Prüfung 1666 Datenbanken in Rechnernetzen

Prüfer: Prof. Wilkes

Datum: 04.03.2016

Allgemeines:

Prof. Wilkes ist ein angenehmer Prüfer, wobei wenn man eine gute Note bekommt empfindet man das wohl meist so. Prüfungsprotokolle sind bei der Prüfungsvorbereitung sehr zu empfehlen.

Fragen:

Warum verteilt man Datenbanken?

Dezentralisierung (Verfügbarkeit, Durchsatz, Skalierbarkeit)
Zentralisierung (Verbesserung Konsistenz, „Single-System-Image“)

Wie kann man heterogene Datenbanken integrieren und welche Probleme gibt es?
LRS

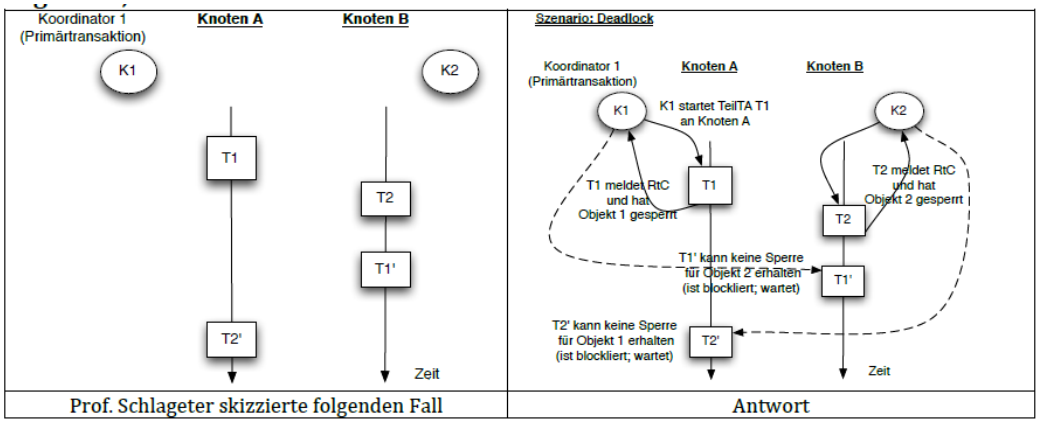
Wie kann man Relationen verteilen bzw. wieder zusammen führen?

Horizontal (Vereinigung)
Vertikal (NJoin)
Horizontal abgeleitete wurde nicht gefragt.

Wie wird Serialisierbarkeit definiert?
Siehe Script.

Serialisierbarkeit bei verteiltem System?

Hier hatte er mir eine Zeichnung gemacht.
Ähnlich dieser:



Prof. Schlageter skizzierte folgenden Fall | Antwort

Was ist 2 PC?

Semi-Join erklären mit Vor- und Nachteilen?

Hashjoin mit Vor- und Nachteilen?

Wie behandelt man Updates bei Repliken?

Rowa, Vorbestimme Kopie, Abstimmverfahren

Prüfungsprotokoll

01666 Datenbanken in Rechnernetzen
Prüfer Dr. Wolfgang Wilkes
13.6.2014

Note: 1.0

Videokonferenzprüfung

Gestellt wurden mir wenige Fragen und dann auch nur die Handvoll Standardfragen, die sich durch alle bisherigen Prüfungsprotokolle (auch die von Dr. Willkes Vorgänger) durchziehen.

Ich möchte darum hier meinen Eindruck von der Prüfung wiedergeben:

Die ganze Prüfung empfand ich so, wie wenn ich in eine wichtige Sitzung als Expertin Einsitz nehme. Entsprechend habe ich auf die Fragen reagiert und so beispielsweise bei der Erklärung des Semi-Joins von mir aus gleich auch Kritikpunkte aufgezählt, auf den Hash-Join verwiesen und folglich erklärt, wie der geht.

