DRAFT

EDS AI Data Platform: Strategic Data Management for AI

Lifecycle Management, Access Control, Scalable Operations

Enterprise data systems require sophisticated orchestration to support AI initiatives at scale. This presentation outlines EDS's strategic approach to centralizing data management, implementing automated lifecycle controls, and establishing governance frameworks that enable rapid AI development while maintaining operational excellence.

Current State: DIY Fragmentation

Critical Pain Points

Each data science team operates independently, building separate data pipelines from Snowflake to GCP with no standardization or reusability. Manual document handling from S3 creates bottlenecks, while lack of governance means no audit trail exists for compliance requirements.

- Individual teams rebuild identical infrastructure
- No centralized access provisioning or security controls
- Weeks-long delays for basic data access requests
- Zero visibility into data lineage or usage patterns

The Strategic Gap

No single entity manages data lifecycle from creation to deletion. Teams operate in silos without coordination, creating technical debt and compliance risks that compound over time.

EDS as Central Data Platform

Data Products

Curated, AI-ready packages with quality validation and monitoring built-in

Lifecycle Management

Complete document control from HOT to WARM to COLD to DELETE with automated transitions

Access Provisioning

Centralized team request processing with EDS-controlled grant authorization

Data Store Management

Automated creation, scaling, and retirement of storage infrastructure

Platform Architecture Overview

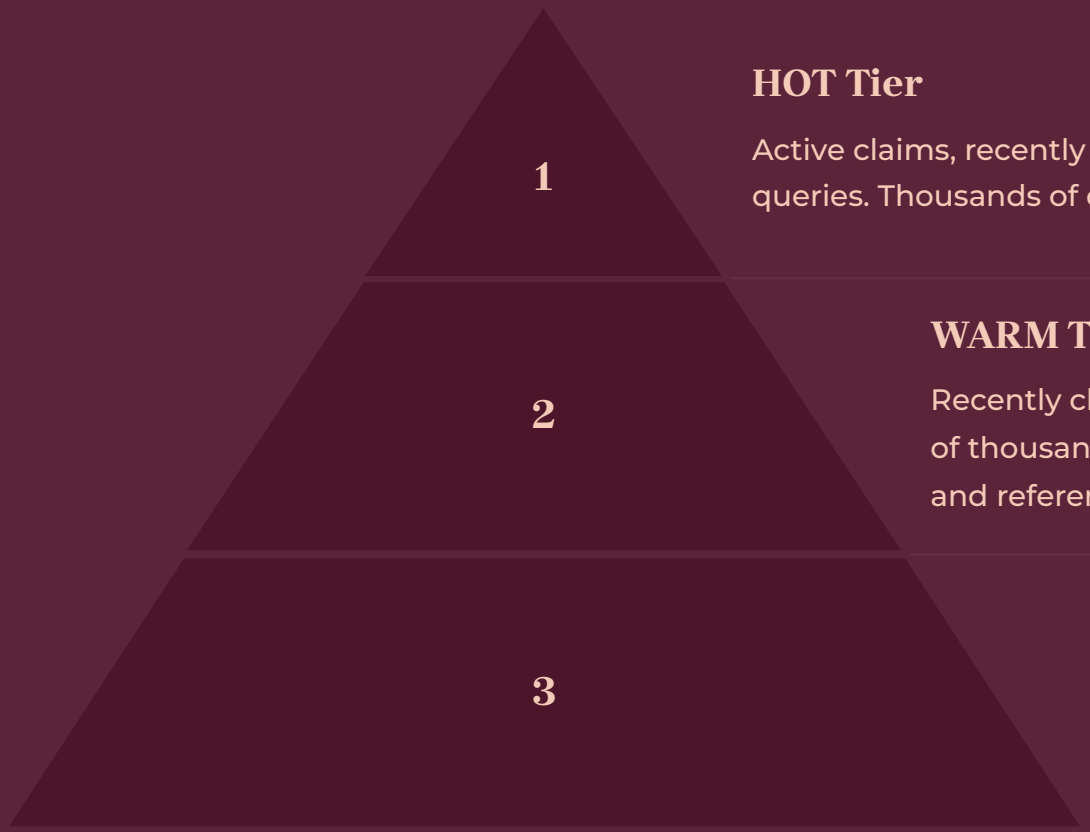
Source Systems: Snowflake (AWS) and S3 (AWS) providing structured and unstructured data inputs

