

Parâmetros de uma Dag



Parâmetro	Função
dag_id	um identificador único para a DAG na cluster
description	descrição da DAG
schedule_interval	o intervalo de tempo no qual a DAG será executada.
start_date	a data e hora em que a DAG deve começar a ser executada
end_date	a data e hora em que a DAG não deve mais ser executada
catchup	determina se a DAG deve executar todas as tarefas que deveriam ter sido executadas desde a data de start_date até o momento atual. Padrão é true
default_view	visualização padrão da interface do Airflow para esta DAG. Por exemplo, “graph”
max_active_runs	máximo de execuções ativas da DAG permitidas


Parâmetros de uma Dag II



Parâmetro	Função
concurrency	máximo de tarefas que podem ser executadas simultaneamente
tags	uma lista de tags para marcar a DAG e suas execuções
default_args	dicionário de argumentos padrão que serão aplicados a todas as tarefas da DAG, a menos que sejam especificamente substituídos
depends_on_past	só inicia se no passado executou com sucesso. Padrão true
email	endereço de email para failure ou retry
email_on_failure	True /False
email_on_retry	True /False
retries	número padrão de novas tentativas
retry_delay	timedelta(minutes=5)

Definidos em default_args

- Alguns destes parâmetros não podem ser definidos diretamente na DAG, mas default_args
- Parametros podem ser sobreescritos diretamente na task



```
default_args = {
    'owner': 'airflow',
    'depends_on_past': False,
    'start_date': datetime(2023, 3, 5),
    'email': ['teste@teste.com'],
    'email_on_failure': False,
    'email_on_retry': False,
    'retries': 1,
    'retry_delay': timedelta(minutes=5),
}

dag = DAG('exemplo_dag_par', description='Exemplo de DAG',
          default_args=default_args,
          schedule_interval="@hourly",
          start_date=datetime(2023, 3, 5), catchup=False,
          default_view="graph",
          tags=["processo", "tag", "pipeline"]
)
```

schedule_interval : Expressão cron

- Minuto, Hora, dia do mês, mês, dia da semana
- 0 0 * * 1 00:00 de segunda feira
- 0 0 5 * * 20:00 do dia 5 do mês



schedule_interval

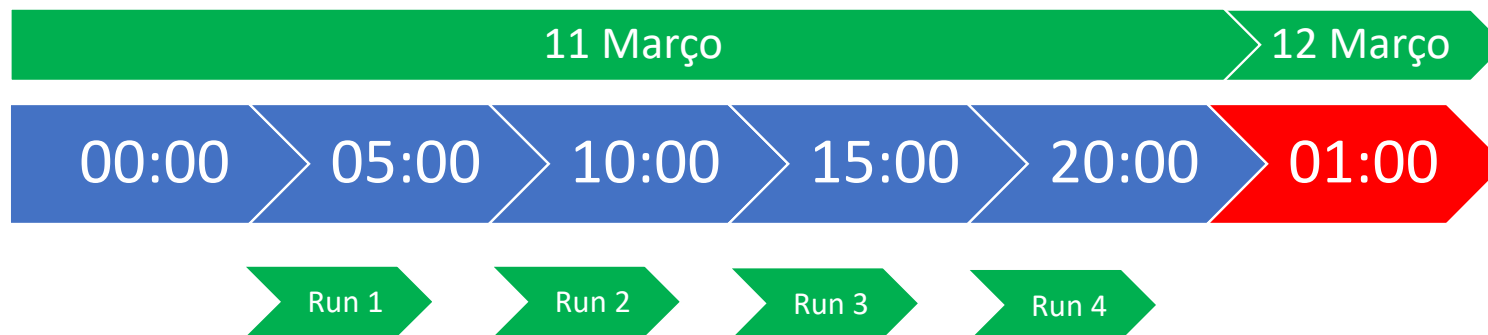
- None
- @once
- @yearly
- @annually
- @monthly
- @weekly
- @daily
- @hourly



schedule_interval: 5 horas

start_date: 11 de março de 2023 às 00:00

end_date: 12 de março de 2023 às 00:00



Catchup e Backfill

- Roda intervalos não executados se definido como verdadeiro
- Da start_date até agora
- Se definido como falso não acontece
- Backfill permite executar intervalos anteriores (airflow backfill)

