Homework 2

Distributed Application using AWS Cloud Services ☁️

> Amarandei Matei Alexandru, Chontas Panagiotis, Tudose Eduard, Zaharia Andrei - MISS

Overview 🌍

This document presents a scientific report based on the services chosen for the first part of the homework. We decided to use Amazon Web Services as our cloud ecosystem of choice, in order to create the distributed application using AWS Amplify, API Gateway, AWS Lambda, IAM Roles, DynamoDB, and, of course WebSockets technology.

What is AWS? 🧐

Amazon Web Services, Inc. (AWS) is a subsidiary of Amazon that provides on-demand cloud computing platforms and APIs to individuals, companies, and governments, on a metered, pay-as-you-go basis. Often times, clients will use this in combination with autoscaling (a process that allows a client to use more compute in times of high application usage, and then scale down to reduce costs when there is less traffic). These cloud computing web services provide various services related to networking, compute, storage, middleware, IOT and other processing capacity, as well as software tools via AWS server farms.

AWS Amplify

AWS Amplify is a complete solution that lets frontend web and mobile developers easily build, ship, and host full-stack applications on AWS, with the flexibility to leverage the breadth of AWS services as use cases evolve. No cloud expertise needed.

AWS 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 containerised and server-less 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, 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.

AWS Lambda

AWS Lambda is a server-less, event-driven compute service that lets you run code for virtually any type of application or backend service without provisioning or managing servers. It is colloquially known as the reference example when discussing Function-as-a-Service (FaaS). You can trigger Lambda from over 200 AWS services and Software-as-a-Service (SaaS) applications, and only pay for what you use.

IAM Roles

An IAM *role* is an IAM identity that you can create in your account that has specific permissions. It is similar to an IAM user, in that it is an AWS identity with permission policies that determine what the identity can and cannot do in AWS. However, instead of being uniquely associated with one person, a role is intended to be assumable by anyone who needs it. Also, a role does not have standard long-term credentials such as a password or access keys associated with it. Instead, when you assume a role, it provides you with temporary security credentials for your role session.

DynamoDB

DynamoDB is a fully managed NoSQL database service, designed to store and retrieve any amount of data, and serve any level of request traffic. It is built on a highly available and scalable architecture that can handle millions of requests per second with single-digit millisecond latency.

DynamoDB automatically scales to accommodate changing workloads and storage needs, without any downtime or performance impact. It is designed to be highly available, with built-in replication and automatic failover capabilities to ensure

uninterrupted service. Amongst Performance, Flexibility and Manageability, it also shines when it comes down to Security.

DynamoDB provides multiple layers of security, including encryption at rest and in transit, access control through IAM, and fine-grained access control through VPC endpoints and conditionals.

Amazon S3 (initially included, but later removed)

Amazon Simple Storage Service (Amazon S3) is an object storage service offering industry-leading scalability, data availability, security, and performance. Customers of all sizes and industries can store and protect any amount of data for virtually any use case, such as data lakes, cloud-native applications, and mobile apps. With cost-effective storage classes and easy-to-use management features, you can optimise costs, organise data, and configure fine-tuned access controls to meet specific business, organisational, and compliance requirements.

While S3 didn't make the final cut and was opted out of in favor of DynamoDB, it is still a technology worth mentioning, as it could have been used to achieve the same end goal.

Real systems that use AWS 🤩

In this part of the document we will talk about companies that used the components mentioned above and how they improved the applications.

AWS Amplify

This component has:

- **Amplify Studio (Visual Interface)**: Point-and-click environment to build and deploy a full-stack app quickly, including frontend UI and backend.
- **Amplify CLI (Command Line Interface** Local toolchain to configure and manage an app backend with just a few commands.
- **Amplify Libraries (Open-Source Client Libraries):** Open-source client libraries to build cloud powered mobile and web apps.
- **Amplify UI Components (Open-Source Design System):** Open-source design system with cloud-connected components for building feature-rich apps fast.

- **Amplify Web Hosting (Managed CI/CD and Hosting):** Fully managed CI/CD and hosting for fast, secure, and reliable static and server-side rendered apps.
- **Works with popular Frameworks and Languages (Angular, React, Java etc)**

Customers that used AWS Amplify in their implementation are Amazon Music, Neiman Marcus, OrangeTheory FITNESS and CreditGenie.

