FANS

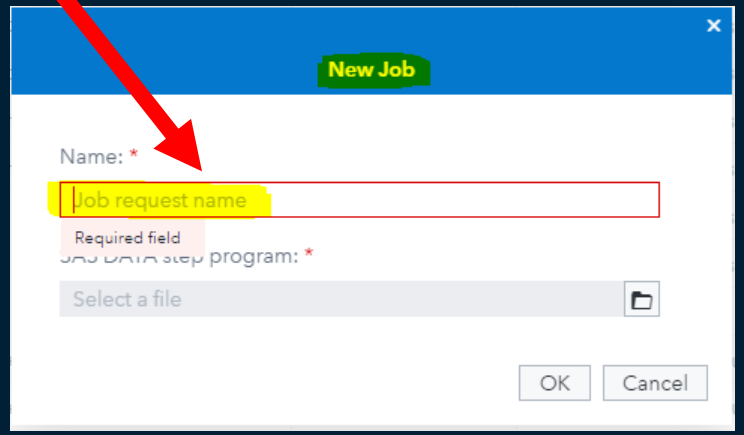
Byg, deploy og skedulér flows i SAS Studio

23. maj 2023

Cecily Hoffritz
Data Engineering Advisor
SAS Nordics

Terminologi DI Studio → SAS Studio

DI Studio	SAS Studio
Job	Flow
Deployed Job	Job / Job request?? (Environment Manager)
Flow (Schedule manager)	Job Flow (Environment Manager)



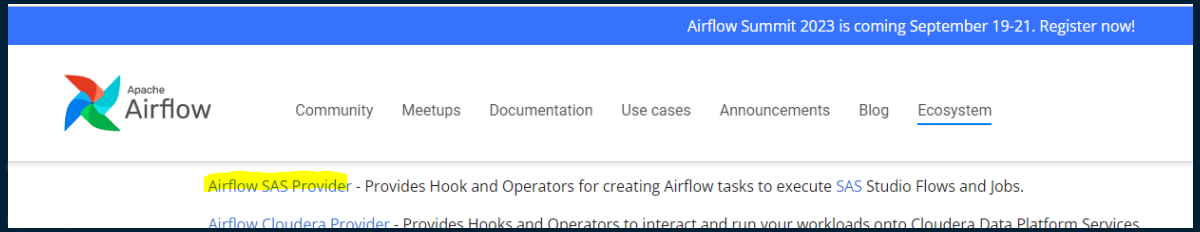
SAS Viya, SAS Studio og Airflow Integration

SAS & Airflow integration

SAS provider for Airflow

Apache Airflow er et af de mest populære open-source-værktøjer til orkestrering, skedulering og overvågning af data-workflows.

Med **Airflow SAS Provider** kan du oprette Airflow tasks til kørsel af SAS Studio Flows, kode og jobs.

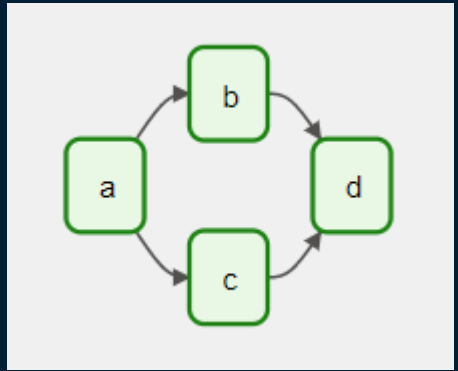


Terminologi

Airflow DAG

En DAG (Directed Acyclic Graph) er kerne i Airflow og samler “Tasks”, som er organiseret med afhængigheder, der viser, hvordan de skal afvikles.

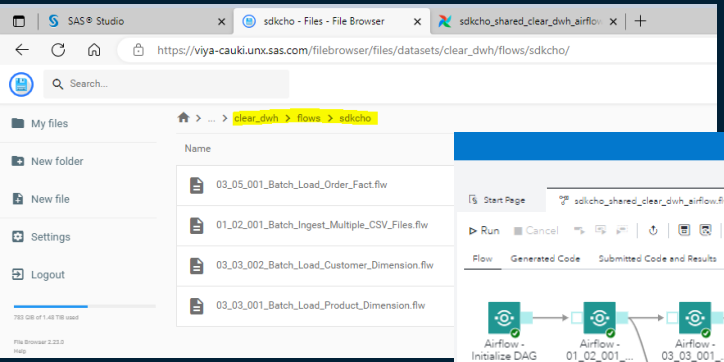
En DAG defineres i et Python program.

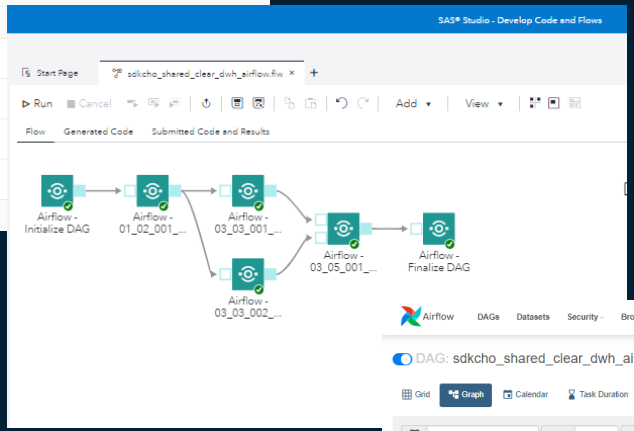


Acyklisk betyder ikke at vise eller indgå i en cyklus.

