Introduction

By Raman Kapoor

Blueprint for the Cloud: Building a Future-Ready Data Strategy

By Raman Kapoor

Conf42 Machine Learning 2025

Introduction

Businesses are racing to the cloud in today's world and Cloud migration without a data strategy is a recipe for risk.

- Data sprawled across legacy systems in numerous formats
- Pressures of security, compliance, and AI adoption across all private and public sectors
- Cloud adoption is accelerating to deliver faster ROI and expand access to the global partners

The Risk of Cloud Migration Without a Strategy

- Data silos and inconsistencies
- Security vulnerabilities
- Regulatory and compliance failures
- Incompatible data formats due to multi-Vendors
- Slower time to value and delay in ROI
- Increased Cost
- Redundant Systems , Inconsistent data and formats
- Without structured data strategies, risks multiply
- Without a clear data strategy and vision, it's impossible to harness the full potential of AI insight

Legacy Tech Challenges

Existing Stack:

- Legacy tech ETLs
- Legacy Integrations Interfaces
- Manual data entry
- Inconsistent Vendor-specific data formats

Impacts:

- Maintenance overhead
- Poor interoperability
- Security risks and attacks to Personal data
- Unable to scale with the vast amounts of data we deal with today

Let's explore how to build secure, scalable, and AI-ready data foundations

Vision: A Future-Ready Data Foundation

Characteristics:

- Secure : Zero trust , Encryption , auditability
- Scalable : Modular Pipelines , Cloud native platforms
- AI-ready :High Quality , Governed , Labelled data

Goals:

- Accelerate migration
- Eliminate inconsistencies
- Unlock innovation through clear insight and data driven analytics

Key Pillars of a Data Strategy

- Data Governance
- Metadata Management
  - Centralize Metadata Catalog
- Data Quality and Lineage
  - Column level lineage and impact analysis
  - Semantic consistency across domains
  - Data Quality Engine Framework
- Security and Access Control
- Secure by Design
- Integration Frameworks
- AI and Analytics Enablement from day 1

Building a Secure Data Foundation

Principles:

- Zero-trust architecture
- Encryption at rest and in transit
- IAM and fine-grained access control
- Automated compliance audits
- Identify all Legacy Data , Legacy Tech Inventories , Issues , map dependencies and their impact
- Create Monitoring and alert dashboards
- Enable Analytics and AI during Monitoring and ingestion to identify the data anomalies and behaviour take corrective actions

Inventory for Cloud Migration

Identify:

- Legacy ETLs and API platforms
- Manual processes and file formats
- Data owners and access dependencies
- All the sources

Map:

- Lineage and critical paths
- Data catalog and Metadata
- AI modeling opportunities
- Clear Data Model

Governance and Ownership

Deep Dive - Governance & Ownership

- Data Stewards and Data Owners
- Data usage policies
- Data Sharing Agreements
- Business glossary and metadata tagging
- Regulatory alignment (GDPR, HIPAA, etc.)

Enabling Scalability

Tools & Practices:

- Serverless data pipelines (Lambda, Glue, Athena)
- Data lake architecture
- Event-driven ingestion
- Containerized workloads

Roadmap to Successful Transformation

- Step 1: Discover & Assess
- Step 2: Define Governance & Security Models
- Step 3: Modernize ETLs and APIs
- Step 4: Migrate in Phases
- Step 5: Enable AI & Analytics

Case Study: Migrating to AWS

Background:

- Legacy systems and processes
- Known Data Issues and processes
- Numerous Manual processes
- Disconnected ETLs & APIs , Outdated tools and services
- Numerous Source systems with multi formats

Migration Path:

- Discover & Assess → Define Governance & Security Models → Modernize ETLs and APIs → Migrate in Phases → Enable AI & Analytics

Outcome

- Enabled cloud hosted API solution on AWS API Gateway and decommissioned all the legacy API
- Continued developing ETLs , ingestion with AWS Glue ,Lambda & Step Functions and feeding the curated data to Athena and Postgres SQL to connect with reports and dashboard
- Continued Replicating Legacy reporting to Power BI and initiated operational readiness process

Thank You!
Q&A?

Email:
ramankapoor19@gmail.com