💡 Amazon Music

The global streaming music platform uses AWS Amplify and AWS AppSync to enable millions of users to sync their music playlist and access it offline from their web and mobile apps at scale. They built a cloud-queuing solution that syncs cloud and local music queues and handles more than 70,000 transactions per second using AWS. The music streaming service wanted to unify its separate, device-specific music-queuing systems under a centralised solution to facilitate a seamless user experience across devices as well as support ongoing service innovation. Building on AWS, it developed a solution that relies on AWS AppSync and AWS Amplify to sync, store, and deliver its curated user experiences. Now, Amazon Music has a scalable solution that supports feature development, both through its technical capabilities and by minimising the maintenance workload for its teams.

💡 Neiman Marcus

A well-known name in luxury retail, Neiman Marcus operates 38 US department stores and a premier digital service for customers around the world. It is the flagship brand of the Neiman Marcus Group, founded in 1907. To speed up its app development time, the Neiman Marcus team chose to build on a server-less architecture using AWS Amplify.

"Using AWS Amplify to build a server-less architecture on AWS, the development team at Neiman Marcus accelerated the launch of our application, reduced development costs, increased agility, and gained the ability to deploy rapid updates. Using AWS Amplify on a server-less architecture cost us 90 percent less than if we had built the app using a more traditional method. This is a huge win for us."

OrangeTheory Fitness

OrangeTheory Fitness is a health and fitness company that combines science, coaching and technology to create maximum results for their customers through heart rate based workouts. OrangeTheory Fitness helps its customers achieve tangible and data driven health outcomes using insights derived from IoT heart rate devices. As of 2020, OrangeTheory Fitness has over one million members.

"We identified Amplify and Amplify Video as the fastest path to integrate video into our application. We estimated that to build and troubleshoot the solution on their own would have taken a full week of dedicated engineering and developer time. In contrast, by using the Amplify CLI and the video plugin code repository, we transcoded the video, generated signed urls, and deployed it to our development environment and then into production – all within an hour. We didn't require any changes to our native iOS/Android code, we simply replaced the backend and Amplify did the work. Our team was able to solve for this complex challenge to keep our customers motivated and working out. We were very pleased to have found AWS Amplify and Amplify Video."

CreditGenie

CreditGenie is a FinTech startup that leverages data and analytics to help borrowers escape the personal debt trap by optimising their debt to manageable levels.

“We found Amplify and played around with it on a weekend, and thought ‘wait a second, you can go from File, New Project to a real production-grade system in minutes.’ From certificate management, hosting, and CloudFront automation — all the pieces were there.” As a fintech startup that stores customer financial data, Amplify also provides critical data protection. Having been at PayPal previously, Kates feels confident using Amplify because Dynamo DB and S3 are backed by AWS Key Management Service (KMS) which is backed by AWS CloudHSM (HSM) that have been validated under FIPS 140-2. Kates states “We’re able to use these services because they’re FIPS compliant so the financial information is encrypted at

rest and in transit with Amplify's SSL management. You would normally have a team of multiple info-sec engineers managing these security concerns alone."

API Gateway

Netflix

Netflix uses Amazon API Gateway to create and manage APIs that provide secure access to their content. The API Gateway helps them manage and scale their APIs as their business grows, and it also allows them to monitor and control access to their content.

Airbnb

Airbnb uses API Gateway to manage and secure access to their micro-services architecture. The API Gateway helps them handle authentication, rate limiting, and other security concerns, while also providing them with detailed analytics and monitoring tools.

Expedia

Expedia uses API Gateway to manage and secure access to their travel booking platform. The API Gateway helps them handle traffic spikes and manage their APIs across multiple regions, while also providing them with real-time analytics and insights into how their APIs are being used.

Lyft

Lyft uses API Gateway to manage and secure access to their ride-sharing platform. The API Gateway helps them handle authentication, rate limiting, and other security concerns, while also providing them with real-time analytics and insights into their APIs.

Twilio

Twilio uses API Gateway to manage and secure access to their communication APIs. The API Gateway helps them handle authentication, rate limiting, and other security concerns, while also providing them with detailed analytics and monitoring tools.

