



AI·DATA
AI & DATA ANALYTICS CENTER

DP(DataPlatform)360 Catalog

Contents.

1. MEGAZONECLOUD & ADC (AI & 데이터분석센터)
2. DP360
3. DP360 Catalog

AI & Data Analytics Platform

01. MEGAZONECLOUD & ADC(AI & Data Analytics Center)

Business & Offerings

메가존클라우드는 기업 고객의 클라우드 도입부터 디지털 트랜스포메이션을 위한 컨설팅, 구축 및 운영 관리를 포함한 다양한 서비스를 통해 기업 고객의 성공적인 디지털 전환과 최적화를 돕고 있습니다.



AI & Data Analytics

데이터 분석 플랫폼 컨설팅 및 구축
AI/ ML 데이터 분석 컨설팅 및 구축

Infrastructure

디지털 혁신을 위한 최적화된 인프라

Platform

디지털 혁신 가속화를 위한
애플리케이션 개발/ 운영/ 관리 플랫폼

Application

신속한 디지털 비즈니스 의사결정 및 관리를
위한 비즈니스 애플리케이션

AI & Data Analytics Center (ADC)

주요 비즈니스 영역



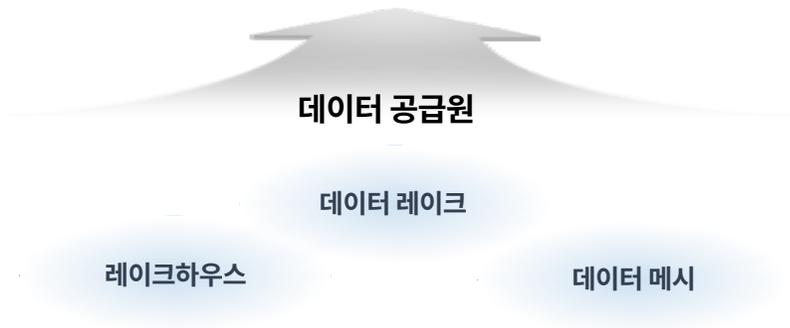
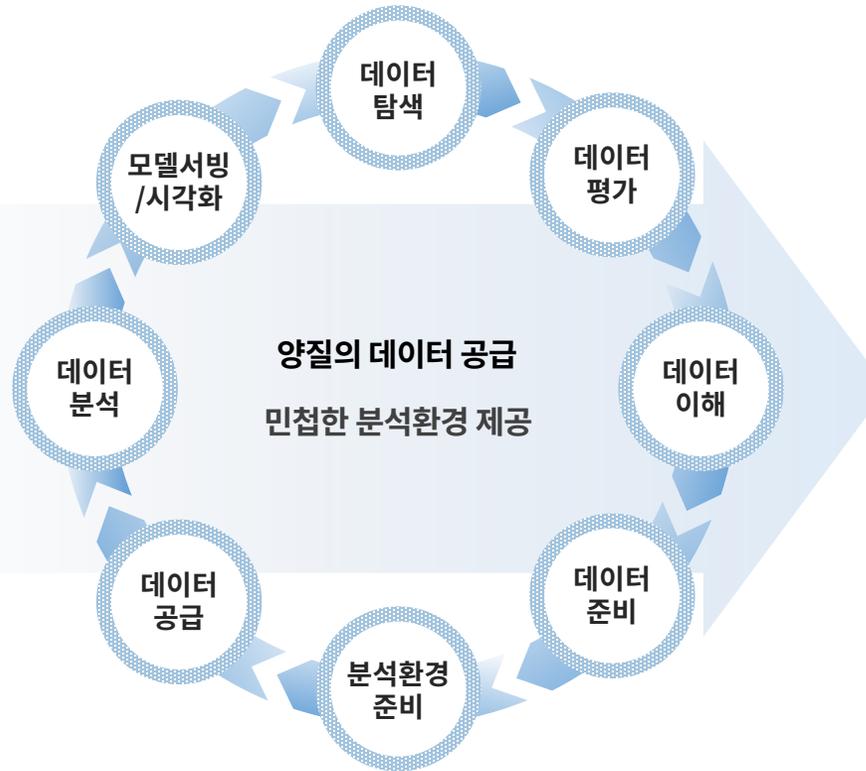
AI & Data Analytics Platform

