

AI 성과를 만드는 데이터 인프라

Transforming industrial data into AI-Ready infrastructure

NUBISON DataLake



Datalake

무엇인가

산업 현장의 데이터를 AI-Ready 상태로 전환하는 자동화 플랫폼
 누비슨 데이터 레이크는 제조·산업 현장의 다양한 원천 데이터를 수집·적재·정제·서빙하는 End-to-End 파이프라인을 제공합니다.
 AI 프로젝트의 가장 큰 병목인 데이터 준비 단계를 자동화하여 기업이 AI 가치를 빠르게 실현할 수 있도록 지원합니다.

AI-Ready Auto Conversion



AI-Ready 자동 변환

비정형 데이터 포함 모든 원천 데이터를 Medallion 아키텍처 기반으로 자동 변환하여 AI 학습·추론에 즉시 투입 가능한 상태로 준비합니다.

- 매뉴얼·SOP·정비 이력 (RAG)
- NUBISON IoT실시간 센서·장비 데이터
- 베테랑 메모·작업일지

End-to-End Pipeline



End-to-End 파이프라인

수집 → 적재 → 정제/변환 → 서빙까지
 파편화된 모든 과정을 단일 플랫폼에서
 유기적으로 통합·자동화합니다.

- 파이프라인 스케줄링·모니터링
- 이기종 시스템 통합 커넥터
- 실시간/배치 처리 동시 지원

Industry-Specific Platform



산업 특화 플랫폼

OT/IIoT 데이터 수집에 최적화된 커넥터와
 ERP·MES·SCM 전용 플러그인 모듈로
 프로젝트 초기 구축 비용을 대폭 절감합니다.

- ISO 62264 제조 표준 준수
- PLC·센서·SCADA 직접 수집
- 산업 현장 검증 100+ 프로젝트

AI 프로젝트 도입은 모델이 아니라 데이터 문제

데이터가 없어서가 아니라, 수집·연계·가공·접근의
전 구간에 걸친 구조적 문제 해결이 우선입니다.

데이터 수집

이기종 시스템 연계

문제점

- ERP·MES·위키 등 파편화된 소스, 공통 스키마 미정으로 시스템 간 연계 불가
- 스케줄 관리·모니터링 부재로 수집 안정성 확보 어려움

NUBISON 해결방향

- 공통 스키마 기반 수집 체계 표준화 및 이기종 시스템 안정적 연계 구축
- 수집 파이프라인 모니터링·장애 대응 체계 마련

데이터 적재

확장형 원천 데이터 보관

문제점

- 무결성·확장성 중심 적재 체계 미흡, 대용량·이기종 데이터 저장 한계
- 메타데이터 기반 이력 및 추적성 미확보

NUBISON 해결방향

- 원천 데이터 무결성·확장성 중심 적재 구조 구축
- 메타데이터 관리 기반 계층별 데이터 추적성 확보

데이터 정제/변환

표준화·정합성 보장

문제점

- 품질 확보를 위한 정제·변환 파이프라인 부재, 레거시 API 연계 미비
- 데이터 의미 불명확으로 AI 모델 정확도 저하

NUBISON 해결방향

- 정합성 보장 정제·변환 시스템 구축 및 AI-Ready 품질 자동 검증
- 레거시 시스템 API 연계 표준화 구조 마련

데이터 서빙

AI 학습·추론 데이터 공급

문제점

- AI-Ready 품질 검증 단계 없음, Feature Store 부재로 피쳐 재사용 불가
- 계층별 학습 데이터셋 공급 체계 미구축

NUBISON 해결방향

- Feature Store 구축·버전 관리 및 계층형 AI 학습 데이터셋 공급 체계
- 실시간 추론 데이터 공급 파이프라인 설계

Ingest to Insight

단 하나의 파이프라인

수집부터 AI 가치 창출까지 파편화된 모든 과정을 유기적으로 통합합니다.

