

**Analytische Funktionen in Oracle 12c**  
**Christian Borghardt**  
**areto consulting gmbh**  
**Schanzenstraße 6-20**  
**51063 Köln**

**Schlüsselworte**

Datenbank, Oracle, Oracle 12c, Analytische Funktionen, Pattern Matching, MATCH\_RECOGNIZE

**Einleitung**

Der Umfang der Daten, welche in einer Datenbank gesichert werden, wird im Rahmen von Big Data und Predictive Analytics immer umfangreicher und größer. Deshalb wird es immer wichtiger, dass die Funktionalität der Analysen ebenfalls umfangreicher und performanter wird. Aktuell ist die Interpolation von Reihen und Folgen nur mittels eigens für die Problemstellung erstellten Skripten und Funktionen möglich. Diese können jedoch nur selten auf andere Aufgabenstellungen übertragen werden.

Durch die neue Datenbankversion 12c liefert Oracle eine analytische Funktion, die dieses Problem nun systematisch angeht, denn durch MATCH\_RECOGNIZE stellt Oracle nun eine Interpolationsmöglichkeit zur Verfügung, welche einheitlich Daten analysieren kann. Dank der neuen Funktionalität ist dies nun auch performant möglich, da die Datenbank alle internen Funktionalitäten nutzen kann, welche zur Verfügung stehen. Zusätzlich ist durch den standardisierten Prozess die Skalierbarkeit gewährleistet. Bei den Eigenkonstruktionen war dies nicht immer der Fall.

Anhand eines Beispiels wird gezeigt, wie die Funktionalität genutzt werden kann und was diese nun liefert. Es wird der Unterschied zwischen einem selbst geschriebenen Skript und der Funktion MATCH\_RECOGNIZE gezeigt und Möglichkeiten zum Einsatz präsentiert.

## **Analytische Funktionen**

An Datenbanken werden aus technischer und fachlicher Sicht verschiedenste Anforderungen gestellt. Es wird nicht nur eine Datendrehscheibe benötigt, welche die vorhandenen Daten sammelt, es werden ebenfalls Informationen gesucht, die in den Daten stecken. Anhand der vorhandenen Daten werden Anhaltspunkte erfragt, um weitere Entscheidungen für das Geschäft zu fällen. Für diese Analyse stellt Oracle standardisierte Funktionen zur Verfügung. Analysen können nun systematisch und performant durchgeführt werden. Analytische Funktionen durchsuchen eine Datenmenge nach vorgegebenen Kriterien und stellen die Ergebnismenge übersichtlich dar, sodass die Informationen optimal ausgewertet werden können.

Im Laufe der Zeit müssen diese Funktionen erweitert werden oder neue Möglichkeiten entwickelt werden, um noch mehr aus den vorhandenen Daten zu gewinnen. Zusätzlich verändern sich die Quellen und es werden zusätzliche Wege der Interpretation gesehen, sodass eine Erweiterung der Funktionen nötig ist. So liefert Oracle 12c nun eine weitere Möglichkeit mit MATCH\_RECOGNIZE und zeigt eine Interpolation von Datenmengen. Die Analyse war bereits schon vorher möglich, jedoch wurde dazu zu jedem Problem eine individuelle Lösung gefunden, welche nicht auf andere Probleme übertragbar war. Zusätzlich waren diese Lösungen schwer wartbar und waren nur sehr schlecht optimierungsfähig. So musste Mehrarbeit geleistet werden, welche nun durch die Datenbankfunktion entfällt. Skalierbarkeit und Übertragbarkeit werden nun durch die Datenbank gegeben.

## **MATCH\_RECOGNIZE**

MATCH\_RECOGNIZE erkennt Muster innerhalb von Datenblöcken und kann zudem angeben, wie sich eine Datenfolge entwickelt. So können Veränderungen in den Daten, wie z.B. von einer Laufzeitanalyse gemacht werden und es wird innerhalb von einem Statement gezeigt, wie sich die Dauer der ETL-Prozesse entwickelt.

Es werden 4 Schritte für die Analyse benötigt:

- 1) Definition der Partitionen und Ordnungsreihenfolge  
Die Partitionen werden mittels PARTITION BY definiert.  
ORDER BY gibt die Reihenfolge an, in der die Mustererkennung stattfindet.
- 2) Definition der Muster, Variablen und Auswahlkriterien  
PATTERN gibt den Ausdruck für die Muster an und die Reihenfolge, in der diese auftreten sollen. PATTERN ist abhängig von den definierten Ausdrücken.  
DEFINE benennt die Ausdrücke und definiert deren Kriterien.
- 3) Definition der Messwerte: Datenpunkte, Musterpunkte und deren Beziehung  
MEASURES zeigt die individuellen Datenpunkte in den Mustern an und definiert die Bedingungen für das Anzeigen.
- 4) Definition der Ausgabe  
[ ALL ROWS | ONE ROW ] PER MATCH bestimmt, ob für jeden Match eine einzelne Zeile wiedergegeben wird, oder ob diese aggregiert werden sollen in einer Zeile.

An einem Beispiel werden diese Optionen gezeigt und im Anschluss im Detail durchgegangen.

**Kontaktadresse:**

Christian Borghardt  
areto consulting gmbh  
Schanzenstraße 6-20  
D-51063 Köln

Telefon: +49 (0) 221 66 95 75 - 0  
Fax: +49 (0) 221 66 95 75 - 99  
E-Mail: [Christian.Borghardt@areto-consulting.de](mailto:Christian.Borghardt@areto-consulting.de)  
Internet: [www.areto-consulting.de](http://www.areto-consulting.de)