
AI Powered Speech Analytics for Amazon Connect Implementation Guide



AI Powered Speech Analytics for Amazon Connect: Implementation Guide

Copyright © Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

Welcome	1
Overview	2
Cost	2
Cost estimate when deploying using AWS Lambda	3
Cost estimate when deploying using AWS Fargate	4
Architecture overview	7
Solution components	8
Live Media Streaming Feature	8
Web UI for Amazon Connect agents	8
Contact flows	8
Security	9
User Authentication	9
IAM roles	10
Amazon CloudFront	10
Design considerations	11
Deploy an alternative CloudFormation stack	11
Regional deployment	11
Updating the solution	12
AWS CloudFormation template	13
Automated deployment	14
Prerequisites	14
Find your Amazon Connect instance alias and ID	14
Deployment overview	14
Step 1. Launch the stack	15
Step 2. Create or update contact flow(s)	17
Step 3. Connect to the Solution's Web UI	18
Additional resources	19
Controlling the solution's behavior	20
Limitations	21
Update an Amazon Elastic Container Service task definition	22
Update an Amazon Elastic Container Service service	23
Edit an Amazon CloudWatch alarm	24
Update the stack	25
Uninstall the solution	26
Using the AWS Management Console	26
Using AWS Command Line Interface	26
Deleting the Amazon S3 Buckets	26
Collection of operational metrics	27
Source Code	28
Revisions	29
Contributors	30
Notices	31

Use AWS artificial intelligence services to transcribe, translate, and analyze your customer interactions

Publication date: *June 2019 (last update (p. 29): May 2021)*

This implementation guide discusses architectural considerations and configuration steps for deploying the AI Powered Speech Analytics for Amazon Connect solution on the Amazon Web Services (AWS) Cloud. It includes a link to an [AWS CloudFormation](#) template and instructions for manual steps to launch, configure, and run the AWS security, compute, storage, and other services required to deploy this solution on AWS, and using AWS best practices for security and availability.

The guide is intended for IT infrastructure architects, administrators, and DevOps professionals who have practical experience architecting on the AWS Cloud.

Overview

The AI Powered Speech Analytics for Amazon Connect solution provides customer insights in real time, helping agents and supervisors better understand and respond to customer needs. Using this solution, contact center agents can focus on the caller and track customer sentiment to resolve issues more effectively, improving the overall customer experience. Supervisors can review the call data, analyze the information, and provide timely feedback to agents. This solution includes pre-trained AWS artificial intelligence (AI) services that help you to transcribe, translate, and analyze each customer interaction in [Amazon Connect](#), and presents this information to assist contact center agents during their conversations

This solution uses the following AWS services:

- [Amazon Transcribe](#) to perform real-time speech recognition and create a high-quality text transcription of each call into text.
- [Amazon Comprehend](#) to analyze the interaction, detect the sentiment of the caller, and identify keywords and phrases in the conversation.
- [Amazon Translate](#) to translate the conversation into an agent's preferred language.

You can use the transcribed information with your knowledge management and customer relationship management (CRM) systems to suggest responses or recommended solutions to better respond to the caller's needs. Additionally, this solution stores the original audio files and conversation transcripts for later review or analysis.

This guide provides infrastructure and configuration information for planning and deploying this solution in the AWS Cloud.

Cost

You are responsible for the cost of the AWS services used while running this solution. As of March 2021, the cost for running this solution with default settings in the US East (N. Virginia) Region (excludes free tier) for a five-minute call is approximately **\$0.15 for the call and \$0.06 per day for the Direct Inward Dial (DID) number** for Amazon Connect.

Table 1: Amazon Connect costs for a five-minute call

AWS service	Dimensions	Cost
Amazon Connect	Voice usage service charge: at \$0.018 per minute * 5 Minutes	\$0.09000/5 min call
	Telephony pricing/per minute inbound usage: at \$0.0120 per minute * 5 Minute	\$0.06000/5 min call
	At \$0.06 per day * 1 day	\$0.06000/day

The additional costs for the five-minute call when deploying the AWS CloudFormation template using the default configuration (AWS Lambda) and other AWS services is approximately **\$0.67 for the call**. Refer to Table 2 for a cost breakdown.

If deploying the template using the [alternative configuration \(p. 11\)](#) (AWS Fargate), the cost for the five-minute call is approximately **\$1.06 for the call**. Refer to Table 3 for a cost breakdown. These cost estimates assume the solution processes 180 words per minute and two segments (customer and agent) per minute.

Cost estimate when deploying using AWS Lambda

By default, this solution deploys with AWS Lambda to process the audio from customer calls. Besides Amazon Connect costs provided in Table 1, the following table provides example costs based on a five-minute call for the other AWS services needed to run the solution.

