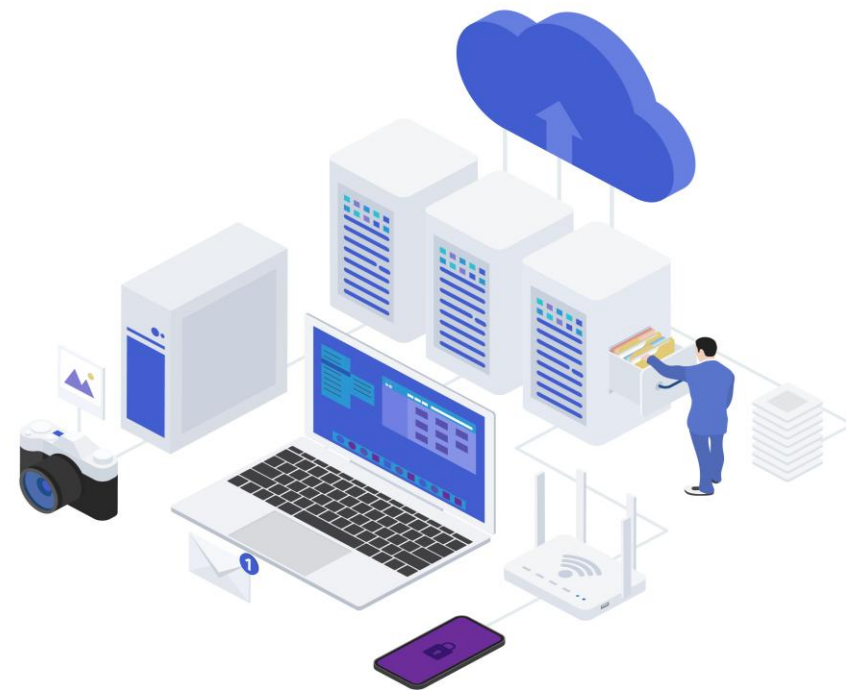


3세대 AI / 빅데이터 플랫폼

Tarantula Lakehouse

UN°NET 유엔넷(주)

© 2026 UNNET. 무단 복제를 금합니다. | All Rights Reserved.



1. 유엔넷(주) 소개

2001년에 설립되어 20년 이상 네트워크부터 데이터까지, 엔터프라이즈 IT 인프라를 통합 설계, 구축, 운영하는 토탈 인프라 솔루션 기업

UN°NET

네트워크에서 데이터까지, 기업의 디지털 전환을 지원하는 파트너

332억

2025년 매출

60% ↑

전년 매출 대비

25년

업력

115 +

임직원수(자회사 포함)

