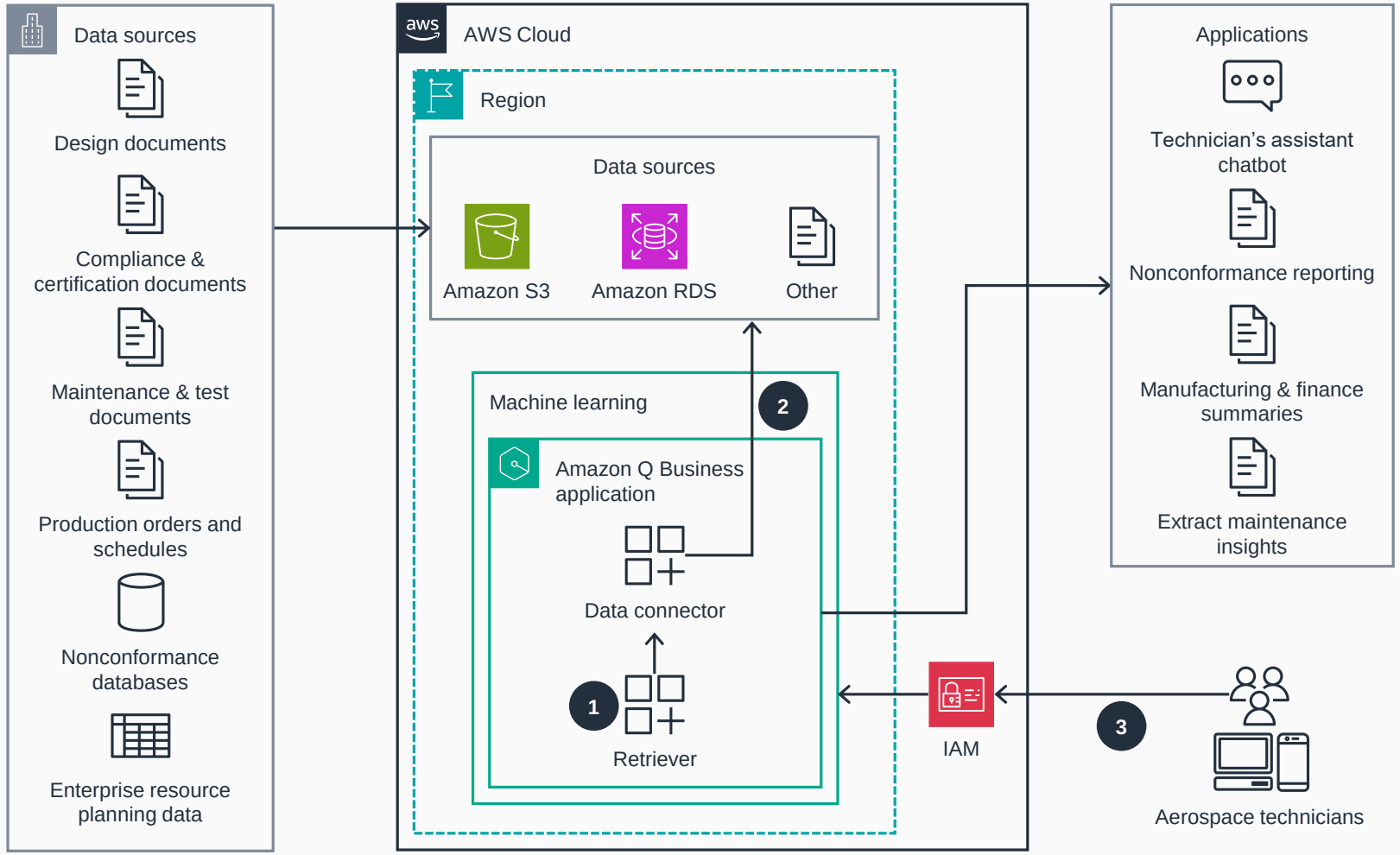
# Guidance for Aerospace Technician's Assistant on AWS

## Amazon Q

This architecture diagram shows how to use Amazon Q to enable natural language searching of paper documents.




