ATLAS DATA SCIENCE

Tech Stack Update

June 2025

Technology Stack

Atlas-DS features a modern, scalable tech stack with React, Node.js, and Python, uses PostgreSQL and Amazon S3 for storage, runs on Docker in AWS, and supports automated deployments with Terraform. It offers strong data governance and search, plus real-time monitoring with AWS CloudWatch and Sentry.

Potential MVP Tech Stack

Core Platform
- React/Next.js: Modern web interface the is fast and polished
- Node.js & Python: lightweight engines powering business logic
- Open MetaData with custom extensions

Data & Storage
- PostgreSQL + PostGIS: Strong relational database for core records
- Amazon S3: Secure cloud bucket for files, images, and back ups

Infrastructure & Operations
- AWS Lambda
- Docker on AWS: Runs code in containers without requiring servers to maintain
- Terraform IaC & GitLab CI/CD: Push button deployment for code

Data Governance & Search
- Open MetaData: Single source of truth for data assets
- OpenSearch Discovery Engine with Geospatial Extensions: Lightning-fast search across the platform

Observability
- AWS Cloudwatch: Real-time monitoring and error alerts

Potential Future Release Tech Stack

Advanced Data & Analytics
- AWS Glue + Apache Spark: High volume data prep and ETL
- Redshift/Lake Formation Lakehouse: Scalable warehouse for dashboards

Real-Time & Streaming
- Amazon Kinesis/Apache Kafka: Live event pipeline for instant insights

AI/ML Enablement
- Amazon SageMaker: managed machine learning to power smart recommendations

Scalability & Multi-Cloud
- Kubernetes (EKS) Auto-Scaling: Elastic orchestration as usage grows
- Azure Extension: Optional second cloud for reach and resilience

Observability & Insights
- Prometheus + Grafana Dashboards: Deep metrics visualized in real-time

Technology Stack

Atlas-DS features a modern, scalable tech stack with React, Node.js, and Python, uses PostgreSQL and Amazon S3 for storage, runs on Docker in AWS, and supports automated deployments with Terraform. It offers strong data governance and search, plus real-time monitoring with AWS CloudWatch and Sentry.

Back-End Development & API
- NodeJS / NextJS: used for back-end development and API endpoints.
- Python: used for general-purpose programming, data processing, automation.
- Java: used for latency-sensitive components.

Containerization & Orchestration
- Docker: containerization using Dockerfiles.

Database
- PostgreSQL: relational database for complex queries and data integrity.
- Elastic: Open search.

AWS Services
- S3: object storage for data.
- Lambda: serverless compute for event-drive functions.
- EC2: virtual server instances for running applications.
- ECR: doctor container registry for storing container images.
- CloudFormation: infrastructure as a code for managing AWS resources.

Big Data & Analytics
- Apache Parquet: columnar storage format optimized for analytics and big data.
- CSV for file formats.

