dgPad

Full Stack
Data Engineering
(Python/SQL)

Intensive Bootcamp

2021 Syllabus

> *The dgPad syllabus is carefully designed, above all else, to foster developers' success in today's tech industry.*

dgPad's syllabus is designed to shorten the time to employment by quickly accumulating experience and acquiring skills through guided learning and hands-on experience. Our syllabus is living and evolving. We constantly improve it by making the necessary changes based on the ever-changing needs of the technology industry, and certainly feedback from employers.

Our goal is to contribute to filling the increasing demand gap for skilled developers by preparing our students with the right knowledge and skills to get employed in the best companies in our region and internationally.

› 12 Weeks Intensive: Creating Top Notch Developers

Students will be spending most of their time coding in pairs or individually to solve daily challenges and submit weekly and bi-weekly projects, putting 8 hours a day, at a minimum, 5 days a week.

Our instructors will be introducing new concepts and challenges on a daily basis and we expect students to have the motivation to discover the extra material that will be given for each lesson. Practicing and digging deeper is what makes you stand out from the crowd.

Duration:
12 Weeks | 9-5 Pm | Monday To Friday

Location:
On-site @ Albab Coworking Space

Registration Fees:
LBP 400,000

Tuition:
0$ (Pay Nothing Until You're Hired)

> Who is it for?

This bootcamp is at an advanced level. We expect participants to know the basics of computer science. Participants should know how to code with the main language of the chosen bootcamp. However, we encourage anyone to learn programming so we have put together a preparatory course for beginners and people with no prior experience to get the programming foundations before participating in the bootcamp.

We Welcome:

✓ *People with no prior experience*

✓ *Beginners who have a brief background about programming*

✓ *Professionals who want to strengthen their programming knowledge*

✓ *Rigorous and motivated career shifters and entrepreneurs*

✓ *Fresh CS/CE graduates who want to gain in-demand skills and build a portfolio to launch their career*

> "
>
> *The best way to learn coding is by spending many many hours actually coding.*
>
> "

> Overview

Data engineering is the practice of designing and building systems that collect, process, store, analyse and visualise massive data sets. It's the foundation for the new world of Big Data with various implementations in just about every industry. The Data Engineering bootcamp is an amazing opportunity for students who want to learn how to build data infrastructures to solve real world challenges.

Students will learn industry's most sought-after skills through working on a data analytics platform from design to deployment while applying the tools, concepts, and processes learned during the bootcamp.

Our Experienced instructors will supervise the work and provide feedback to make sure students understand the data pipeline life cycle and apply best practices.

› Bootcamp Outline

WEEK 1:   **DATA ENGINEERING SKILLS**

Learn about the core skills to become a data engineer and how this role differs from software engineers and data scientists. Get an overview about data pipelines, python programming language and the tools needed throughout the bootcamp.

- Data Engineer Vs Software engineer Vs Data scientist.
- Data pipelines.
- PC Setup (Install Python, Pycharm Ide, Node)
- Git and Source Control.
- Python Fundamentals and libraries.
- Web technologies and Python Flask
- Understand cloud computing
- Create a Google cloud account and understand their services.

TOPICS:
Git/Source Control, Python, Flask, Cloud Computing

PROJECT:
Build a demo web application using python flask and deploy it to Google cloud.

WEEK 2-3:   **DATA COLLECTION**

Learn about different data formats like JSON, XML, CSV, and build scalable data collectors for receiving user events and items metadata in addition to saving the collected data into local machines and to Google Cloud Storage.

- Build a Python Web scraper to extract items metadata.
- Build a Flask data collector that is responsible for receiving user events.
- Build a Client tracker that collects events.
- Build a data generator to create dummy user events (login, view, clicks, purchase, etc.)

TOPICS:
Data Formats, Behavioral data, Scalable Web apps, Web scraping, Client Tracking, Data Generator

PROJECT:
Data Collector, Web Scraper, Client Tracker

WEEK 4:   **DATA PROCESSING**

Learn about the various techniques to process, clean, enrich and transform data and build a pipeline to process the previously collected web pageviews and website items metadata.

