



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) 



New Job

Name: *

Job request name

Required field

SAS DATA step program: *

Select a file

OK Cancel

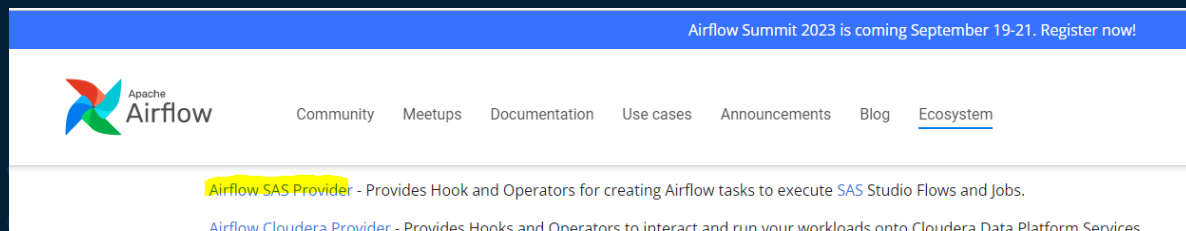
SAS Viya, SAS Studio og Airflow Integration

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.

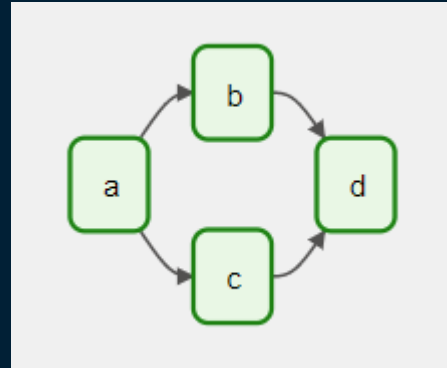


Terminologi

Airflow DAG

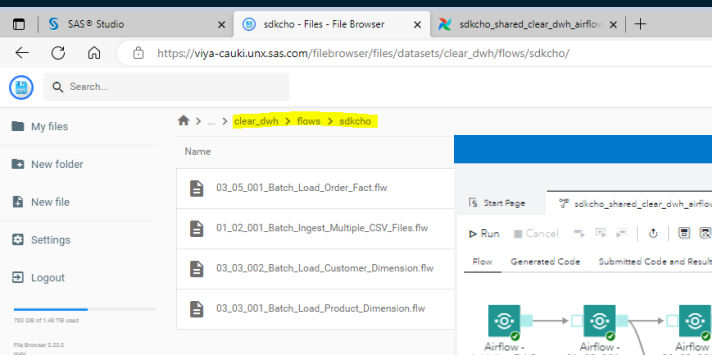
En DAG (Directed Acyclic Graph) er kerne i Airflow og samler “Tasks”, som er organiseret med afhængigheder, der viser, hvordan de skal afvikles.

En DAG defineres i et Python program.

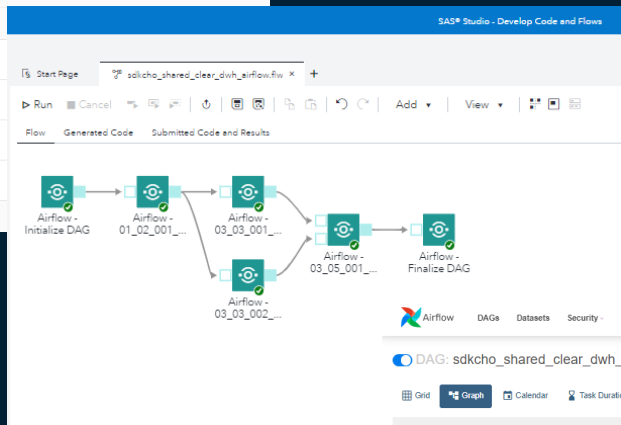


SAS & Airflow integration

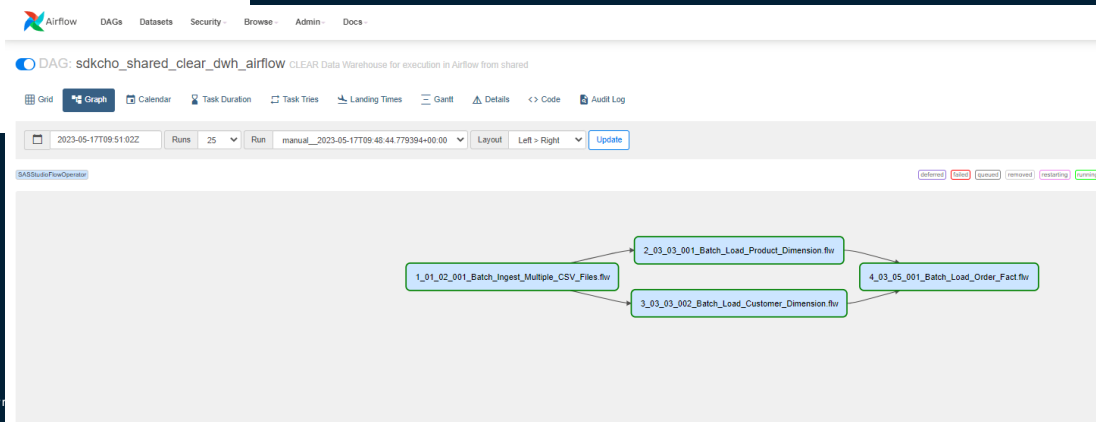
Oversigt over SAS Studio Flows



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).



