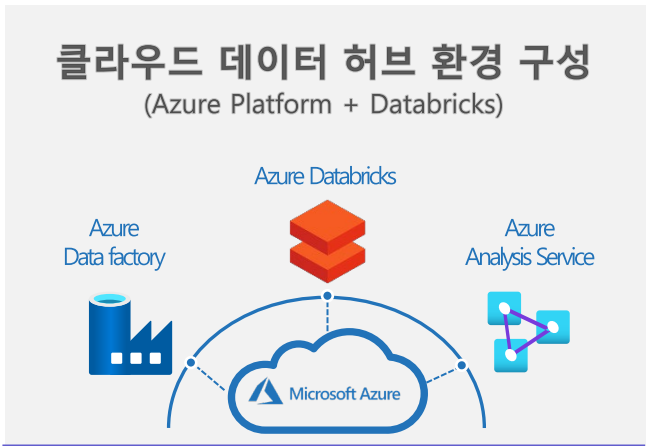
Azure, Databricks 및 보고서 개발 교육 제공

300만원

*해당 PoC는 마이크로소프트, 데이터브릭스 그리고 파트너가 무상 제공합니다.
 프로그램 신청을 원하시는 고객께서는 엠클라우드브리지 (Tel: 070 8808 9779, Email Info@mcloudbridge.com)로 연락주시기 바랍니다.

본 PoC 프로그램은 혜택으로 SAP 데이터와 이외의 내/외부 정형/비정형 등 모든 데이터 보안성, 정확성, 가용성, 사용성을 보장하기 위해 Azure의 인프라(Azure VM), 데이터브릭스(Azure Databricks) 통합 관리 환경 구성, 모니터링 및 퍼포먼스를 확인하여 안정적 클라우드 데이터 분석 환경을 제공 합니다.

PoC 프로그램 혜택 1. Azure Cloud 데이터 환경 제공

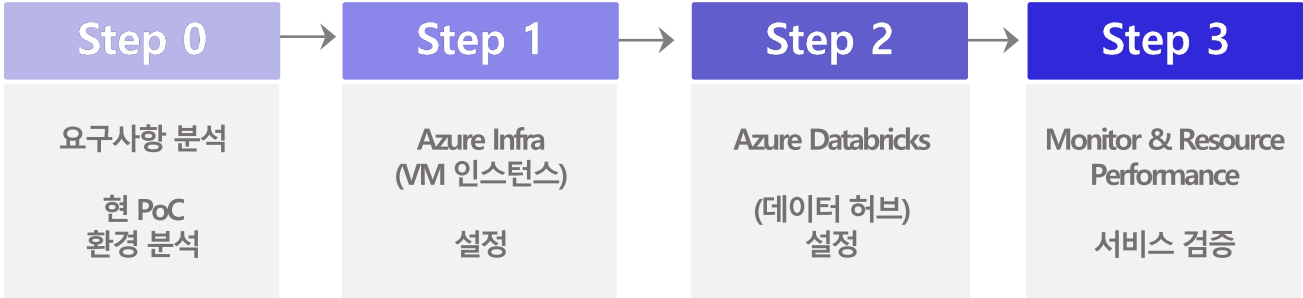


기업 데이터 통합 분석 환경 효율 증명

Azure 클라우드 데이터 환경 제공
300만원



클라우드 데이터 허브 환경 구성



본 PoC 프로그램은 혜택으로 기업이 보유한 SAP 데이터 및 외부의 정형, 반정형, 비정형 등 모든 데이터 정보가 클라우드 환경에서 통합 관리되는 데이터브릭스(Databricks)의 구축 및 기업 보유 데이터를 필요성에 따라 가공 및 분석할 수 있는 데이터 운영 환경 및 BI 표준 보고서를 제공합니다.

