

# Zalando nutzt Pentaho als Datenintegrationstool

Grafische Benutzeroberfläche vereinfacht ETL-Prozess für Business Analysten

## Kunde

Zalando ist einer von Europas führenden Online-Anbietern für Schuhe und Fashion. Das Unternehmen wurde 2008 von Robert Gentz und David Schneider gegründet und hat seinen Sitz in Berlin. Nach dem großen Erfolg in Deutschland werden seit 2009 mittlerweile insgesamt 15 europäische Länder beliefert.

## Herausforderungen

Zalando suchte ein Datenintegrationstool, um die Daten aus den verschiedenen operativen Systemen des Unternehmens, aber auch aus anderen externen Datenquellen, ins zentrale Data Warehouse (DWH) zu laden. Da Zalando mit vielen, oftmals selbst entwickelten, maßgeschneiderten Systemen arbeitet, war eine hohe Konnektivität zu unterschiedlichen Quell-Systemen ein wichtiges Kriterium. Da viele Systeme des Unternehmens auf Open Source Technologien basieren, suchte man nach einer entsprechenden Open Source Lösung. Darüber hinaus sollte das Datenintegrationstool einfach zu bedienen sein, sodass auch Business Analysten ohne größere Programmierkenntnisse auf die Datenquellen zugreifen können.

## Lösung

Zalando entschied sich zunächst für die Community Edition von Pentaho Data Integration aka Kettle und migrierte dann zur Enterprise Edition PDI, die über umfassendere Funktionalitäten verfügt und es zum Beispiel ermöglicht, Extract Transform Load (ETL) Jobs und Transformationen zentral zu speichern und zu managen sowie eine detaillierte Einsicht in Änderungsverläufe zu erhalten. Aufgrund seiner Funktionen zum Teilen, Blockieren, Einsammeln und Abfragen von Log-Daten, erleichtert PDI darüber hinaus das Überwachen von ETL-Jobs.



### FAKTEN

- Branche: eCommerce
- Unternehmenssitz: Berlin
- Internetseite: [www.zalando.de](http://www.zalando.de)
- Produkte: Pentaho Data Integration

## Ergebnisse

Zalando nutzt PDI seit 2012 als Datenintegrations- und ETL-Tool in einer klassischen Data Warehouse Umgebung mit 25-30 verschiedenen IT-Systemen, die vorwiegend auf PostgreSQL basieren und die gesamte Wertschöpfungskette des Unternehmens abdecken. Die Daten werden je nach Bedarf ins DWH geladen und die Anwendung reicht von klassischen nächtlichen Batch-Prozessen bis zum viertelstündlichen Laden.

Aufgrund seiner vielen Schnittstellen und offener Standards ermöglicht PDI einen reibungslosen ETL-Prozess, bei dem auch neue Systeme einfach und schnell eingebunden werden können. So hat das BI-Team eine wiederverwendbare Workflow-Blaupause kreiert, die es ermöglicht, den ETL-Prozess zu automatisieren, sodass auch komplett neue Datenquellen innerhalb von kurzer Zeit (1-2 Stunden) ins DWH geladen werden können.

Pentahos Konnektivität ermöglicht uns, die verschiedensten Datenquellen ins zentrale Data Warehouse zu laden. Die grafische Entwicklungsoberfläche von PDI macht es nicht nur einfach, ETL-Prozesse zu automatisieren, sondern ermöglicht es auch Business Analysten ohne größere Programmierkenntnisse Datentransformationen durchzuführen.

- ERIC VON CZAPIEWSKI  
Head of Business Intelligence, Zalando

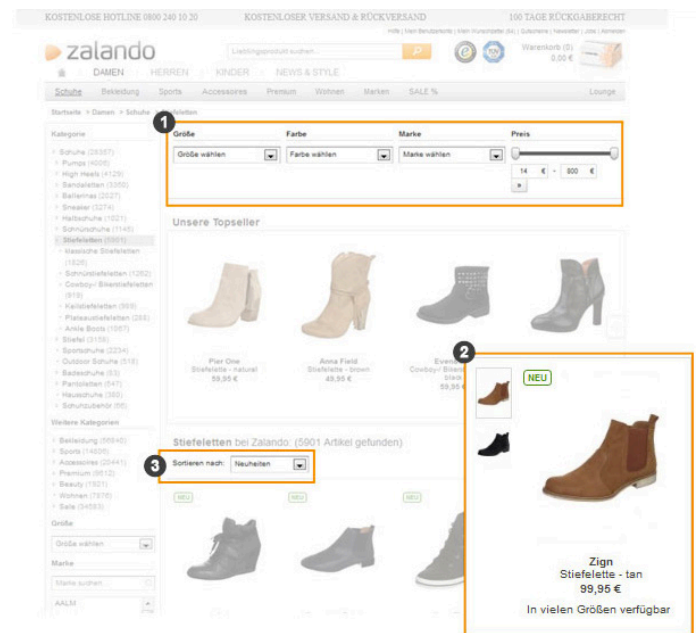
Pentahos offene Standards ermöglichen ferner, dass das BI-Team bereits relativ früh eigene PDI-Scripts für die Integration von MongoDB-Daten erstellen konnte.

Dank der grafischen Benutzeroberfläche von PDI können auch Business Analysten, ohne über große Programmierkenntnisse zu verfügen, Daten-Transformationen durchführen, sind also unabhängiger von den Datenintegrationsentwicklern. So ist es kein Wunder, dass sich unter Zalandos PDI-Nutzern gleichviel Business Analysten wie Datenintegrationsentwickler finden.

Zalando nutzt zudem Pentahos Reporting-Funktionen, um Berichte zum Status der Quellsysteme und zu den Datenintegrationsprozessen zu erstellen

## Zusammenfassung

Zalando nutzt seit 2012 Pentaho Data Integration (PDI) als Datenintegrations- und ETL-Tool in einer klassischen Data Warehouse Umgebung mit 25-30 verschiedenen IT-Systemen, die zumeist auf PostgreSQL basieren. Aufgrund seiner vielen Schnittstellen und offenen Standards ermöglicht PDI einen reibungslosen ETL-Prozess, bei dem auch neue Systeme einfach und schnell eingebunden werden können. So hat Zalandos BI-Team eine wiederverwendbare Workflow-Blaupause kreiert, die es ermöglicht, den ETL-Prozess zu automatisieren, sodass auch komplett neue Datenquellen



innerhalb von kurzer Zeit ins DWH geladen werden können. Pentahos offene Standards ermöglichten, dass das BI-Team bereits relativ früh eigene Scripts für die Integration von MongoDB-Daten erstellen konnte. Dank der grafischen Benutzeroberfläche von PDI ist die Datentransformationen besonders leicht.