- Clean malformed data
- Parse User agent to extract browser info
- Get user location from IP address
- Join pageviews and metadata

TOPICS:
Regex, Pipeline, Data Cleaning, Data Enrich

PROJECT:
Data pipeline, IP to Location Lookup, User Agent Parser

WEEK 5:   **BIG DATA PROCESSING**

Learn more about the big data ecosystem and how to use Apache Spark and Apache Beam to work with massive datasets and get familiar with parallel processing, batch and realtime systems.

- Process Data using Apache spark.
- Deploy Spark pipeline to Google Data proc.
- Process data using Apache Beam
- Deploy Beam pipeline to Google Dataflow.

TOPICS:
Distributed Storage, Parallel Processing, Apache Spark, Apache Beam, Google Dataflow and Data Proc.

PROJECT:
BigData Pipeline

WEEK 6-7:   **DATA MODELING & STORAGE**

Identify the strengths and weaknesses of different types of databases and data storage techniques, and learn more about using Sql and NoSql databases.

- Understand when to use a relational database
- Store data into relational databases like MySQL.
- Understand the difference between OLAP and OLTP databases
- Create normalized data tables.
- Implement denormalized schemas (e.g. STAR)
- Overview of NoSQL Database (MongoDB)
- Create a MongoDB database and store data in it

TOPICS:
Relational Database, SQL/NoSQL, OLAP/OLTP, Normalization/Denormalized, MongoDB, Star schema/Cubes, Facts/Dimensions

PROJECT:
MySql Datastore, MongoDb Datastore

WEEK 8:   **DATA STORAGE USING CLOUD DATA WAREHOUSES**

Understand Data Warehousing architecture and get an overview about Google Bigquery. Trainer will also build a bigquery dataset to save Enriched User Events and items Metadata.

- Data Warehousing architecture
- Build BigQuery Datasets and tables

TOPICS:
Data Warehouse, Google BigQuery, Data Partitioning, Batch loads

PROJECT:
BigQuery data warehouse

WEEK 9: **DATA ANALYSIS USING SQL AND PYTHON**

Get a deep understanding about SQL commands to filter, sort, group and join data. Also learn how to use Python pandas to analyse small data sets and Spark/Beam to analyse massive datasets.

• SQL in depth
• Analyze data using SQL queries.
• Analyze data using Python Pandas.
• Analyze data using Spark and Apache beam.

TOPICS:
Analytics Continuum, SQL Queries/Views, Query Optimization, Indexing, Python Pandas

PROJECT:
Data Analyser

WEEK 10: **DATA API & SECURITY**

Learn how to build data services or data APIs that access, transform, and deliver analytic data in Json format using Python flask framework in addition to securing the APIs from unauthorized access.

• Build Json Data API.
• Secure the data API.
• Execute SQL Queries/Jobs and return data in Json format.

TOPICS:
Web APIs, Authentication and Authorization, Flask

PROJECT:
Data Serving

WEEK 11: **DATA VISUALIZATION**

Learn the fundamentals of data visualization and design. You'll learn to select the most appropriate data visualization based on the goal of the analysis. You'll also learn important design elements for graphically representing data in addition to using Google data studio to build interactive and engaging dashboards.

• Data Visualization Fundamentals
• Design Principles
• Creating Visualizations in Google Data studio
• Telling Stories with Datastudio

TOPICS:
Story Telling, Charts/Maps, Google Data studio

PROJECT:
Data Visualizer

WEEK 12: **DATA INTELLIGENCE & MACHINE LEARNING**

Students will get a general overview about machine learning and play around with some libraries related to ML like scikit learn.