SAS & Airflow integration

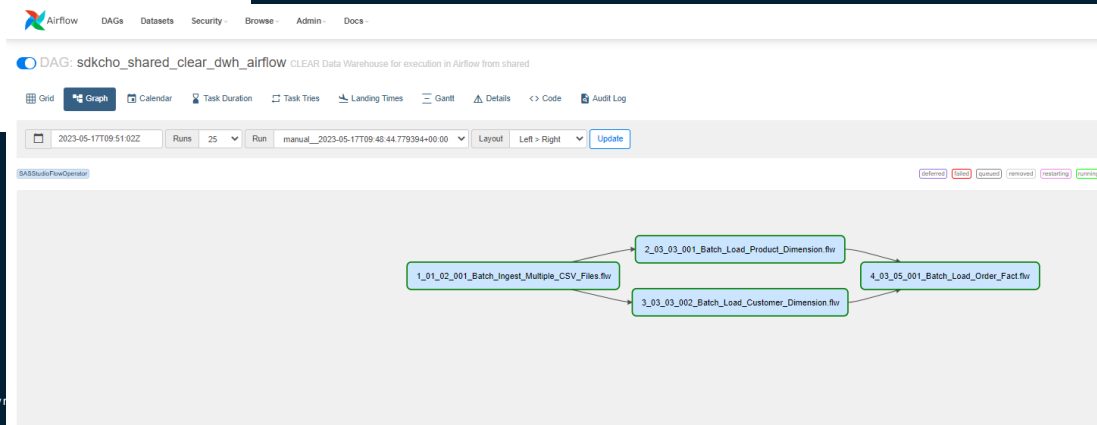
Oversigt over SAS Studio Flows





I SAS Studio bruges custom steps til at oprette Airflow DAG til orkestrering af SAS Studio flows.

I Airflow eksekveres SAS Studio flows (grøn ramme = success).



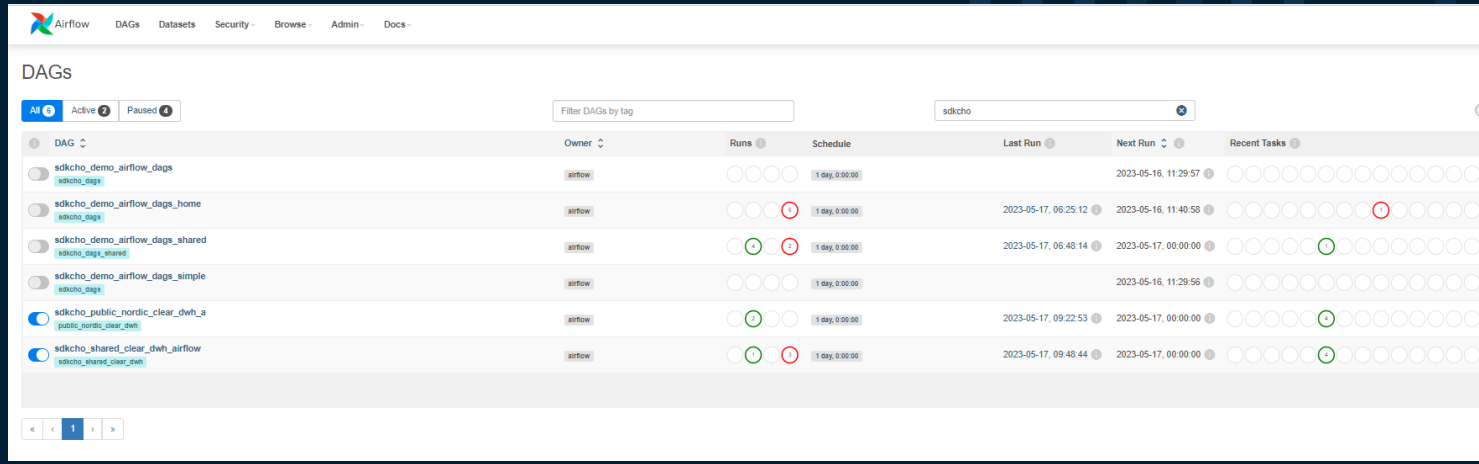
Airflow DAGs Datasets Security Browse Admin Docs

```
# specific language governing permissions and limitations
# under the License.

from datetime import datetime
from airflow import DAG, Dataset
from sas_airflow_provider.operators.sas_studioflow import SASStudioFlowOperator
from sas_airflow_provider.operators.sas_jobexecution import SASJobExecutionOperator

dag = DAG(dag_id="sdkcho_shared_clear_dwh_airflow",
          description="CLEAR Data Warehouse for execution in Airflow from shared",
          start_date=datetime(2023,5,17),
          tags=["sdkcho_shared_clear_dwh"],
          catchup=False)

task1 = SASStudioFlowOperator(task_id="1_01_02_001_Batch_Ingest_Multiple_CSV_Files.flw",
                              flow_path_type="compute",
                              flow_path="/data/home/shared/datasets/clear_dwh/flows/sdkcho/01_02_001_Batch_Ingest_Multiple_CSV_Files.flw",
                              compute_context="SAS Studio compute context",
                              connection_name="sas_default",
                              flow_exec_log=True,
                              flow_codegen_init_code=False,
                              flow_codegen_wrap_code=False,
                              dag=dag)

task2 = SASStudioFlowOperator(task_id="2_03_03_001_Batch_Load_Product_Dimension.flw",
                              flow_path_type="compute",
                              flow_path="/data/home/shared/datasets/clear_dwh/flows/sdkcho/03_03_001_Batch_Load_Product_Dimension.flw",
                              compute_context="SAS Studio compute context",
                              connection_name="sas_default",
                              flow_exec_log=True,
                              flow_codegen_init_code=False,
                              flow_codegen_wrap_code=False,
                              dag=dag)

task3 = SASStudioFlowOperator(task_id="3_03_03_002_Batch_Load_Customer_Dimension.flw",
                              flow_path_type="compute",
                              flow_path="/data/home/shared/datasets/clear_dwh/flows/sdkcho/03_03_002_Batch_Load_Customer_Dimension.flw",
                              compute_context="SAS Studio compute context",
                              connection_name="sas_default",
                              flow_exec_log=True,
                              flow_codegen_init_code=False,
                              flow_codegen_wrap_code=False,
                              dag=dag)

task4 = SASStudioFlowOperator(task_id="4_03_05_001_Batch_Load_Order_Fact.flw",
                              flow_path_type="compute",
                              flow_path="/data/home/shared/datasets/clear_dwh/flows/sdkcho/03_05_001_Batch_Load_Order_Fact.flw",
                              compute_context="SAS Studio compute context",
                              connection_name="sas_default",
                              flow_exec_log=True,
                              flow_codegen_init_code=False,
                              flow_codegen_wrap_code=False,
                              dag=dag)

task1 >> task2
task1 >> task3
task2 >> task4
task3 >> task4
```

