

# Data Engineering och Agile Methods

[modulai.io](https://modulai.io)

## Olof Harrysson

Your teacher for the second half of Data engineering and agile methods.

Taking over from Kokchun who's on parental leave.

Master in **Computer Science** from Lund University  
4 years of work experience as a programmer in ML.  
ITHS board member for this education

### ML Programmer at Modulai

Modulai is ML consultancy ~20 people.

Questions are welcomed!



# Data Engineering

## Needed in all ML projects!

Data Engineering builds good software skills that can be applied elsewhere.

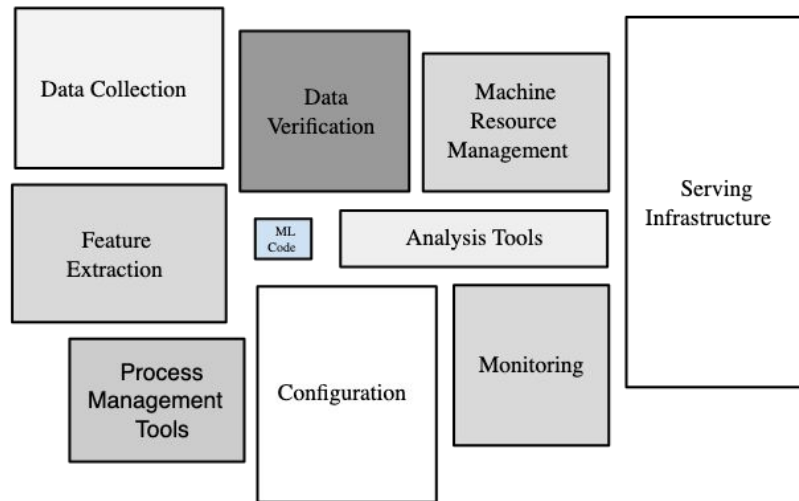
## Practical field where you can be **useful quickly**.

More programming/infrastructure. Less math/modeling.

## Good **job opportunities**.

Less competition from theoretical educations e.g. math

PhD students. They usually do more data science.



## Course Schedule

Week	Content
21	course intro, why data engineering, Linux, data pipelines, docker intro, dockerfile, docker compose, study visit Ericsson
22	containers, data engineering lifecycle, code review, ELT, Airflow intro, DAGs, operators, tasks, Xcom, study visit Olsaro
23	schedule data ingestion, TaskGroup, SCB API, orchestrate data pipeline,
33	repetition, guest lecture agile theory, github actions CI/CD, pre-commit, code tests
34	<b>project</b> start, data version control (DVC), postgres database
35	<b>project</b> , Intro to Azure and deploy a pipeline to Azure.
36	<b>project</b> , ethics and data security - guest lecture?
37	<b>project</b> - presentation and report
38	inspiration: spark, databricks, data lake, data factory, data pipeline on Azure. Modern data stack - study visit Knowit

School starts.  
Repetition and  
project  
preparation.



Project deadline  
with presentation  
and technical  
report.



## Today

Preparation for the project.

1. Repetition lecture
2. Repetition exercise Bash
3. Pimp my terminal
4. Bandit game

<https://overthewire.org/wargames/bandit/bandit0.html>

`ssh bandit0@bandit.labs.overthewire.org -p 2220`