

Productionalizing Data Pipelines with Apache Airflow

INTRODUCING APACHE AIRFLOW



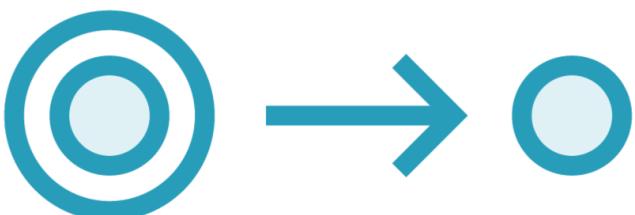
Axel Sirota



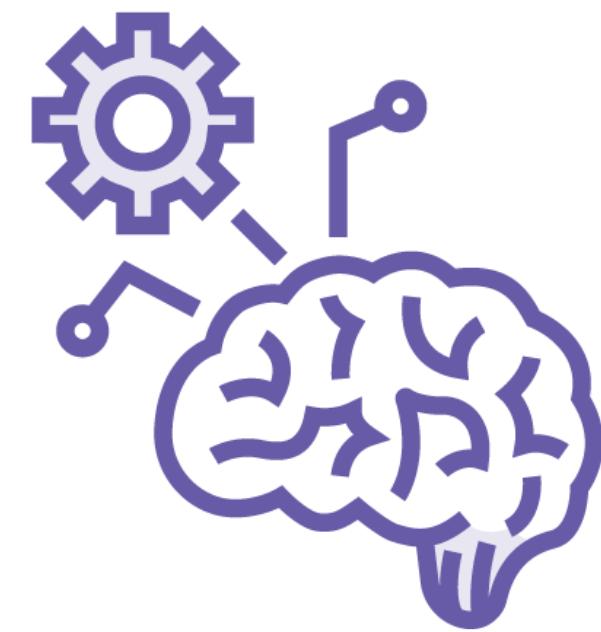
Services



Event Data



Transform



Machine
Learning



What Is Apache Airflow?

- Develop and schedule workflows
- Workflows as Code
- Workflows interact nicely with each other

