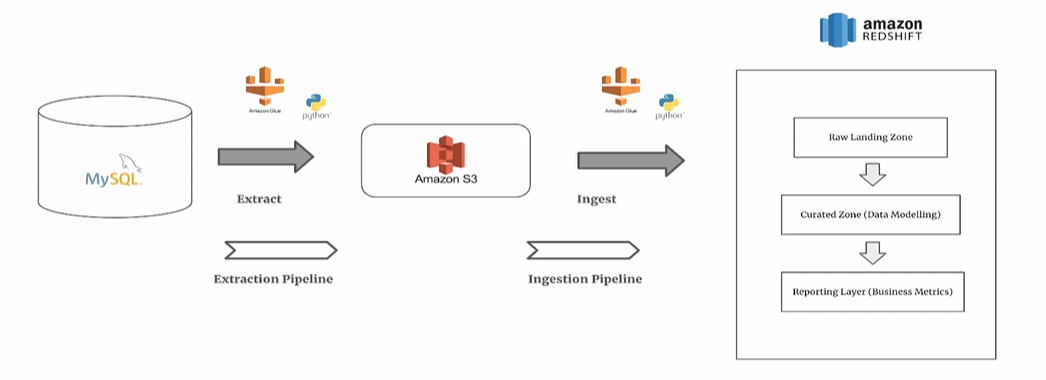
# ETL Data Processing of Rental Apartments Using Glue, Redshift, and Step Functions

## Introduction

This report outlines the batch data processing of rental apartments using AWS Glue, Redshift, and Step Functions. It details the implementation of ETL (Extract, Transform, Load) pipelines for efficient data management and analytics.

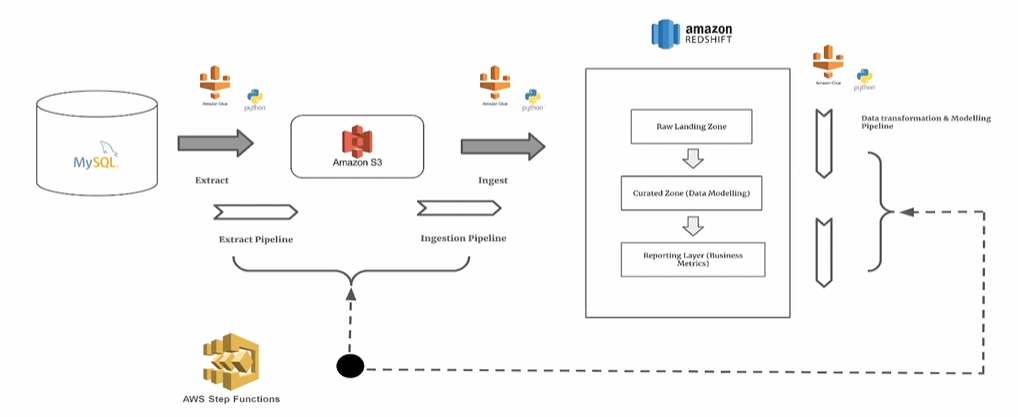
## Architecture Diagram

Below is the architecture diagram representing the ETL process:



## Workflow Overview

The data pipeline follows these major steps:  
1. Data extraction from AWS Aurora MySQL.  
2. Storing extracted data in Amazon S3.  
3. Ingesting raw data into Amazon Redshift.  
4. Transforming data using AWS Glue jobs.  
5. Orchestrating the workflow using AWS Step Functions.  
6. Running analytics using Amazon Redshift.



## AWS Components Used

- \*\*Amazon S3:\*\* Storage layer for raw and processed data.  
- \*\*AWS Glue:\*\* ETL processing using Python Shell jobs.  
- \*\*Amazon Redshift:\*\* Data warehouse for analytics and reporting.  
- \*\*AWS Step Functions:\*\* Orchestration of ETL workflows.  
- \*\*AWS Aurora (MySQL):\*\* Source relational database for transactional data.

## Step Functions for Workflow Automation

AWS Step Functions automate the ETL process by executing jobs in sequence and handling dependencies. It ensures that data flows smoothly from extraction to transformation and ingestion.

## Conclusion

This ETL pipeline provides a robust solution for processing rental apartment data, leveraging AWS services for efficient data transformation and reporting.