

# ) Seminar )

## Datenintegration (ETL) mit Kettle

Pentaho Data Integration PDI (besser bekannt unter dem Namen Kettle) ist ein modular aufgebautes und erweiterbares Open Source Werkzeug zur Datenintegration (ETL). Es ermöglicht das Extrahieren, Transformieren und Laden großer Datenmengen und dient deshalb oftmals als Middleware für Data-Warehouses. Mit Hilfe des grafischen Editors Spoon können komplexe Datentransformationen aus einer vordefinierten Menge, sogenannter Schritte (Steps), zusammengeklickt und in einem zentralen Datenbankrepository oder in einer XML Datei abgespeichert und über mitgelieferte Werkzeuge ausgeführt werden. Das Seminar führt systematisch anhand praktischer Beispiele in das Thema Datenintegration mit Kettle ein und zeigt die Möglichkeiten dieser Technologie auf.

### Seminar-Inhalt

#### Einführung

- Einführung in Jobs und Transformationen
- Rundgang durch die Datenküche und kennenlernen ihrer Werkzeuge
- Einführung in die Entwicklungsumgebung Spoon
- Anbindung von Datenquellen
- Einrichten eines zentralen Repositories

#### Grundlagen

- Kennenlernen der einzelnen Schritttypen
- Abrufen von Daten aus verschiedenen Quellen
- Steuern des Datenfluss
- Filtern, Gruppieren und Aggregieren von Datensätzen anhand von Bedingungen
- Normalisieren von Daten mit Hilfe von Lookups
- Daten anreichern
- Features zur Sicherstellung der Datenqualität und Datenbereinigung
- Laden der Daten in die Zieldatenbank

#### Fortgeschrittene Features für ETL

- Überwachung und Orchestrierung von Transformationen mit Hilfe von Jobs
- Debugging und Monitoring von Jobs und Transformationen
- Fehlerbehandlung und Logging
- Parametrisieren von Transformationen mit Hilfe von Variablen und Argumenten
- Wiederverwendung und Gliederung durch Einsatz von Sub-Transformationen
- Verwendung von Java und JavaScript in Transformationen
- Kennenlernen weiterer Features zur Performance Optimierung

#### Ausführen von Transformationen und Jobs auf dem Server

- Kennenlernen der Kommandozeilentools
- Ausführen von Jobs mit Hilfe von kitcher
- Ausführen von Transformationen mit Hilfe von pan
- Deployment der ETL-Lösung auf dem Server

### Zielgruppe

Als Entwickler, Berater oder Projektleiter möchten Sie die Probleme Datenintegration und Datenmigration datenbankübergreifend lösen. Als technischer Entscheider wollen sie das Potential von Pentaho Data Integration (PDI) kennenlernen.

### Voraussetzungen für Seminar-Teilnahme

SQL-Grundkenntnisse sind erforderlich. Kenntnisse in Datenbankdesign sind von Vorteil.

### Seminar-Dauer

2 Tage

### Vorteile einer Seminar-Teilnahme

- Nach dem Seminar sind Sie in der Lage, komplexe Datentransformationen zu gestalten und auszuführen
- Sie ersparen sich lange Einarbeitungszeiten, die aufgrund fehlender Literatur zum Thema anfallen
- Lernen Sie ein mächtiges Open Source Werkzeug kennen, das Ihnen viele händische IT abnimmt.
- Unsere Kurs-Unterlagen sind immer auf dem aktuellsten Stand

### Seminar-Preis

1.230 EUR pro Person  
(inkl. Unterlagen u. Tagesverpflegung zzgl. MwSt.)

### Seminar-Termine

14.9. - 15.9.2017

### Inhouse-Kurse

Alle unsere Seminare können wir Ihnen auch für eine Durchführung in Ihrem Hause mit einer speziellen Kalkulation für Exklusiv-Seminare anbieten. Hierbei können die Inhalte aller unserer Seminare beliebig für Ihr individuelles Training zu einem Wunschseminar zusammengestellt werden.