

# Data Engineering Roadmap 2025

This roadmap is a part of video on Darshil's channel

I suggest watching above to understand how to follow this roadmap, this is in-depth roadmap that list major tech stack you need to focus on.

Use this roadmap as navigator to know where you stand and what you can focus on next!

## Combo Package Python + SQL + Data warehouse (Snowflake) + Apache Spark + Airflow



### DataVidhya

DataVidhya is your go-to resource for top-notch data engineering content. Our platform offers a wealth of knowledge and insights, tailored for both beginners and

dataVidhya <https://datavidhya.com/combo-pack>

**USE CODE: COMBO50 for 50% off**

## 1. Computer Science Fundamentals (If you don't have a CS background)

If you are just getting started and don't have any understanding about computer science then start by watching this amazing video series by Harvard University CS50 on YouTube

Watching first 5 videos are enough

a.

### CS50 2023

#### CS50x 2023 - Lecture 0 - Scratch

\*\*\*

This is CS50, Harvard University's introduction to the intellectual enterprises of computer  
 [https://www.youtube.com/watch?v=IDDmrzzB14M&list=PLhQjrBD2T380F\\_inVRXMIHCqLaNUd7bN4&pp=iAQB](https://www.youtube.com/watch?v=IDDmrzzB14M&list=PLhQjrBD2T380F_inVRXMIHCqLaNUd7bN4&pp=iAQB)



## b. Grokking Algorithms (best book)

#### Manning Publications Grokking Algorithms: An illustrated guide for programmers and other curious people

Manning Publications Grokking Algorithms: An illustrated guide for programmers and other curious people : Bhargava, Aditya: Amazon.in: Books

 <https://amzn.to/36pEfFV>

# Foundation Skills for Data Engineers

## Python

You can do one course from this list, I recommend my course as it is specially designed for Data Engineers with End-To-End Project

### 1. Python for Data Engineering (by Darshil)

Python for Data Engineering | Data Vidhya  
Join the Python for Data Engineering course by Data Vidhya  
<https://dataavidhya.com/courses/python>

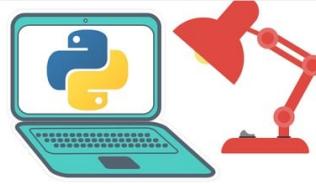
GET 50% off using code: PYT50

### 2. DataCamp - Data Engineering with Python

<http://datacamp.pxf.io/zaO9Gm>

### 3. Python Bootcamps: Learn Python Programming and Code Training

Python Bootcamps: Learn Python Programming and Code Training  
Learn Python like a Professional Start from the basics and go all the way to creating your own applications and games  
 <https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.567828&type=2&murl=https://www.udemy.com/course/complete-python-bootcamp/>



## SQL

Learn about the basics of SQL and how to write queries, once you complete the course make sure you do hands-on practice on Hackerrank or any website you like!

### 1. SQL for Data Engineering (by Darshil)

SQL FOR DATA ENGINEERING | Data Vidhya  
Join the SQL FOR DATA ENGINEERING course by Data Vidhya  
<https://dataavidhya.com/courses/sqldata>

GET 50% off using code: SQLT50

### 2. The Complete SQL Bootcamp for the Manipulation and Analysis of Data

The Complete SQL Bootcamp for the Manipulation and Analysis of Data  
Become an expert at SQL!  
 <https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.762616&type=2&murl=https://www.udemy.com/course/the-complete-sql-bootcamp/>



**Practice SQL here**

Solve Programming Questions | HackerRank

A special-purpose language designed for managing data held in a relational database.

 <https://www.hackerrank.com/domains/sql>



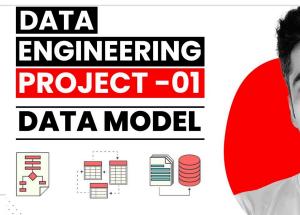
## Do Hands-On Project

Complete a project with me! - Building Data Model and Database | #LearnByDoing (Part 1/3)

Join Course Python for Data Engineering - <https://bit.ly/4140KI9>

Collecting knowledge is of no use if you don't know how to put it into use. With the motive of learning of

 <https://www.youtube.com/watch?v=2xyoz0T47Bs&list=PLBje2dFl4sgukOW600B-OVvX9c6fQKJ2N>



## Core skills for Data Engineers

Here you will be doing two things at the same time, reading book for strong foundation and doing courses to learn latest tools