step 01

데이터 소스

IoT · MES

설비 · 공정 실시간 데이터

ERP · SCM

비즈니스 핵심 데이터

문서 · 이메일

비정형 문서 데이터

이미지 · 영상

비정형 미디어 데이터

step 02

NUBISON Datalake

Bronze Layer

원본 데이터

원형(Raw) 보존 · 언제든지
재가공 가능한 핵심 자산

Silver Layer

AI-Ready Data

결측 · 이상치 자동 보정
즉시 모델링 투입 가능

Gold Layer

즉시 반영 데이터

비즈니스 로직 결합
고차원 지표 · 즉각 의사결정

step 03

Ops

MLOps

AI 모델 학습 · 배포 · 성능 관리
CI/CD 자동화

LLMOps

대규모 언어모델 학습
프롬프트 · 추론 통합 운영

step 04

비즈니스 인사이트

· 예측 분석

불량 · 장애 사전 예측

· AI Agent

자율 의사결정

· 경영 인사이트

KPI 자동 리포트

· PLC 제어

자율제조 기반 설비제어

 파편화된 모든 과정을 하나의 파이프라인으로 통합 · 자동화하여, AX의 가치를 완성합니다.

Datalake

6가지 핵심 기능 모듈

산업 데이터의 수집부터 AI배포까지 전 주기를 커버하는 통합 기능을 제공합니다.

Multi-Source Ingest

멀티소스 수집 · 30+ 커넥터

DB · S3 · MinIO · API · Webhook · SFTP · 이메일
첨부부터 이미지 · 영상 · 음성 비정형까지, RPA 기반
자동 수집과 공통 스키마 표준화를 적용합니다.

30+ 커넥터

RPA 자동화

스키마 표준화

상태 모니터링

Real-time Quality

실시간 품질 관리 · AI-Ready 검증

결측 · 중복 · 이상값을 자동 탐지하고 도메인 규칙 기반
품질 점수로 AI-Ready 합격 기준을 자동 판정합니다.
변환 전후 품질을 추적합니다.

자동 탐지

품질 점수

프로파일링

전후 비교

Medallion Transform

Medallion 변환

형식 변환 · 결측 · 중복 제거, 다중 소스 결합 · 집계,
도메인 코드 표준화에 더해 비정형 데이터의 자동
구조화와 텍스트 정제 · 벡터화를 수행합니다.

코드 표준화

자동 구조화

벡터화

자동 구조화

Data Governance

데이터 거버넌스 · 보안 · 규정 준수

결측 · 중복 · 이상값을 자동 탐지하고 도메인 규칙 기반
품질 점수로 AI-Ready 합격 기준을 자동 판정합니다.
변환 전후 품질을 추적합니다.

마스킹

IAM

Lineage

ISO 62264 · 5259

Feature Store

AI 학습 데이터 허브

피처 등록 · 버전 관리, 모델별 학습 데이터셋 공급,
재사용 · 공유 허브와 Point-in-time 정합성으로
배치 · 실시간 피처를 서빙합니다.

버전 관리

피처 재사용

PIT 정합성

드리프트 감시

LLMOps

AI 에이전트 운영

RAG 기반 의미 검색 파이프라인, Vector DB 연동 ·
임베딩 관리, 프롬프트 버전 · 성능 관리, 산업 특화 LLM
파인튜닝과 응답 품질 모니터링을 지원합니다.

RAG 검색

Vector DB

프롬프트 관리

LLM 파인튜닝

Datalake

특장점 및 가치

AI-Ready? You are Ready! 준비된 데이터가 현장의 해답을 만듭니다.

도입 효과 | 하나의 파이프라인이 만드는 변화

70%↓

Time to Value

AI 가치 창출 시간 단축

∞

Scalability

무한 확장성 · AI 즉시 대응

1

Single Truth

전사 단일 기반 의사결정

30+

Connectors

ERP · MES · IoT · 문서 등

OT Data Strength

OT 데이터 수집 강점

물리(OT) 데이터 수집에 강한 산업 특화 데이터레이크입니다. OT/IIoT 프로토콜의 다양성과 제약을 극복하고, PLC · 센서 · SCADA의 데이터를 직접 수집하여 공정 · 설비 · 환경을 통합 분석하고 ERP · MES · SCM 플러그인을 즉시 적용합니다.

AI Auto Transform

AI 기반 변환 자동화

데이터 변환 단계에 AI 기반 자동화를 적용합니다. 복잡한 전처리 작업을 최소화하고, 이미지 · 문서 등 비정형 데이터를 AI로 자동 처리하여 기존 스크립트 대비 효율을 대폭 높이고 AI-Ready 데이터를 신속하게 확보합니다.

