DataCore (2026)

Technical Product Brief

DataCore is LatentWorlds' data, operations, and intelligence backbone for robotics deployments. It connects robots in the field to a unified platform that captures multi-modal data reliably, makes it usable through synchronized retrieval and reproducible datasets, and supports the operational workflows required to run fleets safely and efficiently.

What DataCore delivers

DataCore is designed as an integrated platform with four product surfaces that reinforce one another.

Data plane (DataCore): reliable edge-to-cloud ingestion, robotics-native cataloging, and synchronized retrieval for debugging and model development.

Processing plane: reproducible pipelines that transform raw recordings into curated, versioned datasets and derived artifacts with traceable lineage.

Ops plane: fleet monitoring, logs and event correlation, incident workflows, and teleoperation primitives with auditability and safety constraints.

Intelligence plane: partner-driven modules that increase yield per fleet hour—automation for QA, episode extraction, anomaly detection, and assistive operational tooling.

A web console exposes the workflows where a UI improves day-to-day work, while APIs remain first-class for integration with your autonomy and ML stack.

How the system works

On the robot, an ingest client buffers and uploads prioritized streams with resumable transfers and explicit backpressure control. In the cloud (or your VPC/on-prem

deployment), DataCore persists raw payloads, indexes them with robotics context, and exposes retrieval, processing, and operations interfaces.

The platform is organized around **recording sessions** (start/stop, manifests, provenance, lifecycle state) and **streams** (time-bounded channels such as video, lidar, IMU, state, events). This structure supports two critical loops: rapid incident investigation (find → slice → replay) and repeatable dataset production (select → curate → version → export).
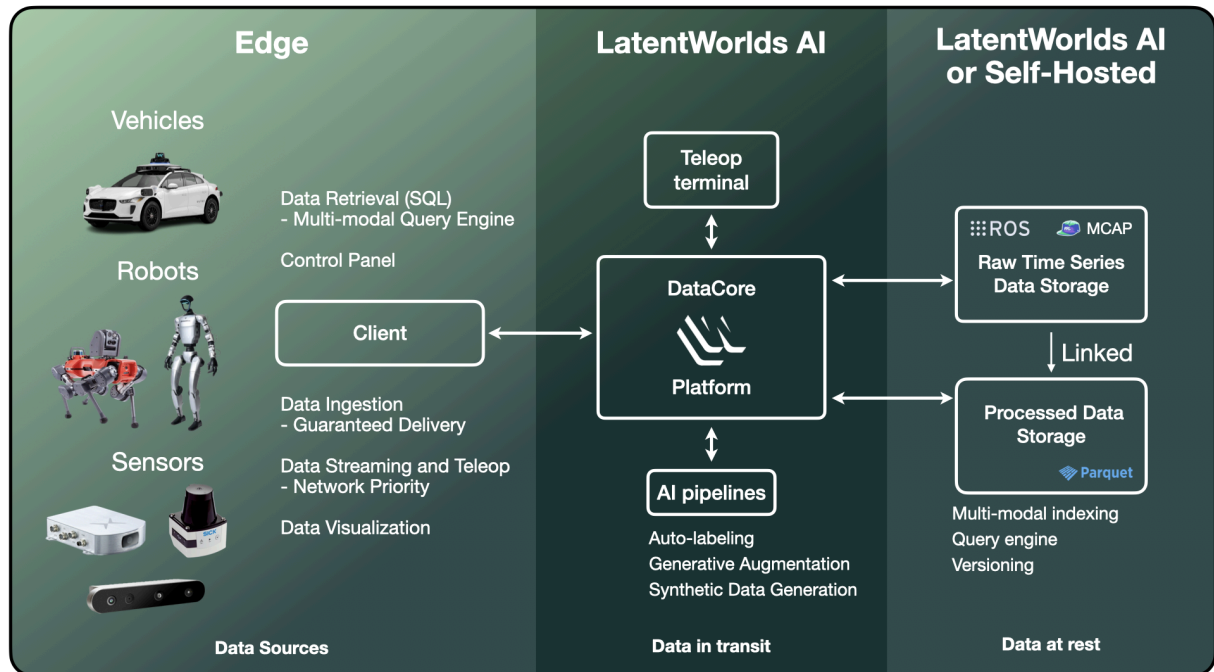
Core capabilities

Data plane: ingestion, cataloging, and retrieval

DataCore is built for real deployment conditions: intermittent connectivity, constrained edge resources, and high-volume multi-sensor payloads.

Ingestion supports high-throughput uploads for binary payloads and metadata, client-side buffering with resumable transfers, and server-side throttling that propagates backpressure cleanly to the edge. Streams can be prioritized so critical telemetry remains reliable when bandwidth is constrained. Device identity and enrollment are handled with certificate-based mutual TLS and project-scoped credentials, with lifecycle controls for rotation and revocation.