### Get The Fundamentals of Data Engineer book here

Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Grayscale Indian Edition) 

Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Grayscale Indian Edition) : Joe Reis, Matt Housley: Amazon.in: Books

 <https://amzn.to/3zFBnjz>

Read this book in background while doing other courses

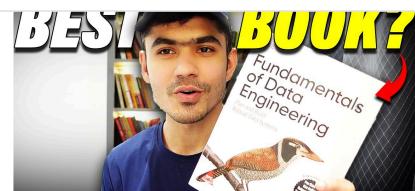
### Learn more about this book here

The only Data Engineering book you'll ever need

Join Python for Data Engineering Course: <https://dataavidhya.com/courses>

Best book for Data Engineering? In this video, I have reviewed The Fundamentals of Data

 <https://youtu.be/leElg0VdfG4>



## Data Warehouse (do one course)

### 1. Data Warehouse using Snowflake for Data Engineers (by Darshil)

Data Warehouse for Data Engineering with Snowflake

Join the Data Warehouse for Data Engineering with Snowflake course by Data Vidhya

 <https://dataavidhya.com/courses/dataswarehouse>

GET 50% off using code: DWT50

### 2. Data Warehousing for Business Intelligence Specialization

**Data Warehousing for Business Intelligence**  
Offered by University of Colorado System. Harness ... Enroll for free.

 <http://coursera.pxf.io/15zQYB>

University of Colorado  
Home | Connect | Study | Research | Medical Center

★★★★★ 4.5 (2.4K) | 64K Students  
**Data Warehousing for Business Intelligence**

coursera

### 3. Data Warehouse Fundamentals for Beginners

**Data Warehouse Fundamentals for Beginners**  
Best Practices and Concepts for Architecture and Dimensional Design

 <https://click.linksynergy.com/link?id=HTtUFxqjt0c&offerid=1060092.2752048&type=2&murl=https://www.udemy.com/course/data-warehouse-fundamentals-for-beginners/>



You can get complete combo package of Python, SQL and Data Warehouse here:

 <https://courses.dataavidhya.com/checkout/zwbzl8e2Cx>

GET 50% off using code: **COMBO50**

### Apache Spark (do one course)

Watch this intro video first

Learn Apache Spark in 10 Minutes | Step by Step Guide  
Join My Data Engineer Courses Here: <https://dataavidhya.com/courses>  
What is Apache Spark and How To Learn? This video will discuss Apache Spark, its  
 [https://youtu.be/v\\_uodKAYwXA](https://youtu.be/v_uodKAYwXA)



#### 1. DataCamp - Big Data Fundamentals with PySpark (Recommended)

<http://datacamp.pxf.io/P0nNQq>

#### 2. Udemy - Spark and Python for Big Data with PySpark

**Spark and Python for Big Data with PySpark**  
Learn how to use Spark with Python, including Spark Streaming, Machine Learning, Spark 2.0 DataFrames and more!  
 <https://click.linksynergy.com/link?id=HTtUFxqjt0c&offerid=1060092.980798&type=2&murl=https://www.udemy.com/course/spark-and-python-for-big-data-with-pyspark/>



### Databricks (do one course)

### Microsoft Azure Databricks for Data Engineering

Offered by Microsoft. In this course, you will learn how ... Enroll for free.

 <http://imp.i384100.net/WDbj5P>

 Microsoft  
21K Students  
**Microsoft Azure Databricks for Data Engineering**  
coursera

### Online Courses - Learn Anything, On Your Schedule | Udemy

Udemy is an online learning and teaching marketplace with over 213,000 courses and 62 million students. Learn programming, marketing, data science and more.

 <https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.5015186&type=2&murl=https://www.udemy.com/course/databricks-certified-data-engineer-associate-practice-exams/>



## Apache Kafka

### Apache Kafka Course for Beginners: Learn Kafka Online

START HERE: Learn Apache Kafka 3.0 Ecosystem, Core Concepts, Real World Java Producers/Consumers & Big Data Architecture

 <https://bit.ly/3qdhRqa>



## Data Orchestration (AirFlow)

### The Complete Hands-On Introduction to Apache Airflow

Learn to author, schedule and monitor data pipelines through practical examples using Apache Airflow

 <https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.1919064&type=2&murl=https://www.udemy.com/course/the-complete-hands-on-course-to-master-apache-airflow/>



## Cloud Computing

### AWS (Amazon Web Services)

#### AWS Cloud Solutions Architect

Offered by Amazon Web Services. Start here to become an ... Enroll for free.