Bei der Primary Copy habe ich dann etwa erklärt, wie die Kopien direkt mit ihr verbunden sind (das ist vielleicht ein Detail, aber so speziell, dass es auffallen sollte!), was asynchrone Updates bedeuten und dann kam logischerweise die Diskussion über Lese-TRX und ob die immer aktuelle Daten bräuchten - dann ist es natürlich hilfreich, Beispiele parat zu haben und die gleich aufzuführen. Und dann bin ich zurückgesprungen zu den Update-Verfahren, hab da weitergemacht.

Dr. Wilkes gehört also nicht zu jenen Prüfern, die Kapitel um Kapitel abfragen und sich auf Details stürzen. Es geht darum, Zusammenhänge zu erkennen und sich sicher durch die Kapitel zu bewegen und sie zu vernetzen.

Und dann, so habe ich das an der Prüfung erlebt, ist es auch nicht schlimm, klar zuzugeben, wo man was nur mit Hilfe der Unterlagen machen kann: so bei mir und dem DDD und diesen WAGs mit den EX; da gibt es ein Beispiel im Kurstext und in den EAs. Das war ich meine Antwort zu dem Thema! Ich hab dann schon gesagt, dass Zyklen damit erkennbar sind. Und dann noch einige kritische Bemerkungen hinzugefügt, wie „Momentaufnahme" und dass das Opfer, welches so gefunden wird, nicht unbedingt das „beste" ist, und das natürlich erläutert

Da war die Prüfung schon vorbei, die Note kam fast unmittelbar und Dr. Wilkes meinte, ich könnte wohl noch stundenlang mit ihm über das Thema (des Kurses) reden.

So ist es, aber: als IT Projektleiter habe ich mit dem Thema im praktischen Alltag nicht viel zu tun.

Ich hoffe, dass mein Bericht zur Prüfungsvorbereitung hilft

Prüfungsprotokoll

01666 Datenbanken in Rechnernetzen
Prüfer Dr Wilkes

Mai 2014
Note 1,7

Einstiegsfrage durch Übergang von privaterem Gespräch, in dem ich etwas über meine lange zurückliegende Berufserfahrung im konzeptionellen Datenmanagement eines Großunternehmens noch mit zentraler DBMS gesagt hatte:

Hier jetzt nicht zentraler Fall, sondern verteilt. Welche Gründe kann das haben?

Ich hatte mich auf die Kernelemente globale und lokale Komponente = Anfragemöglichkeiten gestützt. Reichte nicht. Dann noch allgemein aus wirtschaftlichen, historischen, technischen Gründen genannt. Prüfer wollte konkret auf Verteilung der Aus-/Belastung hinaus.

Dies war dann auch schon die einzige Frage, bei der man sagen konnte, dass der vorher so viel beschworene allgemeine Überblick abgefragt wurde. Danach wurde es richtig kleinkariert:

Voting-Verfahren erklären, vor allem das Dynamic Voting am Beispiel auf dem Papier. Dann noch auf das Tree Quorum eingegangen und wollte es genau erklärt haben. (Anzahl Ebenen = Höhe des Baumes und Verzweigungstiefe spielen eine Rolle.)

Warum man Abstimmungsverfahren macht.

Kopierverfahren erklären, vor allem das Primary Copy – Verfahren. Neu gelernt hatte ich in der Prüfung, dass das Weiterreichen der Änderung an die Nicht-Primär-Knoten keine Transaktionen im Sinne des 2 PC sind. Das wird durch die Zeichnung im Kursheft nicht ganz klar und ist auch nicht beschrieben, wie das funktioniert.

Wie das 2 PC genau ! funktioniert.

Was Serialisierbarkeit genau bedeutet.

Wie man Deadlock in verteilten Datenbanken erkennt = die WAG – Geschichte mit den EX-Knoten erläutern.