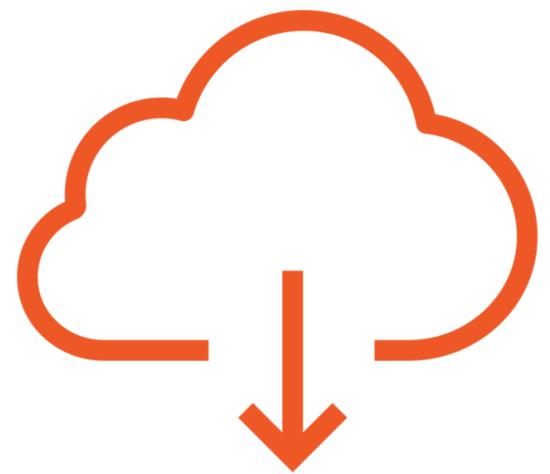


An Example Workflow

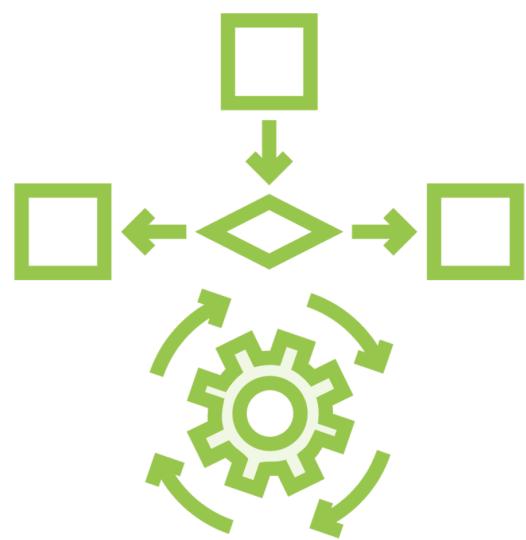


**Download
invoices**

An Example Workflow

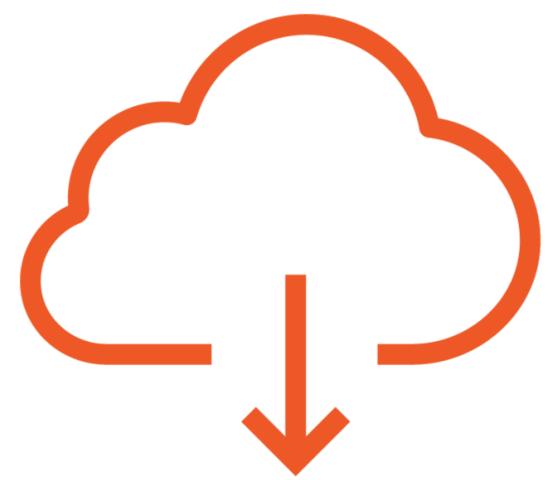


**Download
invoices**

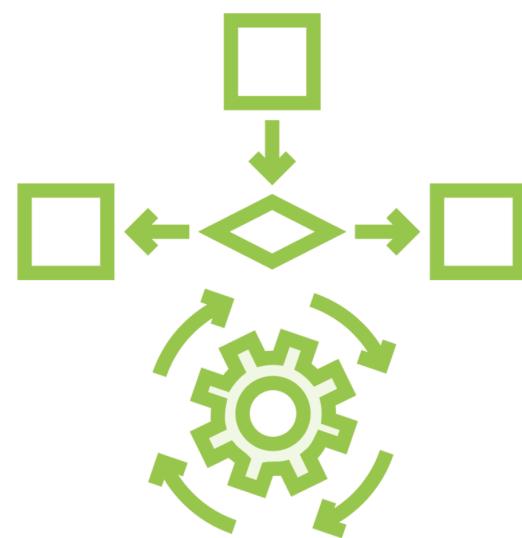


Transform

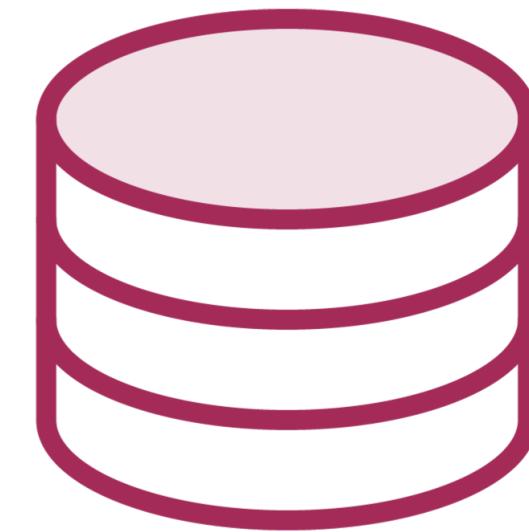
