

PROJECT NAME-  
AUTOMATING DATA AUDIT WITH GEN-AI

POWERED APPLICATION

Created by - Mihir Kumar Mallick

College – National Institute of Technology, Rourkela

Employee no. - 255371

Mentor – Mr. Satya Ranjan Harichandan

Manager – Mr. Srinivas Vadde

### ACKNOWLEDGEMENT

The internship opportunity I had with TVS Motor Company Pvt. Ltd. was a great chance for learning professional development. I am grateful to have met so many wonderful people and professionals who led me through the internship period.

I would like to express my deepest gratitude and special thanks to my mentor **Mr. Satya Ranjan Harichandan** andmanager **Mr. Srinivas Vadde** for constantly motivating me and sharing his experience in this field.

I would also like to thank all other team members who have been working with me on the project.

I perceive this opportunity as a big milestone in my career development. I will strive to use the gained skills and knowledge in the best possible way, and I will continue to work on the improvement to achieve desired career objectives.

Mihir Kumar Mallick

**ETL Audit Application - User Manual & Report**

**1. Introduction**

The **ETL Audit Application** is designed to help data engineers and governance teams automate the auditing process of ETL scripts. It evaluates compliance against predefined **Data COE (Center of Excellence) checklists**, **custom organizational policies**, and **industry best practices**. The tool allows users to upload various script files, analyze them using **GenAI**, and generate structured reports with detailed insights.

**2. Supported File Types for audit**

Users can upload ETL scripts in the following formats:

| **Category** | **File Extensions** |
| --- | --- |
| **Python Scripts** | .py |
| **SQL Scripts** | .sql |
| **Shell Scripts** | .sh |
| **YAML Configs** | .yaml, .yml |
| **JSON Data Files** | .json |
| **CSV Files** | .csv |
| **XML Files** | .xml |
| **Java Programs** | .java |
| **Jupyter Notebooks** | .ipynb |
| **Batch Scripts** | .bat |
| **PowerShell Scripts** | .ps1 |
| **Perl Scripts** | .pl |
| **Ruby Scripts** | .rb |
| **PHP Scripts** | .php |
| **R Scripts** | .r |
| **Scala Scripts** | .scala |
| **Go Programs** | .go |
| **C Programs** | .c |
| **C++ Programs** | .cpp |
| **TypeScript Files** | .ts |
| **JavaScript Files** | .js |
| **Big Data & Cloud** | .hql, .bql, .pig, .spark, .pyspark |
| **Workflow/Orchestration** | .airflow, .nifi, .oozie, .dag |
| **Configuration Files** | .toml, .ini, .properties, .cfg |
| **Data Exchange** | .parquet, .avro, .orc, .xls, .xlsx |
| **Machine Learning** | .rmd (R Markdown) |

Additionally, **compressed ZIP files** containing multiple ETL scripts can also be uploaded.

**3. Audit Process**

The tool evaluates ETL scripts based on the **Data COE Checklist**, ensuring compliance with critical aspects of ETL processes:

| **Category** | **Audit Focus** | **Result Format** |
| --- | --- | --- |
| **Auditability** | Evaluates implementation of logging mechanisms, timestamps, process identifiers, and record count tracking for comprehensive monitoring and troubleshooting. | ✅ Pass / ❌ Fail |
| **Reconcilability** | Assesses whether the script includes source-to-target validation checks, checksums, or other mechanisms to verify data completeness and accuracy. | ✅ Pass / ❌ Fail |
| **Restartability** | Examines whether the ETL process can recover from failure points through checkpointing, idempotent operations, or state management capabilities. | ✅ Pass / ❌ Fail |
| **Exception Handling** | Verifies implementation of robust error handling, including try-catch mechanisms, error logging, notifications, and graceful failure modes. | ✅ Pass / ❌ Fail |

Each category assessment includes detailed evidence with specific references to code implementation patterns or deficiencies.

**4. Downloading & Viewing Reports**

Once the audit is complete, users have two options:

**4.1 View Detailed Report**

* The audit results are displayed on the **Results Page**.

**4.2 Download Report**

* Users can download the audit report as a **CSV file** or **PDF file** for documentation and compliance purposes.
* The CSV report includes:
  + **Audit Timestamp**
  + **File Name**
  + **Audit Categories & Results**
  + **Evidence & Recommendations**

**5. How to Use the ETL Audit Tool**

**Step 1: Upload ETL Scripts**

* Navigate to the **Audit Page**.
* Click **Upload** and select ETL scripts (individual files or ZIP format).

**Step 2: Specify Questions (prompt)**

* Users can enter any **audit questions** before running the analysis.  
  Prompt for this - Evaluate the script based on four key parameters: Auditability, reconcilability, restartability and exception handling.

**Step 3: Run the Audit**

* Click **Submit button** to begin the analysis.
* The system processes the script and generates a **structured report**.

**Step 4: Review Results**

* The **Audit Results Page** displays the findings, categorized into predefined questions.

**Step 5: Download Report**

* Click the **Download Report** button to save the audit report in CSV format to use for later reference.

**8. Web page: User Interface**

A screen shot of a computer screen

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**8. Conclusion**

The **ETL Audit Tool** helps automate **compliance verification**, **governance enforcement**, and **best practices validation** for ETL scripts. With structured analysis and AI-driven insights, users can ensure **audit readiness**, **minimize risks**, and **improve data quality**.

THANK YOU FOR YOUR PATIENCE READING.