**AWS Data Analytic Platform for The City of Vancouver**

**Implementation and Analysis of 3-1-1 Inquiry Volumes Using AWS**

**Introduction**

The City of Vancouver has migrated its data processing to Amazon Web Services (AWS) to optimize the speed and quality of data analysis for better decision-making. The 3-1-1 inquiry system is a crucial public service that allows citizens to report issues, ask questions, and request services from various city departments. This project aims to establish a Data Analytics Platform (DAP) on AWS for sourcing, storing, processing, and analyzing 3-1-1 inquiry volumes to improve resource allocation and service delivery.

**Part 1: DAP Design and Implementation**

**1. Data Analytical Question Formulation**

Key questions driving the analysis include:

* To which departments do citizens most frequently report inquiries?
* What are the monthly trends in inquiry volumes?
* How does the distribution of inquiry volumes vary by year?

**2. Data Discovery**

Data was sourced from the City of Vancouver's 3-1-1 inquiry volume dataset, available at [OpenData Vancouver](<https://opendata.vancouver.ca/explore/dataset/3-1-1-inquiry-volume/information/?disjunctive.department&disjunctive.type&disjunctive.channel>).

**3. Storage Design**

AWS S3 was used for data storage with three distinct buckets:

* Raw Bucket: Stores unprocessed data.
* Landing Bucket: Contains data post-initial processing and staging.
* Curated Bucket: Holds processed data ready for analysis.

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**4. Data Preparation**

AWS DataBrew was used to clean and transform the data, applying standardization and filtering irrelevant data points.

**5. Data Injection**

The cleaned data was ingested into the S3 Raw bucket for storage and further processing.

**A screenshot of a computer

Description automatically generated**

**6. Data Pipeline Design**

AWS Glue was employed to create a data pipeline for ETL (Extract, Transform, Load) processes, ensuring the data flows efficiently to the desired destinations.

**A screenshot of a computer

Description automatically generated**

**7. Data Cleaning**

Data cleaning involved removing null values and resolving formatting issues using AWS DataBrew.

**8. Data Structuring**

The data was restructured by department and monthly trends to fit the analytical questions posed at the beginning of the project.

A screenshot of a computer

Description automatically generated

**9. Data Pipeline Implementation**

AWS Glue Jobs were set up to ensure the correct transformation and loading of data into the designated S3 buckets.

**A screenshot of a web page

Description automatically generated**

A screenshot of a computer

Description automatically generated

**10. Data Analysis**

AWS Athena was utilized for SQL queries to analyze the data. For example, the following query was executed to retrieve relevant information:

**```sql**

**SELECT FROM "group2\_part1\_govfinance\_3\_1\_1\_inquiry\_volume\_database\_hemanth"."group2\_part1\_govfinance\_3\_1\_1\_inquiry\_volume\_table\_hemanth";**

**```**

**A screenshot of a computer

Description automatically generated**

**11. Data Visualization**

The comparison of inquiry volumes between 2023 and 2024 was visualized to analyze departmental resource allocation and performance. This was critical for stakeholders to monitor and report on department performance.

**12. Data Publishing**

An AWS EC2 instance was deployed to publish the analysis results, enabling remote access for stakeholders.

**A computer screen with a blue background

Description automatically generated**

**Part 2: Data Protection, Governance, and Monitoring**

**1. Data Protection**

Ensuring the security of the data was a priority. Several AWS services were used, including:

- S3 Bucket Policies: Enforced strict access control to secure the data.

- S3 Encryption: Enabled AWS KMS-based encryption for all data stored in S3 buckets.

- Data Replication: Set up data replication across regions for redundancy and disaster recovery.

2. Data Governance

Data governance was managed using AWS CloudTrail, which logged all API requests and changes made in the AWS environment. CloudTrail's multi-region support ensured comprehensive monitoring.

3. Data Monitoring

AWS CloudWatch was employed to monitor the health and performance of the system. Custom dashboards tracked critical metrics such as:

- Estimated Charges: Monitoring cost variations associated with S3 storage.

- Bucket Size (in Bytes): Monitoring the growth of data storage over time.

---

Screenshots and Visuals

- S3 Bucket Management and Replication Rules: Demonstrates the lifecycle and replication configurations for data redundancy (refer to the attached screenshots).

- CloudTrail Logging and Governance: Shows the CloudTrail setup for monitoring API requests.

- CloudWatch Dashboard: Highlights metrics such as estimated charges and bucket size for system monitoring.

---

Summary of Insights

- The analysis revealed which departments received the most inquiries, allowing for better resource allocation.

- Trends over time showed monthly and yearly fluctuations, helping to predict inquiry patterns.

- Data security and governance were successfully implemented, ensuring the privacy and integrity of the data.

- The monitoring setup provided real-time insights into system performance and usage, enabling proactive cost and performance management.

---

Conclusion

This project effectively demonstrates how AWS services can be used to build a scalable, secure, and efficient Data Analytic Platform for a municipal 3-1-1 inquiry system. The design, implementation, and governance steps ensured that the City of Vancouver could make data-driven decisions while maintaining the security and integrity of its data. The visualizations and reports generated provided stakeholders with valuable insights into service demand and resource needs, ultimately leading to better public service delivery.

---

The complete document includes both parts, providing a thorough overview of the entire project from design and implementation to governance and monitoring. The screenshots provided demonstrate various configurations made during the project for better visualization【8†source】【7†source】【14†source】.