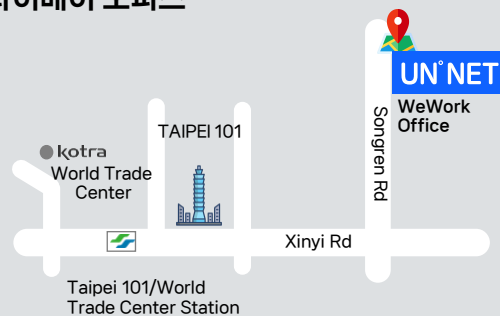
75%

기술인력 비중

서울 오피스



타이베이 오피스



주요 사업 분야

데이터 플랫폼 솔루션

타란툴라DB, 타란툴라Lakehouse

네트워크 솔루션

HPE Juniper, Fortinet, Citrix, IoT

레드햇 솔루션

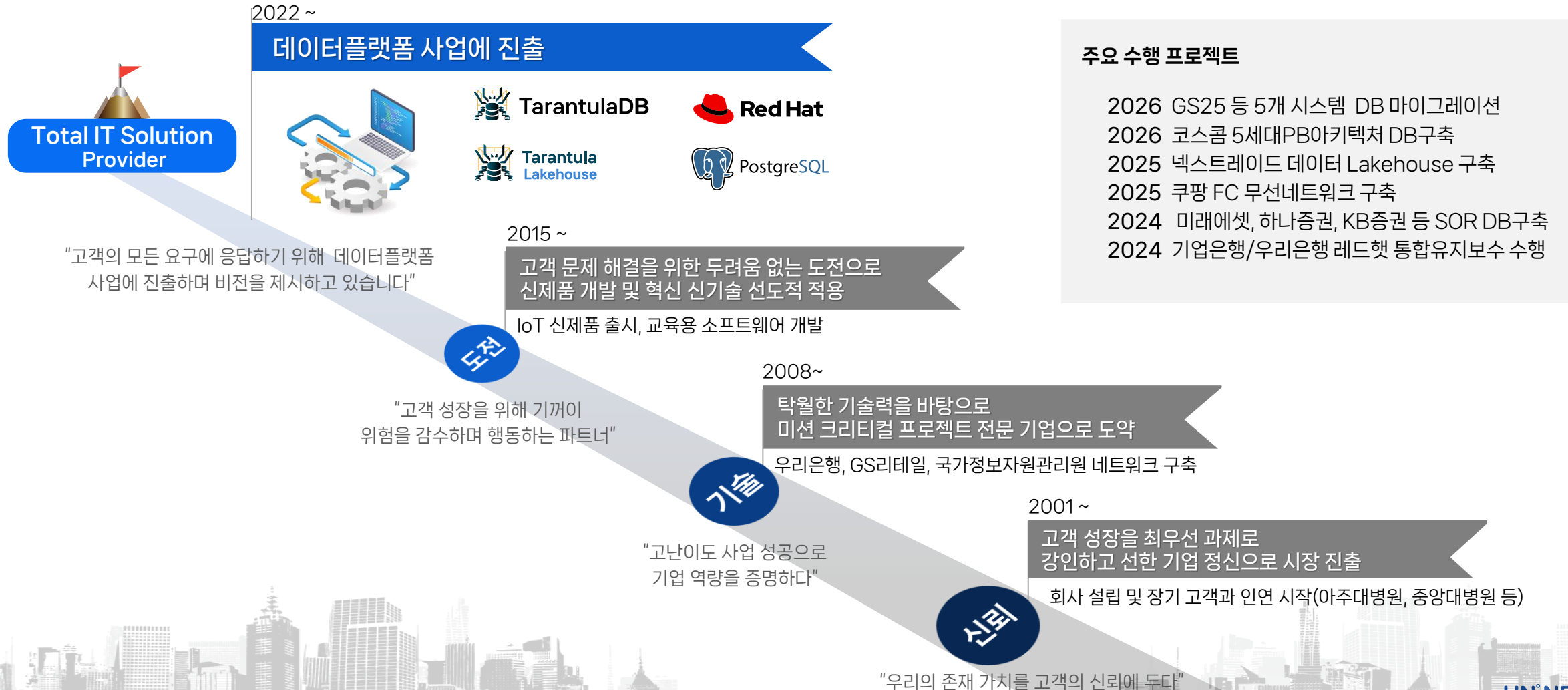
Redhat OS, Openshift

서버 솔루션

HPE, Dell

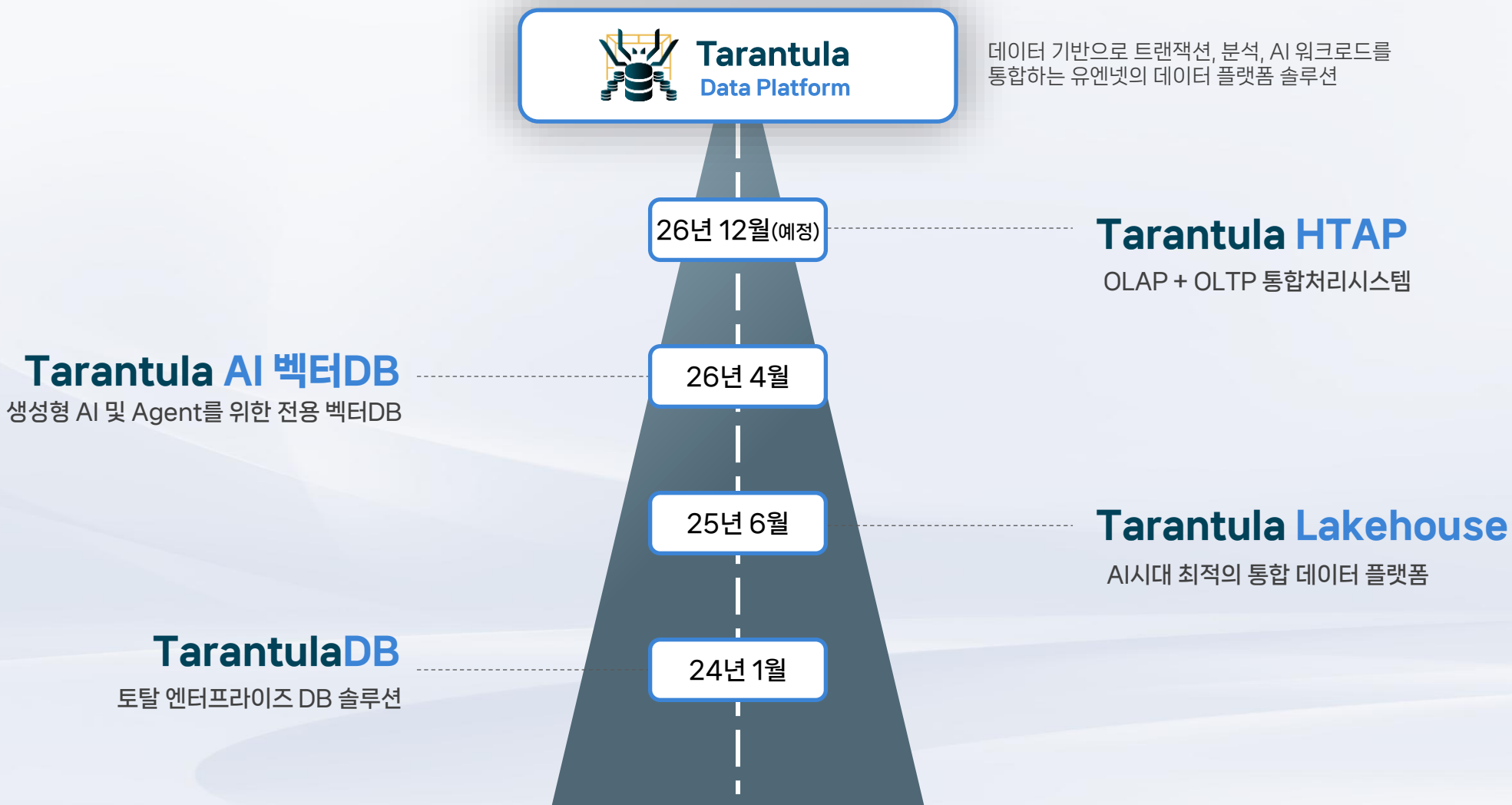
1. 유엔넷(주) 소개

