

기업 데이터 분석 혁신을 위한

Tarantula Lakehouse 소개

2025년 v2.0

UN°NET

© 2025 UNNET. 무단 복제를 금합니다. | All Rights Reserved.

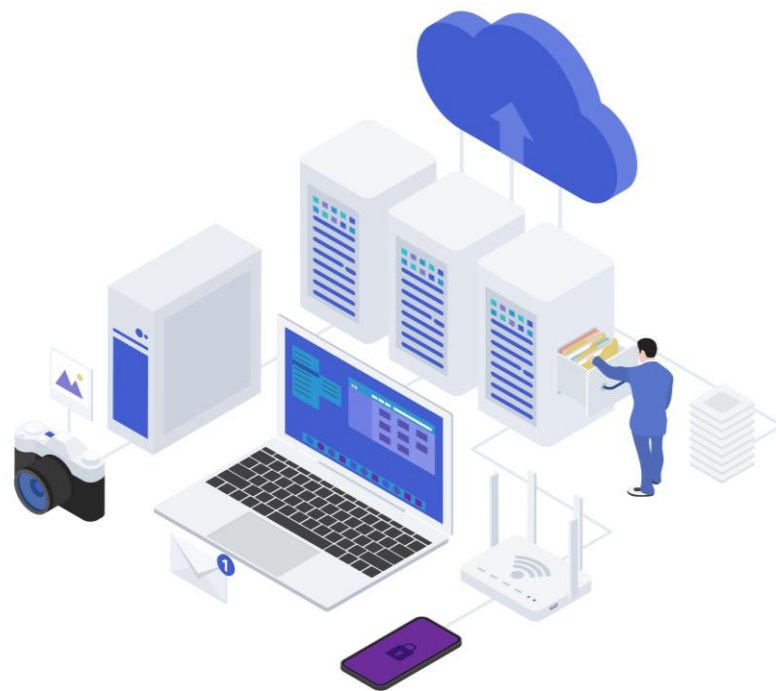


Table of Contents

1

유엔넷 소개

- 1.1 회사소개 및 연혁
- 1.2 조직도 및 인력현황
- 1.3 사업 영역
- 1.4 금융권 사업 경험
- 1.5 넥스트레이드 솔루션 도입사례

2

솔루션 소개

- 2.1 타란툴라 Data Platform 제품 포트폴리오
- 2.2 타란툴라DB 소개
- 2.3 타란툴라 Lakehouse
- 2.4 Minio AIStor



1. 유엔넷 소개

1.1 회사소개 및 연혁

2001년 설립 후 24년간 다양한 IT 사업 경험을 축적하여 금융 전문 종합 IT 기업으로 성장

Total IT Solution
Provider

오픈소스 소프트웨어 사업에 진출

- 2025 넥스트레이드 Lakehouse 구축
- 2025 쿠팡 국내최대 AI 기반 네트워크 구축
- 2024 넥스트레이드 및 4개 증권사 자동주문시스템(SOR) 타란툴라DB 구축
- 2024 PostgreSQL 기반 타란툴라DB 출시 및 GS 인증 취득
- 2024 기업은행 레드햇 통합유지보수 사업 수주
- 2023 우리은행 레드햇 통합유지보수 사업 수주
- 2022 레드햇 파트너 계약

"고객의 모든 요구에 응답하기 위해 오픈소스 소프트웨어
사업에 진출하며 비전을 제시하고 있습니다"

고객 문제 해결을 위한 제품 개발 및 신기술 적용

- 2020 유엔넷(주) 사명변경
- 2020 국가정보자원관리원 지능형클라우드 국내 최초 주니퍼 SDN 구축
- 2019 GS리테일 비용효율적 회선 이중화를 위한 IoT 기반 NetSavior 통신장비 개발 및 특허 취득
- 2019 JTBC 유무선 통합 네트워크 국내 최초 주니퍼 AI기반 MIST 구축

도전

"고객 성장을 위해 기꺼이
위험을 감수하며 행동하는 파트너"

금융 분야 전문 기업으로 도약

- 2018 우리카드 차세대 네트워크 구축
- 2017 메리츠화재 차세대 네트워크 구축
- 2016 우리은행 차세대 네트워크 구축

기술

"고난이도 사업 성공으로 기업
역량을 증명하다"

신뢰

공공/대학/병원 고객과 길게 이어갈 인연 형성

- 2008 기업부설연구소 인가, 경영혁신형 중소기업 선정
- 2008 Juniper Networks Elite 파트너 계약
- 2007 Dacorder 출시 및 전자신문 상반기 인기상품 선정
- 2001 (주)유엔넷시스템즈 설립 및 Cisco 파트너 계약

"우리의 존재 가치를 고객의 신뢰에 두다"

회사명	유엔넷 주식회사
대표이사	김주섭
본사소재지	서울시 강남구 역삼동
전화번호	02-565-7034
임직원수	85명
설립연도	2001년 7월 30일
매출액	208억 (2024년)
신용등급	BB+ (유효기간 2026년 4월 6일)
홈페이지	www.unnet.co.kr

1. 유엔넷 소개

1.2 조직도 및 인력현황

4개 본부 하에 연구개발 자회사를 포함하여 전체 **110명**의 임직원으로 구성

■ 조직도 및 위치

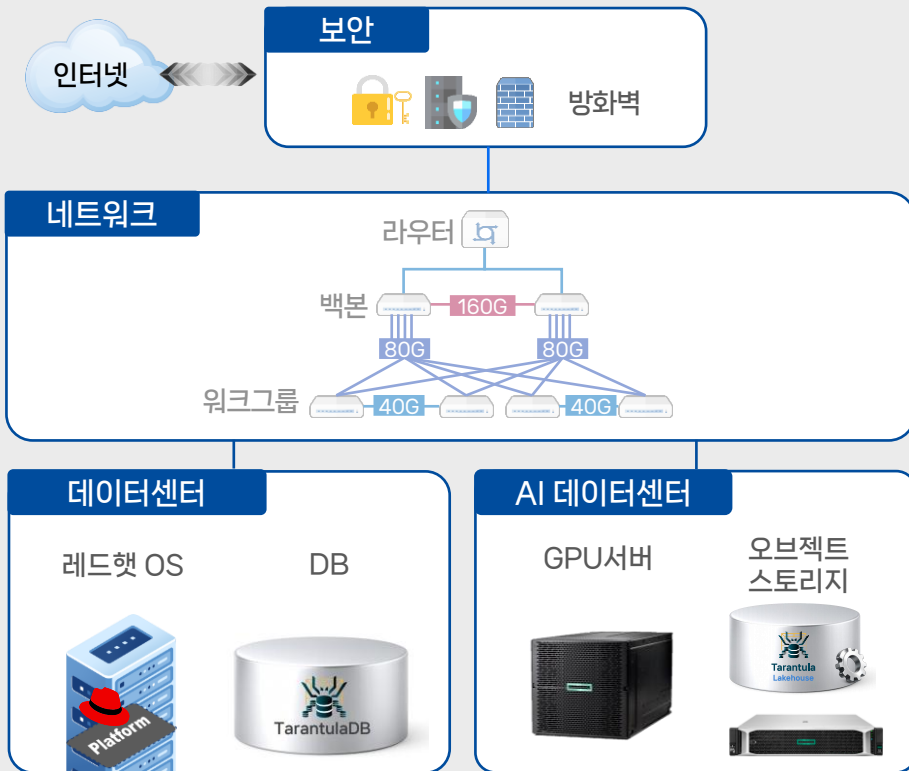


1. 유엔넷 소개

1.3 사업 영역

클라우드 / 데이터센터의 "IT Infrastructure"에 대한 토탈 솔루션 제공

기업의 정보통신 시스템 영역



1. 자사 솔루션

오픈소스 1등 PostgreSQL기반 기업용 DBMS
처리속도 1등 오픈소스 데이터 레이크 하우스 솔루션



2. SW 솔루션

오픈소스 점유율 1등 RedHat OS 및 가상화 플랫폼
오픈소스 기반의 오브젝트 스토리지 솔루션



3. 인프라 솔루션

세계 시장 점유율 1등 네트워크 보안솔루션
유/무선 네트워크, 데이터센터/SDN 솔루션



4. AI 솔루션

GPU서버 및 RoCE 네트워크 솔루션
오픈소스 기반의 온프레임 데이터 레이크 하우스 솔루션



1. 유엔넷 소개

1.4 금융권 사업 경험

넥스트레이드, KOSCOM, 미래에셋증권 등 다수 증권사에 DBMS 구축사업을 성공적으로 수행

■ 주요 금융 사업 실적

증권사 DBMS 구축



오픈소스 기반 DBMS 도입 전체 사업 수행

- 매매체결시스템 DB
- 홈페이지 시스템 DB
- 회원사 시스템 DB
- SOR(Smart Order Routing) 시스템 DB
- 주요 내부 업무 DB



Tarantula Lakehouse

25년 10월 도입



SOR 시스템 DB



SOR 시스템 DB



SOR 시스템 DB



SOR 시스템 DB



SOR, CRM 시스템 DB



자산배분시스템 DB



정보전달체계(UMS) DB



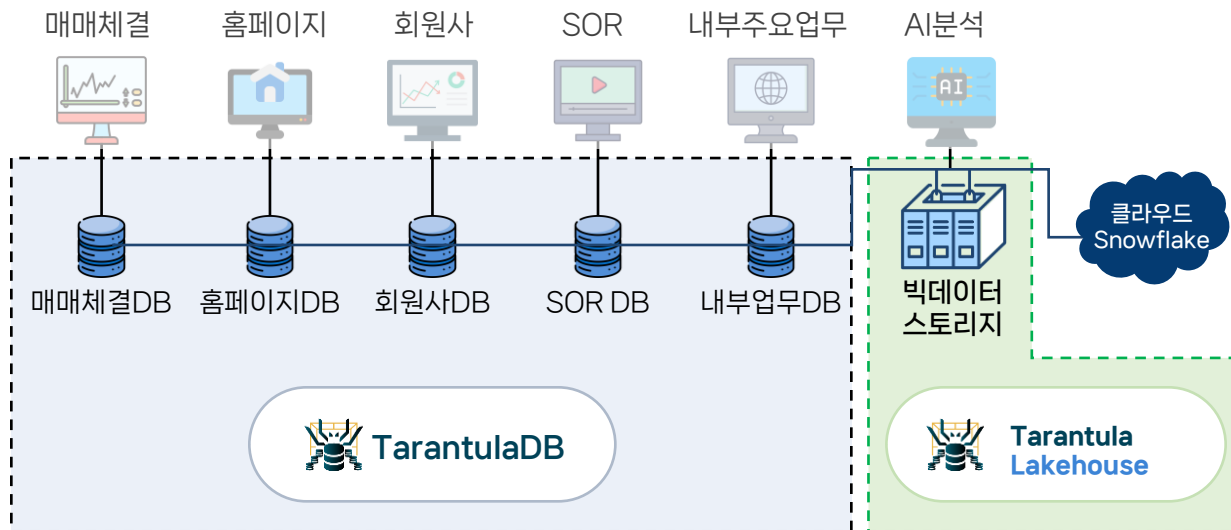
자금세탁방지(AML) DB

1. 유엔넷 소개

1.5 넥스트레이드 솔루션 도입사례 (DB 및 Lakehouse)