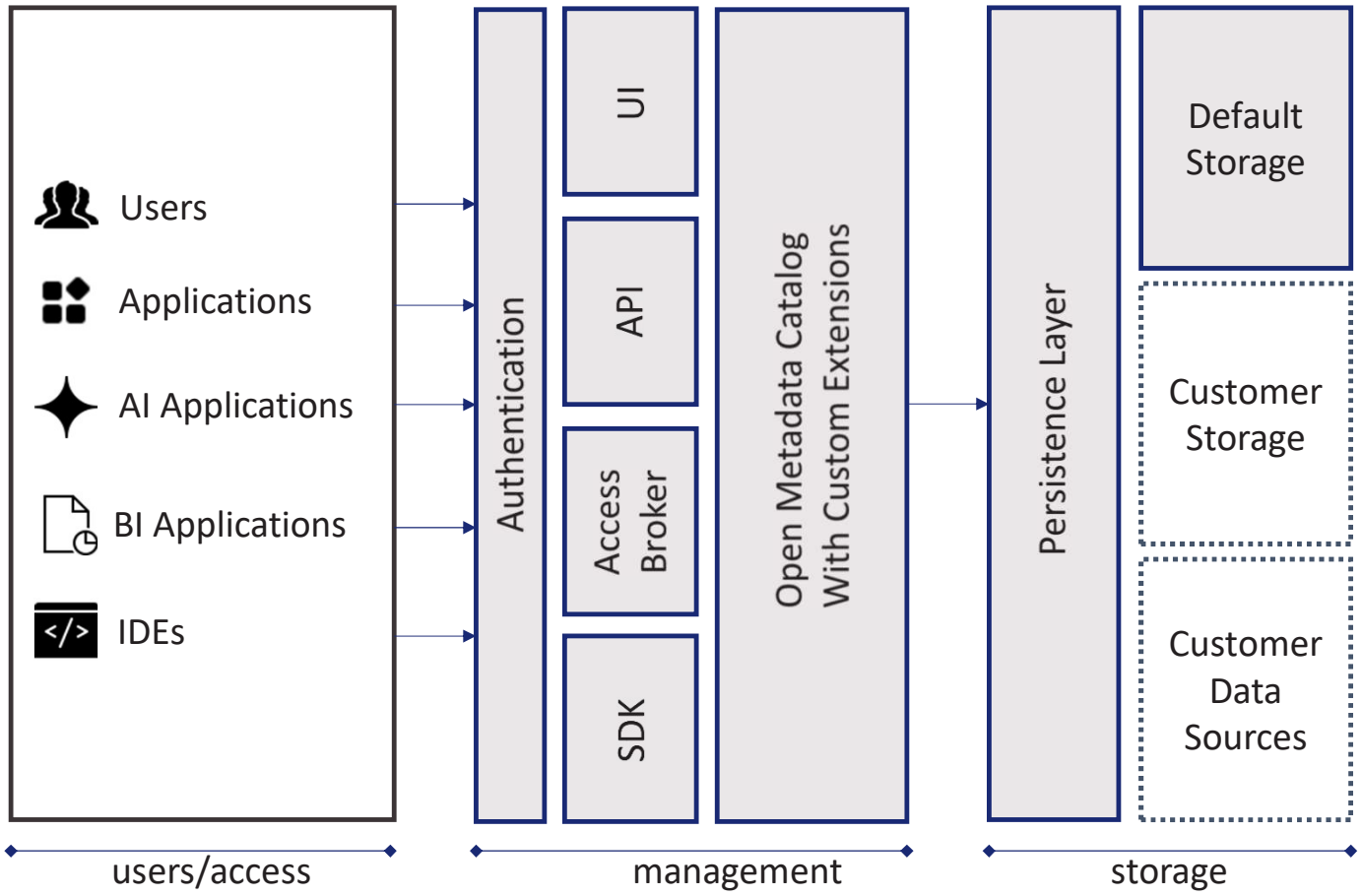
Infrastructure
- Terraform: open-source infrastructure as code for provisioning infrastructure.

Version Control
- GitLab / GitHub: version control and code development and DevOps.

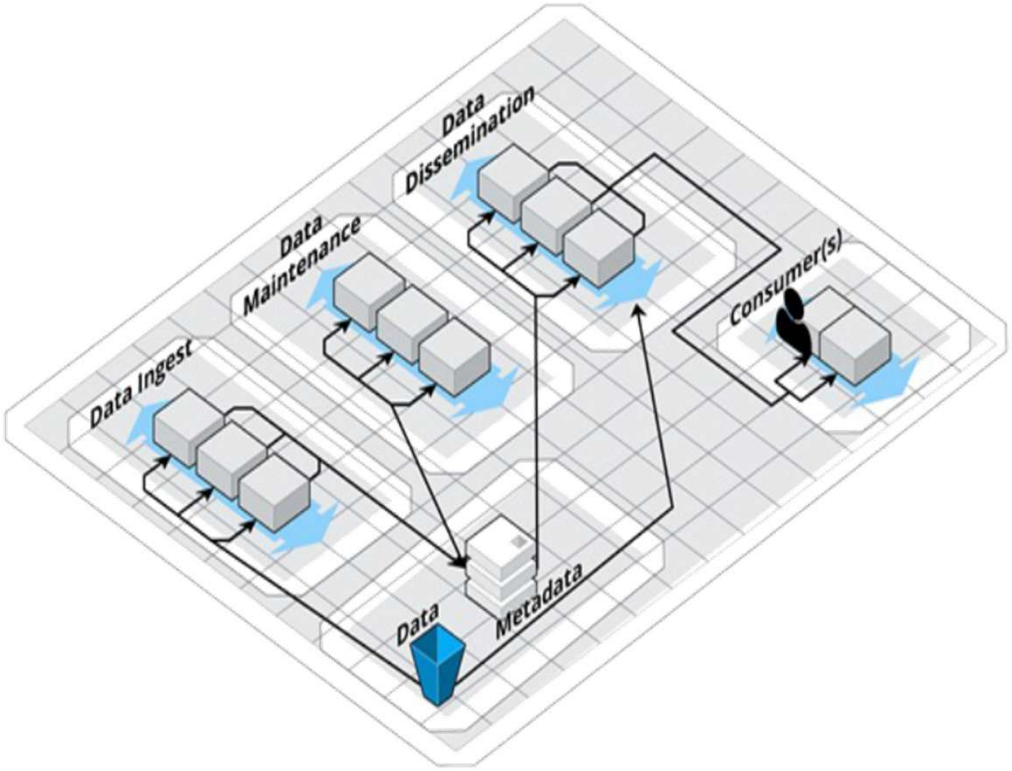
Operating Systems
- Linux: patching, troubleshooting, and security.