Ich persönlich fand die Prüfung viel zu eng geführt. Ich hatte jederzeit das Gefühl, auf Defizit hin geprüft worden zu sein, nicht auf das, was ich vielleicht gut kann und was mir Spaß gemacht hat beim Belegen des Kurses, zumal ich ca. 80% der Einsendeaufgaben erledigt hatte. Das interessierte keinen. Von einer Prüfung, die Gesamtzusammenhänge erfassen will, kann keine Rede sein. Nach der Einleitungsfrage war einfach zusammenhanglos zu den einzelnen Konzepten gesprungen worden.

Aber die Note wird einem hinterhergeworfen. Man scheint schon zu honorieren, dass man sich überhaupt zur Prüfung meldet. Am besten schneiden nach Auskunft des Prüfers diejenigen ab, die schon in der Praxis mit Datenbanken zu tun haben. Es hatte sich für mich so angehört, als ob dies sogar Voraussetzung für den Studiengang sei. Die meisten seien schon im Beruf da drin und legten jetzt nur noch den Master nach.

Mündliche Prüfung
Kurs 1666 Datenbanken in Rechnernetzen
Prüfer Dr. Wolfgang Wilkes
18.10.2013

Warum verteilt man Datenbanken?
- Ausfallsicherheit durch Redundanz
- Parallele Verarbeitung
- Erweiterbarkeit
- Zusammenführen ehemaliger zentraler Datenbanken

Wie kann man sehr verschiedene Datenbanken integrieren, wenn es alte Anwendungen gibt, die nicht verändert werden dürfen?
- Lokales Repräsentationsschema
- Ich habe eine kleine Skizze gezeichnet und genau erklärt, was das LRS ist, wie alte Anwendungen weiterhin arbeiten können und wie die neuen Anwendungen über das globale Schema und das LSR an ihre Daten kommen.

Welche Probleme gibt es bei der Homogenisierung durch das LSR?
- Homonyme, Synonyme
- Unterschiedlich aufgebaute Primärschlüssel
- Sachverhalt ist in der einen Datenbank eine Relation, in der anderen wird er als Attribute dargestellt.

Aus welchen Schichten besteht das globale Schema?
- Globales konzeptuelles Schema
- Globales Partitionsschema
- Globales Allokationsschema

Welche Möglichkeiten der Partitionierung gibt es?
- Horizontale Partitionierung
  - Partitionierung mit Hilfe eines Selektionsprädikats
  - Selektion
- Vertikale Partionierung
  - Partitionierung mit Hilfe einer Projektion
  - Entlang einer funktionalen Abhängigkeit, am besten Primärschlüssel verwenden
- Die abgeleitete horizontale Partitionierung wollte er nicht mehr hören.

Wie kann man die vertikal partitionierten Tabellen wieder zusammenführen?
- Über einen natürlichen Verbund

Wie kann man verteilte Joins durchführen?
- Nested-Loop-Join (hier hat Herr Wilke abgewunken)
- Semi-Join
- Hashfilter
- Sort-Merge-Join

Erklären Sie den Semi-Join.
- Am besten mit einer kleinen Skizze erklären
- Die erste Relation wird mit Hilfe einer Projektion verkleinert
- Die zweite Relation wird komplett übertragen und wird nur um Zeilen verkleinert

Wann ist der Semi-Join besonders gut und wann bringt er Nachteile?

- Der Semi-Join bringt besonders viel, wenn die Relationen stark verkleinert werden können
- Hier wollte er vor allem hören, dass das Senden der zweiten Relation ein Knackpunkt ist
- Wenn die Relationen nicht stark verkleinert werden können (vor allem die zweite), kann der Aufwand erheblich höher sein als bei anderen Verfahren
- Weiteres Problem liegt vor, wenn die Joinattribute viele verschiedene Ausprägungen haben

Den Hashfilter erklären.

- Hashfilter kurz mit einer Skizze erklärt. Hier unbedingt Pseudotreffer erwähnen.

Wann ist der Hashfilter besonders gut und wann nicht?

- Hier müssen wieder die Pseudotreffer erwähnt werden
- Das Problem der Pseudotreffer macht sich bemerkbar, wenn die verkleinerte Relation zurückgeschickt wird.