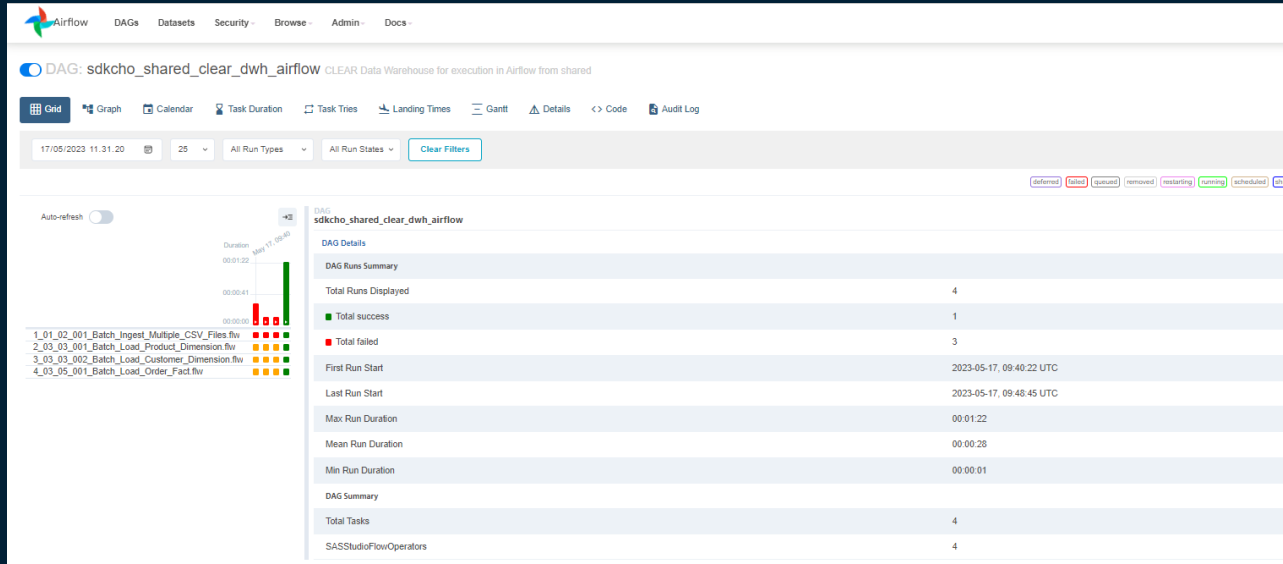
Genereret DAG-kode

sas

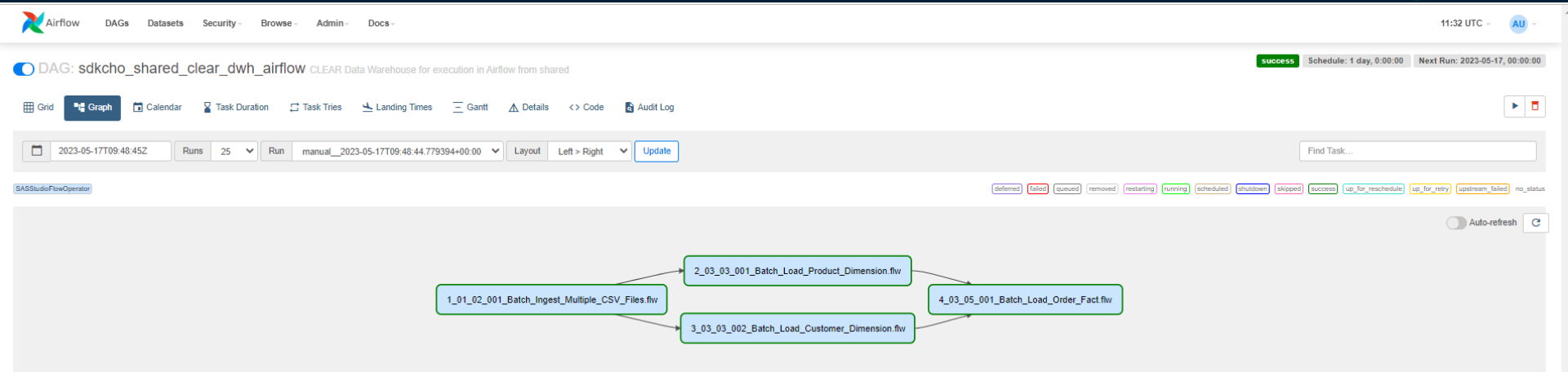
En liste af DAGS med præfiks **sdkcho**