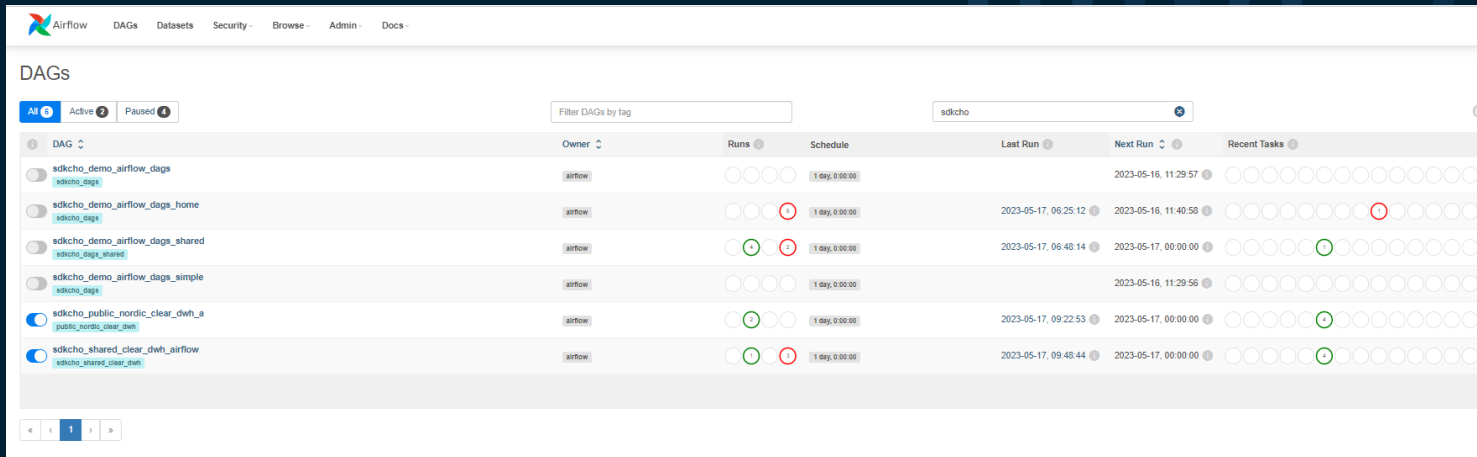
```

16 # under the license.
17
18 from datetime import datetime
19 from airflow import DAG, Dataset
20 from sas_airflow_provider.operators.sas_studioflow import SASStudioFlowOperator
21 from sas_airflow_provider.operators.sas_jobexecution import SASJobExecutionOperator
22
23 dag = DAG(dag_id="sdkcho_shared_clear_dwh_airflow",
24          description="CLEAR Data Warehouse for execution in Airflow from shared",
25          start_date=datetime(2023,5,17),
26          tags=["sdkcho_shared_clear_dwh"],
27          catchup=False)
28
29 task1 = SASStudioFlowOperator(task_id="1_01_02_001_Batch_Ingest_Multiple_CSV_Files.flw",
30                              flow_path_type="compute",
31                              flow_path="/data/home/shared/datasets/clear_dwh/flows/sdkcho/01_02_001_Batch_Ingest_Multiple_CSV_Files.flw",
32                              compute_context="SAS Studio compute context",
33                              connection_name="sas_default",
34                              flow_exec_log=True,
35                              flow_codegen_init_code=False,
36                              flow_codegen_wrap_code=False,
37                              dag=dag)
38
39 task2 = SASStudioFlowOperator(task_id="2_03_03_001_Batch_Load_Product_Dimension.flw",
40                              flow_path_type="compute",
41                              flow_path="/data/home/shared/datasets/clear_dwh/flows/sdkcho/03_03_001_Batch_Load_Product_Dimension.flw",
42                              compute_context="SAS Studio compute context",
43                              connection_name="sas_default",
44                              flow_exec_log=True,
45                              flow_codegen_init_code=False,
46                              flow_codegen_wrap_code=False,
47                              dag=dag)
48
49 task3 = SASStudioFlowOperator(task_id="3_03_03_002_Batch_Load_Customer_Dimension.flw",
50                              flow_path_type="compute",
51                              flow_path="/data/home/shared/datasets/clear_dwh/flows/sdkcho/03_03_002_Batch_Load_Customer_Dimension.flw",
52                              compute_context="SAS Studio compute context",
53                              connection_name="sas_default",
54                              flow_exec_log=True,
55                              flow_codegen_init_code=False,
56                              flow_codegen_wrap_code=False,
57                              dag=dag)
58
59 task4 = SASStudioFlowOperator(task_id="4_03_05_001_Batch_Load_Order_Fact.flw",
60                              flow_path_type="compute",
61                              flow_path="/data/home/shared/datasets/clear_dwh/flows/sdkcho/03_05_001_Batch_Load_Order_Fact.flw",
62                              compute_context="SAS Studio compute context",
63                              connection_name="sas_default",
64                              flow_exec_log=True,
65                              flow_codegen_init_code=False,
66                              flow_codegen_wrap_code=False,
67                              dag=dag)
68
69 task1 >> task2
70 task1 >> task3
71 task2 >> task4
72 task3 >> task4

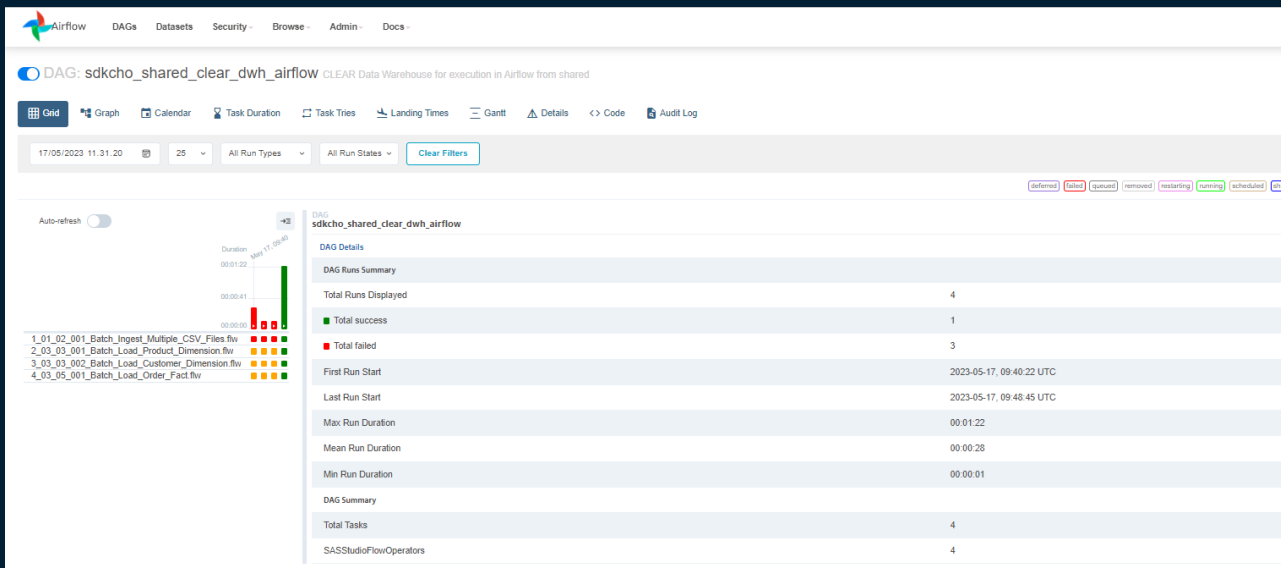
```

