Jigsaw Labs

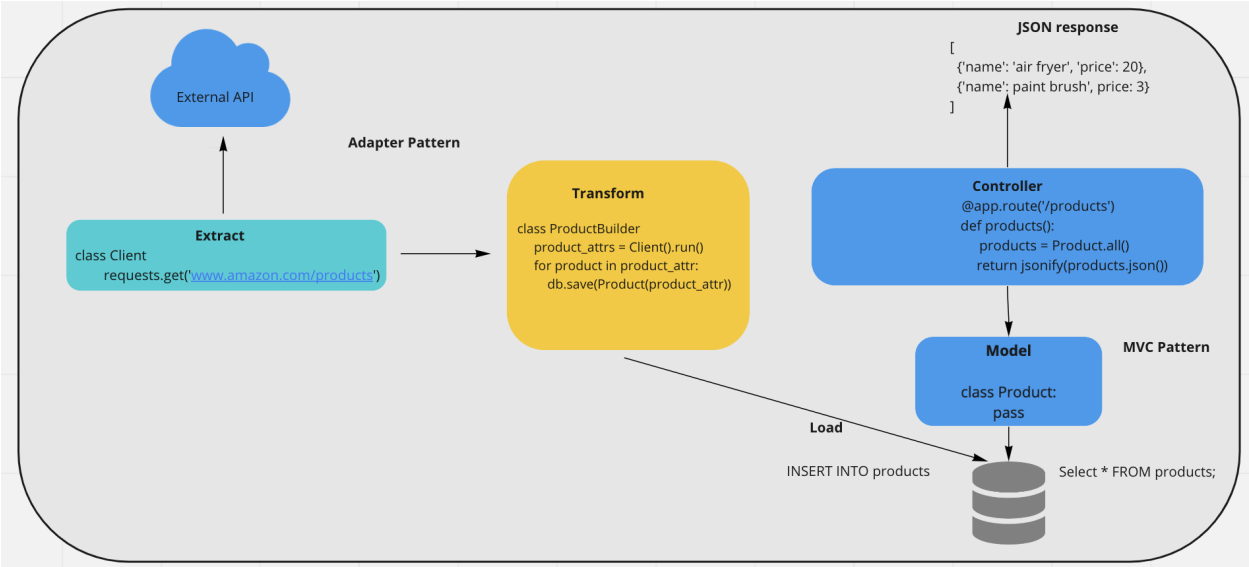
Data Engineering Career Course Syllabus

Module 1 - Backend Engineering & Analytics

This module prepares students to navigate and build large codebases in Python and SQL. It trains students on three main subjects: (1) Software Engineering fundamentals and (2) Object Oriented Design Patterns and (3) Data Analytics.

Software Engineering Fundamentals

Here, students learn to retrieve and manipulate data in Python, focusing on writing clean functions to retrieve and coerce the data. We'll also cover SQL, where we will learn both single table queries, and relational queries, and practice data modeling in a normalized database.



Object Oriented Design Patterns

We then focus on object oriented programming in Python, and related design patterns. Here, students learn about the MVC design pattern for building backend Flask applications, the object relational mapping pattern for mapping data from Python to SQL, and the ETL pattern in Python. Students will learn to read, contribute to, and debug a large code base.

Data Analytics
We'll use Pandas as our data analytics, using it when learning to efficiently extract data insights, and tell a story with data. We'll then develop our first data pipeline with Prefect and DBT.

Module 2 - Cloud Computing and Data Pipelines

In this module, students will begin their internships, and deepen their data engineering skills by learning cloud computing and data pipelines.

Internships - Production Projects

Throughout the second half of the course, students are partnered with outside companies or open source projects where they will practice working with a larger engineering team, and contributing to a professional codebase. We allocate six hours of in class time per week – Wednesday and Sunday evenings – and many students devote additional hours outside of class. The externships give students the opportunity to gain technical and non-technical skills involved when working with a project to be used in production.

Cloud Computing
Students will deepen their knowledge of working in the terminal, and then deploy their Flask API to AWS instances, connecting the API to an RDS database in AWS. Afterwards we will learn about Docker – using both pre-existing images and building our own.

Data Pipelines

Serverless Pipelines
We'll interact with AWS through Python with Boto3, and learn to use AWS Lambda, Athena and Eventbridge and S3 to build serverless pipelines. We'll deploy our code using Docker and ECR, as well as infrastructure as code tool serverless.

Traditional Pipelines
With traditional pipelines, we'll move data from a transactional to an analytics based database, and the data modeling concepts and queries that come with it. We'll write SQL to export our data from RDS, to S3, and finally to Redshift. We'll automate this workflow with Airflow.

