

Verschlagwortungstabelle

[Stand: 23.05.2019 | Programmversion: 12.00.000]

Die Suche über mehrere Indexfelder kann sehr lange dauern, wenn der SQL-Optimizer die starken Selektionskriterien in der Abfrage nicht erkennt. Die Organisation der Indexwerte als Verschlagwortungstabelle kann hier die Performance verbessern. Im Gegensatz zur „objkeys“-Tabelle sind dabei die Indexwerte in den Spalten einer DB-Tabelle statt in den Zeilen abgelegt.



Information: Diese Dokumentation gilt für die Programmversionen ab ELO 10.

Inhalt

| | | |
|-----|--|----|
| 1 | Problem | 2 |
| 2 | Lösungskonzept | 3 |
| 3 | Realisierung | 4 |
| 3.1 | Neue Eigenschaft einer Verschlagwortungsmaske: Datenorganisation | 4 |
| 3.2 | Konvertierung vorhandener Indexdaten | 4 |
| 3.3 | Darstellung in der Datenbank | 6 |
| | 3.3.2 Datenbankindizes | 6 |
| | 3.3.3 Case-insensitive-Suche bei ORACLE | 6 |
| 3.4 | Bearbeiten der Maskendefinition | 7 |
| 3.5 | Suchen über Verschlagwortungstabelle mit findfirstSords | 8 |
| 4 | Auswirkung auf andere ELO Module | 9 |
| 4.1 | Indexserver-Client-Anwendungen, ELO Java Client, ELO Web Client | 9 |
| 4.2 | ELO Windows Client | 9 |
| 4.3 | ELO Replikation | 9 |
| 5 | Einschränkungen | 10 |

1 Problem

Die Suche über die *objkeys*-Tabelle mit mehreren Suchkriterien kann sehr lange dauern, wenn der Optimizer die starken Selektionskriterien in der Abfrage nicht erkennt.

In einem konkreten Anwendungsfall werden Suchen über Banknummer, Kontonummer und einen Zeitbereich abgesetzt. Die *objkeys*-Tabelle enthält 370 Mio. Zeilen. Die Suchen benötigen 1-3 Sekunden, wenn der SQL-Server mit 48 GB RAM betrieben wird. Mit 16GB RAM dauern die Suchen ca. 50 Sekunden. Im Oracle Enterprise Manager kann man erkennen, dass zuerst über Banknummer, dann Zeitbereich und erst zuletzt über Kontonummer gesucht wird. Der Optimizer erkennt also nicht, dass die Suche über die recht eindeutige Kontonummer besser zuerst ausgeführt werden sollte.

2 Lösungskonzept

Statt in der *objkeys*-Tabelle mit einer Zeile pro Indexwert und vielen Zeilen pro Dokument werden die Indexwerte in einer „flachen“ Tabelle gespeichert mit einer Spalte je Indexwert und einer Zeile je Dokument. Performance-Tests auf der Basis der E-Mail-Daten in unserem Produktivarchiv haben eine fünffache Geschwindigkeitssteigerung gezeigt.

3 Realisierung

3.1 Neue Eigenschaft einer Verschlagwortungsmaske: Datenorganisation

Die Verschlagwortungsmaske bestimmt mit dem Feld `DocMask.dataOrganisation`, wie Indexwerte gespeichert werden sollen:

| Datenorganisation | Speicherung in DB |
|---|---|
| <code>DATA_ORGANISATION_OBJKEYS = 0;</code> | Herkömmliche Art der Speicherung als Name-Wert-Paare in der <i>objkeys</i> -Tabelle |
| <code>DATA_ORGANISATION_TABLE = 1;</code> | Verschlagwortungstabelle, eine Tabelle für die Verschlagwortungsmaske (flache Tabelle), je Indexfeld eine Spalte. |

3.2 Konvertierung vorhandener Indexdaten

Die Indexwerte einer bereits vorhandenen Verschlagwortungsmaske können mit dem folgenden Code in eine Verschlagwortungstabelle konvertiert werden.

Die Konvertierung erfolgt dadurch, dass das Element

```
dataOrganisation = DocMaskC.DATA_ORGANISATION_TABLE
```

gesetzt und die Verschlagwortungsmaske eingecheckt wird.