고객의 신뢰와 성장을 최우선으로 두고 금융, 공공, 의료, 리테일 분야에서 다양한 프로젝트를 수행하며 성장



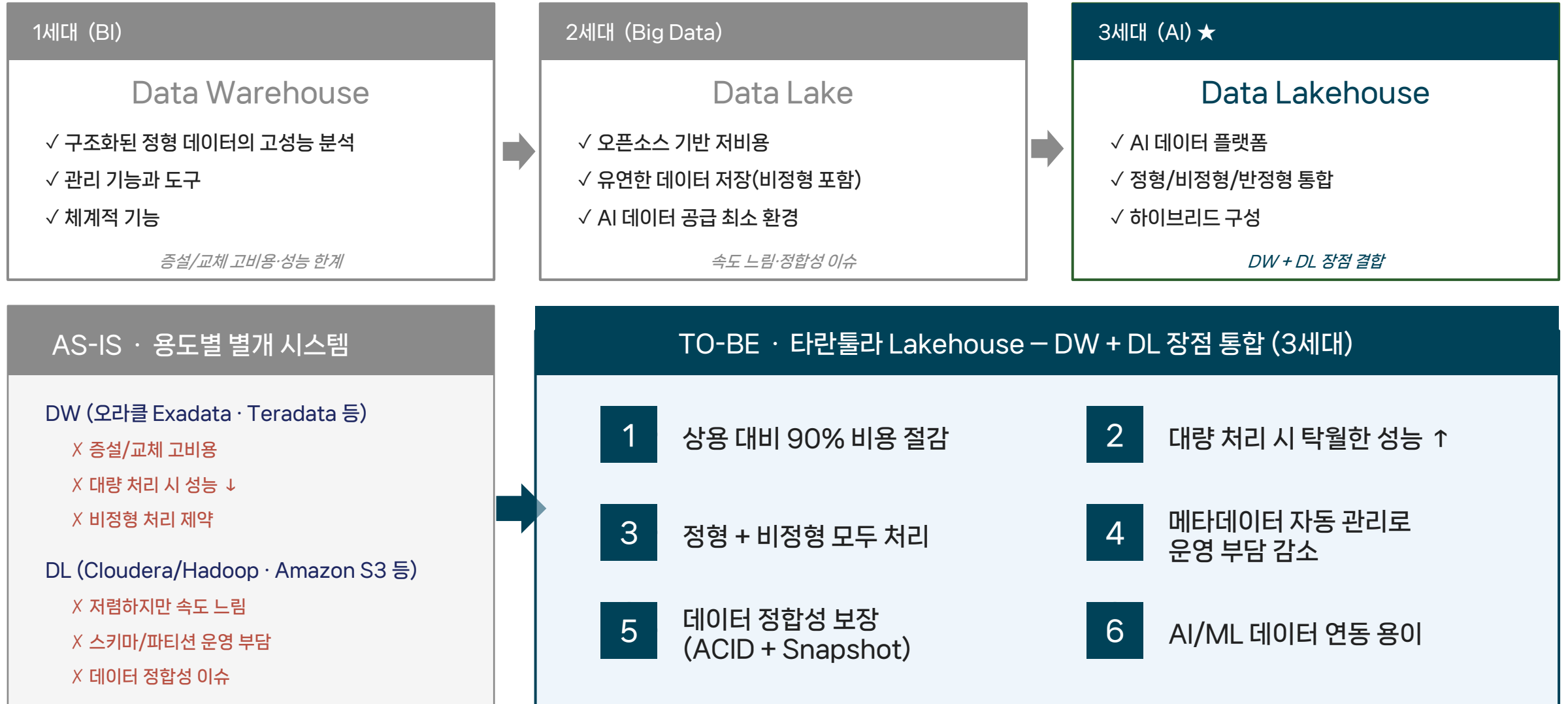
2. 타란툴라 Data Platform 포트폴리오

타란툴라 Data Platform은 타란툴라DB 및 Lakehouse, HTAP(개발예정)에 이르는 유엔넷의 데이터 솔루션 브랜드



3. 왜 Lakehouse인가?

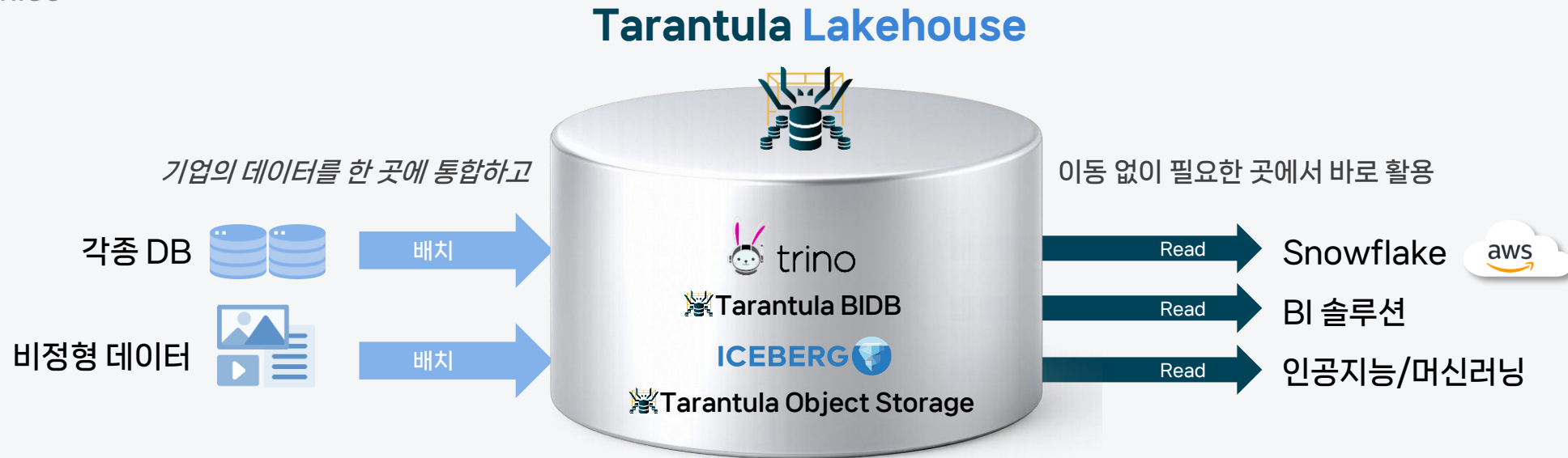
분석과 AI를 아우르는 차세대 데이터 플랫폼 – DW와 DL의 장점을 통합한 3세대 Lakehouse



