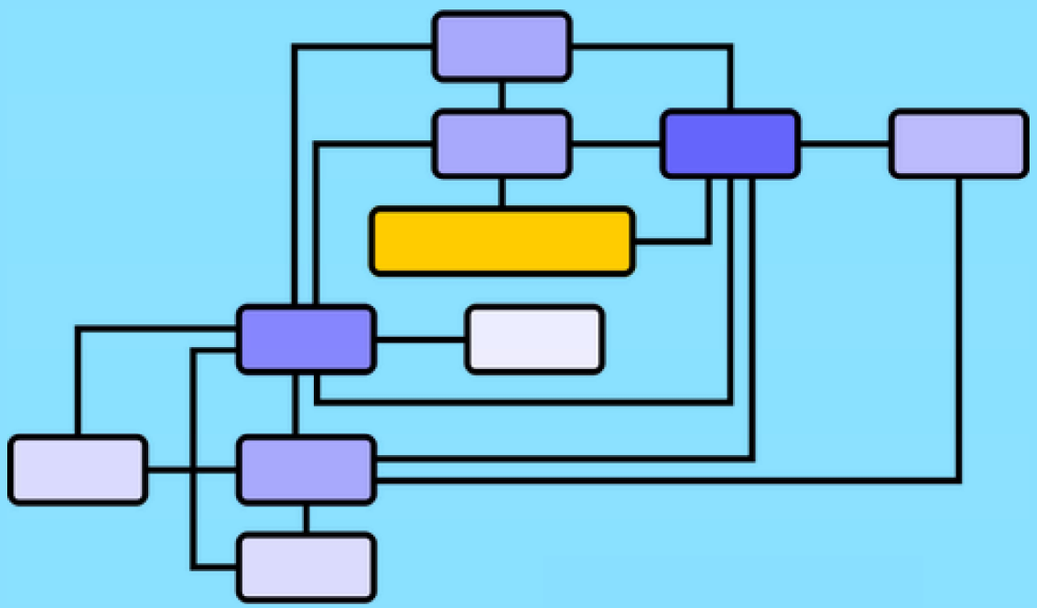
Databricks Workflows

databricks

New Features!

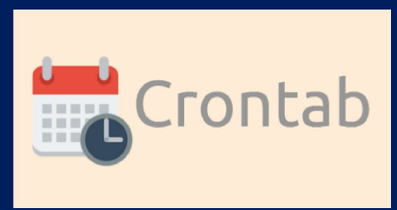


Where Are We Going?

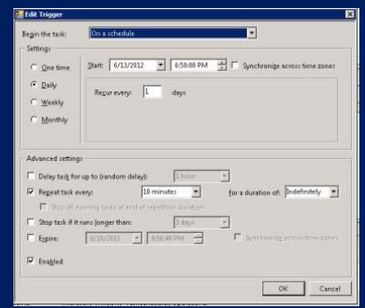
- What are Databricks Workflows?
- Riveting Demo
- Why Use Workflows?
- Limitations
- Use Cases

What are Databricks Workflows?

- Define Series of Activities to Complete a Task
- Scheduling
- Logging and Alerting
- Parallel Task Execution
- Is NOT a Data Movement and Transformation Service (DMAT)



