Proposed Architecture Design

1. Data Ingestion Layer

Source: Library databases or spreadsheets (e.g., CSV, Excel, SQL).

Tool: Python scripts using pandas, openpyxl, or SQLAlchemy.

2. Data Processing & Filtering

Tool: Python

Functionality:

Clean and validate data.

Apply filtering logic (e.g., quality thresholds, missing values).

Generate summary statistics or reports.

3. Automation & CI/CD

Tool: Azure DevOps Pipelines

Functionality:

Trigger Python scripts on a schedule or when new data is available.

Use YAML pipeline definitions to manage steps (e.g., install dependencies, run scripts, store results).

Integrate with Git for version control.

4. Storage & Output

Options:

Store results in Azure Blob Storage, SharePoint, or a local network drive.

Export filtered data and reports as Excel, CSV, or PDF.

5. Monitoring & Logging

Tool: Azure Monitor or custom logging in Python

Functionality:

Track pipeline runs and script execution.

Log errors and performance metrics.

6. Optional Enhancements

Dashboarding: Use Power BI or Streamlit for visualizing quality metrics.

Notifications: Send email alerts via Azure Logic Apps or Python (e.g., smtplib) when issues are detected.