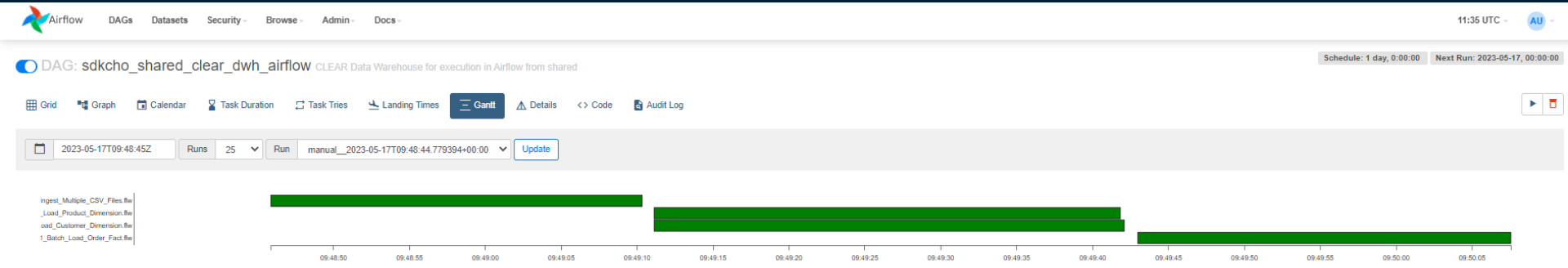


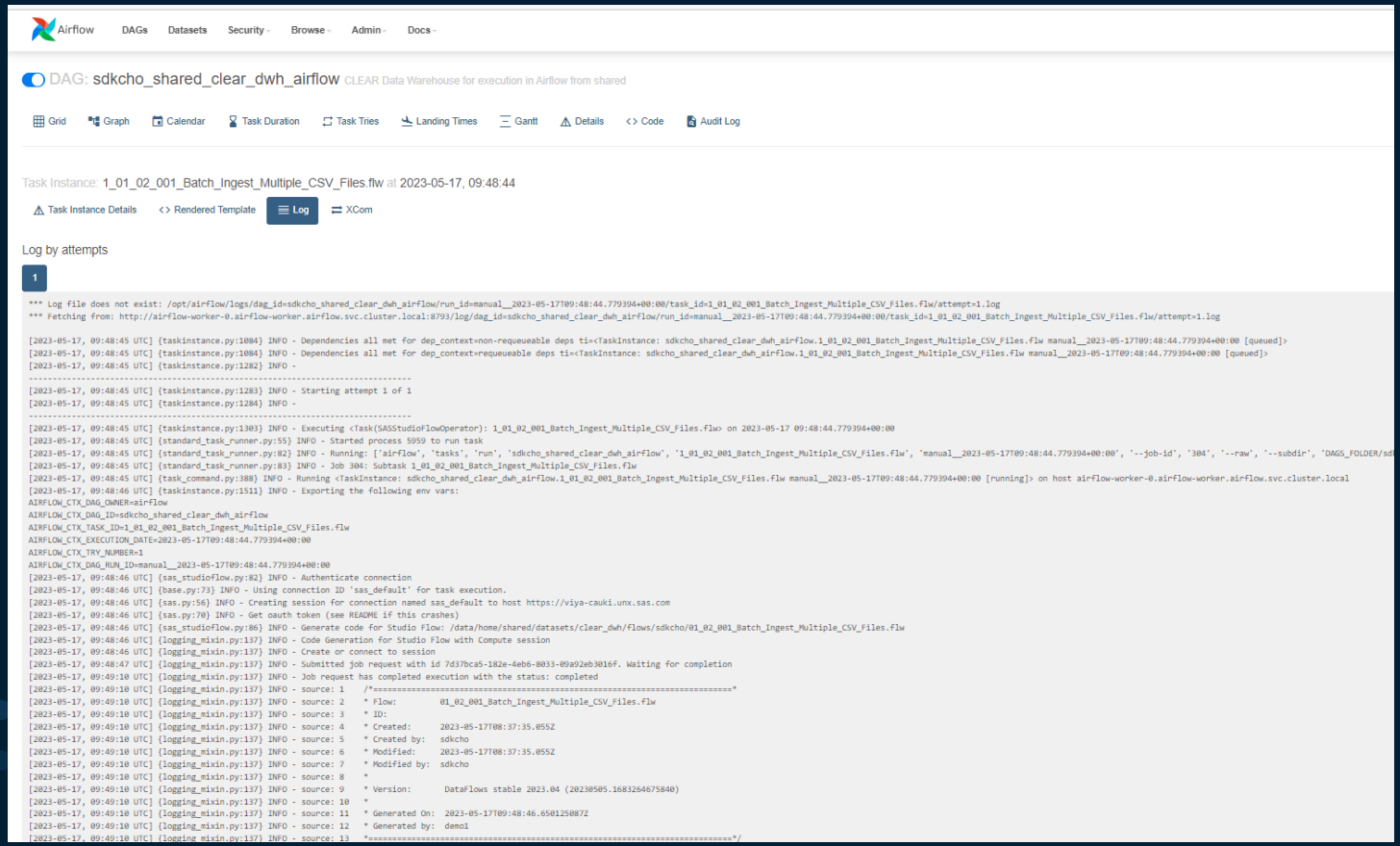
En liste af tasks i en DAG og deres kørselstilstand (Sidste kørsel grøn for alle 4 tasks)