02. DP360 DataPlatform360





AI가 공기같은 존재가 되어가는 지금 ‘데이터 플랫폼’이 갖추어야 하는 역할 고품질 데이터와 민첩한 분석환경 필요



* ANI(Artificial Narrow Intelligence, 제한적인공지능/약인공지능) 게임, 의학, 법학 등 특정 분야의 제한된 소재를 다루는 인공지능. 예) ChatGPT, AlphaGo, IBM Watson 등

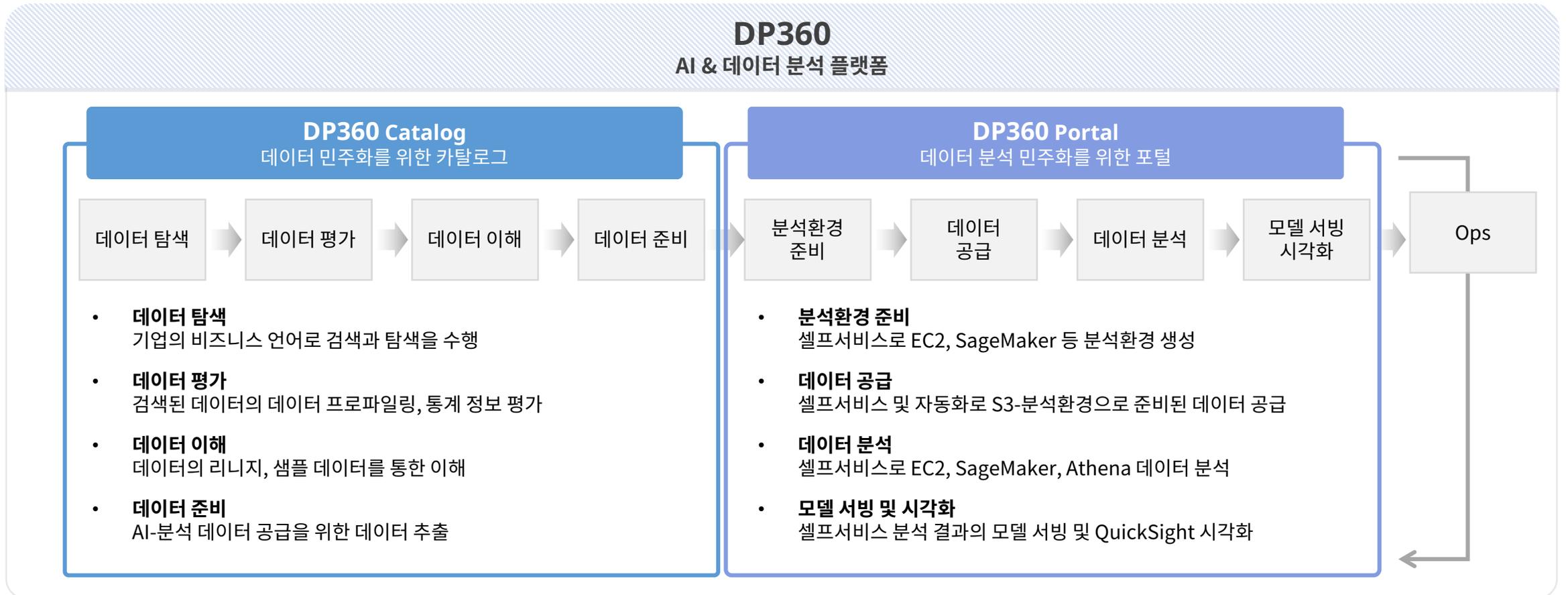


DP360 (DataPlatform360)

