



SNOWPRO® SPECIALTY: GEN AI EXAM STUDY GUIDE

Last Updated: February 10, 2026



TABLE OF CONTENTS:

| | |
|--|-----------|
| SNOWPRO® SPECIALTY: GEN AI STUDY GUIDE OVERVIEW..... | 3 |
| RECOMMENDATIONS FOR USING THIS GUIDE..... | 3 |
| SNOWPRO® SPECIALTY: GEN AI EXAM CONTENT AND FORMAT..... | 4 |
| IMPORTANT UPDATE..... | 4 |
| CERTIFICATION OVERVIEW..... | 4 |
| SNOWPRO® SPECIALTY: GEN AI PREREQUISITE..... | 5 |
| STEPS TO SUCCESS..... | 5 |
| EXAM DOMAINS AND WEIGHTINGS..... | 6 |
| Domain 1.0: Snowflake for Gen AI Overview..... | 7 |
| Domain 2.0: Snowflake Gen AI & LLM Functions..... | 9 |
| Domain 3.0: Snowflake Gen AI Governance..... | 11 |
| Domain 4.0: Snowflake Document AI..... | 13 |
| SNOWPRO® SPECIALTY: GEN AI SAMPLE QUESTIONS..... | 14 |
| MAINTAINING YOUR CERTIFICATION..... | 15 |

SNOWPRO® SPECIALTY: GEN AI STUDY GUIDE OVERVIEW

This is a self-paced guide that will outline the Snowflake domains, objectives, and topics that will be covered on the **SnowPro Specialty: Gen AI Certification Exam**. Use of this study guide does not guarantee certification success.

Holding **either** the SnowPro Associate: Platform **or** the SnowPro Core Certification in good standing are prerequisite options for taking the Specialty: Gen AI Certification Exam.

For an overview and more information on all of the available SnowPro Certifications, please navigate [here](#).

RECOMMENDATIONS FOR USING THIS GUIDE

This guide outlines the Snowflake topics and subtopics covered on the exam. Following the topics will be additional resources consisting of links to documentation, blogs, and/or exercises that are designed to support candidates' understanding of Gen AI on Snowflake.

Estimated length of study time required to complete the guide: 10 – 13 hours

Some links may have more value for exam preparation than others, depending on the candidate's level of experience. The same amount of time should not be spent on each link. Some links may appear in more than one content domain.

SNOWPRO® SPECIALTY: GEN AI EXAM CONTENT AND FORMAT

IMPORTANT UPDATE

Document AI will be decommissioned on February 28, 2026. Starting **March 1, 2026**, the SnowPro Specialty: Gen AI Exam will be updated as follows:

- **Domain 4.0 will be removed completely**
- **All questions about Document AI will be removed from the exam**

Review the updated exam topics and weightings below. Adjust your exam preparation based on when you plan to take the exam.

CERTIFICATION OVERVIEW

The **SnowPro Specialty: Gen AI Certification Exam** tests specialized knowledge, skills, and best practices used to leverage Gen AI methodologies in Snowflake including key concepts, features, and programming constructs. The exam will assess skills through scenario-based questions, interactive questions, and real-world examples.

This certification will test the candidate's ability to:

- Define and implement Snowflake Gen AI principles, capabilities, and best practices related to infrastructure, data governance, and cost governance
- Leverage Snowflake Cortex AI features, Large Language Models (LLMs), and offerings to support customer use cases (for example, Cortex Analyst, Cortex Search, Cortex Fine-tuning, Snowflake Copilot)
- Build open-source models with Snowpark Container Services and Snowflake Model Registry (for example, Hugging Face)

The candidate is expected to have knowledge of:

- Snowflake Cortex suite of Gen AI features and their underlying models
- Retrieval Augmented Generation (RAG) applications leveraging LLMs

Target Audience:

Candidates with one or more years of Gen AI experience with Snowflake, in an enterprise environment. In addition, successful candidates may have advanced proficiency writing code in Python. Previous data engineering and SQL knowledge is assumed.