Wo werden Abstimmungsverfahren eingesetzt?

- Bei Replikationsverfahren

Wie funktioniert ein Abstimmungsverfahren im Groben?

- Bei Replikation muss die Konsistenz der Kopien gewährleistet werden
- Aus diesem Grund stimmen die anderen Kopien darüber ab, ob eine Kopie aktualisiert werden soll, wenn eine Update-Transaktion an einem Knoten eintrifft
- Das hat ihm gereicht und die Prüfung war vorbei

Herr Wilkes ist ein wirklich guter Prüfer. Durch ein kurzes Gespräch am Anfang der Prüfung hat er eine lockere Atmosphäre geschaffen und gelegentlich hat er einen kleinen Scherz eingeworfen. Wenn der Prüfling etwas auf dem Schlauch steht, lässt Herr Wilkes ihn nicht im Stich, sondern versucht ihn durch gezielte Fragen noch einmal auf den richtigen Weg zu bringen. Es war eine angenehme Prüfung und ich kann Herrn Wilkes als Prüfer auf jeden Fall weiterempfehlen. Ich empfand ihn als sehr fair.

Mündliche Prüfung
Kurs 1666 Datenbanken in Rechnernetzen.
Prüfer Dr. Wolfgang Wilkes
6.7.2012
Note 1,3

Warum verteilt man Datenbanken?
Je nachdem, ob Dezentralisierung oder Integration notwendig ist, gibt es verschiedene Gründe.
Dezentralisierung: Antwortzeit zu minimieren, Verfügbarkeit zu erhöhen, Gesamtdurchsatz verbessern, Erweiterbarkeit, günstige Hardwaremöglichket zu nützen. (Bevor ich über Integration weiter erzählen konnte, wurde ich unterbrochen.)

Wie kann man Verfügbarkeit erhöhen?
Redundante Datenspeicherung auf mehreren Knoten.

Wie kann man Durchsatz verbessern?
Durch parallele Ausführung.

Bei der Integration..., es existieren schon hetrogene Systeme. Wie wird Integration ausgeführt?
Durch das lokale Repräsentations-Schema (ich musste Homogenisierung der lokalen Schemata erklären. Hier hat er sehr detailliert gefragt inkl. Details aus der graphischen Darstellung (ich musste skizzieren).
Auch die interne Struktur des Globalen Schemas: globales konzeptuelles Schema, globales Partitionierungsschema (Info wie partitioniert), globales Allokationsschema (Info wo gespeichert)

Welches Problem entsteht bei der Redudanz?
Konsistenzhaltung der redundant gespeicherten Daten.

Wie wird das Problem gelöst?
Es gibt verschieden Kopien – Updates Verfahren. Alle genannt.

Wie funktioniert ROWA?
Von einer Kopier lesen und alle Kopien schreiben, konsistentes Lesen.

Warum ist es unpraktisch?
Updateproblematik erklärt, ineffizient.

Erklären Sie mir das Primary Copy Verfahren.
Änderung an der Hauptkopie zuerst durchgeführt, danach werden die restliche Kopien aktualisiert.

Was passiert wenn die Primary Copy ausfällt?
Eine andere Kopie wird zur Primary Copy gemacht.

Worauf muss man hierbei achten?

Dass die Primary Copy wirklich ausgefallen ist, also damit es nicht zwei davon gibt.

Ja..das auch, aber noch was?
Inkonsistentes Lesen, die nächte Kopie kann möglicherweise noch nicht aktualisiert sein.

Wie garantiert 2PC Serialisierbarkeit?
Sperren werden erst nach dem Abschluss des globalen Commit freigegeben, nicht beim Ready-to-Commit Zeitpunkt.
(Es gab auch viele detaillierte Fragen über das 2PC Protokoll, an die ich mich nicht mehr genau erinneren kann, also alles lernen.)