‘데이터 자산화 - 가치 탐색 - 가치 평가 - AI 분석 활용’ 등 기업의 데이터 혁신을 위한 프로세스 전반을 지원하는 AWS 기반의 엔터프라이즈 AI & 데이터 분석 플랫폼

DP360

AI & 데이터 분석 플랫폼



AI & Data Analytics Platform

03. DP360 Catalog

DataPlatform360 Catalog



기업의 데이터 제공 프로세스

데이터 디스커버리 부재 및 메타데이터 관리시스템 의존

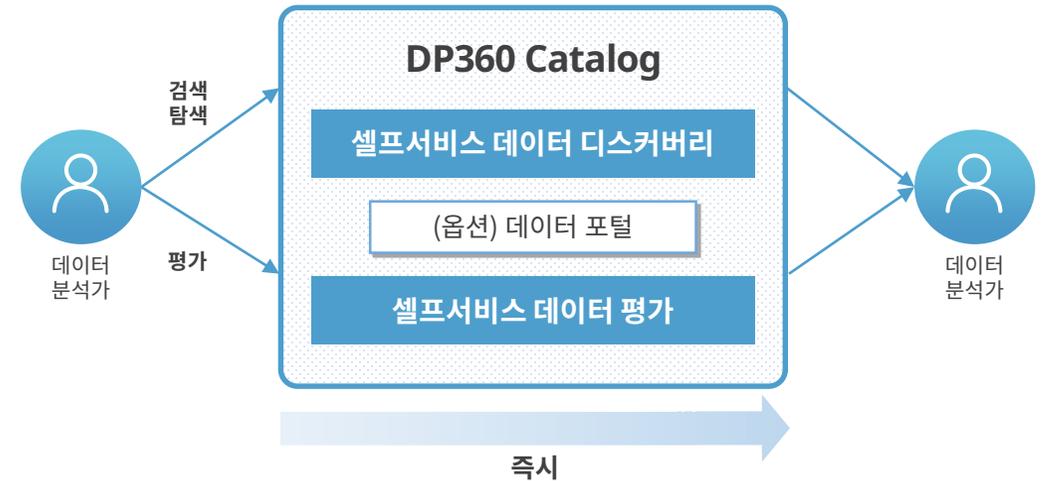


문제점

- 메타데이터 관리시스템은 전문가의 메타 정보 제공 및 통제를 위한 용도
- 검색을 이용한 기술 메타 및 비즈 메타 검색 후 필요 데이터 선정, 입수, 평가 및 재 요청 과정이 단순 반복
- 중앙 관리되는 검색어 및 카테고리, 태그에 기반한 폐쇄형 데이터 디스커버리
- DBMS 기반의 정형 데이터 저장소에 특화

DP360 Catalog

셀프서비스 데이터 디스커버리



개선사항

- 분석을 위한 데이터를 검색, 탐색, 플랫폼 분류 및 비즈니스 용어 등의 다양한 방식으로 **데이터 디스커버리**
- 관련된 데이터셋에 대한 개인별, 그룹별로 셀프서비스 태깅으로 언제든지 재활용 시 **신속한 접근성 확보**
- 데이터 자산의 **단일 진실 공급원**으로 동작
- 정형저장소 및 비정형 저장소인 **데이터 레이크 지원**



데이터 민주화, 강력한 데이터 거버넌스를 위한 차세대 데이터 카탈로그 ‘DP360 Catalog’

전통적인 IT 전문가 대상의 메타데이터 관리 시스템이나 중앙 집중식 데이터 제공자를 위한 데이터 인벤토리 카탈로그가 아닌 데이터 소비자들을 위한 혁신적인 데이터 카탈로그 구축

다양한 데이터 자산 지원

- 과거의 테이블 기반의 데이터 자산이 아닌 DBMS, BI, 데이터 레이크, 머신러닝, Hadoop 등 다양한 데이터 자산 통합
- 최신의 메타데이터 발생 제품 지원
Athena, Trino, S3, Databricks, Snowflake, BigQuery 등

