

Data Engineering Boot Camp | Transforming Data Into Insights - TTML5519

Empowering Data Mastery: Driving Innovation and Insights in Your Organization

Duration: 3 Days

Skill Level: Intermediate

Available Format: Instructor-Led Online; Instructor-Led, Onsite In Person ; On Public Schedule

The Data Engineering Boot Camp is a workshop style program that combines essential skills and practical experience geared to advance your expertise in data management and analysis. This course is carefully structured to equip you with the ability to orchestrate efficient data workflows, optimize data warehousing, and execute advanced analytics solutions.

What You'll Learn

Overview

The Data Engineering Boot Camp is a workshop style program that combines essential skills and practical experience geared to advance your expertise in data management and analysis. This course is carefully structured to equip you with the ability to orchestrate efficient data workflows, optimize data warehousing, and execute advanced analytics solutions. Ideal for individuals looking to elevate their organization's data strategies or manage complex data-centric projects, the program focuses on imparting practical skills such as sophisticated data processing, insightful analytics, and strategic data handling, empowering you to immediately apply these skills to benefit your day-to-day work or your organization's objectives.

The hands-on experience is an immersive journey into real-world applications. With a 50% lab to lecture ratio, you're actively engaging with complex, real-life data scenarios

under the mentorship of an industry veteran. These labs are designed to mirror the challenges encountered in professional environments, offering a deeply practical and immersive learning experience. You'll not only learn the theoretical aspects but also apply them in scenarios such as building data lakes, managing streaming data, and integrating various data technologies. This approach ensures that you're not just learning, but also acquiring the confidence and problem-solving skills necessary for immediate application in your professional life.

Completing this program equips you with a comprehensive toolkit for effective data engineering. You'll emerge with the knowledge to manage and analyze large data sets, integrate and apply sophisticated data processing technologies, and provide strategic insights in diverse professional contexts. This course goes beyond imparting knowledge; it instills the confidence and practical know-how to promptly implement these skills in your role, significantly enhancing your organization's data handling and analytical capabilities.

Flexible Scheduling, Additional Hands-on: Depending on the needs of your team and availability, this camp can run in sequential full days, half day format in back to back days, or spread out over time with homework and project work schedule in between sessions.

Objectives

Working in a hands-on learning environment you'll learn to:

- Harness the art of orchestrating complex data workflows to boost efficiency in organizational data processes.
- Dive into the world of data warehousing, developing skills to craft and optimize strategic data storage solutions.
- Uncover the secrets of advanced data analytics for transformative insights and informed decision-making.
- Become adept at merging and leveraging cutting-edge data technologies for streamlined data management.
- Sharpen your problem-solving prowess and apply strategic insights to real-world data challenges.

- Master real-time data streaming and processing to confidently manage high-velocity data challenges.
- Cultivate the skill to create and implement dynamic data models for scalable, high-quality solutions.

If your team requires different topics, additional skills or a custom approach, our team will collaborate with you to adjust the course to focus on your specific learning objectives and goals.

Audience

The Data Engineering Boot Camp is an intermediate level program ideal for IT professionals, data analysts, and aspiring data engineers who are looking to deepen their understanding and skills in data management and analytics. This program is particularly beneficial for those in roles that involve handling large datasets, optimizing data processes, and driving data-driven decision-making within their organizations.

Pre-Requisites

This is an intermediate-level program, designed to prepare attendees for a deeper dive into next-level, heavy hands-on machine learning courses and workshops. Attendees should have practical, hands-on experience working with Python for Data Science.

- Basic understanding of database concepts and data structures, enabling you to grasp complex data engineering principles more effectively.
- Familiarity with programming languages such as SQL or Python, as they are fundamental in managing and manipulating data.
- Some experience with cloud computing platforms, which will aid in comprehending the cloud-based data engineering tools and technologies taught in the course.

Next Steps / Follow-on Courses: We offer a wide variety of follow-on courses and learning paths for Generative AI, AI for Business, GPT, Applied AI, Azure OpenAI, Google BARD, AI for developers, testers, data analytics, machine learning, deep learning, programming, intelligent automation and many other related topics. Please see our catalog for the current **AI & Machine Learning Courses, Learning Journeys & Skills Roadmaps**, list courses and programs.

