Architecture Diagrams

Plagiarism Detection Architecture

Plagiarism Detection Architecture: Architecture Diagrams

Table of Contents

Plagiarism Detection Architecture

Publication date: **July 26, 2021 ([Diagram history](#))**

This architecture helps you create a plagiarism-detection service using AWS Step Functions, AWS Lambda, Amazon SageMaker AI, and OpenSearch Service.

Plagiarism Detection Architecture Diagram

1. Copy the document you'd like to run plagiarism detection on to **Amazon Simple Storage Service** (Amazon S3).

2. **Amazon S3** event triggers start of **AWS Step Functions** workflow.

3. **AWS Lambda** function extracts text from document using Tika (a content analysis toolkit that detects and extracts metadata and text from over a thousand different file types.

4. For each paragraph in the document, text is passed to a pre-trained Bidirectional Encoder Representations from Transformers (BERT)-based model to extract word embedding vectors.

5. For each word embedding vector, a K-Nearest Neighbor (KNN) search is run using a cosine-similarity algorithm.

6. **Amazon OpenSearch Service** (OpenSearch Service) domain stores an index of pre-processed works that have been converted into word embedding vectors and indexed.

7. Based on the configured similarity threshold that is compared against the **OpenSearch Service** query result score, an event bridge event is raised, specifying source document information that has possibly been plagiarized with reference to relevant works.
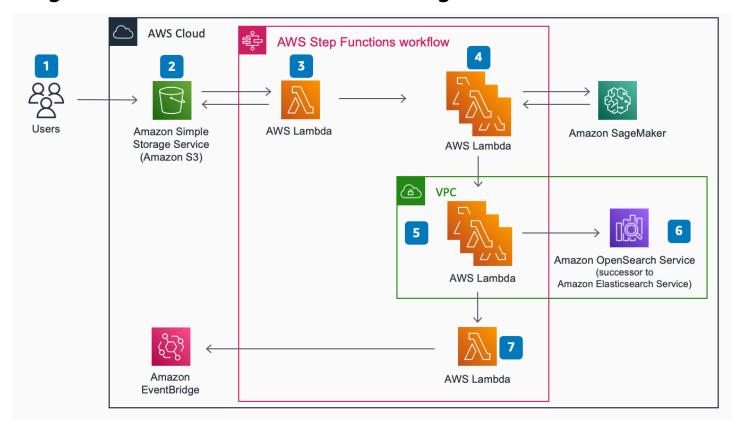
Download editable diagram

To customize this reference architecture diagram based on your business needs, download the ZIP file which contains an editable PowerPoint.

Create a free AWS account

Sign up now

Sign up for an AWS account. New accounts include 12 months of AWS Free Tier access, including the use of Amazon EC2, Amazon S3, and Amazon DynamoDB.

Further reading

For additional information, refer to

- AWS Architecture Icons
- AWS Architecture Center
- AWS Well-Architected

Diagram history

To be notified about updates to this reference architecture diagram, subscribe to the RSS feed.

Change	Description	Date
Initial publication	Reference architecture diagram first published.	July 26, 2021

> **ⓘ Note**
>
> To subscribe to RSS updates, you must have an RSS plugin enabled for the browser you are using.