1. Create an **Amazon Q Business** application and an associated retriever. **Amazon Q** provides retrievers both for **Amazon Kendra** indexes and for native indexes. You can select an **Amazon Q** retriever to use with an **Amazon Q** native index, or you can use an already-configured **Amazon Kendra** index as a retriever. The retriever pulls data from the index in real-time during a conversation.

2. Connect your data sources to your retriever. Available data sources will vary based on your choice of retriever. You can use **Amazon Simple Storage Service (Amazon S3)**, **Amazon Relational Database Service (Amazon RDS)**, or other data sources, and you can choose from supported data connectors.

3. Use **AWS Identity and Access Management (IAM) Identity Center** as a gateway for managing user access. For applications that use legacy identity management, **Amazon Q Business** requires that you integrate your web experience with an identity provider (IdP) that's compliant with SAML 2.0. **Amazon Q Business** can work with any IdP that's compliant with SAML 2.0. **Amazon Q** uses service-initiated single sign-on to authenticate users.

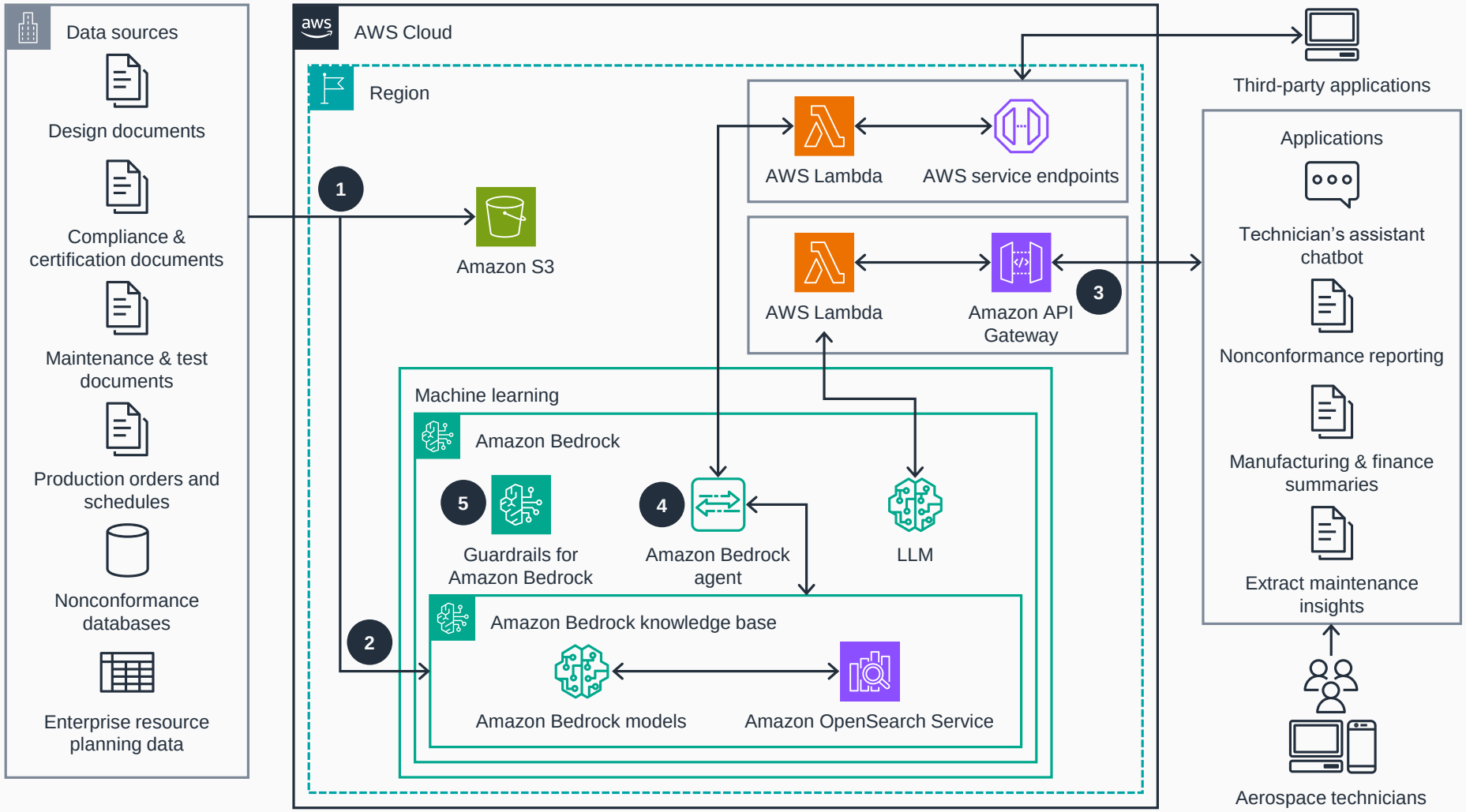
*Reviewed for technical accuracy July 8, 2024*