This exam is designed for:

- AI or ML Engineers
- Data Scientists
- Data Engineers
- Data Application Developers
- Data Analysts with programming experience

SNOWPRO® SPECIALTY: GEN AI PREREQUISITE

To be eligible to take this exam candidates must hold an active SnowPro Associate: Platform or SnowPro Core Certification.

STEPS TO SUCCESS

1. Review this Gen AI Exam Study Guide
2. Attend Snowflake's Instructor-Led [Snowflake Gen AI Training](#)
3. Get hands-on experience with our [Snowflake x Gen AI: A Cool Collab](#) and [Snowflake X Gen AI: LLM Functions](#)
4. [Review and study applicable white papers and documentation](#)
5. Get hands-on practical experience with relevant business requirements using Snowflake
6. [Attend free Snowflake Webinars](#)
7. [Attend Snowflake Virtual Hands-on Labs](#) to gain practical experience
8. Practice with sample questions [here](#)
9. [Schedule the exam](#)
10. [Take the exam!](#)

EXAM DOMAINS AND WEIGHTINGS

Before March 1, 2026: See Table 1 for exam domains and weightings.

Starting March 1, 2026: See Table 2 for updated exam domains and weightings.

The objectives listed under each domain are examples only. The exam may include additional related topics not specifically listed.

Table 1. Domains and Weightings **Before March 1, 2026**

| Domain | Domain Weightings |
|--------------------------------------|-------------------|
| 1.0 Snowflake for Gen AI Overview | 26% |
| 2.0 Snowflake Gen AI & LLM Functions | 40% |
| 3.0 Snowflake Gen AI Governance | 22% |
| 4.0 Snowflake Document AI | 12% |

Table 2. Domains and Weightings **Starting March 1, 2026**

| Domain | Domain Weightings |
|--------------------------------------|-------------------|
| 1.0 Snowflake for Gen AI Overview | 30% |
| 2.0 Snowflake Gen AI & LLM Functions | 44% |
| 3.0 Snowflake Gen AI Governance | 26% |



► Domain 1.0: Snowflake for Gen AI Overview

1.1 Define Snowflake's Gen AI principles, features, and best practices.

- Snowflake Cortex
 - LLMs
 - Cortex Search
 - Cortex Analyst
 - Cortex Fine-tuning
 - Cortex Agents (Public Preview)
- Snowflake Copilot
- Security, privacy, access, and control principles
 - Role-Based Access Control (RBAC)
 - Guardrails
 - Required privileges
 - Cortex LLM Functions
 - Control model access
 - CORTEX_MODELS_ALLOWLIST parameter
- Different interfaces
 - Cortex LLM Playground (Public Preview)
 - SQL
 - REST API
- Different ways of bringing your own models into Snowflake (for example, from Hugging Face)
 - Using Snowflake Model Registry (custom model)
 - Using Snowpark Container Services

1.2 Outline Gen AI capabilities in Snowflake.

- Cortex LLM functions (for example, task-specific, general)
 - Vector-embedding
 - Fine-tuning
- Cortex Search
 - RAG use cases
 - Unstructured data use cases
 - REST APIs
- Cortex Analyst
 - Semantic model generation
 - Stored in YAML files in a stage
 - Stored natively in semantic views (Public Preview)
 - Structured/text-to-SQL use cases
 - REST APIs
- Cortex Agents (Public Preview)
 - REST APIs
- Cross-region inference
 - CORTEX_ENABLED_CROSS_REGION parameter
 - Considerations (for example, latency, availability)

Domain 1.0 Study Resources

Snowflake Documentation

[Document AI](#)

[Cortex LLM REST API](#)

[Cortex Search](#)

[Cortex Analyst](#)

[Cortex LLM Playground](#)

[Vector data types](#)