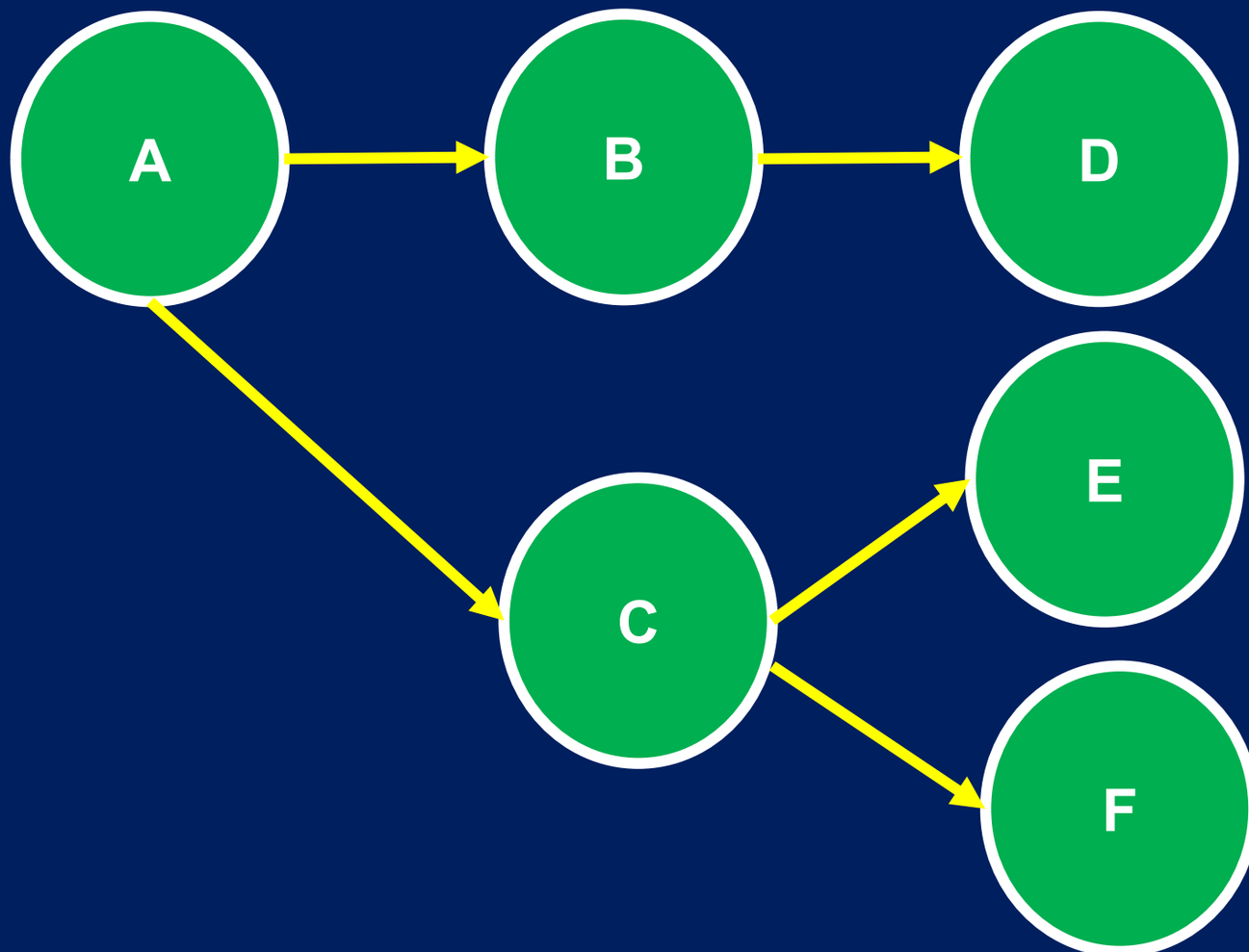
SQL Server Agent



Windows Scheduler

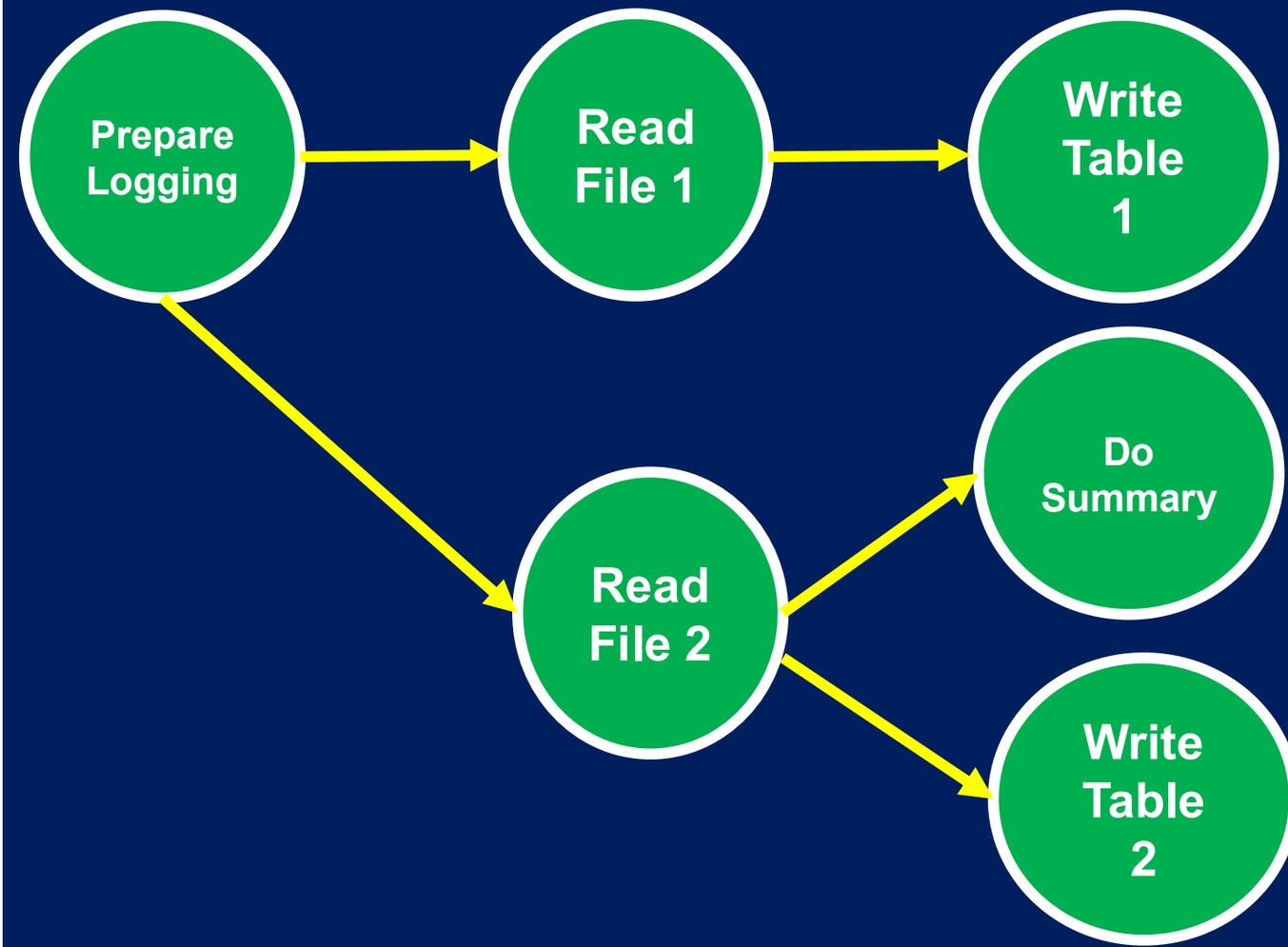
Workflows Use a DAG

Directed Acyclic Graph



- Vertices (Objects)
- Edges (Relationships)
- Objects and Their Relationships to Other Objects
- Directional (Directed)
- No Loops!!! Acyclic

Example Workflow



- Start: Prepare Logging.
- Concurrently Run:
  - Read File 1
  - Read File 2
- After Read File 1 Completes:
  - Write Table 1
- After Read File 2 Completes:
  - Do Summary
  - Write Table 2

Demo

Let's See It Work!

Why Use Workflows?

- Zero Effort to Set Up
- Can Execute Many Types of Tasks
- Highly Configurable
- Fits into the Databricks Roadmap
- Powerful No Code Solution
- Full Supported by the REST API

Limitations

- Only Available on Databricks
- Ideally Suited for Databricks Workloads
- May Not Fit Extremely Complex Workflows

What are Workflows Good For?

- Data Movement and Transformation (DMAT)
- Machine Learning Pipelines
- Preparing Data for Visualization Tools
- Databricks Automation