Research MLOps

Leander Oudakker

December 8, 2021

I researched 3 data processing frameworks, I chose Airflow, Luigi and Prefect. Airflow and Luigi are currently the most used frameworks and Prefect is a new framework that claims to use a more modern approach for data processing.

Airflow and Luigi are the most mature frameworks and are highly tested and trustworthy, while Prefect is still a young framework that is still being developed.

Airflow is the most feature-rich framework of the 3, but this comes at the cost of simplicity and away steeper learning curve than Luigi and Prefect. Specifically, airflow is far more powerful when it comes to scheduling. The biggest strength of Luigi is the ease of use and how it makes connecting to a myriad of different filesystems and databases trivial. Prefect has the most fully-featured dashboards, which can be fully customized with an optional orchestration layer.

Luigi and Prefect have a similar set of features and both use a simple approach to building applications. Airflow and Luigi use inheritance while Prefect uses annotations. This means that Airflow and Luigi force you to create your application in a certain way, while Prefect is much more free form.

I think the best choice is Luigi because it has all the necessary features while remaining very simple, mature and tested. It will also make it easy to work with commonly used filesystems and databases. I think it is a better choice than Airflow, because Airflow has a way steeper learning curve, while it doesn't have any functions that we need to use that Luigi doesn't include. Prefect on the other hand has a comparable set of features but is less mature and trustworthy than Luigi.