An Example Workflow



**Download
invoices**



Transform

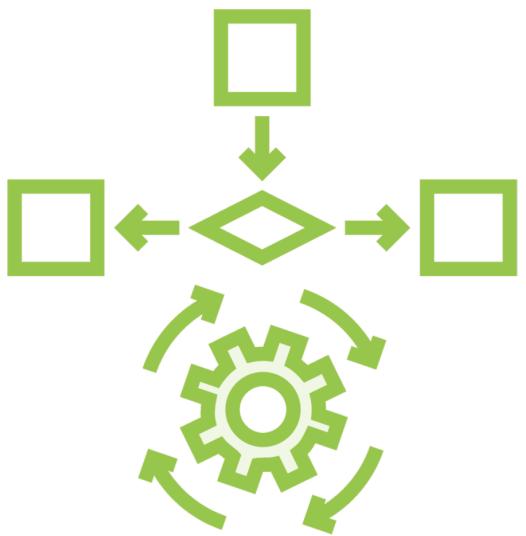


Save

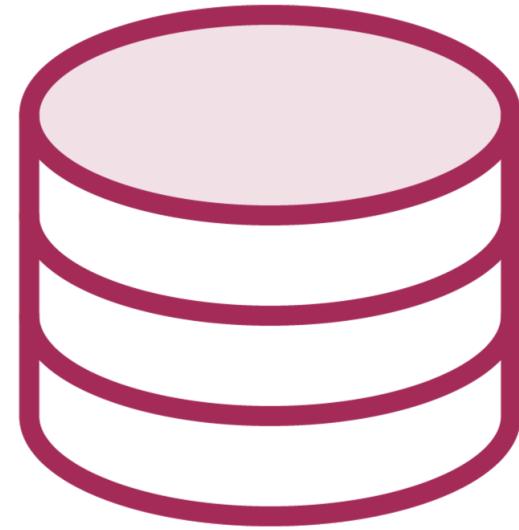
An Example Workflow



**Download
invoices**



Transform

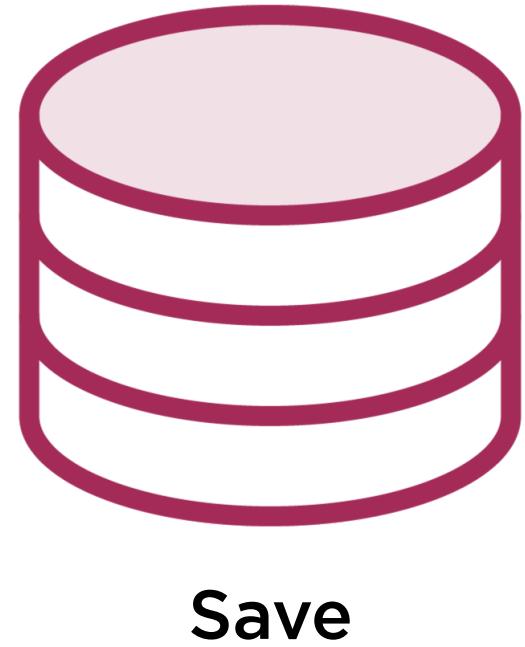
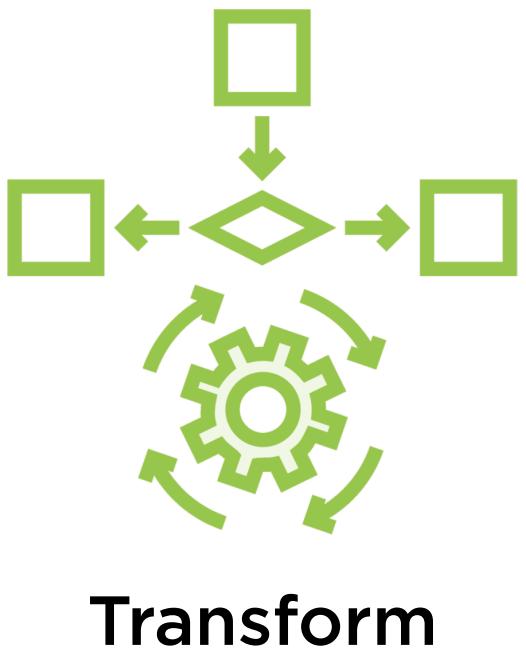
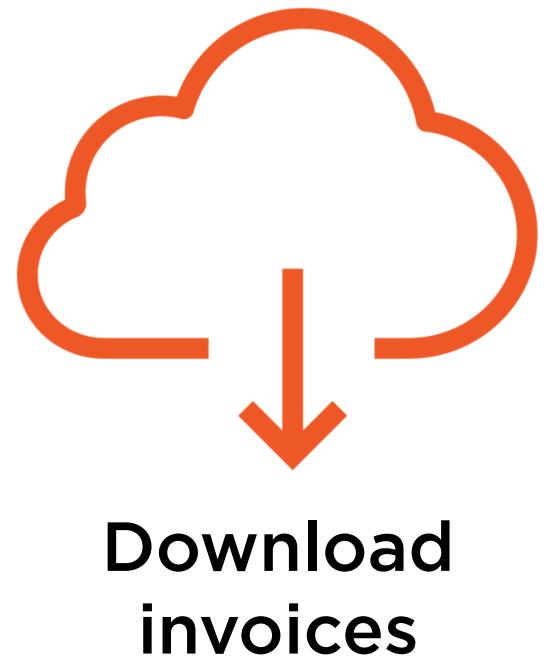


