Course description

In this advanced course, you will delve into the critical aspects of securing generative AI applications and models on Google Cloud's Vertex AI platform. You will gain a comprehensive understanding of the unique security challenges posed by generative AI, including prompt injection, data poisoning, and model exfiltration. The course will cover best practices for implementing robust security controls, ensuring data privacy, and maintaining compliance throughout the entire generative AI lifecycle, from data preparation and model training to deployment and monitoring. You will acquire practical insights into leveraging Vertex AI's native security features, Google Cloud's broader security ecosystem, and responsible AI principles to build secure and trustworthy generative AI solutions.

- Course level: Advanced
- Duration: 2 days

Activities

This course includes presentations, hands-on labs, demonstrations, and group exercises.

Course objectives

In this course, you will learn to:

- Identify and analyze specific security threats and vulnerabilities unique to generative AI models and applications on Vertex AI.

- Design and implement robust Identity and Access Management (IAM) controls for Vertex AI resources and generative AI workflows.
- Secure data at rest and in transit, and apply data privacy-preserving techniques relevant to generative AI training and inference.
- Utilize Vertex AI's built-in security features and Google Cloud's security services to protect generative AI models and endpoints.
- Develop strategies for mitigating common generative AI attacks, such as prompt injection and adversarial examples.
- Implement responsible AI practices and content moderation to ensure safe and ethical deployment of generative AI solutions.
- Establish governance frameworks, compliance strategies, and incident response plans for generative AI security.

Intended audience

This course is intended for professionals involved in securing, developing, or managing AI systems on Google Cloud:

- AI/ML Engineers and Practitioners
- Data Scientists
- Cloud Security Architects
- DevOps and MLOps Engineers
- Security Operations (SecOps) Analysts
- Compliance and Governance Officers
- IT Security Professionals
- Solution Architects

Prerequisites

We recommend that attendees of this course have:

- Basic understanding of Generative AI concepts and technologies.
- Familiarity with Google Cloud Platform (GCP) fundamentals, including IAM, networking, and core services.
- Conceptual knowledge of security principles and best practices.
- Basic experience with Python for hands-on lab exercises.

Course outline

Module 1: Introduction to Vertex AI and Generative AI Security Landscape

- Overview of Vertex AI platform and its Generative AI capabilities (e.g., PaLM 2, Imagen, Codey).
- Introduction to the unique security challenges and attack vectors specific to generative AI.
- The evolving threat landscape for Large Language Models (LLMs) and foundation models.
- Importance of a defense-in-depth strategy for GenAI on Vertex AI.
- Demo: Exploring Vertex AI Generative AI Studio and inherent security controls.

Module 2: Understanding Generative AI-Specific Threats

- Detailed analysis of prompt injection techniques (direct, indirect, jailbreaks).
- Data poisoning and model manipulation attacks during training and fine-tuning.
- Model exfiltration, inversion, and membership inference attacks.
- Adversarial attacks against generative models (e.g., text, image generation).
- Understanding ethical hacking and penetration testing methodologies for GenAI.
- Use case for class: Analyzing a real-world prompt injection vulnerability and its impact.

Module 3: Securing the Vertex AI Platform for Generative AI

- Deep dive into Google Cloud Identity and Access Management (IAM) for Vertex AI.

–Fine-grained permissions for Generative AI APIs, notebooks, and custom models.

–Best practices for service accounts and user-managed identities.

- Network security for Vertex AI workloads:

– VPC Service Controls for perimeter defense.

–Private Service Connect and Private IP for secure endpoint access.

- Data encryption strategies for GenAI:

– Encryption at rest (CMEK, CMK) for training data, model artifacts, and inference data.

–Encryption in transit (TLS/SSL) for API calls and data movement.

- Hands-on Lab: Configuring least privilege IAM roles for a Generative AI project on Vertex AI.