타란툴라DB가 표준DB로 선정되어 전체 시스템의 DB로 도입



넥스트레이드 DB 특징

- ☑ 일평균 10조원 대규모 거래처리 (25년)
- ☑ 추가, 등락률, 거래량 등 실시간 데이터 처리
- ☑ 금융 IT 에서 최고 수준의 신뢰성 요구
- ☑ 25년 3월 오픈 후 무중단 안정적 운영중

✓ 타란툴라DB를 선택한 이유

1. 상용 대비 약 90% TCO 절감
2. PoC를 통한 성능과 기술력 검증
3. 타사 대비 뛰어난 HA 이중화

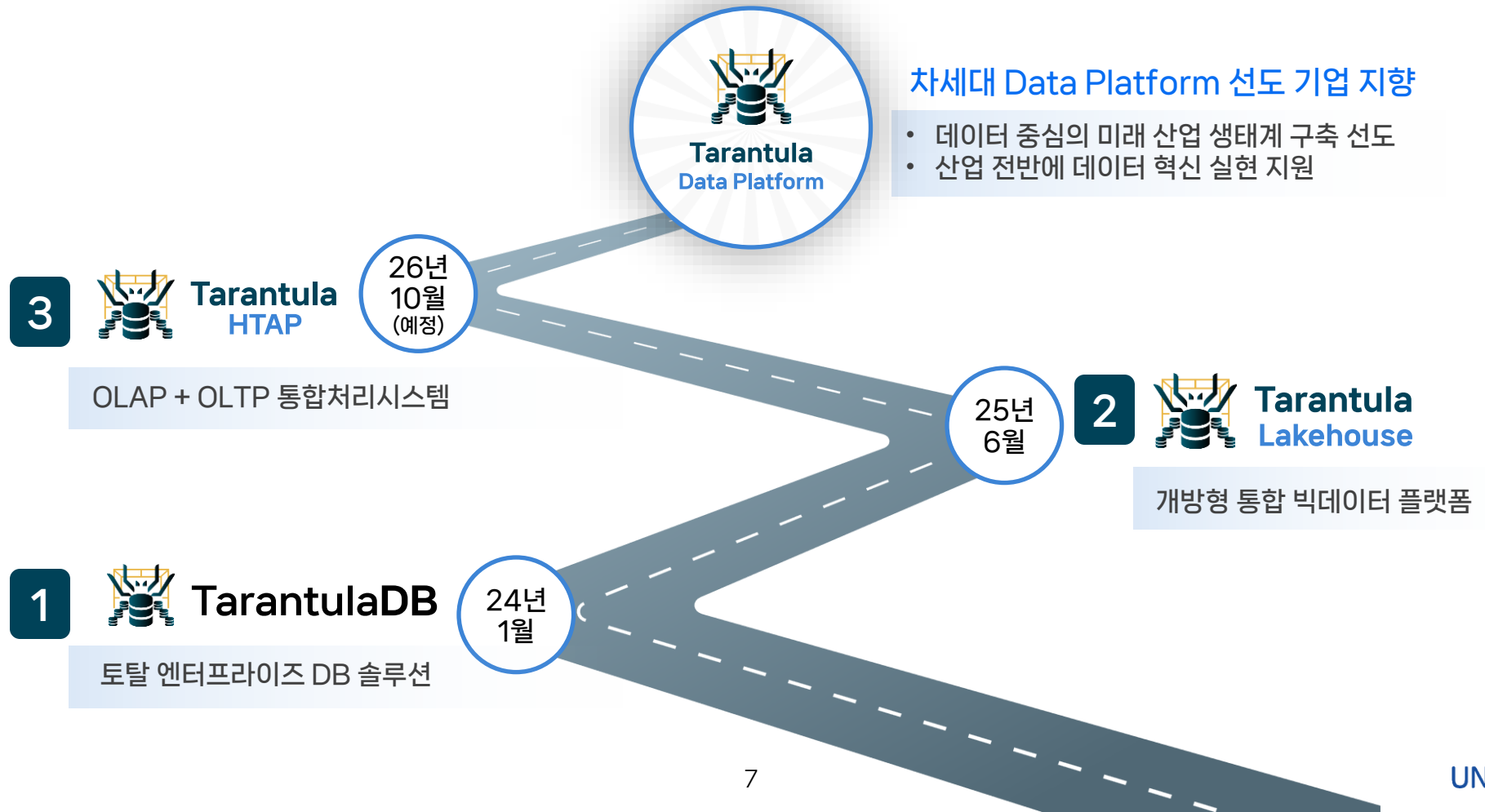
✓ 타란툴라 Lakehouse를 선택한 이유

1. AI, BI, Snowflake 모두 연동되는 국내 최초 솔루션
2. 민감한 데이터는 On-prem 저장
3. 오라클, Snowflake 등 글로벌 기업과 PoC 통해 안정성/성능 검증

2. 솔루션 소개

2.1 타란툴라 Data Platform 포트폴리오

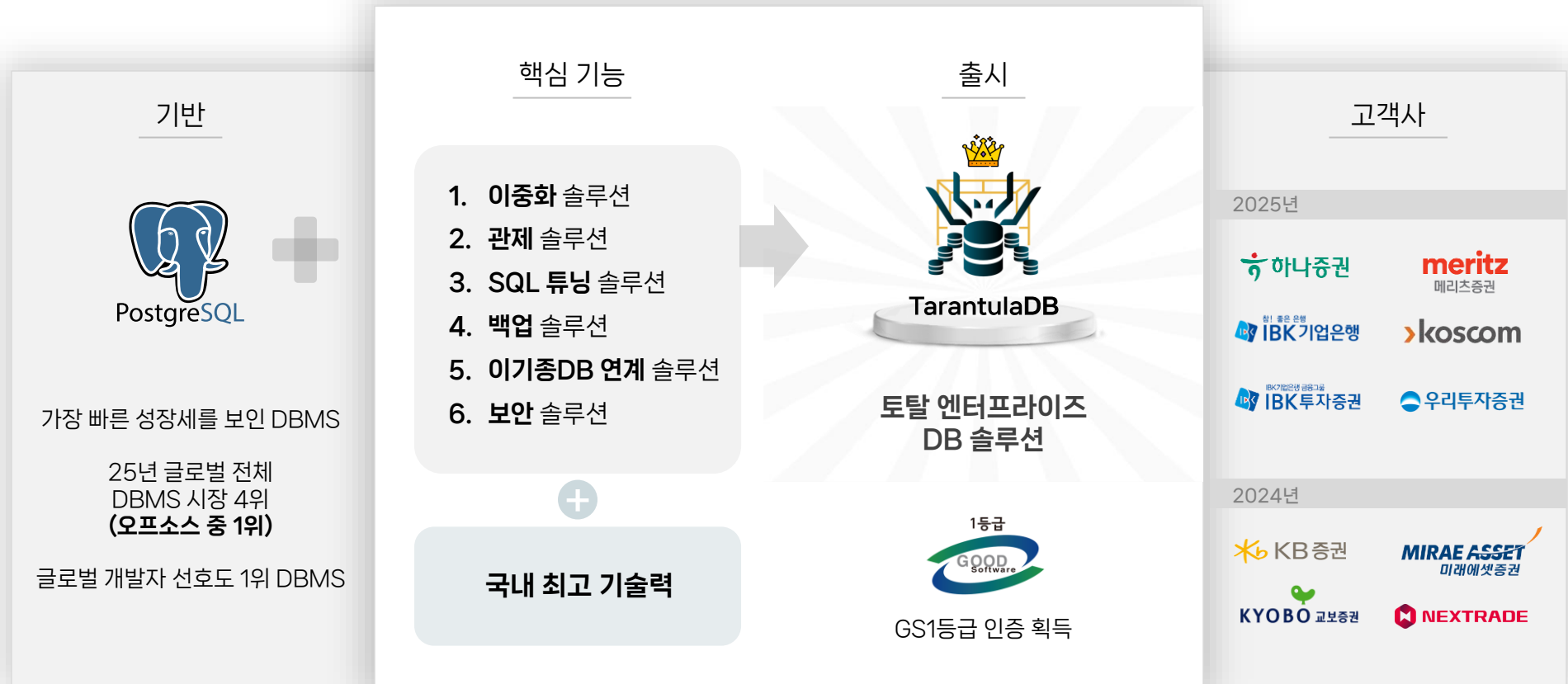
타란툴라DB 및 Lakehouse 제품이 있으며, 향후 개발예정인 HTAP까지 포함한 데이터플랫폼 솔루션을 보유



2. 솔루션 소개

2.2 타란툴라DB 소개

PostgreSQL을 기반으로 기업에서 요구하는 핵심 기능과 국내 최고 기술 지원을 제공하는 '토탈 엔터프라이즈 DB 솔루션'



2. 솔루션 소개 ▶ 2.3 타란툴라 Lakehouse

2.3.1 타란툴라 Lakehouse 소개

데이터 레이크의 민첩성과 비용효율성, 데이터 웨어하우스의 강력한 분석 기능을 결합한 **통합 빅데이터 플랫폼**
AI 시대 최적의 데이터 저장소로, BI 연동 뿐 아니라 snowflake와 AI/ML 연동까지 자유로운 **국내 유일 솔루션**

※ 넥스트레이드와 3주간 PoC를 통해 안정성과 성능을 입증

On-premise



2.3.2 Lakehouse 개념

최근 AI 데이터 아키텍처에서 가장 각광받는 개념으로 데이터 레이크의 민첩성과 비용효율성, 데이터 웨어하우스의 강력한 분석 기능을 결합한 데이터 저장/처리 기술



