XaasIO AI Lake Platform

Enterprise Hadoop Data Platform, Extended for AI

Bring enterprise data engineering, SQL analytics and AI data preparation onto an integrated Hadoop ecosystem. Build on open storage, shared metadata and familiar processing engines, with architecture, deployment and support from XaasIO. Retain control of data location and operating choices as workloads evolve.



© 2026 XaasIO. All rights reserved.

PLATFORM OVERVIEW

Ingest - Connect Enterprise Data

Bring database changes, application events and files into repeatable pipelines. Use NiFi for data flows and Kafka for durable event streams.

Process - Create Usable Data

Transform datasets with Spark, process streams with Flink and expose analytical tables through Hive and Trino. Allocate shared cluster resources with YARN.

Govern - Control Access and Use

Establish authentication, authorization, metadata and audit controls across the selected services. Make access policies, data ownership and retention part of the platform design.

DEPLOYMENT ARCHITECTURE

Storage - HDFS and S3 Interfaces

Use HDFS distributed storage, Ozone object storage or an integrated Ceph S3 tier. Select storage services to match capacity, access and protection requirements.

Resources - YARN and Kubernetes

Schedule Hadoop processing with YARN queues and resource policies. Use XaasIO Kubernetes for selected containerized services alongside the cluster, where required.

Deployment - Your Operating Environment

Deploy on bare metal or validated VMs in private, sovereign or provider environments. Define availability, recovery and network boundaries during design.

ONE DATA FOUNDATION | HADOOP, ANALYTICS AND AI

Combine storage, processing, SQL and governance with the extensions your workloads need.

CORE DATA PLATFORM CAPABILITIES

Build a validated data platform for batch analytics, streaming, operational applications and AI data preparation.

Distributed Storage and Open Formats

Store raw and curated data in HDFS or S3-compatible Ozone. Use open formats such as Parquet and ORC for analytical datasets.

Batch Engineering and Data Preparation

Use Spark for cleansing, enrichment and feature preparation. Build repeatable transformations with familiar programming interfaces and shared cluster resources.

Streaming and Event Processing

Retain and distribute events with Kafka. Use Flink for stateful stream processing, event-time handling and windowed analysis over changing data.

SQL Analytics and Federation

Query distributed data with Hive and federate sources through Trino. Connect BI tools and notebooks to the analytical services selected for your deployment.

Open Tables and Time Travel

Use Iceberg tables for schema evolution, snapshots and time travel. Maintain table and partition metadata through Hive Metastore and the selected catalog.

Operational Data and Metadata

Use HBase for random reads and writes, with Phoenix for compatible SQL workloads. Use Atlas for cataloging, classifications and supported lineage integrations.

Workflow Orchestration

Schedule and monitor pipelines with Airflow. Define dependencies, retries and operational checkpoints so teams can manage recurring processing as code.

Search and AI Extensions

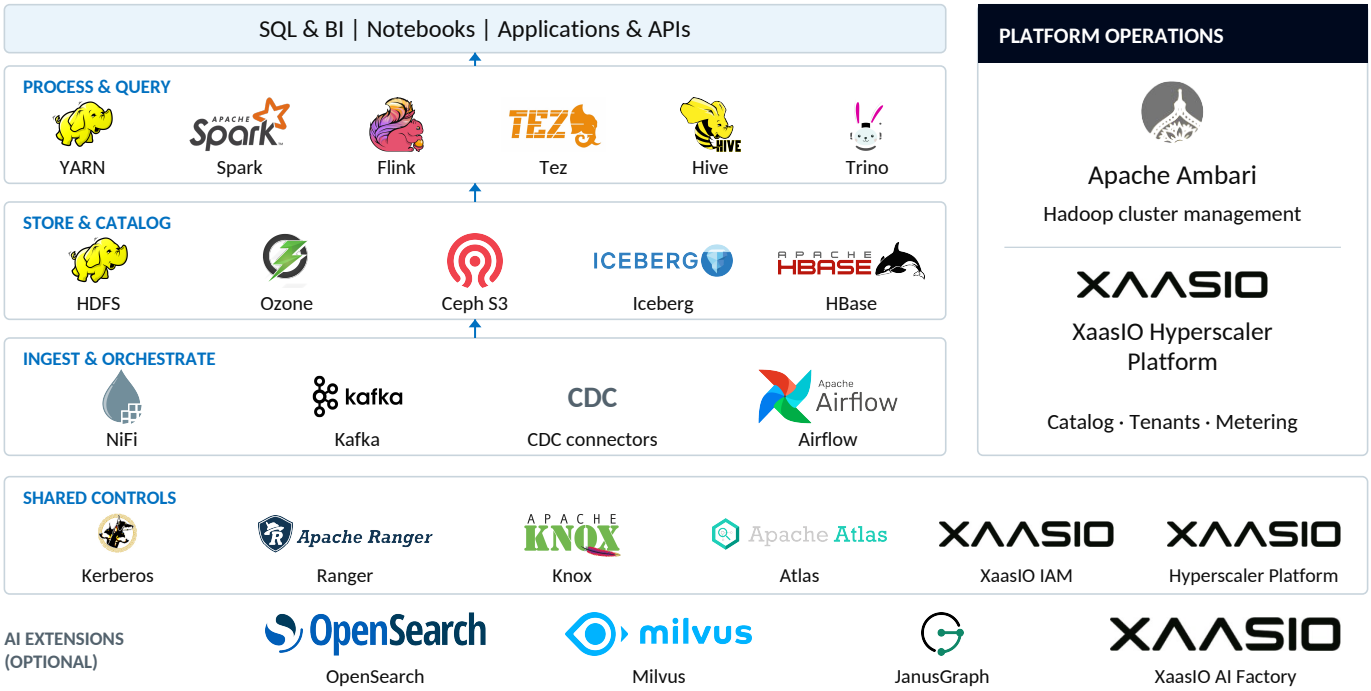
Add OpenSearch for keyword and vector retrieval, Milvus for embedding search, or JanusGraph for connected entities. Integrate private AI through XaasIO AI Factory.

