



# 제조 기업을 위한 INDUSTRY INSIGHT DAY

THE ERA OF ENTERPRISE AI

2024.12.11

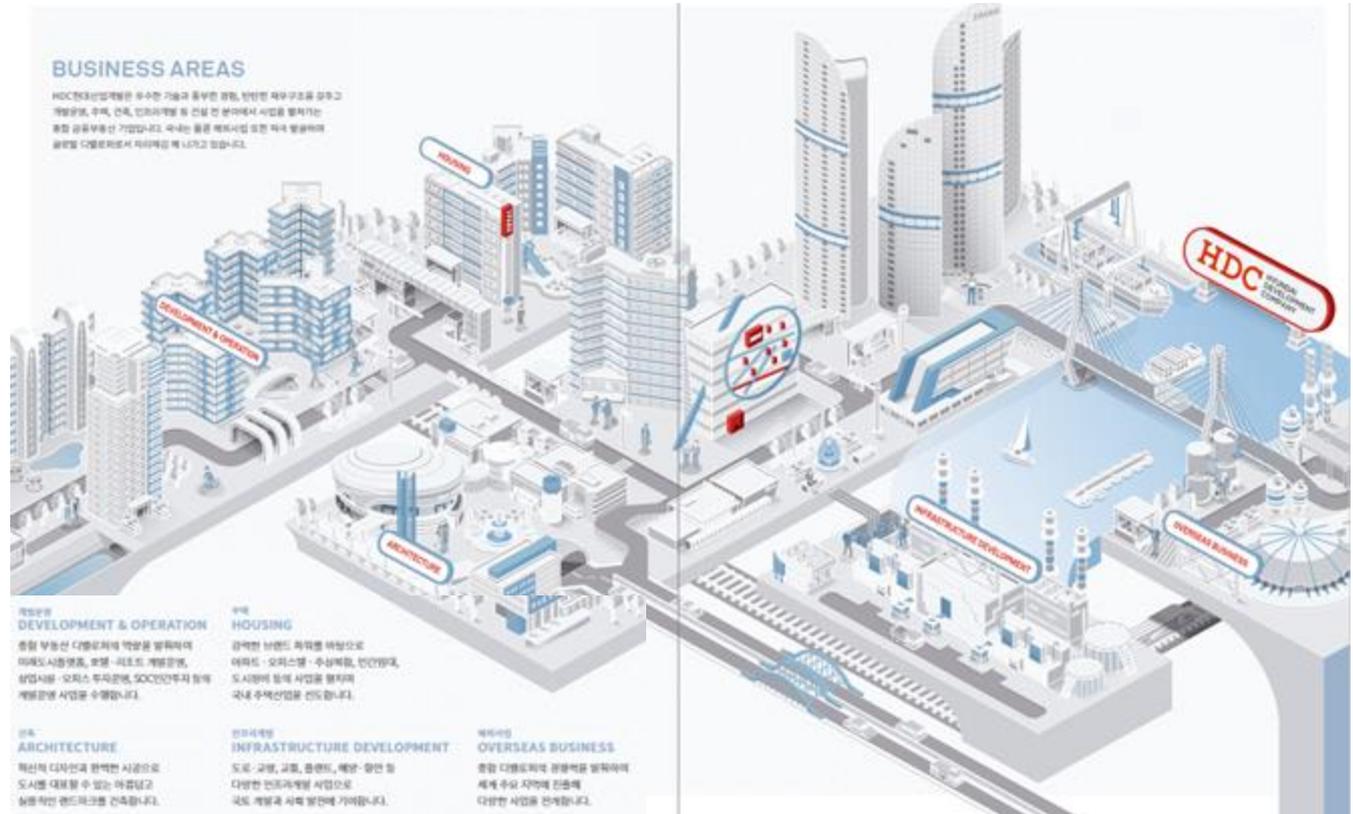
# Agenda

- > Introduction
- > Background & Purpose
- > Snowflake 사용 구조
- > 추진 프로젝트 개요
- > 추진 프로젝트 내용
- > 프로젝트 효과
- > 향후 계획



# Introduction - I PARK

- **도급순위 10위 종합부동산디벨로퍼**  
: 개발뿐만 아니라 운영까지.  
>>> 서울원(광운대)
- **Data 기반 전략 수립 NEEDS 증가**
  - 상품 전략
  - 운영 최적화
  - 건설 품질 향상
- **DX팀-디지털팀 중심 전사 혁신 전략 추진**

