

# Migration of MySQL Databases to Cloud AWS using AWS DMS

## Business Overview

Today's data has become a critical asset for businesses to make informed decisions and gain a competitive edge. However, with the exponential growth of data, managing and storing it on-premise can be expensive and time-consuming. Cloud computing offers an efficient and scalable solution for this challenge. In this project, we will discuss the necessity and importance of real-time data migration from an on-premise database to the cloud, specifically AWS, and how AWS DMS and SCT can be used for migrating data from a data source (MySQL) to a target (Aurora Postgres).

### Real-time Data Migration from Data Source to Target:

- Real-time data migration from a data source to a target is critical for businesses that require real-time insights to remain competitive
- AWS DMS and SCT can be used to migrate data from an on-premise database to AWS in real-time
- DMS listens for new data at the source (MySQL) and migrates it in real time to the target (Aurora Postgres)
- This ensures that businesses can take advantage of their data as soon as it is generated

### Necessity and Importance of Real-time Data Migration to AWS:

- Real-time data migration to AWS is critical for businesses for several reasons.
- Cloud-based databases provide scalability, enabling businesses to handle large volumes of data without expensive hardware.
- Cloud-based databases can offer cost savings compared to on-premise databases, eliminating the need for additional IT staff, hardware, and software upgrades.
- AWS provides robust security measures, ensuring data security for businesses
- Cloud-based databases offer backup and disaster recovery options that can help ensure business continuity.

### Using Migrated Data for Analysis, Reporting, and Visualization:

- Once the data has been migrated to AWS, it can be used for analysis, reporting, and visualization.
- AWS offers a range of analytics tools, such as AWS Glue, Amazon Athena, and Amazon Redshift, which can be used to perform data analysis and generate insights
- AWS also provides reporting tools, such as Amazon QuickSight, which can be used to create and share reports and dashboards
- Additionally, AWS offers visualization tools, such as Amazon QuickSight, which can be used to create interactive dashboards and visualizations.

Real-time data migration from an on-premise database to AWS using AWS DMS and SCT is critical for businesses that generate large volumes of data. Real-time migration ensures that businesses can take advantage of their data as soon as it is generated. Furthermore, cloud-based databases offer scalability, cost savings, improved data security, and disaster recovery options. Finally, migrated data on the cloud can be used for analysis, reporting, and visualization using AWS analytics tools and visualization tools. By migrating data to the cloud, businesses can gain a competitive edge and make informed decisions based on data insights.

## Aim

This project is the second part of the IoT Data Migration series using AWS CDK. The [first project](#) uses an AWS IoT device Simulator to replicate an On-Premise Data Center infrastructure by ingesting real-time IoT-based data. The services used in the first project were AWS CDK (CloudFormation), AWS IoT core, Kinesis Firehose, Lambda, AWS S3, EC2, MariaDB, and AWS Secrets Manager. **In this project, we will migrate and analyze the data, schemas, tables, and functions from the MySQL server to “AWS Cloud Premise” using DMS, RDS, AWS Glue, AWS Timestream, AWS S3, and QuickSight.** Finally, the [third project](#) 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:

Language: Python, SQL

Services: AWS Glue, AWS SCT, AWS DMS, AWS Timestream, AWS S3, AWS Secrets Manager, AWS RDS, AWS Athena, Aurora Postgres, QuickSight, AWS SSM, Apache Spark

## AWS DMS

AWS DMS (Database Migration Service) and AWS DMS SCT (Database Migration Service Schema Conversion Tool) are two services offered by Amazon Web Services (AWS) to simplify the process of migrating databases to the cloud.

AWS DMS is a fully managed service that enables customers to migrate their existing databases to AWS with minimal downtime. It supports migration from various sources, including on-premises databases, Amazon RDS instances, and other cloud providers. AWS DMS can

also replicate ongoing changes to the source database to the target database, allowing for real-time replication and minimal data loss.