Genereret DAG-kode

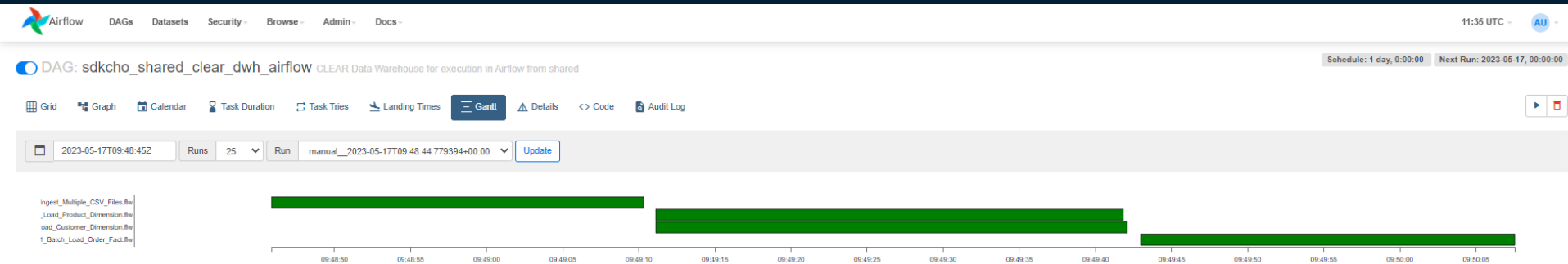
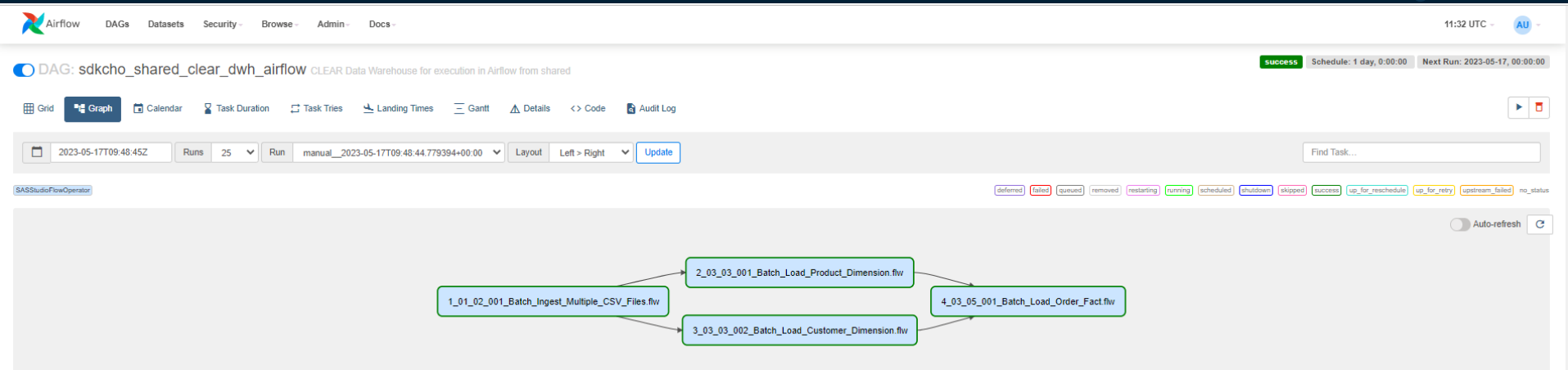
En liste af DAGS
med præfixs sdkcho