Save

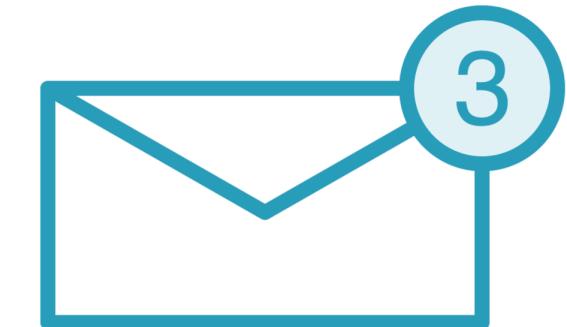
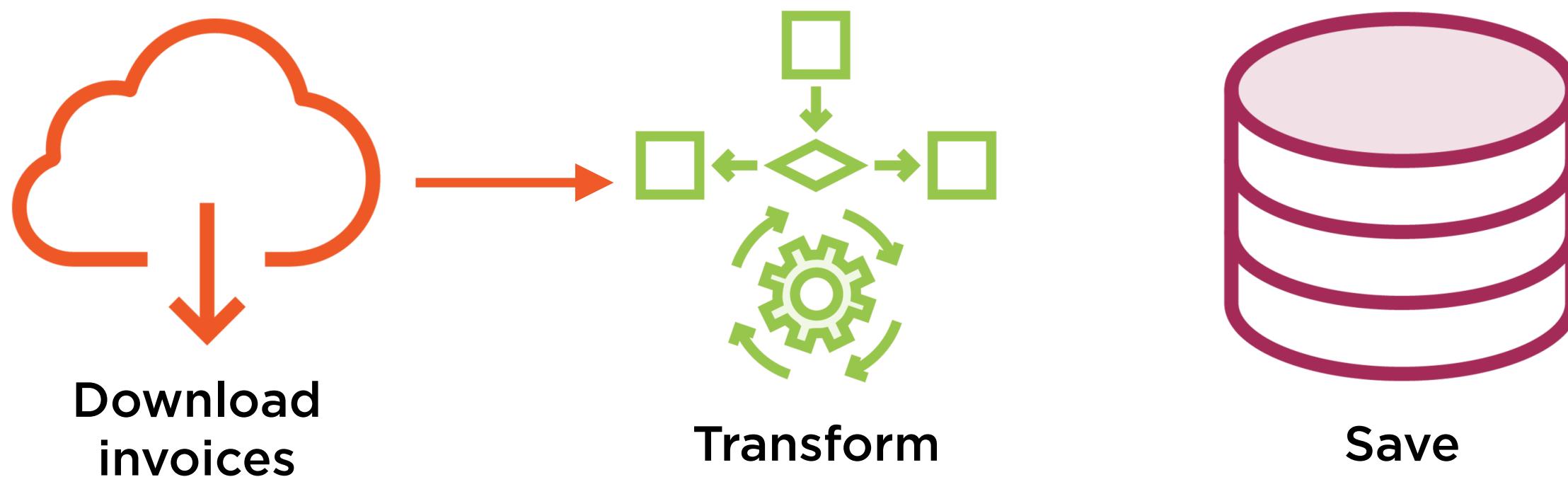


