Get started Op

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

towards
data science

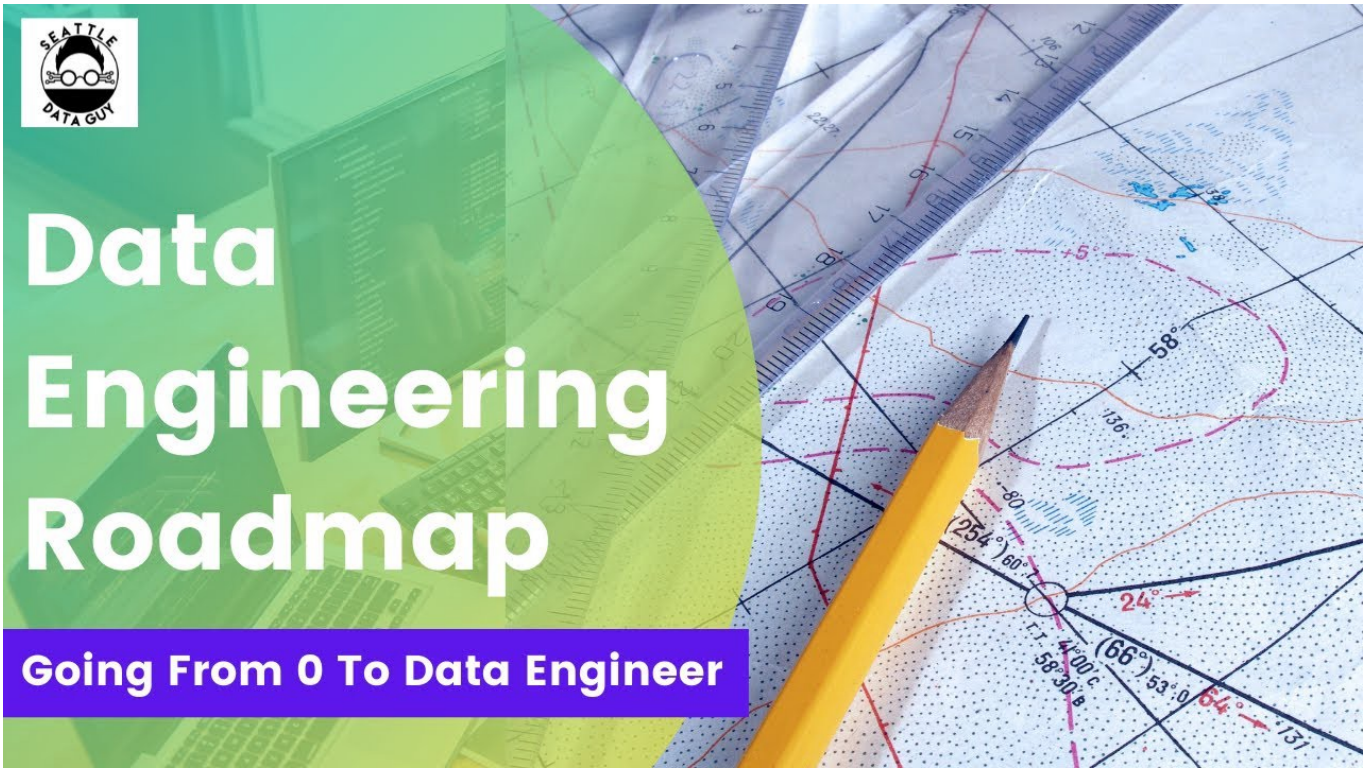
Follow 576K Followers

You have **2** free member-only stories left this month. Sign up for Medium and get an extra one

Data Engineering Roadmap For 2021

12 Steps To Help You Go From 0 To Data Engineering

SeattleDataGuy Jul 9 · 9 min read ★



Maybe it's the 6 figure salaries, the opportunity to work with cool technology or people are finally learning that data engineering is where everything starts in the data field.

Whatever the reason, people are noticing.

VCs are inve
interviewin

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

But how does one become a data engineer? If you were to Google data engineering roadmap, then you would find a very large image of an overwhelming roadmap that has been going around Linkedin for the past few weeks with over a decade of learning.

It's too much.

So in this article, we will provide the steps of how you can go from 0 to data engineer with a combination of free courses as well as paid that can help you gain the skills you need to be a data engineer.

But before diving into that, let's make sure you know what a data engineer is.

. . .

What Is Data Engineering

Data engineers move, remodel, and manage data sets from 10s if not 100s of internal company applications so analysts and data scientists don't need to spend their time constantly pulling data sets.

They may also create a core layer of data that lets different data sources connect to it to get more information or context.

These specialists are usually the first people to handle data. They process the data so it's useful for everyone, not just the systems that store it.

