Combo Package Python + SQL + Data warehouse (Snowflake) + Apache Spark
https://courses.datavidhya.com/package/YF6cVZOWsg
USE CODE: COMBO50 for 50% off

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

   a. CS50 2023
   b. **Book -** Grokking Algorithms: An illustrated guide

2. 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)**
      GET 50% off using code: PYT50
   b. **DataCamp** - Data Engineering With Python
   c. **freeCodeCamp** - Learn Python - Full Course for Beginners

3. 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!

   a. **Darshil - SQL for Data Engineering (Recommended)**
      GET 50% off using code: SQLT50
   b. **Udemy** - The Complete SQL Bootcamp for the Manipulation and Analysis of Data (Recommended)

Practice SQL here

- Hackerrank SQL

Do Hands-On Project

- Beginner Data Engineering Portfolio Project (Recommended)

4. Data Warehouse Fundamentals + Tool

Learn Fundamentals and then learn one tool, Snowflake, BigQuery, Redshift, etc… Just learn one and you are good!

One course for everything in one place
Data Warehouse with Snowflake for Data Engineers (Recommended)
GET 50% off using code: DWT50

OR

a. **Fundamentals**
   i. **Coursera** - Data Warehousing for Business Intelligence Specialization (recommended for deep dive)
   ii. **Udemy** - Data Warehouse Fundamentals for Beginners
b. **Tools**
   i. **Snowflake** - Snowflake – The Complete Masterclass
   ii. **Snowflake Doc** - https://www.snowflake.com/certifications/

5. Learn Batch Processing + Tool
   a. Spark Fundamentals
      Apache Spark with Databricks for Data Engineers (Recommended)
      i. **DataCamp** - Big Data Fundamentals with PySpark (recommended)
      ii. **Udemy** - Spark and Python for Big Data with PySpark
   b. Databricks
      i. **Udemy** - Azure Databricks & Spark Core
      ii. **Udemy** - Databricks Certified Data Engineer Associate
      iii. **Coursera** - Databricks for Data Engineering

6. Learn RealTime Streaming
   a. Realtime Streaming (Kafka)
      i. **Udemy** - Apache Kafka Course for Beginners: Learn Kafka Online (check this)
      ii. **edX** - Building ETL and Data Pipelines with Bash, Airflow, and Kafka

Do Hands-On Project - Stock Market Real-Time Streaming Pipeline

7. Data Orchestration (AirFlow)
    a. Udemy - The Complete Hands-On Introduction to Apache Airflow
    b. DataCamp - Airflow

Do Hands-On Project - Twitter Data Pipeline using Airflow

8. 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. Udemy - Ultimate AWS Certified Cloud Practitioner
    ii. Udemy - Ultimate AWS Certified Solutions Architect Associate (SAA)
    iii. Coursera - AWS Solution Architect Associate

b. GCP (Google Cloud Platform)
    i. Coursera - Cloud Data Engineer Professional Certificate

c. **Microsoft Azure**
    i. Coursera - Microsoft Azure Data Engineering Associate
    ii. Udemy - AZ-900: Microsoft Azure Fundamentals
    iii. Udemy - Azure Data Engineer Certified:8 COURSE BUNDLE

Do Hands-On Project

1. Build ETL Pipeline Using AWS Cloud
2. Covid Data Analysis Project
3. YouTube Data Analysis (End-To-End Data Engineering Project)
4. Olympic Data Analytics | End-To-End Azure Data Engineering Project
5. Uber Data Analytics Project On GCP

9. Open Table Format
    a. Open Table Formats—Delta, Iceberg & Hudi
    b. Open Table Formats for Efficient Data Processing: Delta Lake vs Iceberg vs Hudi
    c. What is an Open Table Format? & Why to use one?
10. Data Observability
    a. What is Data Observability? What You Need to Know
    b. Observability Platform | Datadog

11. Learn Modern Data Stack
    a. **Learn Basics -** https://analyticsindiamag.com/modern-data-stack-and-what-we-know-about-it/
    b. **Dbt** - https://www.getdbt.com/dbt-learn/
    c. **Airbyte** - https://airbyte.com/
    d. **Fivetran** - https://www.fivetran.com/

12. DataOps
    a. Docker Guide - https://www.coursera.org/projects/docker-for-absolute-beginners
    b. Udemy - Docker & Kubernetes: The Practical Guide

Recommended Books

1. Designing Data-Intensive Applications
2. Fundamentals of Data Engineering
3. The Data Warehouse Toolkit

Read Real-World Case Studies

1. **Netflix -** https://netflixtechblog.medium.com/
2. **AWS** - https://aws.amazon.com/solutions/case-studies/
3. **GCP** - https://cloud.google.com/customers
4. **Azure** - https://azure.microsoft.com/en-us/resources/customer-stories/

Follow Me Here:

1. Twitter - https://twitter.com/parmardarshil07
2. Linkedin - https://www.linkedin.com/in/darshil-parmar/
3. YouTube - https://www.youtube.com/c/DarshilParmar

All the best <3