Key Message

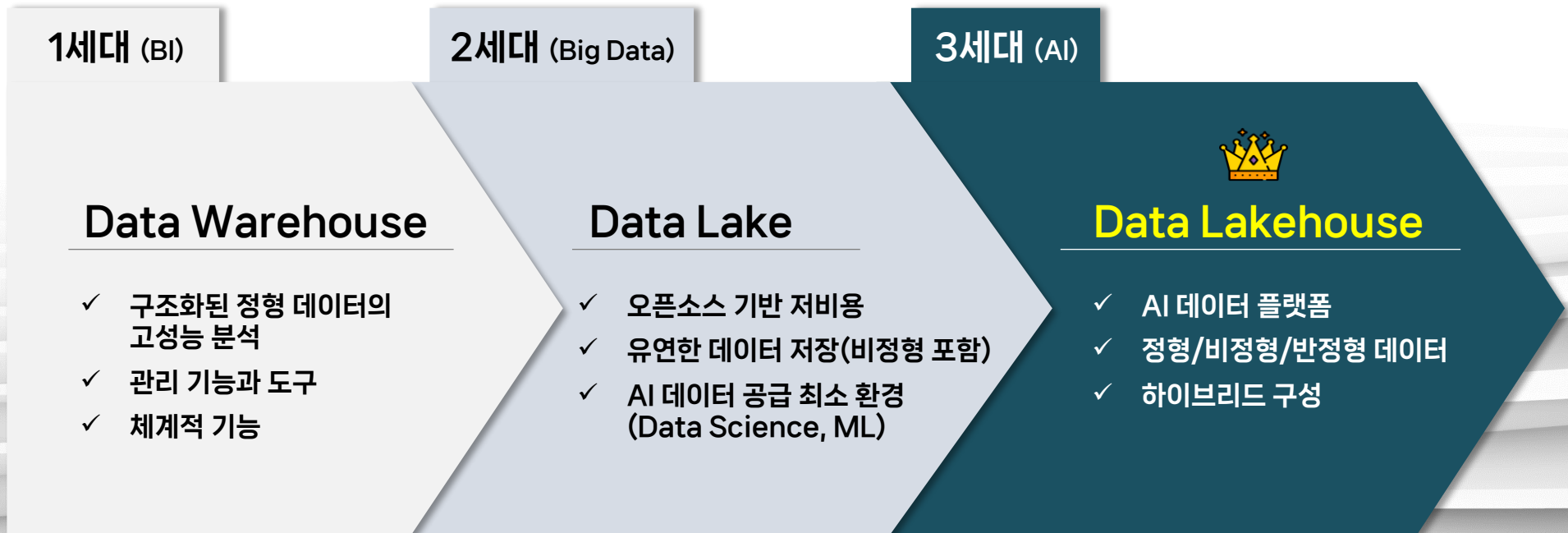
Lakehouse = Data Lake + Data Warehouse의 합성어

데이터 레이크의 유연함 + 데이터 웨어하우스의 관리·분석 기능을 합친 아키텍처

	Data Lake	Data Warehouse
1. 데이터 저장	<ul style="list-style-type: none"> 조직의 모든 데이터를 원시적이고 비정형적인 형태로 담고 있으며, 즉시 또는 나중에 사용할 수 있도록 무기한 저장 	<ul style="list-style-type: none"> 사전 정의된 비즈니스 요구 사항에 따라 전략적 분석을 위해 정리 및 처리된 구조화된 데이터를 포함
2. 사용자	<ul style="list-style-type: none"> 구조화되지 않은 대량 데이터 포함 원시 형태의 데이터를 연구하여 새로운 고유한 비즈니스 통찰력을 얻고자 하는 데이터 과학자와 엔지니어가 사용 	<ul style="list-style-type: none"> 비즈니스 KPI에서 통찰력을 위해 관리자와 최종 사용자가 액세스 데이터는 분석을 위해 미리 결정된 질문에 대한 답을 제공하도록 이미 구조화
3. 분석	<ul style="list-style-type: none"> 예측 분석, 머신 러닝, 데이터 시각화, BI, 빅데이터 분석. 	<ul style="list-style-type: none"> 데이터 시각화, BI, 데이터 분석.
4. 스키마	<ul style="list-style-type: none"> 스키마는 데이터가 데이터웨어하우스가 아닌 데이터 레이크에 저장된 후에 정의되므로 데이터를 캡처하고 저장하는 프로세스 가속화 	<ul style="list-style-type: none"> 데이터가 저장되기 전에 스키마가 정의됨 데이터 처리 시간이 길어지지만, 일단 완료되면 조직 전체에서 일관되고 안정적으로 사용할 수 있는 데이터가 준비
5. 처리	<ul style="list-style-type: none"> ELT(추출, 로드, 변환). 이 과정에서 데이터는 소스에서 추출되어 데이터 레이크에 저장되고 필요할 때만 구조화 	<ul style="list-style-type: none"> ETL(추출, 변환, 로드). 이 프로세스에서는 데이터를 소스에서 추출하고 정리한 후, 비즈니스 분석에 적합하도록 구조화
6. 비용	<ul style="list-style-type: none"> 데이터웨어하우스에 비해 저장 비용이 상당히 저렴 데이터 레이크는 관리 시간도 짧아 운영 비용도 절감 	<ul style="list-style-type: none"> 데이터 레이크보다 비용이 많이 들고, 관리하는 데도 시간이 더 오래 걸리므로 운영 비용이 추가로 발생

2.3.3 데이터 저장소의 진화 – 개요

전통적인 1세대 Data Warehouse에서 빅데이터를 위한 2세대 Data Lake로 발전해 왔으며, 현재는 이 두 기술의 장점을 통합한 3세대 Data Lakehouse가 등장해, 분석과 AI를 아우르는 새로운 패러다임을 제시



2. 솔루션 소개 ▶ 2.3 타란툴라 Lakehouse

2.3.4 데이터 저장소의 진화 – Details

구분		1세대 (BI 시대)	2세대 (빅데이터 시대)	3세대 (AI 시대)
저장소 타입		Data Warehouse (DW)	Data Lake	Lakehouse
Time		2000년 부터 지금까지 Traditional (Legacy)	2013 – 2016년 정부 빅데이터 활성화	2023부터 시작 Gen AI 본격 확산
사용 기술		ETL, Data Mart, BI	Hadoop 기반	Iceberg / S3 / Minio 기반
특징		상용 벤더 기술 중심	오픈소스 중심	오픈소스 중심
주요 제품		오라클 Exadata, IBM DB2 Teradata, HP Vertica	Cloudera	Tarantula Lakehouse (국내최초) Snowflake, Databricks
특징	장점	1. 구조화된 정형 Data만 저장 2. BI 및 정형 분석에 최적화된 성능	1. 낮은 비용과 높은 확장성(오픈소스) 2. 비정형 Data처리와 다양한 소스 지원	<ol style="list-style-type: none"> DW + Data Lake 장점 결합 정형 + 비정형 Data 모두 처리 성능과 비용효율성을 모두 확보 성능향상 (10배 향상) 비용개선 (80% 절감) 다양한 데이터 유형 처리(영상, 이메일, 로그, IoT 센서데이터 등) AI 연동 최적화 (NVIDIA GPU 연동) 강력한 거버넌스, 보안 기능
	단점	1. 높은 비용과 비정형 Data 처리 한계 2. Data증가 시 성능이 느려짐 3. 확장성과 운영성에 제약 존재	1. 느린 쿼리 성능 2. 파티션/스키마 운영에 대한 노력 높음 3. 데이터 품질, 정합성 이슈 (비정상 종료 시) 4. 신기술에 따른 기술 전문인력 필요	

2.3.5 Lakehouse의 포지션

AS-IS

용도에 따라 별개의 시스템 사용
→ 막대한 도입비용과 유지비용

데이터 웨어하우스(DW)

- 오라클 Exadata
- Teradata 등

한계

- ① 증설/교체 고비용
- ② 대량 처리시 성능 ↓
- ③ 비정형 처리 제약

데이터 레이크 (DL)

- 클라우데라 / 하둡
- Amazon S3 등

- ① 저렴하지만 속도 느림
- ② 스키마, 파티션 관리 등 운영 부담
- ③ 데이터정합성 이슈 (비정상 종료시)

TO-BE

DW와 DL을 통합한 '3세대' 데이터 플랫폼

→ Lakehouse를 통해 TCO 절감, AI 기반 확대

타란툴라 Lakehouse

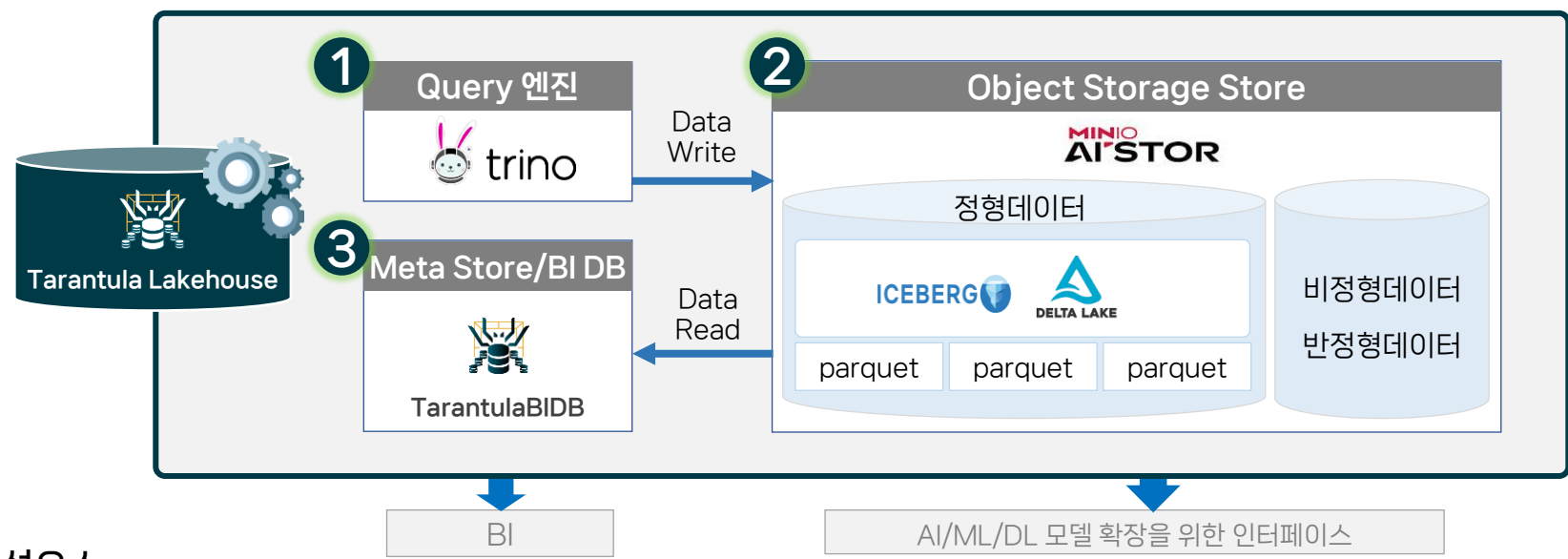


기존 한계 극복

- 1 상용 대비 90% 비용 절감
- 2 대량 처리시 탁월한 성능 ↑
- 3 정형 + 비정형 모두 처리
- 4 메타데이터 자동 관리로 운영 부담 감소
- 5 데이터 정합성 보장
- 6 AI/ML 데이터 연동 용이

