Securing Data Pipelines (ETL/ELT) in the Cloud Training Course

Professional Development Training Course Outline

Category	Data Security	**Duration**	5 Days
Base Fee	$1,500 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

In the era of Big Data and digital transformation, cloud-based data pipelines are the lifeblood of modern analytics, driving crucial business intelligence and AI/ML initiatives. However, migrating or building these pipelines in Multi-Cloud or Hybrid Cloud environments introduces a complex, expanded attack surface. Data is at its most vulnerable during Extract, Transform, and Load phases, especially when dealing with Sensitive Data like PII/PHI across distributed cloud services. This necessitates a shift from traditional perimeter-based security to a proactive, automated, and Zero Trust Architecture (ZTA) approach. Securing Data Pipelines (ETL/ELT) in the Cloud Training Course provides Data Engineers, Cloud Security Architects, and DevSecOps professionals with the in-depth knowledge and Cloud-Native strategies required to design, implement, and monitor truly secure and Compliant Data Pipelines against evolving threats like Data Tampering and Insider Threats.

This training is engineered to move beyond basic cloud security, focusing specifically on the data lifecycle security within modern pipeline orchestration tools like Airflow, Dagster, and cloud services like AWS Glue, Azure Data Factory, and Google Cloud Dataflow. By emphasizing Automated Governance, Secrets Management, and Continuous Monitoring (Observability), learners will acquire the practical skills to embed security controls directly into the pipeline's Infrastructure as Code (IaC), ensuring Data Integrity and meeting stringent regulatory requirements such as GDPR and HIPAA in a scalable, performant, and cost-efficient manner. Master the intersection of Cloud Security Posture Management (CSPM) and Data Governance to build Resilient Data Architectures that protect your organization's most valuable asset: its data.

Programme Curriculum

Securing Data Pipelines (ETL/ELT) in the Cloud Training Course

Introduction

In the era of Big Data and digital transformation, cloud-based data pipelines are the lifeblood of modern analytics, driving crucial business intelligence and AI/ML initiatives. However, migrating or building these pipelines in Multi-Cloud or Hybrid Cloud environments introduces a complex, expanded attack surface. Data is at its most vulnerable during Extract, Transform, and Load phases, especially when dealing with Sensitive Data like PII/PHI across distributed cloud services. This necessitates a shift from traditional perimeter-based security to a proactive, automated, and Zero Trust Architecture (ZTA) approach. Securing Data Pipelines (ETL/ELT) in the Cloud Training Course provides Data Engineers, Cloud Security Architects, and DevSecOps professionals with the in-depth knowledge and Cloud-Native strategies required to design, implement, and monitor truly secure and Compliant Data Pipelines against evolving threats like Data Tampering and Insider Threats.

This training is engineered to move beyond basic cloud security, focusing specifically on the data lifecycle security within modern pipeline orchestration tools like Airflow, Dagster, and cloud services like AWS Glue, Azure Data Factory, and Google Cloud Dataflow. By emphasizing Automated Governance, Secrets Management, and Continuous Monitoring (Observability), learners will acquire the practical skills to embed security controls directly into the pipeline's Infrastructure as Code (IaC), ensuring Data Integrity and meeting stringent regulatory requirements such as GDPR and HIPAA in a scalable, performant, and cost-efficient manner. Master the intersection of Cloud Security Posture Management (CSPM) and Data Governance to build Resilient Data Architectures that protect your organization's most valuable asset: its data.

Course Duration

5 days

Course Objectives

Upon completion, participants will be able to:

1. Implement Zero Trust Architecture (ZTA) principles across data ingestion and transformation stages.
2. Master Cloud-Native Identity and Access Management (IAM) for securing pipeline Service Accounts.
3. Design and deploy pervasive Data Encryption using Key Management Services (KMS).
4. Apply DevSecOps best practices to embed security testing into CI/CD Pipelines for data workflows.
5. Leverage Infrastructure as Code (IaC) tools for secure, compliant pipeline deployment.
6. Configure advanced Secrets Management solutions for credentials.
7. Establish robust Data Governance and Compliance as Code rules for PII/PHI handling.
8. Implement Network Segmentation and Private Endpoints to isolate cloud data flows.
9. Utilize Cloud Security Posture Management (CSPM) tools to detect and remediate pipeline misconfigurations.
10. Develop real-time Security Observability via comprehensive logging, auditing, and SIEM integration.
11. Secure Containerized and Serverless pipeline components
12. Define and execute a strong Incident Response plan for data breach scenarios in a cloud environment.
13. Mitigate risks associated with Data Tampering and Supply Chain Attacks in third-party data sources.

Target Audience

1. Data Engineers/Architects.
2. Cloud Security Architects/Engineers.
3. DevSecOps Engineers.
4. Security Analysts.
5. Compliance and Risk Officers

6. Senior Software Engineers.
7. Data Scientists/ML Engineers.
8. Technical Leaders/Managers.

Course Modules

Module 1: Cloud Data Pipeline Threat Modeling & Architecture

- Data Lifecycle stages and their associated threat vectors.
- Understanding the Shared Responsibility Model in the context of data services
- Principles of Data Tampering and Data Exfiltration prevention in ETL/ELT flows.
- Implementing the Principle of Least Privilege for all pipeline components.
- Cloud Architecture Patterns.
- Case Study: Analyzing a major financial institution's S3 Bucket/Blob Storage Misconfiguration leading to a massive data leak.

