

Major Capstone

This major Capstone is a 3 part project-

In the 1st part we will learn how to use AWS CDK and various AWS services to replicate an On-Premise Data Center infrastructure by ingesting real-time IoT-based data.

In the 1st part of this project, we will cover using an AWS IoT device Simulator to replicate an On-Premise Data Center infrastructure by ingesting real-time IoT-based data. The services we'll be using are AWS CDK (CloudFormation), AWS IoT core, Kinesis Firehose, Lambda, AWS S3, EC2, MariaDB, and AWS Secrets manager. The second project in this series will be migrating and analysing the data to "AWS Cloud Premise" using DMS, RDS, Glue, AWS Time stream, and QuickSight. There is a third project also which will act as a mini-course to dive deep into the concepts and workings of Infrastructure-as-a-Code (IaC) using AWS CDK (Cloud Development Kit).

Data Description

Using the Device Simulator, we will simulate and deal with the geoLocation data of multiple devices parallelly within the vicinity of a popular entertainment complex in London called The O2 Arena.

Tech Stack:

Framework: AWS CDK

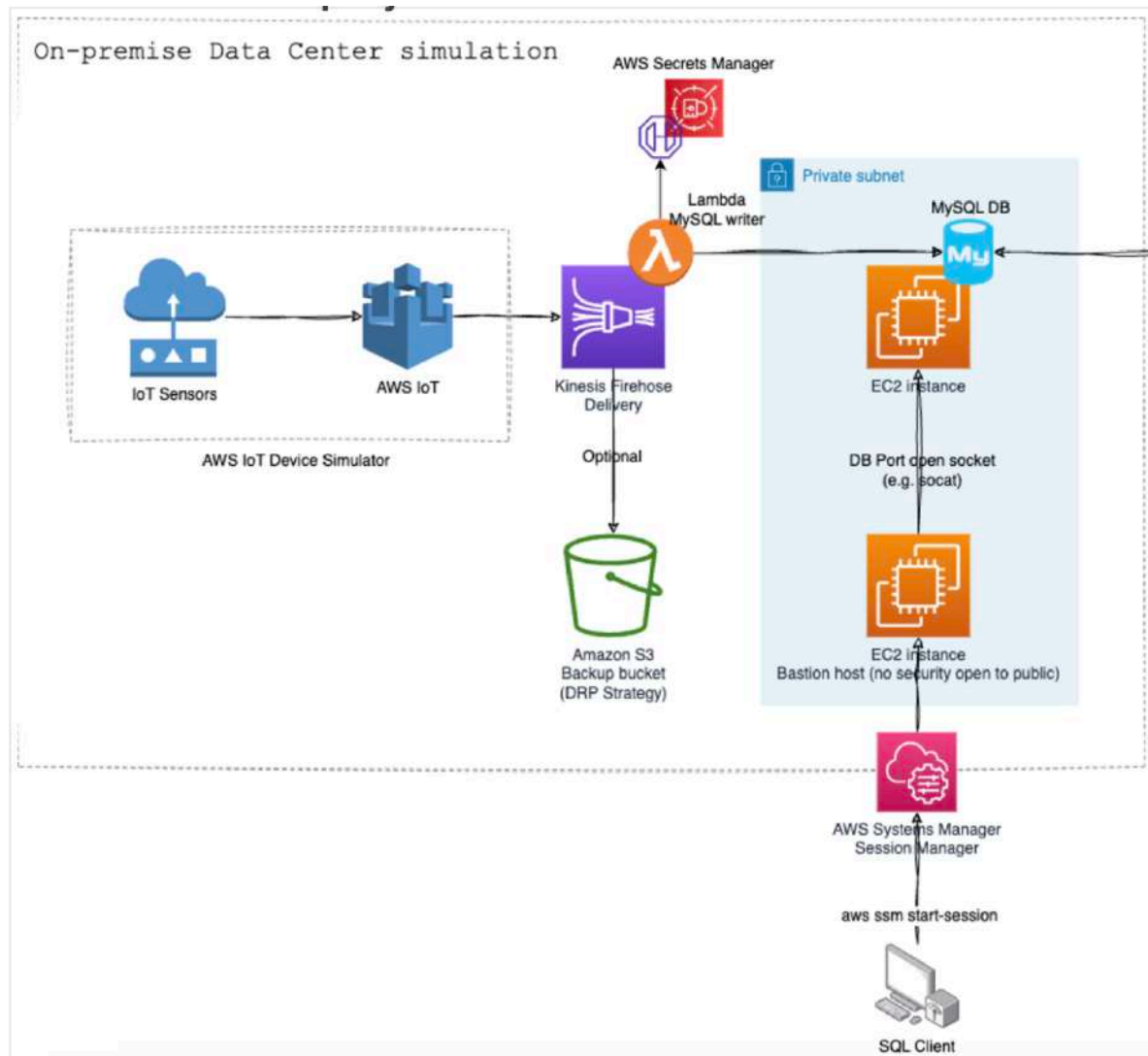
Language: Python

Services: AWS IoT core, Kinesis Firehose, AWS Lambda, MariaDB, AWS S3, AWS Secrets Manager

Key Takeaways:

- Creating AWS Account and following Best Practices
- Installation of AWS CLI for Programmatic access
- Exploring AWS IoT Device Simulator
- Understanding and Creating Amazon Kinesis Firehose stream
- Creating an EC2 machine using the Console
- Simulating an On-Premise MySQL Database on EC2
- Introduction to AWS CDK using Python
- Installation of AWS CDK
- Using MQTT protocol with SQL in AWS IoT Core
- Creation of AWS S3 and Lambda stacks using CDK
- Setting up AWS Secrets and VPC Endpoint
- Deploy the project using CDK
- Deploy the pipeline using the Console

Architecture for the 1st part



- Simulate the IoT data using AWS IoT device simulator and ingest that data to a MySQL DB (we will use Maria DB, but we can use MySQL also) instance running on an EC2 instance in order to replicate an on-premise infrastructure. The ingestion will happen via Kinesis Firehose. We will use a Lambda function where firehose will act as a trigger for it and the Lambda function will eventually write data to the DB.
- We will access the Maria DB instance via bastion host (this is just to follow best practice. Feel free to access it otherwise). Optional - We will also dump the IoT data from Firehose to S3 bucket, which will act as our DR setup.