PoC 프로그램 혜택 2. Data Hub & Proto Type Report 제공

**SAP 데이터 마이그레이션 및
 데이터 허브 분석 플랫폼 구축**
 (Databricks Lakehouse)




Lakehouse

기업 빅데이터 분석 플랫폼 효율 증명

마이그레이션 및 데이터 허브 구축 제공
1,000만원



현황 및 요구사항
분석

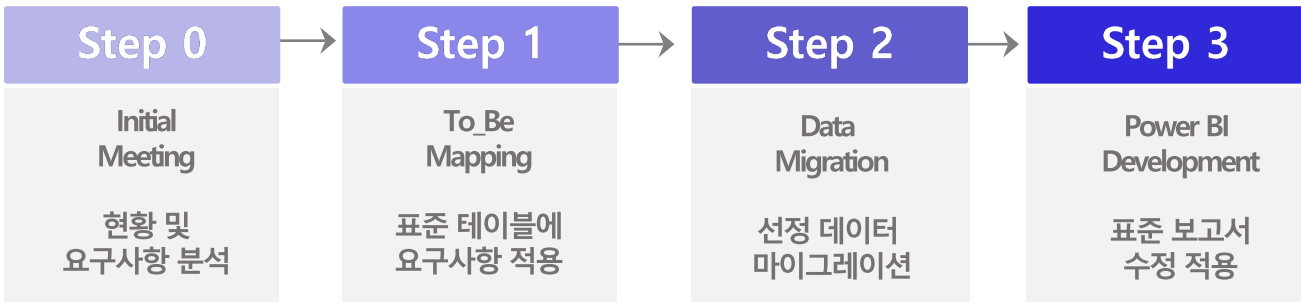


데이터 허브 구축
및 마이그레이션



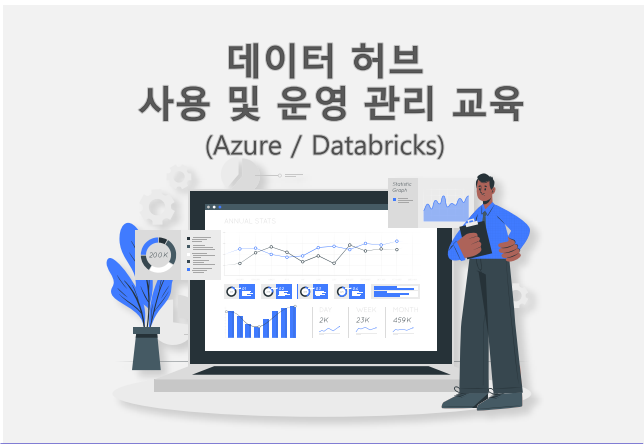
보고서 프로토타입
개발 제공

SAP 데이터 마이그레이션 및 데이터 허브 분석 플랫폼 구축

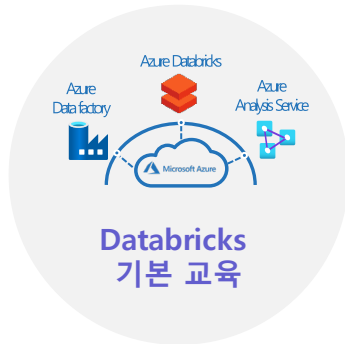


본 PoC 프로그램은 혜택으로 운영자 관리 교육을 통해 Azure 아키텍처를 이해하고, 데이터브릭스(Databricks) 환경이 최적의 성능을 발휘 할 수 있도록 데이터브릭스(Databricks), Azure 클라우드 관리자, 보고서 개발 사용자교육을 제공하여 효율적인 빅데이터 플랫폼 운영 및 관리가 될 수 있도록 합니다.

PoC 프로그램 혜택 3. Azure Databricks & BI 개발 교육 제공

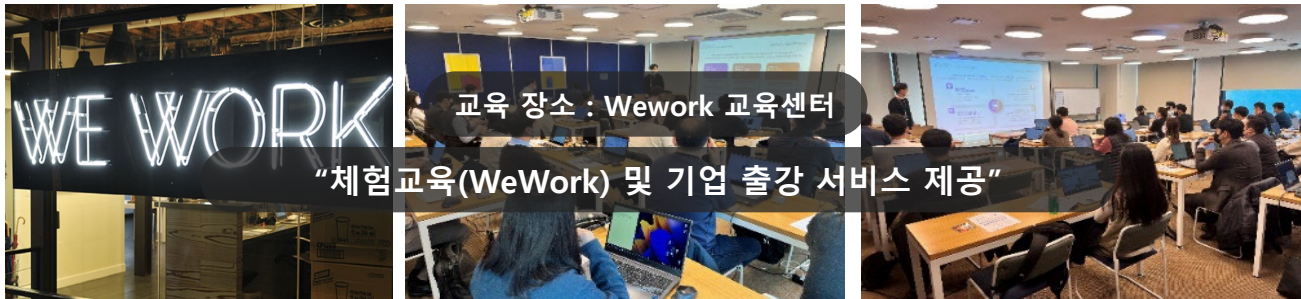


사용 및 운영 효율 증명



Azure & Databricks 관리 교육

Azure 및 Databricks 교육 제공
300만원





6. 엠클라우드브리지 소개

엠클라우드브리지는 디지털 전환시대(Digital Transformation)의 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 회사로서 라이선스 공급, 적용&구축, 유지보수 그리고 고객 맞춤 교육 서비스를 통합 지원하는 Data & AI 전문 기업입니다.