There are obvious reasons to become a data engineer — like a high salary and numerous opportunities due to limited competition within the job market — but we're not focusing on those today. Instead, consider the following thoughts, which are a bit more relevant to the job description.

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

Now before going to far on this data engineering roadmap. We need to answer a very important question.

. . .

0. Is Data Engineering For You?

For that, I have put together a video that may help you.

It will hopefully provide you some context on data engineering and if you would want to do it in the future.

If you're still here, then let's break down the roadmap to become a data engineer.

. . .

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

1. Building

Before getting deep into data engineering specifics you need a solid base.

It can be tempting to start learning some of the concepts and skills that are further along the lines of distributed computing or streaming. But that's like learning words and sentences before you learn what letters are.

That's why you need to start with SQL, programming, and some form of server/Linux basics.

You need to be able to speak to computers in their language and these three skills will help you understand how to communicate with computers from various layers.

Building this solid foundation will ensure that you reduce your future learning curves because to interact with many of the other technical components, you will need to understand some form of programming language or command line basics.

Also, learning the basics in terms of servers such as SFTP, firewalls, PGP, and other technical components will go a long way.

1. SQL Cours On Khan Academy
2. Python On Youtube
3. Data Structures And Algorithms(Edited in)
4. Linux On Udemy
5. FTP, SFTP, and TFTP (Edited in)

• • •

2. Try Building Your First Project — Building A Flask API

You will need to interact with APIs on a daily basis if you become a data engineer. Either to automate processes or pull data.

In that way, building an API is a great first project because it will force you to use several layers of technology.

You will nee mand line, and if you re e cloud by spinning up
But that's a stretch goal. Let's start easy.

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

Flask is a great python library that you can quickly spin up an API in no time. But I don't expect you to just know how to build your first API.

I like freeCodeCamp's Flask tutorial. Now, this is focused on building a site, but you can still use this tutorial to build out a lot of the backend infrastructure.

So for this project, you can follow along with freeCodeCamp and then try to add in your own end-points for your Flask API Project.

. . .

3. Learn About Data Warehousing And Data Pipelines(ETLs, ELTs, and ELs)

When you look at the skill sets of data engineers, software engineers, and data scientists, there is a lot of cross-over.

All three tend to use Python, both data scientists and data engineers tend to use SQL pretty heavily and all three rely to some degree on some understanding of Linux.

So what differentiates data engineers?

One of the big differentiators is the focus on data warehouses and data pipelines.

But what are these?

Data warehouses and data pipelines. At least to start.

Data warehouses and data pipelines are concepts that data engineers need to understand. They are the bread and butter of any good DE.

Luckily there are tons of resources that cover these concepts. But let's start with the granddaddy of all data warehousing resources. Kimball's Data Warehouse Guide.

This is a book written by one of the people who built much of the foundation for data warehouses. There is a lot of history there, but we won't go into that now.

If you need eriously.
Then check

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

These combined should cover most of the conceptual basics.

Of course, now we need to apply it.

. . .

4. Build Your Second Project — Applying Coding And Data Warehousing

Now that you have learned about data pipelines and data warehouses, it would be a great idea to apply this knowledge.

So let's build your second project, to solidify that knowledge. Let's aim to implement these 4 concepts below.

- Scrape an online data source
- Store encrypted data into SFTP
- Create dimensional model
- Pull data from SFTP and load into Data Warehouse(Don't worry too much about Workflows just yet)

At this point, this will bring many of the skills you have learned together. Whether it be learning about PGP encryption, SFTP, or data modeling.

All of which will improve your overall skillsets and provide some form of the final project.

. . .

5a. Learn About Testing

Testing.

You may have never even learned about testing in school or maybe you had that one course that had one unit for one week that just started to touch the surface of testing.

Now in a wo t part of the CI/CD p

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

You need to know how to write test cases.

You need to know the difference between unit tests and integration tests.

To do so, Udemy has a great course on test-driven development.

Check out TDD Course Like Unit TDD For Python

Or if you want a free version you can use this Youtube tutorial.

• • •

5b. Airflow And Docker

You will notice I have 2 step 5s. Well, that's because we are getting to the point where order matters a little less. Steps 6, 7, 8, and so on. Could probably get a little jumbled and you would be fine.

At this point, you should have a solid enough base that any new technology that comes your way shouldn't have the same learning curve.

That's why for this second step 5 I suggest you learn Airflow + Docker.

Why?

Because you st.

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

. . .

✉ Subscribe to my Mailing List For Community Updates And Freebies

6. Cloud and NoSQL

At this point, you have probably already done a little on the Cloud and maybe even played around with a NoSQL database.

But, let's round out that knowledge.