Report

An Example Workflow



An Example Workflow

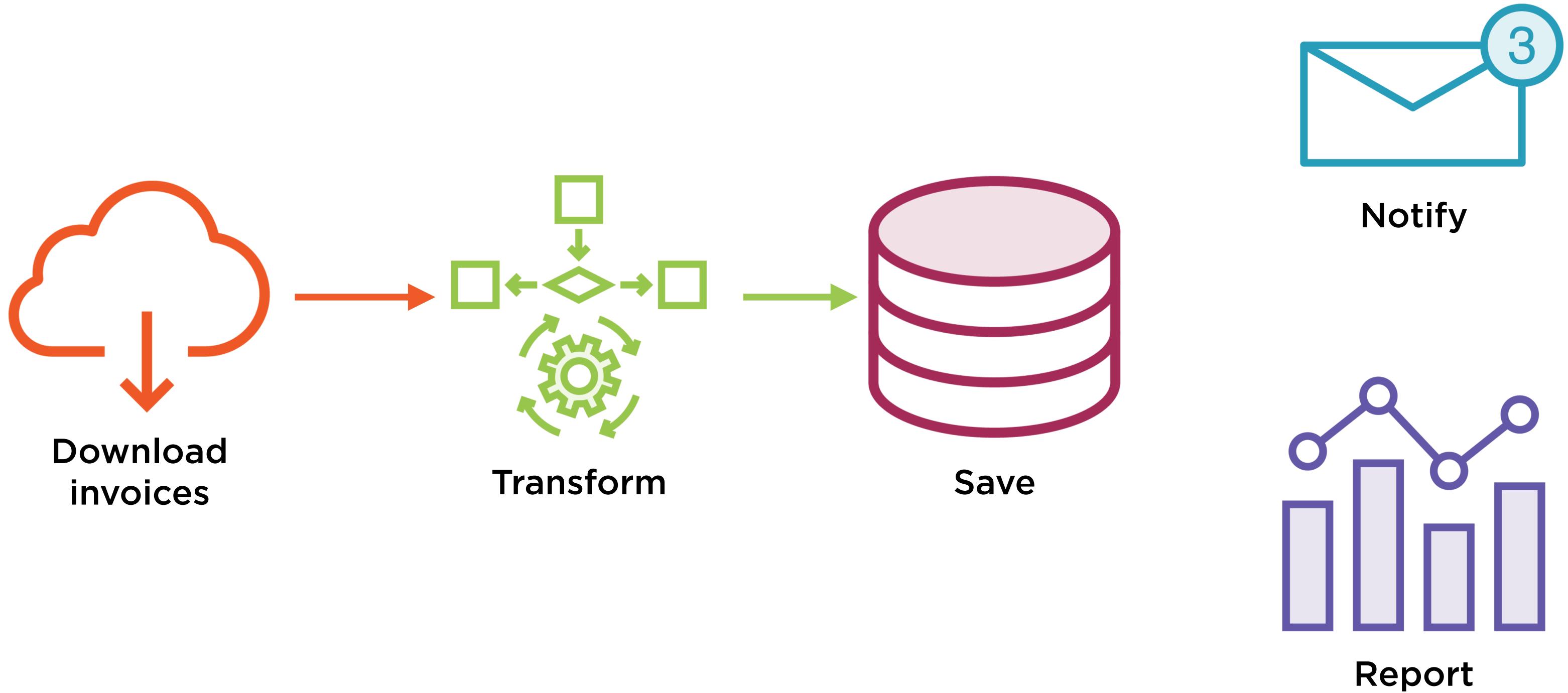


Notify

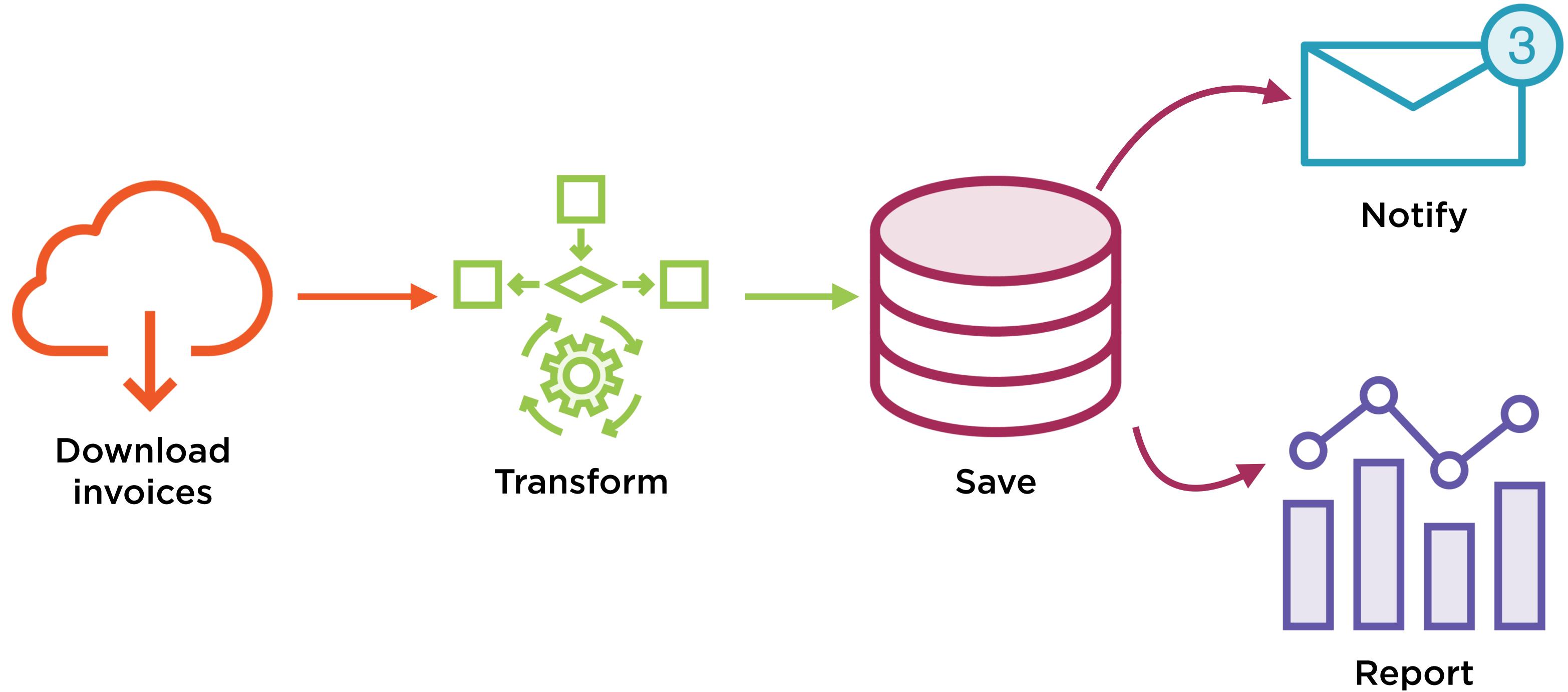


Report

An Example Workflow



An Example Workflow



Benefits Of Airflow



Extensibility

Benefits Of Airflow



Extensibility



Reliability

Benefits of Airflow



Extensibility



Reliability



Scalable



Getting the Best out of This Course

Required

- Proficiency in a programming language
- ETL
- Statistics
- Docker

Required

- Proficiency in a programming language
- ETL
- Statistics
- Docker

Course Suggestions

- Interpreting Data With Statistical Models
- Building Data Pipelines with Luigi and Python
- Core Python: Getting Started
- Docker: The Big Picture

Required

- Proficiency in a programming language
- ETL
- Statistics
- Docker

Course Suggestions

- Interpreting Data With Statistical Models
- Building Data Pipelines with Luigi and Python
- Core Python: Getting Started
- Docker: The Big Picture

Not Needed

- Apache Airflow

More Resources

[Table of contents](#)[Description](#)[Transcript](#)[Exercise files](#)[Discussion](#)[Learning Check](#)[Related Courses](#)



Ask me on Twitter

@AxelSirota

GitHub Repository for the Course

 [axel-sirota / productionalizing-data-pipelines-airflow](https://github.com/axel-sirota/productionalizing-data-pipelines-airflow) Unwatch 1 Star 0 Fork 0

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main 1 branch 3 tags Go to file Add file Code

