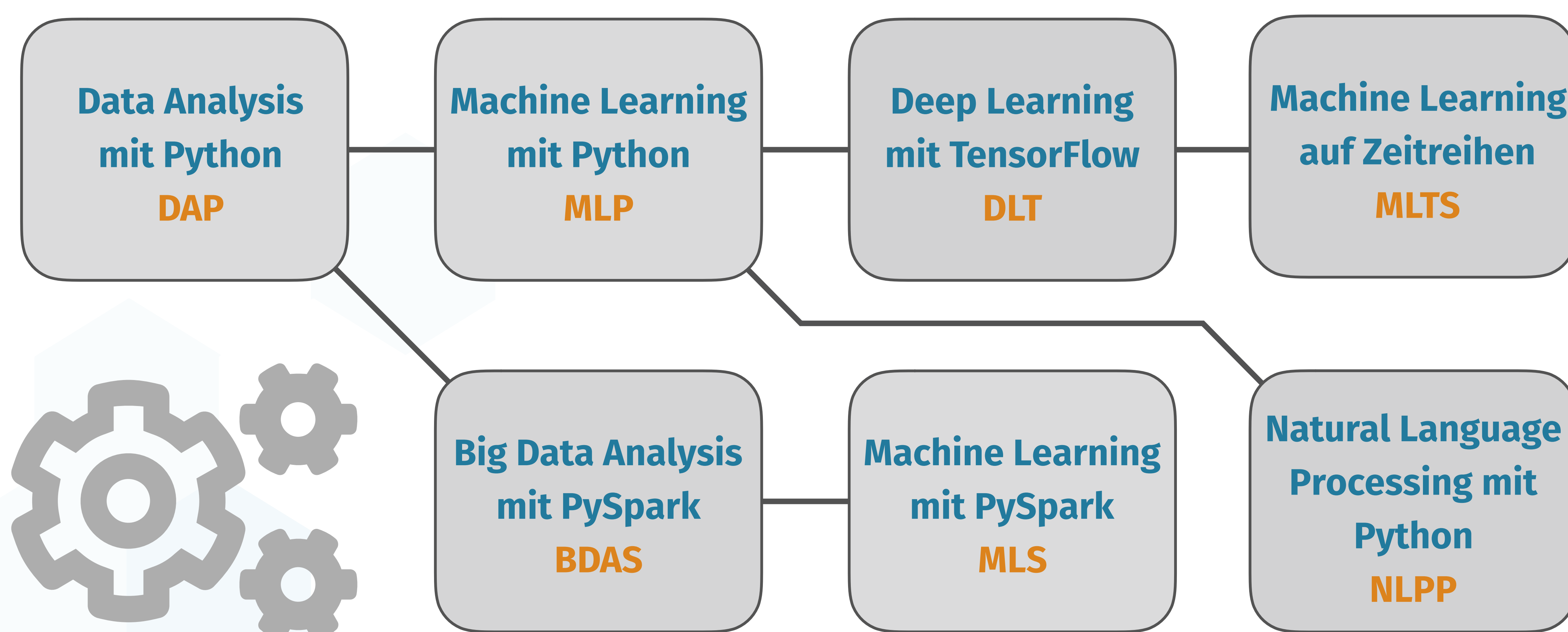


In unseren **Trainings** geht es darum, Grundlagen und Tools der Datenanalyse kennenzulernen und dabei Ihren datengetriebenen Use Case nicht aus den Augen zu lassen. Wir zeigen, wie Sie von Big Data und Datenanalyse profitieren und wie Sie Hype von Potential unterscheiden können. Alle Trainings führen wir **in einem virtuellen Raum** oder **bei Ihnen vor Ort** durch. Im Fokus steht immer die Anwendung und das Machen. Wir arbeiten interaktiv mit praktischen Hands-On Sessions. Unsere Trainings werden von erfahrenen Data Scientists gegeben, die auf eine breite Projekterfahrung sowie technisches Know-How zurückgreifen können.



Unsere **TechTrainings** sind fokussiert auf Anwender, Umsetzer und alle, die selber Hand anlegen wollen.



