Workflow Processor & Diagram Generator Documentation

Overview

The **Workflow Processor & Diagram Generator** is a powerful analytical tool designed to clean, process, and visualize complex batch or job dependencies sourced from Excel files. It automates common data cleanup tasks (like handling merged cells and standardizing delimiters) to generate actionable dependency reports and visual workflow diagrams.

---

1. Input File Requirements

For the application to function correctly, the input Excel file (.xlsx) must adhere to a minimal structure. Column matching is performed **case-insensitively**.

Mandatory Columns

Your file **must** contain columns corresponding to the following three core concepts:

Concept	Description	Data Format Notes
Batch	The unique identifier for the current process, job, or script.	Must be a unique identifier (e.g., BATCH_001).
Predecessor	A list of Batches that **must complete** before the current Batch can start.	Multiple dependencies must be separated by a **slash (/)**. Newlines (\n) will be automatically converted to slashes.
Successor	A list of Batches that are triggered **after** the current Batch completes.	Multiple dependencies must be separated by a **slash (/)**.

Optional Columns (Contextual Data)

If present, these columns will be automatically included in the output tables and diagrams, adding crucial context:

Column Headers Essential To Be Present in XLSX for program to run.

- **Chain:** Used as categories.
- **Batch:** It contains the batch name.
- **Predecessor:** It involves previous dependent batch name.,
- **Successor:** It involves all the successor batches.
- **Start Time:** It involves the time period when a batch is executed.
- **Path, Condition, Parameter 1, Parameter 2, Parameter 3:** Included in the final **Summary Table** for administrative context.

2. Step-by-Step Usage Guide

The application is controlled entirely through the sidebar interface.

Step 1: Upload and Review Cleanup

1. Click **"Upload Input Excel File (.xlsx)"** in the sidebar.
2. The application immediately runs the **data cleaning process** (un-merging cells and standardizing delimiters).
3. Verify the **Data Cleanup Status** is **Complete**. You can download the processed file using the **"⬇️ Download Cleaned Excel File"** button for audit or reuse.

Step 2: Configure Analysis Scope

Before clicking "Generate," define the scope of your analysis:

Sidebar Control	Description
Filter by Chain	Use this if your optional Chain column is present. Selecting a chain limits all subsequent analysis (Batch selection, Diagram, Tables) to only batches belonging to that chain.
Select Batch to Map Around	Select a specific batch to act as the **center** of your analysis. The resulting diagram will trace dependencies inward and outward from this batch. Select **"— Show All Batch —"** to analyze all available batches within the current Chain filter.
Diagram Dependency Depth	Adjust the slider to define the maximum number of **hops** (levels of connection) to trace from the selected batch. A depth of 1 shows immediate predecessors/successors.

Step 3: Generate and Interpret Results

1. Click the **"Generate Summary & Diagram"** button.

---

3. Output Interpretation

Dependency Insights (Metrics & Chart)

This section provides a high-level view of complexity:

- **Total Batches Analyzed:** The number of unique batches included in the current filtered scope.
- **Top 10 Most Critical Batches:** A bar chart visualizing the batches with the highest count of **Post Dependencies**. These jobs are considered the most critical "fan-out" points, as the failure or delay of one impacts the greatest number of downstream jobs.

SUMMARY TABLE (Calculated Dependencies)

This is the core analytical output. It calculates the network connections for every batch in the scope:

Column	Calculation	Purpose
PRE DEPENDENCY	The complete list of unique **Predecessors** for the batch.	What the batch is waiting for.
NO OF PREDEPENDENCY	The count of unique predecessors.	Complexity metric (Input).
POST DEPENDENCY	The complete list of unique batches that cite the current batch as their predecessor.	What is waiting for the batch.
NO OF POST DEPENDENCIES	The count of unique post-dependencies.	**Criticality Metric** (Output).

Batch Dependency Diagram (Visualization)

The diagram visually maps the relationships using **Graphviz**.

Visual Element	Meaning
Nodes (Boxes)	Represent individual batches.
Arrows (Edges)	Represent the dependency link (Predecessor → Successor).
Highlighting	The **selected batch** is highlighted in **light blue**.
External Nodes	Nodes not found in the Batch column of your filtered data (but referenced in the Predecessor column) are shown in a lighter gray color.
Node Labels	Include the **Batch Name** on the top line and the **Time/Frequency** (if available) below it.

Download Options: Use the buttons below the diagram to download the image in standard or high resolution (ideal for presentations and large prints).