Table 2: Cost estimate using AWS Lambda

AWS service	Dimensions	Cost for a five-minute call
Amazon API Gateway	\$0.25/million connection minutes * 5 minutes	\$0.00000
	\$1.00/million messages * 500 messages	\$0.00050
Amazon CloudFront	\$0.0100 per 10,000 HTTPS requests * 52,000 requests	\$0.00005
	\$0.085 per GB * 0.006 GB data transfer out	\$0.00051
Amazon CloudWatch	\$0.50 per GB * 0.002 GB for log data ingestion	\$0.00100
	\$0.03 per GB * 0.000016 GB/month for logs storage	\$0.00000
Amazon Comprehend	\$0.0001 per unit * 168 units for entity detection	\$0.01680
	\$0.0001 per unit * 168 units for key phrase detection	\$0.01680
	\$0.0001 per unit * 489 units for sentiment analysis	\$0.04890
Amazon DynamoDB	\$0.00013 per RCU * 360 read capacity	\$0.04680
	\$0.00065 per WCU * 360 write capacity	\$0.23400
	\$0.02 per 100,000 DynamoDB streams read request units * 1	\$0.00000
	\$0.25 per GB * 0.00000144 GB/month storage	\$0.00000

AI Powered Speech Analytics for
Amazon Connect Implementation Guide
Cost estimate when deploying using AWS Fargate

AWS service	Dimensions	Cost for a five-minute call
Amazon Kinesis Video Streams	\$0.0085 per GB of data read from Kinesis Video Streams * 0.009 GB	\$0.00008
	\$0.0085 per GB of data ingested into Kinesis Video Streams * 0.005 GB	\$0.00004
	\$0.023 Per GB/month of data stored in Kinesis Video Streams * 0.000234 GB/month	\$0.00001
AWS Key Management Service	\$0.03 per 10,000 requests * 577 requests	\$0.00173
AWS Lambda	\$0.0000166667 for every GB/second * 466.592 seconds	\$0.00778
	\$0.20 per 1M requests * 527.000 requests	\$0.00011
Amazon Simple Storage Service	\$0.005 per 1,000 PUT, COPY, POST, or LIST requests * 90.000 requests	\$0.00045
	\$0.004 per 10,000 GET and all other requests * 7,517.000 requests	\$0.00301
	\$0.023 per GB - first 50 TB/month of storage used * 0.006 GB/month	\$0.00014
Amazon Transcribe	\$0.00040 streaming transcriptions per second * 208.770 seconds	\$0.08351
Total cost for a 5-minute call:		\$0.67222

Important

Different AWS services have different units of measurement for determining a cost estimate. As a result, do **NOT** multiply the example call time in this example cost breakdown table to calculate an estimate for longer calls. We recommend creating a [budget](#) through [AWS Cost Explorer](#) to help manage costs.

Cost estimate when deploying using AWS Fargate

This solution provides an alternative deployment configuration using AWS Fargate to process the audio from customer calls. Besides Amazon Connect costs provided in Table 1, the following table provides example costs based on a five-minute call for the other AWS services needed to run the solution.

Table 3: Cost estimate using AWS Fargate

AI Powered Speech Analytics for
Amazon Connect Implementation Guide
Cost estimate when deploying using AWS Fargate

AWS service	Dimensions	Cost for a five-minute call
Amazon API Gateway	\$0.25/million connection minutes * 5 minutes	\$0.00000
	\$1.00/million messages * 486 messages	\$0.00049
Amazon CloudFront	\$0.0100 per 10,000 HTTPS requests * 132 requests	\$0.00013
	\$0.085 per GB * 0.003 GB data transfer out	\$0.00026
Amazon CloudWatch	\$0.10 per alarm * 0.099 alarms	\$0.00990
	\$0.30 per metric * 0.817 metrics	\$0.24510
	\$0.50 per GB * 0.008 GB log data ingestion	\$0.00400
	\$0.03 per GB * 0.000048 GB/month logs storage	\$0.00000
Amazon Comprehend	\$0.0001 per unit * 162 units for entity detection	\$0.01620
	\$0.0001 per unit * 162 units for key phrase detection	\$0.01620
	\$0.0001 per unit * 372 units for sentiment analysis	\$0.03720
Amazon DynamoDB	\$0.00013 per RCU * 360 read capacity	\$0.04680
	\$0.00065 per WCU * 360 write capacity	\$0.23400
	\$0.02 per 100,000 DynamoDB streams read request units * 1	\$0.00000
	\$0.25 per GB * 0.000000510 GB/month storage	\$0.00000
Amazon Kinesis Video Streams	\$0.0085 per GB of data read from Kinesis Video Streams * 0.010 GB	\$0.00009
	\$0.0085 per GB of data ingested into Kinesis Video Streams * 0.005 GB	\$0.00004
	\$0.023 Per GB/month of data stored in Kinesis Video Streams * 0.000104 GB/month	\$0.00000
AWS Key Management Service	\$0.03 per 10,000 requests * 601 requests	\$0.00180

AI Powered Speech Analytics for
Amazon Connect Implementation Guide
Cost estimate when deploying using AWS Fargate

AWS service	Dimensions	Cost for a five-minute call
AWS Lambda	\$0.0000166667 for every GB/second * 11.479 seconds	\$0.00019
	\$0.20 per 1M requests * 493.000 requests	\$0.00010
Amazon Simple Storage Service	\$0.005 per 1,000 PUT, COPY, POST, or LIST requests * 90.000 requests	\$0.00045
	\$0.004 per 10,000 GET and all other requests * 7,517.000 requests	\$0.00301
	\$0.023 per GB - first 50 TB/month of storage used * 0.006 GB/month	\$0.00014
Amazon Transcribe	\$0.00040 streaming transcriptions per second * 250.487 seconds	\$0.10019
Amazon Elastic Container Registry	\$0.10 per GB/month of data storage * 0.004 GB/month	\$0.00040
Amazon Elastic Compute Cloud	\$0.045 per GB data processed by NAT Gateways * 0.485 GB	\$0.02183
	\$0.045 per NAT Gateway hour * 1 hour	\$0.04500
Amazon Elastic Container Service	\$0.004445 per GB per hour * 1 hour	\$0.00445
	\$0.04048 per vCPU per hour * 1 hour	\$0.04048
AWS Key Management Service	\$0.03 per 10,000 requests * 601 requests	\$0.00180
Amazon Simple Queue Service	At \$0.40 per 1 million requests * 15 requests	\$0.00001
AWS CodeBuild	At \$0.005 per build minute * 3 minutes	\$0.01500
Total cost for a 5-minute call:		\$1.05525

Important

Different AWS services have different units of measurement for determining a cost estimate. As a result, do **NOT** multiply the example call time in this example cost breakdown table to calculate an estimate for longer calls. We recommend creating a [budget](#) through [AWS Cost Explorer](#) to help manage costs.