4. 타란툴라 Lakehouse

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

On-premise



핵심 기능

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

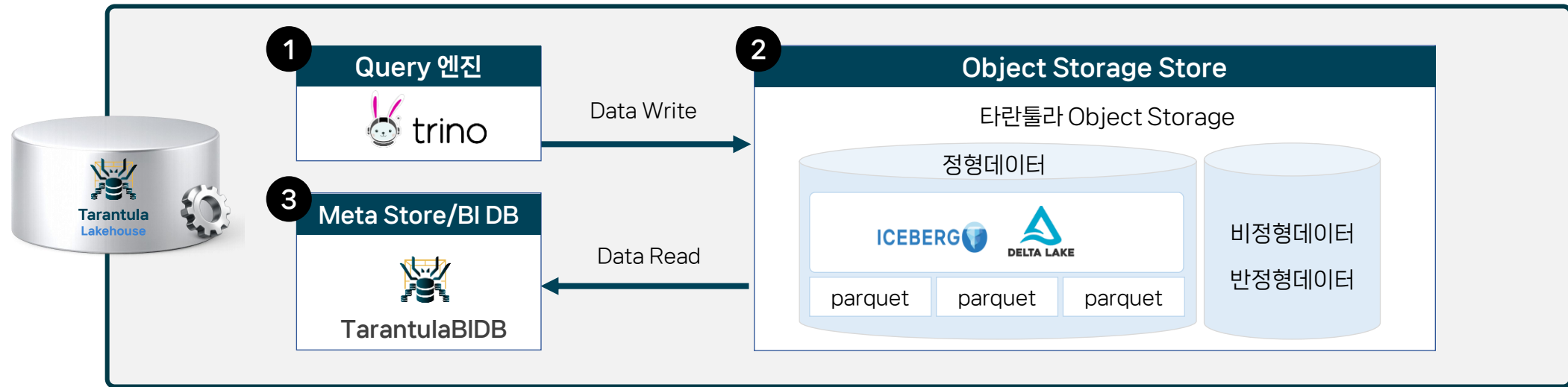


주요 장점

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

4. 타란툴라 Lakehouse - 구성요소

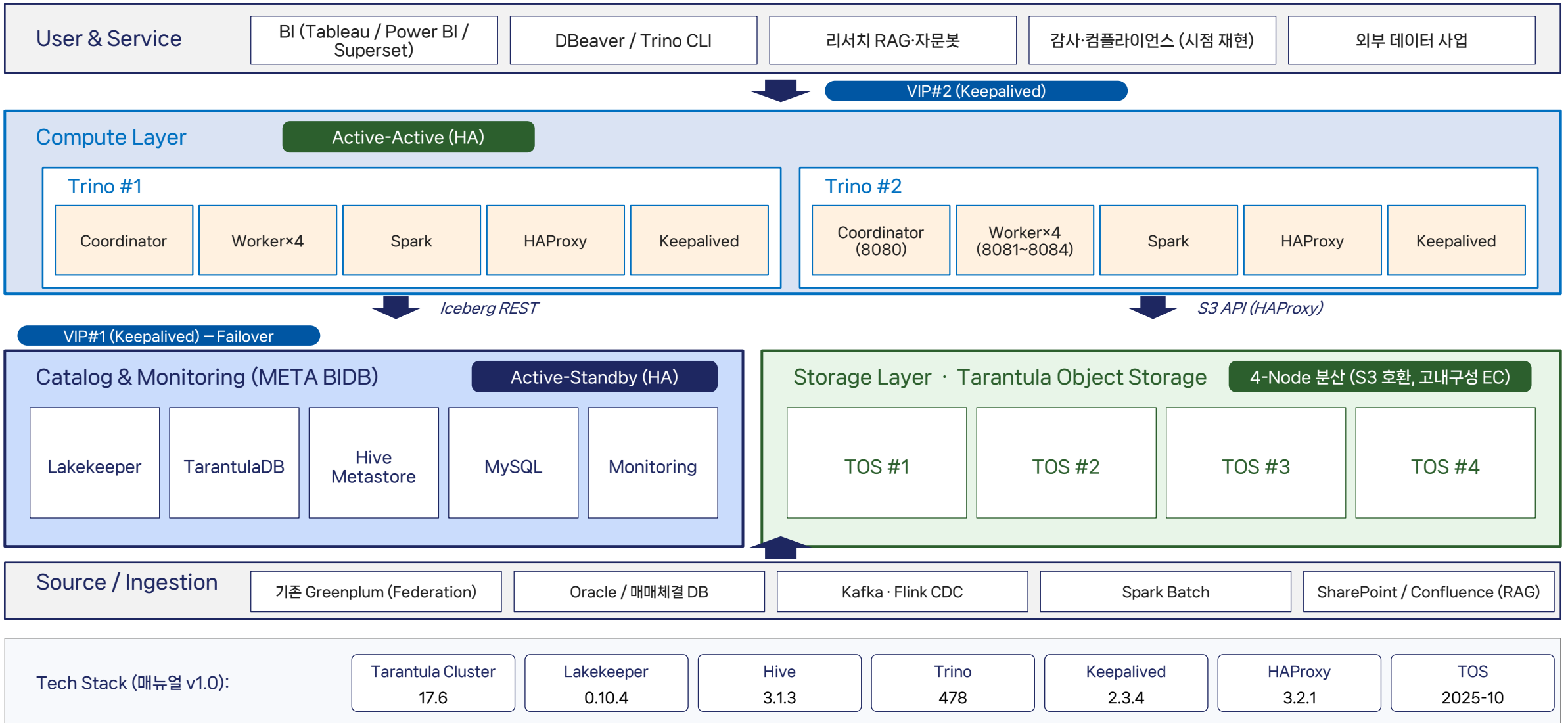
데이터 초고속 쿼리와 메타데이터 관리, 저장까지 각 영역에 특화된 구성요소를 결합한 통합 데이터 아키텍처로, 대규모 데이터 환경에서도 신뢰성과 성능을 균형 있게 제공



■ 구성요소

핵심 컴포넌트	솔루션명	역할 및 기능
1. Query 엔진	trino	- 다양한 데이터 소스를 대상으로 SQL을 사용해 빠르게 분석하는 분산형 *MPP SQL 쿼리 엔진 * MPP, Massively Parallel Processing: 대규모 병렬 처리
2. Meta Store/BI DB	TarantulaBIDB	- ICEBERG 카탈로그 저장 Meta DB - ICEBERG 테이블 조회 용도(Trino Bridge 자체 개발) 및 향후 Vector DB Extension
3. Object Storage Store	Tarantula Object Storage	AI 및 대규모 데이터 워크로드를 위해 설계된 온프레미스 고성능 오브젝트 스토리지 플랫폼으로, 다양한 AI 기능을 통합하여 단일 인프라에서 효율적인 데이터 관리와 분석 제공

4. 타란툴라 Lakehouse – 아키텍처



4. 타란툴라 Lakehouse – 가용성(High Availability)