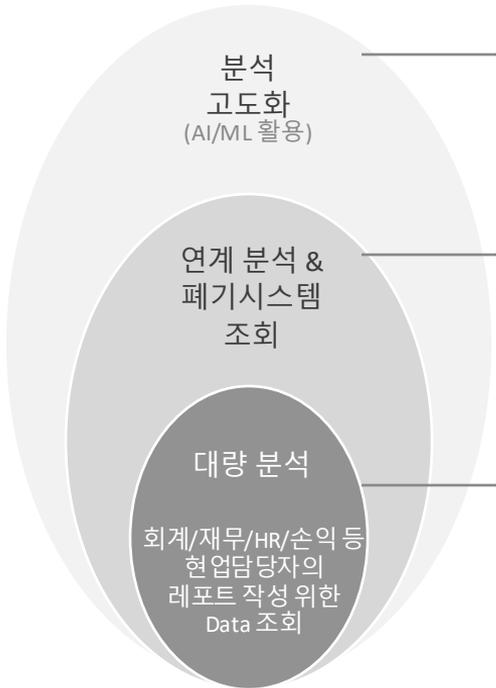


# Background & Purpose

주제

AS-IS 이슈

Snowflake를 사용하게된 이유



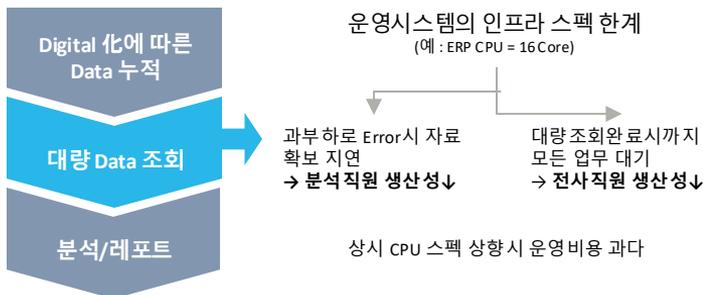
### ③ 장기적인 Data의 활용성 강화



### ② 데이터 접근성, 비용 이슈



### ① 자원 효율화 이슈



### ✓ 전사 분석 문화 확대 위한 인프라 최적화

- IT담당자들의 개입을 최소화하면서, 직원들이 신뢰성 있는 Data의 조합을 쉽게 얻을 수 있는 Data Sandbox 마련
- 다양한 분석 Needs에 대응할 수 있는 비용 효율적인 분석 체계 마련 (순수운 시간단위 스펙 조정 및 과금 체계)

