Data Science at Scale using Spark and Hadoop (DSSH)

ID DSSH   **Preis** auf Anfrage   **Dauer** 3 Tage

Zielgruppe

- Developers
- Data analysts
- Statisticians

Voraussetzungen

- Proficiency in a scripting language
    - Python is strongly preferred
    - Perl or Ruby is sufficient
- Basic knowledge of Apache Hadoop
- Experience working in Linux environments

Kursziele

After completing this class, you will learn:

- How to identify potential business use cases where data science can provide impactful results
- How to obtain, clean and combine disparate data sources to create a coherent picture for analysis
- What statistical methods to leverage for data exploration that will provide critical insight into your data
- Where and when to leverage Hadoop streaming and Apache Spark for data science pipelines
- What machine learning technique to use for a particular data science project
- How to implement and manage recommenders using Spark's MLlib, and how to set up and evaluate data experiments
- What are the pitfalls of deploying new analytics projects to production, at scale

Kursinhalt

Data Science at Scale using Spark and Hadoop is a 3 day instructor-led class where you will learn how scientists use data to solve problems by understanding the tools and techniques they use. Through in-class simulations, participants apply data science methods to real-world challenges in different industries and prepare for data scientist roles in the field.

Data Science at Scale using Spark and Hadoop
(DSSH)

Weltweite Trainingscenter

Fast Lane Institute for Knowledge Transfer GmbH

Husacherstrasse 3
CH-8304 Wallisellen
Tel. +41 44 832 50 80

info@flane.ch, https://www.flane.ch