Prices are subject to change. For full details, refer to the pricing webpage for each AWS service you will be using in this solution.

Architecture overview

Deploying this solution builds the following environment in the AWS Cloud.

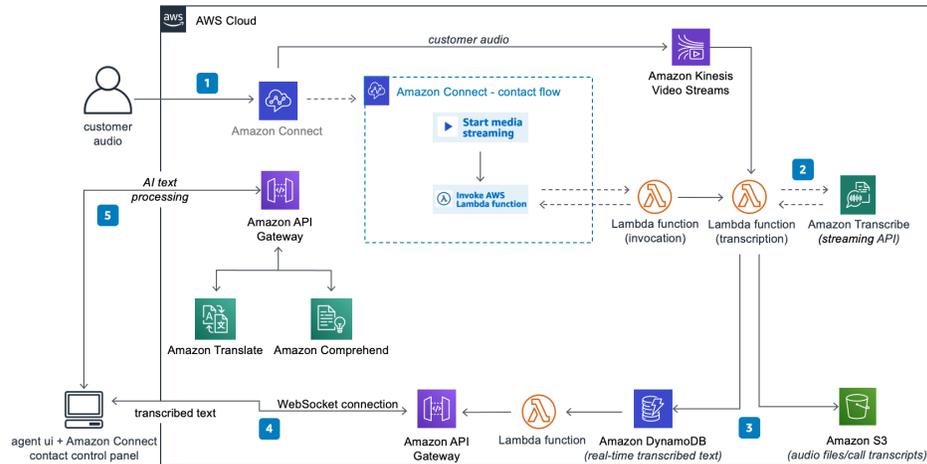


Figure 1: AI Powered Speech Analytics for Amazon Connect architecture on AWS

This solution extends your existing Amazon Connect environment by deploying the AWS services necessary to transcribe, translate, and analyze customer interactions.

1. When a customer calls into your Amazon Connect call center, their call progresses through a [contact flow](#). In the contact flow, the [Start media streaming contact block](#) captures customer audio. Then the [Invoke AWS Lambda function contact block](#) activates the invocation [AWS Lambda](#) function. As the call runs, the customer audio is streamed in real time to [Amazon Kinesis Video Streams](#).
2. The transcription Lambda function consumes the audio stream and uses Amazon Transcribe to convert the audio into text.
3. The transcription Lambda function then stores the transcript segments and contact ID in an [Amazon DynamoDB](#) table. After the call ends, the captured transcripts will be uploaded to the solution-created [Amazon Simple Storage Service \(Amazon S3\)](#) bucket, with the Amazon S3 location added as an attribute to the Amazon Connect contact trace record. Contact center supervisors can mine the contact trace records for additional insights to improve the overall customer experience.
4. When the call is routed to an agent, the agent's call center application establishes a [WebSocket](#) connection to an [Amazon API Gateway](#).
5. The customer audio transcript is provided in real time to the agent. Additionally, Amazon Translate and Amazon Comprehend can provide translated and annotated transcripts, allowing the agent to efficiently find relevant information and perform recommended actions.

Note

AWS Lambda has a limitation where it stops audio processing when calls are longer than 15 minutes. To overcome this limitation, you can deploy an alternative setup. For more information, refer to [Deploy an alternative CloudFormation stack \(p. 11\)](#).

Solution components

Live Media Streaming Feature

This solution uses the Amazon Connect [live media streaming](#) feature to capture the customer audio during an interaction with your contact center by sending the audio to an Amazon Kinesis video stream.

Web UI for Amazon Connect agents

This solution provides a sample web UI for Amazon Connect agents to answer calls and receive transcripts, translations, and call sentiment analysis. You can create and manage agents using the User Management section of the Amazon Connect portal.

You can control the behavior of this solution using the Amazon Connect contact flow attributes. For more information about how these attributes are set in the sample contact flow, refer to [Controlling the solution's behavior \(p. 20\)](#).

Contact flows

Customer calls to be recorded, transcribed, and translated must be routed to this solution through appropriately configured Amazon Connect contact flows. Contact flows must be configured to start media streaming, invoke the `initContactDetails` and `kvsConsumerTrigger` AWS Lambda functions, and store the AWS Security Token Service (AWS STS) tokens in contact attributes. For more information, refer to [Step 2 \(p. 17\)](#) in the Automated deployment section.

Security

When you build systems on AWS infrastructure, security responsibilities are shared between you and AWS. This [shared model](#) can reduce your operational burden as AWS operates, manages, and controls the components from the host operating system and virtualization layer down to the physical security of the facilities in which the services operate. For more information about AWS security, visit [AWS Cloud Security](#).