In einem Hintergrund-Job verschiebt der ELO Indexserver die Daten aus der Tabelle *objkeys* in die Verschlagwortungstabelle. Der Status des Jobs wird unter der GUID der Verschlagwortungsmaske geführt. Er kann mit `queryJobState` ausgelesen werden.



Information:

- a) Eine Konvertierung von Indexfeldern mit ungültigen Gruppennamen ist nicht möglich.
- b) Eine Konvertierung bei doppelten Gruppennamen ist nicht möglich.
- c) Eine Konvertierung zurück, von einer Verschlagwortungstabelle in die *objkeys*-Tabelle, ist aktuell nicht möglich.
- d) Für ein Indexfeld mit mehreren Indexwerten (Spaltenindex) wird nur der erste Wert in die Verschlagwortungstabelle übernommen.

- e) Während des Updates sollten keine Änderungen an der Verschlagwortung von Ordnern und Dokumenten vorgenommen werden, die mit der Verschlagwortungsmaske verschlagwortet sind. Das Hinzufügen von Ordnern und Dokumenten ist möglich.
- f) Der Originaldateiname ist weiterhin in der *objkeys*-Tabelle gespeichert (ELO_FNAME).
- g) Beim Start prüft der Indexserver, ob eine Konvertierung unterbrochen wurde und führt sie gegebenenfalls fort. Hierzu wertet er in der *eloixopt*-Tabelle einen Eintrag aus mit

```
IXID=_ALL,  
OPTNAME=dataOrganisationTransfer_<maskid>,  
OPTVALUE=<ixid>.
```

3.2.1.1 Beispiel

```
public static void changeDataOrganisation(IXConnection conn,  
    String maskId) {  
    log.info("start change data organisation");  
    try {  
        long t1 = System.currentTimeMillis();  
  
        DocMask dmTable = conn.ix().checkoutDocMask(maskId, DocMaskC.mbAll,  
            LockC.NO);  
        dmTable.setDataOrganisation(DocMaskC.DATA_ORGANISATION_TABLE);  
        conn.ix().checkinDocMask(dmTable, DocMaskC.mbAll, LockC.NO);  
  
        JobState jobState = conn.ix().queryJobState(dmTable.guid, true, true,  
            true);  
        while (jobState.isJobRunning()) {  
            Thread.sleep(1000);  
            jobState = conn.ix().queryJobState(dmTable.guid, true, true, true);  
            log.info("jobState=" + jobState);  
        }  
  
        long t2 = System.currentTimeMillis();  
        log.info("end change data organisation, ms=" + (t2 - t1));  
    } catch (Exception e) {  
        log.error("Failed to convert data organisation", e);  
    }  
}
```

3.3 Darstellung in der Datenbank

Die Verschlagwortungstabelle wird mit dem Namen „Keywording_“ plus Verschlagwortungsmasken-ID angelegt. Die Indexfelder finden sich als Spalten in dieser Tabelle wieder.

3.3.1.1 Beispiel:

```
CREATE TABLE [dbo].[keywording_239] (
    [parentid] [int] NULL,
    [rechnr] [nvarchar](255) NULL,
    [rechdat] [nvarchar](255) NULL,
    [liefernrv] [nvarchar](255) NULL,
    [bestellnr] [nvarchar](255) NULL,
    [betreff] [nvarchar](255) NULL,
    [kdnrv] [nvarchar](255) NULL,
    [lieferant] [nvarchar](255) NULL,
    [barcode] [nvarchar](255) NULL,
    [betrag] [nvarchar](255) NULL,
    [mitarb] [nvarchar](255) NULL
)
```

3.3.2 Datenbankindizes

Vom Indexserver werden keine Datenbankindizes auf Verschlagwortungstabellen angelegt. Dies obliegt allein dem Datenbankadministrator.

3.3.3 Case-insensitive-Suche bei Oracle

Für eine case-insensitive Suche unter Oracle müssen funktionsbasierte Indexe verwendet werden. Textspalten im Index werden als Funktionsargument von UPPER aufgenommen.