### ✓ 필요 Data의 자동 업데이트 체계

- 분석 주제의 Dataset 구성 속도 향상

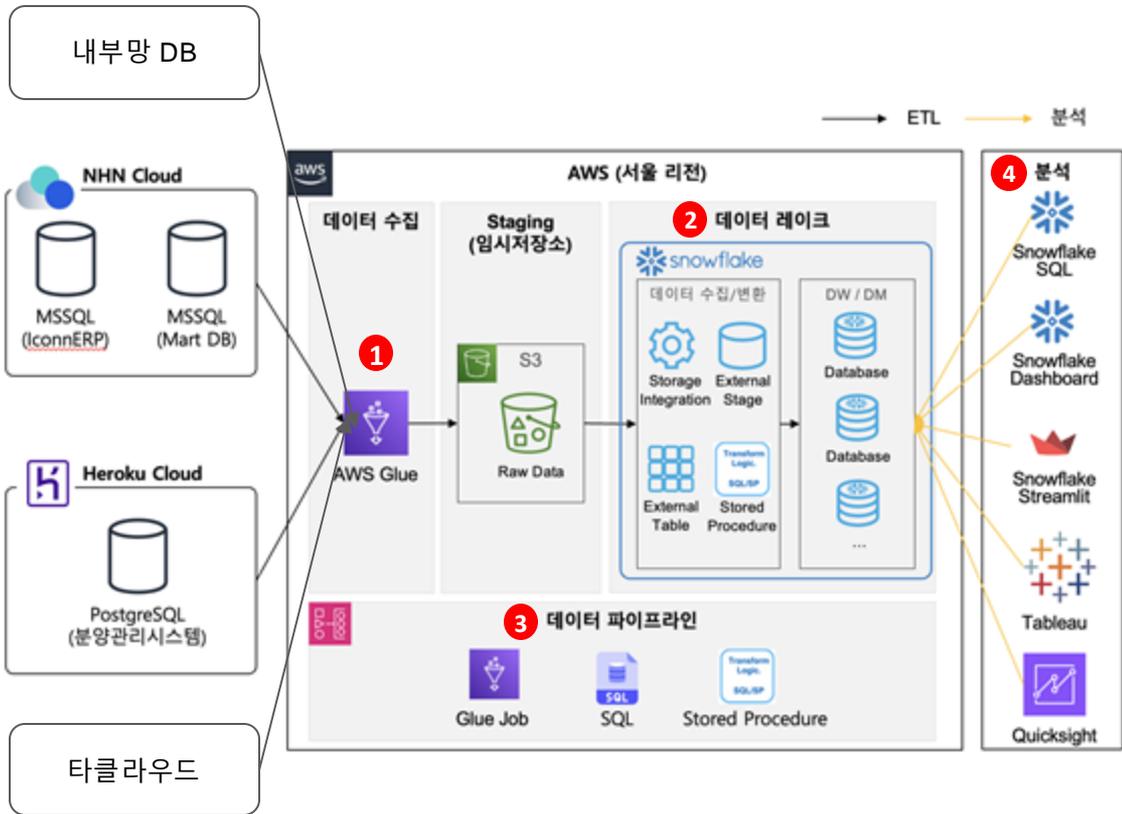
### ✓ 클라우드 환경에서 효과적인 시스템 라이프사이클 관리

- 폐기 시스템 Data 보관 및 조회 체계

### ✓ AI 도입을 위한 Data 정비

- Hallucination 현상 최소화를 위한 Data Lineage 관리 강화

# Snowflake 사용 구조



## 1. 데이터 수집

- 여러 클라우드 소스의 시스템의 데이터들을 Glue를 통해 S3로 수집 환경 구축

## 2. 데이터 레이크

- S3에 수집된 데이터를 Snowflake DW 영역에 적재
- DW 데이터를 기반으로 Mart 데이터 구성

## 3. 데이터 파이프라인

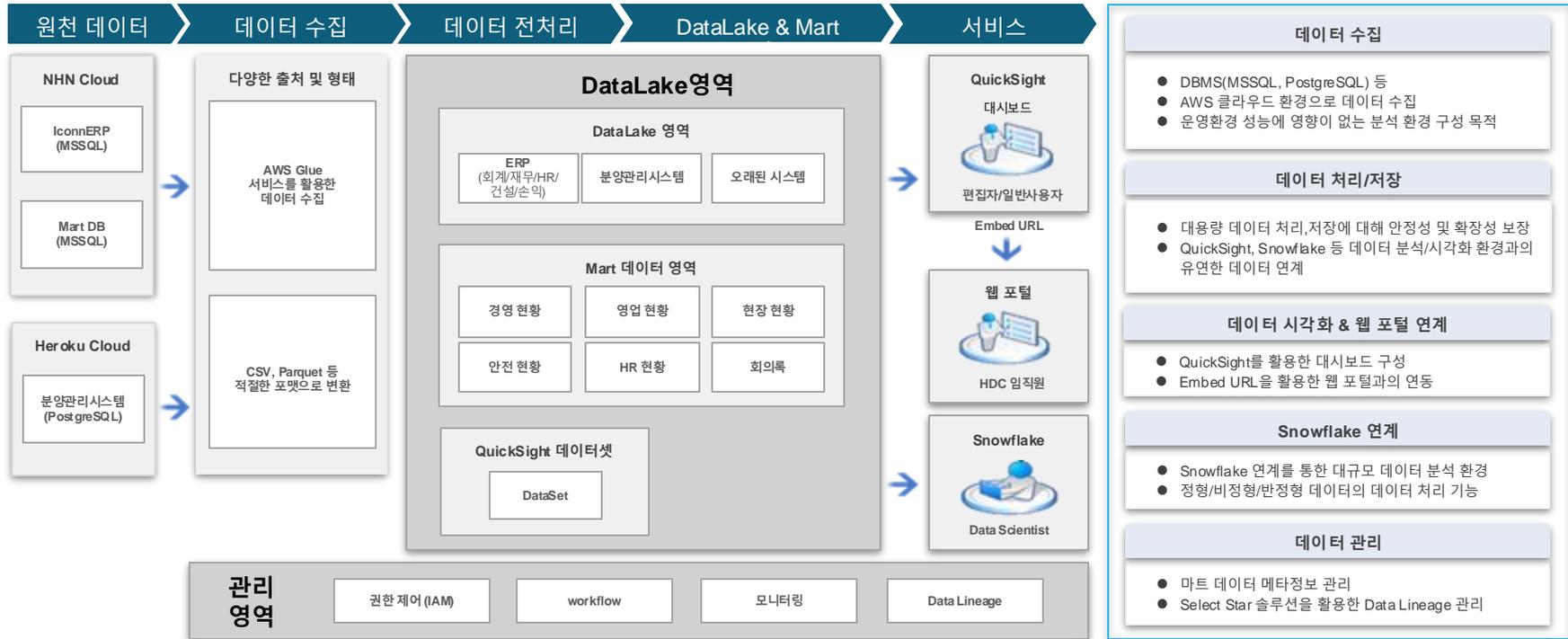
- 데이터 수집/저장을 위한 ETL 선/후 관계를 MWAA를 활용하여 Flow형태로 구성 및 스케줄링 적용

## 4. 데이터 분석

- 마트 데이터를 기반으로 SQL, Dashboard, Streamlit, Tableau를 통한 데이터 분석 환경 구축

# 프로젝트 개요 - 범위

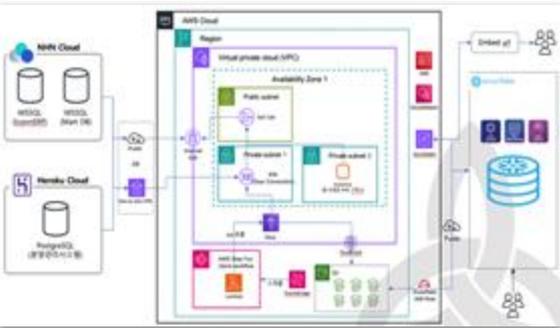
- 프로젝트 기간 : 2024.06월 ~ 2024.11월 (5개월)
- 파트너사 : 메가존 클라우드
- 프로젝트 수행 범위