User Authentication

The AI Powered Speech Analytics for Amazon Connect solution uses a combination of Amazon Connect and AWS Security Token Service (AWS STS) to authenticate users. Agents log into a web UI using Amazon Connect user credentials to set their availability and answer calls. AWS STS provides temporary credentials for agents to access Amazon API Gateway, Amazon Comprehend, Amazon Translate, and to store the complete customer transcription history.

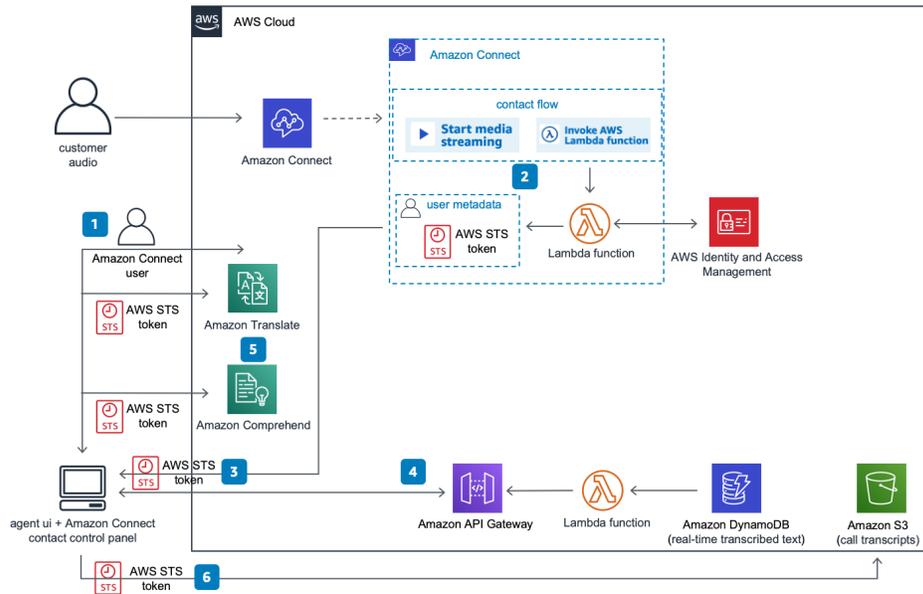


Figure 2: User authentication diagram

1. Agents log into the web client using their Amazon Connect credentials and set their availability to answer calls.
2. When an end user calls into the call center, Amazon Connect invokes an AWS Lambda function to create an AWS STS token, and stores the token in the call attributes.
3. The agent uses the web client to accept a call. After accepting a call, the web client uses the Amazon Connect SDK to retrieve the AWS STS token from the call attributes and updates the AWS SDK configuration to use the AWS STS token for authenticating AWS API calls.
4. The web client uses the AWS STS token to create an authenticated web socket connection to Amazon API Gateway, provides their 16-character connection ID and 32-character contact ID, and receives real-time transcriptions as they are stored in Amazon DynamoDB.
5. The web client uses the AWS STS token to send the transcriptions to Amazon Comprehend and Amazon Translate for sentiment analysis and translation throughout the call.

6. Once the call is completed, the web client uses the AWS STS temporary credentials to store the transcription in the customer-defined Amazon Simple Storage Service (Amazon S3) bucket, and the AWS STS credentials are removed from the call attributes.

IAM roles

AWS Identity and Access Management (IAM) roles allow customers to assign granular access policies and permissions to services and users on the AWS Cloud. This solution creates IAM roles that grant the solution's AWS Lambda functions access to create Regional resources.

Amazon CloudFront

This solution deploys a web console [hosted](#) in an Amazon Simple Storage Service (Amazon S3) bucket. To help reduce latency and improve security, this solution includes an Amazon CloudFront distribution with an origin access identity, which is a CloudFront user that provides public access to the solution's website bucket contents. For more information, refer to [Restricting Access to Amazon S3 Content by Using an Origin Access Identity](#) in the *Amazon CloudFront Developer Guide*.

Design considerations

Deploy an alternative CloudFormation stack

By default, this solution deploys with AWS Lambda to process the audio from customer calls. However, AWS Lambda has a limitation where audio processing stops if a call lasts longer than 15 minutes. To avoid this limitation, you can deploy this solution using [Amazon SQS](#) and [Amazon ECS](#) that process the audio for up to one hour. To launch this alternative setup, you must change the **ComputeType** parameter from the default `Lambda` setting to `Fargate`. For more information about this parameter, refer to [Step 1. Launch the stack \(p. 15\)](#) in the Automated deployment section.

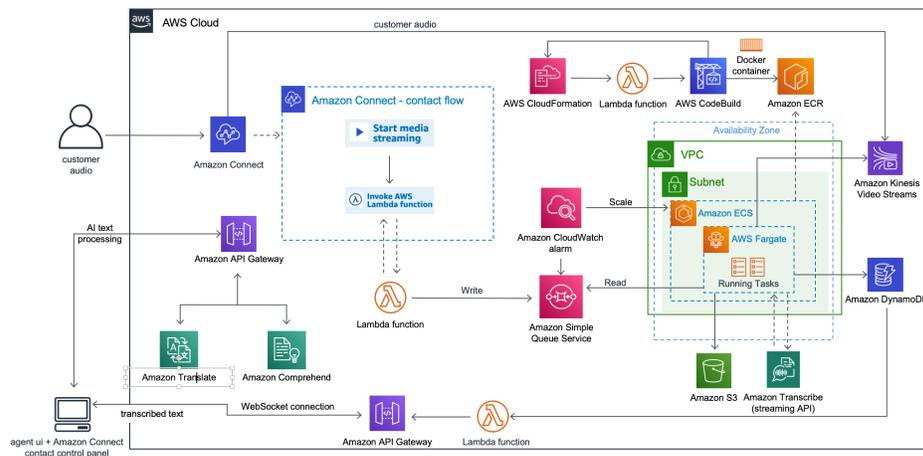


Figure 3: AI Powered Speech Analytics for Amazon Connect architecture on AWS using AWS Fargate

This alternative deployment deploys [AWS CodeBuild](#) to create a [Docker image](#) for the application and uploads it to an [Amazon ECR](#) repository to allow an [AWS Fargate](#) task to locate it. An Amazon ECS cluster is created along with an associated Fargate task definition to run the container's logic.

