



Übung zur Vorlesung
Einsatz und Realisierung von Datenbanksystemen im SoSe15

Moritz Kaufmann (moritz.kaufmann@tum.de)
<http://db.in.tum.de/teaching/ss15/impldb/>

Blatt Nr. 11

Hausaufgabe 1 - Machbarkeit von Hauptspeicherdatenbanken

Berechnen Sie, wie groß ein Data Warehouse für ein Handelsunternehmen wie Zalando oder Amazon.com wäre, wenn die Bestelldaten der letzten 3 Jahre enthalten sind. Versuchen sie ihre Schätzung basierend auf öffentlichen Daten wie Umsatz o.Ä. durchzuführen.

Hausaufgabe 2

HyPer schafft 120.000 Transaktionen pro Sekunde. Pro Transaktion werden 120 Byte in die Log geschrieben. Berechnen Sie den benötigten Durchsatz zum Schreiben der Log.

Die Datenbank läuft für einen Monat und stürzt dann ab. Es wurde kein Snapshot erstellt. Berechnen Sie die Recoveryzeit. Gehen Sie davon aus, dass die Recovery durch die Festplatte limitiert ist (100 Mbyte / s). Wieviel Log Einträge werden pro Sekunde reconvert?

Hausaufgabe 3

Wie ändert sich die Bedeutung des Redo-Log und Undo-Log in Hauptspeicherdatenbanken im Vergleich zu klassischen Datenbanken? Wo werden sie gespeichert?

Hausaufgabe 4

In traditionellen Datenbanksystemen sind die Festplatte und der Buffermanager oft der Hauptgrund für Performanceengpässe. Wie ändert sich dies in Hauptspeicherdatenbanken, wo sind die neuen Flaschenhälse? Unterscheiden Sie auch zwischen Analytischen und Transaktionalen Workloads.

Gruppenaufgabe 5 - Lösung während der Übung

Sie sollen für ein Versandhaus die Datenbank für ein Hauptspeicherdatenbanksystem optimieren. In dem System sind die Daten der letzten drei Jahre gespeichert. Das Schema der verschiedenen Relation ist unten beschrieben. Wählen sie jeweils eine Repräsentation der Daten für die Relationen (z.B. Spalten- oder Zeilenorientert), so dass zur Beantwortung der unten beschriebenen Anfragen möglichst wenig Daten in den CPU-Cache geladen werden müssen. Es existieren Indexe auf VerkaufsId in Verkauf, (KundenId, VerkaufsId) in KundenKäufe und KundenId in Kunde. Es können aber keine neuen Indexe definiert werden. Text wird direkt innerhalb der jeweiligen Spalte / Zeile gespeichert. '

Relationen

Verkauf

VerkaufsId (8 byte), Datum (16 byte), Uhrzeit (16 byte), IP (16 byte), Betrag (16 byte), Versandart (8 byte), Kommentar (48 byte)

KundenKäufe

KundenId (8 byte), VerkaufsId (8 byte)

Kunde

KundenId (8 byte), Anrede(8 byte), Vorname(8 byte), Nachname(8 byte), Einstufung(8 Byte), Anschrift(256 byte), Land(64 byte), Email (16 Byte), Facebook(32 byte), GPlus(32 Byte), WerbungsFrequenz(8 byte)

Anfragen mit Ausführungshäufigkeit:

1. 10000x select * from Verkauf
where Datum > '%d'::date and Datum < '%d'::date + interval '1' month;
2. 100x select * from Verkauf;
3. 100x select count(*) from KundenKäufe group by KundenId;
4. 10000x select Anrede, Vorname, Nachname, Einstufung, Email, WerbungsFrequenz
from Kunde where KundenId = '%id';

Würden Sie ihre Entscheidung ändern, wenn zusätzlich noch die folgenden Anfragen ausgeführt werden müssten?

1. 5000x insert into Verkauf VALUES (...);
2. 5000x insert into KundenKäufe VALUES (...);
3. 100x insert into Kunde VALUES (...);