



Serverless x GenAI BKK Workshop



Generative AI at scale: Serverless workflows for enterprise-ready apps

[Pre-requisite] Enable foundation model access in Amazon Bedrock

[Pre-requisite] Configuring the front-end application

► Playground

▼ Use cases

▼ Building a RAG pipeline

Testing the need for RAG

Building the RAG Pipeline

Running the Pipeline

Testing the RAG Inference

▼ **AWS account access**

[Open AWS console \(us-west-2\)](#)

[Get AWS CLI credentials](#)

Exit event

[Event dashboard](#) > [Use cases](#) > [Building a RAG pipeline](#) > **Summary**

Summary

Congratulations on successfully completing the module!!!

In this module, you learned how to build a custom knowledge base using the OpenSearch Serverless database. You explored how to create a Retrieval-Augmented Generation (RAG) pipeline for processing thousands of documents using AWS Step Functions distributed map, Lambda functions, and LLamaIndex. Additionally, you learned how to manage pipeline failures during processing and how to control concurrency to prevent downstream services, such as Amazon Bedrock, from being overwhelmed by excessive requests.

Additional resources

- [Large scale data processing with distributed map](#)
- [Deploying Lambda function using container format](#)
- [Amazon Bedrock Knowledge base](#)

[Previous](#)

[Next](#)