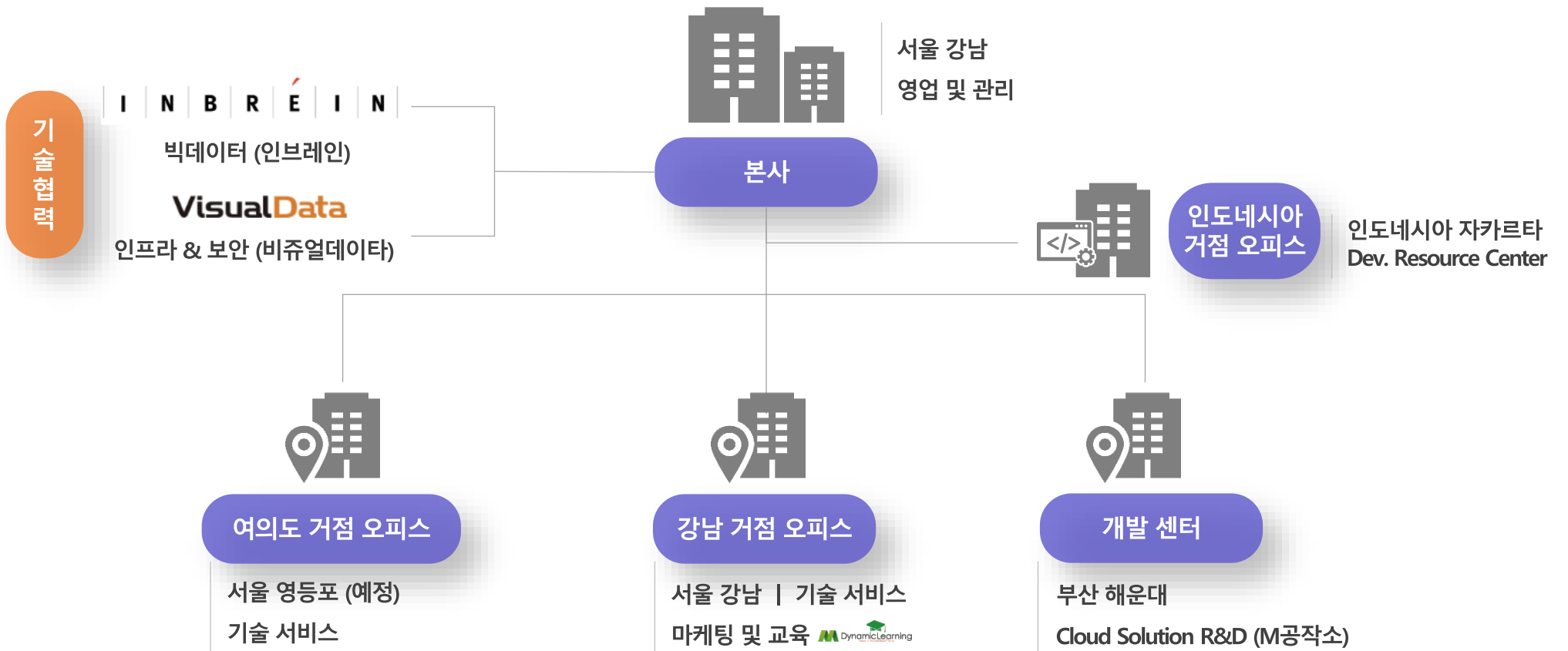
회사명
 엠클라우드브리지(주)



