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***

1. [CS50 2022](https://youtube.com/playlist?list=PLeLzIg9tqA3LQW-RiFA8zJUBcTKqUVLMU)
2. **Book -** [Grokking Algorithms: An illustrated guide](https://amzn.to/36pEfFV)
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!*

* 1. **Darshil -** [Python for Data Engineering](https://learn.datawithdarshil.com/courses/Python-for-Data-Engineering)  **(Recommended)**
  2. **DataCamp** - [Data Engineering With Python](http://datacamp.pxf.io/zaO9Gm)
  3. **Coursera** - [Python for Everybody Specialization](http://coursera.pxf.io/LPE3La) (Do this if you don’t know anything about python)
  4. **Udemy** - [Python Bootcamps: Learn Python Programming and Code Training](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.567828&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fcomplete-python-bootcamp%2F)
  5. **freeCodeCamp** - [Learn Python - Full Course for Beginners](https://www.youtube.com/watch?v=rfscVS0vtbw)

**Practice Projects:**

* Scrape Data Using BeautifulSoup Library eg. Amazon, Covid, Wikipedia, or any website you like
* Build A Calculator Using Python

1. 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!*

* 1. **Udemy** - [The Complete SQL Bootcamp for the Manipulation and Analysis of Data](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.762616&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fthe-complete-sql-bootcamp%2F) (Recommended)
  2. **Coursera** - [SQL for Data Science](http://coursera.pxf.io/jWaA7n)
  3. **DataCamp** - [Intro To SQL DataCamp](http://datacamp.pxf.io/kjAGKM)

**Practice SQL here**

* [Hackerrank SQL](https://www.hackerrank.com/domains/sql)

1. 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*

* 1. **Udemy** - [Linux for Beginners: Linux Basics](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.3945922&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Flinux-for-beginners-2021%2F)
  2. **Coursera** - [Linux Fundamentals](http://coursera.pxf.io/x9oDKO)
  3. **freeCodeCamp** - [Top 50 Most Popular Linux Commands](https://www.youtube.com/watch?v=ZtqBQ68cfJc) (Recommended)

**Do Hands-On Project**

* [Beginner Data Engineering Portfolio Project](https://www.youtube.com/watch?v=2xyoz0T47Bs&list=PLBJe2dFI4sgukOW6O0B-OVyX9c6fQKJ2N) (Recommended)

1. Big Data Fundamentals

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

* 1. **Coursera** - [Big Data Specialization](http://coursera.pxf.io/9Wqa6Q) (Recommended)
  2. **Udemy** - [Learn Big Data: The Hadoop Ecosystem Masterclass](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.768670&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Flearn-big-data-the-hadoop-ecosystem-masterclass%2F) (Do this if you want to learn about legacy systems)

1. Data Warehouse Fundamentals + Tool

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

* 1. **Fundamentals**
     1. **Coursera** - [Data Warehousing for Business Intelligence Specialization](http://coursera.pxf.io/15zQYB) (recommended for deep dive)
     2. **Udemy** - [Data Warehouse Fundamentals for Beginners](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.2752048&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fdata-warehouse-fundamentals-for-beginners%2F) (recommended for quick learning)
  2. **Tools**
     1. **Snowflake** - [Snowflake – The Complete Masterclass](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.4019302&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fsnowflake-masterclass%2F)
     2. **Snowflake Doc** - <https://www.snowflake.com/certifications/>