AWS Lambda

This component is used for:

- **Process data at scale**: Execute code at the capacity you need, as you need it. Scale to match your data volume automatically and enable custom event triggers.
- **Run interactive web and mobile backends:** Combine AWS Lambda with other AWS services to create secure, stable, and scalable online experiences.
- **Enable powerful ML insights:** Preprocess data before feeding it to your machine learning (ML) model. With Amazon Elastic File System (EFS) access, AWS Lambda handles infrastructure management and provisioning to simplify scaling.
- **Create event-driven applications:** Build event-driven functions for easy communication between decoupled services. Reduce costs by running applications during times of peak demand without crashing or over-provisioning resources.

Customers that used AWS Lambda in their implementation are Fender, Nielsen, The Coca Cola Company and Stedi.

Fender

Fender went all in on AWS to better engage with customers and gain supply-chain efficiencies. It 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. The company also uses IoT on AWS to monitor factory conditions and Amazon SageMaker to choose the best wood to use in its guitars.

💡 Nielsen

Nielsen's system, called "DataOut," is processing 250 billion events-per-day, which translates to 55 terabytes (TB) of data. The system can automatically scale up and down, thanks to the capabilities of server-less architectures, processing from 1TB to 6TB of data per hour, and costing "only" $1,000 per day. (If this sounds like a lot to you, think again about how much computing power it takes to process this amount of data and how much workforce time the system saves.).

💡 Coca Cola Freestyle

As the COVID-19 pandemic continues to reshape the world and shift consumer behavior, the Coca-Cola Company has developed a new pouring option to meet consumer needs with its latest Coca-Cola Freestyle technology innovation—contactless mobile pouring by smartphone.

Built on Coca-Cola's existing Amazon Web Services (AWS) serverless architecture, the new contactless Coca-Cola Freestyle solution enables consumers to choose and pour drinks from their phones in just a few seconds, without having to create an account or download an app. The mobile experience is currently rolling out to all Coca-Cola Freestyle dispensers across the United States.

After developing the successful prototype in April 2020, Coca-Cola Freestyle went through an AWS Well-Architected Review to determine whether the design would function well and securely at a greater scale. It introduced the prototype to customer testing locations by early June and publicly launched the web app a month later. By

mid- to late July—within 100 days of conception—500 machines were running using the new web application, and 10,000 were running by the end of August. By the end of 2020, all 52,000 Freestyle machines in the United States are expected to be operational, with the remaining 8,000 worldwide to follow.

To deliver a nearly instantaneous user experience, the Freestyle team built a backend on AWS Lambda, which provides server-less compute, and Amazon API Gateway, a fully managed service that makes it easy for developers to create, maintain, and secure APIs at any scale. WebSocket APIs in Amazon API Gateway provide a server-less, bidirectional, real-time connection between the Freestyle machines and users' phones, with a latency of milliseconds rather than seconds. Without that AWS feature, customers would have to wait for inventory updates at the dispenser; the pour would be slow; and lines would form.

IAM Roles

Adobe

Adobe uses IAM Roles to manage access to their Adobe Cloud Platform. IAM Roles allow them to grant granular access to different services and resources, and also ensure that access is granted based on least privilege.

Spotify

Spotify uses IAM Roles to manage access to their cloud infrastructure. IAM Roles help them enforce security policies, prevent unauthorized access, and also allow them to grant granular access to different services and resources.

Atlassian

Atlassian uses IAM Roles to manage access to their cloud infrastructure. IAM Roles allow them to grant granular access to different resources and services, and also help them ensure that access is granted based on least privilege.

💡 Dropbox

Dropbox uses IAM Roles to manage access to their cloud infrastructure. IAM Roles help them enforce security policies and prevent unauthorized access, while also allowing them to grant temporary access to specific resources for specific tasks or periods of time.

💡 Philips

Philips uses IAM Roles to manage access to their cloud infrastructure. IAM Roles allow them to grant granular access to different resources and services, and also help them ensure that access is granted based on least privilege. IAM Roles also enable Philips to audit and monitor access to their cloud resources.

DynamoDB

💡 Samsung

Samsung uses DynamoDB to store and manage user data for their Samsung Pay mobile payment platform. DynamoDB's scalability and flexibility allow Samsung to store and process large volumes of data quickly and easily, while also ensuring high levels of availability and performance.

💡 Toyota

Toyota uses DynamoDB to store and manage vehicle data for their Connected Vehicles platform. DynamoDB's low-latency and high-throughput capabilities allow Toyota to store and process large amounts of data in real-time, enabling them to deliver personalised services and experiences to their customers.

