PRODUCT · OPERATIONAL ML PLATFORM

Aizen Foresight

An end-to-end platform for building, deploying, and scaling production ML on real-time data — with full lineage and auditability from ingestion to inference.

<1 ms	<50 ms	<100 ms	4×
FEATURE RETRIEVAL	PREDICTION LATENCY	FRAUD SCORING P99	FEATURE STORE LAYERS

§ 01

Platform overview

01 / 06

Foresight is a no-code, streaming-first platform built on a containerized service mesh. Visual pipeline construction in ML Studio coordinates ingestion, feature engineering, AutoML, and serving on a single canvas. Unified feature logic eliminates training-serving skew. All execution is governed: every transformation, model invocation, and prediction is traceable through the integrated metadata store.

§ 02

Core pipeline stages

02 / 06

STAGE	SERVICE	CAPABILITY
01	**Data Collection**	Real-time and batch ingestion from databases, Kafka, REST APIs, and cloud object storage. Sub-second streaming and scheduled batch jobs configured visually.
02	**Feature Engineering**	Streaming and historical feature computation with unified training/serving logic. Eliminates training-serving skew through a single transformation code path.
03	**Feature Store**	Four-layer architecture: Memory, Offline, Online, Metadata. Apache Arrow columnar format for zero-copy reads and sub-millisecond lookups.
04	**AutoML**	Automated algorithm selection and hyperparameter tuning across gradient boosting, random forests, SVMs, FFN, RNN, and 1D CNN architectures. One-click GPU training.
05	**Serving**	Containerized model deployment with REST endpoints, version management, and built-in monitoring for latency, throughput, and error rates.

§ 03

Feature store · architecture

03 / 06

MEMORY In-memory cache, Apache Arrow columnar format. Sub-millisecond retrieval for live inference. Distributed, fault-tolerant.

OFFLINE Schema-flexible row tables for raw and computed features. Native AWS S3 and Google Cloud Storage integration.

ONLINE Time-series and aggregated features in event-table format, indexed by timestamp plus optional entity keys. Powers real-time inference.

METADATA Feature definitions, lineage relationships, and data quality records. Key-value tables for fast discovery and dependency search.

FEATURE SERVE RPC endpoint for real-time retrieval. Integrates with model inference for personalization, anomaly detection, and optimization in request paths.

UNIFIED LOGIC Same transformation code generates features at training and serving time — preventing skew in production.

§ 04

ML Studio · pipeline workflow

04 / 06

STEP	NODE	FUNCTION	OUTPUT
01	**Data Source**	Connection, schema, ingestion mode, and refresh frequency. DBs, Kafka, REST APIs, object storage.	`ingest.stream`
02	**Data Sink**	Routes ingested data to Offline, Online, or external destinations on a defined schedule.	`sink.persist`
03	**Dataset**	Defines four feature types: basis, aggregate, join-key, expression. Validated before training.	`dataset.v1`
04	**Training**	ML or DL training paths. Cost estimates surfaced before submission. Auto-versioned models.	`model.v1`
05	**Real-Time Source**	Connects events-based sinks to live data so inference features reflect current state.	`stream.live`
06	**Prediction Deployment**	Provisions container, exposes REST endpoint, surfaces health metrics, validates with sample requests.	`predict.api`

§ 05

Integration · runtime

05 / 06

API CONTRACT

`POST /predict`

REQUEST JSON · feature key-value pairs

RESPONSE JSON · prediction + confidence score

AUTH Bearer token / API key

LATENCY < 50 ms end-to-end

Foresight handles feature retrieval, model inference, and response formatting transparently — making predictions consumable by any web, mobile, microservice, or pipeline application.

RUNTIME

CONTAINER Every job — feature compute, training, serving — runs identically in dev and prod.

ORCHESTRATION Kubernetes manages scheduling, scaling, and fault recovery. Foresight abstracts cluster operations.

GPU Major cloud GPU clusters. UI-driven instance selection with cost estimate before submit.

STORAGE AWS S3 and Google Cloud Storage. Apache Arrow zero-copy interchange.

DEPLOY Public cloud or on-premises for regulated and sovereign environments.

§ 06

Deployments · industry

06 / 06

INDUSTRY	APPLICATION	PROFILE
Financial Services	Real-time fraud detection on transaction streams.	`<100ms · Kafka · in-memory`
Telecommunications	Predictive churn and network anomaly detection.	`CDR streams · time-series`
Manufacturing	Predictive maintenance on equipment sensor streams.	`1D CNN · vibration · temp`
Retail / E-Commerce	Dynamic pricing and personalized recommendations.	`clickstream · inventory`
Healthcare	Patient risk stratification on EHR data.	`batch · daily retrain`
Logistics	Route optimization and demand forecasting.	`batch + traffic events`

CONTACT

Schedule a technical demonstration.

Request sandbox access, or engage the Aizen solutions team for a proof-of-concept tailored to your use case.

aizencorp.com

sales@aizencorp.com