- 단순하게 데이터의 테이블, 컬럼(column) 등의 정보 제공을 위한 메타데이터의 저장소가 아닌 메타데이터의 라이프사이클 관리
- 데이터셋, 리포트 등 다양한 유형의 자산이 통일된 검색 및 탐색체계에서 통합

메타데이터의 데이터화

엔드투엔드 데이터 가시성 제공

- 단순한 데이터의 물리, 논리 정보 및 위치 정보가 아닌 데이터 360 뷰 제공
- 데이터셋, BI 리포트, 머신러닝 모델 등 모든 엔티티에 대한 데이터 리니지, 데이터 프로파일링 등의 가시성 제공

- 중앙화된 데이터의 검색 및 탐색 정책에 의존하지 않는 데이터 소비자 관점의 데이터 디스커버리 방식 제공
- 실시간 데이터 카탈로그에 사용자별 태깅과 검색, 데이터셋에 대한 커뮤니티 등 협업 기능 제공

임베디드 협업





데이터 민주화, 강력한 데이터 거버넌스를 위한 데이터 카탈로그



데이터 디스커버리

- DBMS, BI, ML, 데이터 레이크 기능 저장소, ETL 등에 대한 통합 검색
- 기술, 논리 메타데이터 결합으로 데이터 360° 뷰 제공

최신의 데이터 거버넌스

- 태그, 용어집, 도메인 기반의 엔티티 관리
- 사용자, 그룹, 조직 단위 엔티티 소유권 관리

데이터 리니지 및 프로파일링

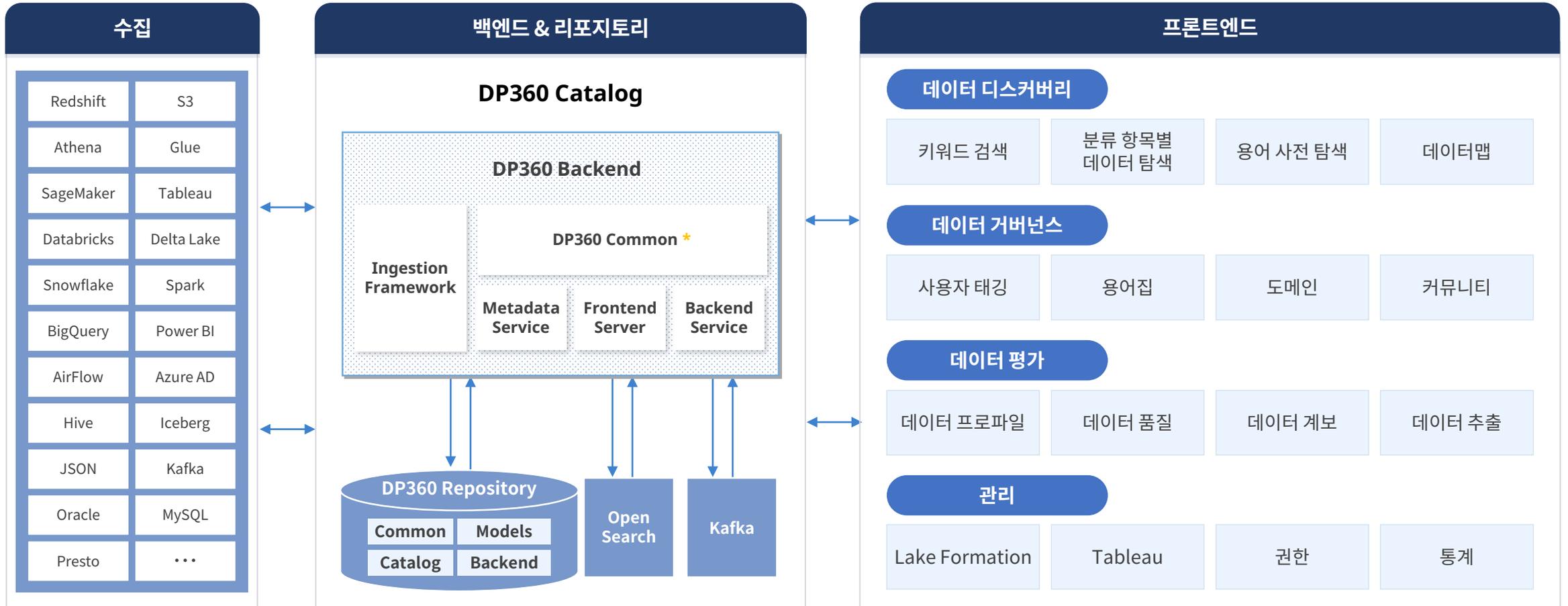
- 플랫폼, 데이터셋, ETL, BI 리포트 등 계보 추적
- 프로파일링을 통한 데이터의 형태, 분포 등 데이터 평가
- 샘플데이터 추출을 통한 데이터 평가

