Cloudera Introduction to Data Science: Building Recommender Systems (CIDS)

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

Zielgruppe

This course is suitable for developers, data analysts, and statisticians.

Voraussetzungen

This course is suitable for developers, data analysts, and statisticians with basic knowledge of Apache Hadoop: HDFS, MapReduce, Hadoop Streaming, and Apache Hive as well as experience working in Linux environments. Students should have proficiency in a scripting language; Python is strongly preferred, but familiarity with Perl or Ruby is sufficient.

Kursziele

Data scientists build information platforms to provide deep insight and answer previously unimaginable questions. Spark and Hadoop are transforming how data scientists work by allowing interactive and iterative data analysis at scale. Learn how Spark and Hadoop enable data scientists to help companies reduce costs, increase profits, improve products, retain customers, and identify new opportunities.

Cloudera University's three-day course helps participants understand what data scientists do, the problems they solve, and the tools and techniques they use. Through in-class simulations, participants apply data science methods to real-world challenges in different industries and, ultimately, prepare for data scientist roles in the field.

Through instructor-led discussion and interactive, hands-on exercises, participants will navigate the Hadoop ecosystem, and develop concrete skills such as:

- 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

- Introduction
- Data Science Overview
- Use Cases
- Project Lifecycle
- Data Acquisition
- Evaluating Input Data
- Data Transformation
- Data Analysis and Statistical Methods
- Fundamentals of Machine Learning
- Recommender Overview
- Introduction to Apache Spark and MLlib
- Implementing Recommenders with MLlib
- Experimentation and Evaluation
- Production Deployment and Beyond
- Conclusion

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