School of Engineering
InIT Institute of Computer Science

Big Data Project (PM4)

Introduction to Course

Jonathan Fürst

Let's get to know each other

Introduction

School of Engineering
InIT Institute of Computer Science

Dr. Jonathan Fürst

- Email: fues@zhaw.ch
- Office: TD (Ob. Kirchgasse), please send an email before.
- At ZHAW since September 2022
  - Senior Lecturer: Teaching & Research at InIT (Data Integration, Data Interfaces, LLMs)

Pre-ZHAW

- 2018-2022: Research Scientist, Senior Researcher, NEC Laboratories Europe, Heidelberg
  - IoT Platform Group: Data Integration, Derive insights from IoT data, Workload orchestration between edge and cloud, Weakly-supervised ML…
- 2016-2018: PostDoc IT University of Copenhagen, Denmark
- 2013-2016: PhD in Computer Science at IT University of Copenhagen and UC Berkeley
  - Software Defined Buildings, Human-Building Interaction
- 2010-2013: MSc. in Computer Science at IT University of Copenhagen
- 2006-2010: BSc. in Industrial Engineering at Aalen University, Germany

Let’s get to know each other

Introduction

School of Engineering
InIT Institute of Computer Science

Nico Spahni

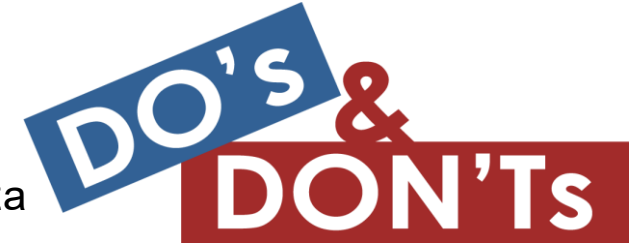
- Email: nico.spahni@zhaw.ch
- Office: SF, Institute of Language Competence
- at ZHAW since 2007
  - Senior Lecturer for Communication Competence

Course Organization

Introduction

School of Engineering
InIT Institute of Computer Science

This is a project course!



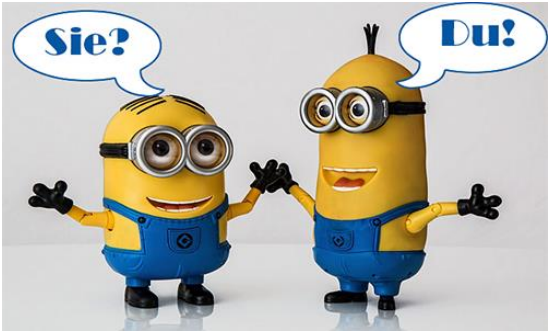
+ Lots of **hands-on** (smaller exercises/labs and mostly project)
+ Application of theories from **Data Engineering 1&2** to Big Data
+ Freedom to **choose project** based on your **own interests**
  → use opportunity to explore some area that you might continue in next semester (Projektarbeit Data Science) and for your Bachelor Thesis or job.
+ Be active during the course period, continuously work on your projects.

- Don't be passive and wait until the end of the course to somehow catch up---it won't work.
- Not typical course with lectures, exercises & written exam
- Grade will depend on how well you do in the project, don't just do the bare minimum

Course Organization

Introduction

School of Engineering
InIT Institute of Computer Science



In-person Sessions:

- Part-time students:
  - **Thursdays 08.00-09.35 TE 626**
- Full-time students:
  - **Thursdays 10.00-11.35 TE 626**
- For project supervision (both technical and from a communication side).
- Take the opportunity to **physically attend**!
- If you have a question, please interrupt anytime → better during than after class

Course Material and **Communication**

- All materials on Moodle
- No dedicated course book, but we will use some online materials.
- Git repositories for your project and project documentation.
- Teams for asynchronous communication, potentially some streaming
  Please use **Teams channel** to to ask your **questions**
  → others might have the same question

Learning Objectives
Introduction

School of Engineering
InIT Institute of Computer Science

Kurzbeschrieb
Der Schwerpunkt dieses Moduls liegt darauf, praktische Erfahrungen im Umgang mit Big-Data-Problemen zu sammeln. Das Modul basiert auf den theoretischen Grundlagen von "Data Engineering 1" und "Data Engineering 2". Die Studierenden werden ausgewählte Themen dieser Grundlagenkurse analysieren und dann skalierbare Anwendungen unter Verwendung der neuesten Big-Data-Technologien implementieren.

