NewsT (News Translator)

Diana-Isabela Crainic, Ioan Sava

March 2022

Abstract

News are important for a number of reasons within a society. Mainly to inform the public about events that are around them and may affect them. *NewsT*, the proposed solution in this report aims to make news from various sources more accessible to people around the world by translating them. A reliable system has been built based on the services provided by the Amazon Web Services (AWS) cloud computing platform.

1 Introduction

Nowadays, it is essential that important news to be accessible to anyone, regardless of their nationality or culture. The language barrier can represent an obstacle to this goal. The proposed solution in this report, *NewsT*, aims to address this problem through a system that collects news from various sources, translates them into several languages and make them available to the interested customers.

This solution is based on numerous cloud computing services provided by *Amazon Web Services*[1] (AWS). Benefits such as flexibility, cost-effectiveness, reliability, scalability, performance and security led to the choice of this platform for the project development.

2 Components

2.1 Overview

The application is composed of 3 main components, fully decoupled: the ***News collector*** component, the ***News processor*** component and the ***News collection*** component. The architecture is illustrated in the diagram below. The communication is done in an asynchronous way. The ***News collector*** component is an on-premise application, which consumes events from an

[1] `https://aws.amazon.com/`

external API and publishes messages in an *Amazon Simple Notification Service*[2] topic. The messages are polled by ***News processor*** from an *Amazon Simple Queue Service*[3] queue, processed and inserted in an *Amazon DynamoDB*[4] database. The ***News collection*** component fetches the news from this database based on specific criteria.

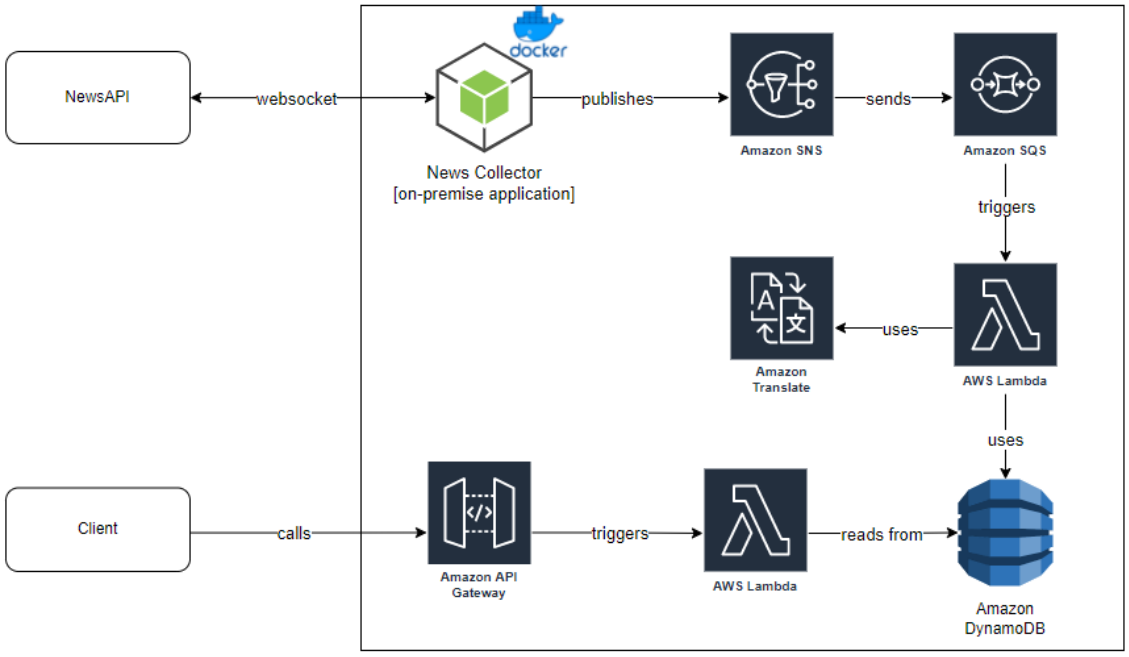

Figure 1: Application architecture

2.2 News collector component

News collector represents the first component of the system. This is an on-premise application implemented using *Node.js*[5]. Its responsibility is to consume events published by an external API (News API[6]) through WebSockets[7]. These events represent news that contain various attributes associated with them, such as title, author or url. An example of such an event is presented below.

[2] https://aws.amazon.com/sns
[3] https://aws.amazon.com/sqs/
[4] https://aws.amazon.com/dynamodb/
[5] https://nodejs.org/en/
[6] https://newsapi.org/docs/client-libraries/node-js
[7] https://developer.mozilla.org/en-US/docs/Web/API/WebSockets_API

```
{
  "title": "If everyone on Earth sat in the ocean at once, how much would sea level rise?",
  "author": "Yahoo News",
  "description": "Hypothetical questions, like what would happen if everyone on Earth went for an ocean swim at once, are fun to think about. And using math, you can get pretty close to a real answer.",
  "publishedAt": "2022-02-25T20:11:35Z",
  "url": "https://news.yahoo.com/everyone-earth-sat-ocean-once-120708401.html"
}
```

These events are published as messages in an *Amazon Simple Notification Service*[8] topic through the *aws-sdk*[9] *SNS* client.