Grafisk & Gantt overblik over tasks i en DAG







SAS log fra en DAG task.

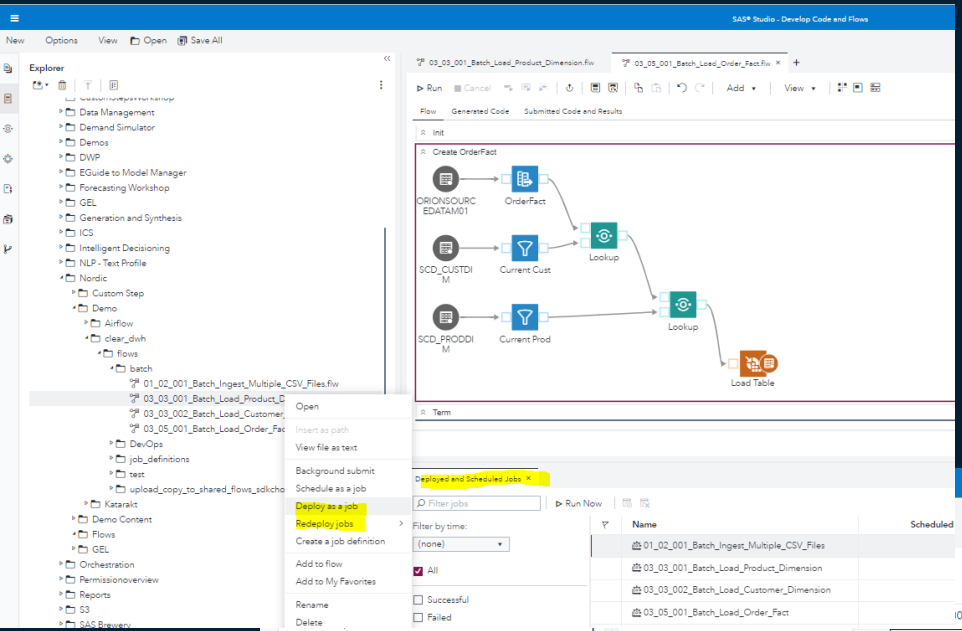
Live Demo

Deploy, redeploy og skedulere fra SAS Studio

- Via SAS Studio menuer
- Programmatisk gennem SAS Studio flows

SAS Studio menu - deploy/redeploy som job

- Man kan deploye et SAS Studio flows, et SAS program m.fl.
- Man kan redeploy et job
  - Ved redeploy opdateres indhold og UID ændres ikke.
- Man kan skedulere de enkelte deployed jobs.
- Man kan skedulere deployed jobs på ny.



Dette sker, når man skedulerer i SAS Studio

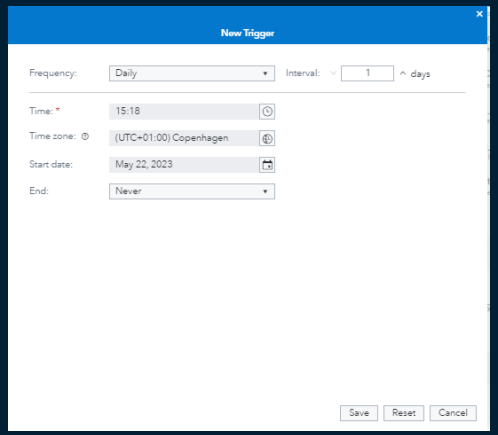
Du får et jobrequest objekt, som bliver skeduleret til kørsel i SAS Viyas Job Scheduler