AWS Lake Formation 접근 권한

- AWS Lake Formation의 데이터 접근 권한 관리 통합
- 데이터셋의 Tag, 리소스 기반 데이터 접근 권한 관리

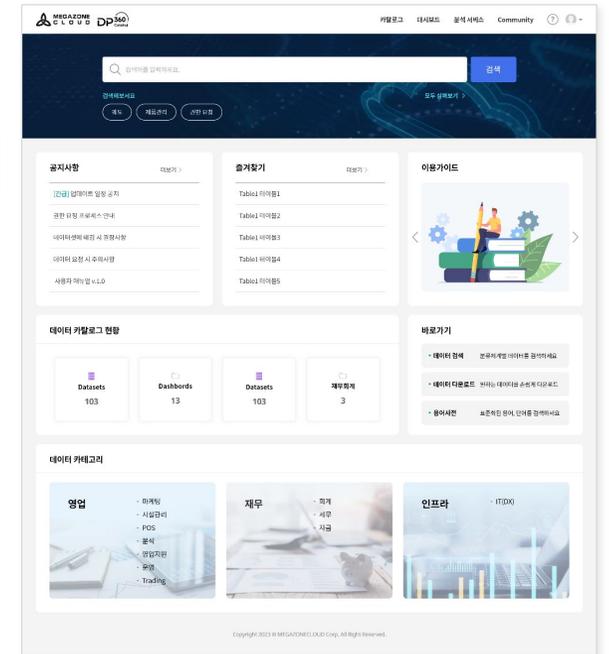


엔드투엔드 데이터 검색, 데이터 디스커버리 및 데이터 거버넌스 지원



* DP360의 Common 모듈을 DP360 Portal과 공유

데이터 & AI 분석을 위한 진실 데이터 공급 및 공유 단일화를 통해 분석 거버넌스 구축



<h3>데이터 디스커버리 & 평가</h3>	<ul style="list-style-type: none"> • 사용자의 편리한 데이터 디스커버리 및 데이터 거버넌스 • 데이터의 프로파일링 및 통계, 샘플데이터 등을 통한 데이터 평가 • AWS Lake Formation 데이터 접근 권한 	<p>검색</p>	<p>데이터 리니지</p>
<h3>데이터 협업</h3>	<ul style="list-style-type: none"> • 사용자별 데이터 디스커버리를 사용자화할 수 있는 사용자 태깅 • 사용자가 직접 입력할 수 있는 데이터셋 정보 WIKI 페이지 • 협업을 위한 커뮤니티 	<p>탐색</p>	<p>데이터 프로파일링</p>
		<p>데이터 다운로드</p>	<p>데이터 접근 권한</p>
		<p>태그</p>	<p>데이터 정보</p>
		<p>용어사전</p>	<p>도메인</p>
		<p>데이터 오류 게시판</p>	<p>데이터 요청 게시판</p>