System Plugins

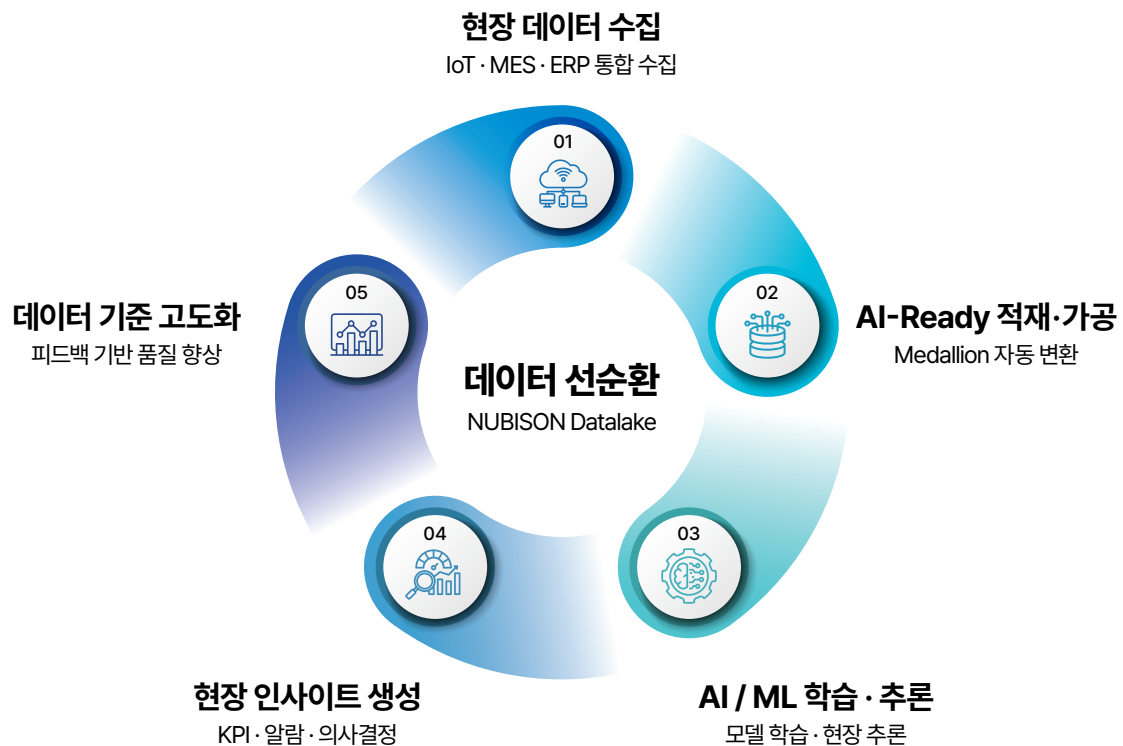
ERP · MES · SCM 플러그인

주요 업무 시스템 전용 커넥터 · 템플릿을 제공합니다. 인터페이스 분석과 ETL 개발 비용을 절감하고, AI-Ready 데이터 기반으로 빠르게 구축하며, 100+ 산업 프로젝트에서 검증된 ISO 62264 · 5259 국제 표준을 준수합니다.

Datalake

데이터 선순환

AI가 진화할수록 더 가치 있는 데이터가 만들어집니다.



BEFORE vs AFTER | 데이터레이크 도입 전후 비교

구분	데이터레이크 도입 전	데이터 레이크 도입 후
AI 가치 창출 속도	AI 가치 창출 속도	AI-Ready 데이터 수주 내 확보
데이터 품질 관리	데이터 품질 관리	AI 기반 자동 검증 · 이슈 감지
현장 AI 연동	현장 AI 연동	단일 파이프라인 즉시 연동



SimPlatform
www.simplatform.com

(주)심플랫폼

(본사) 서울시 금천구 가산디지털1로 226 에이스하이엔드타워5차 2003호

(AX본부) 서울시 강남구 삼성로 555 알앤텍빌딩 4층

Tel. **02-6405-4338** | email **ax@simplatform.com**