2.3.6 솔루션 구성요소

3개의 핵심 컴포넌트로 구성되며 각각 업계 최고의 오픈소스 기반으로 구성되어 개방성, 확장성, 기능, 성능, 안정성 확보



구성요소

핵심 컴포넌트	솔루션명	역할 및 기능
1. Query 엔진	trino	<ul style="list-style-type: none">다양한 데이터 소스를 대상으로 SQL을 사용해 빠르게 분석하는 분산형 *MPP SQL 쿼리 엔진* MPP, Massively Parallel Processing: 대규모 병렬 처리
2. Object Storage Store	MinIO AIStor	<ul style="list-style-type: none">AI 및 대규모 데이터 워크로드를 위해 설계된 온프레미스 고성능 오브젝트 스토리지 플랫폼으로, 다양한 AI 기능을 통합하여 단일 인프라에서 효율적인 데이터 관리와 분석 제공
3. Meta Store/BI DB	TarantulaBIDB	<ul style="list-style-type: none">ICEBERG 카탈로그 저장 Meta DBICEBERG 테이블 조회 용도(Trino Bridge 자체 개발) 및 향후 Vector DB Extension

2. 솔루션 소개 ▶ 2.3 타란툴라 Lakehouse

2.3.7 솔루션 특징점 – 1) 실제 PoC 결과 우위포인트

1. 뛰어난 쿼리 성능

수억~수십억건의 쿼리를 안정적으로 처리

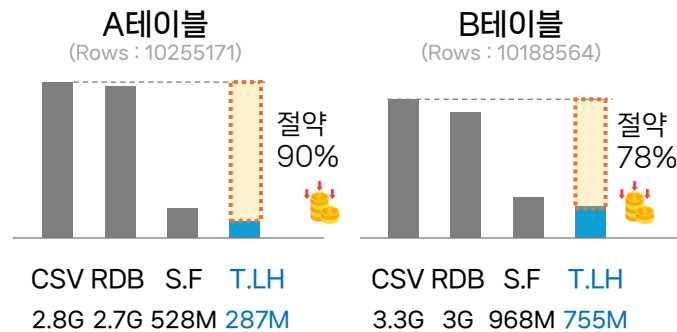
단위: 초

쿼리 부하 유형	A사	B사	타란툴라 Lakehouse
A 하루 집계 (천만)	8(*23)	0.7(*3)	0.4 ^{1등}
A 한달 집계 (수억)	13	1	0.8 ^{1등}
A 일년 집계 (수십억)	26	2	4 ^{2등}
A,B 조인 하루 집계	12	1.2	1 ^{1등}
A,B 조인 한달 집계	29	3.3	2.5 ^{1등}
A,B 조인 일년 집계	52	15	19 ^{2등}

✓ 조인은 DW 방식을 사용 * 데이터 캐싱 전 측정치

2. 월등한 저장공간 효율성

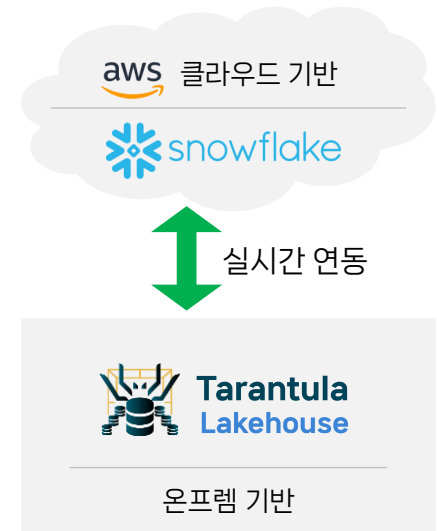
같은 데이터라도 가장 적은 공간 사용



CSV: 값을 쉼표로 구분한 오픈포맷 RDB: 관계형 데이터
S.F: Snowflake T.LH: 타란툴라 Lakehouse

3. Snowflake 완벽 연동

실시간 연동성 검증 완료





2.3.7 솔루션 특징점 – 2) 하나로 끝내는 데이터 플랫폼

통합된 쿼리엔진, 고성능 스토리지 등 빅데이터 시스템에 필요한 모든 구성요소를 모두 포함하고 있어 BI(경영분석) 뿐 아니라 향후 인공지능 데이터 워크플로우까지 확장 가능

■ 제안 솔루션 특징점

하나로 끝내는 데이터 플랫폼



- **통합 빅데이터 플랫폼**

Tarantula Lakehouse는 빅데이터 시스템을 구성요소로 포함하고 있어 별도로 빅데이터 시스템 구축 불필요



- **뛰어난 성능**

노드 수 증가에 따라 성능이 비례적으로 증가하는 수평 확장성 제공하며, 고성능 Object Storage 성능 자랑 (벤치마크 테스트 성능 보유)

폭넓은 개방성 제공



- **개방성과 유연성**

AWS, Snowflake 등 클라우드 서비스와 실시간 연동할 수 있으며, 기업 내부의 AI/ML 연동이 용이

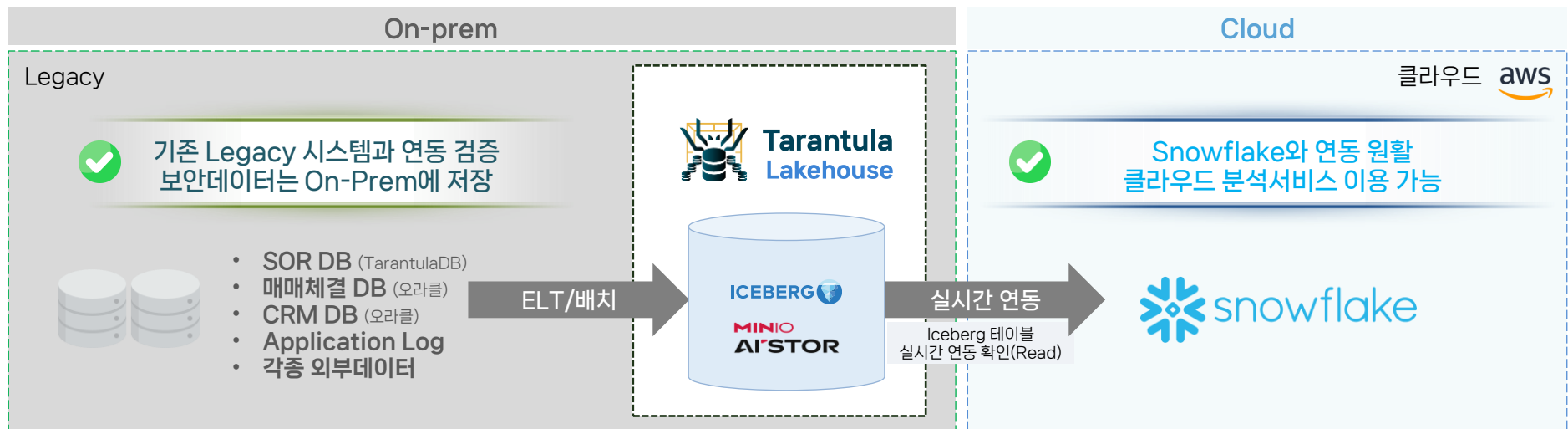


- **오픈소스 기반의 투명한 기술구조와 개방성**

솔루션의 구성요소가 전부 오픈소스로 이루어져 기술구조가 투명하고, 벤더 종속 없이 고객사의 환경에 맞춘 최적의 구성 가능

2.3.7 솔루션 특징점 – 3) 온프레임과 클라우드 데이터의 자유로운 연계

On-Prem 기반의 저장소이지만 뛰어난 개방성을 가지고 있으며, 클라우드 분석서비스인 Snowflake와 연동사례 보유
온프레임과 클라우드를 포괄하는 데이터 통합의 기반 저장소로 활용



■ 핵심 기능

- ✓ 통합된 SQL 쿼리 엔진
- ✓ ACID 트랜잭션 지원
- ✓ 고급 쿼리 최적화
- ✓ 스토리지-컴퓨팅 영역분리 아키텍처



■ 주요 장점

- ✓ 비용 효율성 극대화
- ✓ 대규모 분석 등 성능 우수성
- ✓ 개방성(클라우드 연동)
- ✓ 유연성(AI/ML 연동)



2.3.8 Tarantula BIDB 기능 및 역할

Tarantula Lakehouse 구성 요소 중 TarantulaBIDB는 **안정성과 확장성**을 고려한 설계로 사용자에게 **운영 편의** 제공

Meta Store/BI DB



TarantulaBIDB

- OLAP 분석을 위해 특화된 DB엔진으로, 대용량의 정형/비정형 데이터를 안정적으로 연산하여 쿼리결과를 제공
- 기존 BI도구를 유지하면서도 보안, 감사에 대한 기능은 물론 API 기반 통합을 제공하므로 서비스 확장성이 뛰어남

■ TarantulaBIDB의 특징점

1. 기존 BI 활용	• PostgreSQL을 지원하는 BI 툴 (예: Metabase, Superset, Tableau 등)에서 Trino 쿼리 결과를 별도 커넥터 없이 바로 조회 가능
2. 트랜잭션 데이터와 분석 결과 융합	• PostgreSQL 기존 데이터 + Trino 결과 를 조인하여 트랜잭션+분석 기반의 복합 쿼리 수행
3. 보안 및 감사 관리 통합	• PostgreSQL에서 통합된 접근 제어 및 감사 로깅 수행
4. API 기반 통합 편의	• REST API 로 PostgreSQL을 호출해 Trino 결과를 조회하므로 앱 통합이나 마이크로서비스 연동이 간단

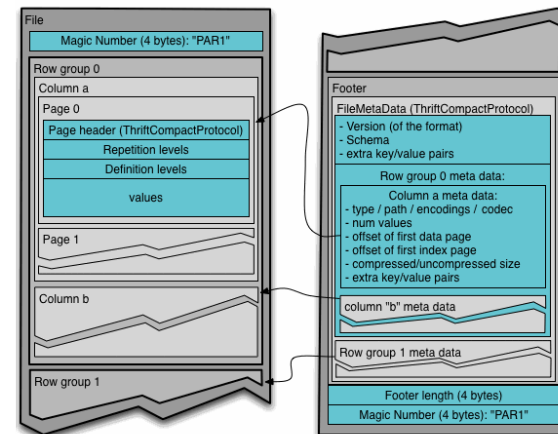
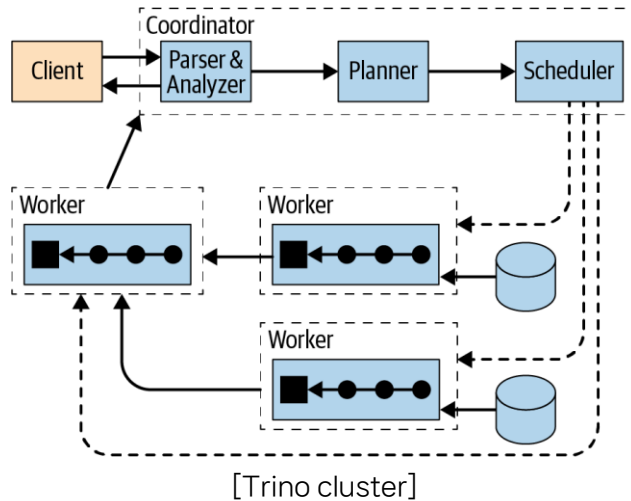
2. 솔루션 소개 ▶ 2.3 타란툴라 Lakehouse

