

## Do you want to automate infrastructure or workflow?



Infrastructure  
Automation

### Deployment Manager

- Terraform
- Ansible
- Chef
- Puppet



Workflow  
Orchestration

### Cloud Composer

- Apache Airflow



Infrastructure automation tools like Deployment Manager make creating cloud infrastructure manageable and repeatable and provide a method of documenting the infrastructure which is a common business requirement.

Workflow orchestration is the idea that most business applications rely on passing data from one area of work concentration to another. For example, in the supply chain raw materials are ordered. In the factory, the materials are manufactured into products. At the loading dock the products are shipped. And in the store the products are sold. Each of these parts of the business have systems associated with them: resource management, manufacturing control, shipping and receiving, sales. And those systems need to share information with one another to operate the business. The overall solution is more about workflow between these systems than about the infrastructure of any single system.

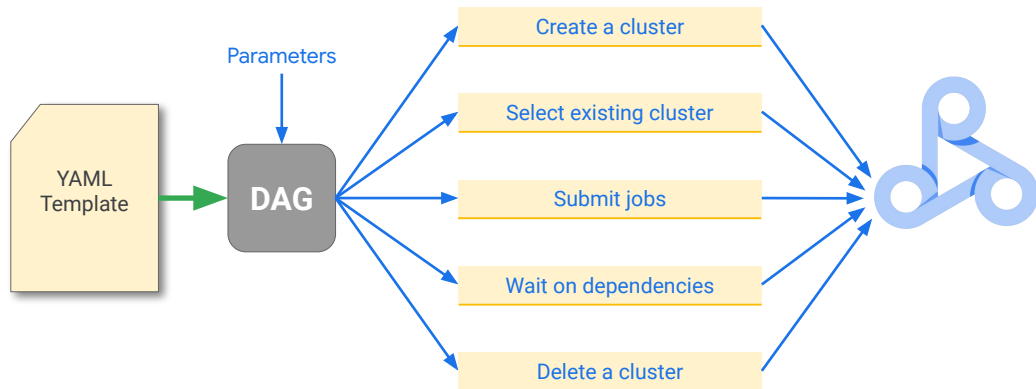
If you start to consider workflow in your overall design, you will start to take control over the overall solution and the reliability and methods of operations of the total system.

Q: Why should you have a conductor for your orchestra?

A: To ensure solution and operations reliability.

# Create data infrastructure when the workflow requires it

Dataprocc Workflow Template



A Dataprocc Workflow Template is a simple example of the crossover between workflow and infrastructure. In this case, the workflow (scheduled or on demand) causes the service to create a Dataprocc cluster to process data, and then deletes the cluster when the work is done. **Infrastructure becomes dynamic in a workflow context.**

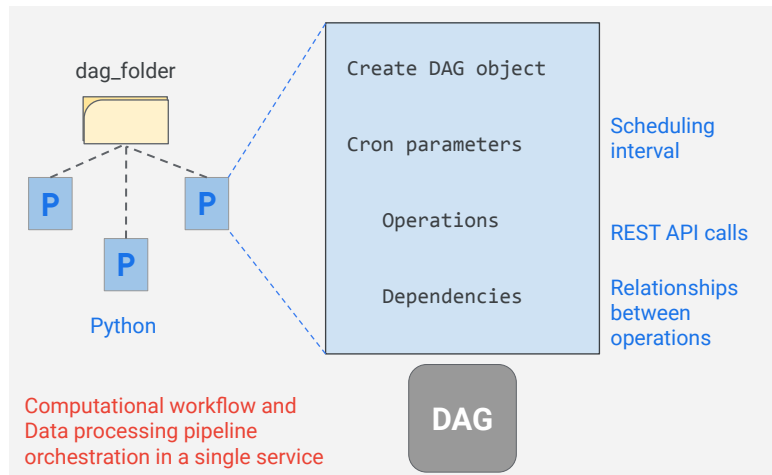
Details:

The Dataprocc Workflow Template is a YAML file that is processed through a Directed Acyclic Graph (DAG). It can create a new cluster, select from an existing cluster, submit jobs, hold jobs for submission until dependencies can complete, and it can delete a cluster when the job is done.

It is currently available through the `gcloud` command and the REST API, but not through Console.

The Workflow Template becomes active when it is instantiated into the DAG. The Template can be submitted multiple times with different parameter values. You can also write a template inline in the `gcloud` command, and you can list workflows and workflow metadata to help diagnose issues.

# Cloud Composer: Extensible workflow orchestration



Cloud Composer, based on Apache Airflow, is a service that provides extensible dependency management for complex workflows. Because the Directed Acyclic Graphs (DAGs) are written in Python, Cloud Composer can be extended to coordinate and orchestrate anything with a Python-compatible API -- *which are most services today*.

In Airflow/Composer, there can be multiple DAGs, and they are defined in Python. Each DAG lives in the dag\_folder.

A few notes about Cloud Composer:

## Workflow orchestration

- Concerned with the instructions necessary to complete each step.
- Computational workflow
- Data processing pipeline
- Dependency management
- Extensible operators
  - operations --> REST APIs