The application was deployed in a *Docker*[10] container, using a **node:lts-alpine** image. The multiple benefits of the containers have led to the implementation of this approach: isolation, cost-effectiveness, ability to run anywhere, flexibility.

2.3 News processor component

This component is a *Node.js Amazon Lambda* function that processes messages in an *Amazon Simple Queue Service* (Amazon SQS) queue. These are the news emitted by the ***News collector*** component through the *Amazon Simple Notification Service* topic mentioned above. Message processing consists of translating the title and the content of the news using the *Amazon Translate*[11] neural machine translation. The translated news are stored in a *DynamoDB* table (*NewsTable*). The languages the news are translated into are also fetched from a *DynamoDB* table (*LanguagesTable*). The interaction with the *Amazon Translate* and *Amazon DynamoDB* is performed with the help of the *DynamoDB* and the *Amazon Translate* aws-sdk clients.

2.4 News collection component

The ***News collection*** component is composed of a *Node.js Amazon Lambda function* which reads data from a *DynamoDB* table (*NewsTable*). This lambda is triggered by an *API Gateway* (*NewsAPI*) which has a single endpoint

[8] https://aws.amazon.com/sns
[9] https://www.npmjs.com/package/aws-sdk
[10] https://www.docker.com/
[11] https://aws.amazon.com/translate/

/news. A call to this api may contains 2 query parameters: *lang*, the wanted language of the news and *num* which represents the number of news it is intended to be returned. Depending on these parameters, the lambda function queries the *DynamoDB* table using the *aws-sdk DynamoDB* client.

3 Services

3.1 Overview

Amazon Web Services offers a broad set of global cloud-based products including compute, storage, databases, analytics, networking, mobile, developer tools, management tools, IoT, security, and enterprise applications: on-demand, available in seconds, with pay-as-you-go pricing. From data warehousing to deployment tools, directories to content delivery, over 200 AWS services are available. New services can be provisioned quickly, without the upfront fixed expense. This allows enterprises, start-ups, small and medium-sized businesses, and customers in the public sector to access the building blocks they need to respond quickly to changing business requirements. [12]

3.2 Amazon Simple Notification Service

Amazon Simple Notification Service (Amazon SNS) is a fully managed messaging service for both application-to-application (A2A) and application-to-person (A2P) communication.

The A2A pub/sub functionality provides topics for high-throughput, push-based, many-to-many messaging between distributed systems, microservices, and event-driven serverless applications. Using Amazon SNS topics, your publisher systems can fanout messages to a large number of subscriber systems, including Amazon SQS queues, AWS Lambda functions, HTTPS endpoints, and Amazon Kinesis Data Firehose, for parallel processing. The A2P functionality enables you to send messages to users at scale via SMS, mobile push, and email.[13]

The NASA Image and Video Library provides easy access to more than 140,000 still images, audio recordings, and videos—documenting NASA's more than half a century of achievements in exploring the vast unknown. The architecture includes Amazon SNS to trigger the processing pipelines when new content is updated, and Amazon SQS to decouple incoming jobs from pipeline processors.[14]

[12] `https://docs.aws.amazon.com/whitepapers/latest/aws-overview/introduction.html`

[13] `https://aws.amazon.com/sns`

[14] `https://aws.amazon.com/solutions/case-studies/nasa-image-library/`

3.3 Amazon Simple Queue Service

Amazon Simple Queue Service (SQS) is a fully managed message queuing service that enables you to decouple and scale microservices, distributed systems, and serverless applications. SQS eliminates the complexity and overhead associated with managing and operating message-oriented middleware, and empowers developers to focus on differentiating work. Using SQS, you can send, store, and receive messages between software components at any volume, without losing messages or requiring other services to be available.[15]

The BMW Group is using AWS for its connected car application that collects sensor data from BMW 7 Series cars to give drivers dynamically updated map information. BMW built its new car-as-a-sensor (CARASSO) service in only six months leveraging Amazon SQS, Amazon Simple Storage Service (S3), Amazon DynamoDB, Amazon Relational Database Service (RDS), and AWS Elastic Beanstalk.[16]

3.4 AWS Lambda

AWS Lambda is a serverless, event-driven compute service that lets you run code for virtually any type of application or backend service without provisioning or managing servers. You can trigger Lambda from over 200 AWS services and software as a service (SaaS) applications, and only pay for what you use.[17]

