# **ZILLOW REAL ESTATE DATA PIPELINE WITH AIRFLOW, AWS LAMBDA, REDSHIFT, AND QUICKSIGHT**

In this data engineering project, we demonstrate how to build and automate a Python-based ETL pipeline that extracts real estate property data from the **Zillow Rapid API**, processes it through AWS services, and visualizes it with **Amazon QuickSight**.

|  |  |
| --- | --- |
| Tool/Service | Purpose |
| Python | API call & scripting |
| Apache Airflow (on EC2) | DAG orchestration |
| BashOperator | Move files from EC2 to S3 |
| AWS Lambda | File copy & transformation automation |
| Amazon S3 | Staging, transforming, storing files |
| S3KeySensorOperator | Detect presence of file before Redshift load |
| Amazon Redshift | Data warehouse |
| Amazon QuickSight | BI tool for visualization |

1. **Python ETL Pipeline Design**

### ****Data Extraction & Staging:****

* **Zillow Rapid API** serves as the primary data source.
* Using **Python**, the API is queried to extract structured real estate property information such as number of bedrooms, bathrooms, price, rent estimates, property type, and location.
* The raw data is then loaded into an **Amazon S3 bucket** referred to as the **Landing Zone**.