# 프로젝트 개요 - 목표

**목표 01** 최적의 데이터 아키텍처 구성

AWS 기반 효율적인 데이터 파이프라인 아키텍처 구성



The diagram illustrates an AWS-based data lake architecture. It shows data being ingested from external sources like 'North Cloud' and 'Hemby Cloud' into 'S3 buckets' (S3-Standard and S3-Infrequent Access). The data then flows into an 'Amazon S3 Data Lake' within an 'AWS Cloud' region. This lake is connected to 'Amazon Athena' for querying and 'Amazon Redshift' for analytics. The architecture is designed for high availability and scalability across multiple availability zones.

- 안정적이고 비용 효율적인 DataLake 구성
- workflow 관리가 가능한 파이프라인 구성
- AI/ML 활용 고려, 유연하고 확장성 있는 DataLake 구축

**목표 02** 비용 효율적인 시각화 구현 및 연계

AWS QuickSight를 활용한 대시보드 구성 및 웹 포털 연계



The screenshot shows a dashboard from AWS QuickSight. It features several key performance indicators (KPIs) at the top, such as '매출/손익 현황' (Sales/Profit Status) with values like 18,174억 원 and 7.19%. Below the KPIs are several charts, including bar charts and line graphs, representing different data trends over time. The dashboard is part of a web portal that integrates with other systems.

- QuickSight를 활용하여 Tableau를 대체하는 대시보드 구성
- QuickSight를 비용 효과적으로 사용할 수 있는 체계
- Embed URL을 활용한 웹 포털과의 연동