Authentication & Authorizations
- Okta, Auth0, KeyCloak: potential identify providers that the platform integrates with customer IDP via OIDC or SAML.

ATLAS-DS Logical Architecture



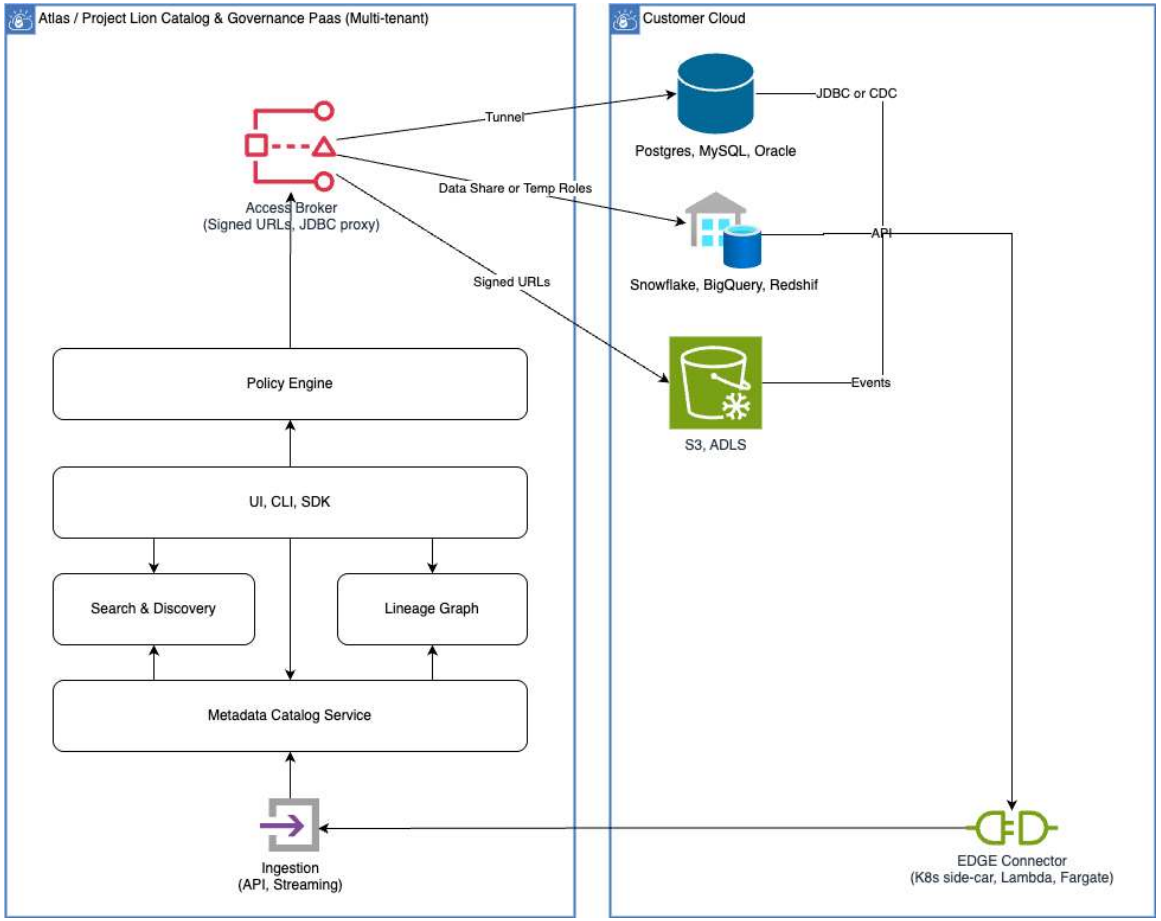
Atlas-DS Functional Overview



Architecture Overview

- Data Ingest
  - Provides the ability to ingest, extract, and relate/match any metadata content/type (digital/non-digital sources) into a Single Metadata Content Library and Data Repository.
- Metadata Catalog
  - Centralized Metadata Catalog for all data holdings of any data product type or format. Provides a central master record continually updated for all data holdings.
- Data Maintenance
  - Provides data stewardship over the data supply chain including data conditioning, data labeling, and data security.
- Data Dissemination
  - Rapid/robust data access for users with the right level of access with built in dissemination/access standards.

ATLAS-DS Functional Architecture



ATLAS-DS Data Workflow

Analysis/measurement of data delivery and workflow utility based on multiple points throughout the data workflow. ATLAS-DS includes automation and performance analysis built into our solution at every design stage that enables measurement of the complete data supply chain from ingest through dissemination, data usage and consumer value derived from data insight, access, and need-to-source matching capabilities.