The Pokemon Company

The Pokemon Company uses DynamoDB to store and manage game data for their Pokemon Go mobile game. DynamoDB's ability to handle millions of requests per second and provide consistent low-latency access to data helps ensure that players have a seamless and enjoyable gaming experience.

Dow Jones

Dow Jones uses DynamoDB to store and manage metadata for their Factiva news and information platform. DynamoDB's scalability and ability to handle large amounts of data allows Dow Jones to easily manage and analyse vast amounts of content, providing their customers with accurate and timely news and information.

Capital One

Capital One uses DynamoDB to store and manage customer data for their banking and financial services platforms. DynamoDB's scalability, flexibility, and high availability enable Capital One to quickly process large volumes of data and provide their customers with fast and reliable access to their financial information.

Amazon S3 (initially included, but later removed)

This component is used for:

- **Building a data lake**: Run big data analytics, artificial intelligence (AI), machine learning (ML), and high performance computing (HPC) applications to unlock data insights.
- **Back up and restore critical data**: Meet Recovery Time Objectives (RTO), Recovery Point Objectives (RPO), and compliance requirements with S3's robust replication features.
- **Archive data at the lowest cost**: Move data archives to the Amazon S3 Glacier storage classes to lower costs, eliminate operational complexities, and gain new insights.
- **Run cloud-native applications:** Build fast, powerful mobile and web-based cloud-native apps that scale automatically in a highly available configuration.

Customers that used Amazon S3 in their implementation are NASCAR, Snap Inc., Shutterstock and Adidas.

Nascar

Migrating all their video, audio, and image libraries to the industry-leading scalability, data availability, security, and performance of Amazon S3 allows them to spend more time developing additional workflows that help their business. While they maintain the NASCAR historical media archive, Amazon S3 Glacier Flexible Retrieval, Amazon S3 Glacier Deep Archive, and now Amazon S3 Glacier Instant Retrieval, have enabled them to put the endless lifecycle management of LTO tape formats in the past.

In doing so, Amazon S3 storage provides them great cost savings by using the many helpful S3 storage classes to optimise costs and performance for their video, audio, and image-based workflow. All while keeping the data protected with 11 9s of durability across all storage classes which is a significant improvement over their legacy LTO storage library.

Snap Inc

Snap migrated more than 2 exabytes of data—roughly equivalent to 1.5 trillion media files—seamlessly to Amazon S3 Glacier Instant Retrieval from Amazon S3 Standard-IA. As a result of the migration, the company saved tens of millions of dollars on storage. Snap has configured Amazon S3 in 20 AWS Regions around the world so that customers anywhere can retrieve data in milliseconds. The AWS Global Infrastructure is the most secure, extensive, and reliable Global Cloud Infrastructure for a business's applications. The global reach of AWS lets Snap store media closer to the place where Snapchatters are creating it for optimal performance.

Migrating Snap's content to Amazon S3 has also improved operations and visibility. Using Amazon S3 Storage Lens, a feature that delivers organisation-wide visibility into object storage usage, the company has better insight into what it's storing so that it can make more informed, data-driven decisions.

Snap also migrated to AWS to scale its infrastructure to support its growth: the amount of content that it stores has grown by 5–10 percent each year. Meanwhile, Snap transitioned other parts of its infrastructure from its previous monolithic architecture to one based on micro-services to host many of the services that powered its app. Snap saw a 20–30 percent reduction in download latency in certain Regions for refreshing feeds, downloading media, and doing near-real-time communications.

Shutterstock

One of the challenges of a fast-growing library of content like theirs is ensuring that ample storage is available well ahead of demand. Maintaining their own storage infrastructure for over a decade has given them valuable insight into the logistics of on-premises hardware maintenance, expansion, and refreshes.

Starting in 2019, all their storage requests were being served from Amazon S3. The storage proxy, now in its second generation and built with code portability in mind, continued mapping requests as needed but no longer had to keep track of multiple storage systems. The greatest initially observed gain was that reduced maintenance of S3 allowed us to regain 20% – 50% of their team sprint time for other tasks. They reinvested this time into optimising other areas of our architecture.

As their content footprint increased to roughly 16 PB by late 2021, significant changes were announced for the S3 Intelligent-Tiering storage class. A new Archive Instant Access tier offered additional savings of up to 68% with no impact on performance. This would allow them to have collections of objects, which may be infrequently accessed, to reach the Archive Instant Access tier and move to the Frequent Access tier in the event of the topic warming up.