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



Grafisk & Gantt overblik over tasks i en DAG





DAG: sdkcho_shared_clear_dwh_airflow CLEAR Data Warehouse for execution in Airflow from shared

Grid Graph Calendar Task Duration Task Times Landing Times Gantt Details <> Code Audit Log

Task Instance: 1_01_02_001_Batch_Ingest_Multiple_CSV_Files.flw at 2023-05-17, 09:48:44

Task Instance Details <> Rendered Template Log XCom

Log by attempts

1

```

*** Log file does not exist: /opt/airflow/logs/dag_id=sdckho_shared_clear_dwh_airflow/run_id=manual_2023-05-17T09:48:44.779394+00:00/task_id=1_01_02_001_Batch_Ingest_Multiple_CSV_Files.flw/attempt=1.log
*** Fetching from: http://airflow-worker-0-airflow-worker.airflow.svc.cluster.local:8793/log/dag_id=sdckho_shared_clear_dwh_airflow/run_id=manual_2023-05-17T09:48:44.779394+00:00/task_id=1_01_02_001_Batch_Ingest_Multiple_CSV_Files.flw/attempt=1.log

[2023-05-17, 09:48:45 UTC] [taskinstance.py:1084] INFO - Dependencies all met for dep_context=non-requeueable deps ti=<taskinstance: sdckho_shared_clear_dwh_airflow.1_01_02_001_Batch_Ingest_Multiple_CSV_Files.flw manual_2023-05-17T09:48:44.779394+00:00 [queued]>
[2023-05-17, 09:48:45 UTC] [taskinstance.py:1084] INFO - Dependencies all met for dep_context=requeueable deps ti=<taskinstance: sdckho_shared_clear_dwh_airflow.1_01_02_001_Batch_Ingest_Multiple_CSV_Files.flw manual_2023-05-17T09:48:44.779394+00:00 [queued]>
[2023-05-17, 09:48:45 UTC] [taskinstance.py:1282] INFO - .....
[2023-05-17, 09:48:45 UTC] [taskinstance.py:1283] INFO - Starting attempt 1 of 1
[2023-05-17, 09:48:45 UTC] [taskinstance.py:1284] INFO - .....
[2023-05-17, 09:48:45 UTC] [taskinstance.py:1303] INFO - Executing <task(SASStudioFlowOperator): 1_01_02_001_Batch_Ingest_Multiple_CSV_Files.flw on 2023-05-17 09:48:44.779394+00:00
[2023-05-17, 09:48:45 UTC] [standard_task_runner.py:55] INFO - Started process 5959 to run task
[2023-05-17, 09:48:45 UTC] [standard_task_runner.py:82] INFO - Running: ['airflow', 'tasks', 'run', 'sdckho_shared_clear_dwh_airflow', '1_01_02_001_Batch_Ingest_Multiple_CSV_Files.flw', 'manual_2023-05-17T09:48:44.779394+00:00', '--job-id', '304', '--raw', '--subdir', 'DAGS_FOLDER/']
[2023-05-17, 09:48:45 UTC] [standard_task_runner.py:83] INFO - Job 304: Subtask 1_01_02_001_Batch_Ingest_Multiple_CSV_Files.flw
[2023-05-17, 09:48:45 UTC] [task_command.py:388] INFO - Running <taskinstance: sdckho_shared_clear_dwh_airflow.1_01_02_001_Batch_Ingest_Multiple_CSV_Files.flw manual_2023-05-17T09:48:44.779394+00:00 [running]> on host airflow-worker-0-airflow-worker.airflow.svc.cluster.local
[2023-05-17, 09:48:46 UTC] [taskinstance.py:1511] INFO - Exporting the following env vars:
AIRFLOW_CTX_DAG_OWNER=airflow
AIRFLOW_CTX_DAG_ID=sdckho_shared_clear_dwh_airflow
AIRFLOW_CTX_TASK_ID=1_01_02_001_Batch_Ingest_Multiple_CSV_Files.flw
AIRFLOW_CTX_EXECUTION_DATE=2023-05-17T09:48:44.779394+00:00
AIRFLOW_CTX_TRY_NUMBER=1
AIRFLOW_CTX_DAG_RUN_ID=manual_2023-05-17T09:48:44.779394+00:00
[2023-05-17, 09:48:46 UTC] [sas_studioflow.py:82] INFO - Authenticate connection
[2023-05-17, 09:48:46 UTC] [base.py:73] INFO - Using connection ID 'sas.default' for task execution.
[2023-05-17, 09:48:46 UTC] [sas.py:56] INFO - Creating session for connection named sas.default to host https://vja-cawki.unx.sas.com
[2023-05-17, 09:48:46 UTC] [sas.py:70] INFO - Get oauth token (see README if this crashes)
[2023-05-17, 09:48:46 UTC] [sas_studioflow.py:86] INFO - Generate code for Studio Flow: /data/home/shared/datasets/clear_dwh/flows/sdkcho/01_02_001_Batch_Ingest_Multiple_CSV_Files.flw
[2023-05-17, 09:48:46 UTC] [logging_mixin.py:137] INFO - Code Generation for Studio Flow with Compute session
[2023-05-17, 09:48:46 UTC] [logging_mixin.py:137] INFO - Create or connect to session
[2023-05-17, 09:48:47 UTC] [logging_mixin.py:137] INFO - Submitted job request with id 7d37bcas-182e-4e86-8033-09d92e6301ef. Waiting for completion
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - Job request has completed execution with the status: completed
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 1 /*****
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 2 * Flow: 01_02_001_Batch_Ingest_Multiple_CSV_Files.flw
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 3 * ID:
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 4 * Created: 2023-05-17T08:37:35.055Z
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 5 * Created by: sdckho
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 6 * Modified: 2023-05-17T08:37:35.055Z
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 7 * Modified by: sdckho
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 8 *
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 9 * Version: DataFlows stable 2023.04 (20230605.1683264675840)
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 10 *
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 11 * Generated On: 2023-05-17T09:48:46.650125087Z
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 12 * Generated by: demol
[2023-05-17, 09:49:10 UTC] [logging_mixin.py:137] INFO - source: 13 *****/