2.3.9 BI를 통한 조회 시 성능향상을 위한 튜닝방안

대량의 데이터를 빠르게 조회하기 위한 여러가지 튜닝방안을 보유하고 있으며, 실제 사용사례를 통해 검증한 바 있음

주요 성능 튜닝 방안

- 1) Trino MPP 방식으로 서버 대수 추가 또는 리소스 추가
- 2) 테이블 설계 시 비정규화
- 3) 테이블 파티셔닝 설계
- 4) Iceberg table 저장 시 parquet 파일 포맷으로 저장 (iceberg 는 parquet 파일 말고도 여러 파일 포맷이 가능)
- 5) SQL 작성시 꼭 필요한 컬럼명을 명시하여 조회 (* 사용금지)
- 6) Aggregation table 활용



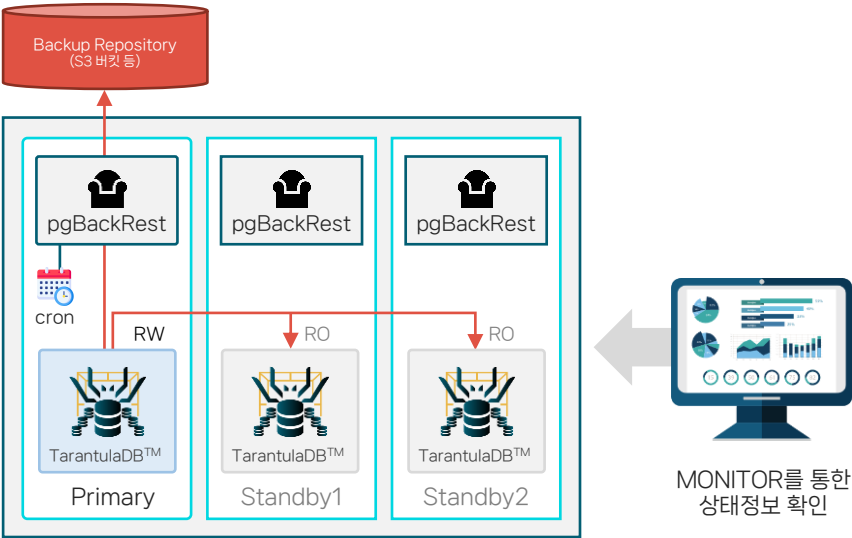
[Parquet 포맷]

2. 솔루션 소개 ▶ 2.3 타란툴라 Lakehouse

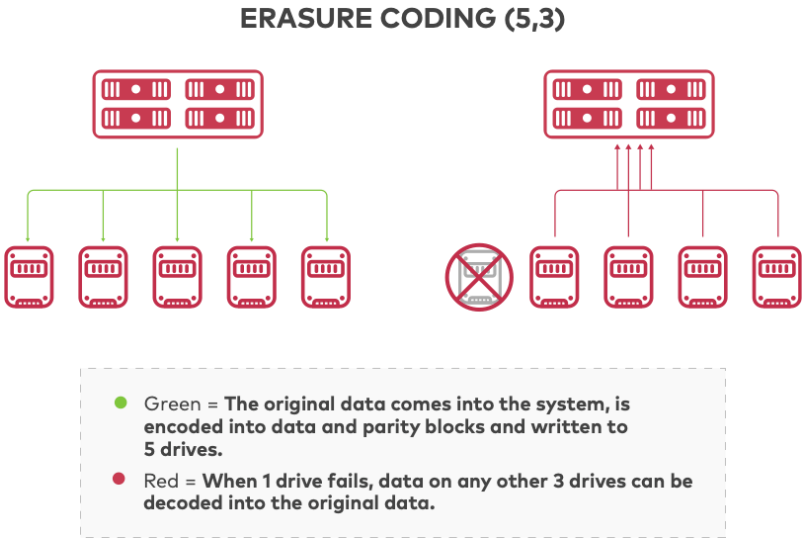
2.3.10 백업 방안

백업 대상 데이터 특성에 맞게 최적의 백업 방안을 제공

백업 대상 데이터	백업 방안	세부 방안
Meta 데이터	아카이브 모드 Hot / Cold 백업	<ul style="list-style-type: none">메타DB 백업은 중요도가 대단히 높으므로 복원 테스트와 무결성 검증을 통해 백업의 안정성을 철저히 확보TarantulaBIDB 아카이브 모드 운영으로 Hot / Cold 백업 수행
S3 데이터	Erasure coding 방식 데이터 분산저장으로 데이터를 복원	<ul style="list-style-type: none">S3 데이터는 대량인 경우가 많으므로 재생산 전력이 가장 효율적임전체 백업이 아닌 복원 가능한 설계가 핵심S3 versioning 기능 이용 시 스토리지 사용량이 급증



[Tarantula DB 백업 모드]



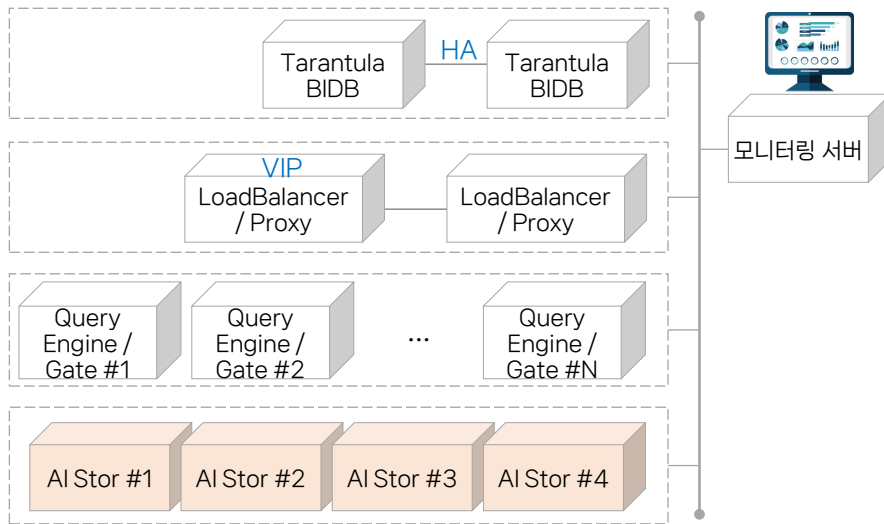
[이레이저 코딩]

2. 솔루션 소개 ▶ 2.3 타란툴라 Lakehouse

2.3.11 Tarantula Lakehouse H/W 구성도

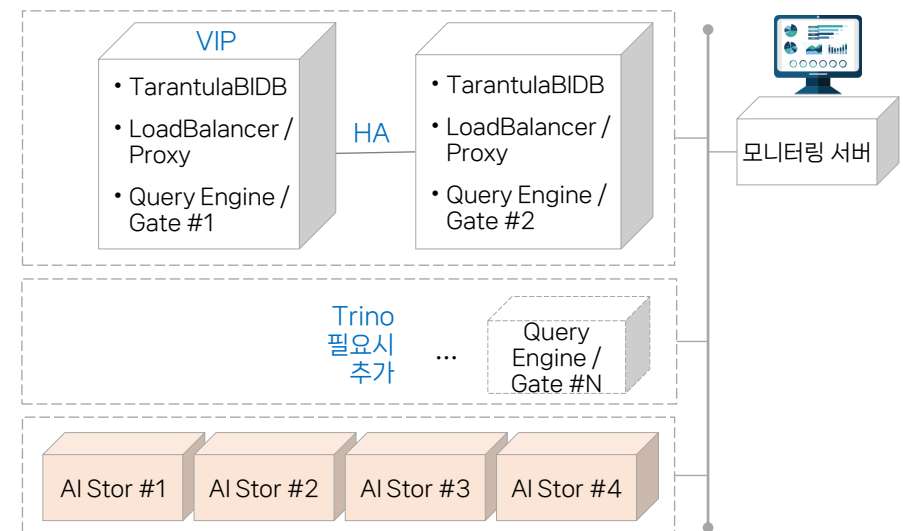
일반적인 권장 구성이며, 단계적 확장을 고려하여 초기 비용 부담을 줄인 최소화 구성안 선택 가능

1안 - 권장 구성도



- 성능과 안정성 최대화 구성

2안 - 최소화 구성도



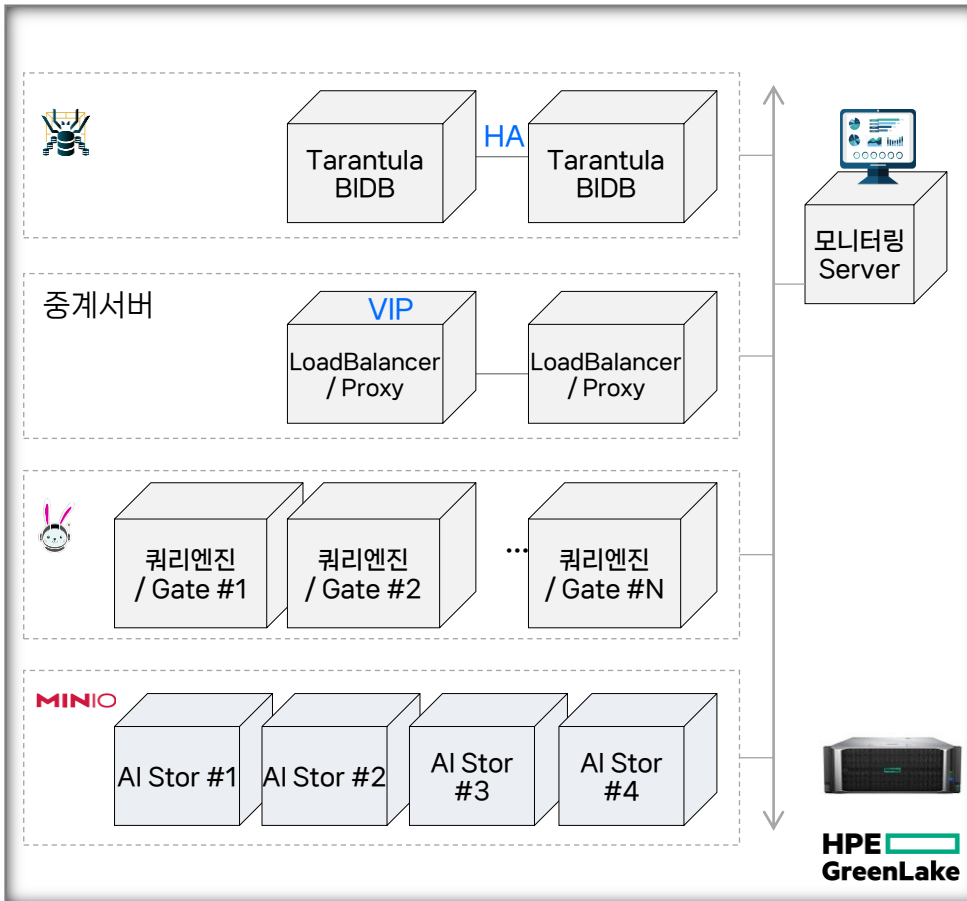
- 하나의 서버에서 동작할 수 있는 여러 기능들을 묶어 서버의 수량을 최소화 한 구성 (3가지 기능을 하나의 서버에 구성)
- 운영상 문제는 없으며 데이터 처리범위를 단계적으로 확장 시, 초기 단계 구성에 적합

