

Module 1: Fundamentals of Data Engineering



Overview

In Module 1, we introduce Data Engineering, discuss its cross-disciplinary nature, introduce tasks and tools, and present the course's custom pipeline.



Objectives

Upon completion of this module, students should be able to:

1. Define data engineering and its characteristics
 2. Identify the components and tasks of a data engineering project lifecycle
 3. Comprehend the cross-disciplinary nature of data engineering
 4. Explain the roles and responsibilities of a Data Engineer
 5. Understand the architecture of a custom pipeline comprising open-source tools and commercial platforms
-



Readings (1 hour)

- A Conversation with chatGPT on Data Engineering, Koushik Thota, 2022, <https://medium.com/towards-data-engineering/a-conversation-with-chatgpt-on-data-engineering-5e9434d0f4bc> ↗ (<https://medium.com/towards-data-engineering/a-conversation-with-chatgpt-on-data-engineering-5e9434d0f4bc>)
 - Data science is different now, Vicki Boykis, 2019. <https://vickiboykis.com/2019/02/13/data-science-is-different-now/> ↗ (<https://vickiboykis.com/2019/02/13/data-science-is-different-now/>), accessed 03/13/2022 (optional)
-



Watch (1 hour)

- **Lecture 1 Video** (https://canvas.vt.edu/media_objects_iframe/m-6du2EzkMSo1vNN4v6wxWmPcxiSPpFPij?type=video?type=video)

- [Lecture 1 Slides \(https://canvas.vt.edu/courses/176740/files/28995307?wrap=1\)](https://canvas.vt.edu/courses/176740/files/28995307?wrap=1)
-



Additional Course Material

- [Course Terminology \(https://canvas.vt.edu/courses/176740/files/28995309?wrap=1\)](https://canvas.vt.edu/courses/176740/files/28995309?wrap=1)
 - [Link to Course Tools and Documentations \(https://canvas.vt.edu/courses/176740/files/28995316?wrap=1\)](https://canvas.vt.edu/courses/176740/files/28995316?wrap=1)
-



Lab - Setting up your environment (1 hour)

Go through the [Lab \(https://canvas.vt.edu/courses/176740/assignments/1816315\)](https://canvas.vt.edu/courses/176740/assignments/1816315) to set up your environment



Class Activity (30 minutes)

- Introduce Yourself in the [Introductions \(https://canvas.vt.edu/courses/176740/discussion_topics/1648227\)](https://canvas.vt.edu/courses/176740/discussion_topics/1648227) Discussion Forum
- Use the [setting up the environment \(https://canvas.vt.edu/courses/176740/discussion_topics/1648226\)](https://canvas.vt.edu/courses/176740/discussion_topics/1648226) Discussions Forum to help each



Figure out



Recitation (1 hour)

The recording of the optional synchronous ZOOM session for this lecture will be linked here.