데이터 카탈로그 메인

로그인

- 기업 포털과 SSO 연동 및 계정 통합
- Azure Active Directory 연동 가능
- AWS IAM Identity Center (IDC) 연동 가능
- DP360 내장 계정 기능 이용 가능

메인 화면

이트 일정 공지 Table1 테이블1

데이터 카탈로그 현황

Datasets	Dashboards	Datasets	재무회계
103	13	103	3

바로가기

- **데이터 검색** 분류체계별 데이터를 검색하세요
- **데이터 다운로드** 원하는 데이터를 손쉽게 다운로드
- **용어사전** 표준화된 용어, 단어를 검색하세요

데이터 카테고리

영업 마케팅 재무 회계 인프라 IT(DX)



데이터 맵 및 데이터 목록 기반 탐색

데이터 맵

MEGAZONE CLOUD DP360 Catalog

검색어를 입력하세요.

카탈로그 대시보드 분석 서비스 Community

카테고리

본문검색

10. 영업 > 11. 마케팅 > Link-hover > M01.B2C EDW

데이터맵 목록

DW 345 12 4

ERP 76 5 23

계약시스템 74 80 32

상담시스템 153 254 341

BI 9 6 0

영업분석 8 6 3

상품 1.2k 80 987

✓ 데이터 맵 기반 데이터 탐색

데이터 목록

MEGAZONE CLOUD DP360 Catalog

검색어를 입력하세요.

카탈로그 대시보드 분석 서비스 Community

카테고리

본문검색

10. 영업 > 11. 마케팅 > M01.B2C EDW

데이터맵 목록

Table 영업 > 마케팅 > B2C EDW > DW

신계약 실적 14 행 수: 12,999,999 Legacy Data Ingestion

Table 영업 > 마케팅 > B2C EDW > DW

청약 14 행 수: 12,999,999 Legacy Data Ingestion

Table 영업 > 마케팅 > B2C EDW > DW

유지율 14 행 수: 12,999,999 Legacy Data Ingestion

Table 영업 > 마케팅 > B2C EDW > DW

Call인입 14 행 수: 12,999,999 Legacy Data Ingestion

Table 영업 > 마케팅 > B2C EDW > DW

수수료 지급 14 행 수: 12,999,999 Legacy Data Ingestion

Table 영업 > 마케팅 > B2C EDW > DW

설계사 정보 14 행 수: 12,999,999 Legacy Data Ingestion

✓ 목록 기반 데이터 탐색



데이터셋 상세 조회 및 용어사전

데이터셋 상세 조회

MEGAZONE CLOUD DP360 CATALOG

카탈로그 대시보드 분석 서비스 Community

10. 영업 > 11. 마케팅 > M01_B2C EDW

데이터셋 목록

Table > L2 > SALE_MASTER

SALE_MASTER (매출 마스터)

당첨자: (현업) A티 홍길동 [운영] 마스터 Tag: Legacy Data Ingestion

컬럼: 14 행 수: 12,988,000

관련 설명 계보 데이터

Q 검색 최종 업데이트: 2023-01-01 13:00 2.0.0-7일 전

데이터 컬럼명	데이터 타입	관급명	영문명	컬럼 설명	Tag	History
Customer_ID	String	고객번호	Customer_ID	고객번호	#고객번호	Customer Identifier Number
Customer_Name	String	고객명	Customer_Name	고객명	#고객이름	Customer Name, Name Customer
Item_Name	String	상품명	Item_Name	상품명	#상품 #판매단	Item_Name, Item
Contract_Shipment_Number	String	계약번호	Contract_Shipment_Number	계약번호	#증권번호	Contract_Shipment Number, Number

