what is workflow?

Quick Answer:

A workflow is simply a **sequence of steps or tasks** that need to be completed in a specific order to achieve a goal. In data engineering, workflows describe how raw data is collected, cleaned, transformed, and delivered for analysis Dagster +1 .

🔄 General Definition of Workflow

- Sequence of tasks: A workflow is like a checklist of actions that must be done in order.
- Actors: The people or systems responsible for carrying out each task.
- Inputs: The resources or data needed to perform the tasks.
- Outputs: The final result after all tasks are completed (e.g., a report, a cleaned dataset).
- Workflow management: The process of defining, monitoring, and optimizing these steps to ensure efficiency GeeksForGeeks .

📊 Workflow in Data Engineering

- Data acquisition: Collecting raw data from sources like APIs, databases, or files.
- Data transformation: Cleaning, filtering, and reshaping data into usable formats.
- Data storage: Saving processed data into databases, warehouses, or data lakes.
- Data delivery: Making data accessible to analysts, dashboards, or machine learning models.
- Automation: Tools like Apache Airflow automate workflows so they run reliably and on schedule Dagster .

🧩 Analogy

Think of a workflow as a **recipe**:

- Ingredients = inputs (raw data)
- Steps = tasks (cleaning, transforming, loading)
- Chef = actor (data engineer or system)
- Final dish = output (usable dataset or report)

Just like a recipe ensures the dish comes out right, a workflow ensures data is processed correctly.

🚀 Why Workflows Matter

- Consistency: Ensures tasks are done the same way every time.
- Efficiency: Automates repetitive steps, saving time.
- Reliability: Reduces errors by enforcing order and dependencies.
- Scalability: Handles growing amounts of data without breaking.