TOPICS:
Machine Learning, Scikit Learn, Predictive Analysis, Recommendation engine

PROJECT:
Content Recommendation, Predictive analyser, Students can choose up to their use case

❯ Learning Outcomes

By the end of the bootcamp you will have:

✔ **Build end-to-end data platforms**
✔ **Create user-friendly relational and NoSQL data models**
✔ **Work efficiently with massive datasets**
✔ **Build and interact with cloud-based data warehouses**
✔ **Automate and monitor data pipelines**
✔ **Develop proficiency in Apache beam, Google BigQuery**

❯ Data Engineering Technologies & Skills

• **Tech Stack:**

Python, SQL/NoSQL, MySql, Mongodb, Apache Spark, Apache Beam, Google Storage, Google BigQuery, Google Cloud run, Flask, Kafka

• **Skills:**

Data Cleaning, Data Scraping, Parallel Processing, Query Optimization, Data Partitioning, Distributed storage, ETL, Orchestration, Cloud Computing, Data Modeling, Data Security

› Projects

During the bootcamp students will be building a Data Analytics Platform (DAP), which is an ecosystem of services and technologies that performs analysis on massive datasets allowing organisations to retrieve, combine, explore, and visualize data from the various sources a company might have. Its main scope is to turn any kind of data into actionable insights for real business outcomes.

The DAP consists of 5 different modules (mini projects) that students will be building throughout the bootcamp:

- **Data Collector:** Collect events from users and track the way they engage with a website or an application.
- **Data Processor:** Process the raw events using big data tools and enrich the events with external datasets.
- **Data Store:** Stores data in a scalable event data warehouse.
- **Data Analyzer:** Analyse data to prepare insights provided by the analytical team.
- **Data Visualizer:** Visualize the data using ready and custom reports.

Data Analytics Platform

Data Collection	Data Processing	Data Storage	Data Analysis	Data Visualization

The DAP can be implemented in several use cases, so at the beginning of the bootcamp, each student will choose one of the following:

The DAP consists of 5 different modules (mini projects) that students will be building throughout the bootcamp:

- **Content Recommendation System**
  - ✔ Increase user engagement with data-driven content recommendations
  - ✔ Get a 360 view of how users are engaging with your content to build a recommendation engine that drives up engagement.

- **Product Recommendation System**
  - ✔ Increase revenue with data-driven product recommendations
  - ✔ Companies like Amazon make up to 35% of their revenue from product recommendations

- **Web Analytics System**
  - ✔ Assemble the web analytics that make sense for your website

- **Mobile App analytics System**
  - ✔ Capture data from mobile apps to identify unique users, track their journeys, and record their behavior to optimize the mobile app experience

- **Customer Data Platform**
  - ✔ Deliver personalized experiences by developing a single customer view
  - ✔ Insights into how your customers interact with you across platforms and channels over time enable you to deliver excellent customer experiences.

- **Advertisement Delivery optimization**
  - ✔ Help the Ad agencies to get more clicks with minimum ad impressions.
  - ✔ Target Ads to the audience who are most likely to click the ads' links.

> Career Services & Support

A dedicated Career success manager will be following up with students along the bootcamp to prepare them for post-graduation challenges through:

1- Resume Help

You will learn how to craft a professional resume that is ready to share with employers.

2- Interview Training

During the bootcamp you will participate in mock interviewing. You will learn the do's and don'ts of the interview process and important tips to help you conduct successful interviews with employers.

3- LinkedIn Profile

It's important to know how to best position yourself for competitive opportunities. We teach you how to build a LinkedIn profile that is specifically suited for the tech career you are aiming for.

4- Connect with Employers

One of our main goals in this bootcamp is to help our students land a job in the tech industry in a record time. You will benefit from our network of companies who are eager to hire new talents to enrich their teams. Your evaluation and work will be shared with potential employers to facilitate the recruitment process.

5- Post-Graduation Support

As a dgPad alumni you will always have access to your career success advisor and you will keep receiving information on specific job opportunities that we think will be fit for you until you get hired. An alumni network will be available at the first anniversary, which will bring networking and collaboration opportunities.

6- Soft Skills

Knowing how to code is not enough in the modern workforce. During the bootcamp students will be evaluated and soft skills workshops (Ex: public speaking and presentation skills) will be provided based on the need. Other skills that you'll be honing during the bootcamp are: Strategic problem solving, Product and project management, Learning to learn, Teamwork and collaboration, Self-discipline and determination.