- 100Tera 기준 구성이며, AI Stor 외에 모든 서버를 가상화(Vmware etc.) 시스템으로 적용 가능
- AI Stor 권장 스펙
 - 서버당 고성능 NVMe PCIe 4.0 디스크 11개 장착되며, 기존 SSD 대비 5배 이상 성능 확보 됨(7,500MB/Sec)
 - 100Gbps 초고속 네트워크 지원을 통해 대규모 병렬 처리 인프라 환경 제공

2. 솔루션 소개 ▶ 2.3 타란툴라 Lakehouse

2.3.11 Tarantula Lakehouse H/W 구성도 – 실제 적용된 구성 사례

100TB 규모로 구성하였으며, MinIO AIStor 서버(물리서버)는 4대이며, 그 외 서버는 VM서버로도 구성 가능



데이터 규모 (Sample)

- 데이터 100TB 기준

서버 유형

- MinIO AIStor 서버 : 물리서버 4대
- 그외 서버 : 물리서버 또는 가상화 서버 (Vmware 등)

스토리지 (MinIO AIStor) 요구 사양

- 서버당 11개의 고성능 NVMe PCIe 4.0 디스크 장착
- 기존 SSD 대비 5배 이상의 성능 제공 (최대 7,500 MB/sec)
- 대규모 병렬 처리 환경을 위한 100Gbps 초고속 네트워킹 지원

구성의 특징점 및 기대효과

- 클라우드와 온프레미스 **하이브리드 Lakehouse** 구성
- 노드 증설과 성능이 비례적으로 상승되는 **수평 확장성** 제공
- 하이브리드 구성으로 서비스 안정성과 **획기적인 비용 절감** 효과

2. 솔루션 소개 ▶ 2.3 타란툴라 Lakehouse

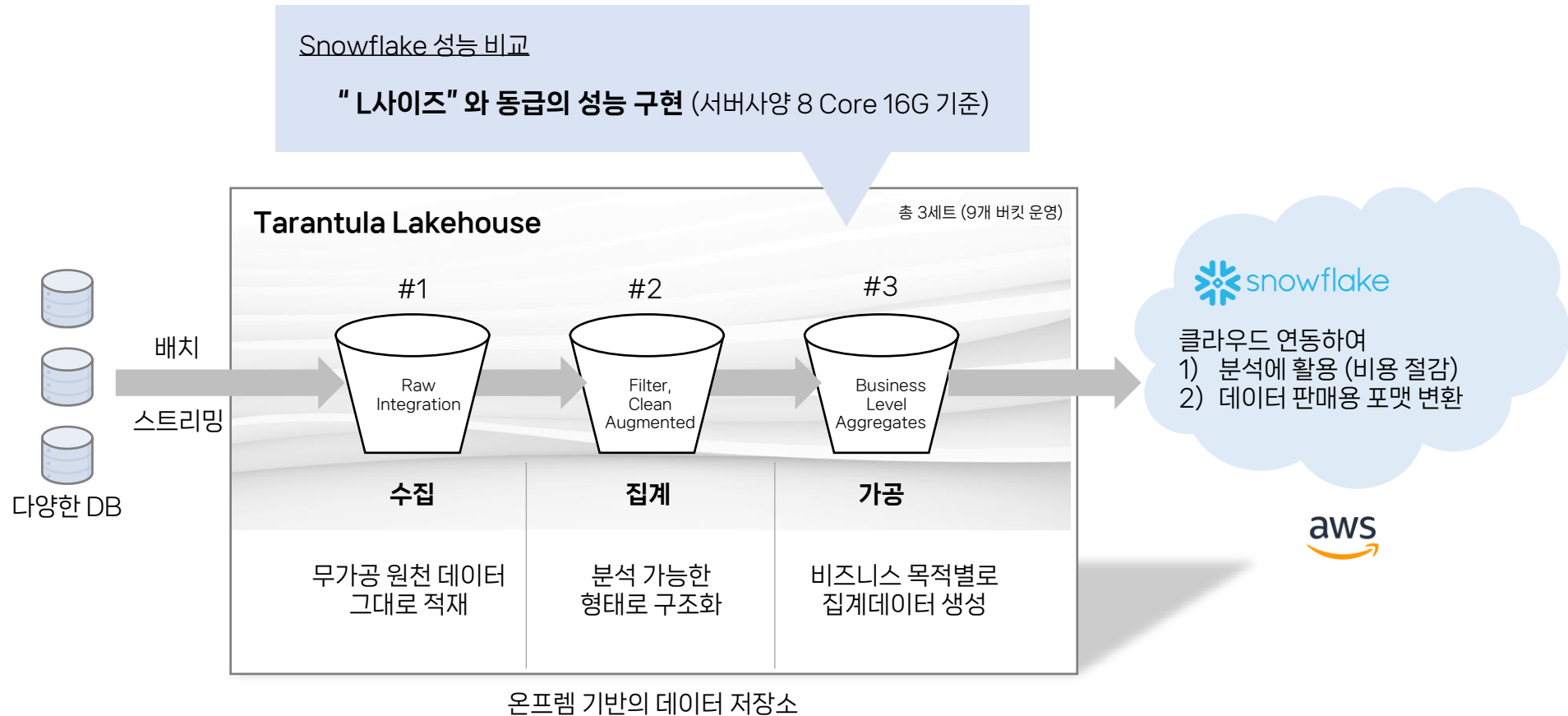
2.3.13 Tarantula Lakehouse H/W 권장사양(Specifications)

※ 서버 OS 는 모두 RHEL9 으로 구성. (서브스크립션 필요)

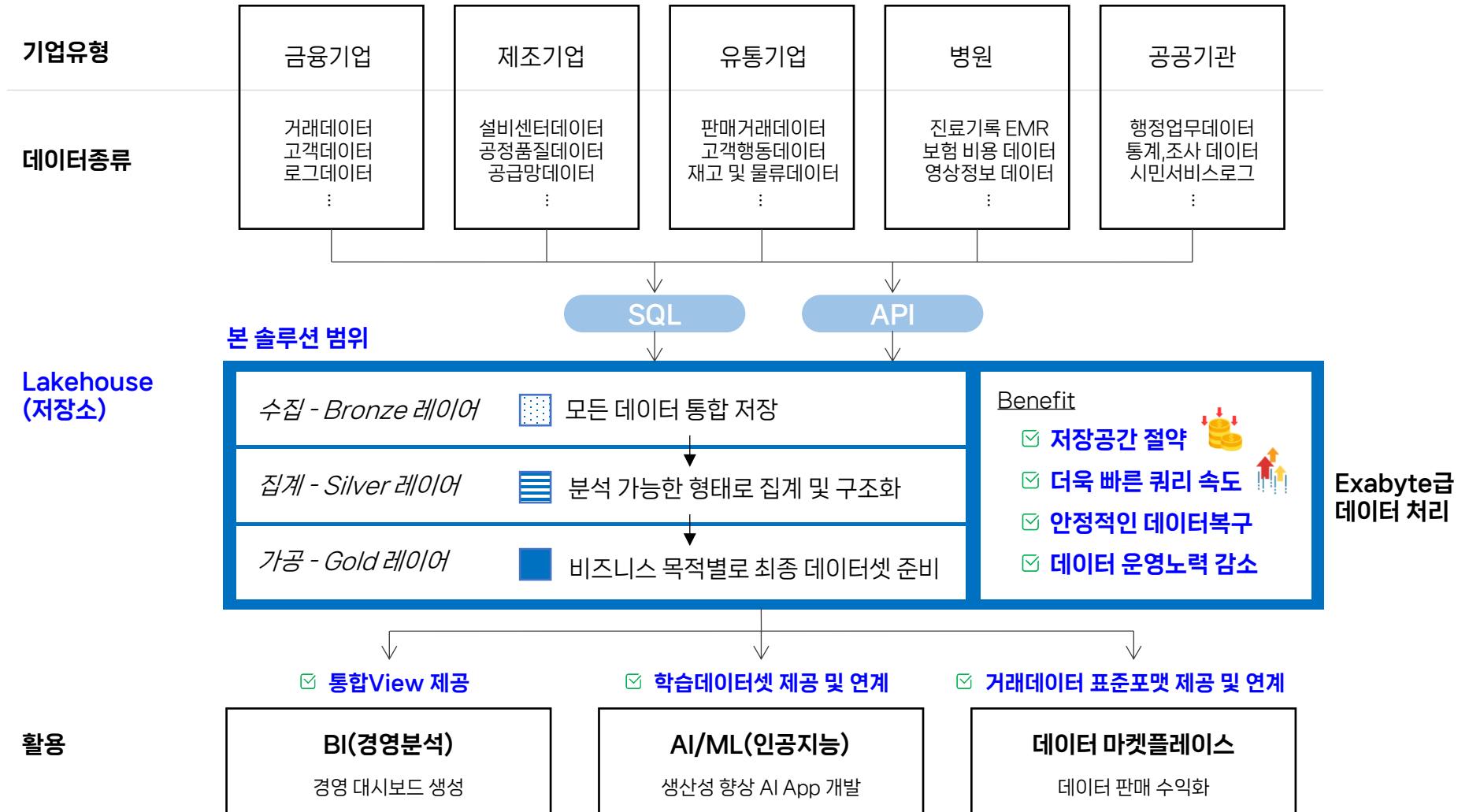
구분	항목	권장사양	수량	
TarantulaBIDB 서버	프로세서	16 Core	1	<ul style="list-style-type: none"> HA 구성으로 동일사양 2대 필요 Pacemaker 필요 공유 스토리지 영역 필요
	메모리	128 GB Mem	1	
	OS 디스크	480GB SSD	2	
	NIC 1	4포트 1GbE	2	
	NIC 2	듀얼 포트 10/25GbE	2	
	HBA	듀얼 포트	2	
Load Balancer/ Proxy 서버	프로세서	8 Core	1	<ul style="list-style-type: none"> 동기화 구성으로 동일사양 2대 필요
	메모리	64 GB Mem	1	
	OS 디스크	480GB SSD	2	
	Data 디스크	240GB SSD	2	
	NIC 1	4포트 1GbE	2	
	NIC 2	듀얼 포트 10/25GbE	2	
Query Engine/ Gate 서버	프로세서	8 Core	1	<ul style="list-style-type: none"> 동일사양 최소 2대 필요 (데이터 부하에 따라 가변적)
	메모리	256 GB Mem	1	
	OS 디스크	480GB SSD	2	
	Data 디스크	240GB SSD	2	
	NIC 1	4포트 1GbE	2	
	NIC 2	듀얼 포트 10/25GbE	2	
AI Stor 서버	프로세서	16 Core	1	<ul style="list-style-type: none"> 동일사양 4대 필요 Data 영역은 별도 Raid 구성 필요 없음 (MinIO 자체 Parity 설정) 전체 용량 중 약 63% 정도 Usable (100TB Usable 필요 시 약 160TB 물리 디스크 필요) 제안 스토리지 가용영역은 최소 100TB 기준임 NVMe 디스크 사용 권장
	메모리	256 GB Mem	1	
	OS 디스크	480GB SSD	2	
	Data 디스크	3.84TB NVMe	11	
	NIC 1	4포트 1GbE	2	
	NIC 2	듀얼 포트 10/25GbE	2	
Monitoring 서버	프로세서	8 Core	1	<ul style="list-style-type: none"> 1대 단독 서버
	메모리	128 GB Mem	1	
	OS 디스크	480GB SSD	2	
	Data 디스크	1.92TB SSD	2	
	NIC 1	4포트 1GbE	2	
	NIC 2	듀얼 포트 10/25GbE	2	