Module 4: Responsible AI, Trust & Safety for Secure Generative AI

- Integrating Responsible AI principles into the security framework of GenAI.
- Leveraging Vertex AI's Trust & Safety APIs and content moderation features.
- Strategies for detecting and mitigating the generation of harmful, biased, or toxic content.
- Addressing fairness, transparency, and accountability as security considerations.
- Human-in-the-loop validation and oversight for generative AI outputs.
- Hands-on Lab: Implementing and evaluating content filtering and safety attributes for a Vertex AI Generative AI model.

Module 5: Data Security and Privacy in Generative AI Workflows

- Securing the entire data lifecycle for generative AI:

–Data ingestion and preparation (Cloud Storage, BigQuery).

– Training and fine-tuning data protection.

–Inference data handling and retention policies.

- Privacy-enhancing technologies (PETs) for generative AI:

– Data anonymization, pseudonymization, and tokenization.

–Introduction to differential privacy and federated learning in GenAI contexts.

- Data lineage, auditing, and logging for compliance with privacy regulations (GDPR, HIPAA, CCPA).
- Establishing clear data governance policies for sensitive information used by GenAI.

Module 6: Secure Generative AI Model Deployment and MLOps

- Implementing secure MLOps practices for Vertex AI Generative AI.
- Securing Vertex AI Model Registry and versioning of foundation and fine-tuned models.
- Secure deployment of generative AI models as endpoints:

–Private endpoints and authenticated access for custom models.

–Access controls for Generative AI Studio APIs.

- Monitoring, logging, and auditing for security anomalies:

–Integrating with Cloud Logging, Cloud Monitoring, and Security Command Center.

– Detecting unusual usage patterns or potential attacks.

- Vulnerability scanning and penetration testing for GenAI applications and APIs.
- Demo: Setting up security logging and alerting for a Vertex AI Generative AI endpoint.

Module 7: Incident Response and Advanced Threat Mitigation for Generative AI

- Developing a comprehensive incident response plan for generative AI security breaches.
- Strategies for detecting and responding to prompt injection, data exfiltration, and model theft.
- Advanced mitigation techniques for adversarial attacks and model manipulation.
- Leveraging threat intelligence specific to generative AI.
- Continuous security improvement, regular audits, and vulnerability assessments.
- Legal and ethical considerations in responding to GenAI security incidents.
- Hands-on Lab: Capstone – Designing an Incident Response Playbook for a simulated prompt injection attack on a Vertex AI Generative AI application.

Module 8: Course Wrap-up

- Review of key security concepts and best practices for Vertex AI Generative AI.
- Discussion of emerging threats and future trends in GenAI security.
- Next steps and additional resources for ongoing learning (e.g., Google Cloud security certifications, advanced topics in AI security research).
- Course summary and Q&A.

cloudthat

Your course overview ends here.

But your cloud journey doesn't!

Unlock advanced training and certification pathways with **CloudThat.**

About CloudThat

Established in 2012, CloudThat is an award-winning company and the first in India to offer cloud training and consulting services for individuals and enterprises worldwide. Recently, it won Google Cloud's New Training Partner of the Year Award for 2025, becoming the first company in the world in 2025 to hold awards from all three major cloud giants: AWS, Microsoft, and Google. CloudThat notably won consecutive AWS Training Partner of the Year (APJ) awards in 2023 and 2024 and the Microsoft Training Services Partner of the Year Award in 2024, bringing its total award count to an impressive 12 awards in the last 8 years.

CloudThat specializes in Cloud Migration, Data Platforms, DevOps, IoT, and cutting-edge technologies like Gen AI & AI/ML. As a Microsoft Solutions Partner, AWS Premier Tier Training Partner, Google Cloud Platform Partner, and collaborator with leading organizations like HPE, Oracle, Nvidia and Databricks, CloudThat has trained over 850,000 professionals across 600+ cloud certifications, empowering students and professionals worldwide to advance their skills and careers.

To know more about our cloud certification training, email at sales@cloudthat.com or call us at +91-8880002200 or +1-8555588830 for those from outside India.