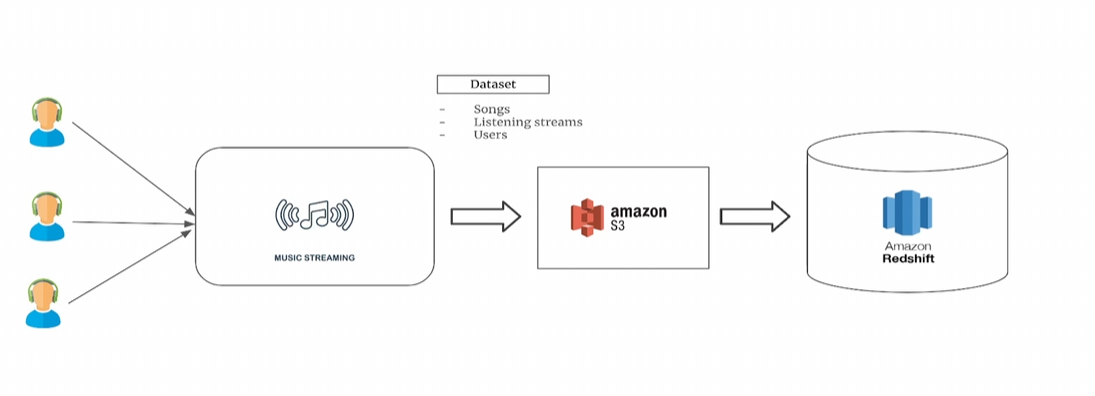
Batch Data Processing Using Airflow & Redshift

# Project Overview

This project processes music streaming data using AWS services. Data is ingested from Amazon S3, transformed using Python and Apache Airflow, and stored in Amazon Redshift for analytics.

# Architecture

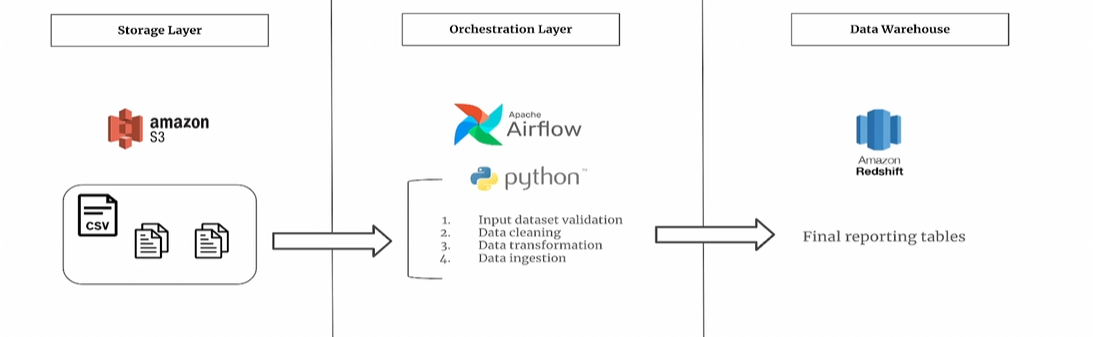
The solution follows this architecture:



# Data Pipeline Steps

1. \*\*Ingest Data\*\*: Fetch CSV files from S3.  
2. \*\*Validate Data\*\*: Ensure required columns exist.  
3. \*\*Transform Data\*\*: Compute music streaming KPIs.  
4. \*\*Load to Redshift\*\*: Store transformed data.  
5. \*\*Archive Processed Files\*\*: Move files post-processing.

# Workflow



# Tools & Technologies

- \*\*Amazon S3\*\*: Data storage.  
- \*\*Apache Airflow\*\*: Orchestration.  
- \*\*Python & Pandas\*\*: Data transformation.  
- \*\*Amazon Redshift\*\*: Data warehouse.

# Conclusion

The implementation of Airflow and Redshift streamlines data processing for music analytics. Future improvements could include real-time processing and predictive analytics.