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

   Watch this if you don't have a computer science background, as a Data Engineer having good knowledge of CS fundamentals is important to understand big systems and how they work

   Watching these videos will give you a basic understanding of CS fundamentals

   You can watch the first 7 lectures from this playlist

   CS50 2022

2. Introduction to Data Engineering

   Before learning different tools for Data Engineering, let's first understand "What is Data Engineering? And other topics"

   This section will give you a basic understanding of Data Engineering

   a. Introduction to Data Engineering
   b. Data Engineering in 7 Minutes
   c. Data Engineer (Expectation vs Reality)
   d. Day in a life of a Data Engineer
   e. Should You Become A Data Engineer?
   f. Different Types of Data Engineering Jobs
   g. Software Engineer vs Data Engineer

3. Programming Language

   Do any courses, your main goal here is to understand how to write basic Python code and how to work with different datasets!

   a. **Darshil -** Python for Data Engineering **(Recommended)**
   b. freeCodeCamp - Learn Python - Full Course for Beginners
   c. Programming with Mosh - Python Tutorial for Beginners

4. SQL (Structured Query Language)

   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!

   First, watch this - How I use SQL as Data Engineer

   a. freeCodeCamp - SQL Tutorial - Full Database Course for Beginners
   b. Programming with Mosh - MySQL Tutorial for Beginners [Full Course]

   Practice SQL here

   - Hackerrank SQL
   - SQL Tips and Tricks by Ankit Bansal
   - SQL Medium Interview Questions by Ankit Bansal

5. Basics Of Linux

Why Linux? Because you will be working with many remote machines, doing SSH to access them, and performing operations so it's important to learn them.

You don't have to remember all the commands but just understand what they do and how to write them

a. Kunal Kushwaha - Introduction to Linux and Terminal Commands
b. freeCodeCamp - Top 50 Most Popular Linux Commands

Your First Data Engineering Project

Building Data Model and Writing ETL Job

Data modeling is an essential part of Data Engineering, DO NOT SKIP THIS!!!

What will you learn?

✅ Python
✅ SQL
✅ Building Data Models
✅ Basics of DBMS
✅ Writing ETL Job
✅ Querying Data Programmatically
✅ PostgreSQL

Project Link - Beginner Data Engineering Portfolio Project

6. Core Data Engineering Concepts

This section is theoretical and you need to understand how big data system works and their history of them

a. What is Big Data
b. Database vs Data Warehouse vs Data Lakes
c. What is a Data Pipeline in Data Engineering
d. Different Types of file formats (CSV/ORC/Avro/Parquet)
e. Different Types of Data
f. OLAP vs OLTP
g. Batch vs Stream Data Processing

Big Data Fundamentals Full Course

a. Simplilearn **-** Big Data Full Course 2022

7. Data Warehouse Fundamentals

Same as the previous section, more theory, and understanding of concepts

a. Data Warehouse Fundamentals
b. Data Warehouse Tutorial for Beginners

8. Learn Batch/Realtime Streaming Pipeline Building
    a. Batch Pipeline (Spark)
        i. Great Learning - Spark Tutorial
        ii. Data Engineering - Apache Spark Tutorial
        iii. Learning Journal - Apache Spark Tutorial
    b. Realtime Streaming (Kafka)
        i. Learning Journal - Kafka Tutorial
        ii. Intellipaat - Kafka Tutorial

9. Data Orchestration (AirFlow)
    a. Airflow Tutorial for Beginners

10. Cloud Computing

Advance section, do courses, and then do the certification to add value in your Resume, If you are new then start with AWS but if you know about other clouds then you can do that too!

a. AWS (Amazon Web Services)
    i. AWS Certified Cloud Practitioner
    ii. AWS Certified Solutions Architect Associate (SAA)
b. GCP (Google Cloud Platform)
    i. Cloud Data Engineer Professional Certificate
c. Microsoft Azure
    i. AZ-900: Microsoft Azure Fundamentals
    ii. Azure Data Engineer Certified

Projects for Hands-on Practice

1. **ETL Pipeline on AWS Cloud**
   What will you learn?
   ✅ Python
   ✅ SQL
   ✅ Cloud Computing Basics
   ✅ AWS Services - Athena, Glue, Redshift, S3, IAM
   ✅ Creating Data Pipeline
2. **Covid Data Analysis Project**
   What will you learn?
   ✅ Python
   ✅ SQL
   ✅ Building Data Model
   ✅ AWS Services - Athena, Glue, Redshift, S3, IAM
   ✅ Creating Data Pipeline
   ✅ PostgreSQL
3. **YouTube Data Analysis (End-To-End Data Engineering Project)**

What will you learn?
✅ Python and PySpark
✅ SQL
✅ How to understand the business problem
✅ AWS Services - Athena, Glue, Redshift, S3, IAM, Lambda, Quicksight
✅ Building Data Pipeline and Scheduling it

4. **Twitter Data Pipeline using Airflow**

What will you learn?
✅ Python
✅ Basics of Airflow
✅ Working with Twitter Data and Package - Tweepy
✅ Python Package - Pandas
✅ Writing ETL job and storing data on S3