

Datenbanken II B: DBMS-Implementierung — 11. Übungsblatt: Heap-Dateien, Zeilenformat —

Hausaufgaben

Geben Sie die Aufgaben dieses Abschnitts bis Mittwoch, 26.01.2022, 18⁰⁰, über die Übungsplattform in StudIP ab. Schreiben Sie die Lösungen in eine `.txt`-Datei bzw. `.sql`-Datei (kein PDF und schon gar kein Word).

Sie haben in den Übungen Zugriff auf eine Oracle-Datenbank (Version 18c) über die Adminer-Webschnittstelle bekommen:

```
https://dbs.informatik.uni-halle.de/db2b/adminer?  
oracle=oracle-18.4-xe-db2b%2FXEPDB1&username=&db=USERS
```

Für Teil c) müssen Sie mit Administrator-Rechten arbeiten. Die Tabelle `USER_EXTENTS` enthält nicht die Blockadresse (`BLOCK_ID`) des Extents, die Tabelle `DBA_EXTENTS` schon.

- a) **(3 Punkte)** Wie beim letzten Übungsblatt sei die Tabelle

`HEAP(ID, A, B, C°)`

betrachtet. Sie finden die `CREATE TABLE` Anweisung in der bekannten Datei:

```
[http://www.informatik.uni-halle.de/~brass/dbi21/homework/heap.sql]
```

Die Spalten `A` und `B` sind jeweils mit 64 Zeichen langen Strings gefüllt. Die Spalte `C` enthält zunächst einen Nullwert, später eine 30 Zeichen lange Zeichenkette.

Wie viele Zeilen müssten vor dem Update in einen Block passen? Rechnen Sie mit einer Größe des „Block Header“ von 119 Byte. `PCTFREE` hat den Default-Wert 10. Die Zeilenlänge wurde in der letzten Übung zu 137 Byte berechnet, das wurde auch durch die Tests verifiziert. Vergessen Sie nicht den Platz für das „Row Directory“ (das ist in den 137 Byte noch nicht enthalten).

- b) **(3 Punkte)** Schreiben Sie ein Programm in `C++` oder einer anderen Sprache, das `INSERT`-Anweisungen der gleichen Art wie in der gegebenen Datei `heap.sql` erzeugt. Erzeugen Sie damit ausreichend viele Anweisungen, dass ein zweiter Extent begonnen werden muss. Geben Sie Ihr Programm als Lösung von Teil b) ab. Löschen Sie den Inhalt der vorhandenen Tabelle mit einem `TRUNCATE`-Befehl (Sie dürfen alternativ auch die ganze Tabelle löschen und sie neu anlegen). Führen Sie dann die von Ihnen erzeugte Datei aus.

- c) **(3 Punkte)** Was ist die erste ID einer Zeile, die nicht mehr im ersten Extent gespeichert wurde?

Man kann vielleicht annehmen, dass die Blöcke des ersten Extents gefüllt werden, bevor ein zweiter Extent begonnen wird. (Möglicherweise werden auch nur einige Blöcke gefüllt, so dass weiter mehrere einfügbereite Blöcke für parallele Transaktionen zur Verfügung stehen.)

- d) **(3 Punkte)** Wie viele Zeilen sind nun in den Blöcken gespeichert? Geben Sie pro Block, der mindestens eine Zeile enthält, die Anzahl der Zeilen aus, sowie die minimale und maximale ID (aus der Tabelle `HEAP`). Bestimmen Sie die maximale Anzahl Zeilen pro Block (entweder manuell oder als Teil der SQL-Anfrage). Geben Sie die SQL Anfrage mit dem Maximum als Kommentar ab.

- e) Es sei die folgende Relation gegeben:

`R(A: NUMERIC(5), B: VARCHAR(10), C: VARCHAR(50))`

Es ist kein Schlüssel deklariert, daher können auch lauter gleiche Zeilen eingefügt werden. `PCTFREE` sei 10. Nun werden 10000 Tupel der folgenden Form eingefügt:

`(99999, null, 'abcdefghijklmnopqrstuvwxyz')`

Beantworten Sie folgende Fragen:

1. Was ist die Speichergröße der Zahl 99999 in Oracle? (Zählen Sie dabei das Längenbyte noch nicht mit.)
2. Was ist die Zeilenlänge in Oracle?
3. Wie viele Bytes benötigt die Speicherreserve für verlängernde Updates? Die Blockgröße ist 8 KByte, also 8192 Byte (Runden Sie bei Bedarf auf.).
4. Wie viele Tupel passen in einen Block? (Rechnen Sie mit 119 Bytes für den Block Header.)
5. Wie viele Blöcke wird die Tabelle ungefähr benötigen?

Führen Sie diese Berechnungen theoretisch durch. Wenn Sie wollen, können Sie hinterher einen praktischen Test machen (Sie müssen aber nicht.). Leider werden die Ergebnisse des praktischen Tests nicht ganz passen. Falls Sie herausfinden, wie die Angaben im Skript aktualisiert werden müssen, würde mich das interessieren. Früher (Oracle 8, vermutlich auch 8i) stimmte das Berechnungsergebnis genau mit dem realen Wert überein. Das ASSM hat aber zu deutlichen Änderungen geführt.

