Summer 2023 Data Engineering Internship at A.M. Best Rating Services

Vikram Rangarajan

CPSP359G

College Park Scholars – Science & Global Change Program
Computer Science - Machine Learning Track
vikramr@terpmail.umd.edu

College Park Scholars Academic Showcase, May 3, 2024

Issues Confronted

The A.M. Best Data Engineering team were tasked with a hybrid cloud migration which would be done over the span of several months to years. I was asked to aid in this and learn many real-world techniques used in the field of data and cloud engineering. Many of the required tasks involved creating data pipelines and managing database schemas.

The team wanted to use Apache Airflow to orchestrate scheduled data pipelines and other tasks. It was my responsibility to learn what Airflow was, how to use it, and the potential applications it had with their workflow.

Activities

I started off this internship having to learn everything there was to data engineering and Apache Airflow, so I spent the first 4 weeks learning SQL, Docker (to install Airflow), and learning how to make simple Airflow DAGs. I presented what I learned to my teammates with a demo which queried the NWS weather API and stored it into a PostgreSQL database on an hourly basis. Reading data from a remote source and storing it all on a schedule is what Airflow is all about, so it served as a good demo.

I was a part of an internship program with 9 other interns, and we participated in many team building activities such as community service (park cleanup), getting ice cream, and being on the A.M. Best private news channel.

Impact

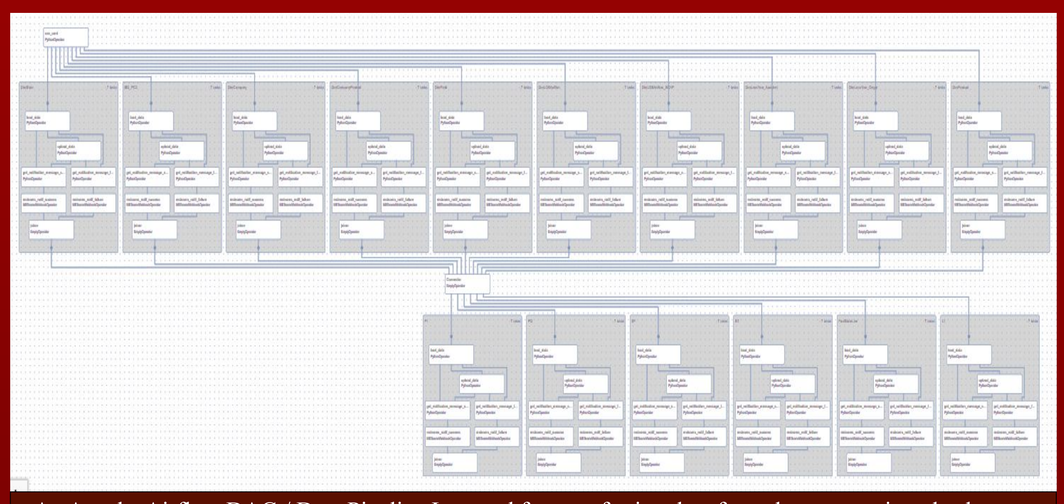
I instructed my team members on how to use specific useful Airflow tools, such as plugins, notifications, and encryption. I also documented all of my work for if the team ever needs it as a reference for a similar task. In fact, the data pipeline shown in the picture is a very general purpose pipeline which can transfer any query from the on-prem database to the datalake, and can be used continuously.

Technologies Learned

In order to thrive in this experience, I had to very quickly learn many different tools in the data engineer's toolkit. These include:

- Python, Pandas, DBAPI, SQLAlchemy
- SQL, Databases
- Linux, Docker, Remote Development, Git
- Azure Cloud Technologies (Data Factory, Data Lake, Devops, etc.)
- Apache Airflow (Data Orchestrator)
- APIs
- Writing efficient, documented code

Me (left) with the 9 other Interns in the program



An Apache Airflow DAG / Data Pipeline I created for transferring data from the on-premises database to the Azure Data Lake. It was created dynamically, with each one of the gray blocks being a template, and the block being repeated.

Site Information:

A.M. Best Rating Services (https://web.ambest.com/home)

1 Ambest Rd, Oldwick, NJ 08858

Supervisors: Rohit Motiani & Matthew Coppola

Company Mission: Provide rating services for insurance companies worldwide

Goals of Data Engineering Team: Manage the flow of data in order to do a hybrid cloud migration

Future Work

After my internship in the summer of 2023, I was asked by my manager, Rohit, if I wanted to work part-time at A.M. Best again, but as a contractor. After a few technical struggles due to the fact that an intern working part time right after the internship hadn't been done before, I started working during the fall semester (October 2023) and only then was I able to really get hands on experience with Azure Cloud services. I ended my contract in January 2024 due to scheduling conflicts, but if the situation allows, I would be greatful to work there again!

Acknowledgments:

Rohit Motiani, Matthew Coppola: Excellent managers, guided me to success

A.M. Best Data Engineering Team: Assisted and taught me good business practices, teamwork, and many tools

Dr. Holtz, Dr. Merck: Directors of SGC, gave us a thorough and fun education on the topic of climate change and human impacts on the world