Weekly Schedule

The class is seventeen hours per week with classes running from 12:30 pm to 9:30 pm on Sundays (one hour break), and from 6:30 pm - 9:30 pm on Tuesdays, Wednesdays and Thursdays. All classes are held via Zoom.

Office hours are available to students weekly to review material or receive other assistance.

Module 1 - Backend Engineering (Weeks 1 - 13)			
Week	**Chapter**	**Topic**	**Description**
Week 1	Functional Python	Functions and APIs	Learn to write clean, small functions We'll learn how to retrieve and coerce data from APIs
Week 2	Scripting	Command Line, Python Scripts, Tests	Start working in the terminal, and running our code in scripts. We'll learn how to write tests for our functions.
Week 3	Intro to Databases	Single Table and Relational Queries	Learn SQL fundamentals including selecting data, performing aggregates, joins and data modeling
Week 4		Postgres	Learn how to use the postgres from the CLI and the psycopg2 library, practicing our relational queries and database design.
Week 5	Object oriented programming	Intro to Objects	Learn object fundamentals, including self, instance and class methods, and testing objects
Week 6		Object Relations	Relate various classes together with relational queries in Python and practice testing.
Week 7		OOP Projects	Build an OOP project from scratch, isolating models and from service objects
Week 8	Intro to building APIs	Flask and Github	Learn both git and github. We'll introduce Flask and MVC in Python.

Week 9		ORMs from Scratch	Write out and then use ORMs in our Flask application. We'll integrate relational queries in our codebase.
Week 10	ETL with APIs	ETL in Python Pattern	Use the the adapter pattern to query data from APIs, selecting and coercing the data to load into each corresponding table
Week 11	Data Analysis	Data Analysis with Pandas	We'll introduce pandas, move through techniques for checking a dataset and data exploration
Week 12	Data Pipelines with ETL and ELT	Prefect	We'll learn how to automate our pipeline with Prefect. Applying the ETL pattern.
Week 13		DBT Core	We'll learn to use DBT, applying classic ELT patterns to build data marts.
Week 14		DBT to Streamlit	We'll learn to connect our DBT pipeline to Streamlit to build dashboards in Streamlit.
Week 15	Pipeline Projects	Prefect/ETL	Develop the Prefect portion of your project.
Week 16		DBT to Streamlit	Coerce data to marts with DBT
Week 17		Data Storytelling	Perform data storytelling work and create data visualizations
Week 18		Presentations	Complete your data pipeline and present your project.

Module 2 - Cloud Computing and Data Pipelines (Weeks 14 - 24) *Externships: Sunday evenings and Wednesdays*			
Week	**Chapter**	**Topic**	**Description**
Week 19	AWS Fundamentals Big O	Advanced terminal, EC2, RDS	Learn how to our API to an EC2 machine, and connect the database to an AWS RDS database Learn about Big O, and review Python with the brute force approach.
Week 20	Docker ETL Pipelines	Docker CLI and containers	Learn the differences between containers and images, and learn to use bind mounts and volumes

			Practice writing and building our own images in Docker
Week 21		Docker Images	Learn to build images from scratch
Week 22	Serverless Pipelines	Boto3, Athena	Use Boto3 to store in S3 and query with Athena.
Week 23		Lambda, ECR	Deploy dockerized lambda functions, triggered via Boto3
Week 24	ETL Pipelines Review ETL and Analytics Engineering	Docker Compose/Airflow	We'll move onto Docker Compose
Week 25		OLTP to OLAP	Review querying and modeling data in 3NF. We'll then move to modeling and querying data using the star schema.
Week 26		ETL with Cloud Services	Introduction to Amazon Redshift, and learn how to migrate data from a transactional RDS instance, to S3 to redshift. Learn how to migrate our data from OLTP to OLAP in SQL.
Week 27		Airflow and Python	Work with Airflow 2, working with DAGs, tasks and connections Review Python with scraping, and CS fundamentals like lists, dictionaries and histograms
Week 28	Database Architectures	Snowflake	Set up snowflake and learn about snowflake warehouses, computes, and worksheets, and column based storage
Week 29		Spark	Learn about features of RDDs, MapReduce, and the Pyspark interface.
Week 30	Finals	SQL, Python, OOP Finals	We'll review our fundamentals in SQL, Python and object oriented design.

Post Graduation Classes

After completing the core material, we continue to meet twice weekly as you begin your job search. While we'll keep a degree of flexibility to respond to the current job market, in the past topics have included the following:

(1) Python Data Structure, Algorithms
(2) Python Leetcode Problems and Problem Solving
(3) SQL Leetcode and SQL Topics (Outer, Cross & Self Joins, Correlated Subqueries)
(4) Spark/Pyspark, Review Database Architectures
(5) Tableau
(6) Working with Open Source Projects
(7) Networking and Behavioral Interviews