Data Engineering SS25

Week 1 – Course Organization

Dr. Sucheta Ghosh

HOCH
SCHULE
OFFEN
BURG

Introduction and Organization

- Introduction
- Course layout
- Course dates
- Exam
- Lab
- ...

Introduction

- Dr. Sucheta Ghosh
  - Researcher at IWR University of Heidelberg
  -
  - Lecturer in Data Engineering, Software Eng. Database Programming

Course Layout

Hochschule Offenburg
offenburg.university

Planned: 12+ blocks

each with 2 SWS lecture + 2 SWS lab

- *Lecture*
  - Thursdays (12:00)
  - Room B122
- *Lab*
  - Thursdays
  - Room B106
  -

Communication

Hochschule Offenburg
offenburg.university

- Contact
  - *sucheta.ghosh@lehrbeaufrag.hs-offenburg.de*
  - **For technical questions: please use the anonymous Moodle forums!**

Online Course Materials

Hochschule Offenburg
offenburg.university

Please sign up to the course Moodle:

~~https://elearning.hs-offenburg.de/moodle/course/view.php?id=6852~~

There is a technical problem, I have contacted the HelpDesk

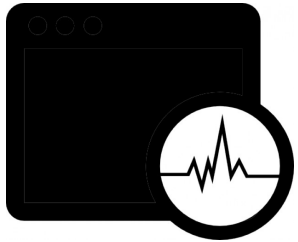
Code + Lab materials on GitHub:

https://github.com/ghoshsucheta/DataEng25

Exam

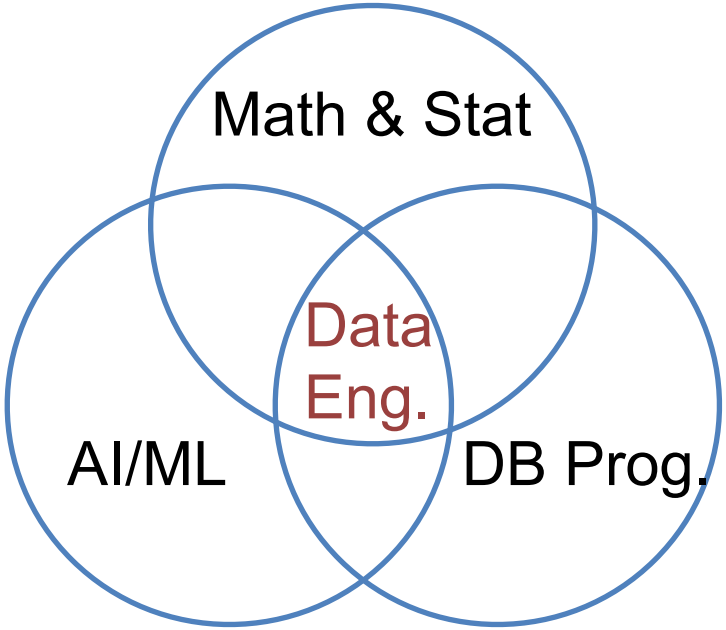
- 60/90min written exam (details later)