 <http://imp.i384100.net/b3LZM9>

 4.8 (1.4K) | 68K Students  
**AWS Cloud Solutions Architect**  
coursera

#### Ultimate AWS Certified Solutions Architect Associate (SAA)

Full Practice Exam | Learn Cloud Computing | Pass the AWS Certified Solutions Architect Associate Certification SAA-C03!

 <https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.2196488&type=2&murl=https://www.udemy.com/course/aws-certified-solutions-architect-associate-saa-c02/>



## Do Projects

### YOUTUBE DATA ANALYSIS - DATA ENGINEERING PROJECT

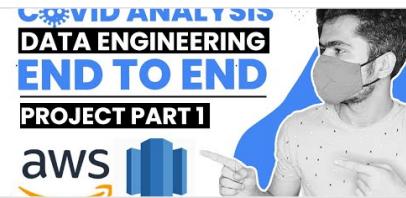
Execute the END TO END DATA ENGINEERING PROJECT using Kaggle YouTube Trending Dataset.

▶ <https://youtube.com/playlist?list=PLBJe2dFl4sguF2nU6Z3Od7BX8eALZN3mU>



### COVID ANALYSIS - END TO END DATA ENGINEERING PROJECT

▶ <https://youtube.com/playlist?list=PLBJe2dFl4sgvavQzL2Hm5CsnoIWHY5fl3>



### Data Engineering Project - Cloud Series

This playlist is the continuation of my Data Engineering Project Series! You will do end to end data engineering project on AWS Cloud where you will create d...

▶ <https://youtube.com/playlist?list=PLBJe2dFl4sgt-9GR2j-rTeKtimE9pfqyt>

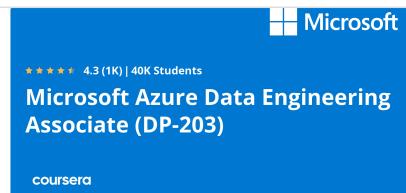


## Microsoft Azure

### Microsoft Azure Data Engineering Associate (DP-203)

Offered by Microsoft. Launch Your Career in Data ... Enroll for free.

⌚ <http://imp.i384100.net/YgO5MJ>



### Olympic Data Analytics | Azure End-To-End Data Engineering Project

Join My Data Engineer Courses - <https://datavidhya.com/courses>

In this video, you will analyze Olympic data using various tools and technologies,

▶ <https://youtu.be/laA9YNlg5hM?si=6Ez3JV8Bxf3a8IAF>



## GCP (Google Cloud Platform)

### Preparing for Google Cloud Certification: Cloud Data Engineer

Offered by Google Cloud. Advance your career in data ... Enroll for free.

⌚ <http://coursera.pxf.io/4eMjNG>



### Uber Data Analytics | End-To-End Data Engineering Project

Join My Python for Data Engineering Course: <https://datavidhya.com/courses>

In this video, you will analyze Uber data using various tools and technologies, including

▶ [https://youtu.be/WpQECq5Hx9g?si=TIFz\\_QRakv-HAkD9](https://youtu.be/WpQECq5Hx9g?si=TIFz_QRakv-HAkD9)



# Advance Data Engineering

## Open Table Format

Open Table Formats—Delta, Iceberg & Hudi

Brief intro

 <https://medium.com/geekculture/open-table-formats-delta-iceberg-hudi-732f682ec0bb>



Open Table Formats for Efficient Data Processing: Delta Lake vs Iceberg vs Hudi

Compare three popular open table formats to see which one will give you better performance based on your data processing needs.

 <https://medium.com/starschema-blog/open-table-formats-for-efficient-data-processing-delta-lake-vs-iceberg-vs-hudi-b1107141e9a6>



What is an Open Table Format? & Why to use one?

Do you need clarification about what Open Table Formats (OTF) are? Is it more than just a pointer to some metadata files that helps you sift through the data quickly? What is the difference between table formats (Apache Iceberg, Apache Hudi, Delta Lake) & file

 [https://www.startdataengineering.com/post/what\\_why\\_table\\_format/](https://www.startdataengineering.com/post/what_why_table_format/)



What is an open table format, and why to use one?



## Data Observability

What is Data Observability? What You Need to Know

What is data observability? Learn how data observability gives data teams the ability to maximize data reliability, eliminate operational problems, and perform cloud cost optimization.