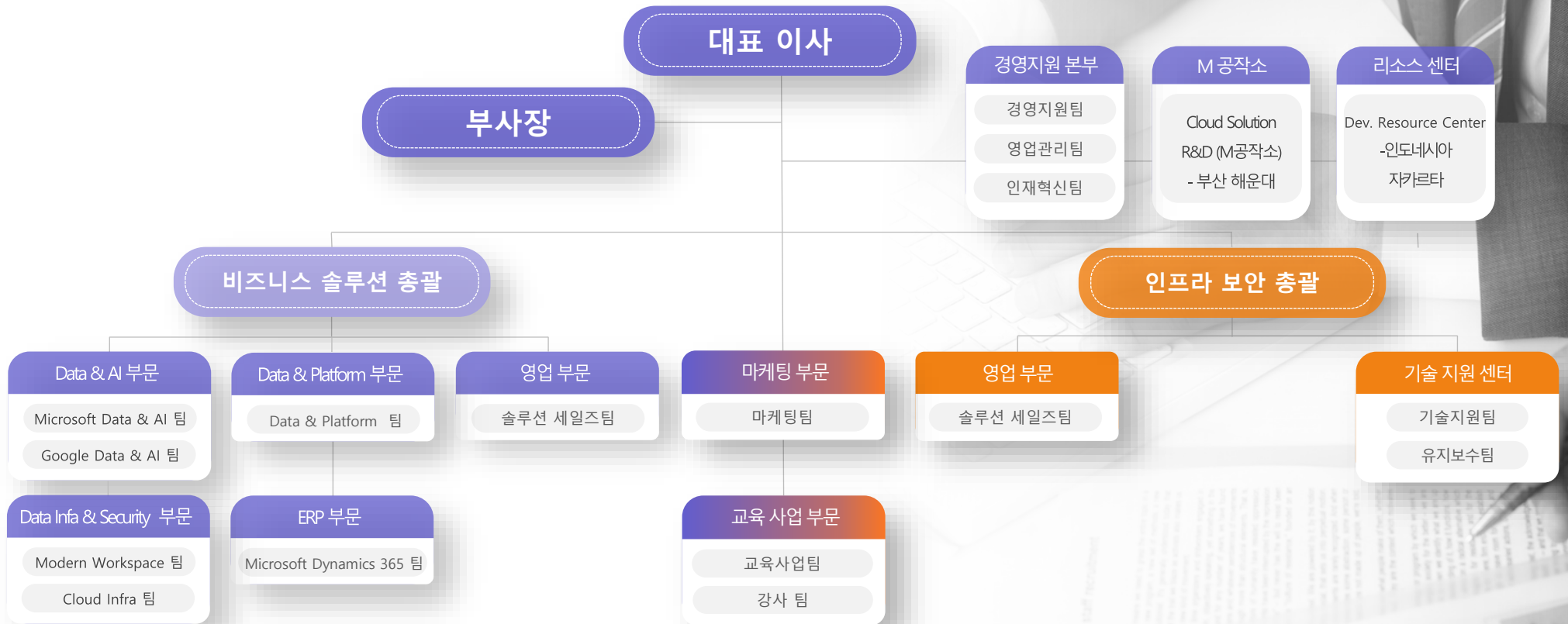
대표이사
 이 혁재

+ 설립일 Establishment	2022년 09월 19일
+ 자본금 Capital	11억
+ 관계사 Affiliated company	다이나믹러닝 인브레인 비주얼데이터 M공작소
+ 업종 Type of business	클라우드 서비스 전문 및 교육 사업
+ 주소 Address	서울특별시 강남구 테헤란로86길 15, 6층
+ 연락처 Tel Fax	Tel. 02. 552.9700 Fax. 02. 552. 9799
+ 홈페이지 Homepage	www.mcloudbridge.com

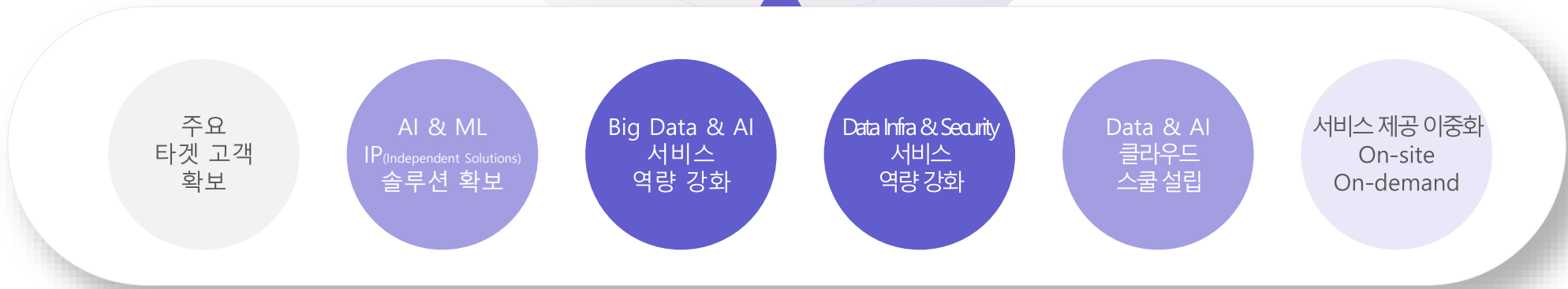
엠클라우드브리지는 고객 환경에 적합한 클라우드 데이터 서비스 제공을 위해 빅데이터(인브레인) 및 인프라&보안(비주얼데이터) 기술 협력사, 개발센터(M공작소), 각 지역 거점 오피스와 유기적으로 협업하고 있으며, 자체 보유하고 있는 교육센터를 통한 고객 맞춤 교육 서비스를 제공하고 있습니다.