```

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

The screenshot shows the SAS Studio interface with a flow diagram in the center. The flow starts with three data sources: 'ORIGNSOURCEDATAM01', 'SCD_CUSTDIM', and 'SCD_PRODDIM'. Each source feeds into a 'Current Cust' or 'Current Prod' node, which then connects to a 'Lookup' node. The 'Lookup' nodes feed into a 'Load Table' node. The 'Deploy as a job' menu is open, showing options like 'Deploy as a job', 'Redeploy jobs', 'Run now', 'Schedule', and 'Create job definition'. The 'Deploy as a job' menu is highlighted in yellow.

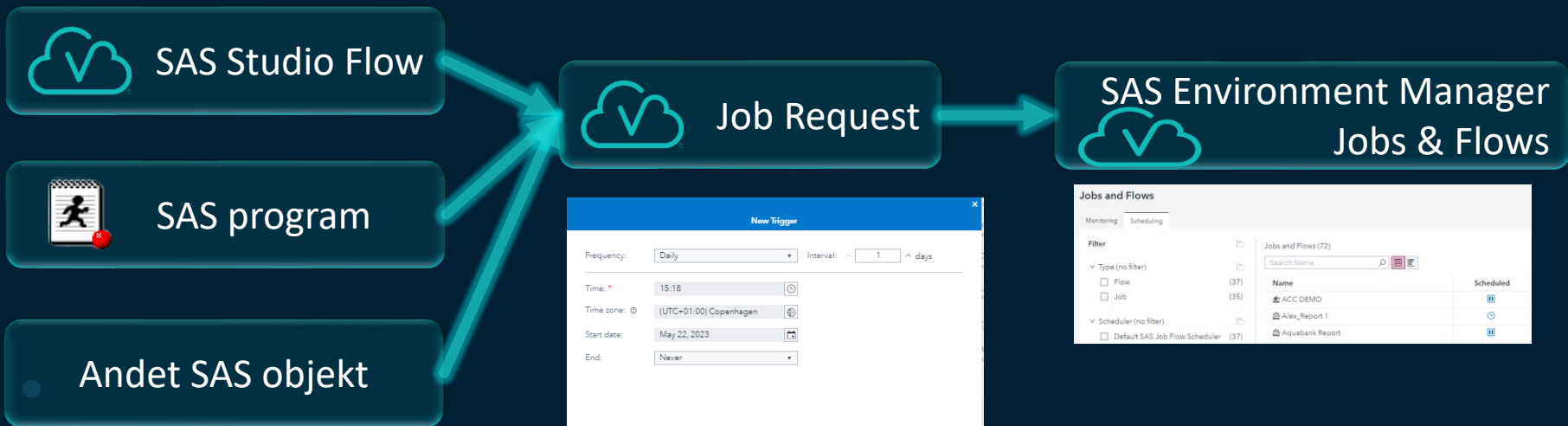
- Man kan deploye et SAS Studio flows, et SAS program m.fl.
- Man kan redeploy et job
 - Ved redeploy opdateres indhold og UID ændres ikke.
- Man kan skedulere de enkelte deployed jobs.
- Man kan skedulere deployed jobs på ny.