The savings they realised from using S3 Intelligent-Tiering, up to 60% in some buckets, allowed them to further reinvest in our storage infrastructure and replicate their storage environment to a second AWS Region. The timing aligned well with a recent feature release for Amazon S3 Batch Replication for existing objects that would allow them to turn on S3 Cross-Region Replication (CRR) for both existing and newly uploaded objects. Single Region S3 is already designed for 99.999999999% (11 9s) durability and 99.9% availability when utilising the S3 Intelligent-Tiering class. Furthermore, by employing S3 Intelligent-Tiering in their second Region, they keep the cost to a minimum when the data is not being accessed.

💡 Adidas Runtastic

Runtastic has served fitness enthusiasts since 2009, creating a community of users around its health and wellness apps. The company now has tens of millions of customers across Europe, South America, and the US. But its rapid growth brought challenges for the firm’s aging technology infrastructure. By migrating to AWS, Runtastic has reduced latency, improved the customer experience, and saved more than €300,000 in new infrastructure spending alone.

In 2015, Runtastic was acquired by international fitness brand adidas. The acquisition boosted Runtastic’s profile globally. A growing numbers of user registrations and activations, and an increased volume of activity tracking also stressed infrastructure and slowed down processing. To serve its enlarging user base, Runtastic migrated to Amazon Elastic Compute Cloud (Amazon EC2)—which provides secure and resizable compute capacity for workloads—and expanded its use of Amazon S3 and Amazon CloudFront.

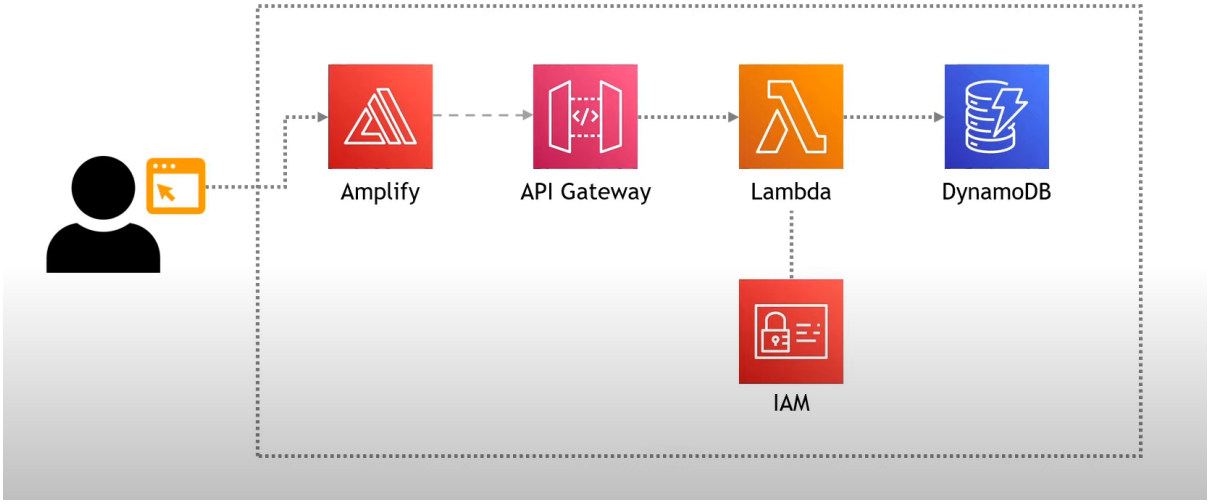
Runtastic soon faced a new challenge when the COVID-19 pandemic hit. It saw use of its apps surge. Whenever a country locked down, demand spiked as users suddenly had more free time to exercise. To maintain performance, Runtastic decide

to spin up more AWS services, as this strategy had previously worked well in responding to spikes in demand.

Runtastic, despite still having on-premises solutions, knows that the cloud is where business is headed and is committed to it.

App Description 📝

A Group-Messaging app that allows users to send messages to everybody online or only to selected users.



WebSockets

The app uses WebSockets for real-time interaction, enabled by API Gateway's WebSockets API.

AWS Amplify

The web-app is hosted on AWS Amplify using a GitHub Repo connection.

API Gateway

All of the requests are made through API Gateway on one of the six (*3 default, 3 custom*) exposed routes.