[VECTOR_INNER_PRODUCT](#)

[Known limitations to Document AI](#)

[Snowflake Cortex AISQL \(including LLM functions\)](#)

[Anomaly Detection \(Snowflake ML Functions\)](#)

[Time-Series Forecasting \(Snowflake ML Functions\)](#)

[Using Snowflake Copilot](#)

[Bring your own model types via serialized files](#)

[Snowflake Model Registry user interface](#)

[COMPLETE \(SNOWFLAKE.CORTEX\)](#)

[FINETUNE \('CREATE'\)](#)

[\(SNOWFLAKE.CORTEX\)](#)

[VECTOR_L1_DISTANCE](#)

[VECTOR_L2_DISTANCE](#)

[VECTOR_COSINE_SIMILARITY](#)

[Cross-region inference](#)

[Fine-tuning \(Snowflake Cortex\)](#)

[GRANT DATABASE ROLE](#)

[Vector Embeddings](#)

Reading Assets

[Best Practices for Getting Started with Snowflake's Document AI \(Article\)](#)

[Snowflake AI Trust and Safety FAQs \(Article\)](#)

[Snowflake's Data Architecture: Enabling AI Apps, Next-Gen Lakehouse Analytics And More \(Article\)](#)



Domain 2.0: Snowflake Gen AI & LLM Functions

2.1 Apply Gen AI and LLM functions in Snowflake.

- Snowflake Cortex
 - General
 - COMPLETE
 - COMPLETE Structured Outputs
 - Task-specific functions
 - CLASSIFY_TEXT
 - EXTRACT_ANSWER
 - PARSE_DOCUMENT
 - SENTIMENT
 - SUMMARIZE
 - TRANSLATE
 - EMBED_TEXT_768
 - EMBED_TEXT_1024
- Cortex Search
- Cortex Analyst
- Cortex Fine-tuning
- Cortex Agents (Public Preview)
- Vector functions
 - VECTOR_INNER_PRODUCT
 - VECTOR_L1_DISTANCE
 - VECTOR_L2_DISTANCE
 - VECTOR_COSINE_SIMILARITY
- Helper functions
 - COUNT_TOKENS
 - TRY_COMPLETE
 - SPLIT_TEXT_RECURSIVE_CHARACTER

- Choosing a model
 - Considerations (e.g. capability, latency, and cost)

2.2 Perform data analysis given a use case.

- Use fully-managed LLMs, RAG, and text-to-SQL services
 - Unstructured data
 - CORTEX_PARSE_DOCUMENT
 - Structured data
 - Cortex Analyst
 - Cortex Analyst Verified Query Repository (VQR)
 - Integration with Cortex Search
 - Suggested Questions
 - Custom_instructions field
- Performance considerations
 - Latency (for example, fine-tuning, model size)

2.3 Build chat interfaces to interact with data in Snowflake.

- Set up the Snowflake environment
 - Required privileges
- Invoke Cortex functions within the application code (for example, Streamlit)
- Chat conversations
 - Multi-turn architecture
 - Update parameters

2.4 Use Snowflake Cortex functions in data pipelines.

- Snowflake Cortex
 - SQL interface
 - Extracting data from text using COMPLETE
 - Transcripts
 - Data enrichment
 - Data augmentation
 - Data transformations

2.5 Run third-party models in Snowflake.

- Using Snowpark Container Services
 - Environment setup
 - Docker images
 - Specification files
 - Create compute pool
 - Create image repository
- Using the Snowflake Model Registry
 - Logging the model
 - Calling the model

Domain 2.0 Study Resources

Snowflake Documentation

[Cortex Agents](#)

[Cortex PARSE_DOCUMENT](#)

[Cortex Analyst Verified Query Repository](#)

[Suggested Questions in Cortex Analyst](#)

[Improve literal search to enhance Cortex Analyst responses](#)

[Cortex Analyst semantic model specification](#)

