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

**1. Introduction**

The **ETL Audit Tool** 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**

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 audit process involves two types of questions:

1. **Static Questions** - Based on the **Data COE Checklist** (Predefined audit parameters).
2. **Dynamic Questions** - User-defined questions related to organizational policies or specific ETL requirements.

**3.1 Static Audit Questions (Mandatory Checks)**

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

| **Category** | **Audit Focus** | **Result Options** |
| --- | --- | --- |
| **Auditability** | Ensures logs, timestamps, and record counts are maintained for tracking. | ✅ YES /  ⚠️ PARTIAL /  ❌ NO |
| **Reconcilability** | Validates whether data transformation/movement logs provide a clear source-to-target traceability. | ✅ YES /  ⚠️ PARTIAL /  ❌ NO |
| **Restartability** | Checks if the ETL job can resume from a failure point instead of restarting from scratch. | ✅ YES /  ⚠️ PARTIAL/ ❌ NO |
| **Exception Handling** | Ensures errors are handled with alerts, retries, and fail-safe mechanisms. | ✅ YES /  ⚠️ PARTIAL / ❌ NO |
| **Follows Best Practices** | Checks if the script is structured optimally, using modularization, comments, and performance optimization. | ✅ YES /  ⚠️ PARTIAL /  ❌ NO |

Each category is assessed based on the **technical evidence** extracted from the script.

**3.2 Dynamic Questions (Custom Checks)**

In addition to the predefined checks, users can ask any additional questions specific to their **organizational policies** or **custom governance rules**.

* These questions can be asked in free-text format during the audit process.
* The system will analyze the script and generate **structured responses** for the additional questions.
* These responses will be stored in a separate **"Additional Questions"** section.

**Example of Additional Questions**

1. **Does the script encrypt sensitive data before storage?**
2. **Is there a retry mechanism for failed database connections?**
3. **Are there comments explaining each step of the ETL process?**
4. **Does the script follow the organization’s naming conventions?**

**4. Understanding the Audit Results**

The audit results provide a **structured JSON report** with three key elements:

* **Category** - The rule being evaluated.
* **Result** - Whether the script follows the rule (YES / PARTIAL / NO).
* **Evidence** - A technical explanation of why the script passed or failed.

**5. Downloading & Viewing Reports**

Once the audit is complete, users have two options:

**5.1 View Detailed Report**

* The audit results are displayed on the **Results Page**.
* Users can review both **predefined audit checks** and **additional custom questions**.

**5.2 Download Report**

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

**6. 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 Additional Questions (Optional)**

* Users can enter any **custom audit questions** before running the analysis.

**Step 3: Run the Audit**

* Click **Start Audit** 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 and additional questions.

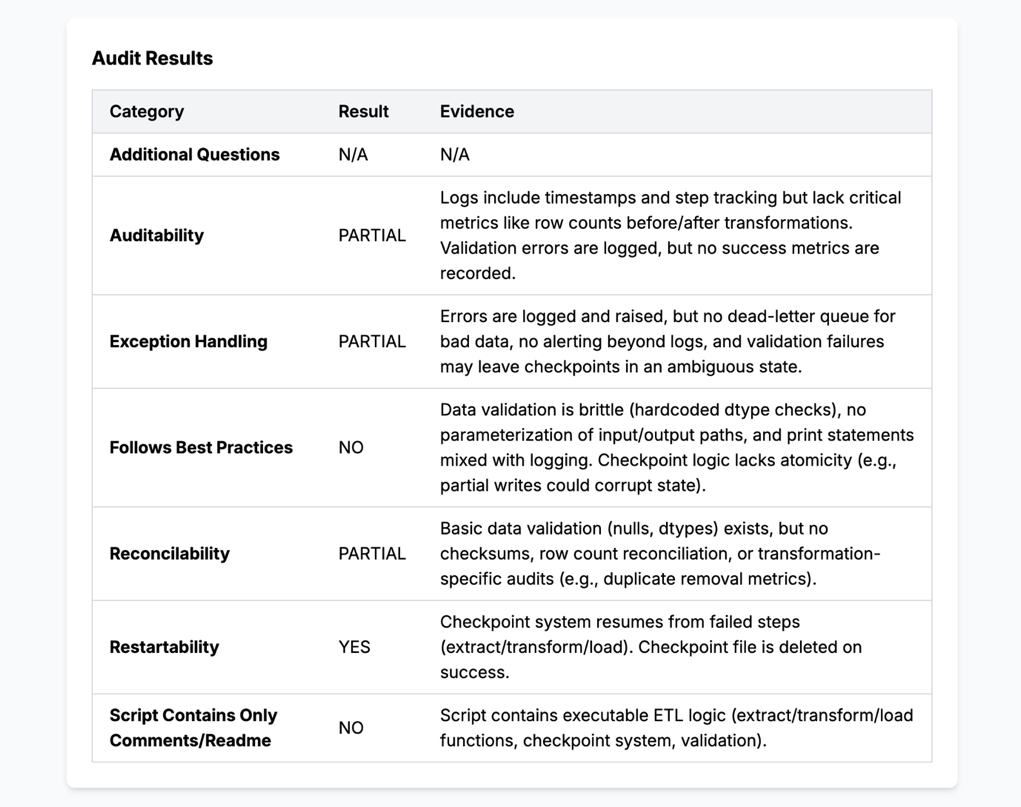
**Step 5: Download Report**

* Click the **Download Report** button to save the audit report in CSV format.

**8. Web page : User Interface**

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**.