Airflow Plugins:

Airflow is built with the intention of allowing its users to extend and customize its functionality through plugins. The most common types of user-created plugins for Airflow are Operators and Hooks. These plugins make DAGs reusable and simpler to maintain.

To create a custom operator:

1. Identify Operators that perform similar functions can be consolidated.
2. Define a new Operator in the plugins folder
3. Replace the original Operators with your new custom one, re-parameterize, and instantiate them.

TASK Boundaries:

DAG tasks should be designed such that they are:

- Atomic and have a single purpose.
- Maximize Parallelism
- Make Failure States Obvious

Every task in your DAG should only do one job.

Benefits of Task Boundaries:

- Re-visitable: task boundaries are useful for you if you revisit a pipeline you wrote in the past. You'll have much easier time understanding how it works and the lineage of data if the boundaries between tasks are clear and well defined.

- Parallelization: tasks that are atomic are more easily parallelizable.

SubDAGs:

Advantages:

Commonly repeated series of tasks within DAGs can be captured as reusable SubDAGs; benefits include:

- Decrease the amount of code we need to write and maintain to create a new DAG
- Easier to understand the high level goals of the DAG.
- Bug fixes, speedups, and other enhancements can be made more quickly and distributed to all DAGs that use that SubDAG.

Disadvantages:

General issues:

- Limit the visibility within the Airflow UI
- Abstraction makes understanding what the DAG is doing more difficult.
- Encourages premature optimization
- Produces pipeline coupling

Monitoring:

Airflow can surface metrics and emails to help you stay on top of pipeline issues.

SLAs: can specify a time in which a DAG must be completed.

Emails & Alerts: airflow can be configured to send emails; useful for production support.

Metrics: comes with "out of the box" metrics, using a metric aggregator called statsd.