[Cortex Search tutorials](#)

[Dynamic tables compared to streams and tasks and to materialized views](#)

[AI_COMPLETE Structured Outputs](#)

[Snowpark Container Services](#)

[Working with an image registry and repository](#)

[Working with an image registry and repository](#)

[Snowflake Model Registry](#)

[Snowflake ML: End-to-End Machine Learning](#)

[Model Serving in Snowpark Container Services](#)

[Snowpark Container Services Tutorials](#)

[SENTIMENT \(SNOWFLAKE.CORTEX\)](#)

[TRANSLATE \(SNOWFLAKE.CORTEX\)](#)

[TRY_COMPLETE \(SNOWFLAKE.CORTEX\)](#)

[CLASSIFY_TEXT \(SNOWFLAKE.CORTEX\)](#)

[SUMMARIZE \(SNOWFLAKE.CORTEX\)](#)

[EXTRACT_ANSWER \(SNOWFLAKE.CORTEX\)](#)

[snowflake.ml.registry.Registry](#)

[CREATE COMPUTE POOL](#)

Quickstarts and Tutorials

[Fine-Tuning an LLM in Snowpark Container Services with AutoTrain](#)

[Deploy and Fine-tune Open Source Llama 2 in Snowpark Container Services](#)

[Build a basic LLM chat app](#)

Reading Assets

[Snowflake Cortex Analyst: Support for Multi-turn Conversation \(Article\)](#)

[AI-Infused Pipelines with Snowflake Cortex \(Blog\)](#)

[Snowpark Container Services — A Tech Primer \(Blog\)](#)



Domain 3.0: Snowflake Gen AI Governance

3.1 Set up model access controls.

- Limits on which models can be used
 - Restrict access to specific models
 - CORTEX_MODELS_ALLOWLIST parameter
 - Cortex LLM REST API
 - COMPLETE(SNOWFLAKE.CORTEX)
 - TRY_COMPLETE(SNOWFLAKE.CORTEX)
 - Cortex LLM Playground (Public Preview)
- Data safety and security considerations
 - Is data leaving/going to LLMs?
- REST API authentication methods

3.2 Set guardrails to filter out harmful or unsafe LLM responses.

- Cortex Guard
 - COMPLETE arguments
- Methods to reduce model hallucinations and bias
- Error conditions

3.3 Monitor and optimize Snowflake Cortex costs.

- Cortex Search
 - Different types of costs (virtual warehouse, EMBED_TEXT, Serving)
- Cortex Analyst
 - Snowflake Service Consumption Table
- Cortex LLM functions
 - Minimize tokens
 - Token cost implications
- Tracking model usage and consumption
 - Usage quotas
 - CORTEX_FUNCTIONS_USAGE_HISTORY view
 - CORTEX_FUNCTIONS_QUERY_USAGE_HISTORY view

3.4 Use Snowflake AI observability tools.

- Snowflake AI observability (Public Preview) features
 - Evaluation metrics
 - Comparisons
 - Tracing
 - Logging
 - Event tables
- Implementation methods
 - Trulens SDK

Domain 3.0 Study Resources

Snowflake Documentation

Snowflake Service Consumption Table
Snowpark-optimized warehouse
Evaluate AI applications
AI Observability in Snowflake Cortex
Cortex Analyst administrator monitoring
Understanding cost for Cortex Search Services
[**CORTEX_FUNCTIONS_USAGE_HISTORY**](#)
view

[**METERING_DAILY_HISTORY**](#) view
[**CORTEX_SEARCH_DAILY_USAGE_HISTORY**](#)
view
[**CORTEX_DOCUMENT_PROCESSING_USAGE_HISTORY**](#)
view
`snowflake.cortex.CompleteOptions`



Domain 4.0: Snowflake Document AI

Domain 4.0 will be removed starting March 1, 2026.