2.3.14 Lakehouse 적용 사례

넥스트레이드에 도입된 Lakehouse는 내부적으로 공간을 3계층으로 분리하여 사용하고 있음 (*메달리온 아키텍처 개념)
최종 가공된 데이터는 snowflake 연동과 내부 BI연동에 사용되고 있음



2.3.15 Lakehouse 데이터 파이프라인 아키텍처



2.3.16 글로벌 기업의 Iceberg기반 Lakehouse 적용 사례 (1/2)

NETFLIX



- 1. 저장 비용 90% 절감:**
기존 테이블 포맷 대비 메타데이터 최적화 및 파일 관리 개선
- 2. 쿼리 성능 대폭 향상:**
메타데이터 푸시다운으로 데이터 파일 스캔 완전 제거, 일부 쿼리는 1시간 → 수 초로 단축
- 3. 증분 처리 효율화:**
전체 재처리 대신 변경분만 처리하여 컴퓨팅 비용 대폭 절감

기존 저장소의 한계

- **Hive 기반 시스템의 한계**
대규모 데이터셋에서 스키마 진화, 동시성, 성능 문제 발생
- **엑사바이트급 데이터**
웨어하우스의 확장성 및 신뢰성 요구
- **지연 데이터 처리와 증분**
처리의 효율화 필요

개선

Lakehouse 도입 후

- ✓ Exabytes급 데이터 처리의 안정성 확보
- ✓ 비용 절감
- ✓ 성능 향상
- ✓ 업무 효율화

Apple



- 1. 유지보수 작업 시간 98% 단축:**
2시간 → 수 분으로 감소
- 2. 집계 쿼리 99.9% 단축:**
1시간 이상 → 수 초로 개선, 메타데이터 푸시다운 활용
- 3. 규제 준수 비용 대폭 절감:**
대규모 데이터셋에서 효율적인 행 단위 업데이트 가능

기존 저장소의 한계

- **GDPR, DMA 등 규제 준수**
행 단위 업데이트 필요
- **전통적 Hive 한계**
파티션 단위로만 업데이트 가능해 대규모 데이터셋에서 비효율적
- **페타바이트급 테이블**
유지보수 작업 시간 단축 필요

개선

Lakehouse 도입 후

- ✓ 규제 준수환경 개선
- ✓ 대규모 데이터셋 처리환경 구축
- ✓ 유지보수 시간 단축

2.3.16 글로벌 기업의 Iceberg기반 Lakehouse 적용 사례 (2/2)



- 데이터 적재 시간 90% 단축:**
45분 → 5분으로 감소
- 실시간 분석 역량 강화:**
준실시간 데이터 처리로 비즈니스 의사결정 가속화
- 데이터 일관성 보장:**
스냅샷 격리로 동시성 문제 해결

기존 저장소의 한계

- **Kafka 토픽 처리 병목**
HDFS 데이터 적재 시간이
45분으로 실시간 분석에 병목
- **실시간 처리의 어려움**
수십억 건의 회원 활동 로그
실시간 처리 요구에 대응 필요

개선

Lakehouse 도입 후

- ☑ 보다 빠른 실시간 분석
- ☑ 데이터 일관성 보장



- 데이터 신뢰성 향상:** 타입 불일치 및 손상 문제 해결
- 쿼리 성능 개선:** 효율적인 데이터 스캔으로 분석 속도 향상
- 운영 안정성:** 단일 장애점 제거로 시스템 신뢰성 강화

기존 저장소의 한계

- **이벤트 처리량 급증**
Adobe 경험플랫폼(AEP)에서
일일 100만 배치, 320억
이벤트 처리
- **데이터 신뢰성 문제:**
스키마 강제 부족으로 타입
불일치 및 손상 발생
- **읽기 신뢰성:** 대규모 테이블의
느린 파일 리스팅 및 비효율적
데이터 스캔
- **확장성:** 별도 메타데이터 서비스
의존으로 단일 장애점

Lakehouse 도입 후

- ☑ 원활한 대규모
이벤트 처리
- ☑ 데이터 신뢰성 향상
- ☑ 운영 안정성

개선

2.3.17 국내 기업의 Iceberg 기반 Lakehouse 적용 사례 (1/3)

오브젝트 스토리지와 Iceberg 테이블 포맷을 적용하여 Lakehouse를 구성 후 업무에 활용한 사례

금융기업

기업명	시스템명	도입배경	정량적 효과
K기업	통합 분석 플랫폼	<ul style="list-style-type: none"> • 대규모 결제 이벤트 로그 처리시 성능 한계 • 실시간 이상거래 탐지, 사용자 맞춤형 리포트 	<ul style="list-style-type: none"> ✓ 쿼리성능 30% 향상 ✓ 데이터 복구 시간 90% 향상 (1H 36M)
T기업	금융 데이터 파이프라인	<ul style="list-style-type: none"> • 고객 거래, 이체 로그 실시간 분석 탐지 • 머신러닝 모델 등 데이터 확보 	<ul style="list-style-type: none"> ✓ 준실시간 처리 ✓ 스키마 변경 대응 80% (3일 → 6H) ✓ 운영자동화 (주당 20H → 5H)
M기업	규제준수 거버넌스 강화	<ul style="list-style-type: none"> • 트레이딩 로그 장기보관(금융감독원 요구) • 전통적 DB기반 감사 시스템 비용과 운영의 복잡성 	<ul style="list-style-type: none"> ✓ 감사 이력조회 95% 개선 ✓ 스토리지 비용 60% 개선 ✓ 감사리포트 자동화 (80H → 2H)
H기업	통합 데이터 플랫폼	<ul style="list-style-type: none"> • 여러 시스템에 분산된 고객, 주문, 시장 데이터 통합 및 데이터 사일로 해소 • AI 모델 학습용 데이터셋 구축 	<ul style="list-style-type: none"> ✓ 데이터 통합시간 70% 개선 ✓ AI모델 학습성능 25% 향상 ✓ 운영비용 30% 개선

2. 솔루션 소개 ▶ 2.3 타란툴라 Lakehouse

2.3.17 국내 기업의 Iceberg 기반 Lakehouse 적용 사례 (2/3)

오브젝트 스토리지와 Iceberg 테이블 포맷을 적용하여 Lakehouse를 구성 후 업무에 활용한 사례

테크기업

기업명	시스템명	도입배경	정량적 효과
N기업	대규모 로그 플랫폼	• 매일 수십억건 웹/앱 로그 저장 (안정적, 비용효율적)	✓ 스토리지 비용 40% 개선 ✓ 데이터 처리량 50% 증가 ✓ 운영자동화
K기업	로그 유형별 최적화	• 메시징, 경제, 광고 등 로그의 특성 반영한 최적화된 저장 구조	✓ 쿼리 응답시간 35% 개선 ✓ 실시간 장애탐지, 안정성 증가 ✓ 운영자동화 (90% 개선)
S기업	통합 Lakehouse	• 고객이용 패턴 분석을 위한 데이터 처리	✓ AI기반 미디어 추천 ✓ 신규광고수익 (2억\$) ✓ 쿼리 확장성 3배 증가
L기업	메시징 로그, 광고 Lakehouse	• 메시지 전송량 광고클릭 이벤트 등 Peta byte 규모 스트리밍 데이터 통합	✓ 데이터 파이프라인 지연 80% 절감 ✓ 운영비용 30% 효율화(자동화) ✓ AI모델 학습 데이터 품질 개선 (15%)

2.3.17 국내 기업의 Iceberg 기반 Lakehouse 적용 사례 (3/3)

오브젝트 스토리지와 Iceberg 테이블 포맷을 적용하여 Lakehouse를 구성 후 업무에 활용한 사례

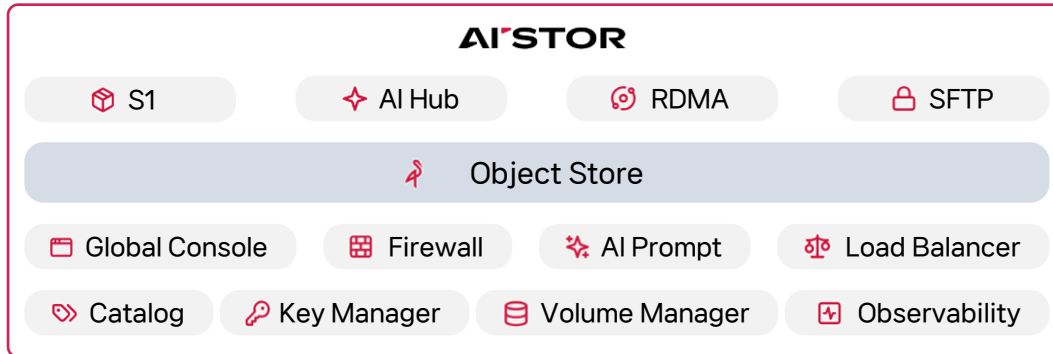
제조기업

기업명	시스템명	도입배경	정량적 효과
L기업	고객경험 개선 운영효율화	<ul style="list-style-type: none">글로벌 단말 및 가전 제품에서 수집되는 로그 원격 진단기존 인프라 확장성과 실시간 분석 한계	<ul style="list-style-type: none">✅ 장애 예측 정확도 20% ↑✅ 데이터 처리 비용 35% ↑✅ 대시보드 응답시간 50% ↑
H기업	커넥티드카 데이터 분석 플랫폼	<ul style="list-style-type: none">커넥티드 차량에서 수집되는 위치, 운행, IoT 센서 데이터 통합 분석OTA 업데이트 및 원격 진단을 위한 실시간 데이터 처리	<ul style="list-style-type: none">✅ 실시간 이벤트 지연 80% 향상✅ 운영 비용 40% 향상✅ 데이터 엔지니어링 시간 60% 절감
H기업	대용량 센서, 예측 정비	<ul style="list-style-type: none">철도/공장 자동화 설비에서 발생하는 IoT 대용량 센서 데이터 분석예측정비(Predictive Maintenance) 고도화	<ul style="list-style-type: none">✅ 설비 가동 중단율 30% 감소 (예측 정확도 향상)✅ 데이터 처리량 2배 증가✅ 운영비용 25% 감소