컬럼 설명 계보 데이터

#	기준일자	datekey	col_02
1	2023-07-10	25	1,500,230
2	2023-07-10	25	1,500,230

다운로드

✓ 데이터셋 상세 정보 및 데이터 추출

용어사전

MEGAZONE CLOUD DP360 CATALOG

카탈로그 대시보드 분석 서비스 Community

용어사전

단어

부모사전

- 단어1
- 단어2
- 단어
- 단어

용어

업무용어

부모사전

벤젠

#사전논리명 #사전물리명 #사전물리전체명 #표준 or 비표준

벤젠(영어: benzene) 또는 벤졸(benzol)은 휘발성 액체 탄화수소로 플라르를 종류, 정제하여 만든다. 6개의 탄소 원자가 동인 평면에 있는 평면 정육각형의 고리구조를 가지며 가장 기본적인 방향족 탄화수소 물질이고 대표적인 아로렌(6-아로렌)이다. 무색이고 가연성이 있으며, 발암 물질로도 알려져 있다. 비등점은 섭씨 80.1도이고 화학식은 C6H6이다. 가끔 페닐기에 수소가 붙어 있다는 의미로 Ph-H로 표기되기도 한다. 상대적으로 높은 끓는점을 가진 것으로 알려져 있으며 달콤한 냄새를 가진 것이 특징이다. 발암성 때문에 가솔린에 첨가는 제한되었지만 화학 공정에서 중요한 용매로 사용되며, 약품, 플라스틱, 인조 고무, 염료, 향료, 폭약 등의 원료로 쓰인다. 벤젠은 원유의 성분이며, 석유에 존재하는 다른 물질들을 통해서 합성할 수도 있다.

벤젠을 구성하는 6개의 탄소 원자는 서로 공유결합을 이루며, 6개의 탄소 원자 끝에 각각 1개의 수소 원자가 공유결합을 이루고 있다. 탄소 원자 간의 결합은 단일결합도, 이중결합도 아닌 중간적 성질을 가지며, 탄소원자 사이의 결합길이는 0.139nm로 모두 같다. 공명 구조를 이루고 있어 화학적으로 매우 안정하다. 벤젠은 무극성 분자이다.

C₆H₆

Benzene
Molecular formula

Kekulé Structures
(Resonance Forms)

Planar Hexagon
Bond Length 140 pm

Sigma Bonds
Hybridized orbitals

6 p_z orbitals

✓ 용어사전 및 커뮤니티



리니지 정보 (데이터베이스, 데이터셋, 리포트, 대시보드 등)

Platform

- Tableau (5.3k)

Type

- Charts (4.5k)
- Datasets (489)
- Dashboards (402)
- Containers (192)

Sub Type

- Embedded Data Source (398)
- Workbook (138)
- Custom SQL (58)
- View (58)
- Project (54)
- Published Data Source (33)

Owned By

- chaesh@mz.co.kr (5k)
- wtkim (156)
- BOMD2 (30)
- MOED_Site_Admin (7)
- Accounting_Explorer (6)

대시보드 리니지 정보

Charts > tableau > 99.BOMD2_Test > DBC > 3. Medium type Customer Report (상세) Details Lineage 1 upstream, 1 downstream

Dashboard > Tableau > ... 3. Medium View in Tableau Columns All Time

AWS Usage Report (상세) - E C2

6 charts Changed 6 months ago

Last synchronized 11 minutes ago

About

AWS Usage Report (상세) - EC2

Chart Info

Tableau | Embedded Data Source
v_cost) (dm) | 프로젝트...

Tableau | Chart
EC2 Cost Type별 - 수...

Tableau | Dashboard
AWS Usage Report (상세...

Tableau | Dashboard Source
v_cost) (dm) | 프로젝트...

- Operation
- Month Nm
- Year Nm
- YYYY_MM
- Product
- Account Nm
- Cost

Tableau | Chart
EC2 Cost Type별 - 수...

- Account Nm
- Company Nm
- Cost Type
- Product
- City Desc
- Yyyymm Dt

Tableau | Dashboard
AWS Usage Report (상세...