AWS DMS SCT is a standalone tool designed to help customers convert their database schema from one format to another. It provides a graphical user interface for converting schemas and automated conversion scripts for commonly used databases such as Oracle, Microsoft SQL Server, and MySQL. AWS DMS SCT also supports the conversion of database code, such as stored procedures, functions, and triggers, to ensure compatibility with the target database.

AWS DMS and AWS DMS SCT provide a comprehensive solution for migrating databases to AWS. With AWS DMS, customers can easily migrate their existing databases to AWS with minimal downtime, while AWS DMS SCT simplifies the process of converting database schemas and code to ensure compatibility with the target database. These tools can help organizations reduce the time and cost associated with database migration while ensuring a smooth transition to the cloud.

## AWS Timestream

AWS Timestream is a fully managed time series database service offered by Amazon Web Services (AWS). It is designed to store and analyze time series data at scale, such as metrics, log data, and IoT sensor data. With Timestream, customers can easily store and query large volumes of time series data in real-time, with the ability to scale up or down as needed. Timestream also integrates with other AWS services, such as Amazon CloudWatch and AWS IoT Analytics, to provide a complete time series data management and analysis solution.

## AWS RDS

AWS RDS (Relational Database Service) is a fully managed database service offered by Amazon Web Services (AWS). It allows customers to easily set up, operate, and scale relational databases in the cloud. AWS RDS supports a variety of database engines, including Amazon Aurora, PostgreSQL, MySQL, MariaDB, Oracle, and SQL Server.

One of the most popular database engines supported by AWS RDS is Aurora PostgreSQL, a high-performance and fully managed PostgreSQL-compatible database engine. Aurora PostgreSQL is designed to deliver high availability, scalability, and security for mission-critical applications. It provides up to five times the performance of standard PostgreSQL, with automatic scaling capabilities and built-in replication across multiple Availability Zones.

AWS RDS and Aurora PostgreSQL provide a comprehensive solution for running and scaling PostgreSQL databases in the cloud. With AWS RDS, customers can easily set up and operate PostgreSQL databases in the cloud, while Aurora PostgreSQL provides a high-performance and fully managed PostgreSQL-compatible database engine. This combination enables organizations to reduce the time and cost of managing databases while ensuring high availability and scalability for mission-critical applications.

## AWS SSM

AWS SSM (Systems Manager) is a service offered by Amazon Web Services (AWS) that enables customers to manage their virtual machines and instances in the cloud. SSM provides a unified user interface for viewing operational data, automating tasks, and managing instances, while also providing access to the underlying data sources for further analysis.

A bastion host, or a jump server, is a special-purpose server designed to be the primary gateway for accessing virtual machines and instances in the cloud. It provides an extra layer of security by controlling access to the instances while also providing a central point for managing and monitoring those instances.

AWS SSM and a bastion host provide a secure and efficient way to manage instances in the cloud. With AWS SSM, customers can easily manage their instances through a unified user interface, while a bastion host provides an extra layer of security and centralizes access to those instances.

## Key Takeaways

- Understanding the Project Overview and Architecture
- Introduction to various AWS Migration services
- Understanding AWS DMS (Database Migration Service)
- Implementation of Bastion Host for Data Security
- Using AWS SCT (Schema Conversion Tool)
- Create Aurora Postgres instance using RDS
- Deploy DMS SCT between MySQL and Postgres
- Deploy MySQL functions for Migration
- Migrate DB Elements from MySQL to Postgres
- Create a DMS Replication Instance
- Analyse DMS Data Migration from MySQL to Postgres
- Load data to AWS S3 using DMS
- Crawl S3 Data using Glue to create Data Catalog
- Query Data using Athena and Glue Data Catalog
- Prepare Data for Timestream using AWS Glue and Spark
- Load Data into Amazon Timestream DB
- Visualize Geospatial Data using QuickSight

# Architecture Diagram