Lab Exercises

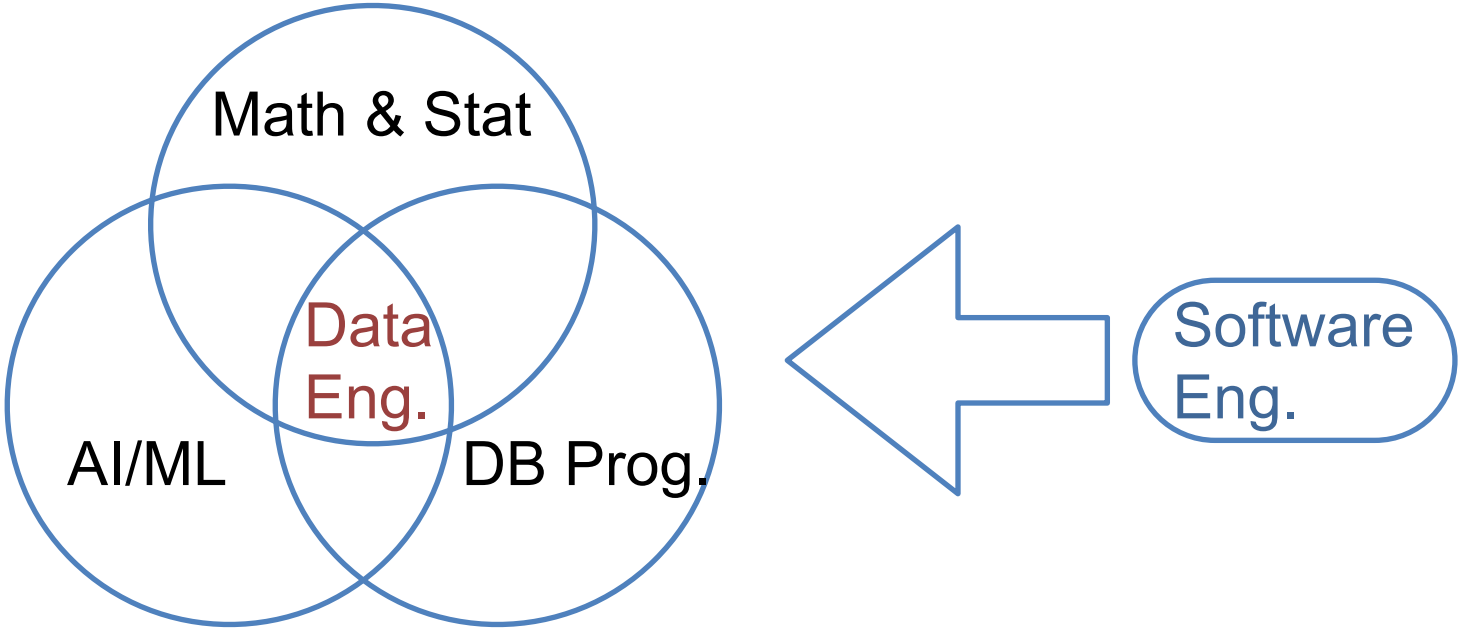


- Lab exercises + home work
  - Mandatory !
  - new exercises every Thursday
  - submit via Moodle by next Wednesday 11.59pm