When a customer calls into your Amazon Connect call center, their call progresses through the contact flow which invokes a Lambda function to emit a message to the SQS queue. Customer audio is streamed in real time through [Amazon Kinesis Video Streams](#).

The Fargate task polls for SQS messages, then consumes the audio stream and uses Amazon Transcribe to convert the audio into text. The Fargate task then stores the transcript segments and contact ID in an [Amazon DynamoDB](#) table.

This alternative deployment uses two CloudWatch alarms to check for the number of items in the queue, and to scale the number of ECS containers. One alarm scales up the number of ECS containers while the other alarm scales the number of containers down.

Regional deployment

AI Powered Speech Analytics for Amazon Connect uses Amazon Connect, Amazon Translate, Amazon Comprehend, and Amazon Kinesis Video Streams, which are currently available in specific AWS Regions only. Therefore, you must launch this solution in an AWS Region where these services are available. For the most current service availability by Region, refer to the [AWS Regional Services List](#).

Note that Amazon Transcribe Streaming is not supported in every region where Amazon Transcribe is supported. Refer to [Amazon Transcribe endpoints and quotas](#) for a list of AWS Regions where Amazon Transcribe and Amazon Transcribe Streaming are available.

Updating the solution

If you have previously deployed the solution, you must update the solution's CloudFormation stack to get the latest version of the solution's framework. For details, refer to [Update the stack \(p. 25\)](#).

AWS CloudFormation template

This solution uses AWS CloudFormation to automate the deployment of AI Powered Speech Analytics for Amazon Connect on the AWS Cloud. It includes the following AWS CloudFormation template, which you can download before deployment:

[View
Template](#)

AI-powered-speech-analytics-for-amazon-connect: Use this template to launch the solution and all associated components. The default configuration deploys Amazon Translate, Amazon Comprehend, AWS Lambda, Amazon Kinesis Video Streams, Amazon API Gateway, Amazon CloudFront, and Amazon DynamoDB. The alternative configuration deploys all the AWS services in the default configuration plus AWS Fargate, Amazon ECS, Amazon ECR, Amazon SQS, and AWS CodeBuild. You can also customize the template based on your specific needs.

Automated deployment

Before you launch this solution, review, review the architecture, prerequisites, and other considerations discussed in this guide. Follow the step-by-step instructions in this section to configure and deploy this solution into your account.

Note

If you have previously deployed the solution, you must update the solution's CloudFormation stack to get the latest version of the solution's framework. For details, refer to [Update the stack \(p. 25\)](#).

Time to deploy: Approximately 20 minutes

Prerequisites

Before you launch this solution, you must have an existing Amazon Connect instance with the live media streaming feature activated. If you don't have an existing Amazon Connect instance, refer to [Getting Started with Amazon Connect](#) in the *Amazon Connect Administrator Guide*. To activate live media streaming, refer to [Use Live Media Streaming in a Contact Flow](#).

After you deploy the solution, you must configure contact flow(s) to activate the speech analytics functionality. The solution provides an example contact flow to get you started. For more information, refer to [Step 2 \(p. 17\)](#).

Find your Amazon Connect instance alias and ID

This solution requires an Amazon Connect instance alias and instance ID. Use this procedure to find your alias and ID.

1. Sign in to the [Amazon Connect console](#).
2. Find and copy the applicable Amazon Connect instance alias.
3. Select the Amazon Connect instance alias, and select **Overview** in the left-hand menu list.
4. In the **Instance ARN** field, copy the 32 character Amazon Connect instance ID. The instance ID can be found at the end of the ARN. For example: `arn:aws:connect:<Region>:<account-id>:instance/<11111111-2222-3333-4444-555555555555>`

Deployment overview

Use the following steps to deploy this solution on AWS. For detailed instructions, follow the links for each step. Before launching the template verify that your Amazon Connect instance is configured correctly.

[Step 1. Launch the stack \(p. 15\)](#)

- Launch the AWS CloudFormation template into your AWS account.
- Enter values for the required parameters: **Stack Name**, **Instance ID**, **Instance Alias**, **Website Bucket Name**, and **Call Audio Bucket Name**
- Review the other template parameters, and adjust if necessary.

[Step 2. Create or update contact flow\(s\) \(p. 17\)](#)

- Create, update, or import sample contact flow(s) for calls to be recorded, transcribed, and translated.

Step 3. Connect to the Solution's Web UI (p. 18)

- Use the example web UI to answer calls and view transcription, translation, and sentiment analysis in action.

Step 1. Launch the stack

This automated AWS CloudFormation template deploys AI Powered Speech Analytics for Amazon Connect in the AWS Cloud. Confirm that you have copied the instance alias and instance ID, and activated the Live Media Streaming feature before launching the stack.