**목표 03** Snowflake 데이터 분석 환경 연계

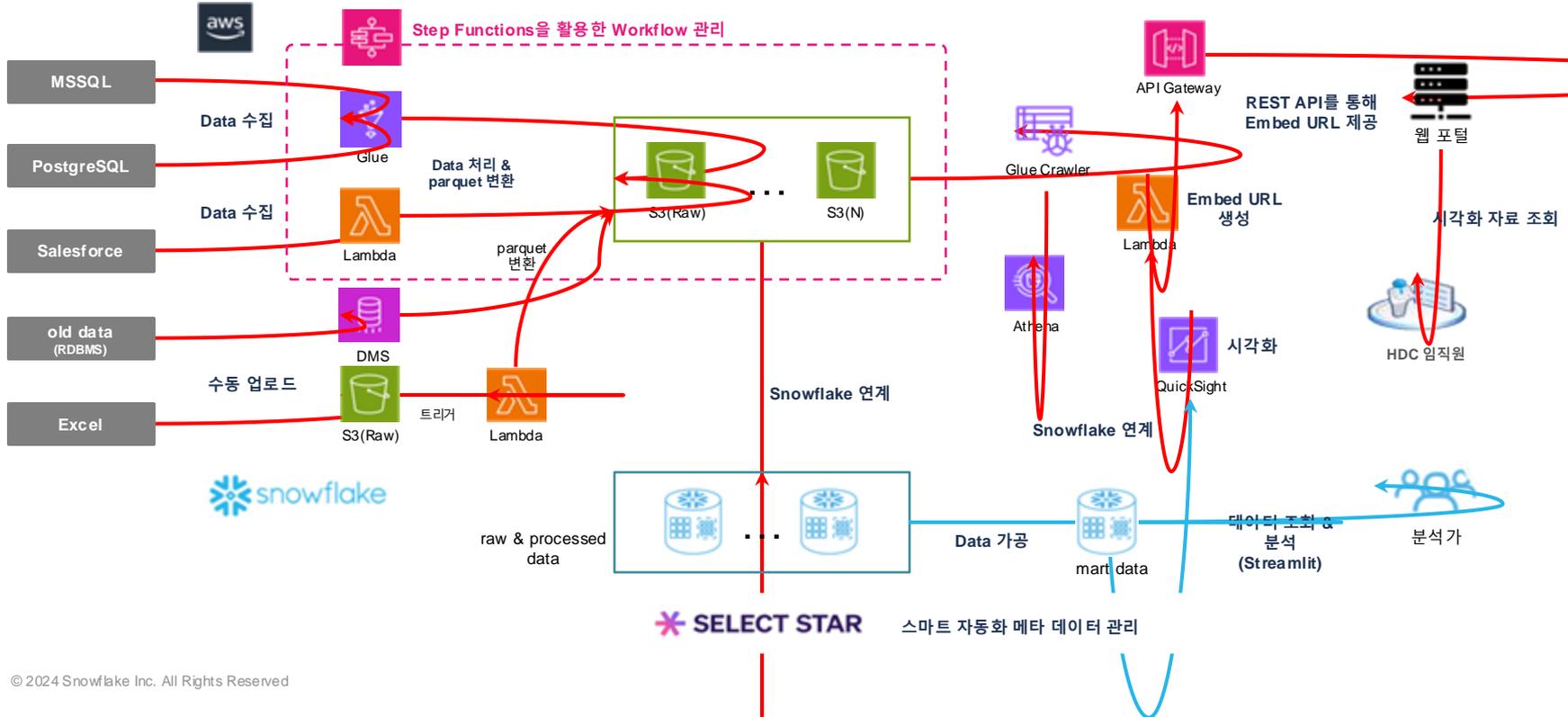
Snowflake 연계를 활용한 데이터 분석 환경 제공



The diagram shows a Snowflake data analysis environment. It is hosted within a 'VPC' (Virtual Private Cloud) and includes 'Authentication & access control' mechanisms. The environment is composed of several layers: 'Cloud services' (Infrastructure manager, Cluster, Metadata manager, Security), 'Query processing' (Virtual warehouse), and 'DataLake storage' (Virtual warehouse, Virtual metadata). This setup enables efficient data analysis and management.

- Snowflake Streamlit 등을 활용한 대량 데이터 분석
- Select Star 솔루션 테스트를 통한 Data Lineage 관리 체계 구현

# 프로젝트 내용 - 논리 아키텍처



# 프로젝트 내용 - Snowflake Data 검증

Snowflake에서 제공하는 웹 기반 SnowSight Worksheet를 통해 SQL기반의 데이터 분석 및 결과 다운로드 가능

The screenshot displays the Snowflake SnowSight Worksheet interface. On the left, there is an 'Object Browser' sidebar showing a tree view of databases and tables. The main area shows a SQL query with a CASE statement filtering by 'TY\_KI\_COLOR'. Below the query, a table of results is displayed with columns for '이월일여부', '이월일타', '이월일타\_2', '입원코드', '거래번호', '역분개번호', '회사이유', and '연도상계번호'. To the right, a bar chart visualizes the data, with a legend for '이월일타' (Yes/No) and '이월일타\_2' (Green/Red). A callout box points to the 'Filter' button in the results table, and another callout points to the chart, indicating that query results can be visualized as charts or graphs.

GUI 형태의 필터링 사용 가능

쿼리 조회 결과를 차트 또는 그래프로 시각화 가능 (Dashboard로 구현 가능)

이월일여부	이월일타	이월일타_2	입원코드	거래번호	역분개번호	회사이유	연도상계번호
55	Y	●	A101	20236000009861	null	null	2023600000058
56	Y	●	A101	20236000008045	null	null	202360000008C
57	Y	●	A101	20236000029128	null	null	202360000291
58	Y	●	A101	20236000034623	null	null	20236000034E
59	Y	●	A101	20236000061037	null	null	20236000061C
20	Y	●	A101	20236000071483	null	null	202360000714
21	Y	●	A101	20236000084005	null	null	20236000084C
22	Y	●	A101	20236000105082	null	null	20236000105C
23	Y	●	A101	20236000105843	null	null	20236000105E
24	Y	●	A101	20236000111932	null	null	20236000111E
2	N	Green	A101	정산완료	A101		202360000553
3	N	Green	A101	정산완료	A101		202360000553
4	N	Green	A101	정산완료	A101		202360000553
5	N	Green	A101	정산완료	A101		202360000553
6	N	Green	A101	정산완료	A101		202360000553
7	N	Green	A101	정산완료	A101		202360000553
8	N	Green	A101	정산완료	A101		202360000553
9	N	Green	A101	정산완료	A101		20236000055390
10	N	Green	A101	정산완료	A101		20236000055405

# 프로젝트 내용 - Snowflake 대량 조회

회계전표 마트 데이터를 활용하여 필요 검색조건으로 조회 및 다개년치 대량 데이터 다운로드 받을 수 있는 화면을 구성함

**전표 데이터**

계정코드: 전산대체(가수금) | 귀속부서: 재무팀 | 기간: 2024/01/23 - 2024/01/30

전표: 39 | 차변: 229,244,835 | 대변: 162,984,603 | 차대차액: -66,260,232

↑ 30

승인	전표번호	전기일자	응답일자
0	Y	20246000008962	20240126
1	N	20246000008972	20240126
2	Y	20246000008968	20240126
3	N	20246000008967	20240126
4	Y	20246000008965	20240126
5	Y	20246000008969	20240126
6	Y	20246000008556	20240125
7	N	20246000008720	20240125
8	N	20246000008557	20240125
9	N	20246000008720	20240125

Download

**개별전표 데이터 다운로드**

• 최대 5년 조회가 가능하고 다운로드만 가능합니다

범위코드: A101 | 귀속부서: 재무팀 | 검색 시작 날짜: 2019/10/11 | 검색 마지막 날짜: 2019/12/31 | 거래처명: AS

계정과목 - Level 1: AS | 계정과목 - Level 2: AS | 계정과목 - Level 3: AS | 계정과목 - Level 4: AS | 계정과목 - Level 5: AS