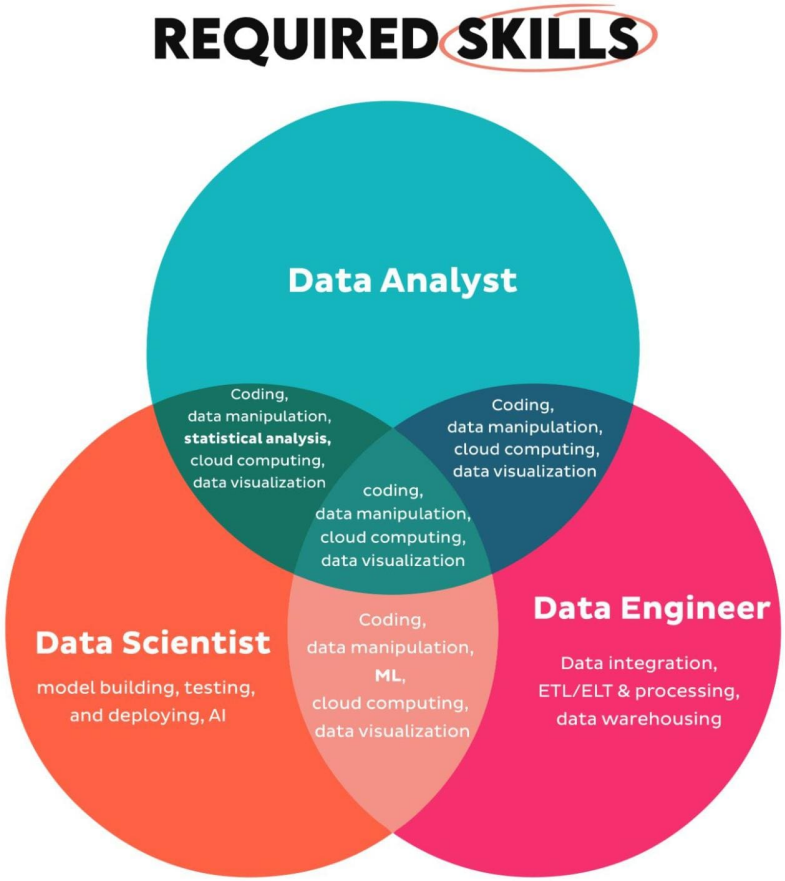
Data Engineering as a Subject



Data Engineering as a Subject



Data Engineering vs. Data Analysis vs. Data Scientist

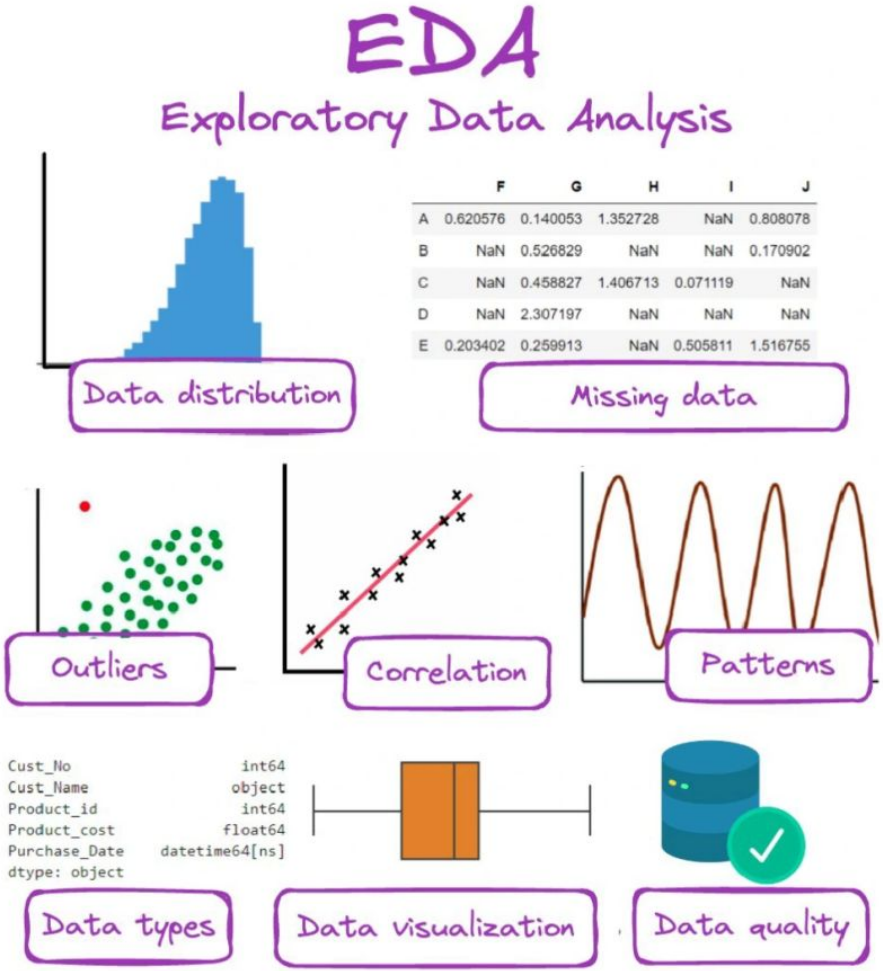


Data Engineering vs. Data Analysis vs. Data Scientist

Hochschule Offenburg
offenburg.university

Dimension	Data Analyst	Data Scientist	Data Engineer
Focus	Data analysis	Predictive models	Data infrastructure
Skills	Statistical analysis, programming in R and Python	Machine learning, programming in R and Python	Database technologies, programming languages
Responsibilities	Collecting, processing, analyzing data, making recommendations	Designing and developing predictive models, providing insights and recommendations	Designing, building, and maintaining data infrastructure, ensuring data accuracy
Technologies	SQL, R, Python	Python, R	SQL, NoSQL, Hadoop, Spark

Exploratory Data Engineering



	F	G	H	I	J
A	0.620576	0.140053	1.352728	NaN	0.808078
B	NaN	0.526829	NaN	NaN	0.170902
C	NaN	0.458827	1.406713	0.071119	NaN
D	NaN	2.307197	NaN	NaN	NaN
E	0.203402	0.259913	NaN	0.505811	1.516755