Data Science, ML &  
KI for Business  
DSB

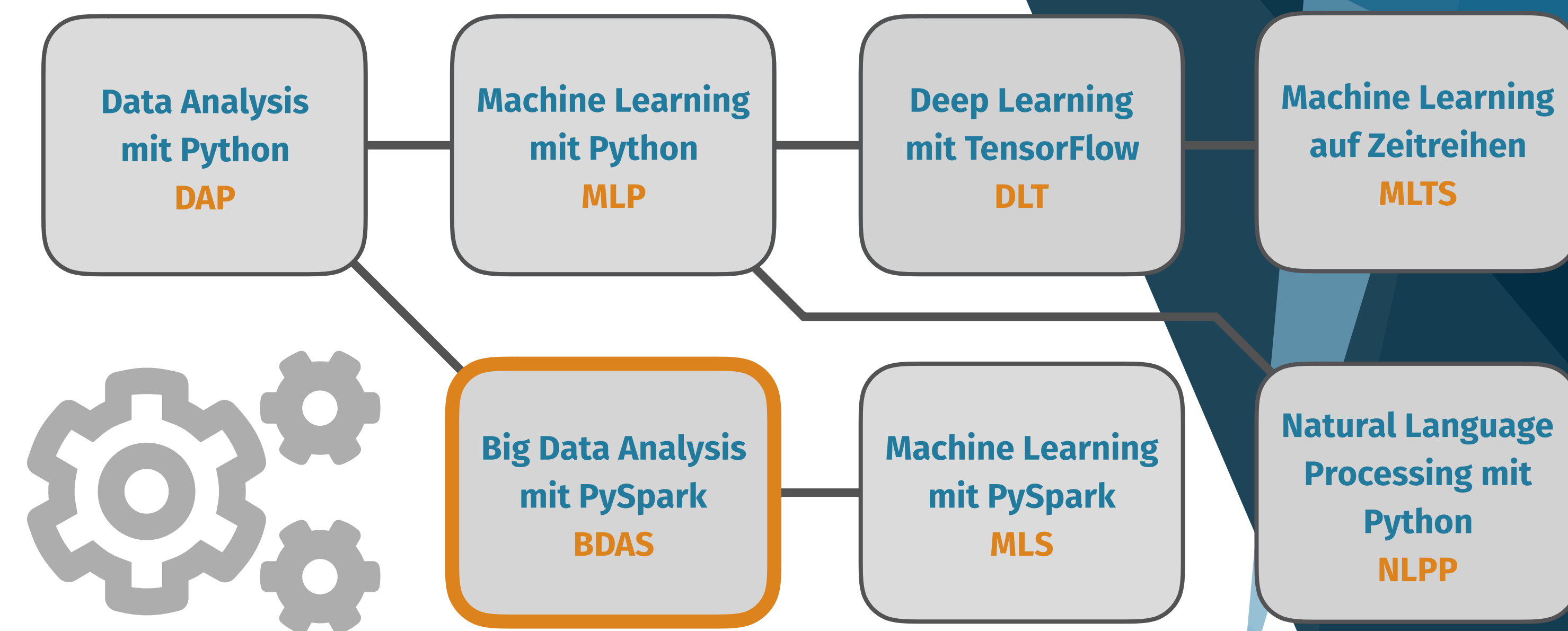
Die **BusinessTrainings** für Entscheider und Steuerer zeigen auf der etwas höheren Flughöhe, was hinter KI & Co. steckt, und wie Sie all dies zukunftsicher in Ihrem Unternehmen einsetzen können.

Wir greifen auf unsere **digitale Laborumgebung in der Cloud** zurück, die jedem Teilnehmer das individuelle und parallele Arbeiten ermöglicht. An erster Stelle stehen stets Interaktivität, Hands-On sowie die individuelle Unterstützung. Dies ermöglichen wir durch selbst entwickelte und mehrfach erprobte Trainingsmaterialien in Jupyter Notebooks, kleine Gruppengrößen sowie genug Zeit zum Üben und Selbermachen.



## Big Data Analysis mit PySpark BDAS

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:** ▶▶ (Intermediate)

**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.
5. **Exercises:**
  - A. **Museums of France**  
An exercise with a clear task, requiring you to apply the learnings from the course.
  - B. **Counting Bigrams**  
Using Spark to count bigrams in big text data.