엠클라우드브리지는 디지털 전환시대(Digital Transformation) 고객 환경에 적합한 클라우드 데이터 분석 환경 제공을 위해 조직된 각 분야 평균 10년 이상의 전문가 그룹으로서, Data 서비스 전문 컨설팅 기업 도약이라는 목표로 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(SaaS) 중심의 서비스 제공이라는 차별화 전략을 기반 Big Data & AI, Data Infra & Security 서비스 역량 강화 투자를 전개할 계획입니다.

















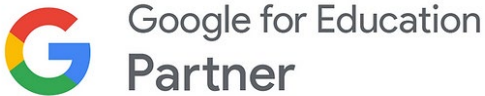
“데이터에 가치를 더하여 고객의 성장에 공헌합니다”

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 인증 자격을 갖춘 전문 컨설팅 조직과 지원조직을 갖추고 라이선스 공급, 적용 & 구축, 유지보수 그리고 교육 서비스를 통합 지원합니다.

보유 완료 및 취득 자격 (Data & AI)

<p>Microsoft Partner</p> <p>CSP & Solutions Partner</p> 		<p>Google Partner</p> <p>Data Cloud & Workspace Sell/Service Partner</p> 		<p>Databricks Partner</p> <p>Consulting & SI Partner</p> 	
				<p>Training Partner</p> 	
					

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>

엠클라우드브리지는 디지털 전환시대(Digital Transformation)의 핵심인 클라우드 기반 데이터 분석 및 시각화(Data & BI), 빅 데이터 및 AI (Big Data & AI), 데이터 관리 자동화(Data Flow & Automation), 이를 위한 데이터 인프라 및 보안(Data Infra & Security) 컨설팅 서비스를 제공하는 전문 회사입니다.

데이터에 가치를 더하여 고객의 성장에 공헌합니다

Data & BI

데이터 분석 및 시각화



- Azure Synapse Analytics
- Azure Stream Analytics
- Azure Databricks
- Power BI
- BigQuery
- Looker
- .
- .
- .

Big Data & AI

빅데이터 및 AI



- Azure Event Hub
- Azure Data Lake Storage
- Delta Lake
- Unity Catalog
- Spark Notebook
- BigQuery
- .
- .
- .

Data Flow & Automation

데이터 관리 및 자동화



- Power Automate
- Power Apps
- Power Virtual Agents
- Datastream
- Apps Script
- .
- .
- .

Data Infra & Security

데이터 인프라 및 보안



- Microsoft 365 & EMS
- Azure
- Google workspace
- Google Cloud Platform
- Kaspersky
- Ahnlab
- .
- .
- .

엠클라우드브리지는 클라우드 기반 Data & AI 전문 서비스를 제공하기 위해 각 분야별 전문 조직을 보유하고 있는 전문 파트너로서 디지털 전환 시대(Digital Transformation)의 기업에 필요한 클라우드 라이선스 공급, 적용&구축, 유지보수 그리고 고객 맞춤 교육 서비스를 주요 사업 영역으로 합니다.

주요 사업 영역 및 제공 서비스

Data & BI

데이터 분석 및 시각화



Big Data & AI

빅데이터 및 AI



Data Flow & Automation

데이터 관리 및 자동화



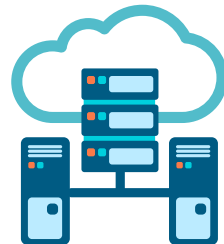
Data Infra & Security

데이터 인프라 및 보안



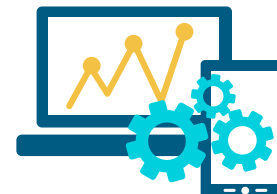
License

라이선스 공급



Service

적용 & 구축 서비스



Support

유지 보수



Training

고객 맞춤 교육

<p>Data & BI 데이터 분석 & 시각화</p>	<p>Big Data & AI 빅데이터 및 AI</p>	<p>Data Flow & Automation 데이터 관리 및 자동화</p>	<p>Data Infra & Security 데이터 인프라 및 보안</p>
 THE HYUNDAI   	    'TORAY' 도레이첨단소재 	 THE HYUNDAI  	 DOHWA      

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