AWS Lambda

The Lambda is responsible for handling the calls received from API Gateway and actually implementing the backend logic of the app.

IAM Roles

The IAM Roles and Policies are used to manage access and allow different components to properly work together (*e.g. Lambda and API Gateway*)

DynamoDB

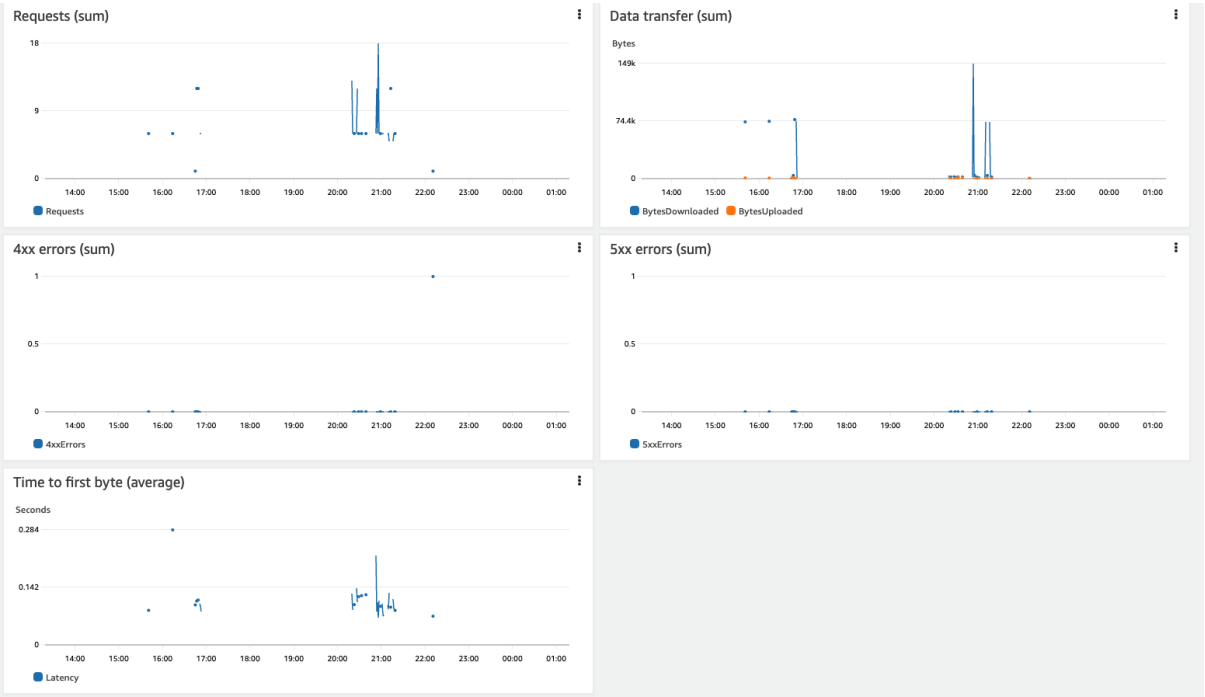
Used for storing a record of every message sent on the platform, including connection activities. The Lambda manages the creation and uploading of objects to the database.

CloudWatch

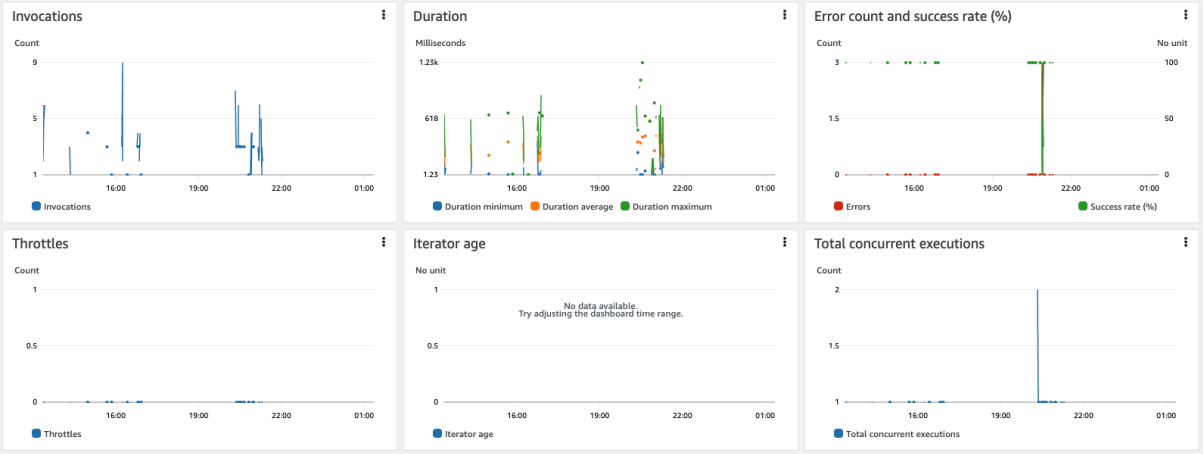
Used for monitoring alarms and general performance, more details below.