**AWS Reference Architecture**

# Guidance for Aerospace Technician's Assistant on AWS

## Amazon Bedrock

This architecture diagram shows how to use Amazon Bedrock to enable natural language searching of paper documents.




1. Upload data to **Amazon S3**.
2. Use **Knowledge Bases for Amazon Bedrock** to index your data by generating embeddings and storing them in a supported vector store, such as **Amazon OpenSearch Service**.
3. Query the large language model (LLM) and return augmented responses to your applications. Provide access to **Amazon Bedrock** APIs using **AWS Lambda** and **Amazon API Gateway** for external and AWS endpoints. **Amazon Bedrock** is a service that offers access to foundation models to build generative AI applications.
4. Use **Amazon Bedrock** agents to orchestrate and analyze tasks to break them down into the correct logical sequence using the foundation model's reasoning abilities. **Amazon Bedrock** agents automatically call the necessary APIs to transact with your systems to fulfill the request, determining along the way if they can proceed or if they need to gather more information.
5. As part of your responsible artificial intelligence (AI) strategy, you can now use **Guardrails for Amazon Bedrock** to promote safe interactions between users and your applications generated with AI. You can do this by implementing safeguards customized to your use cases and responsible AI policies.

*Reviewed for technical accuracy July 8, 2024*

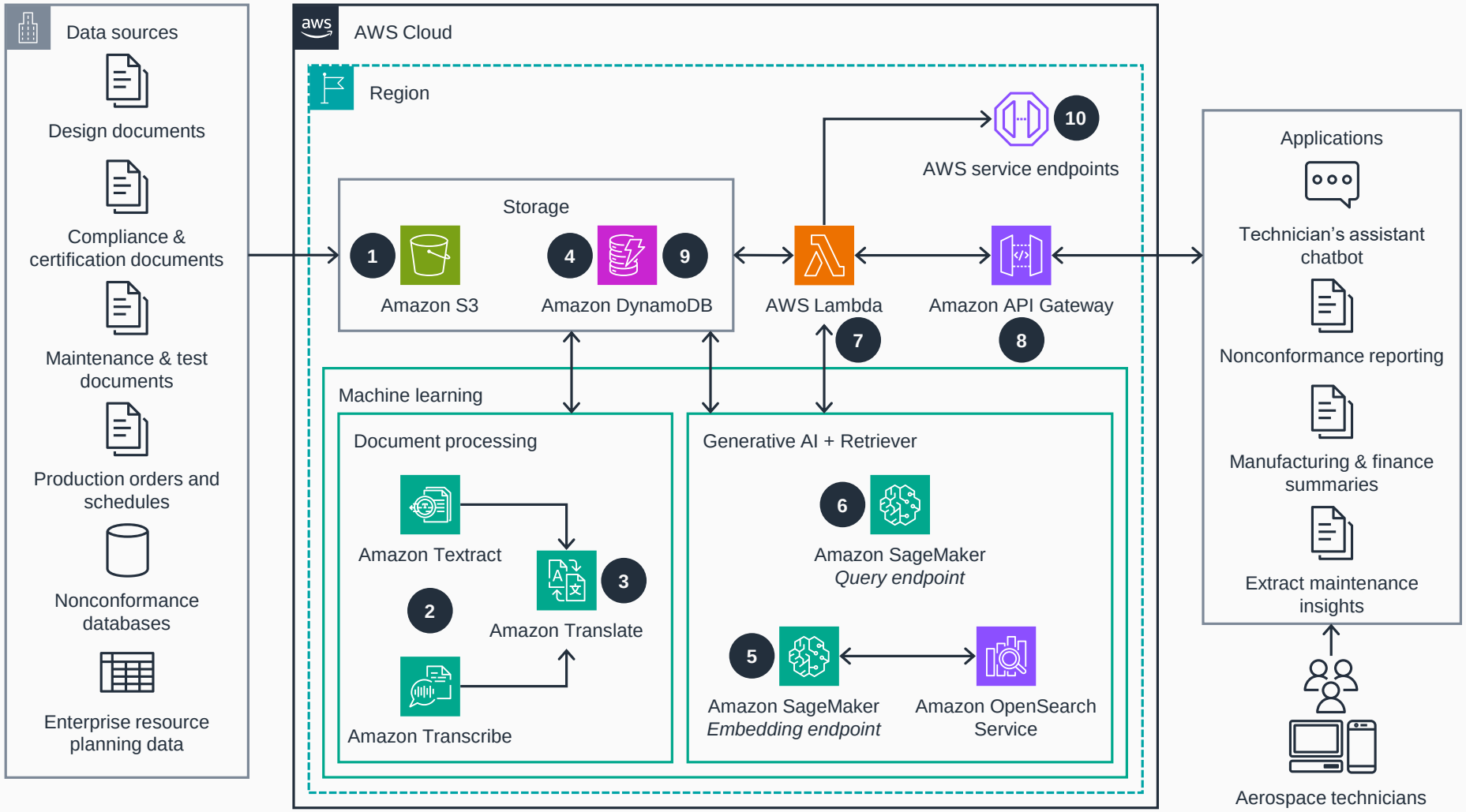


**AWS Reference Architecture**

# Guidance for Aerospace Technician's Assistant on AWS

## Amazon SageMaker

This architecture diagram shows how to use Amazon SageMaker to enable natural language searching of paper documents.




1. Upload data to **Amazon S3**.
2. Convert your files to text using **Amazon Textract** for PDF or image files. Use **Amazon Transcribe** for multimedia files. Store the converted text in **Amazon S3**.
3. If needed, use **Amazon Translate** to translate your text into a language compatible with your chosen LLMs. Store translated text in **Amazon S3**.
4. Store document processing status information and other relevant metadata in **Amazon DynamoDB**.
5. Create embeddings from your processed or translated text data using an embedding model on an **Amazon SageMaker** endpoint. Store the embeddings in an **OpenSearch Service** index.
6. Host an LLM on a **SageMaker** endpoint (the query endpoint) to process user queries.
7. Use **Lambda** to make queries to your **SageMaker** endpoints, your **OpenSearch Service** index, **DynamoDB**, and other AWS services.
8. Use **API Gateway** to provide access control, authorization, and API access to your **Lambda** functions.
9. Use **DynamoDB** to store chat information to enable a stateful chat interface.
10. Integrate your Retrieval-Augmented Generation(RAG) framework with other AWS services using AWS service endpoints.

