**Cloud Computing Portfolio-Omar**

**Project 1: Property Tax Report Analysis**

**Project Description:**

An analysis of a property tax report dataset is done using a bunch of AWS services in this project. This project’s primary objective is to make data handling both efficient and accurate for business decision making purpose. Data ingestion, profiling, cleaning, cataloging, summarization, security, governance, monitoring and cost optimization are what the focus is on. This project hopes to involve a structured approach to check that the data it uses is accurate, available and further protects it from unauthorized access.

**Steps Completed:**

**Data Ingestion:**

Various cloud storage systems, namely AWS S3, Google Cloud Storage and Azure Blob Storage were all used for uploading the property tax report.csv file. It achieves data redundancy and further availability in multiple platforms.

Since it was determined that the data ingestion was probably going to be a long and multi step process, I looked at AWS Glue, Apache NiFi, and Google Dataflow as options for Iooking at doing the data ingestion in. AWS Glue was chosen because it fits well with Amazon Redshift for further data analysis.

Ingestion process enables the consolidation of data on a single instance dose to improve efficiency of analytical and reporting processes.

**Data Profiling:**

Data profiling was performed using AWS Glue DataBrew, Trifacta as well as Pandas Profiling in Python. Then, this step allowed us to understand what the structure and level of quality is of the dataset.

Metrics that were key assessed included data types, number of null values, number of unique entries, and distribution patterns.

This step guaranteed the detection of possible problems such as the presence of missing values, different data types across data sets, and the presence of wrongly entered values in the data before it could reflect in our visualizations.

**Data Cleaning:**

I applied methods like fill null values, remove duplicates, data formats standardization.

In order to address consistency, while removing anomalies which could affect analytical results, data transformation methods were applied.

**Data Cataloging:**

A structured approach to cataloging was carried out using AWS Glue Data Catalog. To ensure consistency and accuracy, there were documented to the metadata entries.

It contained things like the column name, its type, description, and transformation steps.

It solves the problem of fast data discovery and retrieval by authorized users.

**Data Summarization:**

Mean, median, mode and standard deviation were generated on a statistical summary of the dataset.

To increase the visualization of data distribution and pattern, data was made into histograms and box plots.

**Data Analysis using AWS Athena:**

Description analysis of the dataset stored in S3 storage bucket was performed with AWS Athena.

Effectively met key business questions, for example, what are typical, maximum and minimum property tax values

For further review and validation, the results of the analysis were exported.

**Data Security:**

The data was encrypted by using AWS Key Management Service (KMS) to protect data from any unauthorized access.

To allow for data recovery, or at least attempt to, S3 versions were placed in trust on an S3 bucket.

We created S3 bucket replication for disaster recover and backup purpose.

**Data Governance:**

The solution is implemented using AWS Glue Data Catalog and IAM policies for enforcing access control.

An approach to create a structured cataloging system was adopted to increase data integrity;

Data Monitoring and Quality Control:

Continuous monitoring and validation of the data transformation process were enabled with configuring AWS S3 data pipelines.

Data accuracy was checked before being stored with automated quality checks.

**Cost Optimization:**

Automated selection of Frequent, Infrequent, and Archive access for cost efficient storage tiers was applied.

One cost marginalizing function that has been applied is S3 lifecycle policies and Reserved Instances.

**Property Tax Data Security and Management Project 2.**

**Project Description:**

A major contribution of this project is to make property tax data more secure, governed, and monitored. The objective is to keep data availability and confidentiality and yet ensure that data is viable, but recognizes optimum performance and minimal costs. In the course of the project, different AWS services are used to build a complete data management framework which fulfills the strict standards for both reliability and efficiency.

**Steps Completed:**

**Data Security Implementation:**

Data was encrypted to prevent unauthorized access by using AWS Key Management Service.

To safeguard against accidental modification or deletion, we had also activated versioning of S3 buckets so that they are recoverable even if they are accidentally rolled over.

The S3 buckets were replicated to improve data availability as well as disaster recovery.

**Data Governance:**

IAM policies were used to define the user roles and provide them access as allowed.

An efficient metadata management and data discovery was ensured through a structured cataloging system.

**Data Monitoring:**

The main framework to monitor data quality and potential errors is used, which is AWS S3.

Condisonal route was deigned to separate the valid from the non-valid data so that it can be processed in a streamlined fashion.

**Cost Optimization:**

Measures to save costs such as lifecycle policies, reserved instances, and auto tier selection were employed at low cost.

We resorted to continuous monitoring of utilization which ensured financial sustainability.

**Deliverables:**

Descriptive Analysis Results

Data Governance and Monitoring Frameworks

Cost Optimization Strategies

Detailed Reports and Visualizations

**Conclusion:**

This portfolio leverages a variety of AWS services to structure the management of property tax data, and addresses the EXAM’s Topic 12. Coupled processes of these allow for reliable, accessible, and secure data management. Data cataloging, monitoring systems, and cost optimization measures employed here constitute a wide ranging approach towards bettering data quality, security, and financial sustainability through encryption. With its projects, this effectively serves as a good showcase of your hands on cloud based data management solution handling skills which is a total makeable addition as part of your portfolio.