Ziel	Kompetenzen	Taxonomiestufen
Die Studenten vertiefen die Methoden und Tools aus anderen Kursen (z.B., Data Engineering 1 und Data Engineering 2) durch Anwendung in einem grösseren Kursprojekt.	F, M	K3, K4
Die Studenten lernen über die praktische Anwendbarkeit von Big Data Systemen (z.B., Spark) und deren Vor- und Nachteile.	F, M	K3, K4
Die Studenten sind in der Lage, Python und das dazugehörige Data-Science-Ökosystem (z.B. pandas, numpy, scikit-learn) zu nutzen und selbständig in ihrem Kursprojekt anzuwenden.	F, M	K3
Die Studenten lernen, ihre erstellten Prototypen experimentell zu bewerten und mit einer ausgewählten Basislinie zu vergleichen.	F, M	K3, K6
Die Studenten sind in der Lage, einen kompletten Projektlebenszyklus im Team zu durchlaufen, vom Projektvorschlag bis zur Projektdurchführung und Präsentation.	M, SO	K3
Die Studenten sind in der Lage, über ihr Vorwissen hinauszugehen und geeignete Technologien zu wählen, die den Problemen ihres Kursprojekts entsprechen.	SE	K3

Big Data Project Process

Introduction

School of Engineering
InIT Institute of Computer Science

1. Select an interesting **problem** to solve:
   a) **More Application-focused**. Past examples: Analyze chess game openings, Stock prediction from twitter/news data, Football player rating from game data, Snow/ice detection from satellite images, Analysis and price prediction from web-scrapped data on rental properties, Movie score prediction, Percipitation prediction, Analyse des Sprechtempos verschiedener schweizerdeutscher Dialekte, Finding Celebrity Look-Alikes…
   b) **More Technology focused**. Past examples: Scalable Question Answering with Retrievel Augmented Generation, Reducing processing time for Variant Call Format (VCF) files…
2. Select the **datasets**. This includes *potentially* data acquisition (e.g., web scrapping), data curation (cleaning, integration…), data storage (e.g., relational database).
3. **Big Data Aspects**:
   a) Discuss and potentially evaluate the **design decisions** that you made in your data processing pipeline. E.g., Why the chosen data sources? Response time/throughout requirements? Bottlenecks? Parallelization? Big Data framework needed? Database needed?...
   b) Relate the Big Data Dimensions (→ Vs) to your project.

Schedule
Introduction

- Woche 1: Willkommen
- Woche 2: Proposal Erarbeitung
- Woche 3: Proposal Präsentationen (05.03.2026)
- Woche 4: Scaling (single node)
- Woche 5: Scaling (distributed)
- Woche 6: Präsentation-Skills
- Woche 7: Projektbetreuung
- Woche 8: Midterm Präsentationen (09.04.2026))
- Woche 9: Projektbetreuung
- Woche 10: Projektbetreuung
- Woche 11: Projektbetreuung
- Woche 12: Schlusspräsentationen (07.05.2026)
- Woche 13: Unterrichtsfrei (Auffahrt)
- Woche 14: Abgabe Bericht und Github (21.05.2026)

School of Engineering
InIT Institute of Computer Science

- Updated schedule can always be found on Moodle!

Grading
Introduction

School of Engineering
InIT Institute of Computer Science

- **No written exam**, grade will be based on **project work.**
- Group size: 2-3 people (other sizes require approval)

1. Project Presentations

Proposal	(2-3 slides)	
Mid-term	(3-4 slides)	**30%**
Final	(~10 slides)	**(T+C)**

2. Project Report + Code

Final	(8 pages)	**60% (T)**

3. Blog Post

10% (C)

Overall: **75% Technical Content (T)** **(Jonathan Fürst)** **25% Communication (C)** **(Nico Spahni)**