The screenshot shows the 'Deployed and Scheduled Jobs' table in SAS Studio. The table has columns for 'Name' and 'Scheduled'. The table contains several rows of job names. A context menu is open over the table, showing options like 'Redeploy', 'Run now', 'Schedule', 'Remove from schedule', 'Activate', 'Delete job', 'Copy link to Clipboard', and 'Clear all history'. The 'Redeploy' option is highlighted in yellow.

Name	Scheduled
01_02_001_Batch_Ingest_Multiple_CSV_Files	
03_03_001_Batch_Load_Product_Dimension	
03_03_002_Batch_Load_Customer_Dimension	
03_05_001_Batch_Load_Order_Fact	

Dette sker, når man skedulerer i SAS Studio

Du får et jobrequest objekt, som bliver skeduleret til kørsel i SAS Viya's Job Scheduler



New Trigger

Frequency: Interval: ^ days

Time: *

Time zone:

Start date:

End:

Jobs and Flows

Monitoring | Scheduling

Filter

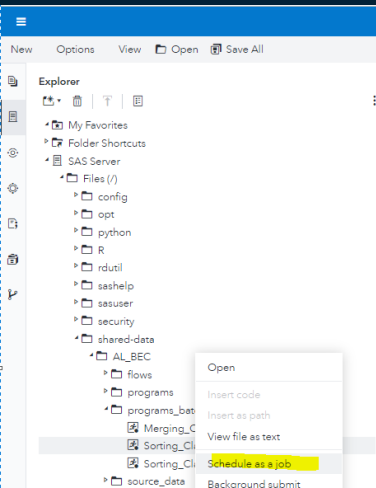
- Type (no filter)
- Flow (37)
- Job (35)
- Scheduler (no filter)
- Default SAS Job Flow Scheduler (37)

Jobs and Flows (72)

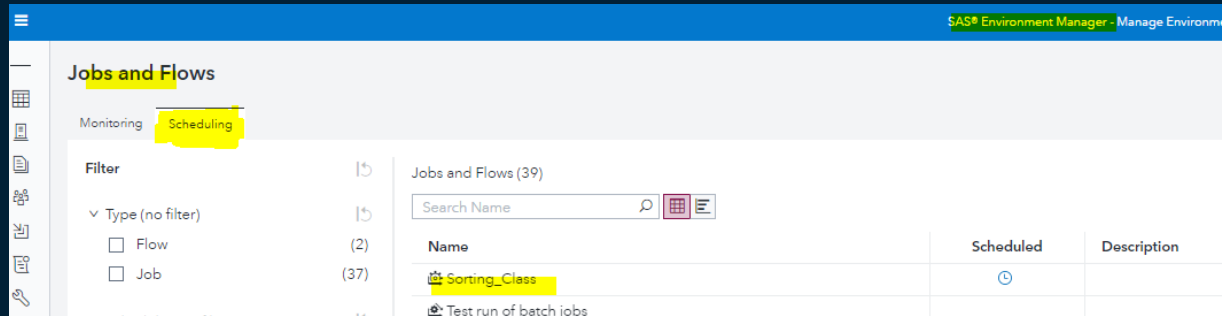
Search Name

Name	Scheduled
ACC DEMO	<input type="button" value="🔍"/>
Alex_Report 1	<input type="button" value="⌚"/>
Aquabank Report	<input type="button" value="🔍"/>