Commit	Message	Time
 axel-jampp	Add demos	07cae26 20 hours ago
 module3-demo1	module3-demo1	4 days ago
 module3-demo2	Add demos	20 hours ago
 module3-demo3	Add demos	20 hours ago
 .gitignore	first working version	10 days ago
 LICENSE	Initial commit	11 days ago
 README.md	first working version	10 days ago
 docker-compose-CeleryExecutor.yml	first working version	10 days ago

About Productionalizing Data Pipelines with Apache Airflow Readme MIT License

Releases 3 tags Create a new release

<https://github.com/axel-sirota/productionalizing-data-pipelines-airflow>

A Production Grade Pipeline in Airflow

Off DAG: invoices_dag

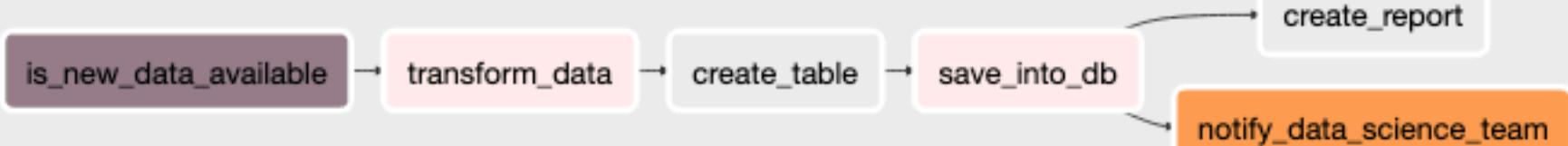
schedule: @daily

Graph View Tree View Task Duration Task Tries Landing Times Gantt Details Code Trigger DAG Refresh Delete

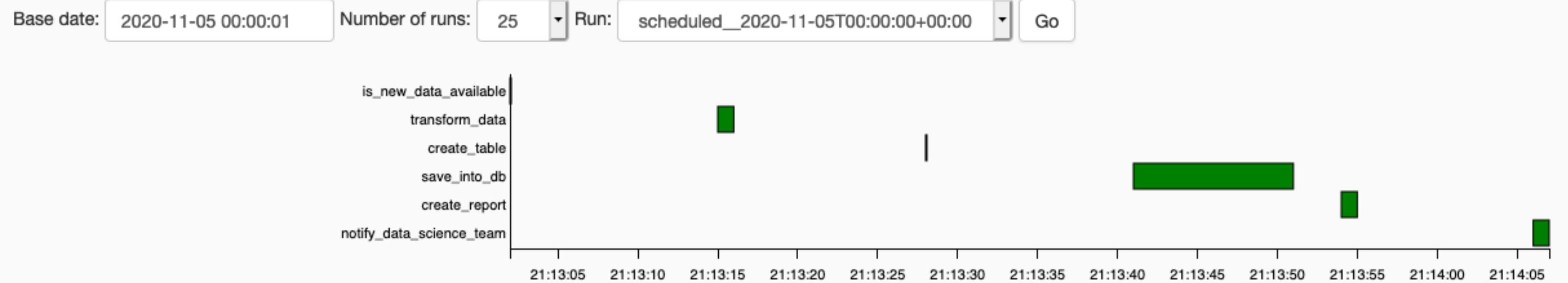
None Base date: 2020-11-05 23:52:26 Number of runs: 25 Run: Layout: Left->Right Go Search for...

FileSensor PostgresOperator PythonOperator SlackWebhookOperator

success running failed skipped upstream_failed up_for_reschedule up_for_retry queued no_status



 Graph View  Tree View  Task Duration  Task Tries  Landing Times  Gantt  Details  Code  Trigger DAG  Refresh  Delete



Outline of the Course

Fundamentals

Architecture and core concepts

Decouple Pipelines

SubDAGs,
Branching, and
XComs

Final Thoughts

How does ML fit?

Scheduling

Backfill and Catchup



Scaling Airflow

LocalExecutor and
CeleryExecutor

More Resources

Where to continue
your journey?