Exploratory vs Explanatory Data Engineering

Hochschule Offenburg
offenburg.university

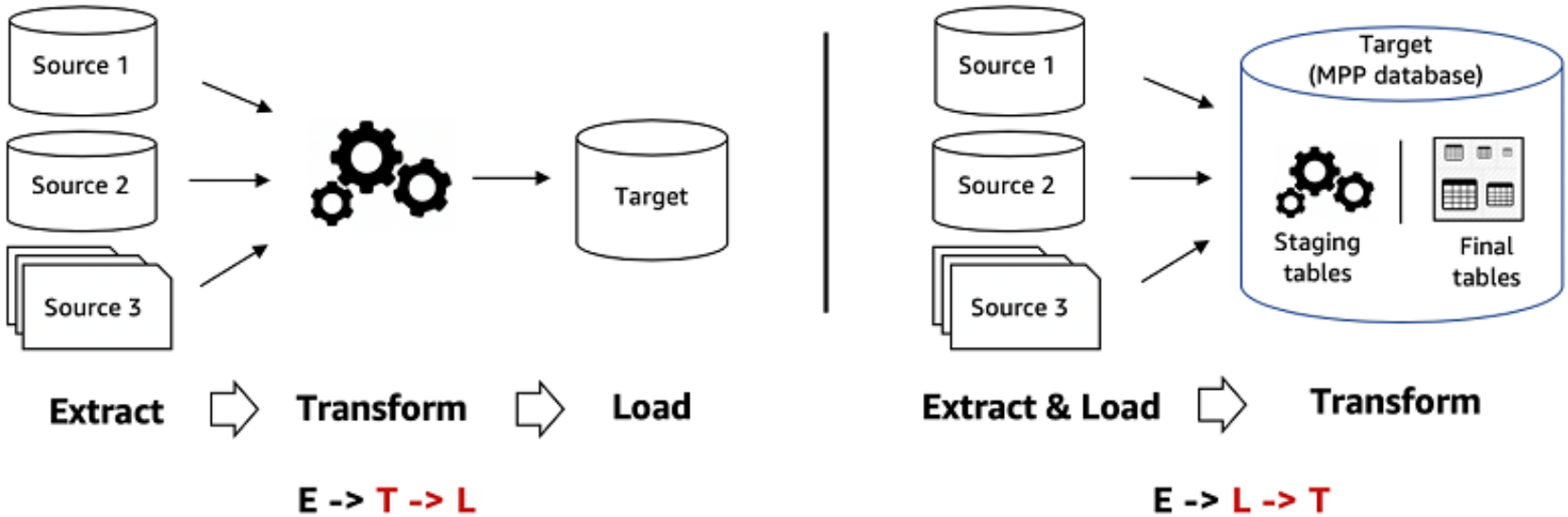
Data Visualization: Key differences between approaches

	Exploratory	Explanatory
Goal	Understand	Communicate
Audience	You	Other people
Data Familiarity	Very Familiar (You)	Less familiar (Others)
Visualization Focus	Flexibility and speed	Simplicity, clarity, and cohesion
Narrative	Unknown	Known
Outcome	Insight	Action

Effectivedatastorytelling.com

Next Part of this presentation is for practical part, now the rest of the lecture today will be borrowed from Prof. Dr.-Ing. Janis Keuper (Last Sem. week 1)

ETL vs ELT Data Engineering



Hierarchical Data Formats (HDF5)

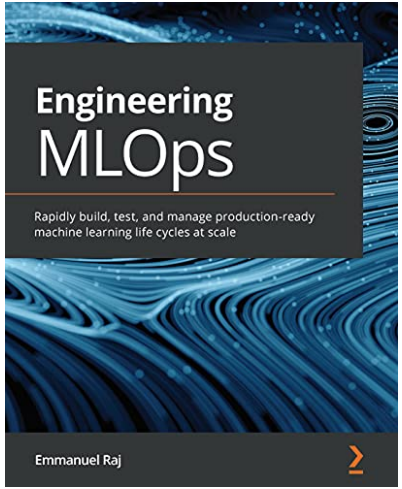
A Tutorial:

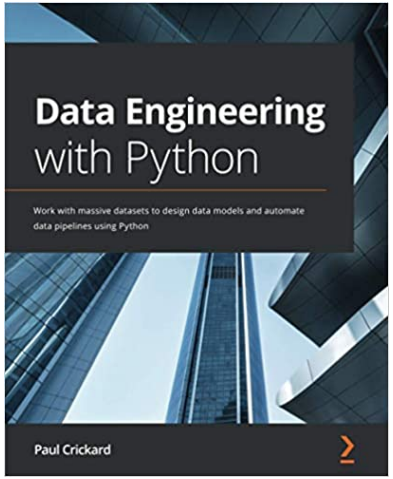
https://www.neonscience.org/resources/learning-hub/tutorials/about-hdf5

Literature

Hochschule Offenburg
offenburg.university

For now I suggest





Non of the books cover this course completely, it is NOT mandatory to have any of them.

Literature

[1] free icons taken from https://www.flaticon.com