How? Well, there are a few great options when it comes to rounding out knowledge. For example, now I think it would be a good time to take a certificate program.

I don't often find certificates useful until you have some experience because I find certificates often help fill in the gaps of knowledge on a particular topic.

One great certificate for data engineers is the Google Data Engineer Certificate.

As for NoSQL check out freeCodeCamp's video on database systems.

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

7. Streaming And Distributed Systems

There are so many ways to process data in the modern world. More importantly, using more complex systems such as streaming or distributed systems is so much easier than it's ever been.

You can spin up a fully managed service on AWS or GCP and you're off to the races. No need to spin up 5 other services just to try to wrangle and manage your streaming system.

So, let's find some courses on this.

Andreas Kretz for example has a why video on Spark and Kafka thats a great primer before you jump into taking courses.

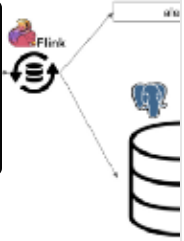
Once you have a good idea of Kafka and why you should learn it. Take, really, any of Frank Kane's courses. He has done a great job producing a whole host of courses that discuss Spark, Kafka and Hadoop.

Also, if you need an example of someone using a streaming component. Check out StartDataEngineering.

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

Data engin
design and

www.startdataengineering.com



8. Start Studying for Interviews

At some point, you need to go out into the real world and attempt to interview.

For that, I have put together a data engineering interview guide. This guide helps breakdown what you will need to study and provides the questions so you don't have to waste time Googling.

The Data Engineering Interview Study Guide

For your FAANG and other technical interviews

betterprogramming.pub

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

9. Build Y

At this point in your learning, you should have a broad knowledge of skills.

You should know about distributed systems, streaming, programing, APIs and so much more.

So now take all of that knowledge and apply it.

But to what?

Well, I put together a video where I discuss 5 different examples of projects that are real, that exist. You can find links to them on the video. There are tons of great examples of people using tons of technologies to build data engineering projects.

10.Learn Enough UI/UX and Dashboarding

Taras Bakusevych still has my favorite article in terms of how to design a dashboard. The way it's broken down, even with the cliche "10 rules" title, really helps you understand how to develop a dashboard.

Honestly, th nts.

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

10 rules for better dashboard design

Practical guide

uxplanet.org

Truthfully, UI/UX isn't always necessary for data engineers. However, for some of you out there, you will fall in love with designing dashboards and displaying data.

And in general, there will always be a need to at least build a good enough dashboard. So take a moment to learn this skill.

. . .

11. Pick Your Own

At this point in your learning, you have probably found areas you enjoy. Maybe you liked distributed computing or putting together an API.

Dig into that.

Figure out what you enjoy learning about and learn more.

Technology is great that way in the sense that it is an overflowing set of learnings. There is always more to dig into always concepts, technologies, design practices that you know nothing about.

So go learn.

. . .

12. Give Yourself Room To Grow

Learning is a process and it should be both fun and frustrating.

So let it be that way.

When you g ce you solve that p

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

You get a brief moment of excitement and dopamine for solving the riddle, but then there is just another riddle to solve.

Another bug to fix.

If you've gotten this far in your learning, then you have probably gotten past the "Hello World" phase. But if you recall, there was something magical about the first time you got code to run.

So don't rush.

And Good Luck!

. . .

Read/Watch These Next

- **Data Science Interview Study Guide**
  121 resources to help you land your data science dream job
- **How I Went From Analyst To Data Engineer**
  How to become a data engineer — and know if it's right for you
- How To Start A Consulting Business As A Consultant
  Getting Your First Client

. . .

✉ Subscribe to my Mailing List For Community Updates And Freebies

. . .

About Me

I have spent my career focused on all forms of data. I have focused on developing algorithms to detect fraud, reduce patient readmission and redesign insurance provider policy to help reduce the overall cost of healthcare. I have also helped develop analytics

for marketi
employees a
both solo as
working hands-on with technical problems as well as helping leadership teams develop strategies to maximize their data.

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

Seattle Data Guy

Welcome to my channel. I make videos on data science, data engineering and consulting. Follow me on: Linkedin...

www.youtube.com

Connect with Me on Social Network

YouTube:https://www.youtube.com/channel/SeattleDataGuy
Website: https://www.theseattledataguy.com/
LinkedIn: https://www.linkedin.com/company/18129251
Personal Linkedin: https://www.linkedin.com/in/benjaminrogojan/
FaceBook: https://www.facebook.com/SeattleDataGuy

Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. Take a look.

Get this newsletter

Data Science | Data Engineering | Programming | Technology | Careers

Medium

About Write He

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

Get the Medium ap

Download on the App Store

GET IT ON Google Play