

Document Title: RAG Workflow Credential Setup

Goal: This guide explains how to generate the required API keys and OAuth credentials and input them into your n8n instance for the RAG workflow.

Service	n8n Node Used	Credential Type	Purpose in Workflow
Google Gemini	Google Gemini Chat Model	API Key	Provides access to the Large Language Model for generating the final chat response.
Pinecone	Pinecone Vector Store	API Key & Environment URL	Connects to the Vector Database for storing and retrieving knowledge embeddings.
Hugging Face	Embeddings HuggingFace...	Access Token	Authenticates the use of the Embeddings model for creating vectors from text.
Google Drive	file-download-gdrive	OAuth 2.0	Secures connection for downloading files from your Google Drive account.

Step-by-Step Credential Generation Instructions

1. Google Gemini API Key

1. **Go to Google AI Studio:** Navigate to [Google AI Studio](#) and sign in with your Google account.
2. **Create Key:** Go to the "API Keys" page (usually found under the Dashboard menu).
3. Click "Create API Key".
4. Copy the generated **API Key** immediately. **You will not be able to view it again.**
5. **Usage:** This key is used in the Google Gemini Chat Model node.

2. Pinecone API Key & Environment

1. **Go to Pinecone Console:** Log in to your Pinecone account (or sign up).
2. **Generate Key:** Navigate to the "API Keys" section in your dashboard.
3. Click "Create API Key" (if not already present).
4. Copy the generated **API Key** and the **Environment URL** (sometimes called the "Host" or "Region"). Both are required.
5. **Usage:** The API Key and Environment URL are used in the Pinecone Vector Store nodes for both ingestion and retrieval.

3. Hugging Face Access Token

1. **Go to Hugging Face:** Log in to your Hugging Face account.
2. **Settings:** Go to your **Profile Settings** and then select "Access Tokens" or "API Tokens."
3. **Create Token:** Click "New token," give it a clear name (e.g., "n8n RAG Embeddings"), and set the Role to "read."
4. Copy the generated **Token**.
5. **Usage:** This Token is used to authenticate the Embeddings HuggingFace Inference node.

4. Google Drive OAuth 2.0 Credentials

This process requires the Google Cloud Console and is the most complex.

1. **Google Cloud Project:** In the [Google Cloud Console](#), create or select a project.
 2. **Enable API:** Go to **APIs & Services > Library** and **Enable the Google Drive API**.
 3. **OAuth Consent Screen:** Navigate to **APIs & Services > OAuth consent screen**.
 - o Configure the screen (App Name, Support Email, etc.). Set User Type to **External**.
 4. **Create Credentials:** Go to **APIs & Services > Credentials** and click "**CREATE CREDENTIALS**" > "**OAuth client ID**."
 - o Select Application Type: **Web application**.
 - o **Crucial Step:** When setting up the credential in n8n (see below), n8n will provide a **Redirect URI** (e.g., <https://yourn8n.com/rest/oauth2-credential/callback>). **You must copy this URL** and paste it into the **Authorized redirect URIs** field in the Google Cloud Console.
 - o Click **Create** and copy the resulting **Client ID** and **Client Secret**.
-

Inputting Credentials in n8n

1. **Navigate to Credentials:** In your n8n instance, click the **Settings icon** () in the sidebar, and then select **Credentials**.
 2. **Create New:** Click "**Create new credential**."
 3. **Search:** Search for the service (e.g., "Google Drive" or "Gemini").
 4. **Input Data:** Paste the corresponding API Key, Token, Client ID, and Secret into the fields.
 5. **Authorize (for Google Drive):** For the Google Drive credential, after pasting the Client ID and Secret, click "**Sign in with Google**." Follow the prompts to grant n8n access to your Drive.
 6. **Save:** Name and save the credential. You will use this saved name when configuring the imported workflow nodes.
-

II. Deployment & Setup Guide (README)

Document Title: RAG Workflow Deployment Guide

This guide provides the necessary steps to import and run the RAG workflow in your n8n environment.

1. Prerequisites

Ensure the following items are prepared *before* importing the workflow:

Prerequisite	Status	Notes
n8n Instance	Active	The n8n instance must be accessible (e.g., via a public URL if using webhooks/chat triggers).
API Keys	Generated	All keys from the Credential Setup Guide must be generated and ready to input.
Pinecone Index	Created	A Pinecone Index must be created manually before the workflow runs. Note its name and the embedding dimension (e.g., 384 or 768).
Google APIs	Enabled	Google Drive API must be enabled in your Google Cloud Project.
Workflow JSON	Available	You must have the exported RAG workflow JSON file ready for import.

2. Workflow Import Instructions

1. Log in to your n8n instance.

2. Go to the **Workflows** section.
3. Click "Import from File" (or similar option, usually found in the top right menu).
4. Select the **RAG workflow JSON file** and upload it.
5. The workflow will appear on the canvas. **Do not execute it yet.**

3. Configuration and Credential Mapping

After import, you must configure every node showing an error or a placeholder credential name:

1. **Select Nodes and Connect Credentials:**
 - o Click on the **Google Gemini Chat Model** node (Pipeline 2).
 - o In the node settings, select your newly created **Gemini API Key** credential from the dropdown menu.
 - o Repeat this process for the **Pinecone Vector Store** nodes and the **Embeddings HuggingFace...** node.
 - o For **file-download-gdrive**, select your saved **Google Drive OAuth 2.0** credential.
2. **Configure Pinecone Index Name:**
 - o In both **Pinecone Vector Store** nodes, ensure the **Index Name** parameter matches the name of the Pinecone index you created in the prerequisites.
3. **Check Splitting:**
 - o Verify the settings in the **Recursive Character Text Splitter** node (Pipeline 1). The **Chunk Size** (e.g., 1000) and **Chunk Overlap** (e.g., 200) are critical for RAG accuracy.

4. Final Deployment Steps

1. **Initial Ingestion (One-Time Setup):**
 - o **Upload Data:** Place a sample document (e.g., a PDF manual) into the Google Drive folder targeted by the file-download-gdrive node.
 - o **Execute Pipeline 1:** Select all nodes in the left (**Ingestion**) pipeline and click "**Execute Selected Nodes**" (or run the entire Ingestion workflow if it has a manual trigger).
 - o **Verification:** Check your Pinecone console to confirm that the vectors were successfully created and stored in your index.
2. **Activate Chat Agent:**
 - o **Save:** Save the entire workflow.

- **Activate:** Toggle the workflow **ON** using the switch in the top right corner. The Chat Trigger node will now be active and listening for messages.
3. **Test:** Send a query to your activated chat endpoint (e.g., "What does the user manual say about installation?"). The agent should retrieve the relevant context and provide a grounded answer.