Note

You are responsible for the cost of the AWS services used while running this solution. For more details, refer to the [Cost \(p. 2\)](#) section in this guide, and refer to the pricing webpage for each AWS service used in this solution.

1. Sign in to the AWS Management Console and select the button to launch the `AI-powered-speech-analytics-for-amazon-connect` AWS CloudFormation template.



Alternatively, you can [download the template](#) as a starting point for your own implementation.

2. The template is launched in the US East (N. Virginia) Region by default. To launch this solution in a different AWS Region, use the region selector in the console navigation bar.

Note

This solution uses Amazon Connect, Amazon Translate, Amazon Comprehend, and Amazon Kinesis Video Streams services, which are currently available in specific AWS Regions only. Therefore, you must launch this solution in an AWS Region where these services are available. For the most current service availability by Region, refer to the [AWS Regional Services List](#).

3. On the **Create stack** page, verify that the correct template URL is in the **Amazon S3 URL** text box and choose **Next**.
4. On the **Specify stack details** page, assign a name to your solution stack. For information about naming character limitations, refer to [IAM and STS Limits](#) in the *AWS Identity and Access Management User Guide*.
5. Under **Parameters**, review the parameters for this solution template and modify them as necessary. This solution uses the following default values.

Parameter	Default	Description
Instance ID	<Requires input>	The Amazon Connect instance ID for the Amazon Connect instance that this solution will use. To find the instance ID, refer to Find Your Amazon Connect Instance Alias and ID (p. 14) .
Instance Alias	<Requires input>	The instance alias for the Amazon Connect instance that this solution will use.

AI Powered Speech Analytics for
Amazon Connect Implementation Guide
Step 1. Launch the stack

Parameter	Default	Description
Website Bucket Name	<Requires input>	Unique name of the solution-created Amazon S3 bucket that will be used to store the website assets used for the demo web UI.
Call Audio Bucket Name	<Requires input>	Unique name of the solution-created Amazon S3 bucket that will be used to store the call recording files. Note that this must be a different name than what is used for the Website Bucket Name .
Audio File Prefix	recordings/	The Amazon S3 prefix where the audio files will be saved (must end in "/")
Test Mono Audio Prefix	audio-file-input/	Specifies the Amazon S3 prefix where raw/wav (audio/L16; mono; 8 kHz) audio recordings are uploaded to process an audio file. This parameter is used mainly for testing purposes or for real-time transcription of audio files. This parameter only works with single channel (mono) files.
Price Class	PriceClass_100/	Specifies the CloudFront price class. Refer to Amazon CloudFront Pricing for a description of each price class.
Compute Type	Lambda	By default, this solution uses an AWS Lambda function to transcribe call audio consumed from Kinesis Video Streams. However, AWS Lambda has a maximum run time of 15 minutes per invocation. To overcome this limitation, change this parameter to Fargate to deploy an alternative set up.
sessionDuration	3600	Specifies the maximum call duration time that will be processed by this solution. The available range is from 900 seconds (15 minutes) to 3600 seconds (1 hour). Note that one hour is the maximum amount of time that can be processed by this solution.

6. Choose **Next**.
7. On the **Configure stack options** page, choose **Next**.
8. On the **Review** page, review and confirm the settings. Check the box acknowledging that the template will create AWS Identity and Access Management (IAM) resources.
9. Choose **Create stack** to deploy the stack.

You can view the status of the stack in the AWS CloudFormation console in the **Status** column. You should receive a **CREATE_COMPLETE** status in approximately 20 minutes.

Note

In addition to the primary AWS Lambda functions, the solution includes the `CustomResourceHelper` Lambda function, which runs only during initial configuration or when resources are updated or deleted.

When running this solution, the `CustomResourceHelper` function is inactive. However, do not delete the `CustomResourceHelper` function as it is necessary to manage associated resources.

Step 2. Create or update contact flow(s)

This solution includes an example contact flow that is configured for your Amazon Connect instance. Use the following steps to import the example contact flow into your Amazon Connect instance.

1. Download the example contact flows, `kvsStreamingSampleFlow` and `agentWhisperKvsStreamingSampleFlow`. The example flows are located in the root of the Call Audio Bucket using the bucket name specified when launching the solution. For example, `call-audio-bucket/kvsStreamingSampleFlow`.
2. **Import the example contact flows** into your Amazon Connect instance, using the following procedure:
 - a. Open the Amazon Connect portal for your instance.
 - b. In the navigation pane, select **Routing**, then choose **Contact flows**. You can log in to your instance by navigating to: `https://<your-instance-alias>.awsapps.com/connect/login`.
 - c. Do one of the following, determined by whether you are creating a new stack or updating an existing one:
 - If creating a new stack, select **Create agent whisper flow** to import the flow.
 - When updating the stack, replace the existing `agentWhisperKvsStreamingSampleFlow` contact flow with the one you are importing.
 - d. Select the `agentWhisperKvsStreamingSampleFlow` file you saved earlier and choose **Import**.
 - e. Choose **Publish**.
 - f. In the navigation pane, select **Routing**, then choose **Contact flows**.
 - g. Do one of the following, determined by whether you are creating a new stack or updating an existing one:
 - If creating a new stack, select **Create contact flow** to import the flow.
 - When updating the stack, replace the existing `kvsStreamingFlow` contact flow with the one you are importing.
 - h. Select the drop-down arrow next to **Save** in the top right-hand corner, and choose **Import flow**.
 - i. Select the `kvsStreamingSampleFlow` file you saved earlier and choose **Import**.
 - j. Choose **Publish**.
 - k. In the navigation pane, select **Routing**, then choose **Phone numbers** and associate a phone number to the newly created `kvsStreamingFlow` contact flow.

