**Data Analytic Platform for City of Vancouver**

**Data Protection**

During our project City Of Vancouver, Property tax Data Protection phase, I put in place numerous crucial security protocols against property tax data meant for the City of Vancouver being stolen. First of all, I developed a KMS key against data encryption meant for the main S3 bucket. This adds to security by encrypting all data automatically at rest. I then made a backup S3 bucket with the duplicated copy of the data for its availability and robustness in should any failure happen. I then set the replication rules, which using the same KMS key securely sent data from the primary bucket to the backup bucket with encryption. This guarantees data security and continuity ensured, encrypted, protected, and securely backed-upon.

A screenshot of a computer

Description automatically generated

*Fig1: Role- Lab Role*

A screenshot of a computer

Description automatically generated

*Fig2: Back up for Property Tax Data*

A screenshot of a computer

Description automatically generated

*Fig 3: Data Encryption with KMS*

A screenshot of a computer

Description automatically generated

*Fig4: Replicated Rules for Property Tax Data*

**Data Governance**

AWS Glue enabled me to create the Data Governance ETL pipeline for my property tax project. Maintaining norms of governance will help to guarantee effective data processing along this pipeline. To ensure conformity to Canadian privacy standards, I began searching the raw property tax collection for sensitive information. In this regard, I have conducted data quality checks; most of the crucial fields, including tax rates and zoning zones, must be exact and whole. I then set up these Monday data quality tests to guarantee constant automated monitoring.

A screenshot of a computer

Description automatically generated

*Fig 5:QRPR ETL*

A screenshot of a computer

Description automatically generated

*Fig6: Data Quality Schedule*

I developed a workflow in AWS Glue to enable this governance process to be taken forward and in far more seamless manner. The process will guarantee that other operations connected to data governance and quality checks of the data follow the correct sequence, therefore guiding the activities of data administration. This will link the ETL pipeline with planned activities so that ongoing data compliance and governance may occur without human involvement. Efficiency will be raised by this automated process, which also helps to maintain data controlled and up to standards.

A screenshot of a computer

Description automatically generated

*Fig7: Workflow for ETL job*

**Data Monitoring**

In data infrastructure, security, efficiency, and regulatory compliances constitute the ongoing control in the phase of data monitoring and management. This requires real-time performance and activity monitoring of the system including Amazon CloudWatch and AWS CloudTrail. A good summary of certain key indicators, including resource consumption and expenses, CloudWatch dashboards offer. Alarms are set to provide alerts in response to threshold breaks, so enabling ideal prompt reaction to avoid problems. For instance, it would be a CloudWatch Dashboard made to view estimations for costs, several objects in S3, and an alert to alert overcharging.

A screenshot of a computer

Description automatically generated

*Fig8: Cloud Watch Data Monitoring*

A screenshot of a computer

Description automatically generated

*Fig 9: Cloud Watch Dashboard for Data Monitoring and Controlling*

*A screenshot of a computer

Description automatically generated*

*Fig10: Created Alarm In CloudWatch*

AWS CloudTrail tracks user activities and API calls inside AWS offerings. All the operations carried out inside the system can thus be recorded and kept in S3 for audit and additional study. Events tracked by CloudTrail create openness and prove rather useful for tracking access, keeping compliance, and looking at security events. CloudWatch and CloudTrail taken together guarantee complete system performance and probable security issues for data governance by means of their continuous monitoring of the data pipeline.

A screenshot of a computer

Description automatically generated

*Fig 11: Cloud Trial*

A screenshot of a computer

Description automatically generated

*Fig 12: AWS User logs In S3 Bucket*