SAS Studio Flow

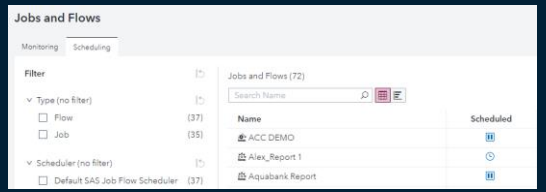
SAS program

Andet SAS objekt

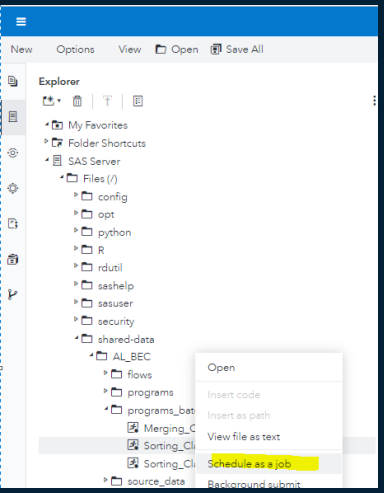
Job Request

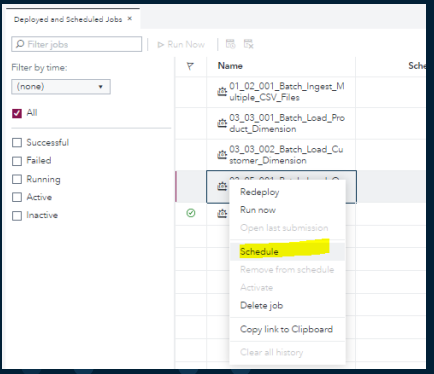


SAS Environment Manager
Jobs & Flows



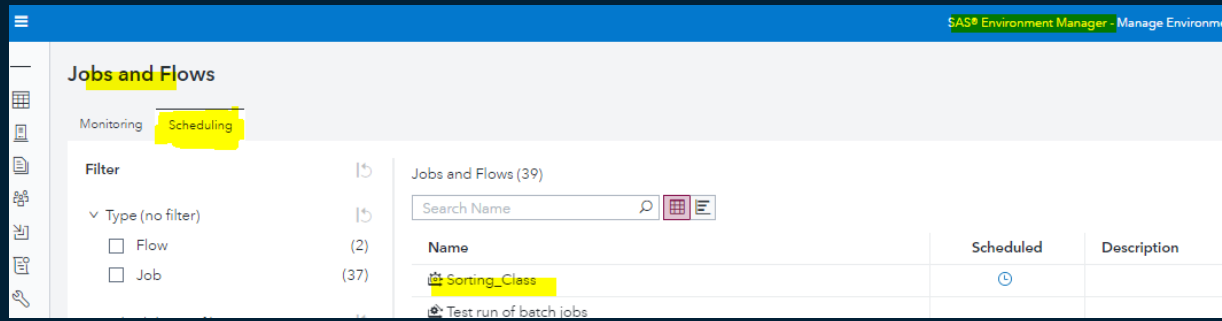
Dette sker, når man skedulerer i SAS Studio



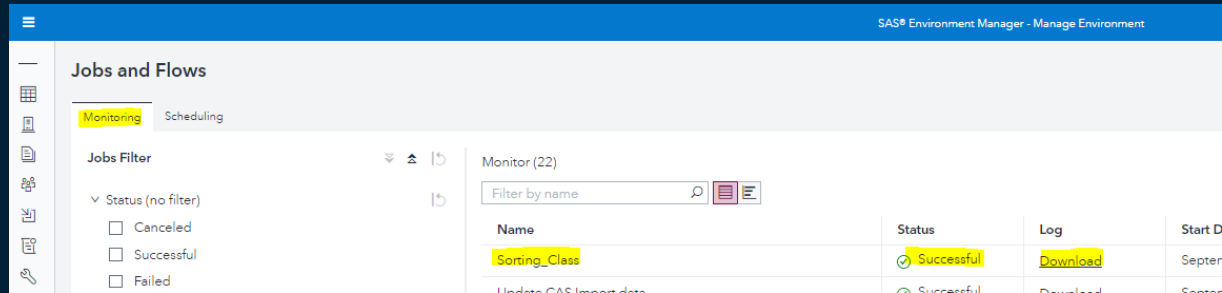


1: I SAS Studio kan du skedulere fra en menu

2: Dit skedulerede job eksekveres i Jobs and Flows i SAS Environment Manager.

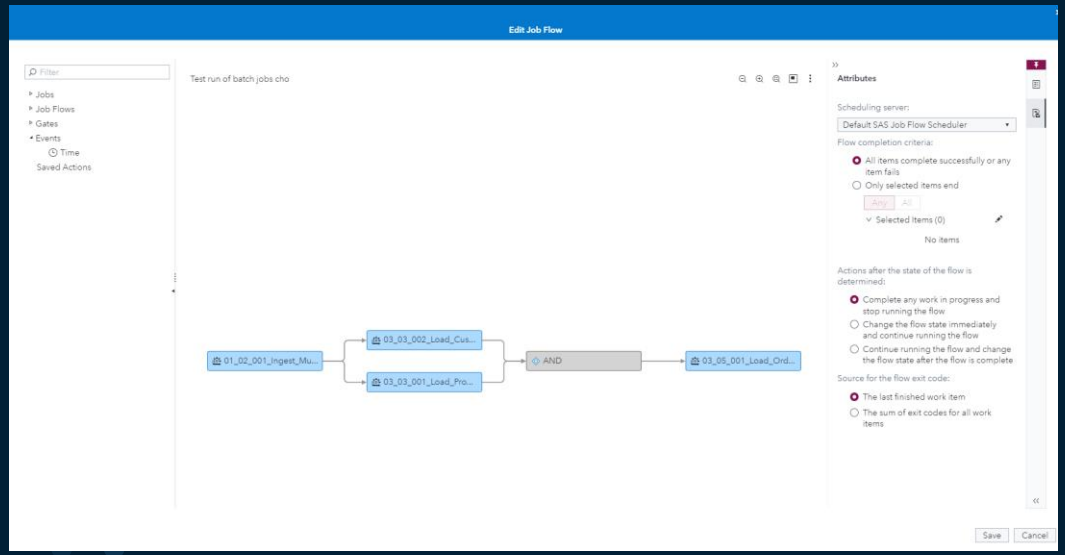


3: I SAS Environment Manager kan du overvåge jobs.