모든 장애 시나리오에서 SQL 정상 수행하는 고가용성 구조

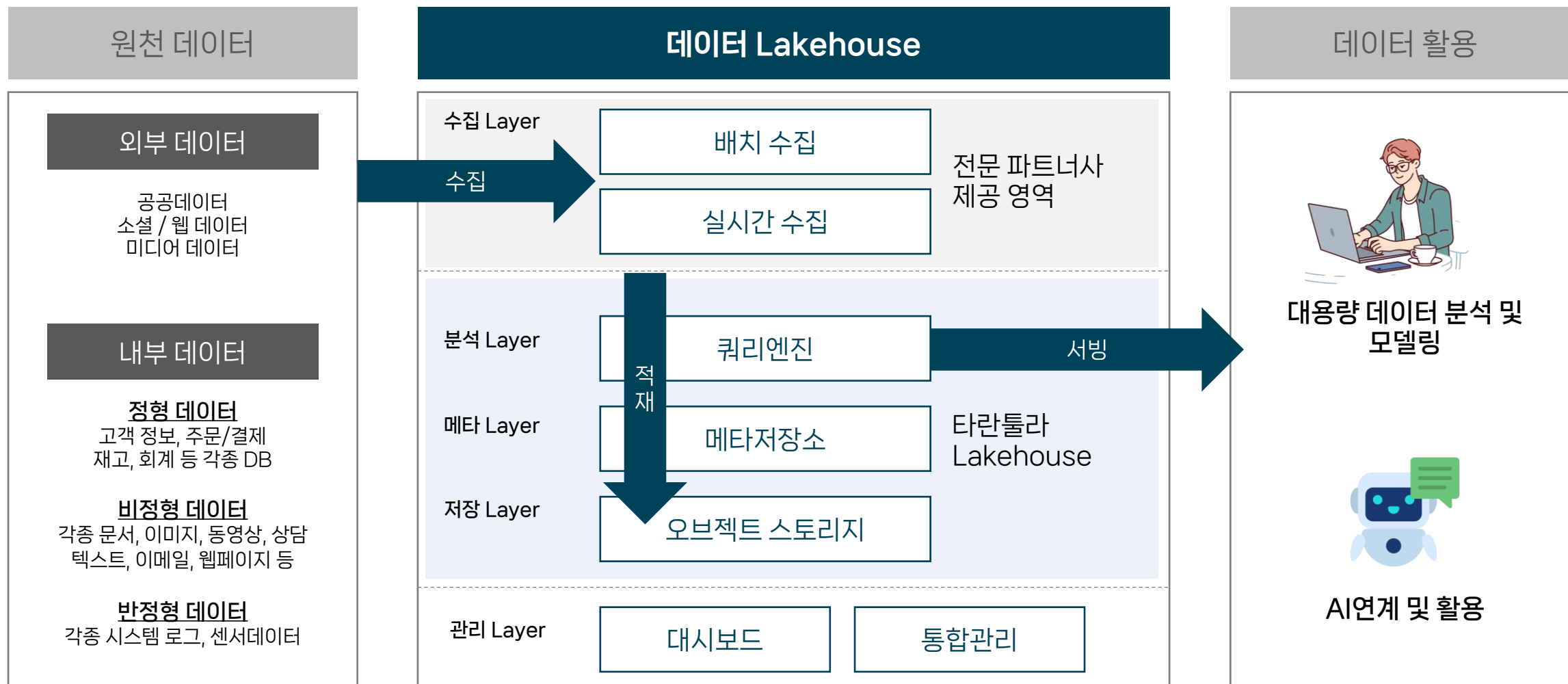


HA 메커니즘 요약

Keepalived(VIP#1~#2) + HAProxy 부하분산 · Iceberg Snapshot Isolation · TOS Erasure Coding · → 시나리오 무중단

4. 타란툴라 Lakehouse – 데이터 전체 흐름

데이터 수집부터 저장, 분석까지 전 영역을 전문 파트너사와 함께 당사에서 토탈 서비스로 제공하며, 고객의 환경에 적합하게 아키텍처 설계



5. 왜 타란툴라 Lakehouse인가?

AI · BI · Snowflake 모두 자유롭게 연동되는 국내 최초 유일 솔루션 (NEXTRADE 검증)



상용 대비 90% TCO 절감

오픈소스 + 효율적 저장/연산 인프라
NEXTRADE 84% 저장비용 절감 검증



대량 처리 탁월한 성능

N사 PoC 1등 (4건/6건). 수십억건 안정 처리
타사 대비 0.4초 vs 8.0초 (천만건 집계)