1. Learn Batch Processing + Tool
   1. Spark Fundamentals
      1. **DataCamp** - [Big Data Fundamentals with PySpark](http://datacamp.pxf.io/P0nNQq) (recommended)
      2. **Udemy** - [Spark and Python for Big Data with PySpark](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.980798&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fspark-and-python-for-big-data-with-pyspark%2F)
   2. Databricks
      1. **Udemy** - [Azure Databricks & Spark Core](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.4182538&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fazure-databricks-spark-core-for-data-engineers%2F)
      2. **Udemy** - [Databricks Certified Data Engineer Associate](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.5015186&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fdatabricks-certified-data-engineer-associate-practice-exams%2F)
      3. **Coursera** - [Databricks for Data Engineering](http://imp.i384100.net/WDbj5P)
2. Learn RealTime Streaming
   1. Realtime Streaming (Kafka)
      1. **Udemy** - [Apache Kafka Course for Beginners: Learn Kafka Online](https://bit.ly/3qdhRqa) (check this)
      2. **edX** - [Building ETL and Data Pipelines with Bash, Airflow, and Kafka](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.1075642&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fapache-kafka%2F)

**Do Hands-On Project -** [Stock Market Real-Time Streaming Pipeline](https://www.youtube.com/watch?v=KerNf0NANMo&t=2882s)

1. Data Orchestration (AirFlow)
   1. Udemy - [The Complete Hands-On Introduction to Apache Airflow](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.1919064&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fthe-complete-hands-on-course-to-master-apache-airflow%2F)
   2. DataCamp - [Airflow](http://datacamp.pxf.io/b3Q9o6)

**Do Hands-On Project -** [Twitter Data Pipeline using Airflow](https://www.youtube.com/watch?v=q8q3OFFfY6c)

1. 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!*

* 1. **AWS (Amazon Web Services)**
     1. Udemy - [Ultimate AWS Certified Cloud Practitioner](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.3142166&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Faws-certified-cloud-practitioner-new%2F)
     2. Udemy - [Ultimate AWS Certified Solutions Architect Associate (SAA)](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.2196488&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Faws-certified-solutions-architect-associate-saa-c02%2F)
     3. Coursera - [AWS Solution Architect Associate](http://imp.i384100.net/b3LZM9)
  2. **GCP (Google Cloud Platform)**
     1. Coursera - [Cloud Data Engineer Professional Certificate](http://coursera.pxf.io/4eMjNG)
  3. **Microsoft** **Azure**
     1. **Coursera -** [Microsoft Azure Data Engineering Associate](http://imp.i384100.net/YgO5MJ)
     2. Udemy - [AZ-900: Microsoft Azure Fundamentals](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.2394982&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Faz900-azure%2F)
     3. Udemy - [Azure Data Engineer Certified:8 COURSE BUNDLE](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.3150056&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fimplementing-real-world-use-cases-in-azure-data-factory-v2%2F)

**Do Hands-On Project**

1. [Build ETL Pipeline Using AWS Cloud](https://youtube.com/playlist?list=PLBJe2dFI4sgt-9GR2j-rTeKtimE9pfqyt)
2. [Covid Data Analysis Project](https://youtube.com/playlist?list=PLBJe2dFI4sgvavQzL2Hm5CsnoIWHY5fI3)
3. [YouTube Data Analysis (End-To-End Data Engineering Project)](https://youtube.com/playlist?list=PLBJe2dFI4sguF2nU6Z3Od7BX8eALZN3mU)
4. Learn Modern Data Stack
   1. **Learn Basics -** <https://analyticsindiamag.com/modern-data-stack-and-what-we-know-about-it/>
   2. **Dbt** - <https://www.getdbt.com/dbt-learn/>
   3. **Airbyte** - <https://airbyte.com/>
   4. **Fivetran** - ​​https://www.fivetran.com/
5. DataOps
   1. Docker Guide - <https://www.coursera.org/projects/docker-for-absolute-beginners>
   2. Udemy - [Docker & Kubernetes: The Practical Guide](https://click.linksynergy.com/link?id=HTtUFxqit0c&offerid=1060092.3490000&type=2&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fdocker-kubernetes-the-practical-guide%2F)

**Recommended Books**

1. [Designing Data-Intensive Applications](https://amzn.to/3XijdOJ)
2. [Fundamentals of Data Engineering](https://amzn.to/3wdBuAU)
3. [The Data Warehouse Toolkit](https://amzn.to/3kqXaHh)

**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