PREFECT

Clearcover Reduces Time-to-Insights by Orchestrating Workflows with Prefect

ROI

"WE HAVE DEFINITELY GOTTEN
THE RETURN ON OUR INVESTMENT"

Efficiency

"WE SAVED DAYS OVER AIRFLOW"

COMPANY
Clearcover

INDUSTRY
Insurance (Auto)

HEADQUARTERS
Chicago, IL

PRODUCTS USED
PREFECTCORE
PREFECTCLOUD

Synopsis

INTRODUCTION

Clearcover's focus on innovative engineering practices provide smarter car insurance choices to consumers, offering better coverage for less money.

CHALLENGE

Limit 'time-to-insights' for data engineers and analysts through a mixture of custom built tooling and vendor relationships, to ultimately drive revenue generating initiatives.

SOLUTION

Implement Prefect Cloud as a central data platform to orchestrate batch processing workflows and to standardize workflows defined with Prefect Core semantics.

RESULTS

An accessible, self-service platform for data analysts to orchestrate their own workflows and reach valuable business insights faster, without the need for complex knowledge of infrastructure deployments as a prerequisite.

Introduction

As a modern insurance company heavily invested in data engineering, Clearcover leverages machine learning models to provide customers low-rates insurance quotes. As an online-only insurer, the company's innovative technology provides reliable car insurance for the modern driver. In an effort to realize new data practices for the entire data team, Lead Data Engineer at Clearcover Braun Reyes sought out accessible workflow orchestration tools to streamline the development of data applications at scale.

Challenge

Initially, the engineering team at Clearcover found success orchestrating their data pipelines with a combination of AWS Step Functions and AWS ECS Fargate, but quickly encountered the high barrier to entry for non-AWS professionals. This technical complexity, which demanded knowledge of cloud resources and infrastructure tooling like Terraform, only increased with the separation of data processing code from the orchestrator. As Clearcover's analysts became increasingly eager to orchestrate their own workflows, a centralized platform was needed to serve across the data division.

> We were getting more and more requests from Data Analysts across the [organization] that wanted to schedule their workflows. They were moving beyond

writing SQL in Snowflake or running reports in Tableau and starting to use the Python language and its data tooling to do more advanced analysis. " –Braun Reyes, Lead Data Engineer at Clearcover

It was clear that accessibility was a critical success metric for any tool under consideration, to enable data analysts to self-serve using custom configurations specific to Clearcover's daily operations without explicit aid from the engineering team.

❝ *At the time we were also considering Dagster, which was another Airflow–esque offering. However, it just wasn't as accessible as Prefect. I am a big proponent of accessibility when choosing tools. Technology changes way too fast and you need tools that you can get up and running for POC quickly. We found Prefect to be way more accessible than Dagster."* –Braun Reyes, Lead Data Engineer at Clearcover

Upon discovering Prefect from an interview with CEO Jeremiah Lowin, Prefect orchestration seemed to hit Clearcover's requirements: a pythonic orchestration tool for data scientists, data engineers, and ML engineers to centralize their work.

Solution

Initially, Braun Reyes and his team needed to build automation infrastructure to get Prefect production ready, and was able to do so thanks to the object–based design of Prefect Core. The goal was to enable users to self-serve easily, for a stakeholder to be able to write a flow locally, test the flow, and get it to Prefect Cloud quickly.

❝ *This required us getting into the weeds of the internals of Prefect Core and the Fargate Agent, which was fine since we had vested interest in contributing to those parts of the open source project. Once we got over that hurdle, it was pretty easy from there."* –Braun Reyes, Lead Data Engineer at Clearcover

The primary driver behind Clearcover's self-service engine was a custom CI/CD process to empower users without waiting on the data engineering team. Users were given a GitHub repository template with configurations to deploy

their own flows to Prefect Cloud and cutting of a release. Additionally, the team was able to subclass the base Prefect Agent for usage in conjunction with AWS ECS Fargate, which was eventually contributed back to the Prefect library and is available as a built–in integration today.

❝ *We were able to use AWS ECS Fargate as our agent and flow execution environment to limit our total cost of ownership on the infrastructure side. It was a huge success and has really enabled our data user to do much more than just waiting on Data Engineering, which was the ultimate goal."* –Braun Reyes, Lead Data Engineer at Clearcover

The object based design of Prefect Core also allowed the team to onboard new engineers quickly and provide a set of tools connected to Clearcover's internal library of resources written for common data tasks. Mark McDonald, a Data Engineer on Braun's team, subclassed the built–in ShellTask to create the DbtShellTask. This Task (now one of Prefect's most popular library Tasks) allowed Mark to incorporate the modern analytics engine for his team and allow them to easily issue dbt commands as Prefect Tasks. Ultimately, the flexibility of Prefect's task decorator allowed the team to quickly compose tasks and integrate existing data engineering work.

Dask also provided a pivotal role in allowing the team to quickly access parallelism for flows. By leveraging Prefect's task map function, which now supports depth-first execution, the team was able to achieve quick concurrency and parallelism for strenuous flows. With the capabilities of Dask firmly in the team's toolset and a standardized means of orchestrating workflows, the team could easily move to more distributed analysis as their data needs continue to grow.

Results

Prefect Cloud provided an accessible, self–service platform for writing and scheduling batch data work flows written in Python.

The object based design of Prefect Core allowed the team to build a custom CI/CD process for flow creation, reusing configured aspects of Prefect based on the needs of each

data engineer/analyst.

Prefect Core's open-source library can be subclassed for custom configurations, which enabled the Clearcover team to build a specialized FargateAgent and DbtShellTask, both of which are available in the Task Library.

The interactive API allowed the development team to test out different GraphQL requests, ultimately aiding open-source contributions.

Prefect Core workflow semantics quickly enable parallelism with it's built-in integrations with Dask distributed processing.

Prefect provided an entry point into using Dask for distributed processing for team members unfamiliar with concurrency and parallelism.

Data engineers preconfigured templates for critical Dask infrastructure to allow data analysts/ML engineers the freedom to experiment and optimize.

Prefect enables guard rails for data engineers to allow data analysts to write their own batch processing workflows and to produce specialized reports.

Prefect Cloud's feature set automated potentially painful details: the ability to run flows on multiple schedules with different parameters, a daylight savings aware scheduler, cloud hook error notifications, and logs/metrics for troubleshooting.

❝ *Saving us days on DAG design vs. Airflow" – Braun Reyes, Lead Data Engineer at Clearcover*

Clearcover can onboard new data engineers and analysts to create data pipelines with ease thanks to Prefect's task decorator, the simplest way to create Prefect tasks, integrating an internal library of resources for common data needs.

"We context-switched to Prefect, got into Prefect Cloud as early adopters. We have definitely gotten the return on our investment. All aspects of our data organization use Prefect for scheduling data workflows. Prefect is now how we write these data applications and people expect to start there for anything new."

BRAUN REYES, LEAD DATA ENGINEER

The self-service data workflow platform provided data analysts and ML engineers preconfigured templates to deploy their own flows to Prefect Cloud, without the direct help of Clearcover's data engineers.