4.1 Set up Document AI.

- Virtual warehouse, database, and schema considerations
- Roles and privileges
 - USAGE
 - OPERATE
 - CREATE SNOWFLAKE.ML.DOCUMENT_INTELLIGENCE
 - CREATE MODEL

4.2 Prepare documents for Document AI.

- Upload documents
- Train the model
- Requirements (for example, formats, size limits)
- Question optimization best practices

4.3 Extract values from documents using Document AI.

- Conditions
- <model_build_name>!PREDICT query
- Automation of data pipelines

4.4 Troubleshoot Document AI given a use case.

- Extracting query errors
- GET_PRESIGNED_URL function
- Requirements and privileges
- Cost and best practices considerations

Domain 4.0 Study Resources

Snowflake Documentation

Document AI

Setting up Document AI

Troubleshooting Document AI

Cost governance of Document AI

Create a document processing pipeline

Prepare your documents for Document AI

Question optimization for extracting information with Document AI

Known limitations to Document AI

Extract information with Document AI

DOCUMENT_INTELLIGE(SNOWFLAKE.ML)

CREATE STAGE

<model_build_name>!PREDICT

SNOWPRO® SPECIALTY: GEN AI SAMPLE QUESTIONS

1. A Gen AI Specialist needs to analyze the daily costs incurred for AI services in Snowflake.

Which query will retrieve the credit consumption from Snowflake's metadata objects for data usage?

- A. SELECT * FROM SNOWFLAKE.ACCOUNT_USAGE.QUERY_HISTORY
WHERE SERVICE_TYPE='AI_SERVICES';
 - B. SELECT * FROM SNOWFLAKE.INFORMATION_SCHEMA.METERING_HISTORY
WHERE SERVICE_TYPE='AI_SERVICES';
 - C. SELECT * FROM SNOWFLAKE.ACCOUNT_USAGE.METERING_HISTORY
WHERE SERVICE_TYPE='AI_SERVICES';
 - D. SELECT * FROM SNOWFLAKE.ACCOUNT_USAGE.METERING_DAILY_HISTORY
WHERE SERVICE_TYPE='AI_SERVICES';*
2. What is the primary role of memory in a multi-turn chat conversation using a Gen AI model in Snowflake Cortex Analyst?
 - A. To securely store user credentials
 - B. To increase the speed of response generation
 - C. To maintain context throughout multiple requests*
 - D. To limit the number of tokens processed for each request
 3. Which parameter can be used by administrators to restrict access to specific LLMs within Snowflake?
 - A. NETWORK_POLICY
 - B. SAML_IDENTITY_PROVIDER
 - C. CORTEX_MODELS_ALLOWLIST*
 - D. CORTEX_ENABLED_CROSS_REGION
 4. Which Snowflake Cortex LLM function should be used to generate an instructional lesson plan based on a prompt?
 - A. COMPLETE*
 - B. EXTRACT_ANSWER
 - C. SUMMARIZE
 - D. TRANSLATE

MAINTAINING YOUR CERTIFICATION

All Snowflake Certifications expire two (2) years after your certification issue date.

SnowPro Certifications can now be recertified through the Snowflake Continuing Education (CE) program which includes these options -

- Completion of eligible Snowflake [Instructor Led \(ILT\) Training Courses](#)
- Earning of an equivalent or higher-level SnowPro Certification

Note: You must have a valid Certification to participate in the Continuing Education (CE) program.

The information provided in this guide is provided for your internal purposes only and may not be provided to third parties.

IN ADDITION, THIS STUDY GUIDE IS PROVIDED “AS IS”. NEITHER SNOWFLAKE NOR ITS SUPPLIERS MAKES ANY OTHER WARRANTIES, EXPRESS OR IMPLIED, STATUTORY OR OTHERWISE, INCLUDING BUT NOT LIMITED TO WARRANTIES OF MERCHANTABILITY, TITLE, FITNESS FOR A PARTICULAR PURPOSE OR NONINFRINGEMENT.