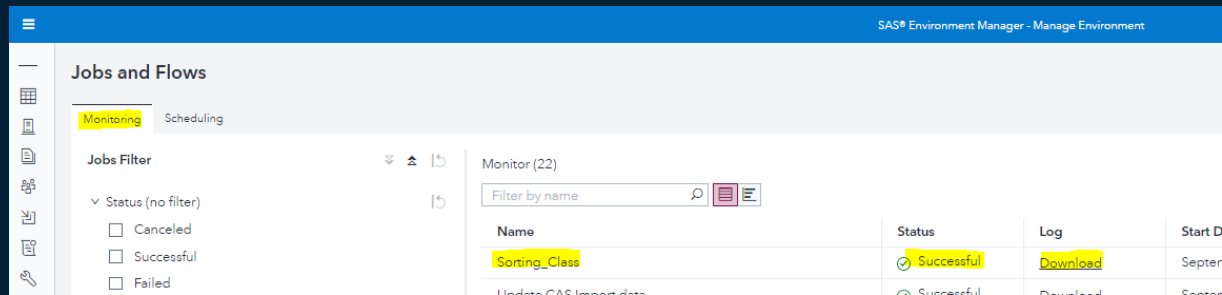
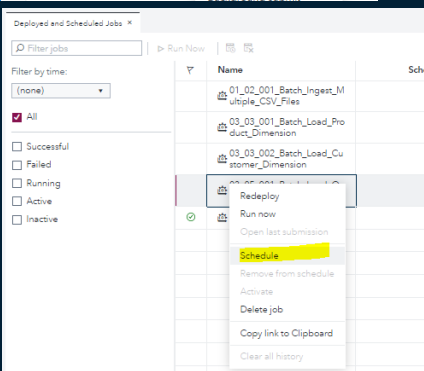
Dette sker, når man skedulerer i SAS Studio



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



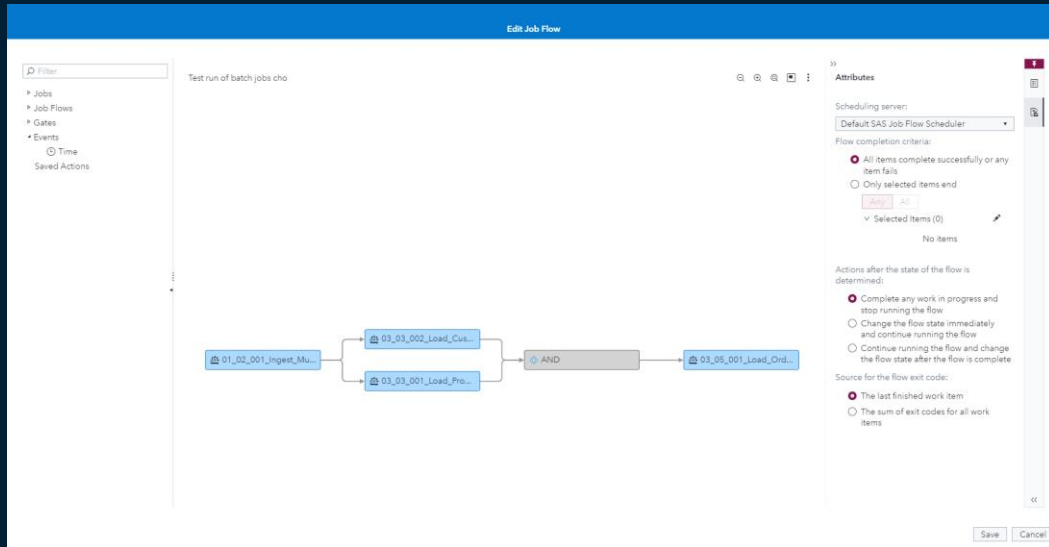
3: I SAS Environment Manager kan du overvåge jobs.



