

# ROADMAP TO BECOME AWS DATA ENGINEER

## PRE-REQUISITES

### 1. SQL

**Basics** - <https://www.youtube.com/playlist?list=PLtgiThe4j67rAoPmnCQmcgLS4ilc5ungg>

**Advanced** - <https://lnkd.in/dPPpttHg>

**Practice problems (this will help you build a logic)** - <https://lnkd.in/dWCVfXsA>

#### Best Platforms to SQL problems daily:

- Dataford - <https://lnkd.in/gtqdw864>
- LeetCode - <https://lnkd.in/gkCpv7NA>
- Namaste SQL - <https://lnkd.in/gQTuepVf>
- HackerRank - <https://lnkd.in/gnFS4frz>
- SQLZoo - <https://sqlzoo.net/>
- Mode Analytics - <https://lnkd.in/gRPrQrf5>
- SQL Bolt - <https://sqlbolt.com/>
- GeeksforGeeks - <https://lnkd.in/ggYbizNB>
- LearnSQL - <https://www.learnsql.com/>
- LearnMode - <https://lnkd.in/gQYCKwS2>
- Strata Scratch - <https://lnkd.in/gYtZQY53>
- DataLemur - <https://datalemur.com/>
- SQL Fiddle - <http://sqlfiddle.com/>
- DB-Fiddle - <https://www.db-fiddle.com/>
- SQL Exercises - <https://lnkd.in/d89TewuQ>
- SQL Practice Set - <https://lnkd.in/ddn7hfeu>
- DataCamp - <https://lnkd.in/dUyvbSwC>
- Kaggle - <https://lnkd.in/dn83kbwv>
- Mode SQL Tutorial - <https://lnkd.in/d53iPD-U>
- SQL Murder Mystery - <https://lnkd.in/dtVqDV-g>

## 2. Python

### Fundamentals -

<https://www.youtube.com/playlist?list=PLTsNSGelpGnEpKQfQuOWmT85Pvsx1BWY>

**Pandas** - [https://lnkd.in/dt\\_-2-Uj](https://lnkd.in/dt_-2-Uj)

**NumPy** - <https://lnkd.in/dqvnAmSK> (Optional)

**Matplotlib** - <https://lnkd.in/gTY4pZgu> (Optional)

## 3. Data Warehousing and Data Modeling

### 20 Video Series by Manish Kumar

<https://www.youtube.com/watch?v=UiTvqSd52akClist=PLTsNSGelpGnGP8A74le1PgqHhewsqD3f>

### 5 Best Articles on Data Modeling

Part 1 - <https://medium.com/@seancoyne/cracking-the-data-modeling-interview-part-1-an-overview-b09e7d5a7938>

part 2 - <https://medium.com/@seancoyne/cracking-the-data-modeling-interview-part-2-normalization-indexes-and-partitioning-fac334d767ca>

Part 3 - <https://medium.com/@seancoyne/cracking-the-data-modeling-interview-part-3-slowly-changing-dimensions-e33118de7fd1>

Part 4 - <https://medium.com/@seancoyne/cracking-the-data-modeling-interview-part-4-designing-efficient-data-structures-625a78d5fe6c>

Part 5 - <https://www.holistics.io/books/setup-analytics/kimball-s-dimensional-data-modeling/>

## 4. Linux

### Master all Linux Commands for Data Engineering in just one video –

<https://www.youtube.com/watch?v=w1apcBWgJ0w>

## AFTER PRE-REQUISITES START WITH BELOW THINGS

### 5. Pyspark

[https://www.youtube.com/playlist?list=PLTsNSGelpGnFiErPovNizG\\_2IP2RvrgnK](https://www.youtube.com/playlist?list=PLTsNSGelpGnFiErPovNizG_2IP2RvrgnK)

### 6. AWS Cloud

**There are around 9 services which we need to learn as an AWS Data Engineer**

1. Glue
2. Step Functions
3. EMR
4. Redshift
5. S3
6. Amazon Kinesis Data Analytics
7. Amazon DynamoDB
8. AWS Lambda
9. Athena

There are no proper resources available for AWS Data Engineering on YouTube. So, you can consider joining some paid course or explore Udemy courses.

### 7. Interview Preparation – Use my guide.

### 8. AWS Projects –

#### 1. Airlines Data Ingestion | End-To-End AWS Data Engineering Project

<https://www.youtube.com/watch?v=35Du026h-KY>

#### 2. Reddit Data Pipeline Engineering | AWS End to End Data Engineering

[https://www.youtube.com/watch?v=LSlt6iVI\\_9Y&list=PL\\_Ct8Cox2p8WjETeFftZePYTvuTRDx5h-&index=1&t=3556s](https://www.youtube.com/watch?v=LSlt6iVI_9Y&list=PL_Ct8Cox2p8WjETeFftZePYTvuTRDx5h-&index=1&t=3556s)

### **3. Apache Spark Processing with AWS EMR**

[https://www.youtube.com/watch?v=ZFns7fvBCH4&list=PL\\_Ct8Cox2p8WjETeFftZePYTvuTRDx5h-&index=4](https://www.youtube.com/watch?v=ZFns7fvBCH4&list=PL_Ct8Cox2p8WjETeFftZePYTvuTRDx5h-&index=4)

### **4. Smart City End to End AWS Realtime Data Engineering Project**

[http://youtube.com/watch?v=Vv\\_fwF41\\_0&list=PL\\_Ct8Cox2p8WjETeFftZePYTvuTRDx5h-&index=8](http://youtube.com/watch?v=Vv_fwF41_0&list=PL_Ct8Cox2p8WjETeFftZePYTvuTRDx5h-&index=8)

### **5. YouTube Data Analysis (End-To-End Data Engineering Project)**

<https://www.youtube.com/watch?v=yZKJFKu49Dk>

### **6. Twitter Data Pipeline using Airflow**

<https://www.youtube.com/watch?v=q8q3OFFfY6c>

### **7. Stock Market Real-Time Data Analysis using Kafka, AWS, and Python**

<https://www.youtube.com/watch?v=KerNf0NANMo>

### **8. Covid Analysis, end to end AWS data engineering project.**

<https://www.youtube.com/playlist?list=PLBJe2dFI4sgvavQzL2Hm5CsnoIWHY5fI3>