2. 솔루션 소개 ▶ 2.4 Minio AIStor

2.4.1 MinIO AIStor 소개

AI 워크로드 지원을 위한 대규모 데이터 인프라 운영에 필요한 모든 구성 요소를 포함하는 오브젝트 스토리지 솔루션



- ✓ 세계에서 가장 빠른 오브젝트 스토어 – 실운영환경 **초당 2.2 TiB***처리량
- ✓ ARM에서 x86까지 **모든 하드웨어에서 실행** – Cloud Native 표준
- ✓ **암호화**와 IAM, ILM, 로드밸런싱, 캐싱, 가시성 등 엔터프라이즈급 제공
- ✓ **수천 개의 분산 노드**에 걸쳐 단일 네임스페이스 생성 – 놀라운 확장성
- ✓ AWS S3, HDFS 대비 **40% ~ 70% 까지 TCO 절감**

■ S3와 비교 시 우위 포인트

솔루션 비교	AWS S3	타란툴라 Lakehouse AIStor	우위 포인트
배포 방식	AWS 클라우드 전용	온프레미스 / 클라우드 / 엣지	✓ 유연한 배포 환경
성능	네트워크와 AWS 내부 한계 존재	초고속 (하드웨어 직접 제어 가능)	✓ 로컬 디스크 최적화, 초고성능
비용 구조	저장량·트래픽에 따라 비용 증가	무제한 저장 (서버만 있으면 됨)	✓ 비용 예측 가능, 대규모 저장에 유리
S3 API 호환성	표준	완벽 호환 (AWS S3와 동일한 API)	✓ 이식성 우수, 기존 S3 기반 앱과 호환
고가용성	리전·AZ 분산으로 고가용성	4노드 이상 분산 구성 지원	✓ 자체적 고가용성 클러스터
오프라인 사용	불가능 (AWS 연결 필요)	가능 (내부망에서 독립적 운영)	✓ 금융기관에 적합

2. 솔루션 소개 ▶ 2.4 Minio AISTOR

2.4.2 일반 스토리지 vs MinIO 스토리지

MinIO AISTOR는 AI시대 최적의 저장소로 특히 빅데이터 저장 시 경제성과 확장성이 뛰어나며 NVMe, 100G 네트워크, 고속병렬처리 기술로 블록 스토리지급 성능을 제공

구분	일반 스토리지 (블록)	오브젝트 스토리지
대표솔루션	Dell EMC	aws S3 (클라우드) MinIO AISTOR (온프레임 Global No.1)
특징	데이터를 '블록' 단위로 저장 고속의 입출력 성능 제공	데이터를 '객체'단위로 저장, 무한확장성 (객체 = 파일+메타데이터+ID)
용도	DB, ERP, VM	각종 데이터 저장소 (정형/비정형)
적합한 업무유형	OLTP	★★★★ OLAP
대용량 저장비용	대용량일 때 특히 비쌈	★★★★ 대용량일 때 특히 저렴함
확장성	중간 (고가의 컨트롤러 추가)	★★★★ 뛰어남 (무제한 확장성)
성능	고성능 (지연 없는 I/O 처리)	★★★★ 고성능 (고대역폭 병렬 스트리밍)
AI 연계 적합도	가성비 측면에서 부적합	최적 (AI API 기본지원, 고속병렬처리)
네트워크 접근방식	SAN (FC, iSCSI)	IP기반 네트워크 (HTTP(s) API 통신)
네트워크 특징	스토리지 전용으로 고성능/고비용	일반적인 네트워크도 지원하나 100G일 때 최고의 성능 발휘

2. 솔루션 소개 ▶ 2.4 Minio AISTOR

2.4.3 커뮤니티 버전 vs 엔터프라이즈 버전 (AISTOR)

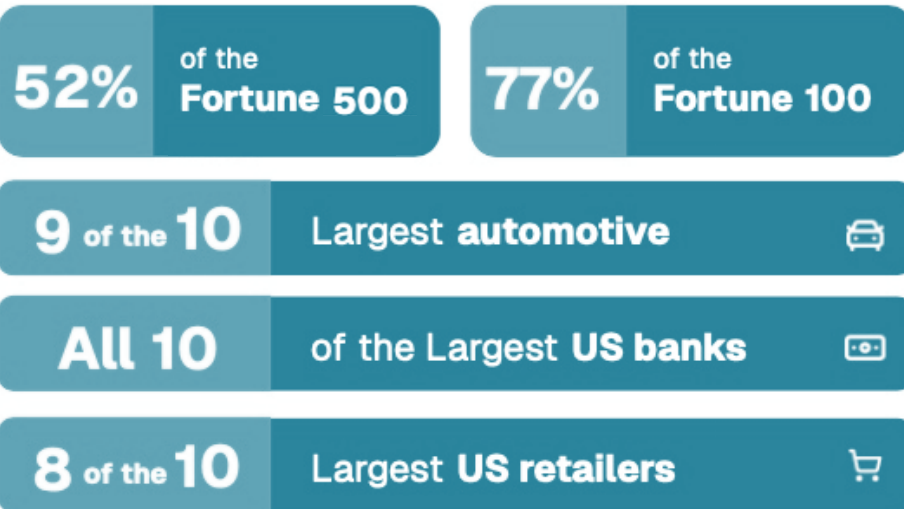
커뮤니티 버전은 테스트 및 개발용으로 기본 기능만 제공하며 실제 비즈니스 운영에서 필요한 보안, 규제준수, 성능 최적화와 전문 지원을 위해 Enterprise Edition인 AISTOR 사용을 권장

구분	MINIO OBJECT STORE Community Edition	MINIO AISTOR Enterprise Edition
라이선스	GNU AGPL v2 라이선스	Enterprise License
용도	테스트 및 개발용	실제 운영환경용 (Production)
보안	네트워크 암호화 : TLS 1.2/1.3	<ul style="list-style-type: none"> 최신 취약점(CVE) 평가 및 보완 네트워크 암호화 : TLS 1.2/1.3 저장 데이터 암호화 : SSE-KMS, SSE-S3, SSE-C 암호화 알고리즘 : AES-256-GCM, ChaCha20-Poly1305, and AES-CBC KMS : AISTOR KMS 및 외부 서드파티 KMS 데이터 방화벽
규제준수	해당 없음	FIPS 140-3, 침투테스트 SOC2, ISO 27001, SEC 17a-4(f), FINRA 4511(c), CFTC 1.31(c)-(d)
지원	<ul style="list-style-type: none"> 커뮤니티 슬랙(Slack) 채널 자원봉사자 지원 	<ul style="list-style-type: none"> SUBNET 엔터프라이즈 지원 포털 SLA: 연중무휴 24/7, 응답시간 4시간 이내, 중요 이슈(PO) 즉각 대응
기능	기본 기능 (지속적 투자 차원)	기본 기능 + 추가 기능(적극적 투자): 콘솔, 암호화, 사이트 간 복제, QoS, 카탈로그, 모니터링, 감사 로그, 헬스 진단, RDMA, GPU Direct, AI 포털, AI 허브
K8s 통합	업스트림 쿠버네티스	업스트림 + Red Hat OpenShift, Rancher RKE2/K3S, VMware Tanzu (TKG)
최적화	해당 없음	소규모 객체 최적화, 대량 삭제, 리스트 작업, 낮은 TTFB, 분산 캐시, 빠른 부팅

2.4.4 레퍼런스

글로벌 No.1 오브젝트 스토리지로 전세계 많은 기업들에 도입된 바 있으며, 유엔넷은 국내 최초 OEM 계약을 통해 판매 및 서비스하며 향후 일본 등 아시아 시장 진출 예정

MinIO에서 가동되는 전세계 기업들



도입 사례



2.4.5 MinIO AIStOR 단독사용 vs 타란툴라 Lakehouse 사용

MinIO는 데이터를 저장하는 데는 강하지만, 분석과 보고 등 데이터의 활용에 한계 존재
Lakehouse는 사용자 친화적 SQL 사용, BI 툴 연계, AI 연동 등 다양한 기능을 종합적으로 제공

구분	MinIO 단독 사용	타란툴라 Lakehouse 사용
구성요소	MinIO AIStor	대용량 쿼리 처리 메타 BI DB MinIO AIStor(오브젝트 스토리지)
기술영역	스토리지 영역	컴퓨팅 영역 + 스토리지 영역
제공 기능	주로 비정형 데이터 '저장/불러오기'에만 초점 분석과 보고 등 데이터 활용 용도는 아니며 데이터 저장소 역할만 제공	★★★★ 다양한 DB에 대한 통합된 SQL 쿼리 ACID 트랜잭션 지원 BI 연동 / AI 연계 데이터 거버넌스 (버전관리, 파티셔닝 등)
용도	데이터레이크 AI학습용 데이터셋 저장소	★★★★ 데이터 웨어하우스, 데이터레이크, 하둡 대체 BI 연계 / AI 연계
접근방식	S3 API (코드 필요)	★★★★ SQL, BI툴, API 등 (사용자 친화적인 접근)
사용자	개발자 / AI 엔지니어	데이터분석가, 개발자, AI 엔지니어, 영업, 경영진 등

감사합니다.

유엔넷 주식회사

솔루션 문의

금융부문

정지원 이사

- 전화 010-4121-1486
- 이메일 j1jung@unnet.co.kr

공공부문

정래혁 이사

- 전화 010-3746-1545
- 이메일 jungrh@unnet.co.kr

엔터프라이즈부문

이강일 전무

- 전화 010-2717-7408
- 이메일 ki.lee@unnet.co.kr

유엔넷 연락처



02-565-7034



www.unnet.co.kr



www.tarantuladb.com

찾아오시는 길

[HQ] 서울 강남구 테헤란로20길 12 안타워 2층

[여의도 Office] 서울 영등포구 여의나루로 71 동화빌딩 7층