EDS Bridge Layer: Managed pipelines orchestrating data transformation and movement

GCP Destinations: Snowflake replica, Vertex AI Search, and BigQuery optimized for AI consumption

Team Consumption: Data science teams access via standardized APIs with complete audit logging

HOT/WARM/COLD: Complete Lifecycle Control



HOT Tier

Active claims, recently uploaded documents in Vertex AI Search optimized for sub-second queries. Thousands of documents per data store, lasting days to weeks.

WARM Tier

Recently closed claims and training data in standard Vertex AI Search. Hundreds of thousands of documents, retained for months to 2 years for model training and reference.

COLD Tier

Archive storage for old claims and compliance retention. Millions of documents stored for 2-7 years with slower but acceptable access for audit purposes.

Automated Transition Logic

EDS manages all tier transitions automatically: claim closure triggers 7-day countdown to WARM tier, documents over 2 years move to COLD storage, and 7-year retention limits trigger compliant deletion. This approach keeps HOT tier small and fast while managing costs effectively across the entire lifecycle.

Managing Vertex AI Search at Scale

The Fundamental Constraint

Each Vertex AI Search data store performs optimally with fewer than 1 million documents, but enterprise AI initiatives require storing millions of documents total. EDS solves this through intelligent multi-store architecture.

01

Organize by Use Case and Tier

Create separate stores like liability-docs-active (50K docs, HOT), liability-docs-recent (500K docs, WARM), and liability-docs-archive (2M docs, COLD)

02

Monitor Capacity Daily

EDS tracks document counts per data store with alerts triggering at 80% capacity (800K documents)

03

Automated Provisioning

Generate Terraform configuration, create Git PR, provision new Vertex data store, and update routing—all automated within hours

04

Transparent Team Experience

Data science teams query the same way with metadata directing them to appropriate stores—no code changes required

How Teams Get Access to Data Products

1

Platform Team Setup (One-Time)

Creates service account ds-auto-liability@project.iam with minimal permissions for query execution only. This foundational step takes 1-2 hours when teams first join the platform.

2

DS Team Requests Access

Teams submit requests via portal specifying "Need Liability Assessment v2.0" with business justification. The entire request process takes approximately 5 minutes.

3

EDS Automated Granting

System reads access template including BigQuery tables, Vertex stores, and appropriate roles, then applies all permissions to service account with complete audit logging. Full automation completes in 5-15 minutes.