SHARED CONTROLS AND CLOUD INTEGRATION

Combine Kerberos, Ranger, Knox, Atlas and XaasIO IAM with XaasIO Hyperscaler Platform for service catalogs, tenants, roles, quotas and usage metering. Integrate XaasIO MLT for monitoring, logging and telemetry according to the approved architecture.

© 2026 XaasIO. All rights reserved.

REFERENCE ARCHITECTURE



PLATFORM OPERATIONS

Apache Ambari - Cluster Management

Provision supported Hadoop services, manage configurations and service state, and inspect cluster health through the management interface and REST APIs.

XaasIO Hyperscaler Platform

Connect AI Lake to the shared console, catalog, accounts, roles and quotas. Enable usage metering, showback or provider billing within the integration scope.

XaasIO MLT - Operational Visibility

Integrate monitoring, logging and telemetry for the selected cluster services, pipelines and infrastructure. Align dashboards and alerts with the operating model.

SECURITY AND GOVERNANCE

Enterprise Identity and Gateway Access

Use Kerberos authentication and Knox gateway services with the enterprise directory and XaasIO IAM integrations selected for the deployment.

Data Policy, Catalog and Audit

Apply Ranger authorization through supported integrations. Use Atlas classifications and lineage, and route audit records to the agreed monitoring or SIEM services.

Encryption and Controlled Deployment

Define encryption, key ownership, security zones, repository access and update procedures during design. Validate these controls against the services included in the solution.

FROM DAY 0 THROUGH DAY 2

- **Assess** - Data sources, workloads, growth and migration dependencies.
- **Design** - Storage, processing, identity, availability and recovery.
- **Deploy** - Install, integrate and onboard initial pipelines.
- **Validate** - Functionality, security and operational readiness.
- **Operate** - Monitor, tune, maintain and plan capacity.

SUPPORT AND SERVICES

- Architecture, sizing and validated component selection.
- Installation, integration and Hadoop migration assessment.
- Pipeline validation, workload tuning and capacity reviews.
- Runbooks, knowledge transfer and team enablement.
- Release maintenance, platform support and managed operations.

© 2026 XaasIO. All rights reserved.

DESIGNED FOR ENTERPRISE DATA WORKLOADS

01 Data Engineering and Business Intelligence

Consolidate analytical datasets, standardize transformation pipelines and serve SQL analytics through open tables and familiar tools.

02 Streaming and Operational Analytics

Combine event streams and database changes for telemetry analysis, operational dashboards and applications that respond to changing data.

03 Private AI and Enterprise Search

Prepare datasets for model training and feature generation. Add semantic search and retrieval over enterprise documents through selected AI extensions.

XAASIO AI LAKE PLATFORM

Build an enterprise data foundation that can grow into AI.

Bring together data engineering, SQL analytics, streaming and AI data preparation. Define the storage, governance, integrations and operating model around your workloads, with architecture, deployment and support from XaasIO.

Explore the Platform at xaasio.com/platform/ai-lake/

Book a data platform workshop at xaasio.com/contact/

Learn more at xaasio.com

AMERICAS
+1 650 523 6628

EMEA
+44 20 3031 1074

APAC
+91 93423 61010

© 2026 XaasIO Group. All rights reserved. Information is subject to change. | xaasio.com
Third-party names and logos are trademarks of their respective owners.