Cataloging preserves operational context so recordings remain searchable and usable over time. Recordings can be queried by robot identity, session, time window, stream/channel, software/config versions, mission/environment tags, and labels. DataCore also supports incremental robotics semantics commonly required at scale, such as calibration artifacts and time-alignment policies.

Retrieval is designed around synchronized slices rather than files. Users can request a time window across multiple streams and receive aligned, streaming-friendly responses suitable for replay, debugging, and construction of training clips. Export paths support interoperable formats aligned with common robotics and ML workflows (for example, structured exports such as Arrow/Parquet and robotics packaging such as MCAP/ROS bag, where appropriate).

Processing plane: reproducible pipelines and dataset versions

DataCore turns raw recordings into curated, repeatable datasets.

Processing runs generate versioned outputs with explicit lineage back to source sessions and selection rules. Runs are observable and idempotent, with well-defined retry behavior and failure modes. Customer-defined transforms can be executed in an isolated processing environment with auditability. Dataset versions are intended to be stable inputs to training and evaluation, reducing ad-hoc glue code and improving reproducibility across teams.

Ops plane: fleet observability and controlled teleoperation primitives

DataCore includes operations primitives that connect production events to the underlying evidence.

Fleet surfaces provide per-robot status and health signals, structured logs, and event correlation to recording sessions and dataset artifacts. Incident workflows support actions like initiating targeted debug recordings, collecting snapshots, and quickly

retrieving synchronized evidence around incidents. Teleoperation primitives are designed to be auditable and safe-by-default, with operator identity, scoped permissions, session lifetimes, rate limits, and an immutable audit trail.

Intelligence plane: yield-boosting automation

DataCore supports intelligence modules that sit on top of the same data, processing, and ops foundations.

These modules are intended to reduce manual work and increase the signal captured per fleet hour: automated QA checks, rule-based and learned episode extraction, anomaly detection for triage, and assistive tooling that improves operator effectiveness during teleop and incident response.

Typical customer workflows

Incident investigation

When an incident is reported or detected, engineers locate the relevant session and time window using metadata (robot, mission tags, software version, labels). DataCore then streams a synchronized slice across the required sensors and logs so the behavior can be reproduced in your replay and debugging stack. The same evidence can be promoted into a dataset version for regression tests or model iteration.

Dataset creation

Teams define selection rules (robots, environments, time ranges, labels, failure modes) and generate a dataset version with provenance. Curation pipelines can align, decode, QA, sample, and export artifacts into the training and evaluation stack, while preserving lineage back to the raw sources.

Fleet operations and teleop

Operators use the console to monitor fleet status, review recent incidents and recordings, and initiate controlled actions such as debug recordings or teleop sessions. Audit trails and access controls provide accountability and reduce operational risk.

Deployment and integration

DataCore integrates through a robot-side ingest SDK (Python and Rust), control-plane APIs (sessions, manifests, identity, governance), retrieval APIs (synchronized slice streaming and export), processing interfaces (pipelines, artifacts, dataset versions), and an operations surface (fleet status, logs, incidents, teleop sessions). A web console provides an opinionated workflow layer for daily use.

Deployment options include managed cloud, customer VPC/hybrid, and on-prem components when required by connectivity constraints or regulation. Data residency, retention, and deletion policies are configurable by deployment model.

Product maturity and roadmap alignment

DataCore is being delivered in phases, with APIs designed to remain consistent as higher-level workflows are introduced.

Near term (pilot readiness): hardening robot enrollment and ingestion, consistent semantics for sessions/streams, initial synchronized retrieval, and a console for projects and recordings.

Mid term (datasets and processing): dataset catalog with versioning and lineage, exporters into robotics and ML toolchains, and a processing plane with observable pipeline runs and durable artifacts.

Expansion (ops and intelligence): fleet observability primitives, incident response workflows, teleop session management with safety constraints, and partner-driven intelligence modules (QA, episode extraction, anomaly detection, assistive ops).

Evaluation criteria

A typical evaluation connects one representative robot end-to-end and validates three outcomes:

1. **Ingestion reliability** under real connectivity conditions (success rate, partial uploads, missing streams).
2. **Incident workflow speed** from incident → synchronized slice → replay.

3. **Dataset repeatability** with clear lineage from source sessions to exported dataset versions.

A production expansion plan typically adds ops-plane requirements (fleet health, log/event correlation, auditability) and a path to intelligence modules that reduce manual triage and curation.

Fit

DataCore is a strong fit for teams operating multi-sensor robots outside controlled networks who need synchronized retrieval, reproducible datasets, and production operations tooling—without building and maintaining bespoke ingestion, indexing, export, monitoring, and teleop infrastructure.