**검색 조건**

범위코드: A101  
 귀속부서: 재무팀  
 거래처명: AS  
 검색날짜: 2019-10-11 ~ 2019-12-31

조회

전체 행수: 352,534 | 차변합: 70,086,870.591,207  
 파일 이름: 전표조회\_2024-10-11\_15:07:34.csv

전표 데이터 다운로드

검색 조건 기반의 데이터 분석

검색 결과 합산 정보

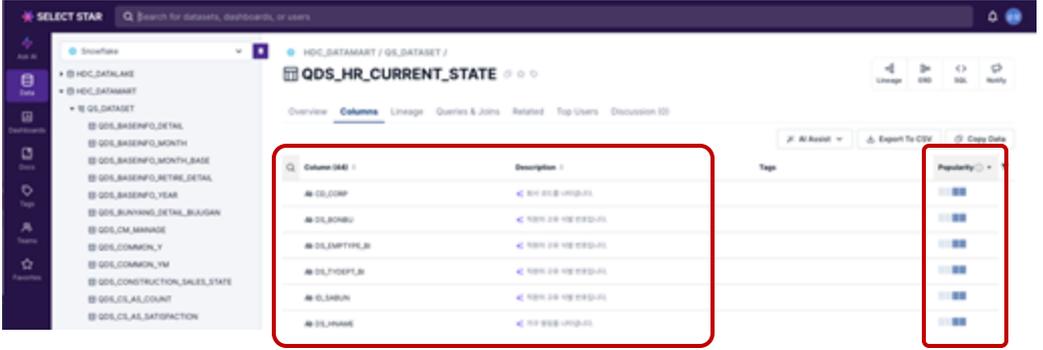
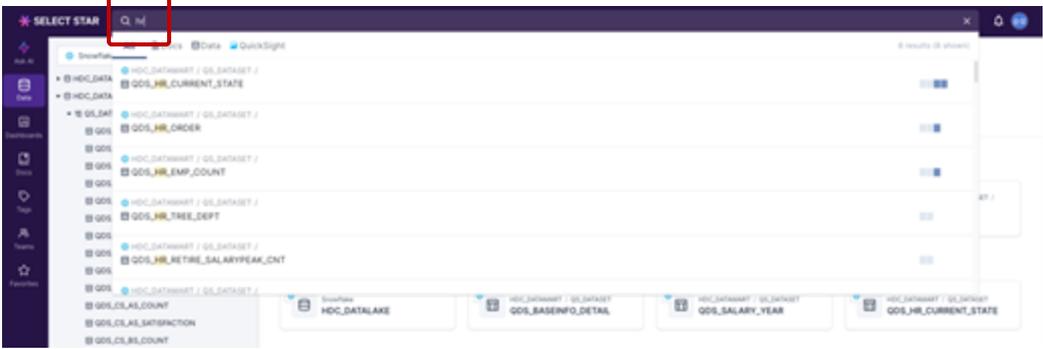
대량 데이터 다운로드

