

Geoanalysen in fünf Minuten mit Oracle R

Detlef E. Schröder
Oracle Deutschland B.V. & Co KG
Düsseldorf

Schlüsselworte

Analysen, Geodaten, R, Oracle Datenbank, einfache Umsetzung, schnelle Ergebnisse

Einleitung

Die meisten Daten die uns im Unternehmensalltag begegnen haben einen Bezug zu dem Ort wo sie entstanden sind, oder dem Ort der Person oder Verkaufsfläche, dem Lager oder einer anderen Organisationseinheit.

Oft werden diese Ortsangaben zwar als Text verwendet und auch in Hierarchien zur Gliederung genutzt, aber noch immer ist es selten, dass diese Informationen für die Analyse und die Berichterstellung verwendet werden.

In diesem Vortrag geht es daher darum zu zeigen, wie einfach es mit Oracle und R ist, aus den ortbezogenen Daten eine Geoanalyse zu gestalten. Dazu werden Daten aus der Oracle DB verwendet und in Oracle R Enterprise zur Geovisualisierung verwendet.

Das Skript steht im Anschluss auch zur eigenen Verwendung zur Verfügung.

Daten mit Ortsbezug

Daten entstehen nicht im ortsleeren Raum, sondern haben nicht zuletzt durch den Ort ihrer Entstehung einen Ortsbezug. Oft sind aber noch viele weitere Ortsbezüge vorhanden, da auch die dimensionalen Attribute weitere Ortsbezüge mit zu den Daten beitragen. Selbst im Internet und auf dem rein Digitalen Weg sind Ortsbezüge zu erfassen und einfach zu verwenden.

In dem Beispiel des Starschemas der Oracle DWH Community ist kein Ortsbezug in den Kundenadressen gegeben. Natürlich liegt hier ein Wohnort vor. Dieser soll zur Analyse und zur Visualisierung aber auch verwendet werden und dies nicht nur als Text einer Hierarchiestufe.

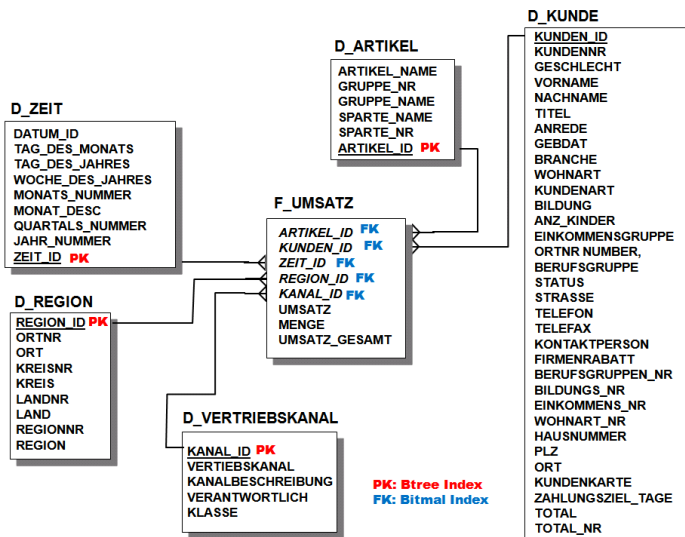


Abb. 1: Starschema für viele Beispiele dem DWH Community (www.oraclewh.de)

Diese Tabelle D_Kunde dient als Basis für die Ortsdarstellung.

Oracle R Enterprise

R ist nicht nur im statistischen Umfeld sehr beliebte Sprache zur Analyse von Daten. Der Vorteil von R liegt in seiner weiten Verbreitung und der dardurch umfangreichen Pakete – Programm Bausteine - die bei der Erledigung der Arbeit sehr hilfreich sind.

Auch für die Geovisualisierung liegen verschiedenen Pakete vor, die verwendet werden können. In diesem Beispiel wird auf GeoR zurückgegriffen und dieses Paket verwendet.

Der Zugriff auf die Tabelle in der Datenbank wird über R Oracle hergestellt und aus dem Paket Oracle R Enterprise (kurz ORE) verwendet. Dadurch ist der Zugriff einfach sicher und funktionsreich. Dazu kann zur Analyse und Verarbeitung der Daten drei Positionen verwendet werden. Position eins ist das R auf dem Client. Dazu müssen alle Daten auf den Client übertragen werden, was oft nicht geht. Position zwei ist die Transparente Abarbeitung von R Befehlen in der Datenbank und Position drei ist die Verwendung von R auf dem Datenbank Server, was schneller, sicherer und transparenter funktioniert. Dies bietet sich für spezielle statistische Berechnungen an, die so in der DB mit SQL nicht möglich sind.

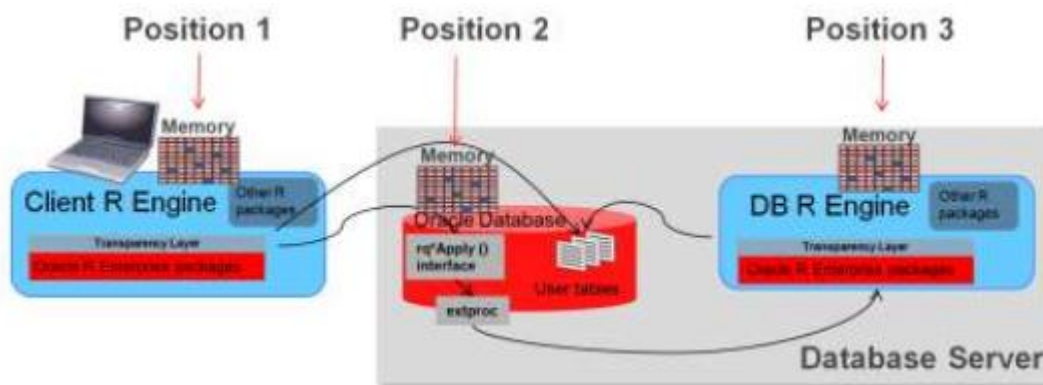


Abb.2: Arbeiten mit R und der Datenbank

Die Visualisierung kann entsprechend ebenfalls in Position eins oder drei stattfinden. Dazu werden nur die notwendigen Daten weitergeleitet.

Dies soll in dem Beispiel mit den Daten der D_Kunden Tabelle geschehen und die Wohnorte einfach geografisch dargestellt werden, auf einer Landkarte.

Diese Grafik kann dann exportiert oder dynamisiert werden und um weitere Komponenten aus der Datenbank ergänzt werden, z.B. Umsatz je Wohnort, Wohnortattribute,

Das komplette Skript wird nach der Veranstaltung auf www.oracledwh.de veröffentlicht und kann für eigene Analysen anschließend verwendet werden.

Kontaktadresse:

Detlef E. Schröder
Oracle Deutschland B.V. & Co KG
Riesstr. 25
D-80992 München

Telefon: +49 (0) 40-89091-423
E-Mail Detlef.E.Schroeder@oracle.com
Internet: www.oracle.de