



# SNOWPRO<sup>®</sup> SPECIALTY: GEN AI EXAM STUDY GUIDE

Early Registration: June 2, 2025  
Beta Delivery: June 23 - July 13, 2025



**EXAM STUDY GUIDE TABLE OF CONTENTS:**

**SNOWPRO® SPECIALTY: GEN AI STUDY GUIDE OVERVIEW.....3**  
**RECOMMENDATIONS FOR USING THE GUIDE..... 3**  
**SNOWPRO® SPECIALTY: GEN AI EXAM CONTENT AND FORMAT.....4**  
    CERTIFICATION OVERVIEW..... 4  
**SNOWPRO® SPECIALTY: GEN AI PREREQUISITE.....5**  
    STEPS TO SUCCESS..... 5  
    EXAM DOMAINS AND WEIGHTINGS..... 5  
        Domain 1.0: Snowflake for Gen AI Overview..... 6  
        Domain 2.0: Snowflake Gen AI & LLM Functions..... 8  
        Domain 3.0: Snowflake Gen AI Governance..... 10  
        Domain 4.0: Snowflake Document AI.....12  
**SNOWPRO® SPECIALTY: GEN AI SAMPLE QUESTIONS..... 13**

## 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 Beta Exam**. Use of this study guide does not guarantee certification success.

Holding either the SnowPro Associate: Platform or SnowPro Core Certification in good standing is a prerequisite for taking the Specialty: Gen AI Certification Beta Exam.

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

## RECOMMENDATIONS FOR USING THE GUIDE

This guide will show the Snowflake topics and subtopics covered on the exam. Following the topics will be additional resources consisting of documentation, blogs, and/or exercises to help you understand Gen AI.

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

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

# SNOWPRO® SPECIALTY: GEN AI EXAM CONTENT AND FORMAT

## CERTIFICATION OVERVIEW

The **SnowPro Specialty: Gen AI Certification Beta 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 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 meet 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)
- Use Document AI to train and troubleshoot models specific to customer use cases

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:

1 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

Eligible individuals 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 Snowflake Webinars](#)
7. [Attend Snowflake Virtual Hands-on Labs](#) to gain practical experience
8. [Schedule your exam](#)
9. Take your exam!

### EXAM DOMAINS AND WEIGHTINGS

The following table contains the domains and weightings covered on the exam. It is not a comprehensive listing of all the content that will be presented on the exam.

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%



## 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](#)  
[Large Language Model \(LLM\) Functions \(Snowflake Cortex\)](#)  
[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](#)  
[Snowflake AI and ML](#)  
[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 Snowflake Model Registry
  - Logging the model
  - Calling the model

## Domain 2.0 Study Resources

### Snowflake Documentation

[Cortex Agents](#)  
[Cortex Agents REST API](#)  
[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](#)  
[Cortex COMPLETE Structured Outputs](#)  
[Snowpark Container Services: Working with an image registry and repository](#)  
[Snowpark Container Services](#)  
[Snowflake Model Registry](#)  
[Snowflake ML: End-to-End Machine Learning Model Serving in Snowpark Container Services](#)  
[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

[Fine-Tuning an LLM in Snowpark Container Services with AutoTrain](#)  
[Deploy and Fine-tune Open Source Llama 2 in Snowpark Container Services](#)  
[Snowpark Container Services Tutorials](#)  
[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\\_HI](#)

[STORY view](#)

[snowflake.cortex.CompleteOptions](#)



## Domain 4.0: Snowflake Document AI

### 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 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_PREIGNED_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\\_INTELLIGENCE \(SNOWFLAKE.ML\)](#)

[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 store user credentials securely
- B. To increase the speed of response generation
- C. To maintain context throughout several requests\*
- D. To limit the number of tokens processed for each request

3. A Gen AI Specialist is using Document AI to create a model. When creating a model build with a name unique to the specified schema, this error is returned:

Unable to create a build on the specified database and schema.  
Please check the documentation to learn more.

What would cause this error?

- A. There is a model build with the same name in another schema in the database.
- B. The `CREATE SNOWFLAKE.ML.DOCUMENT_INTELLIGENCE` privilege has not been granted to the role the Specialist is using.\*
- C. The `USAGE` privilege on the database used to create the model build has not been granted to the role the Specialist is using.
- D. The `SNOWFLAKE.DOCUMENT_INTELLIGENCE_CREATOR` database role has not been granted to the role the Specialist is using.

4. 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`

5. 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

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.