

DATABRICKS FOR SMART MANUFACTURING: IOT PRODUCT SPEC

Databricks provides a unified Data Intelligence Platform designed to handle the high-velocity, high-volume data streams typical of modern manufacturing environments.



CORE ARCHITECTURE



Delta Live Tables (DLT)

Automates ETL pipelines for real-time sensor data.



Unity Catalog

Centralized governance for all IoT assets and ML models.



Mosaic AI

Tools to build, deploy, and monitor predictive maintenance models.



Photon Engine

High-performance vectorization for fast time-series analysis.



MANUFACTURING USE CASES



Predictive Maintenance

Reducing unplanned downtime by 25%+



Quality Assurance

Using Computer Vision to spot defects on the line.



Energy Optimization

Analyzing power consumption patterns across shifts.



Supply Chain Visibility

Integrating ERP data with real-time logistics.



KEY TECHNICAL FEATURES

1

Real-Time Data Ingestion

- **Auto Loader:** Efficiently ingests files from cloud storage (S3/ADLS).
- **Structured Streaming:** Low-latency processing for MQTT/Kafka streams.
- **Schema Evolution:** Handles sensor firmware updates without breaking pipelines.

2

Time-Series Intelligence

- **ASOF Joins:** Precisely aligns telemetry data from disparate machines.
- **Window Functions:** Calculates moving averages for anomaly detection.
- **Native Geospatial:** Tracks assets across logistics and factory floors.

3

Edge-to-Cloud Integration

- **Delta Sharing:** Securely shares production data with vendors/partners.
- **MLflow:** Tracks model versions from lab testing to factory deployment.
- **Lakehouse Monitoring:** Detects data drift in sensor accuracy over time.



SECURITY & COMPLIANCE



SOC 2 & HIPAA: Meets global security standards.



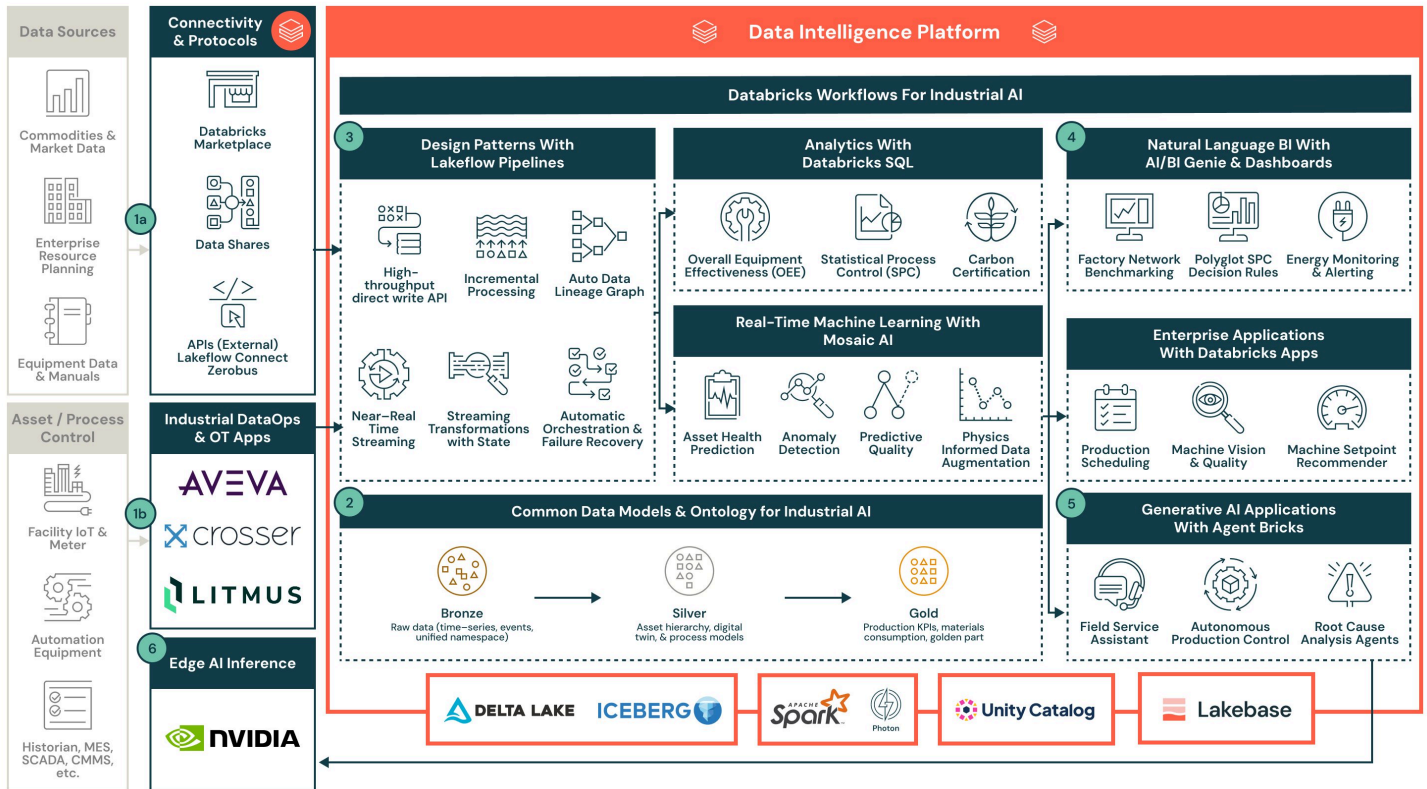
Role-Based Access (RBAC) & Attribute-Based Access (ABAC): Restricts data access to specific roles or based on policies combining specific attributes.



End-to-End Encryption: Protects sensitive proprietary manufacturing IP.

INDUSTRIAL AI REFERENCE ARCHITECTURE

Integrating operational technology (OT) with Databricks requires a robust bridge between factory protocols and cloud-scale analytics.



DATA ACQUISITION LAYER



Field Connectivity:
PLCs (Siemens, Rockwell) communicate via **OPC-UA** or **Modbus**.



Edge Gateway:
Hardware like [Azure IoT Edge](#) or [AWS IoT Greengrass](#) converts protocols.



Message Broker:
Data is published to [Confluent Kafka](#) or [Azure Event Hubs](#) for buffering.



THE MEDALLION PIPELINE



Bronze (Raw):
Databricks Auto Loader ingests raw JSON/Protobuf streams from cloud storage.



Silver (Cleaned):
Delta Live Tables filters noise, deduplicates signals, and enforces sensor schemas.



Gold (Business):
Aggregated metrics (OEE, cycle times) ready for [Power BI](#) or [Tableau](#) visualization.



INTEGRATION COMPONENTS



Databricks SQL:
Provides sub-second query performance for plant manager dashboards.



Unity Catalog:
Manages permissions so "Plant A" cannot see "Plant B" sensitive data.



Model Serving
Deploys [MLflow](#) models to predict motor failure before it happens.