대시보드 상세 정보

Dashboard Details

차트 계보 정보



ML 파이프라인 리니지 정보, 모델 및 피쳐 테이블 정보

Type

- Pipelines (44)
- Tasks (44)
- ML Models (9)
- Feature Tables (4)
- ML Groups (1)

Platform

- SageMaker (102)

Summary Documentation Group Features Properties

Training Metrics

Name	Value
ObjectiveMetric	2.18509984499717
train_rmse	1.623990058989208
validation_rmse	2.18509984499717

Hyper Parameters

Name	Value
_training_objective_metric	validation_rmse
alpha	0.785218850295352

Features Sources Properties Documentation

Feature	Source	Properties
no_previo us_contact	사전 접속 횟수	
not_worki ng	작동 안함	
FS_time	Feature Store 등록시각	

Tasks > training

Details Lineage 2 upstream, 2 downstream

Show Full Titles Compress Lineage Show Columns All Time

데이터셋 상세 정보

Dataset AWS S3 sagemaker-us-east-1-551075-147781/sagemaker/videogames-usb-boost/train

Last synchronized 15 hours ago

About

- Feature Store: videogames-xgboost/train
- 내용: xgboost

attribute_names None

channel_name train

dataset_type s3

Task 상세 정보

Task xgboost-reg-06-12-00-11

Documentation Properties Pipeline Lineage Runs

Best Training Job

CommandExecution

Created Time

HyperParameters

HyperParameters

HyperParameters

HyperParameters

HyperParameters

JobType

모델 상세 정보

MLModel SageMaker videogames-xgb oost-2020-10-05

Last synchronized 15 hours ago

About

XGBoost Algorithm

1. 초기 모델은 상수로 설정하며 다음과 같이 찾는다.

$$F_0(x) = \arg \min_{\sigma} \sum_{i=1}^n L(y_i, \sigma)$$

2. $m = 1, \dots, M$ 에 대하여 다음을 반복한다.

(a) Gradient g_m 와 Hessian h_m 를 계산한다.

$$g_m = \left[\frac{\partial L(y_i, F_{m-1}(x))}{\partial F(x)} \right]_{F(x)=F_{m-1}(x)}$$

$$h_m = \left[\frac{\partial^2 L(y_i, F_{m-1}(x))}{\partial F(x)^2} \right]_{F(x)=F_{m-1}(x)}$$

Copyright 2024 © MegazoneCloud Corp. All Right Reserved.

19

수집 지원되는 메타데이터



 Redshift	 S3 Data Lake	 Athena	 Glue	 SageMaker	 Tableau
 Databricks	 Delta Lake	 Snowflake	 Spark	 BigQuery	 Power BI
 Airflow	 Azure AD	 Hive	 Iceberg	 JSON Schema	 Looker
 Kafka	 Kafka Connect	 LDAP	 MariaDB	 Metabase	 MS SQL Server
 MongoDB	 MySQL	 Oracle	 PostgreSQL	 Presto	 Trino
 Presto on Hive	 Redash	 Salesforce	 SAP HANA	 GCS	 Superset
 Elasticsearch	 Great Expectation	 Feast	 Business Metadata - yaml	 OpenAPI	 CSV



표준화 된
데이터 관리

명확한
데이터 이해

데이터
혁신을 통한
DT 가속화

데이터 디스커버리



데이터, 리포트, ML모델 등에
대한 검색 및 탐색

데이터 거버넌스



단일의 진실 공급원으로 전사
데이터의 일관성 제공

데이터 리니지



데이터, 데이터 파이프라인,
리포트의 계보 추적

데이터 품질



데이터 프로파일링 및 통계,
품질 관리

데이터 접근 제어



Lake Formation과 연계된
데이터 권한 제어



Talk to DATA CREW!

상세한 내용 또는 궁금하신 점이 있으세요?
데이터 크루에게 연락 주세요.

기술 미팅 혹은 찾아가는 세미나를 통해
데이터 크루가 도와드리겠습니다.

AskData@megazone.com

<https://AIDATA.megazone.com>