BrahmDAX Hub

Advanced Data Infrastructure

What is BrahmDAX Hub?

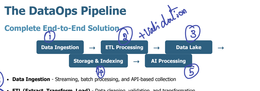
BrahmDAX Hub is a robust DataOps pipeline designed to power:

- **BrahmAI LLM** - Supporting advanced language model training and inference
- **Retrieval-Augmented Generation (RAG)** - Enhancing LLMs with contextual knowledge
- **Comprehensive Data Management** - End-to-end data infrastructure for AI

> **A complete data infrastructure solution for the entire BrahmAI ecosystem**

The DataOps Pipeline

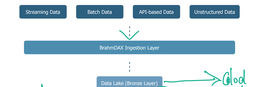
Complete End-to-End Solution



- **Data Ingestion** - Streaming, batch processing, and API-based collection
- **ETL (Extract, Transform, Load)** - Data cleaning, validation, and transformation
- **Three-Layer Data Lake** - Bronze (raw), Silver (validated), Gold (categorized)
- **Storage & Indexing** - Traditional databases and vector search for AI workloads
- **AI Processing** - Fine-tuning, LLM training, and retrieval-augmented generation
- **Data Governance & Security** - Access control, metadata tracking, and monitoring

Data Ingestion Capabilities

Multiple Data Source Support



Streaming Data
- Apache Kafka
- Apache Nifi
- Airbyte

Batch Data
- S3
- Hadoop
- Data Warehouses (BigQuery, Snowflake)

API-based Data
- REST APIs
- Webhooks
- GraphQL

Unstructured Data
- Logstash
- Fluentd
- Elastic Beats

ETL Processing & Transformation

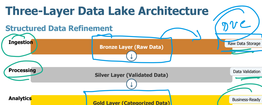
Powerful Data Processing Tools

- **Apache Spark** – Distributed data processing at scale
- **dbt (Data Build Tool)** – SQL-based transformation workflows
- **Apache Beam** – Unified batch & stream processing
- **Pandas / PySpark** – Python-based data manipulation
- **Airflow / Prefect / Dagster** – Pipeline orchestration

> Transforming raw data into valuable insights with industry-standard tools

Three-Layer Data Lake Architecture

Structured Data Refinement



Bronze Layer: Stores raw, unprocessed data from all sources using HDFS, S3, Delta Lake, or Object Storage.

Silver Layer: Cleansed and validated data, ready for processing. Uses Great Expectations, Deequ for data quality checks.

Gold Layer: Curated and structured data optimized for AI and analytics. Stored in PostgreSQL, ClickHouse, Snowflake, or BigQuery.

Data Storage & Indexing

Flexible Storage Solutions

Data Types
- Structured Data
- Semi-Structured
- Unstructured Data
- Vector Embeddings

Storage Solutions
- Relational DBs
- Data Warehouses
- Vector DBs
- Search Indexing

Relational Databases
- PostgreSQL
- ClickHouse

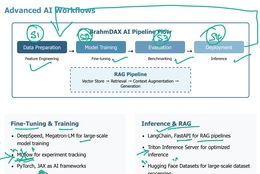
Data Warehouses
- BigQuery
- Snowflake
- Amazon Redshift

Vector Databases
- FAISS
- Weaviate
- Pinecone
- ChromaDB

Search Indexing
- Elasticsearch
- OpenSearch
- LlamaIndex

AI & LLM Processing

Advanced AI Workflows



Fine-Tuning & Training
- DeepSpeed, Megatron-LM for large-scale model training
- MLflow for experiment tracking
- PyTorch, JAX as AI frameworks

Inference & RAG
- LangChain, FastAPI for RAG pipelines
- Triton Inference Server for optimized inference
- Hugging Face Datasets for large-scale dataset processing

> Seamlessly integrating data processing with state-of-the-art AI operations

Data Governance & Security

Enterprise-Grade Protection

- **Metadata Tracking:** DataHub, OpenMetadata
- **Access Control:** Apache Ranger, OAuth
- **Observability & Monitoring:** Prometheus, Grafana, Datadog
- **Data Versioning:** DVC, Delta Lake Time Travel

> Ensuring security, compliance, and traceability across the entire data lifecycle

Complete Technology Stack

Category	Key Technologies
Ingestion	Kafka, Airbyte, Nifi, REST APIs
ETL & Processing	Apache Spark, dbt, Beam, Airflow, Prefect
Validation	Great Expectations, Deequ
Storage	PostgreSQL, ClickHouse, Delta Lake, S3
Vector DB	FAISS, Weaviate, Pinecone, ChromaDB
AI & LLM	DeepSpeed, PyTorch, JAX, MLflow, Triton Server
Governance	DataHub, Apache Ranger, Prometheus, Grafana

Key Benefits

- **Scalable:** Handles massive datasets efficiently from gigabytes to petabytes
- **Flexible:** Supports diverse data types and formats including structured, unstructured, and semi-structured data
- **Integrated:** Seamless workflow from data ingestion to AI processing with minimal friction
- **Governed:** Enterprise-grade security and compliance built into every layer
- **Optimized:** High-performance architecture specifically designed for AI and LLM workloads

> BrahmDAX delivers a complete, enterprise-ready solution for the most demanding AI data requirements

Next Steps

- Technical demonstration and proof-of-concept
- Infrastructure requirements assessment
- Implementation planning and timeline
- Integration with existing systems
- Training and knowledge transfer

> Let's discuss how BrahmDAX can transform your data operations and AI capabilities

Thank You!

BrahmDAX: The complete DataOps solution for BrahmAI LLM