Step 3. Connect to the Solution's Web UI

This solution provides a sample web UI which allows agents to answer calls and view call transcriptions, translations, and sentiment analysis. Use the following steps to configure Amazon Connect and log into the web UI.

1. Sign in to the [AWS CloudFormation console](#).
2. Select the **Outputs** tab and copy the value of the `cloudfrontEndpoint` key.
3. Navigate to the [Amazon Connect console](#).
4. Find and select the applicable **Amazon Connect instance alias**.
5. In the navigation pane, select **Approved origins**. Then select the **Add origin** link.
6. Enter the `cloudfrontEndpoint` domain name (for example, `https://<distribution>.cloudfront.net`) found in [Step 2 \(p. 17\)](#), and select **Add**.
7. Use a browser to navigate to `cloudfrontEndpoint` domain. Log in with your Amazon Connect agent credentials, and set the agent's status to **Available**. Note that it may take up to 30 minutes for the CloudFront endpoint and website to be available after the stack creation is completed.
8. Call your Amazon Connect dial-in number configured previously in [Step 2 \(p. 17\)](#).
9. Answer the call using the solution's web UI.
10. Once the call is completed you can view the transcription in the solution's Amazon DynamoDB table, view the logs in the stream for the `Transcriber` Lambda function, and listen to the call recording in the Amazon S3 bucket you specified in the `Transcriber` Lambda function.

Note

To ensure that transcription is available, verify that the live media streaming option is activated. For information, refer to the [Prerequisites \(p. 14\)](#) section.

Additional resources

AWS services

- [AWS CloudFormation](#)
- [AWS Lambda](#)
- [Amazon Connect](#)
- [Amazon Transcribe](#)
- [Amazon Translate](#)
- [Amazon S3](#)
- [Amazon Kinesis Video Streams](#)
- [Amazon Simple Queue Service](#)
- [Amazon Elastic Container Service](#)
- [AWS Fargate](#)
- [Amazon Elastic Container Registry](#)
- [AWS CodeBuild](#)

Controlling the solution's behavior

The AI Powered Speech Analytics for Amazon Connect solution has flags you can use in the `kvsConsumerTrigger` AWS Lambda function which will then pass the attributes to the `kvsTranscriber` Lambda function. The following flags control the behavior of this solution.

Lambda function	Default	Description
<code>transcribeCall</code>	<code>true</code>	Enable call transcriptions.
<code>saveCallRecording</code>	<code>false</code>	Choose whether to save the audio captured from the live audio stream to the customer-defined Amazon S3 bucket.
<code>languageCode</code>	<code>en-US</code> or <code>es-US</code>	The language the customer selected when calling in, and transcribed to the agent using Amazon Transcribe. For more information on supported languages and codes, refer to Streaming Transcription in the <i>Amazon Transcribe Developer Guide</i> .

The `kvsTranscriber` Lambda function has the following flags that can be modified.

Lambda function	Default	Description
<code>CONSOLE_LOG_TRANSCRIPT_FLAG</code>	<code>true</code>	Activate call logging for transcripts in CloudWatch.
<code>SAVE_PARTIAL_TRANSCRIPTS</code>	<code>true</code>	Logs and saves all partial transcripts from Amazon Transcribe. If you set this to <code>false</code> you will only see the final result returned from Amazon Transcribe.
<code>START_SELECTOR_TYPE</code>	<code>now</code>	Used to tell Transcribe when to start the transcription. Use <code>now</code> to start transcription once the agent and customer are connected. Select <code>FRAGMENT_NUMBER</code> to start transcription once the Start Media Streaming block is invoked in the contact flow.

This solution includes an `agentAssist.html` file that contains a mapping of key phrases to suggested actions. You can customize this file to fit your needs. For example, the file contains a mapping that suggest the action to *offer discount* if Amazon Comprehend detects the key phrase *cancel my service*.

Limitations

Consider the following limitations before using AI Powered Speech Analytics for Amazon Connect in a production environment.

- When a new AWS Lambda container is used for a call, it is considered a cold start and has a latency of approximately 10 seconds. When using this solution in a production environment where latency needs to be at zero seconds, we recommend increasing the memory of the Lambda function so new containers are used less often. Or, you can use a different AWS service such as AWS Fargate. AWS Lambda stops transcribing if a call is longer than 15 minutes. Using the AWS Fargate alternative overcomes this limitation.
- For security purposes, this solution uses role chaining to assume a role with temporary security credentials using the `GetSessionToken` API. It cannot assume the role for more than an hour. As a result, translation and sentiment analysis cannot run for more than an hour. You can customize this solution to implement a refresh mechanism for the temporary token by calling the `GetSessionToken` API or use the long-term AWS security credentials of the AWS account root user or an IAM user to overcome this limitation. For additional information, refer to [GetSessionToken](#) in the *AWS Security Token Service API Reference Guide*.
- The first call using the Amazon Kinesis video stream streaming block, may fail. If this happens wait 10 seconds and try again. Also, the number of concurrent calls cannot exceed the number of KVS streams, so you must manually increase this number of streams based on your estimated traffic.
- The first few seconds of a call may not be transcribed. If this happens and impacts your transcriptions, we recommend changing the `START_SELECTOR_TYPE` Lambda function value to `FRAGMENT_NUMBER`.