Agenda

Please note that this list of topics is based on our standard course offering, evolved from typical industry uses and trends. We'll work with you to tune this course and level of coverage to target the skills you need most. Topics, agenda and labs are subject to change, and may adjust during live delivery based on audience skill level, interests and participation.

Module 1: Foundations of Data Engineering

Gain a foundational understanding of key data engineering concepts and tools. By the end of this module, you will be adept at building and managing cloud infrastructure using Terraform, and setting up efficient data engineering environments, enabling you to confidently handle real-world data engineering tasks.

Topics

- Understanding Google Cloud Platform (GCP) (or AWS or Azure – Please inquire for options)
- Introduction to Docker and Docker-Compose
- Implementing Postgres with Docker for Local Development
- Building GCP Infrastructure Using Terraform
- Setting Up the Data Engineering Environment
- Assignment: Setting up a Local Database using Docker
- Further Exploration

Module 2: Orchestrating Data Workflows

Learn the principles and practical applications of orchestrating data workflows. This module will equip you with the skills to design and manage data lakes, and implement advanced workflow orchestration, enhancing your ability to streamline and optimize data processes.

Topics

- Exploring Data Lakes in Data Engineering
- Principles of Workflow Orchestration
- Implementing Workflow Orchestration with Mage
- Assignment: Designing a Basic Data Lake
- Additional Insights

Module 3: Mastering Data Warehousing

Master the intricacies of data warehousing, with an emphasis on BigQuery. You'll gain the ability to effectively operate, optimize, and integrate data warehousing solutions, particularly with BigQuery and Apache Airflow, boosting your data warehousing proficiency.

Topics

- Fundamentals of Data Warehousing
- Deep Dive into BigQuery
- Techniques in Partitioning and Clustering in BigQuery
- Best Practices for Optimizing BigQuery
- Understanding BigQuery's Internal Mechanisms
- Integrating BigQuery with Apache Airflow
- Introduction to BigQuery ML
- Additional Learning Materials

Module 4: The Art of Analytics Engineering

Explore analytics engineering, focusing on the utilization of dbt and its integration with key data systems. You'll learn to craft and deploy dbt models and ensure their quality, empowering you to build robust analytics solutions and drive insights.

Topics

- Introduction to Analytics Engineering
- Utilizing dbt (Data Build Tool)
- Integrating BigQuery and dbt
- Applying dbt with Postgres
- Crafting dbt Models
- Ensuring Quality: Testing and Documentation
- Deploying Analytics Solutions in the Cloud and Locally
- Data Visualization with Google Data Studio and Metabase
- Extended Resources

Module 5: Batch Data Processing Techniques

Understand the fundamentals and advanced techniques of batch data processing through Apache Spark. This module will enable you to proficiently analyze and process large data sets using Spark DataFrames and Spark SQL, significantly enhancing your data processing capabilities.

Topics

- Overview of Batch Processing
- Introduction to Apache Spark
- Working with Spark DataFrames
- Utilizing Spark SQL for Data Analysis
- Internal Mechanics: Understanding GroupBy and Joins in Spark
- Supplementary Information

Module 6: Streaming Data Fundamentals

Dive into the world of streaming data, focusing on Apache Kafka and related technologies. You'll develop the skill to set up and manage real-time data streaming processes, making you adept in handling modern, high-velocity data challenges.

Topics

- Getting Started with Apache Kafka
- Understanding Schemas in Data Streaming (Avro)
- Exploring Kafka Streams
- Integrating Kafka Connect and KSQL
- Additional Learning

Module 7: Capstone Project

- Application of Learned Concepts in a Comprehensive Project
- Project Development Phase
- Peer Review and Feedback Session
- Further Guidance and Resources

All applicable course software, digital courseware files or course notes, labs, data sets and solutions, live coaching support channels, CodeCoach.AI anytime tutor access, and rich extended learning and post training resources are provided for you in our “easy access, single source, no install required” online **Learning Experience Platform (LXP)**, remote lab and content environment. Access periods vary by course. We’ll collaborate with you to ensure your team is set up and ready to go well in advance of the class. Please inquire about set up details and options for your specific course of interest.

For More Information

Please [contact us](#) or call 844-475-4559 toll free for more information about our training services (instructor-led, self-paced or blended), coaching and mentoring services, public course enrollment or questions, partner programs, courseware licensing options and more.