

Visual Rules in Hadoop



Apache Hadoop (hadoop.apache.org/) ist ein Java-Framework für die Verarbeitung von Software, die nach dem Funktionsprinzip: Map-Reduce (de.wikipedia.org/wiki/MapReduce) entwickelt wurde. Prinzipiell teilt man eine Aufgabenstellung in kleine Portionen auf, bildet diese auf bestimmte Funktionen (map-functions) ab und fasst die erzeugten Zwischenergebnisse mit Reduzier-Funktionen (reduce-functions) zu einem Gesamtergebnis zusammen. Dieser Map-Reduce-Ansatz ist für die leistungsfähige Verarbeitung von großen Datenmengen (Petabytes) geeignet. 2008 und 2009 hat Hadoop als erste Java-Plattform den Terabyte Sort Benchmark gewonnen. Auch Amazon bietet Hadoop in ihrer Cloud Infrastruktur an (aws.amazon.com/elasticmapreduce/), was eine sofortige Nutzung ermöglicht.

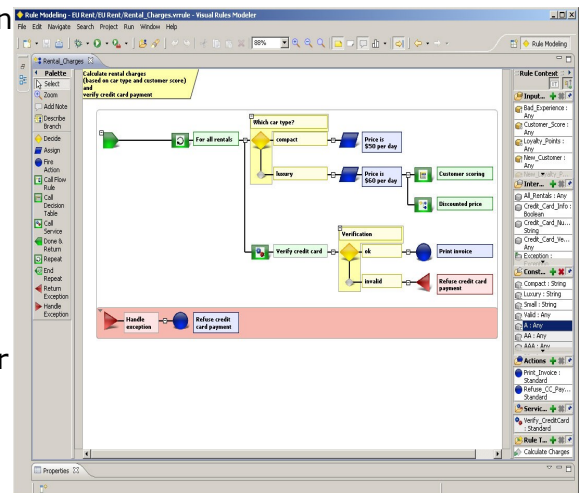
Fachexperten Experten haben erfahrungsgemäß Schwierigkeiten bei der Verwendung solch mächtiger technischer Werkzeuge zur Lösung ihrer fachlichen Fragestellungen. Um sie trotzdem von den Vorteilen der neuen Technologien profitieren lassen zu können, müssen zwei Ziele erreicht werden:

1. Vereinfachung der Beschreibung, wie Daten transformiert und abgefragt werden
2. Optimierung der Abfragezeiten, so dass in kurzer Zeit große Datenmengen verarbeitet werden können

Visual Rules (www.visual-rules.de), ein Software-Produkt der Firma Innovations (www.innovations.de), dient zur Analyse von Geschäftsdaten mithilfe von grafisch von Fachexperten modellierten Regeln (siehe nebenstehende Abbildung).



Diese beiden Technologien bilden eine perfekte Symbiose für die regelbasierte Verarbeitung extrem großer Datenmengen. Der regel-basierte Ansatz kann sehr einfach zur Beschreibung von Daten Mappings, Aggregationen und Bearbeitung von Spezialfällen verwendet werden. Fachexperten werden in die Lage versetzt, ihre Regeln zu definieren, zu testen und in ablauffähiger Form in einer Hadoop-basierten Infrastruktur zu deployen. Dieser Ansatz nutzt auf sehr vorteilhafte Weise die Potentiale des Cloud Computings und vereinigt diese mit den Ideen und Vorstellungen von Fachexperten, um sich einen Überblick über Marktdaten zu verschaffen und schneller auf Änderungen an Märkten reagieren zu können.



Aufgabenstellung: Mit Visual Rules erstellte Regeln sollen in einer Map-Reduce-Infrastruktur zum Einsatz gebracht werden. Dabei bekommen Sie sehr gute Hilfestellung durch die Firma.

Technologien: Java, Map-Reduce, Hadoop, Visual Rules

Ziel des Projektes:

- Aufbau einer Map-Reduce-Infrastruktur mit Hilfe von Hadoop
- Erstellung von Visual Rules Regeln zur Verwendung als Map-Reduce Programme
- Test und Dokumentation des Ansatzes

Teamgröße: 2-5 Studenten

Betreuer: Mike Haller/Markus Schärtel, Prof. Dr. Ch. Reich