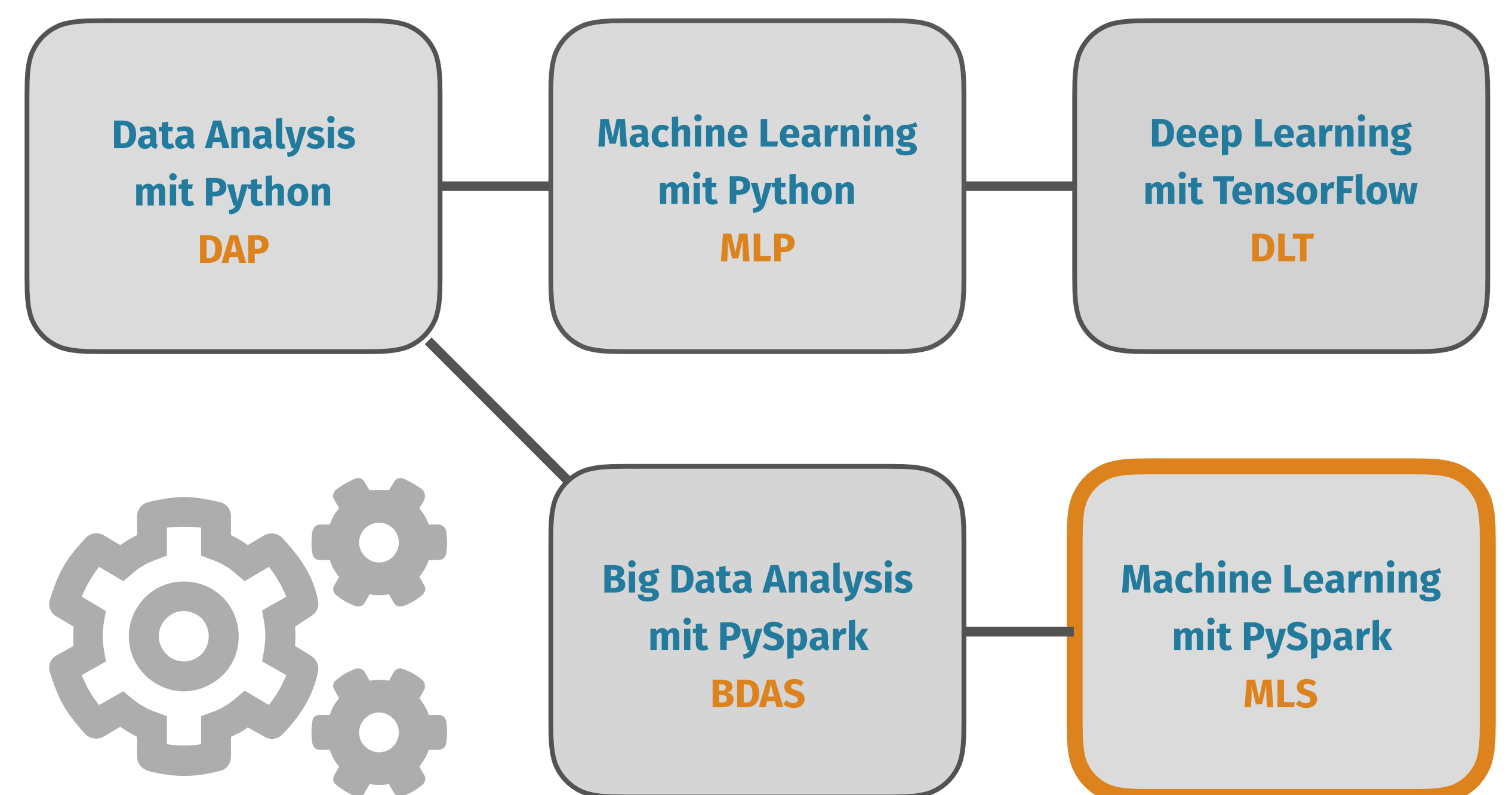


Machine Learning mit PySpark MLS

point8

data matters.

Vertiefungskurs für Machine Learning
Anwendungen auf Big Data mit PySpark.



TechTrainings für Anwender,
Umsetzer und alle, die selber
Hand anlegen wollen.

Level: ▶▶ (Fortgeschrittene)

Länge: 2 Tage

Voraussetzung: DAP + BDAS (oder vergleichbares Know-how)

Kursprache: Deutsch (Englisch auf Anfrage); Materialien auf Englisch

- 1. Introduction to Machine Learning**
An overview over the field of machine learning.
- 2. ML for Classification**
Build a classification model and learn about the building blocks of ML with PySpark
 - A. About Classification**
Learn about classifiers and how to measure the quality of their decisions.
 - B. Feature Engineering**
How to build better features.
 - C. Algorithm Selection and Hyperparameter Tuning**
Select most suitable algorithm for the job and optimize performance.
 - D. Exercise: Classifier**
Build your own classification pipeline from scratch.
- 3. ML for Regression**
Learn about regressors and how to measure the quality of their prediction.
 - A. Exercise: Predicting House Prices**
Apply regression to predict house prices.
- 4. Building a Recommender Engine**
Use machine learning to generate movie recommendations.
- 5. Unsupervised Learning: Clustering**
Apply clustering algorithms to detect structure in the data.
 - A. Exercise: Clustering**
Practice on a clustering problem.