정형/비정형/반정형 통합

각종 DB·텍스트·이미지·영상·로그까지
단일 Lakehouse(Iceberg)에 통합 저장



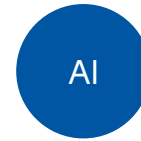
메타데이터 자동 관리

Iceberg Snapshot·Manifest 자동 관리
Lakekeeper 카탈로그로 운영 부담 ↓



데이터 정합성 보장

ACID 트랜잭션 + Snapshot Isolation
비정상 종료에도 무결성 유지



AI/ML 데이터 연동 용이

타란툴라 AI 벡터DB와 즉시 연계
LLM·RAG·자문봇 학습 자산화

NEXTRADE – 타란툴라 Lakehouse 도입 실증

150 TB

5년 운영 규모

8 대

BIDB·Proxy·Mon ×2 /
Query ×2 / AiStor ×4

84% ↓

기존 증설 대비 저장비용 절감

무중단

25년 3월 오픈 후 안정 운영

Tarantula Data Platform 포트폴리오

24.01

TarantulaDB

25.06

Tarantula Lakehouse

26.04

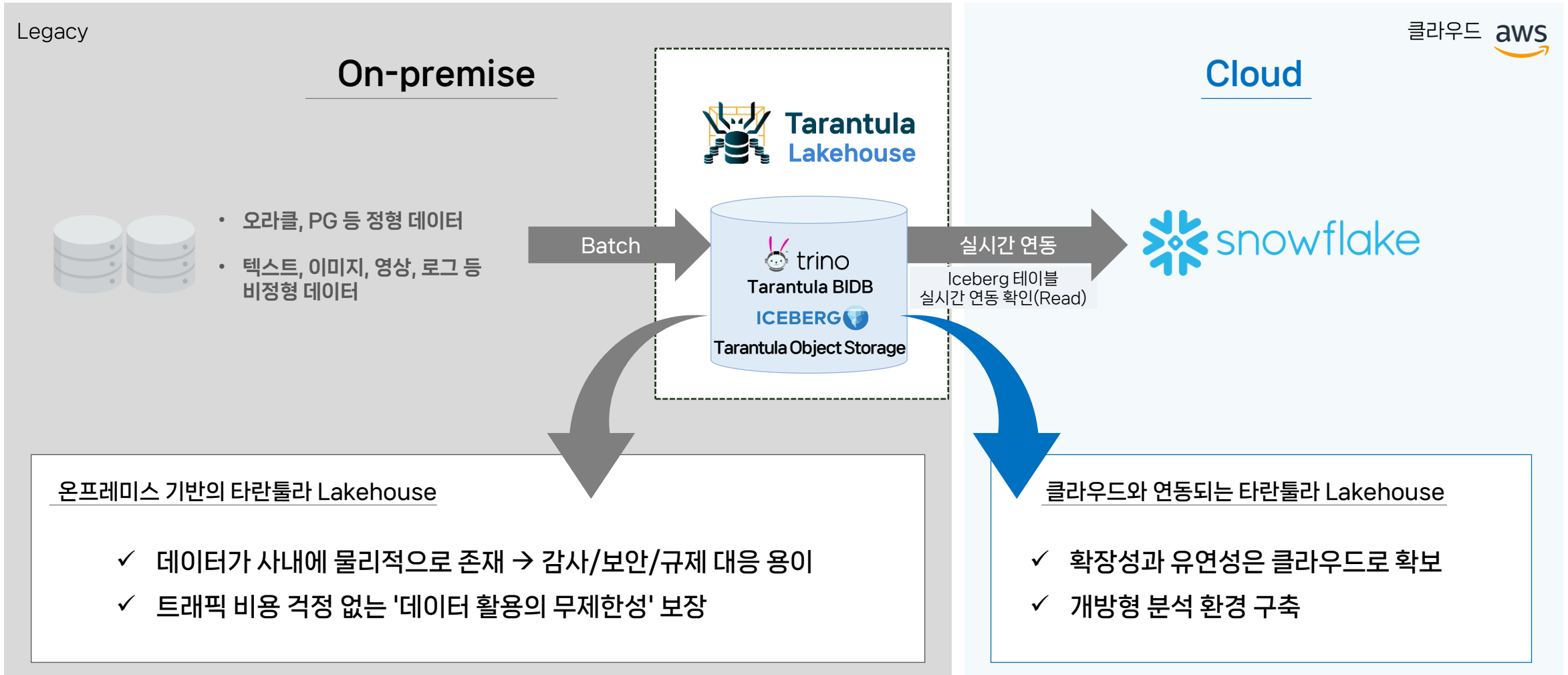
Tarantula AI 벡터DB

26.12

Tarantula HTAP (예정)

5. 왜 타란툴라 Lakehouse인가? – 온프레임과 클라우드를 포괄

온프레임과 클라우드를 포괄하는 데이터 통합 플랫폼으로써 온프레미스의 강력한 보안성과 클라우드의 유연한 확장성을 결합하여 기업의 모든 데이터를 한곳에서 연결하고 비즈니스 가치로 전환

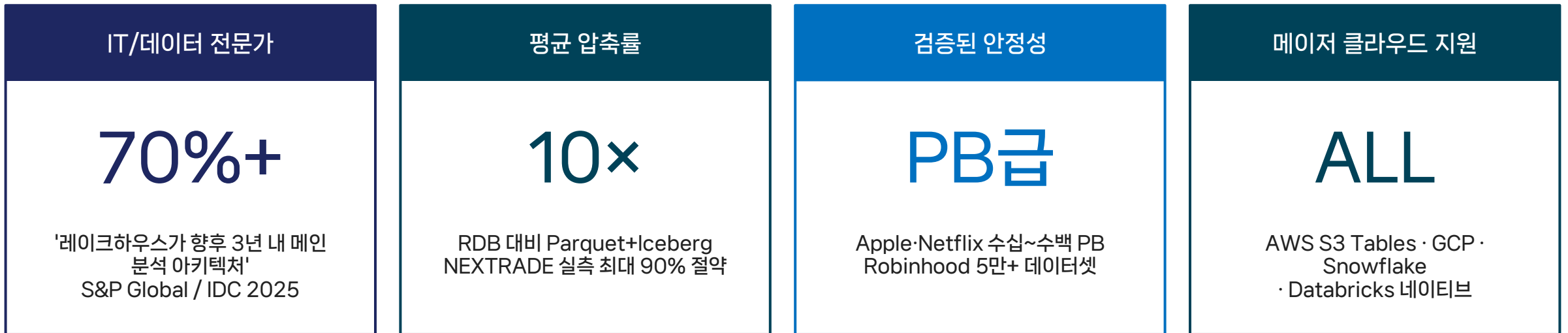


5. 왜 타란툴라 Lakehouse인가? – 사실상 표준 ICEBERG 채택

Format Wars 종료 – Apache Iceberg가 사실상 표준 (AWS·GCP·Snowflake·Databricks 모두 지원)으로 데이터 개방성 보장



Iceberg 채택의 결정적 근거



Iceberg는 "데이터 자산의 헌법" – 쿼리엔진은 갈아낄 수 있어도 데이터 포맷은 바꾸기 어렵다 (Goldman Sachs)

5. 왜 타란툴라 Lakehouse인가? – 사실상 표준 ICEBERG 채택

글로벌 BFSi-빅테크 도입사례를 바탕으로 Iceberg는 단순 트렌드가 아니라 PB급 운영에서 검증된 기술

Goldman Sachs

Legend Lakehouse

Iceberg + Unity Catalog + Databricks

전사 데이터 거버넌스 통합
"벤더 중립성·데이터 자산화 전략"

Source: Data + AI Summit 2025

JPMorgan Chase

Hadoop → Open Lakehouse

Iceberg on S3 + Spark + K8s

Data Mesh 결합
도메인별 자율성 + 중앙 카탈로그
AI 학습 데이터 통합

Source: JPMC Tech Blog · AWS BD Blog

Apple

전사 표준 Lakehouse

Iceberg + Spark/Trino + S3

전 사업부 통합
마이크로배치·스트리밍·ETL을 단일 테이블로 다중 PB

Source: Qlik Tech Blog · Apple Eng.

Netflix

Iceberg 창시 기업

Iceberg + Spark + S3

Iceberg 직접 개발(2018)
추천·스트리밍 분석 백본 수집~수백 PB

Source: Apache Software Foundation

Robinhood

온라인 증권사

Iceberg + CDC + Trino

50,000+ 데이터셋 / 다중 PB / 분 단위 신선도
거래 시간 중 분석.

Source: Onehouse · Robinhood Eng.