Byg jobflow med afhængigheder i SAS Environment Mgr

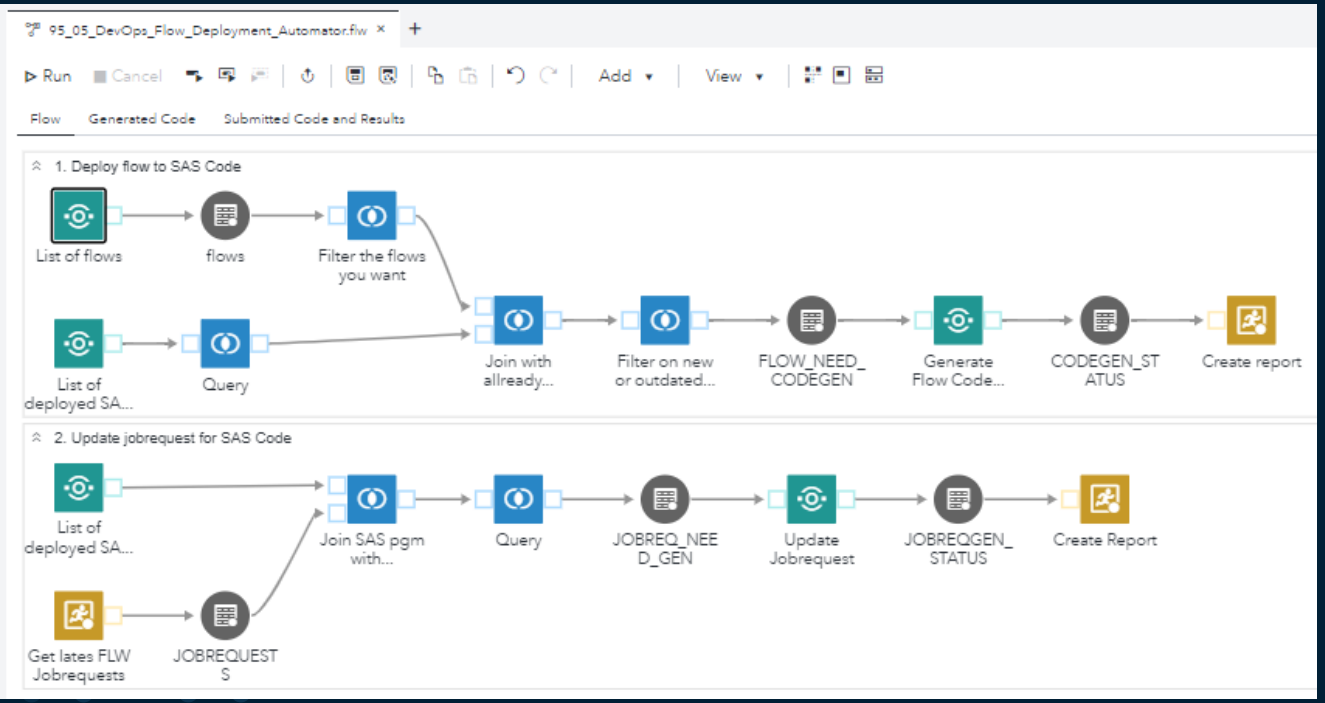
Jobs and Flows menuen



Visuel orkestrering af job flow i Jobs and Flows, SAS Environment Manager

DevOps Flow Deployment Automator

I et SAS Studio flow kan man deploye multiple SAS Studio flows og danne multiple jobrequests.



Deploy .flw til .sas
(med custom steps)

Dan jobrequest
(til brug ved skedulering)

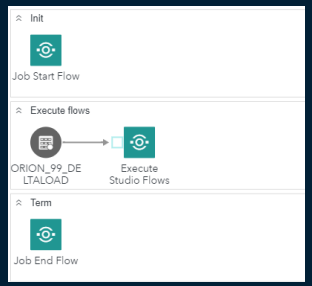
DevOps – orkestrere & skedulere i batch

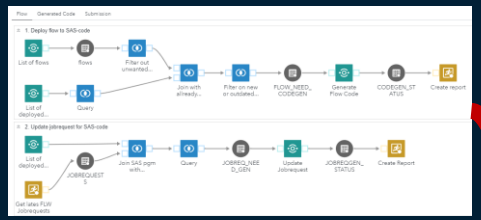
```
data control.ORION_99_DeltaLoad;
    attrib
        flowname informat=$200. length=$200
        async informat=$3. length=$3
    ;
    infile cards4 dsd missover dlm=';';
    input flowname async;
cards4;
01_02_001_Ingest_Multiple_CSV_Files_SL.flw;NO
03_03_001_Load_Product_Dimension_SL.flw;YES
03_03_002_Load_Customer_Dimension_SL.flw;YES
03_05_001_Load_Order_Fact_SL.flw;NO
;;;;
run;
```

