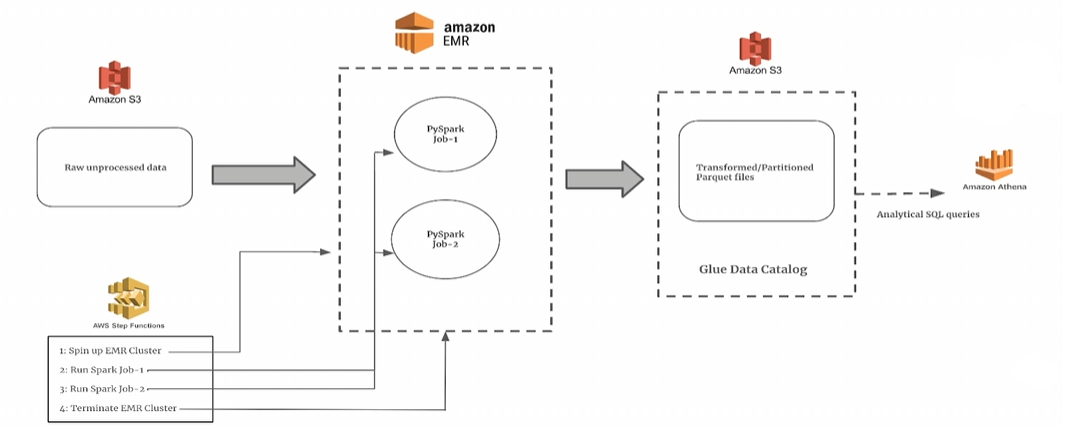
Big Data Processing of Rental Vehicles - Website Friendly Report

# Introduction

This document outlines the implementation of an AWS-based data lake for rental vehicle analytics. The architecture enables large-scale data processing using AWS services like EMR, S3, and Athena.

# Technical Overview

The system leverages:  
- \*\*Amazon S3\*\* for storage  
- \*\*AWS EMR\*\* for distributed data processing  
- \*\*AWS Glue\*\* for schema inference  
- \*\*Amazon Athena\*\* for querying the data lake  
- \*\*AWS Step Functions\*\* for workflow automation



# Workflow Execution

1. \*\*Data is uploaded to Amazon S3\*\*  
2. \*\*EMR executes PySpark jobs to process the data\*\*  
3. \*\*Transformed data is stored back in S3\*\*  
4. \*\*Glue crawlers create metadata for Athena queries\*\*  
5. \*\*Athena enables SQL queries on the data lake\*\*

# Summary

This project highlights how AWS services simplify big data processing by reducing infrastructure management and optimizing query performance.