AWS Quant Research

퀀트 리서치 레퍼런스

Iceberg Time Travel + S3

백테스트 록어헤드 방지
'FOR VERSION AS OF' 시점 재현

Source: AWS Big Data Blog 2025

6. 국내 동향 + 넥스트레이드 검증사례

국내 증권사 데이터 플랫폼 동향

2022

NH투자증권

증권사 최초 클라우드 분석 플랫폼
온프레임+클라우드 하이브리드

2023~25

KB·미래에셋·한투

리서치/AI 자문봇 빅데이터 강화
(Iceberg 미공개)

2025~

NEXTRADE

국내 최초 Iceberg 기반 Lakehouse
매매체결 150TB · 4개월 구축

NEXTRADE — 국내 최초 대체증권거래소 Iceberg Lakehouse 실측 검증

150 TB

최초 구축 용량
매매체결 데이터

4 개월

구축 기간
기획~안정화

8 대

서버 (Trino + Object Storage)
100Gbps × 2

~80% ↓

5년 TCO
기존 대비 20% 수준

쿼리 성능 (NEXTRADE PoC 실측)

- 일별 천만건 집계: 0.4초 (타사 8.0초 / 0.7초)
- 월별 수억건 집계: 0.8초 (타사 13초 / 1.0초)
- 조인 한달 집계: 2.5초 (타사 29초 / 3.3초)

스토리지 효율 (1년 → 10년 보관)

- CSV 2.8 GB → Iceberg 287 MB (90% 절약)
- Snowflake(Iceberg) 528 MB 대비도 절반 수준
- Parquet+Iceberg 압축 + EC(Erasure Coding)

7. 넥스트레이드 활용 시나리오



1

매매·체결 데이터 분석

매매체결·주문·시세·고객 → 부서별 셀프 SQL/BI. DW 부담 분산.

쿼리 응답 50~70% ↓

셀프 분석 사용자 ↑

운영 DB 영향 0건

2

콜드데이터 이관 / 비용 절감

1년+ 미접근/규제 보관 데이터 → Iceberg+Object Storage 이관.

5년 TCO 10억+ 절감

저장공간 80~90% ↓

증설 CapEx 회피

3

AI · RAG · LLM 활용

리서치·고객상담·매뉴얼·약관 → 임베딩 → 타란툴라 AI 벡터DB.

RAG 챗봇/자문봇 학습
자산

모델 재현성 보장

AI 신사업 기반

4

리스크 · 컴플라이언스 · 감사

전자금감규/자본시장법 대응. Time Travel 시점 재현.

감사 응답 시간 단축

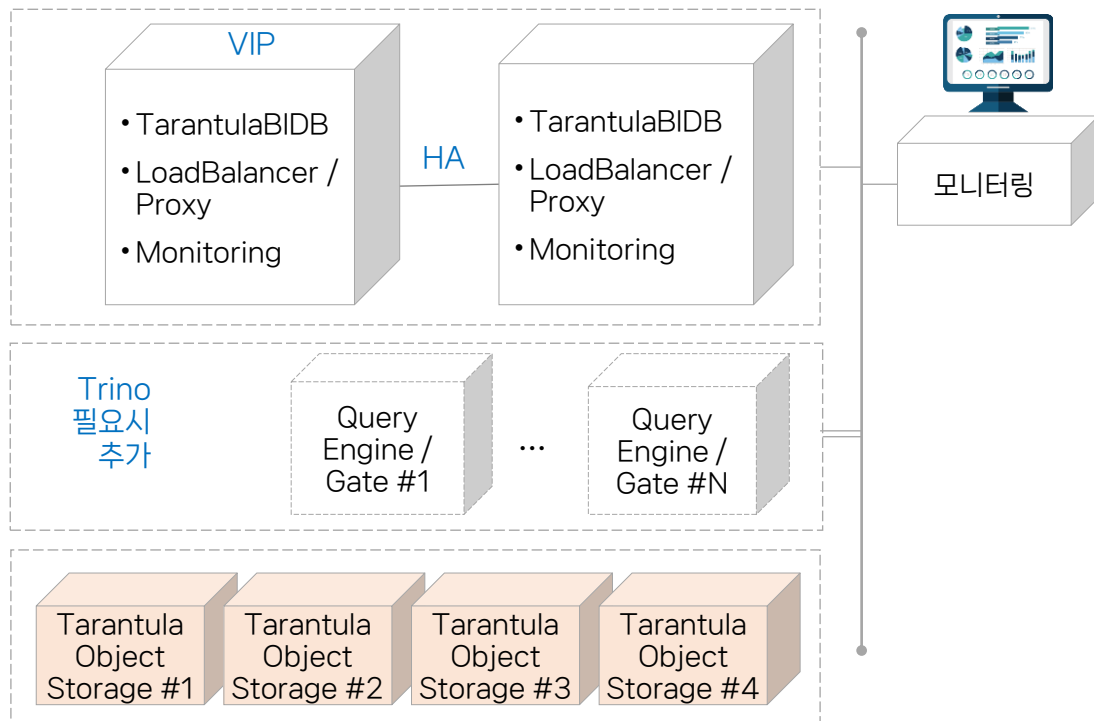
룩어헤드 바이어스 방지

규제·거버넌스 강화

8. 서버 구성 (최소구성 vs 권장구성)

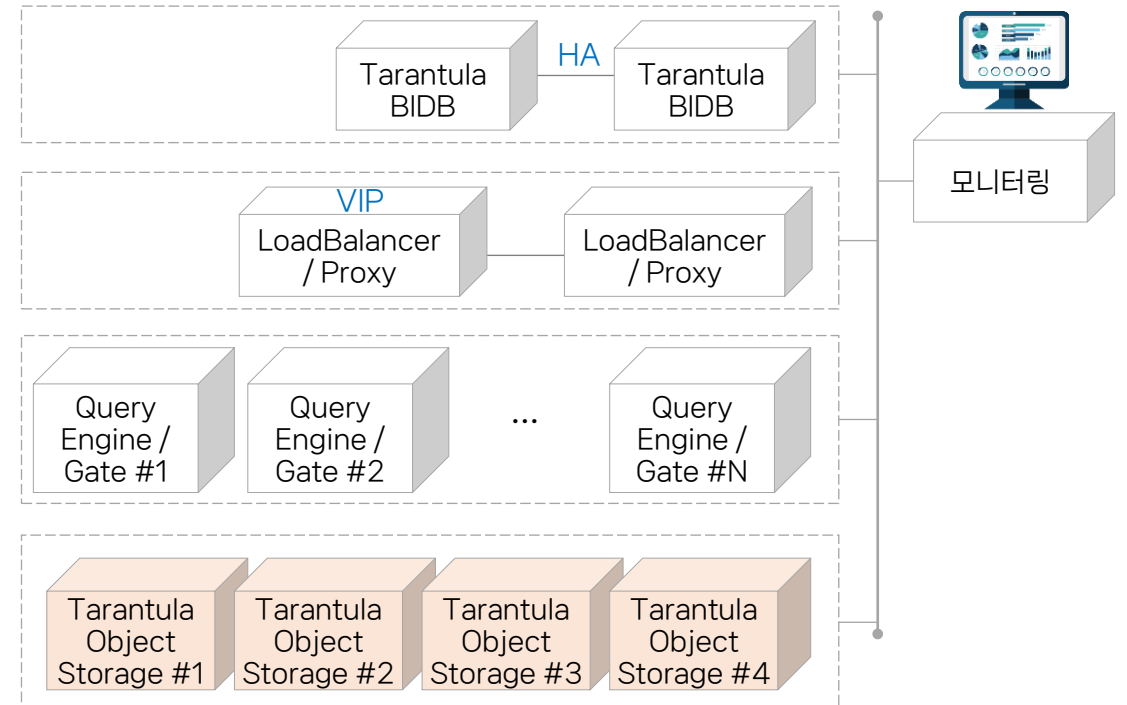
단계적 확장을 고려하여 초기 비용 부담을 줄인 최소화 구성도 가능 (Compute-Storage 분리 아키텍처로 필요한 부분만 확장도 가능)