*Reviewed for technical accuracy July 8, 2024*

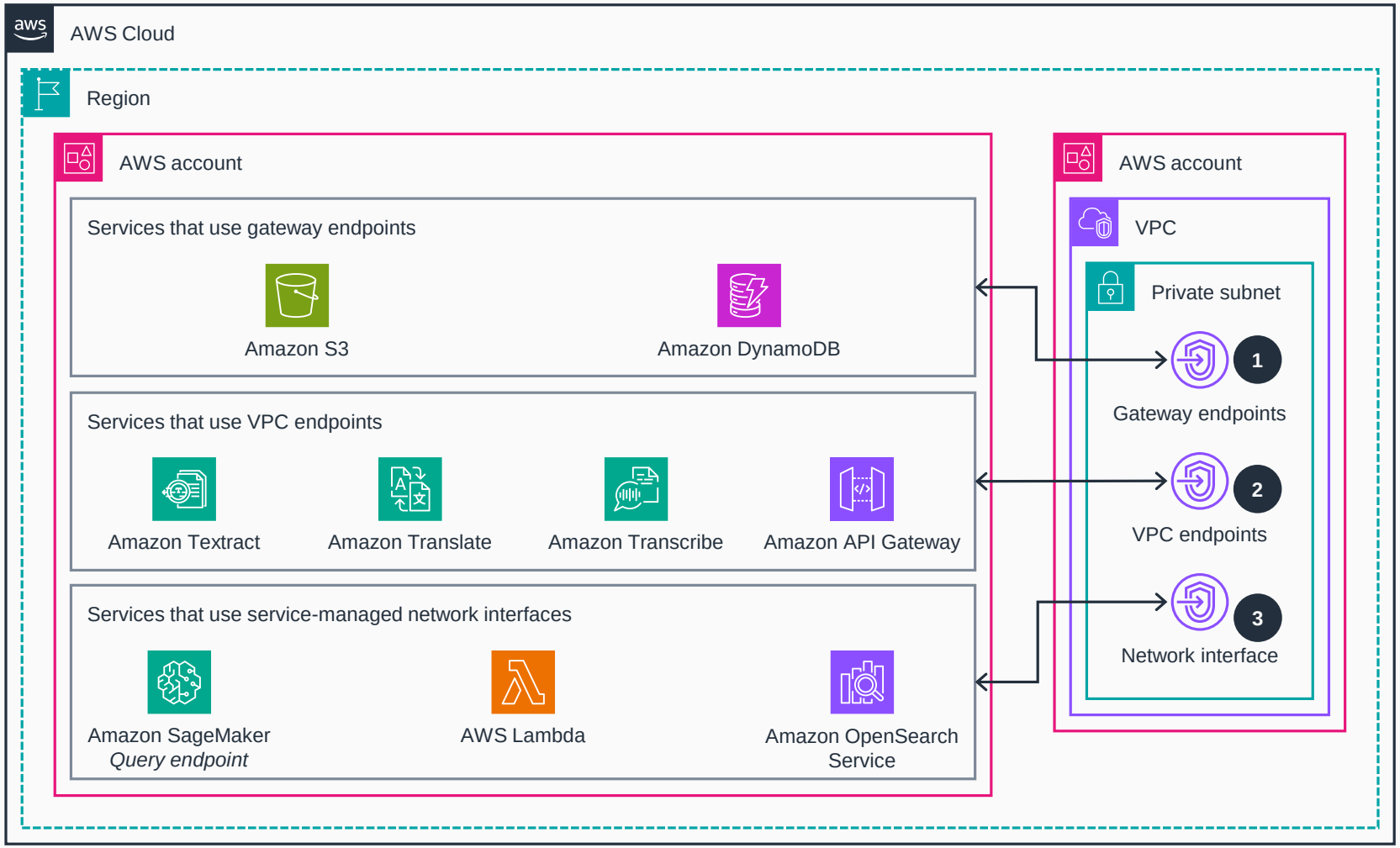


**AWS Reference Architecture**

# Guidance for Aerospace Technician's Assistant on AWS

## VPC Networking

This architecture diagram shows how to enable VPC networking for services used in generative AI solutions.




1. Connect **Amazon S3** and **DynamoDB** to your **Amazon Virtual Private Cloud (Amazon VPC)** environment using gateway endpoints.
2. Connect **Amazon Textract**, **Amazon Translate**, **Amazon Transcribe**, and **API Gateway** to your VPC using VPC endpoints.
3. Use service-managed elastic network interfaces to connect to **SageMaker** inference endpoints, **Lambda** functions, and **OpenSearch Service** domains.

*Reviewed for technical accuracy July 8, 2024*


**AWS Reference Architecture**