Ein passender Index sieht beispielsweise so aus:

```
CREATE INDEX "ARCHIV1"."IXKEYWORDING_32_01" ON "ARCHIV1"."KEYWORDING_32"
(
    UPPER(BANK),
    UPPER(KTO),
    PARENTID
)
TABLESPACE "TS_ELO_INDEX";
```

3.4 Bearbeiten der Maskendefinition

Wird in der ELO Administration Console ein Indexfeld hinzugefügt, so wird in der Verschlagwortungstabelle eine Spalte ergänzt.

| | |
|---|--|
|  | <p>Beachten Sie: Diese Spalte enthält NULL Werte. Das Setzen eines zum Indexfeld festgelegten Standardwertes würde bei einer sehr großen Tabelle sehr lange dauern. Vorhandenen Objekten werden folglich keine Standardwerte zugewiesen. Sie gelten nur für neue Ordner und Dokumente.</p> <p>Umgekehrt, wenn ein Indexfeld entfernt wird, dann löscht der Indexserver die zugehörige Spalte in der Tabelle.</p> <p>Dabei gehen alle Daten des Indexfeldes verloren.</p> |
|  | <p>Information: Wie die Verschlagwortungsmasken mit <i>objkeys</i>-Daten können die Verschlagwortungsmasken mit Verschlagwortungstabelle nur gelöscht werden, wenn kein Objekt mehr damit verschlagwortet ist.</p> <p>Wenn beim Löschen eine alternative Verschlagwortungsmaske angegeben wird, erfolgt keine automatische Konvertierung der Daten.</p> <p>Verschlagwortungsmasken werden nicht dauerhaft entfernt.</p> |

3.5 Suchen über Verschlagwortungstabelle mit „findfirstSords“

- a) Anhand der in `FindByIndex` angegebenen Verschlagwortungsmaske wird entschieden, über welche Tabelle gesucht wird.
- b) Ist in der Suche keine Verschlagwortungsmaske angegeben, dann werden die Verschlagwortungsmasken über die Indexfeld-Gruppennamen aus der Suchanfrage ermittelt.
 - a) Es wird über die Verschlagwortungstabellen gesucht, die alle in `FindByIndex.objKeys` gegebenen Indexfeld-Gruppen enthalten.
 - b) Enthält eine Verschlagwortungsmaske mit `objkeys`-basierter Datenorganisation alle Indexfelder, wird auch über die `objkeys`-Tabelle gesucht.
 - c) Wenn keine Verschlagwortungsmaske alle Indexfeld-Gruppen enthält, wird nur über die `objkeys` gesucht.
- c) Sind in `FindByIndex` Verschlagwortungsmasken angegebenen, dann wird nur über diese Masken gesucht. Die Angabe erfolgt über `FindByIndex.maskId` oder über das Array `FindByIndex.maskIds`.
- d) Eine Suche über alle Indexfelder von Verschlagwortungstabellen ist aktuell nicht möglich.
- e) Eine ODER-Verknüpfung von Indexfeldern ist aktuell nicht möglich.

4 Auswirkung auf andere ELO Module

4.1 Indexserver-Client-Anwendungen, ELO Java Client, ELO Web Client...

Die Indexwerte aus Verschlagwortungstabellen werden ebenfalls in `sord.objKeys` zurückgegeben bzw. entgegengenommen. Eine Änderung in bestehenden Client-Anwendungen ist nicht nötig.

4.2 ELO Windows Client

Es ist nicht geplant, die Behandlung von Verschlagwortungstabellen im ELO Windows Client nachzubauen. Aus diesem Grund können Verschlagwortungstabellen nur in Installationen ohne ELO Windows Client verwendet werden.

4.3 ELO Replikation

Es gibt zum jetzigen Stand kein ELO Replikationsmodul, das Verschlagwortungstabellen übertragen kann.

5 Einschränkungen

- Nur ein Wert pro Indexfeld möglich (kein Spaltenindex)
- Beim Konvertieren wird nur der erste Wert eines Indexfelds aus der Tabelle *objkeys* übernommen
- Keine Suche über alle Indexfelder
- (Zurück-)Konvertierung von Verschlagwortungstabelle in der Tabelle *objkeys* nicht möglich
- ODER-verknüpfte Suche nicht möglich
- Während der Konvertierung dürfen keine Einträge mit der Maske bearbeitet oder eingefügt werden
- Nur Installationen ohne ELO Windows Client und ELO Replikation