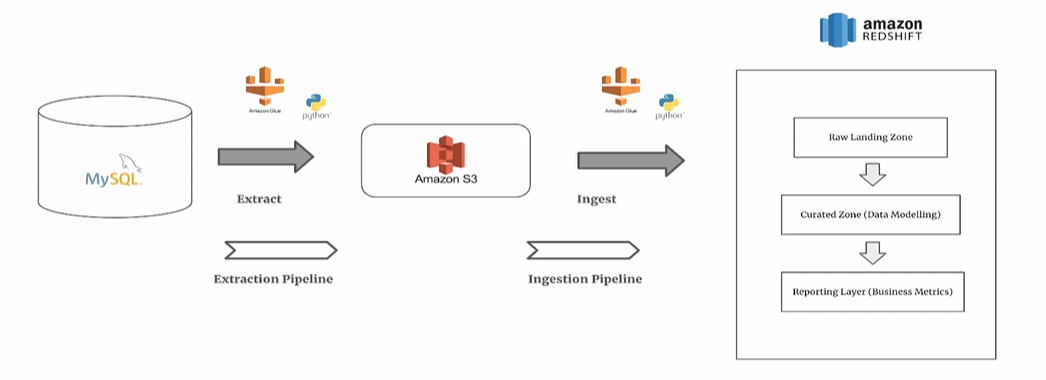
# ETL Pipeline for Rental Apartments - AWS Implementation

## Overview

This document explains how to implement an ETL pipeline for processing rental apartment data using AWS Glue, Redshift, and Step Functions.

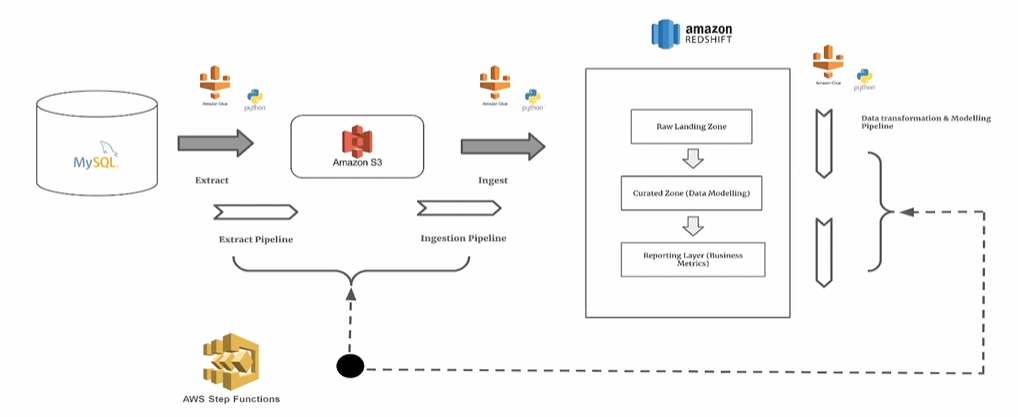
## Architecture



The ETL pipeline consists of data extraction, transformation, ingestion, and orchestration.

## ETL Process

1. \*\*Extract:\*\* Data is pulled from AWS Aurora (MySQL) and stored in Amazon S3.  
2. \*\*Transform:\*\* AWS Glue jobs process and clean the data.  
3. \*\*Load:\*\* The transformed data is ingested into Amazon Redshift.  
4. \*\*Orchestration:\*\* AWS Step Functions manage workflow execution.



## Key AWS Services Used

- \*\*Amazon S3\*\*: Cloud storage for raw and processed data.  
- \*\*AWS Glue\*\*: ETL service using Python-based processing.  
- \*\*Amazon Redshift\*\*: Data warehouse for analytics.  
- \*\*AWS Step Functions\*\*: Workflow automation and execution.

## Benefits of the ETL Pipeline

✅ Scalable data processing  
✅ Cost-efficient cloud-native ETL  
✅ Automated workflow with Step Functions  
✅ Advanced analytics with Redshift