Fender went all in on AWS to better engage with customers and gain supply-chain efficiencies. Fender, made famous by the likes of Jimi Hendrix, is the world's largest maker of stringed instruments by revenue. Fender runs its Fender Play, Fender Tune, and Fender Tone apps on AWS, which help customers learn to play guitar, tune their instruments, and control digital amplifiers. The company uses AWS services including AWS Lambda, Amazon DynamoDB, and Amazon API Gateway to store and deliver more than 700 TB of video and more than 4.9 million lessons to customers.[18]

3.5 Amazon Translate

Amazon Translate is a neural machine translation service that delivers fast, high-quality, affordable, and customizable language translation. Neural machine translation is a form of language translation automation that uses deep learning models to deliver more accurate and more natural sound-

[15] https://aws.amazon.com/sqs/
[16] https://aws.amazon.com/solutions/case-studies/bmw/
[17] https://aws.amazon.com/lambda/
[18] https://aws.amazon.com/solutions/case-studies/fender

ing translation than traditional statistical and rule-based translation algorithms.

With Amazon Translate, you can localize content such as websites and applications for your diverse users, easily translate large volumes of text for analysis, and efficiently enable cross-lingual communication between users. [19]

One customer of many that makes use of this service is *Siemens*, an industry, energy, and healthcare conglomerate based in Germany. It uses Amazon Translate to conduct internal employee surveys. Given the 50+ languages spoken by their 377,000 global employees, the organization could previously only afford to conduct employee surveys once a year—a problem that Amazon Translate has rendered moot.[20]

3.6 Amazon DynamoDB

DynamoDB is a fast and flexible NoSQL database service for all applications that need consistent, single-digit-millisecond latency at any scale. Its flexible data model and reliable performance make DynamoDB a great fit for mobile, web, gaming, advertising technology, Internet of Things, and other applications.[21]

Duolingo uses Amazon DynamoDB to store 31 billion items in support of an online learning site that delivers lessons for 80 languages. The U.S. startup reaches more than 18 million monthly users around the world who perform more than six billion exercises using the free Duolingo lessons. The company relies heavily on Amazon DynamoDB not just for its highly scalable database, but also for high performance that reaches 24,000 read units per second and 3,300 write units per second.[22]

3.7 Amazon API Gateway

Amazon API Gateway is a fully managed service that makes it easy for developers to create, publish, maintain, monitor, and secure APIs at any scale. APIs act as the "front door" for applications to access data, business logic, or functionality from your backend services. Using API Gateway, you can create RESTful APIs and WebSocket APIs that enable real-time two-way communication applications. API Gateway supports containerized and serverless workloads, as well as web applications.

API Gateway handles all the tasks involved in accepting and processing up to hundreds of thousands of concurrent API calls, including traffic management, CORS support, authorization and access control, throttling,

[19] https://aws.amazon.com/translate/
[20] https://aws.amazon.com/solutions/case-studies/siemens-translate/
[21] https://aws.amazon.com/translate/
[22] https://aws.amazon.com/solutions/case-studies/duolingo-case-study-dynamodb/

monitoring, and API version management. API Gateway has no minimum fees or startup costs. You pay for the API calls you receive and the amount of data transferred out and, with the API Gateway tiered pricing model, you can reduce your cost as your API usage scales.[23]

PhotoVogue is an online photography platform. Launched in 2011 and part of Vogue Italia, it allows upcoming photographers to showcase their work. It also gives them the chance to have their photos published in Vogue or be picked by photography agencies. Each picture that's submitted is carefully reviewed by Vogue Italia's editorial staff, who ensure only the highest quality images appear online. According to Marco Viganò, the Head of Digital Development, "With Amazon API Gateway and AWS Lambda, the user experience is up to 90% faster. That's for both photographers uploading images and the editorial team processing them."[24]

4 Conclusion

The proposed solution is composed of three components: the ***News collector*** component, the ***News processor*** component and the ***News collection*** component, harmonized in an *Amazon Web Services*[25] (AWS) infrastructure.

There were used services like AWS Lambda, Amazon Translate, Amazon DynamoDB. These services are also used by big companies like NASA, BMW, Siemens to ensure the scalability and the performance of the systems.

[23] https://aws.amazon.com/api-gateway/
[24] https://aws.amazon.com/solutions/case-studies/photovogue/
[25] https://aws.amazon.com/

References

[1] NASA Case Study (2017)
https://aws.amazon.com/solutions/case-studies/nasa-image-library/.

[2] BMW Case Study (2015) AWS re:Invent 2015
https://aws.amazon.com/solutions/case-studies/bmw/.

[3] Fender Case Study (2018) AWS re:Invent 2018
https://aws.amazon.com/solutions/case-studies/fender/.

[4] Siemens Case Study (2017)
https://aws.amazon.com/solutions/case-studies/siemens-translate/.

[5] Duolingo Case Study (2016)
https://aws.amazon.com/solutions/case-studies/duolingo-case-study-dynamodb/.

[6] PhotoVogue Case Study (2017)
https://aws.amazon.com/solutions/case-studies/photovogue/.