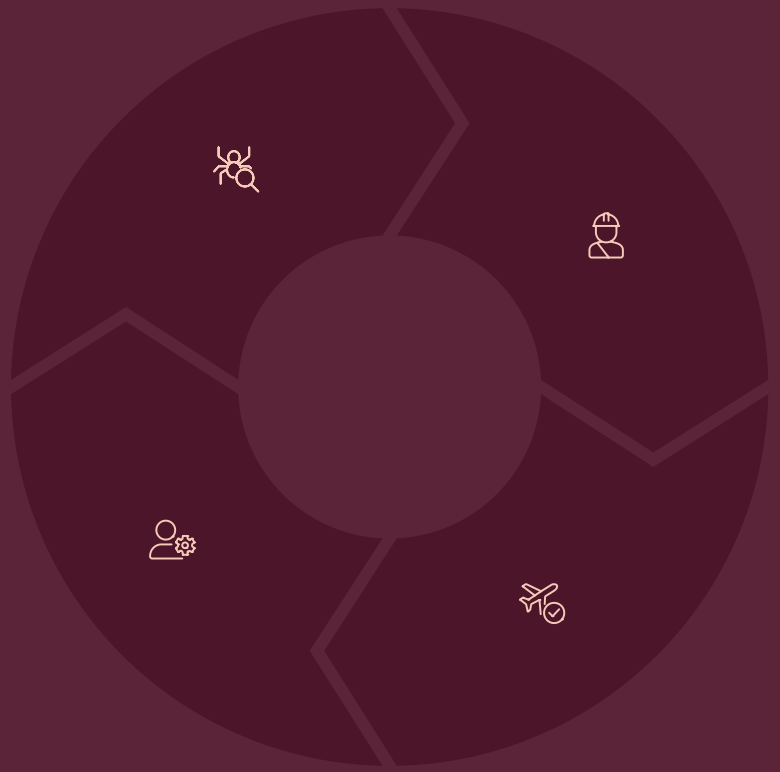
Key Operational Benefits

Teams gain immediate query access upon approval, all interactions are logged for compliance, and standardized templates ensure consistent access patterns across all teams requesting the same data products.

Co-Designing Data Products

Discovery Phase

Joint workshops define specifications based on DS team requirements: data needs, query patterns, and latency requirements like sub-5-second response times.



EDS Development

Build structured data with pre-joined tables, process documents through OCR with quality checks, and implement version-controlled reference data.

Production Operations

Continuous improvement cycle with issue reporting, enhancement requests, and EDS monitoring and optimization of data products.

Team Validation

DS teams test with sample data, verify quality and performance metrics, provide feedback for iterations, and sign-off when requirements are met.

This collaborative approach ensures data products are designed **for** teams based on their specific needs rather than imposed **on** teams with generic solutions. The result is higher adoption rates and more effective AI implementations.

Platform Reusability Across AI Initiatives

First Use Case: Liability Assessment

Build complete foundation including document pipeline, tiering automation, access control, and monitoring systems. Initial development requires months of foundational work.

Second Use Case: Fraud Detection

Reuse existing document pipeline and tiering infrastructure while building fraud-specific features and new data stores. Development time reduced to weeks.

Third Use Case: Property Assessment

Leverage all existing infrastructure, focusing only on property-specific data sources and integrations. Minimal incremental development required.

Compound Value Creation

Each new use case leverages the existing platform foundation, dramatically reducing development time and effort. The platform's value compounds as more teams adopt standardized data products, creating economies of scale and shared operational efficiency.

Real Scenario: Document Replacement

Business Context: CIO Discovers Document Error

When a Claims Investigation Officer discovers an incorrectly uploaded police report for claim CLM-12345, the system must handle document replacement seamlessly while maintaining complete audit trails.

1 Initial State

Police report stored in S3, processed by EDS, residing in "liability-docs-active" Vertex data store with AI agents actively using the document.

2 Error Discovery

CIO uploads corrected version to S3 using same claim ID. S3 event triggers automated EDS response pipeline.

3 Automated Processing

Within 5 minutes: EDS detects update, processes document through OCR and quality checks, marks old version as superseded, ingests new version to Vertex.

4 Seamless Transition

AI agents automatically begin using new version with complete version history preserved and audit trail documenting the change.

Zero manual intervention required. The system maintains operational continuity while ensuring data integrity and compliance requirements are automatically satisfied.

Real Scenario: Document Lifecycle End

Automated Retention Policy Enforcement

EDS implements comprehensive document deletion workflows to meet 7-year retention requirements and handle special compliance scenarios including GDPR "Right to be Forgotten" requests.

01

Daily Retention Assessment

Cloud Scheduler executes daily job querying for documents where claim_closure_date exceeds 7 years, generating deletion candidates list.

02

Multi-System Deletion

For each document: delete from all Vertex AI Search tiers, remove from Cloud Storage, soft delete BigQuery metadata while preserving audit logs.

03

Exception Handling

Litigation hold flags prevent deletion, manual overrides require compliance officer approval, and special processing handles GDPR deletion requests.

04

Proof of Compliance

System provides complete deletion proof for auditors and notifies CIO that corresponding S3 documents can be safely removed from source systems.