Module 2: Identity, Authentication, and Secrets Management

- Using Service Roles and fine-grained permissions for ETL/ELT workers.
- Implementing Multi-Factor Authentication and Session Tags for privileged access.
- Secrets Management.
- Federated Identity for cross-account and third-party access to data sources.
- Automating access reviews and credential rotation policies.
- Case Study: A large retailer's data breach via a hardcoded database password in a pipeline script, and the shift to Azure Key Vault or AWS Secrets Manager.

Module 3: Data Encryption and Key Management (KMS)

- Best practices for Encryption at Rest.
- Mandating Encryption in Transit using TLS/SSL across all data links.
- Cloud KMS management.
- Implementing Client-Side Encryption for maximum control over sensitive data.
- Securing and auditing access to the master Encryption Keys.
- Case Study: A healthcare provider's non-compliant pipeline where PHI data was processed unencrypted due to poor KMS policy configuration.

Module 4: Network Security and Isolation

- Segmenting data pipeline components using Virtual Private Clouds and subnets.
- Securing connection endpoints using Private Link/Private Endpoints to bypass the public internet.
- Configuring Security Groups/Network Security Groups as firewalls for data flow.
- Implementing and monitoring DLP policies on egress points.
- Using Firewall as a Service for granular control over third-party connections.
- Case Study: A manufacturing company's successful use of AWS PrivateLink to secure data transfer from an on-premises data center to its Snowflake ELT environment.

Module 5: DevSecOps and Infrastructure as Code (IaC) Security

- Writing Terraform or CloudFormation templates with embedded security checks.
- Integrating Vulnerability Scanning for container images and pipeline dependencies in the CI/CD phase.
- Enforcing Compliance as Code to ensure pipeline deploys meet regulatory blueprints automatically.
- Code Signing and artifact verification to prevent supply chain attacks.

- Implementing Blue/Green Deployments for secure, low-risk pipeline updates.
- Case Study: A platform's use of GitHub Actions and a CSPM tool to automatically revert pipeline configuration changes that violated a Cloud Policy.

Module 6: Runtime Security and Governance

- Securing Serverless components via resource-level permissions and least privilege.
- Hardening Dockerfiles, using minimal base images, and managing runtime policies for Kubernetes/ECS.
- Enforcing Data Masking, Tokenization, and Format-Preserving Encryption during the transformation phase.
- Utilizing Attribute-Based Access Control for dynamic data access within a pipeline.
- Implementing Data Quality Checks as a security measure against injection or manipulation.
- Case Study: A global tech firm implementing Databricks/Spark cluster security policies and row-level access controls to ensure data residency compliance during processing.

Module 7: Security Observability and Continuous Monitoring

- Establishing Audit Trails and comprehensive logging for all data access and configuration changes.
- Integrating pipeline logs with Security Information and Event Management tools
- Setting up Real-time Alerting for anomalies
- Utilizing Cloud-Native Monitoring tools for pipeline health and security.
- Performing regular Vulnerability Assessments and Penetration Testing on the entire pipeline stack.
- Case Study: A social media company detecting an Insider Threat through a sudden, large-scale data export flagged by their Behavioral Analytics monitoring system.

Module 8: Compliance, Incident Response, and Future Trends

- Mapping pipeline security controls to major compliance frameworks
- Developing and testing a structured Data Breach Incident Response Plan specific to cloud data services.
- Defining Disaster Recovery and Business Continuity for critical data pipelines.
- Exploring the role of AI/ML in Threat Detection for future pipelines.
- Securing advanced architectures like Data Mesh and Data Lakehouse environments.
- Case Study: A financial service company's regulatory audit success following the implementation of "Automated Data Lineage" and Compliance as Code tools to prove control effectiveness.

Training Methodology

This course employs a participatory and hands-on approach to ensure practical learning, including:

- Interactive lectures and presentations.
- Group discussions and brainstorming sessions.
- Hands-on exercises using real-world datasets.
- Role-playing and scenario-based simulations.
- Analysis of case studies to bridge theory and practice.
- Peer-to-peer learning and networking.
- Expert-led Q&A sessions.
- Continuous feedback and personalized guidance.

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.com or call +254769199797

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to FINESKILL TRAINING CENTER account, as indicated in the invoice so as to enable us prepare better for you.

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
07 Sep 2026	11 Sep 2026	Online	$1,100
14 Sep 2026	18 Sep 2026	Online	$1,100
21 Sep 2026	25 Sep 2026	Online	$1,100
28 Sep 2026	02 Oct 2026	Online	$1,100
05 Oct 2026	09 Oct 2026	Online	$1,100
12 Oct 2026	16 Oct 2026	Online	$1,100
19 Oct 2026	23 Oct 2026	Online	$1,100
26 Oct 2026	30 Oct 2026	Online	$1,100

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com