# 프로젝트 내용 - Select Star 개념

## 데이터 디스커버리 및 거버넌스 플랫폼

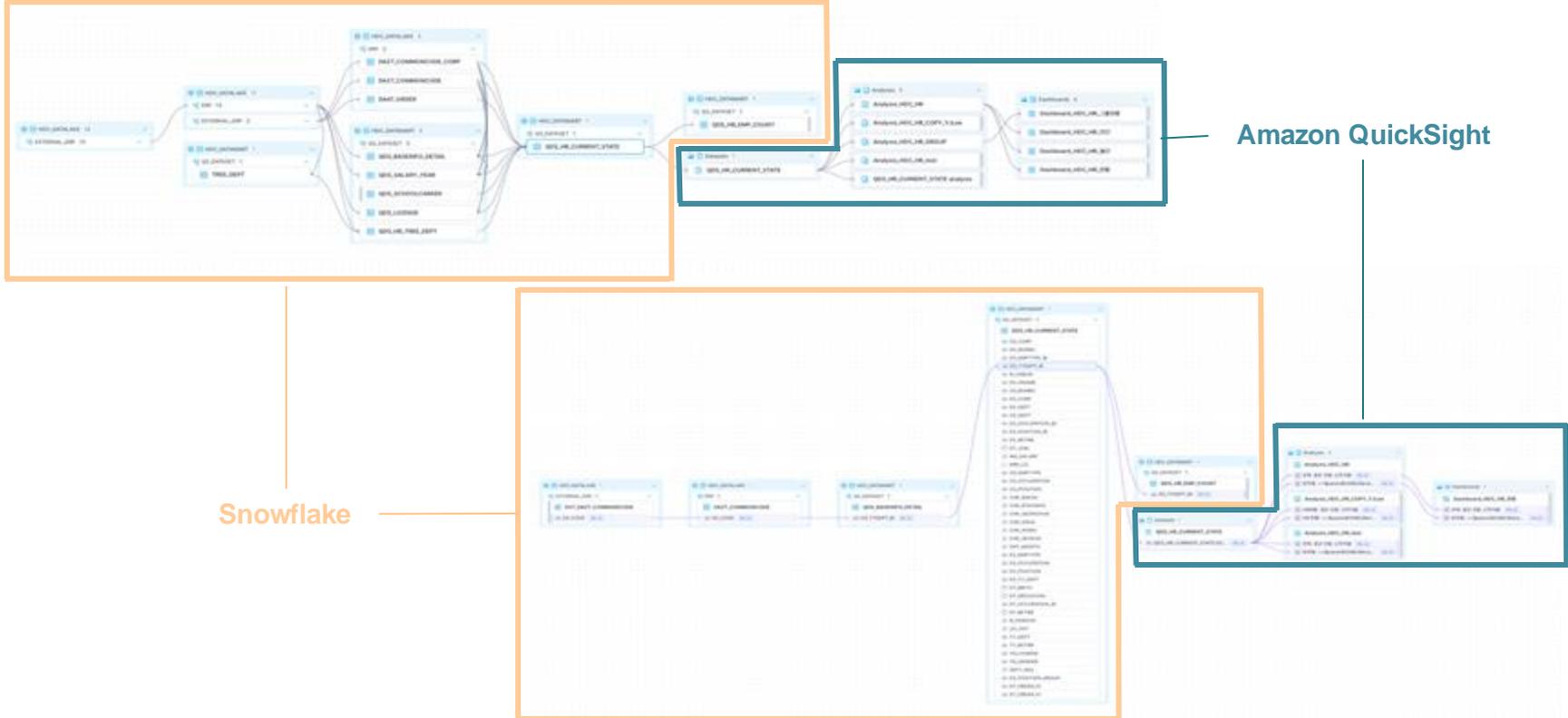


▶ 손쉬운 테이블, 컬럼 검색 / Description 관리 및 Popularity 확인



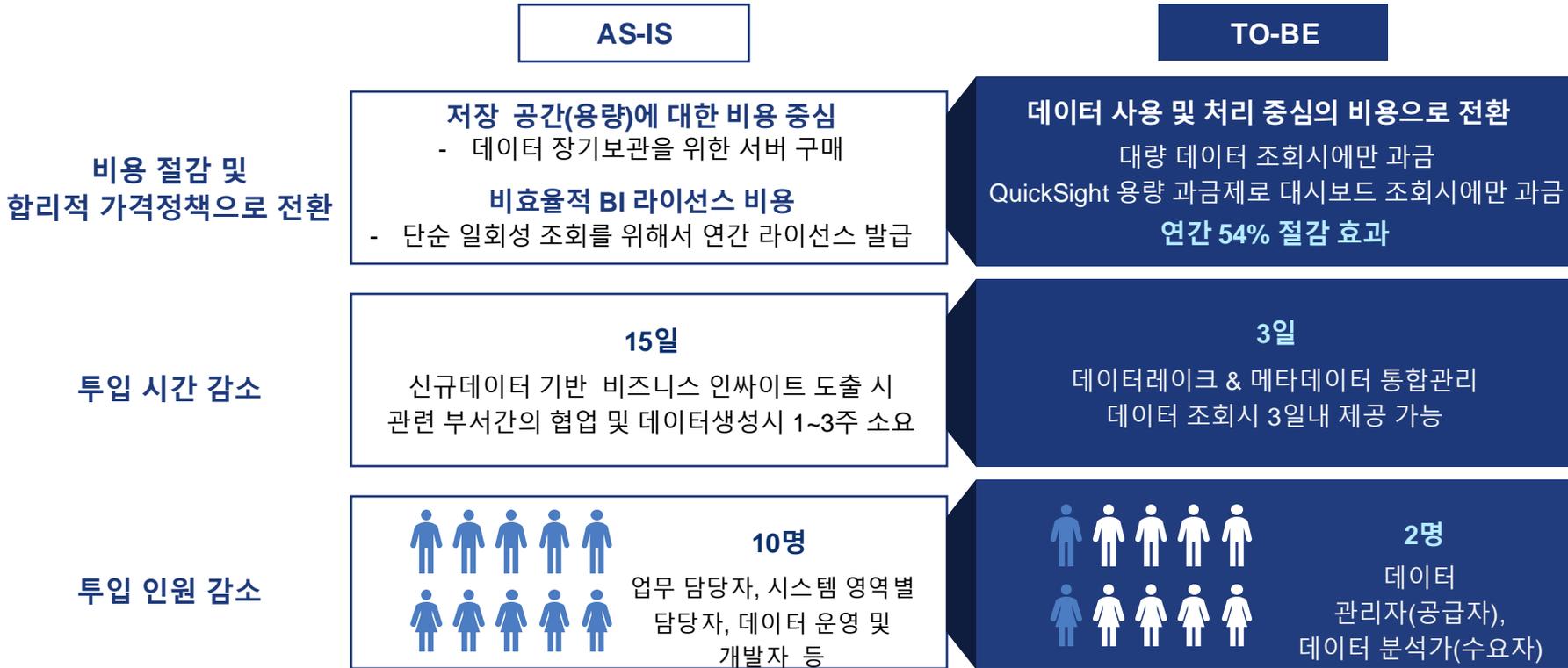
# 프로젝트 내용 - Select Star 테이블/컬럼 뷰

▶ 테이블, 컬럼 단위 All Up & Down Stream Lineage





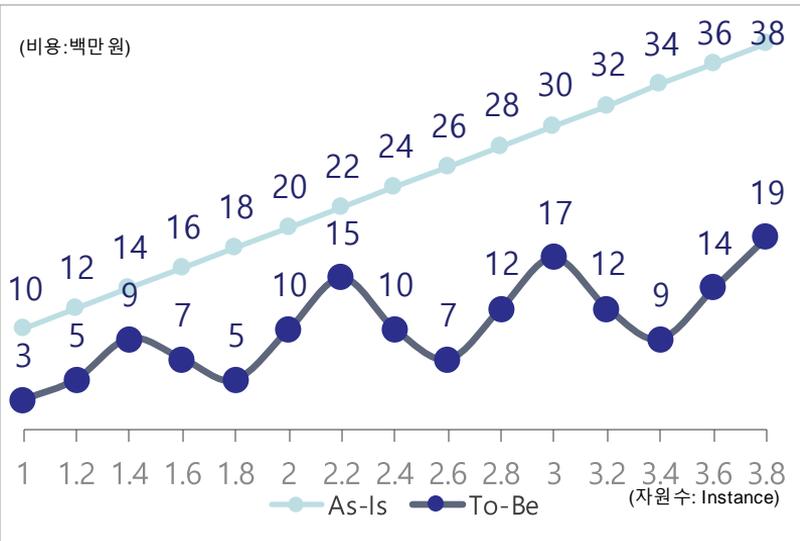
# 프로젝트 효과 1



# 프로젝트 효과 2

- 지속적인 데이터 증가에 따른 자원증설 비용을 사전에 대응, BI 시각화 라이선스 비용 최적화
- 단순히 인프라의 확장을 넘어서, 데이터 중심의 전사적인 역량 강화와 미래 지향적인 비즈니스 전략 수립에 기여

■ 데이터양에 따른 월별 운영비용 Gap.



■ 데이터 관리 핵심역량 5가지 수치화 Gap.



