

# Using AWS Batch and Amazon ECS to implement an AI-powered data transformation pipeline for the Bat Conservation Trust

The Bat Conservation Trust (BCT) needed to accelerate the analysis of their vast library of bat audio recordings to better understand bat populations and inform conservation efforts. Previously, sequentially processing their 100TB+ of annual data on manual instances caused significant bottlenecks, delaying critical insights by weeks.

**100TB+**

of bat call audio data collected annually

**Weeks**

of processing time reduced to minutes

**1.9m+**

audio processing jobs run in AWS Batch

By leveraging AWS Batch and scalable AWS Fargate spot instances, a new AI-powered audio classification pipeline was implemented, enabling highly reliable parallel processing of terabytes of data. This dramatically reduced processing times from weeks to minutes, optimized compute costs, and has allowed BCT to significantly bolster their conservation efforts across the UK.

Turn over for a deep dive

“ Before, running datasets through the pipeline took weeks and hours of engineering effort. Now, data is processed seamlessly via automation in minutes or hours. ”

Lia Gilmour, Research Manager, Bat Conservation Trust

“ Our audio processing pipeline was completely transformed! We simply don't have to worry about scaling or performance issues anymore. ”

Lia Gilmour, Research Manager, Bat Conservation Trust

## Who are Bat Conservation Trust?

The Bat Conservation Trust (BCT) is a non-profit organisation based in the United Kingdom that is dedicated to the conservation of bats and the habitats they reside in. They run a range of monitoring and research programmes to increase our understanding of bats, and engage in education and advocacy to raise awareness and support conservation efforts.

Bat Conservation Trust



## Spotlight: AWS Batch

AWS Batch automates the provisioning and scaling of compute resources for batch jobs. Efficiently run high-volume workloads at any scale while optimizing costs and minimizing management overhead.



## Trusted by leading names on AWS



## We are experts in



### AWS Migrations

From on-premise and other cloud providers to AWS



### AWS Modernization

Future-proof your stack with containerization, IaC and serverless



### Data and Analytics

Unlock and turbocharge your data with pipelines and analytics



### AI on AWS

Stay ahead of the curve with GenAI and ML



### AWS Security

Secure your cloud with tailored security reviews



### AWS Managed Services

Place your cloud in the hands of experts

To find out more please contact [sales@lambertlabs.com](mailto:sales@lambertlabs.com)

# Using AWS Batch and Amazon ECS to implement an AI-powered data transformation pipeline for the Bat Conservation Trust

## Opportunity / Customer Challenge

The Bat Conservation Trust (BCT) collects terabytes of audio recordings to monitor bat populations across the UK. Analysing this vast dataset is crucial for understanding bat behaviour, tracking species distribution, and informing conservation efforts. Their existing process, which relied on manually configured Amazon EC2 instances, proved to be a significant bottleneck. Processing terabytes of data sequentially meant that only a limited number of files could be analysed at any given time, resulting in extremely long processing times and delaying critical insights.

This slow processing speed hampered the BCT's ability to efficiently analyse the growing volume of data, hindering their research and conservation work. They needed a solution that could drastically reduce processing time and enable them to analyse their extensive audio library more effectively.

## Solution

To address the processing bottleneck, a parallelised audio classification pipeline leveraging Batch and Fargate was implemented. The core of this solution lies in the efficient distribution of processing jobs across a scalable container environment. This was made possible after containerizing the audio analysis AI application.

Upon upload to Amazon S3, audio files trigger an event notification that pushes a message to an Amazon Simple Queue Service queue. This queue acts as a buffer, decoupling the upload process from the processing pipeline. An AWS Lambda function monitors the SQS queue and, for each message (representing an audio file), submits a job to a Batch job queue.

Critically, the Batch jobs utilise Fargate spot instances. This serverless compute engine allows for rapid scaling and cost optimisation by leveraging spare compute capacity. Each job spins up a container running the audio analysis AI application. This application uses the latest version of the bat call classifier, trained on the latest data, to analyse the audio file. The container annotates the recording with the predicted bat species and other detected features. The results of the analysis are then stored in an Amazon Aurora Serverless database, providing centralised storage for further analysis and reporting.

## Outcome

The implementation of the parallelised audio classification pipeline has significantly enhanced BCT's ability to analyse their vast audio archive. Processing time for the terabytes of recordings has been drastically reduced, enabling researchers to gain insights from their data much more quickly. Where previously analysis was limited, the new pipeline can now process a vastly greater number of audio files concurrently, scaling automatically with the volume of data uploaded.

The use of Batch with Fargate spot instances has also improved the reliability of the analysis process. Each audio file is now processed independently by a dedicated job, making it easier to track progress, identify any failures, and retry individual files without affecting others. This granular approach ensures that no single point of failure can halt the entire analysis pipeline.

Furthermore, the solution has delivered significant cost benefits. By leveraging spot instances, the BCT now pays only for the compute resources used during each job, eliminating the overhead of running EC2 instances for extended periods of time. This has resulted in lower compute costs and more predictable budget allocation.

Finally, the containerized deployment process has streamlined application updates. Changes to the audio analysis AI application are now deployed by pushing a new Docker image, ensuring a more reliable and automated update process compared to manually configuring and deploying to an EC2 instance.

Trusted by leading names on AWS



BALLOTPEDIA

To find out more please contact [sales@lambertlabs.com](mailto:sales@lambertlabs.com)