Update an Amazon Elastic Container Service task definition

For information about updating an Amazon ECS task definition, refer to [Updating a task definition](#) in the Amazon ECS Developer Guide.

Update an Amazon Elastic Container Service service

For information about updating an Amazon ECS service, refer to [Updating a service](#) in the Amazon ECS Developer Guide.

Edit an Amazon CloudWatch alarm

For information about editing an Amazon CloudWatch alarm, refer to [Editing or Deleting a CloudWatch Alarm](#) in the Amazon CloudWatch User Guide.

Update the stack

Complete the following steps to update your AWS CloudFormation stack to the current version.

1. Sign in to the [AWS CloudFormation console](#).
2. Select your existing AI Powered Speech Analytics for Amazon Connect CloudFormation stack, and choose **Update**.
3. On the **Update stack** page, verify that **Replace current template** is selected.
 - In the **Specify template** section, select **Amazon S3 URL**.
 - Copy the link of the [latest template](#).
 - Paste the link in the Amazon S3 URL box.
 - Verify that the correct template URL shows in the **Amazon S3 URL** text box and choose **Next**.
4. On the **Specify stack details** page, under **Parameters**, review the parameters for the template and modify them as necessary. Refer to [Step 1. Launch the stack \(p. 15\)](#) in the Automated deployment section for details about the parameters.
5. Choose **Next**.
6. On the **Configure stack options** page, choose **Next**.
7. On the **Review** page, review and confirm the settings. Check the box acknowledging that the template might create AWS Identity and Access Management (IAM) resources.
8. Choose **View change set** and verify the changes.
9. Choose **Update stack** to deploy the stack.

You can view the status of the stack in the AWS CloudFormation console in the **Status** column. You should receive an **UPDATE_COMPLETE** status in approximately 20 minutes. After the stack has been updated, proceed to [Step 2 \(p. 17\)](#) in the Automated deployment section to update the contact flows.

Uninstall the solution

You can uninstall the AI Powered Speech Analytics for Amazon Connect solution from the AWS Management Console, or by using the AWS Command Line Interface (AWS CLI). You must manually delete the Amazon Simple Storage Service (Amazon S3) bucket created by this solution. AWS Solutions Implementations do not automatically delete this resource in case you have stored data that you wish to retain.

Using the AWS Management Console

1. Sign in to the [AWS CloudFormation console](#).
2. On the **Stacks** page, select the solution stack.
3. Choose **Delete**.

Using AWS Command Line Interface

Verify that the AWS Command Line Interface (AWS CLI) is available in your environment. For installation instructions, refer to [What Is the AWS Command Line Interface](#) in the *AWS CLI User Guide*. After confirming the AWS CLI is available, run the following command.

```
$ aws cloudformation delete-stack --stack-name <installation-stack-name>
```

Deleting the Amazon S3 buckets

This solution is configured to retain the Amazon S3 buckets if you decide to delete the AWS CloudFormation stack to prevent against accidental data loss. After uninstalling the solution, you can manually delete the S3 bucket if you do not need to retain the data. Follow these steps to delete the Amazon S3 bucket.

1. Sign in to the [Amazon S3 console](#).
2. Choose **Buckets** from the left navigation pane.
3. Locate the *<stack-name>* S3 buckets.
4. Select the S3 bucket and choose **Delete**.

To delete the S3 buckets using AWS CLI, run the following command:

```
$ aws s3 rb s3://<bucket-name> --force
```

Collection of operational metrics

This solution includes an option to send anonymous operational metrics to AWS. We use this data to better understand how customers use this solution and related products and services. When enabled, the following information is collected and sent to AWS:

- **Solution ID:** The AWS solution identifier
- **Unique ID (UUID):** Randomly generated, unique identifier for each solution deployment
- **Timestamp:** Data-collection timestamp
- **Answered Calls:** The number of incoming answered calls
- **Dashboard Views:** The number of times the dashboard is viewed

AWS owns the data gathered through this survey. Data collection is subject to the [AWS Privacy Policy](#). To opt out of this feature, modify the AWS CloudFormation template mapping section as follows:

```
"Send" : {  
  "AnonymousUsage" : { "Data" : "Yes" }  
},
```

to

```
"Send" : {  
  "AnonymousUsage" : { "Data" : "No" }  
},
```

Source Code

You can visit our [GitHub repository](#) to download the templates and scripts for this solution, and to share your customizations with others. Refer to the [README.md file](#) for additional information.

Revisions

Date	Change	
June 2019	Initial release	
March 2020	Bug fixes	
April 2021	Release version 1.1.0: Added an alternative setup option using AWS Fargate. For more information, refer to the CHANGELOG.md file in the GitHub repository.	
May 2021	Release version 1.1.1: Updated the Docker build image. For more information, refer to the CHANGELOG.md file in the GitHub repository.	

Contributors

- Yasser El-Haggan
- Niranjan Hira
- Ramprasad Srirama
- Karan Grover
- Girish Mallenahally

Notices

Customers are responsible for making their own independent assessment of the information in this document. This document: (a) is for informational purposes only, (b) represents AWS current product offerings and practices, which are subject to change without notice, and (c) does not create any commitments or assurances from AWS and its affiliates, suppliers or licensors. AWS products or services are provided “as is” without warranties, representations, or conditions of any kind, whether express or implied. AWS responsibilities and liabilities to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

AI Powered Speech Analytics for Amazon Connect is licensed under the terms of the MIT No Attribution at [The Software Package Data Exchange](#).