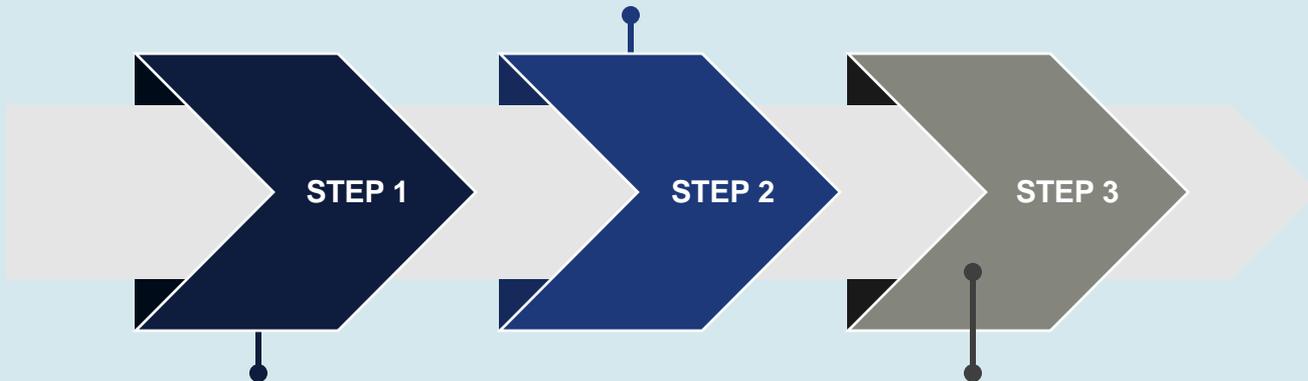
# 향후 계획

HDC현대산업개발의 데이터관리플랫폼 프로젝트는 3단계에 걸쳐 구축 및 발전시킬 예정

- 올해 2단계 일부 진행, 25년부터는 자동파이프라인 관리 범위 확대 및 3단계로의 발전 기반 마련

데이터 제공 범위 확대 및 데이터 연계 분석 강화

- 데이터 자동 파이프라인 관리 범위 확대
- snowflake 연계 데이터 분석 및 비용 효율적 BI 시각화 구현 체계 마련



데이터 기반 건설 DX  
업계 리더십 유지  
지속 가능 성장

대량 데이터 조회 업무 효율화

- 회계 및 재무, 구매, 감사자료 등 대규모 데이터 조회
- 데이터 수요자의 요구사항에 민첩한 서비스 제공

DataLake 기반 AI 활용 서비스 제공

- 현장 (IoT, 영상, 이미지 등), 비정형 데이터 (도면, 결재문서, 지침서 등) 데이터 추가 수집
- 데이터 학습 환경 구성 및 LLM 등 AI 서비스 구축

# 감사합니다.

