

```
# The DAG object; we'll need this to instantiate a DAG
from airflow import DAG

# Operators; we need this to write tasks!
from airflow.operators.bash_operator import BashOperator

from airflow.utils.dates import days_ago

#defining DAG arguments
# You can override them on a per-task basis during operator initialization
default_args = {
    'owner': 'Sagar S Patil',
    'start_date': days_ago(0),
    'email': ['sagarspatil@somemail.com'],
    'email_on_failure': True,
    'email_on_retry': True,
    'retries': 1,
    'retry_delay': timedelta(minutes=5),
}
```

```
# defining the DAG
```

```
dag = DAG(  
    'ETL_toll_data',  
    default_args=default_args,  
    description='Apache Airflow Final Assignment',  
    schedule_interval=timedelta(days=1),  
)
```

```
# define the task 'extract_data_from_csv'
```

```
extract_data_from_csv = BashOperator(  
    task_id='extract',  
    bash_command='cut -f1,2,3,4 -d"," /home/project/airflow/dags/finalassignment/staging/vehicle-data.csv > /home/project/airflow/dags/finalassignment/staging/csv_data.csv',  
    dag=dag,  
)
```

```
extract_data_from_tsv = BashOperator(  
    task_id='extract_data_from_tsv',  
    bash_command='cut -f5,6,7 /home/project/airflow/dags/finalassignment/staging/tollplaza-data.tsv > /home/project/airflow/dags/finalassignment/staging/tsv_data.csv',  
    dag=dag,  
)
```

```
extract_data_from_fixed_width = BashOperator(  
    task_id='extract_data_from_fixed_width',  
    bash_command='cut -c59-61,62-67 /home/project/airflow/dags/finalassignment/staging/payment-data.txt > /home/project/airflow/dags/finalassignment/staging/fixed_width_data.csv',  
    dag=dag,  
)
```

```
consolidate_data = BashOperator(  
    task_id='consolidate_data',  
    bash_command='paste -d "," csv_data.csv tsv_data.csv fixed_width_data.csv | cut -d "," -f 1,2,3,4,5,6,7,8,9 > /home/project/airflow/dags/finalassignment/staging/extracted_data.csv',  
    dag=dag,  
)
```

```
unzip_data >> extract_data_from_csv
unzip_data >> extract_data_from_tsv
unzip_data >> extract_data_from_fixed_width
extract_data_from_csv >> consolidate_data
extract_data_from_tsv >> consolidate_data
extract_data_from_fixed_width >> consolidate_data
consolidate_data >> transform_data
```

---

```
theia@theiadocker-spatil59:/home/project$ cp ETL_DAG.py $AIRFLOW_HOME/dags
theia@theiadocker-spatil59:/home/project$
```



Search

Actions

		State	Dag Id	Task Id	Run Id	Map Index	Logical Date	Operator	Start Date	End Date	Duration	Job Id	
<input type="checkbox"/>	<div><div></div><div></div></div>	success	ETL_toll_data	unzip_data	manual__2023-03-16T04:24:14.322619+00:00		2023-03-15, 23:24:14	BashOperator	2023-03-15, 23:24:16	2023-03-15, 23:24:16	<1s	6	6
<input type="checkbox"/>	<div><div></div><div></div></div>	success	ETL_toll_data	extract_data_from_tsv	manual__2023-03-16T04:24:14.322619+00:00		2023-03-15, 23:24:14	BashOperator	2023-03-15, 23:24:20	2023-03-15, 23:24:23	2s	7	6
<input type="checkbox"/>	<div><div></div><div></div></div>	success	ETL_toll_data	extract_data_from_fixed_width	manual__2023-03-16T04:24:14.322619+00:00		2023-03-15, 23:24:14	BashOperator	2023-03-15, 23:24:21	2023-03-15, 23:24:23	2s	8	6
<input type="checkbox"/>	<div><div></div><div></div></div>	success	ETL_toll_data	extract_data_from_csv	manual__2023-03-16T04:24:14.322619+00:00		2023-03-15, 23:24:14	BashOperator	2023-03-15, 23:24:21	2023-03-15, 23:24:24	3s	9	6
<input type="checkbox"/>	<div><div></div><div></div></div>	success	ETL_toll_data	consolidate_data	manual__2023-03-16T04:24:14.322619+00:00		2023-03-15, 23:24:14	BashOperator	2023-03-15, 23:24:26	2023-03-15, 23:24:27	<1s	10	6
<input type="checkbox"/>	<div><div></div><div></div></div>	success	ETL_toll_data	transform_data	manual__2023-03-16T04:24:14.322619+00:00		2023-03-15, 23:24:14	BashOperator	2023-03-15, 23:24:30	2023-03-15, 23:24:31	<1s	11	6