

LEARNING RESOURCES:

1. Intro to Data Engineering:

- <https://medium.com/codex/what-is-data-engineering-407bcf860baf#:~:text=Data%20engineers%20use%20a%20variety,the%20cloud%2C%20or%20a%20mixture.>
- <https://akaabachi.medium.com/2023-rockstar-data-engineer-roadmap-c990da37f625>
- <https://vinodhini-sd.medium.com/how-to-break-into-data-engineering-in-2023-1b17afc7f5bd>
- <https://medium.com/@rchang/a-beginners-guide-to-data-engineering-part-i-4227c5c457d7>
- <https://www.montecarlodata.com/blog-the-future-of-the-data-engineer/>
- <https://insightsndata.com/how-to-become-a-data-engineer-in-8-steps-697afe32859a>
- <https://medium.datadriveninvestor.com/the-great-debate-data-engineer-vs-data-scientist-29ea06912529>
- <https://www.datacamp.com/blog/what-is-data-engineering>
- <https://heartbeat.comet.ml/data-engineering-roadmap-is-there-any-e2a0f1303012>
- <https://towardsdatascience.com/who-are-you-data-engineer-8febb06552a>
- <https://towardsdatascience.com/tagged/data-engineering>
- <https://betterprogramming.pub/data-engineering-101-from-batch-processing-to-streaming-54f8c0da66fb>
- <https://medium.datadriveninvestor.com/what-is-a-data-engineer-a-guide-to-this-in-demand-career-13489380f06>
-
- https://www.youtube.com/watch?v=qWru-b6m030&t=7s&ab_channel=AltexSoft
- https://www.youtube.com/watch?v=pERD1Ln_Qh8&t=685s&ab_channel=DataScienceDojo

2. Docker and Microservices

- https://www.youtube.com/watch?v=3c-iBn73dDE&ab_channel=TechWorldwithNana
- https://www.youtube.com/watch?v=8vXoMqWgbQQ&ab_channel=TechWorldwithNana
- https://www.youtube.com/watch?v=fqMOX6JhGo&t=5990s&ab_channel=freeCodeCamp.org
- https://www.youtube.com/watch?v=WmcdMiyqfZs&ab_channel=TechWorldwithNana
- https://www.youtube.com/watch?v=MVIcrmeV_6c&ab_channel=TechWorldwithNana
- https://www.youtube.com/watch?v=rv4LlMlMVWk&t=771s&ab_channel=TechWorldwithNana
- https://www.youtube.com/watch?v=IL_j7ilk7rc&ab_channel=5MinutesorLess

3. FAST API:

- HTTP APIs were introduced in 4.1 lecture: <https://emeritus.talentlms.com/unit/view/id:2528>. Due to the time restrictions we didn't go deep into it, but you can start with the slides and then see examples of exposing an API and accessing it from your web browser (linked in the slides).
- One of 4.1 activities was aiming to challenge you to integrate frontend service and inference service. They communicate over HTTP and in solutions directory you can see an example of [frontend making the request](#) and [inference exposing an endpoint for POST](#) (btw. [frontend exposes an endpoint on its own](#), using Flask instead of FastAPI – this is not for communication between these two services though but for accessing the frontend externally).

- This is a good introduction to FastAPI: <https://realpython.com/fastapi-python-web-apis/>. It uses pydantic library for data validation. We talked a bit about this library in our 1.2 lecture.
- If you want to explore FastAPI in more details, I personally found [FastAPI tutorial](#) useful.
- https://www.youtube.com/watch?v=XnYYwcOfcn8&list=PLqAmigZvYxIL9dnYeZEhMoHcoP4zop8-p&ab_channel=JVPDesig
- https://www.youtube.com/watch?v=0sOvCWFmrtA&ab_channel=freeCodeCamp.org