Three-Phase Implementation Plan

Phase 1: Foundation (Months 0-6)

Establish Snowflake replication, build document pipeline from S3 to Vertex, create first data product for Liability use case, implement HOT/WARM/COLD tiering automation, and pilot with single DS team to validate approach.

Phase 2: Scale (Months 6-12)

Deploy additional data products for Fraud and Property assessments, onboard multiple DS teams, launch self-service portal, implement automated data store provisioning, and conduct data-driven evaluation of Snowflake versus BigQuery performance.

Phase 3: Optimize (Year 2+)

Achieve production-grade platform status, implement advanced features based on user feedback, establish continuous optimization processes, and finalize strategic decisions on data warehouse architecture.

Success Metrics by Phase

Phase 1 success measured by pilot team satisfaction and basic functionality. Phase 2 targets multiple team adoption and operational efficiency gains. Phase 3 focuses on platform maturity and enterprise-scale optimization.

Expected Business Impact

Note: The statistics and percentages shown below are placeholder (dummy) values for illustrative purposes only.

85%

Time Reduction

Data access requests fulfilled in days instead of weeks, accelerating AI development cycles

60%

Cost Optimization

Reduced infrastructure duplication and improved storage tier management

100%

Compliance Coverage

Complete audit trail and automated retention policy enforcement

3x

Development Velocity

Reusable platform components accelerate subsequent AI initiative delivery

Strategic Positioning Benefits

EDS becomes the critical enabler for enterprise AI initiatives, establishing centralized expertise that scales across all use cases while maintaining operational excellence and regulatory compliance. This positions EDS as an indispensable strategic partner rather than a supporting service provider.

Risk Mitigation Strategy

Technical Risk Controls

- **Vendor Lock-in:** Multi-cloud architecture with portable data formats
- **Scale Limitations:** Automated provisioning prevents capacity constraints
- **Performance Degradation:** Tier-based storage maintains optimal query speeds
- **Data Loss:** Multi-region replication and comprehensive backup strategies

Operational Risk Management

- **Team Dependencies:** Cross-functional training and documentation standards
- **Adoption Resistance:** Co-design approach ensures user buy-in
- **Compliance Failures:** Automated controls with manual override capabilities
- **Budget Overruns:** Cost monitoring with tier-based optimization

Contingency Planning

Each risk category includes specific escalation procedures, fallback options, and recovery timelines. Regular risk assessment reviews ensure proactive identification of emerging challenges before they impact operations.

Resource Requirements & Success Metrics

Team Structure Requirements

Platform Engineering (2 FTE)

Infrastructure automation, Terraform management, GCP service provisioning, and monitoring system implementation

Data Engineering (2 FTE)

Pipeline development, data transformation logic, quality assurance frameworks, and performance optimization

Compliance & Security (1 FTE)

Access control implementation, audit trail design, retention policy automation, and regulatory requirement validation

Key Performance Indicators

- **Operational Efficiency:** Data access request fulfillment time (target: <3 days)
- **Platform Adoption:** Number of active DS teams and data products (target: 8 teams, 12 products by year-end)
- **Cost Management:** Infrastructure cost per data science team (target: 40% reduction)
- **Compliance:** Zero audit findings related to data lifecycle management

Decision & Next Steps

What We Need to Proceed

1 Executive Authorization

Approve EDS AI Data Platform initiative with committed resources including dedicated team assignments and infrastructure budget allocation

2 Cross-Functional Partnerships

Establish formal collaboration agreements with CIO office for source system access, Platform team for GCP provisioning, and DS teams for requirements validation

3 Immediate Technical Actions

Finalize team assignments by week 2, platform provisions GCP project by week 3, setup Snowflake replication by week 4, and begin document pipeline development

90-Day Milestone Commitments

- **Week 4:** First pilot working with single data product and DS team
- **Month 3:** Complete document lifecycle management operational
- **Month 6:** Multiple teams actively using platform with measurable efficiency gains
- **Month 12:** Production-grade platform supporting all AI initiatives

Expected Outcome: EDS positioned as the strategic AI data partner enabling rapid, compliant, and scalable artificial intelligence development across the enterprise.