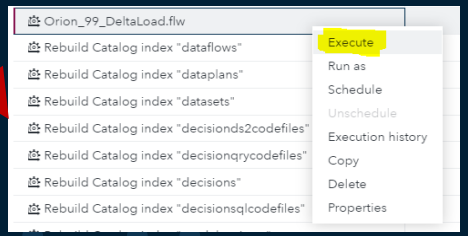
ORION_99_DELTALOAD

Enter expression

	flowname	async
1	01_02_001_Ingest_Multiple_CSV_Files_SL.flw	NO
2	03_03_001_Load_Product_Dimension_SL.flw	YES
3	03_03_002_Load_Customer_Dimension_SL.flw	YES
4	03_05_001_Load_Order_Fact_SL.flw	NO







Name	Start Date	Status	Created By	Log
03_03_002_Load_Customer_Dim...	May 22, 2022 05:20:27 PM	Running	sasdemo	
03_03_001_Load_Product_Dime...	May 22, 2022 05:20:26 PM	Running	sasdemo	
01_02_001_Ingest_Multiple_CSV...	May 22, 2022 05:19:56 PM	Successful	sasdemo	Download
Orion_99_DeltaLoad.flw	May 22, 2022 05:19:53 PM	Running	sasdemo	Download

Kontroltabellen

```
data control.ORION_99_DeltaLoad;
    attrib
        flowname informat=$200. length=$200
        async informat=$3. length=$3
    ;
    infile cards4 dsd missover dlm=';';
    input flowname async;
cards4;
01_02_001_Ingest_Multiple_CSV_Files_SL.flw;NO
03_03_001_Load_Product_Dimension_SL.flw;YES
03_03_002_Load_Customer_Dimension_SL.flw;YES
03_05_001_Load_Order_Fact_SL.flw;NO
;;;;
run;
```

- Definér afhængigheder mellem SAS Studio flows.
  - Skal de køre asynkront – ja eller nej?

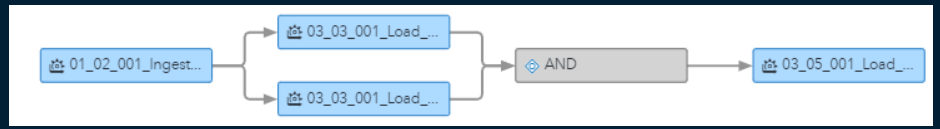
Async - skal et job vente til et andet job er kørt færdigt?

ORION_99_DELTALOAD

Enter expression

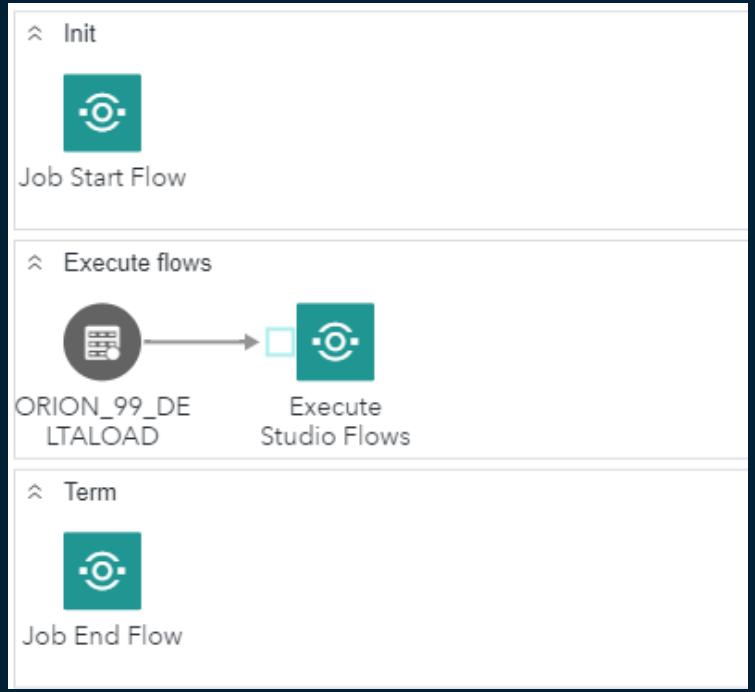
	flowname	async
1	01_02_001_Ingest_Multiple_CSV_Files_SL.flw	NO
2	03_03_001_Load_Product_Dimension_SL.flw	YES
3	03_03_002_Load_Customer_Dimension_SL.flw	YES
4	03_05_001_Load_Order_Fact_SL.flw	NO

=



Byg batchflowet

- Benyt custom step **Execute Studio Flows**.
- Læs kontroltabellen
- Hvis Async = Yes, bliver SAS Studio flows/programmer eksekveret parallel.



Live Demo

Roadmap

Roadmap

Nu

- File-based triggers er supporteret

```
[{
  "name": "file exists trigger",
  "type": "fileevent",
  "active": true,
  "event": {
        "expression": "exists('output.txt')",
        "description": "Determine when the output file exists."
       }
}]
```

Tak

sas.com