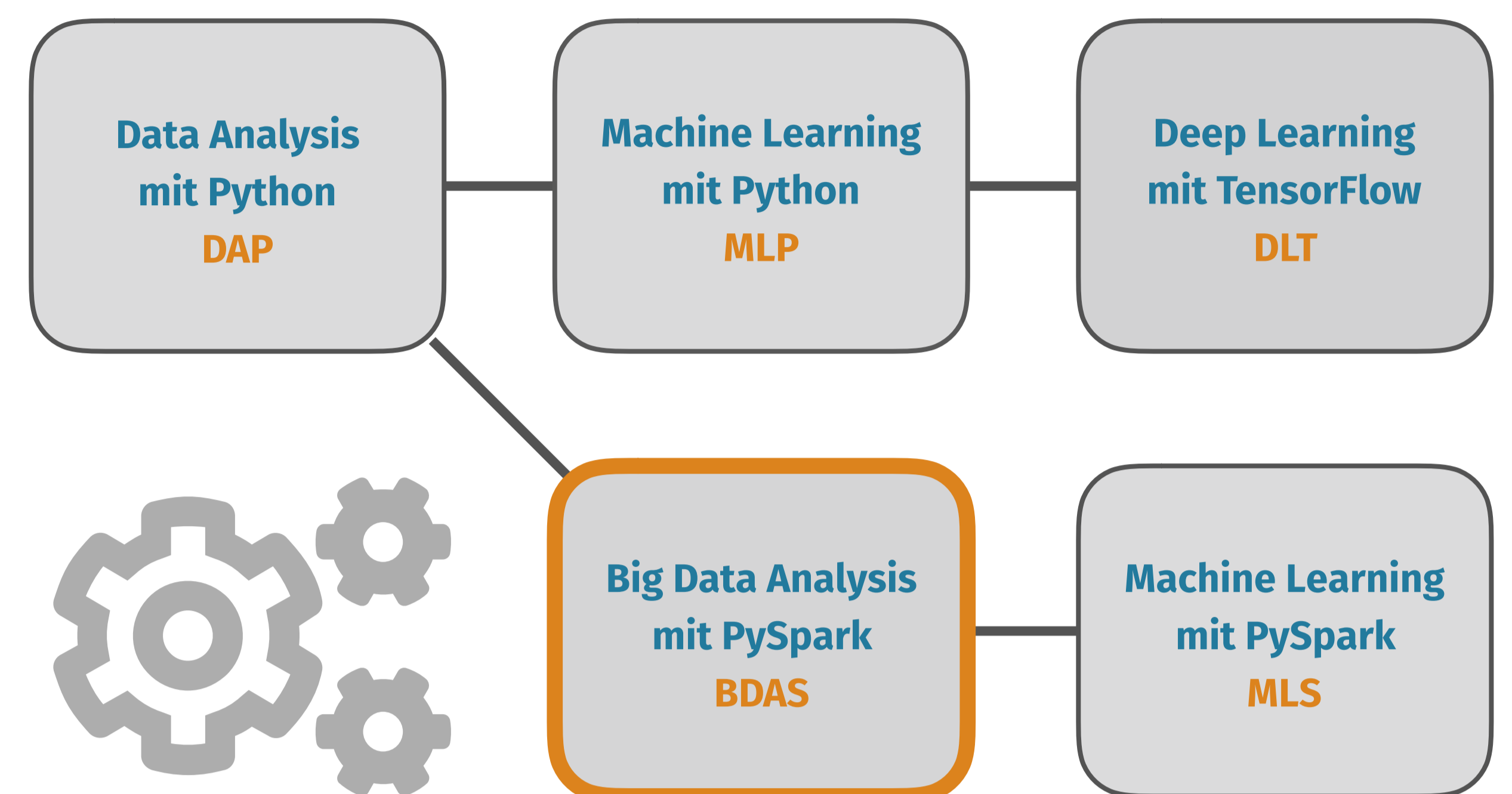


Big Data Analysis mit PySpark BDAS

point8

data matters.

Erweiterungskurs für die Skalierung der Datenanalyse auf große Datenmengen und das verteilte Rechnen auf einem Cluster mittels PySpark.



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

Level: ▶▶ (Fortgeschrittene)

Länge: 1 Tag

Voraussetzung: DAP (oder vergleichbares Know-how)

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

- 1. Processing Big Data**
Which strategies are available to compute efficiently with increasing amounts of data? What is a cluster, and when do we need one?
- 2. Spark Fundamentals**
An overview of Spark - a framework for programming distributed computation, using PySpark, its Python API - core data structures and operations.
- 3. Submitting Spark Jobs**
How to submit jobs to a Spark cluster for batch processing.
- 4. Spark and Structured Data**
Working with structured data in Spark.