**Portfolio: HR - 8002p Conflict Resolution Procedure Analysis**

**Introduction**

This project focuses on analyzing conflict resolution patterns within an HR department using AWS services. The goal is to conduct an exploratory and descriptive analysis to identify trends, resolution effectiveness, and factors influencing conflict outcomes. The analysis leverages AWS services used throughout the term to process and visualize HR data efficiently.

**Project Title:**

Understanding Conflict Resolution in HR: An Exploratory Data Analysis

**Objective:**

The primary objective of this project is to utilize Amazon Web Services (AWS) to perform Exploratory Data Analysis (EDA) on the HR conflict resolution dataset associated with UCW HR-8002P. The goal is to implement cloud-based solutions that will enable efficient data processing, cleaning, transformation, analysis, and reporting of HR-related data.

**Dataset Overview:**

The dataset consists of records related to conflict resolution cases within an organization, including:

* **Case ID:** Unique identifier for each conflict case
* **Employee ID:** Identification number for employees involved
* **Department:** Department where the conflict occurred
* **Role:** Employee’s job role (e.g., Manager, Analyst, Associate)
* **Conflict Type:** Nature of the conflict (e.g., Workplace Harassment, Performance Dispute, Policy Violation)
* **Reported Date:** Date the conflict was reported
* **Resolution Date:** Date the conflict was resolved
* **Resolution Time:** Time taken to resolve the conflict
* **Resolution Method:** Mediation, Investigation, Policy Change, or Other
* **Outcome:** Conflict resolution status (Resolved, Unresolved, Escalated)

The dataset consists of HR conflict resolution records, which are initially stored in AWS S3. The data undergoes transformation using AWS Glue and AWS Glue DataBrew, followed by advanced analysis with Amazon Athena. Monitoring and logging are carried out through AWS CloudWatch and CloudTrail.

**Methodology:**

* **Data Collection:** HR data is stored in **Amazon S3** in raw bucket named hr-raw-rup, as a csv file. S3 provides a scalable and secure storage solution for the large volumes of data, ensuring durability and ease of access (*see Appendix, Screenshot 1*).
* **Data cleaning**: Data quality is ensured by using **AWS Glue DataBrew**, a no-code tool that simplifies the process of data cleaning. This step involves identifying and correcting data anomalies, handling missing values, and standardizing data formats (*see Appendix, Screenshot 2)*.
* **Data Transformation**: **AWS Glue** is employed to process the data using Extract, Transform, and Load (ETL) jobs. This service facilitates the transformation of raw data into a structured format, suitable for querying and analysis (*see Appendix, Screenshot 3)*.
* **Data Analysis:** The cleaned and structured data is analyzed using **Amazon Athena**, a serverless query service that enables users to run SQL queries directly on data stored in S3. Athena’s seamless integration with S3 ensures an efficient querying process. This includes 3 business Questions (*see Appendix, Screenshot 4)*.
* **Data Monitoring**: **AWS CloudWatch** is utilized for monitoring system performance and resource usage during the data analysis process. implemented AWS CloudWatch to monitor S3 bucket activities, track bucket size and resource usage, and set up alarms for critical changes (*see Appendix, Screenshot 5)*. **AWS CloudTrail** provides auditing capabilities, allowing for detailed logs of user and service activity for security and compliance purposes (*see Appendix, Screenshot 6)*. Multi-region tracking is enabled for broader coverage, though log validation and SNS alerts are disabled. To enhance real-time monitoring, I configured AWS SNS and CloudWatch alarms, which send instant notifications if any metric crosses its threshold. This setup ensures proactive security, cost efficiency, and optimal data management in my AWS environment.
* **Security & Access Management**: Security is managed using **AWS KMS (Key Management Service)** for encryption and **AWS Organizations** for controlling access across different accounts. These services ensure that sensitive HR data is protected and accessible only to authorized users.

I configured **Default Encryption, Bucket Versioning, and Replication** for three S3 buckets: hr-trf-rup, hr-cur-rup, and hr-raw-rup.

Replication: Set up an S3 replication rule to automatically copy objects between buckets using AWS KMS encryption. This ensures data availability, disaster recovery, compliance, and backup, making the system more resilient.

Default Encryption: Enabled AWS KMS encryption, so objects are automatically encrypted for data security and protection against unauthorized access.

Bucket Versioning: Activated S3 Versioning to track changes, store encrypted versions and prevent accidental loss. IAM and key policies restrict decryption to authorized users for added security (*see Appendix, Screenshot 7)*.

**Tools and Technologies:**

* AWS Glue
* AWS Glue DataBrew
* Amazon S3
* Amazon Athena
* AWS CloudTrail
* AWS CloudWatch
* AWS Key Management Service (KMS)

**Deliverables:**

* Data ingestion and transformation reports
* Analytical insights using Amazon Athena
* Monitoring logs from AWS CloudWatch and AWS CloudTrail
* Security and compliance report using AWS KMS and AWS Organizations
* Screenshots of AWS Console for different services used

**2. Descriptive Analysis:**

**Project Title:** Exploring HR Conflict Resolution Data using AWS Services

**Objective:** The objective of this phase is to conduct a detailed descriptive analysis of the HR conflict resolution dataset. The goal is to uncover trends, patterns, and statistical summaries that provide valuable insights into HR conflict resolution processes.

**Dataset**: The HR conflict resolution data, stored in AWS S3 and processed using AWS Glue DataBrew, is analyzed using Amazon Athena to generate descriptive statistics.