Performance Analysis (using CloudWatch) 📊

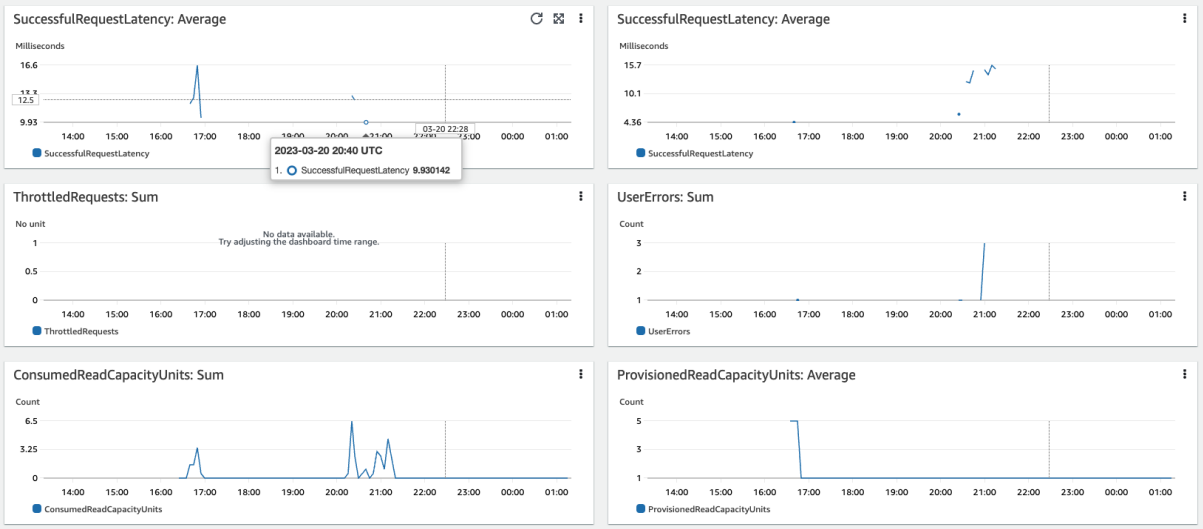
AWS Amplify



AWS Lambda



DynamoDB



Conclusion 🎬

The combination of WebSockets, AWS Amplify, API Gateway, AWS Lambda, IAM Roles and DynamoDB provides a powerful and flexible stack for building a real-time chat application.

The use of WebSockets enables low latency, bidirectional communication between the client and server, making it ideal for real-time applications.

API Gateway provides a scalable and secure entry point for incoming traffic.

AWS Lambda allows for server-less compute power to handle incoming requests and processing data.

AWS Amplify allows for easy building and deploying of the front-end application with pre-built UI components, integration with AWS services, and a robust development and deployment workflow.

IAM Roles ensure secure access control and permission management.

DynamoDB provides a highly scalable and reliable NoSQL database for storing and retrieving data in real-time.

In terms of performance, the combination of these services allows for high throughput, low latency, and efficient processing of data. With AWS's global infrastructure, the application can scale and handle high volumes of traffic from any location around the world.

The use of AWS services also provides high reliability, as these services are designed to be highly available, fault-tolerant, and provide automatic failover and backup capabilities. Additionally, the transparent monitoring and logging provided by AWS allows for easy debugging and troubleshooting, as well as performance optimisation.

Overall, the use of WebSockets, AWS Amplify, API Gateway, AWS Lambda, IAM Roles and DynamoDB provides a robust and efficient stack for building real-time chat applications that are powerful, flexible, and scalable, with high performance, low latency, and high reliability.