Was is Serialisierbarkeit?
Definiton

Wie können die Relationen auf die verschiedenen Knoten verteilt werden?
Horizontal und vertikal

Semi-Join erklären

Vergleich Semi-Join und Hashfilter Join
Beim Senden von der äußeren Relation zur inneren Relation ist der Hashfilter Join besser, da nur ein Bitvektor zu senden ist. Von der inneren zur äußeren Relation kann u.U. eine große Relation übertragen werden wenn man die Länge des Bitvektors zu kurz wählt, wegen Pseudo-Treffern !!! Die Güte dieses Verfahren hängt deswegen von der Wahl einer geeingneten Hashfunktion und damit auch von der Größe des Bitvektors wichtig.

Dr. Wilkes hat sehr viel gefragt und nutzt die 25 Minuten voll aus. Er hat noch mehr Fragen gestellt, an die ich mich nicht mehr erinnern kann. Er fragt sehr detailliert und bohrt auch wenn er das Gefühl hat, dass der Prüfling das Thema nicht gut kennt. Er ist aber sehr freundlich und die Prüfungsathmosphere ist sehr angenehm. Das nimmt die Prüfungsnervosität. Die alten Prüfungsprotokolle sind sehr nützlich, man sollte aber sehr detailliert lernen.

Prüfung: Datenbanken in Rechnernetzen (Kurs 1666)
Prüfer: Dr. Wilkes
Datum: 20.4.2012

Dr. Wilkes ist der Nachfolger von Prof. Schlageter. Da ich die Prüfung erst im zweiten Versuch bestanden habe, kann ich vielleicht etwas zum Unterschied zwischen den Prüfungen von Prof. Schlageter und Dr. Wilkes sagen:

Es gibt keinen.

Die alten Prüfungsprotokolle sind nach wie vor sehr gut zur Vorbereitung geeignet. Die Reihenfolge bei den Fragen ist nicht unbedingt die gleiche, ich wurde kreuz und quer gefragt. Und das Thema „Client/Server“ kann man wohl mittlerweile etwas vernachlässigen, ich wurde dazu auch nichts gefragt.
Eine Frage ist mir noch in Erinnerung geblieben, die ich in keinem Protokoll gefunden habe:

Welche Transaktion wählt beim der DDD als Opfer aus?

Die jüngste. Grund: Sie hat bislang am wenigstens gemacht und es ist am günstigsten, diese Transaktion neu zu starten.

Gedaechtnisprotokoll 1666

Prüfer: Dr. Wilkes
Datum: 08.07.2011
Note: 2.0

Die Prüfung fing etwas früher an als geplant. Zu erst plauderten wir locker über Anreise und Herkunft - das nimmt ein wenig die Nervosität und habe ich als sehr angenehm gefunden.

Die Fragen bzw. Themen sind alle schon in allen anderen Protokollen vorhanden.

Kurz:
- Warum Verteilung?
- Redundanz
- Synchronisationsverfahren
- Replikationsverfahren
- Transaktionen und Serialisierbarkeit
- 2PL, 2PC Protokoll
- Verklemmung

Bei Transaktionen und Serialisierbarkeit wurden Beispieltransaktionen aufgezeichnet. Hier war es ein wenig hakelig bei mir. Es empfiehlt sich genau den Ablauf von Transaktionen und Serialisierbarkeit (auch globale) anzuschauen und zu verstehen.

Fazit: Dr. Wilkes ist ein sehr netter Prüfer und uneingeschränkt zu empfehlen. Er führt ein wenig durch die Fragen und schafft passende Überleitungen. Das logische Überlegen und laute Nachdenken wird als positiv aufgefasst, allerdings sollte man dann auch zum Ergebnis kommen.

Hier sind noch zwei MP3 zur Vorbereitung (für die U-Bahn oder im Auto):

http://www.sobek.info/fileadmin/studium/01666_Kurszusammenfassung.mp3

http://www.sobek.info/fileadmin/studium/01666_Kontrollfragen_KE1.mp3