1: I SAS Studio kan du skedulere fra en menu

Byg jobflow med afhængigheder i SAS Environment Mgr

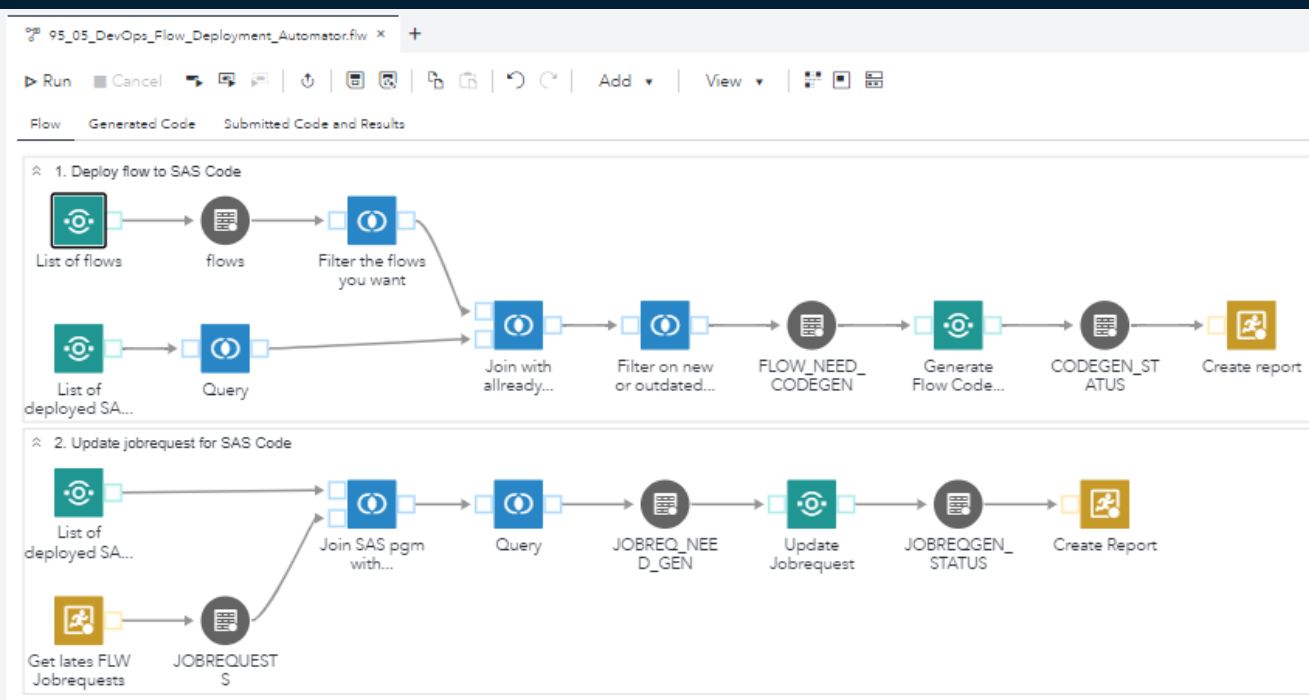
Jobs and Flows menuen



Visual 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.



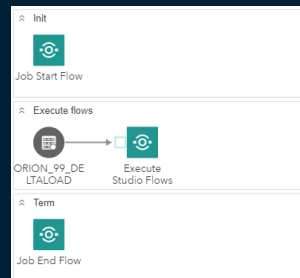
Deploy .flw til .sas
(med custom steps)

Dan jobrequest
(til brug ved skedulering)

DevOps – orkestrere & skedulere i batch

```
data control.ORION_99_DeltaLoad;
  attrlib
    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



Orion_99_DeltaLoad.flw

- Execute
- Run as
- Schedule
- Unschedule
- Execution history
- Copy
- Delete
- Properties

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

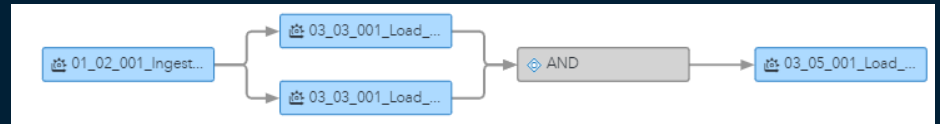
```
data control.ORION_99_DeltaLoad;
  attrib
    flowname informat=$200. length=$200
    async informat=$3. length=$3
  ;
  infile cards4 dsd missover dlm=' ';
  input flowname async;
cards4;
01_02_001_Ingest_Multiple_CSV_Files_SL.flw;NO
03_03_001_Load_Product_Dimension_SL.flw;YES
03_03_002_Load_Customer_Dimension_SL.flw;YES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

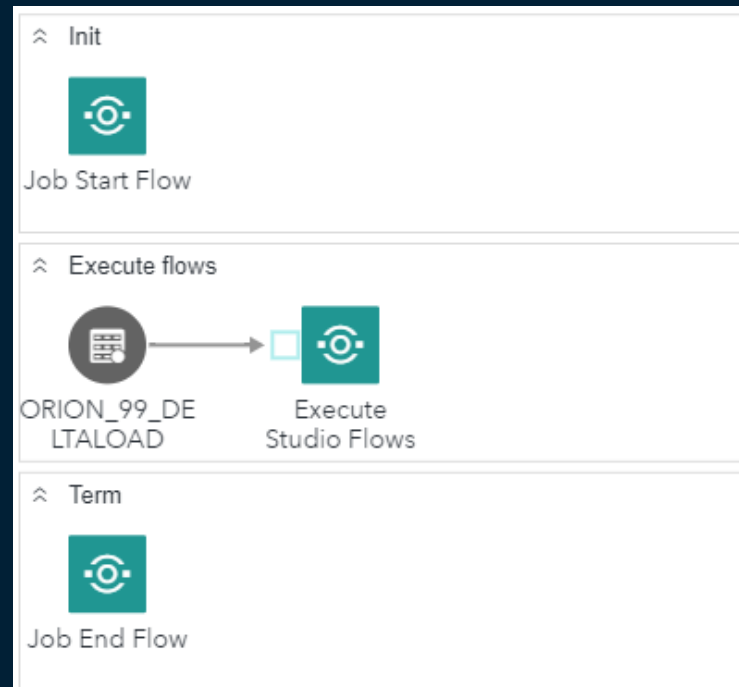
ORION_99_DELTALOAD		
	flowname	async
1	01_02_001_Ingest_Multiple_CSV_Files_SL.flw	NO
2	03_03_001_Load_Product_Dimension_SL.flw	YES
3	03_03_002_Load_Customer_Dimension_SL.flw	YES
4	03_05_001_Load_Order_Fact_SL.flw	NO

=



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