 <https://www.acceldata.io/why-data-observability>

Observability Platform | Datadog

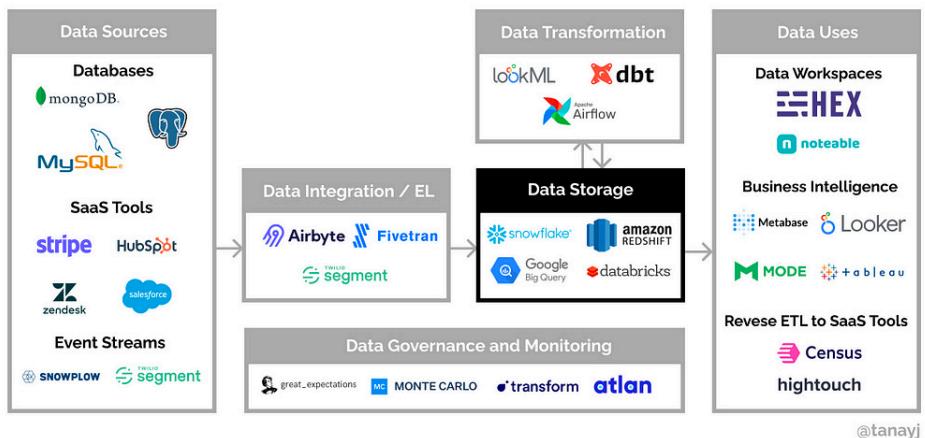
Modern monitoring and security allows you to seamlessly analyze and correlate front and backend data. Try Datadog's observability platform with support for 650+ vendor-backed technologies.

 <https://www.datadoghq.com/monitoring/observability-platform/>



## Modern Data Stack

## Modern Data Stack

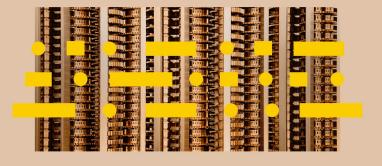


Learn basics about modern data stack

### Modern Data Stack and what we know about it

The Modern Data Stack is a novel data integration method that may save time while also focusing on high-value tasks.

<https://analyticsindiamag.com/modern-data-stack-and-what-we-know-about-it/>



### dbt Learn | dbt Labs | dbt Labs

Explore dbt Learn to find on-demand courses for using dbt.



<https://www.getdbt.com/dbt-learn/>

### Airbyte | Open-Source Data Integration Platform | ELT tool

Explore Airbyte, your go-to data integration platform and ELT tool. Seamlessly integrate, transform, and load data with our powerful, user-friendly solution. Accelerate your data workflows and unlock insights effortlessly.

<https://airbyte.com/>

<https://www.mage.ai/>

## DataOps

### [Project] Docker for absolute beginners | Coursera

Complete this Guided Project in under 2 hours. Docker ...

<https://www.coursera.org/projects/docker-for-absolute-beginners>

25K Students

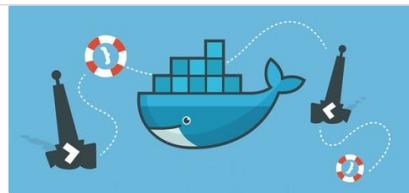
### Docker for absolute beginners

coursera

### Docker & Kubernetes: The Practical Guide [2023 Edition]

Learn Docker, Docker Compose, Multi-Container Projects, Deployment and all about Kubernetes from the ground up!

 <https://click.linksynergy.com/link?id=HTtUFxqjt0c&offerid=1060092.3490000&type=2&murl=https://www.udemy.com/course/docker-kubernetes-the-practical-guide/>



## Real-World Case Studies

### Netflix Technology Blog – Medium

Read writing from Netflix Technology Blog on Medium. Learn more about how Netflix designs, builds, and operates our systems and engineering organizations.

 <https://netflixtechblog.medium.com/>



### Customer Success Stories: Case Studies, Videos, Podcasts, Innovator stories

Learn how organizations of all sizes use AWS to increase agility, lower costs, and accelerate innovation in the cloud.

 <https://aws.amazon.com/solutions/case-studies/>



### Customers | Google Cloud

Learn why many of the world's leading companies are choosing Google Cloud to solve their toughest challenges.

 <https://cloud.google.com/customers>



### Customer and Partner Success Stories | Microsoft Azure

Explore Azure customer success stories and case studies to see how organizations all over the world are optimizing their costs and gaining new capabilities.

 <https://azure.microsoft.com/en-us/resources/customer-stories/>