4. Kafka:

- <https://medium.com/swlh/apache-kafka-what-is-and-how-it-works-e176ab31fcd5>
- <https://medium.com/inspiredbrilliance/kafka-basics-and-core-concepts-5fd7a68c3193>
- <https://betterprogramming.pub/thorough-introduction-to-apache-kafka-6fbf2989bbc1>
- <https://medium.datadriveninvestor.com/an-introduction-to-apache-kafka-43eaad561198>
- <https://levelup.gitconnected.com/apache-kafka-architecture-and-how-does-it-work-54e99f72b9be>
- <https://towardsdatascience.com/how-to-build-a-simple-kafka-producer-and-consumer-with-python-a967769c4742>
- [Real world example of Kafka](#)
- [Apache Kafka Basics](#)
- [Apache Kafka Tutorial \(Similar logic will be implemented in our project\)](#)
- [Integration of Kafka with FAST API](#)
- https://www.youtube.com/watch?v=c2K4tkzK0dA&ab_channel=BhimrajYadav

5. Pyspark/Pandas

- [Pyspark with AWS \(note this tutorial will also helpful in week 4\)](#)
- [How to read kafka stream data event in pandas](#)
- [Handling real time kafka streams using pyspark](#)
- [Integrate kafka with Pyspark](#)
- [Connecting the dots with python, spark and kafka](#)
- [How to test pyspark etl data pipeline](#)

6. MongoDB:

- [MongoDb Tutorial](#)
- [MongoDb Basics](#)
- [Working with MongoDB using Spark to Read/Write the Data](#)
- [Read Collection from MongoDB using PySpark](#)
- [Spark Connector](#)
- [Spark Connector Official MongoDB Documentation](#)

Medium

[What is Data Engineering?](#)

While data engineering is not a new field, it seems to have stepped out from the background recently and started to take center stage. In...

Reading time

13 min read

Apr 7th, 2021

<https://medium.com/codex/what-is-data-engineering-407bcf860baf#:~:text=Data%20engineers%20use%20a%20variety,the%20cloud%2C%20or%20a%20mixture>

Medium

[2023 — Rockstar Data Engineer Roadmap](#)

This article presents a roadmap for those who want to become Data Engineers in 2023. It also serves as a reference to learn and improve...

Reading time

4 min read

Jan 1st

<https://akaabachi.medium.com/2023-rockstar-data-engineer-roadmap-c990da37f625>

Medium

[How to break into Data Engineering in 2023?](#)

Step by step guide to help you get into junior DE roles in 30 days!

Reading time

2 min read

Jan 10th

<https://vinodhini-sd.medium.com/how-to-break-into-data-engineering-in-2023-1b17afc7f5bd>

Medium

[A Beginner's Guide to Data Engineering — Part I](#)

Data Engineering: The Close Cousin of Data Science

Reading time

14 min read

Jun 25th, 2018

<https://medium.com/@rchang/a-beginners-guide-to-data-engineering-part-i-4227c5c457d7montecarlodata.com>

[The Future Of Data Engineering As An Engineer | Monte Carlo](#)

Is the data engineer still the "worst seat at the table?" Maxime Beauchemin, creator of Apache Airflow, weighs in on the future of data engineering.



Medium

[What is Data Engineering?](#)

While data engineering is not a new field, it seems to have stepped out from the background recently and started to take center stage. In...

Reading time

13 min read

Apr 7th, 2021(Not automatically expanded because 2 MB is too large. You can [expand it anyway](#) or [open it in a new window.](#))



Medium

[2023 — Rockstar Data Engineer Roadmap](#)

This article presents a roadmap for those who want to become Data Engineers in 2023. It also serves as a reference to learn and improve...

Reading time

4 min read

Jan 1st

<https://akaabachi.medium.com/2023-rockstar-data-engineer-roadmap-c990da37f625>



Medium

[How to break into Data Engineering in 2023?](#)

Step by step guide to help you get into junior DE roles in 30 days!

Reading time

2 min read

Jan 10th (50 kB)

<https://vinodhini-sd.medium.com/how-to-break-into-data-engineering-in-2023-1b17afc7f5bd>



Medium

[A Beginner's Guide to Data Engineering — Part I](#)

Data Engineering: The Close Cousin of Data Science

Reading time

14 min read

Jun 25th, 2018 (2 MB)

<https://medium.com/@rchang/a-beginners-guide-to-data-engineering-part-i-4227c5c457d7>



montecarlodata.com

[The Future Of Data Engineering As An Engineer | Monte Carlo](#)

Is the data engineer still the "worst seat at the table?" Maxime Beauchemin, creator of Apache Airflow, weighs in on the future of data engineering.