Methodology:

* **Data Ingestion:** Data is ingested from AWS S3 i.e., raw buzcket hr-raw-rup, where it is stored in its raw form (*see Appendix, Screenshot 1*).
* **Data Cleaning and Transformation:** AWS Glue DataBrew is used to clean and preprocess the data by removing duplicates, handling missing values, re-naming the columns and transforming data into an analyzable format (*see Appendix, Screenshot 8)*.
* **Statistical Analysis:** Summary statistics and aggregations are generated through SQL queries in **Amazon Athena** (*see Appendix, Screenshot 4)*.
* **Visualization and Monitoring:** Data trends and metrics are visualized through AWS CloudWatch dashboards to monitor the data's characteristics over time (*see Appendix, Screenshot 5)*.

**Tools and Technologies:**

* AWS S3
* AWS Glue DataBrew
* Amazon Athena
* AWS CloudWatch

**Deliverables:**

* A report containing summary statistics generated using Amazon Athena queries
* Logs detailing the data cleaning and transformation processes
* Screenshots of the AWS Console displaying the data workflows and analysis

**3. Diagnostic Analysis:**

**Project Title:** Diagnostic Analysis of Data Quality Issues in HR Conflict Resolution Records.

**Objective:** The purpose of this analysis is to identify data inconsistencies, missing values, and anomalies in the HR conflict resolution dataset to ensure the quality and accuracy of the data.

**Background:** Accurate data is essential for HR decision-making. This diagnostic analysis seeks to detect data quality issues and recommend necessary corrective actions.

**Dataset:**

The HR conflict resolution dataset stored in **AWS S3** is examined for issues such as missing values, inconsistencies, and duplicates.

**Methodology:**

* **Data Profiling:** **AWS Glue DataBrew** is used to profile the dataset and identify potential anomalies, such as missing or incorrect values (*see Appendix, Screenshot 9)*.
* **Anomaly Detection:** **AWS Glue ETL jobs** are employed to perform anomaly detection and handle any issues found analysis (*see Appendix, Screenshot 3)*.
* **Query-Based Diagnostics:** Queries executed in **Amazon Athena** is used to further identify data inconsistencies (*see Appendix, Screenshot 4)*.
* **Security Auditing:** **AWS CloudTrail** logs any access or modification to the data, ensuring compliance and traceability (*see Appendix, Screenshot 6)*.

**Tools and Technologies:**

* AWS Glue
* AWS Glue DataBrew
* Amazon Athena
* AWS CloudTrail

**Deliverables:**

* A report outlining detected data inconsistencies and anomalies.
* Results from Athena queries used for data quality diagnostics.
* CloudTrail logs documenting access and modifications to the data.
* Screenshots from the AWS Glue DataBrew and CloudTrail dashboards.

**4. Data Wrangling**

**Project Title:** Data Wrangling for HR Conflict Resolution Analytics

**Objective:** The objective of this phase is to preprocess, clean, and structure the HR-8002P dataset in preparation for further analysis.

**Background:** Data wrangling is necessary to ensure that the dataset is consistent, formatted correctly, and free of errors. This process involves handling missing values, removing duplicates, and standardizing data formats.

**Dataset:** The HR-8002P dataset stored in **AWS S3** serves as the input for the data wrangling process.

**Methodology:**

* **Data Ingestion:** The dataset is ingested into **AWS S3** (*see Appendix, Screenshot 1)*.
* **Preprocessing:** The data is cleaned using **AWS Glue DataBrew**, which identifies and resolves inconsistencies (*see Appendix, Screenshot 9)*.
* **Data Structuring:** **AWS Glue** is used to catalog the structured data for ease of access and analysis (*see Appendix, Screenshot 10)*.

**Tools and Technologies:**

* AWS S3
* AWS Glue
* AWS Glue DataBrew

**Deliverables:**

* A cleaned and transformed dataset
* **AWS Glue** data catalog documentation detailing the structure of the dataset
* Screenshots of the **AWS Glue** and **AWS DataBrew** interfaces showing the data transformations and cataloging processes.

**5. Data Quality Control**

**Project Title:** Ensuring Data Quality in HR Conflict Resolution Reports

**Objective:** This phase focuses on establishing data validation processes to ensure that the data used for analysis is of high quality and accuracy.

**Background:** Ensuring data quality is essential for making informed HR decisions. This project establishes a framework for data validation and quality control.

**Scope:** This phase covers identifying and correcting errors in the HR dataset, as well as implementing validation checks to maintain data integrity.

**Methodology:**

* **Data Profiling:** Data is profiled using AWS Glue DataBrew to detect inconsistencies and errors.
* **Validation Rules:** AWS Glue ETL jobs are used to enforce validation rules and correct any issues identified. The system fetches data from the raw S3 bucket, applies transformations to maintain data quality, and records timestamps for entries. It enforces **uniqueness** by preventing duplicate records, ensures **completeness** by verifying mandatory fields, and maintains **freshness** by keeping data up-to-date (*see Appendix, Screenshot 11)*. Records failing these checks are moved to the "failed" folder in the hr-trf-rup bucket for verification, while valid records are stored in the "passed" folder for further processing. This setup ensures data consistency, quality preservation, and accurate analytics (*see Appendix, Screenshot 12)*.
* **Logging:** AWS CloudWatch logs are used to monitor the effectiveness of data quality controls and track improvements over time.

**Deliverables:**

* Data validation and quality control reports
* Logs from AWS CloudWatch showing quality metrics and validation rule outcomes
* Screenshots of the AWS DataBrew profiling and AWS Glue validation rule configurations.