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

권장 구성도



- 성능과 안정성 최대화 구성

8. 서버 구성 (권장 Spec)

※ 서버 OS 는 모두 RHEL9 으로 구성. (서브스크립션 필요), Object Storage를 제외한 모든서버는 가상서버로 구성 가능

구분	항목	권장사양	수량	
TarantulaBIDB 서버	프로세서	16 Core	1	- 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	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	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	
Object Storage 서버	프로세서	16 Core	1	- 동일사양 4대 필요 - Data 영역은 별도 Raid 구성 필요 없음 (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 (서버당 NVMe PCIe 4.0 디스크 11개 장착) 100Gbps 초고속 네트워크 지원
	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	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	

9. 라이선스 유형

구분		Enterprise Edition	Standard Edition
용도		금융, 제조, 유통 등 엔터프라이즈용	병원 연구소, 지자체, 중소기업 등 프로젝트 연구 개발용
권장 데이터 용량		50TB 이상 권장	500GB ~ 10TB
시스템 구성	Computing 영역 - MetaDB (타란툴라BIDB) - Query Engine (Trino) - Proxy	이중화 구성 (서비스 연속성, 운영안정성 확보)	단일 구성 (서비스 다운타임 발생 가능)
	Storage 영역 타란툴라 Object Storage	엔터프라이즈 버전 (공식 지원제품)	Community Version (사내 용도만 사용 가능)
기술지원		24 x 7 기술지원 / 4시간 이내	8 x 5 기술지원 / 1Day

10. 기술 비교 – 1) 그린플럼(Greenplum) vs 타란툴라 Lakehouse



항목	Greenplum	타란툴라 Lakehouse	결론
아키텍처	MPP Shared-Nothing, Compute/Storage 결합	Lakehouse, Compute/Storage 분리 (Iceberg+Object Storage)	분리형 우위
스토리지	노드 로컬 디스크(고정)	Object Storage (PB급 무중단 확장)	확장 우위
테이블 포맷	AOCO(자체 컬럼) / Heap	Apache Iceberg (Parquet/ORC)	오픈 표준
쿼리 엔진	GP 자체 엔진 (Postgres 기반)	Trino, Spark, Flink 멀티 엔진	멀티엔진
ACID/시간여행	ACID O / Time Travel X	Snapshot ACID + Branch + Time Travel	Iceberg 우위
스키마 진화	DDL, 일부 비파괴 변경 제한	스키마·파티션 진화 풀 지원	Iceberg 우위
AI/RAG 연계	공식 미흡	타란툴라 AI 벡터DB 카탈로그 공유	전략 우위
라이선스/지원	VMware Tanzu EOL, OSS Closed (2024)	타란툴라 상용지원 + OSS 생태계	리스크 회피

10. 기술 비교 – 2) 클라우데라 vs 타란툴라 Lakehouse



항목	클라우데라 Open Data Lakehouse	타란툴라 Lakehouse	결론
아키텍처 / 스택	Hadoop 계보 풀스택, 무겁고 광범위	Trino + BIDD + Object Storage 경량 모듈형	경량·단순성 타란툴라 우위
스토리지	Apache Ozone / HDFS(자체) 또는 클라우드 오브젝트 스토리지	S3 호환 Object Storage, 고성능 NVMe, PB급 무중단 확장	동등 (타란툴라 S3 표준)
테이블 포맷	Apache Iceberg (Parquet/ORC)	Apache Iceberg (Parquet/ORC)	동등 (오픈 표준)
ACID / 시간여행 / 스키마·파티션 진화	Iceberg 기반 풀 지원	Iceberg 기반 풀 지원	동등
쿼리 엔진	Impala / Hive / Spark / Flink (자사 통합 엔진 중심)	Trino 중심(ANSI SQL·페더레이션)	개방형 엔진 타란툴라 우위
거버넌스 / 보안	SDX 기반 성숙한 엔터프라이즈 거버넌스·lineage	카탈로그 기반 거버넌스+ 온프레임·망분리	충분
AI / RAG 연계	AI Workbench·MCP, Iceberg 데이터 공유	타란툴라 AI 벡터DB 카탈로그 공유	동등 (양사 Iceberg 기반)
라이선스 / 비용 모델	CCU(코어+메모리)·TB·BI 사용자·GPU 단위 구독 → 규모 따라 비용 확장, 외산	OSS 기반 + 정액형 상용지원, 코어/TB 종량과금 없음 → 예측 가능한 비용	비용 타란툴라 우위
운영 부담 / 도입	풀스택 운영 복잡, 전문 인력 필요	경량 구성 + 유엔넷 매니지드, AI로 관리 용이, 빠른 도입	운영 타란툴라 우위
지원 / 데이터 주권	외산 벤더, 글로벌 지원	국산, 국내 직접 지원·한국어 대응, 데이터 주권	국산·주권 타란툴라 우위

감사합니다.

유엔넷 주식회사

문의사항

이메일 info@unnet.co.kr

홈페이지 문의



<https://tarantuladb.com/inquiry/>

유엔넷 연락처

[HQ] 서울 강남구 테헤란로20길 12 안타워 2층

[여의도 Office] 서울 영등포구 여의나루로 71 동화빌딩 7층



02-565-7034



www.unnet.co.kr



www.tarantuladb.com