- f) Angenommen, man setzt den Wert von B für alle Tupel auf `'ABCDEFGHIJ'` (10 Zeichen). Wird es migrierte Zeilen geben? Beantworten Sie auch diese Fragen durch Rechnung, und nicht (nur) über einen praktischen Test:
- Wie lang werden die Zeilen nach dem Update, d.h. wie viele zusätzliche Bytes werden benötigt?
 - Ist die `PCTFREE` Platzreserve ausreichend für das Wachstum aller Zeilen im Block?

- Falls nicht, was wäre ein guter Wert für `PCTFREE`, um migrierte Zeilen durch den Update zu vermeiden?
- g) **(3 Punkte)** Beschreiben Sie kurz ein alternatives Speicherformat für Zeilen der Relation `R`. Auch den `NUMERIC(5)` Wert könnten Sie anders speichern als in Oracle. Welche Information müsste im Data Dictionary stehen, damit die Zeilen decodiert werden können?

Wiederholungsaufgaben

Die „Wiederholungsaufgaben“ brauchen Sie nicht abzugeben. Beschäftigen Sie sich aber bitte auch mit diesen Aufgaben. Notieren Sie sich Fragen, die Sie gerne in der Übung geklärt haben wollen.

- h) Wie würden Sie in einer mündlichen Prüfung auf folgende Fragen antworten?
- Was sind die Vor- und Nachteile von Zeilen fester Länge im Vergleich zu Zeilen variabler Länge? Bedenken Sie auch, was bei einem `ALTER TABLE`-Befehl getan werden muss, der eine Spalte zu der Tabelle hinzufügt. Vielleicht denken Sie auch an Techniken der Daten-Kompression.
 - Welche Blöcke liest Oracle bei einem „Full Table Scan“? Geben Sie ein Beispiel, bei dem das sehr ineffizient wirkt. Was ist die Ursache für das Problem? Wenn Sie Ihr eigenes DBMS entwickeln würden, würden Sie das gleiche Verfahren wie Oracle verwenden?
 - Betrachten Sie ein Blockformat für Zeilen fester Länge und eine verkettete Liste von Blöcken, die noch Platz für die Einfügung von Zeilen haben. Wie wird diese Liste verwaltet, d.h. wo finden Sie den Beginn der Liste und was passiert bei Einfügungen und Löschungen von Zeilen? Wenn ein Block voll war, und dann eine einzelne Zeile gelöscht wurde, warum könnte es eventuell geschickt sein, den Block nicht sofort wieder auf die Freispeicherliste zu setzen (sondern erst nach Löschung mehrerer Zeilen)? Diskutieren Sie Vor- und Nachteile dieser Technik.
 - Erläutern Sie das Zeilenformat von Oracle. Wie findet man den Wert einer bestimmten Spalte? Diskutieren Sie Alternativen. Was sind die Vor- und Nachteile des Oracle-Formats? Wenn Sie Ihr eigenes DBMS entwickeln würden, welches Zeilenformat würden Sie wählen?
 - Warum brauchen Spalten am Ende einer Zeile, die einen Nullwert enthalten, in Oracle keinen Speicherplatz? Welchen nützlichen Effekt hat das neben der Einsparung von Speicherplatz?
 - Wie kann man die Speichergröße einer Festkommazahl (`NUMERIC`) in Oracle berechnen? Was sind die Komponenten des Wertes in der internen Repräsentation? Wieviel Platz würde der Wert `12345` benötigen? Und der Wert `-12.34`?

- Erklären Sie das „ANALYZE TABLE“ Kommando in Oracle. Warum ist es kein Problem, wenn die Statistiken im Data Dictionary nicht 100% aktuell sind? Wann sollte der DBA den „ANALYZE TABLE“ Befehl ausführen? Warum sollte das nicht während der Haupt-Geschäftszeiten geschehen?

Für Interessierte

- i) Recht detaillierte Angaben zum Block Header finden Sie auf dieser Seite:

[http://www.adp-gmbh.ch/ora/concepts/db_block.html]

Die letzte dort erwähnte Version ist 10g.

- j) Angaben von Oracle (leider für eine ganze alte Version, 8.0) sind auf dieser Seite:

[https://docs.oracle.com/cd/A58617_01/server.804/a58397/apa.htm]

Dort wird insbesondere auch die Tabelle `V$TYPE_SIZE` verwendet, in der die Größen von verschiedenen internen Datentypen gespeichert sind:

[https://docs.oracle.com/en/database/oracle/oracle-database/18/refrn/V-TYPE_SIZE.html]

In neueren Oracle-Versionen wird auf Prozeduren verwiesen, die diese Berechnung durchführen (insbesondere `DBMS_SPACE.CREATE_TABLE_COST`):

[<https://docs.oracle.com/en/database/oracle/oracle-database/18/admin/managing-space-for-schema-objects.html>]

Das Paket `DBMS_SPACE` ist hier beschrieben:

[https://docs.oracle.com/en/database/oracle/oracle-database/18/arpls/DBMS_SPACE.html]

Leider weiss ich nicht, wie man die Code-Blöcke zum Prozeduraufruf über den Adminer eingeben kann.

- k) Mehr Informationen zu „Locally Managed Tablespaces“ und „Automatic Segment Space Management“ finden Sie auf dieser Webseite:

[http://www.dba-oracle.com/art_dbazine_ts_mgt.htm]