

Übung 5. – Leistungsmodellierung

Dimensionsanalyse. Während der Lösung der Leistungsmodellierungsaufgaben lohnt es sich die aus der Physik bekannte Dimensionsanalyse¹ zu benutzen. Sehen wir uns das quadratische Weggesetz an: $s = v_0 t + \frac{a}{2} t^2$

Mit Dimensionen: $s[\text{m}] = v_0 \left[\frac{\text{m}}{\text{s}} \right] t[\text{s}] + \frac{a}{2} \left[\frac{\text{m}}{\text{s}^2} \right] t^2[\text{s}^2] = v_0 t[\text{m}] + \frac{a}{2} t^2[\text{m}]$

Die Motivation für Dimensionsanalyse ist, dass wenn die Dimensionen nicht passen, dann haben wir die Formel sicher falsch aufgeschrieben.² Mit Dimensionsanalyse können wir die richtige Formel leichter auswählen. Es ist wichtig, dass die Maßeinheiten wie „Stück“, „Anfrage“ usw. bedeuten keine eigene Dimension, deshalb sind beispielsweise die Dimensionen $\frac{\text{Stück}}{\text{s}}$ und $\frac{1}{\text{s}}$ gleich.

Die grundlegende Formel. Der Satz von Little: $N = X \cdot T$, $N [1] = X \left[\frac{1}{\text{s}} \right] \cdot T [\text{s}]$

Die Auslastung intuitiv und von dem Satz von Little bei einer *einzigsten exklusiven* Ressourceninstanz: $U = \frac{X}{X^{\max}} = \frac{T_{\text{busy}}}{T_{\text{measured}}} = N = X \cdot T$

Der Grenzdurchsatz nach der Ausführungszeit bei einer *einzigsten exklusiven* Ressourceninstanz (der Grenzdurchsatz ist der größte erreichbare Durchsatz, also dann ist die Auslastung 100%): $X = \frac{U}{T} \Rightarrow X^{\max} = \frac{1}{T}$

1 Leistung einer Festplatte

Eine Festplatte bedient 50 Anfragen pro Sekunde. Die Bedienung einer Anfrage braucht 0,005 Sekunden. Die Anfragen werden ohne Überlappung bedient.

- Wie hoch kann die maximal bedienbare Belastung (Ankunftsrate) sein?
- Wie hoch ist die Auslastung der Festplatte?

2 Einsichtnahme der Klausurarbeiten

Während der Einsichtnahme der Klausuren haben die Studenten die Möglichkeit, wegen eventueller Korrekturfehler zu reklamieren. Nach erfolgreicher Reklamation werden die Punktzahlen modifiziert. Einer der Lehrer kann 10 Stück pro Stunde von der ersten großen Aufgabe (F1) bewerten, oder 20 Stück von der zweiten großen Aufgabe (F2). Zu jeder Aufgabe gehört je ein Lehrer, der die gegebene Aufgabe korrigiert hat. In den Folgenden entwickeln Sie zu jeder Frage ein Prozessmodell, und bestimmen Sie, wie viele Klausuren pro Stunde in den einzelnen Fällen reklamiert werden können.

- Die Studenten reklamieren zuerst über F1, und dann F2.
- Die Aufgaben sind auf separaten Blätter geschrieben, deshalb können die einfallsreichen Studenten die einzelnen Aufgaben gleichzeitig zu verschiedenen Lehrern geben. Was können wir mit Parallelisierung gewinnen?
- Wegen der großen Masse lassen die Studenten nur eine Aufgabe überprüfen, und zwar die, deren Lehrer eben frei ist.
- Es wird unter den Studenten herumgesprochen, dass der Lehrer von F2 viel weniger streng ist, deshalb warten 80% der Studenten diese Schlange ab. Die restliche 20% reklamieren bei dem anderen Lehrer wegen F1.
- Nach der Reklamation bleibt noch 10% der Studenten, die nur 1-2 Punkte für eine bessere Note brauchen, deshalb versuchen sie wieder und wieder die Reklamation. Nehmen wir an, dass die Studenten die in der Aufgabe a. beschriebenen Strategie verwenden.
- Was würde anders passieren, wenn beide Lehrer bereit wären, Reklamationen für beide Aufgaben zu beantworten (die Überprüfung der einzelnen Aufgaben braucht dieselbe Zeit wie bisher angenommen), und die Studenten möchten wie zuerst geplant, zuerst die erste, dann die zweite Aufgabe vom Lehrer überprüfen lassen?

¹Dimensionsanalyse (Wikipedia), <https://de.wikipedia.org/wiki/Dimensionsanalyse>

²Buchtip: what if? – Droppings, <http://what-if.xkcd.com/11/>

3 Zweischichten Architektur

Sei ein Webserver (WS) und ein Datenbankverbund mit zwei Datenbankserver (DB1 und DB2) gegeben. Die Belastung wird unter den zwei Datenbankservern mit dem Round-Robin-Verfahren in einem Verhältnis 1:2 verteilt. Die Bedienung jeder Anfrage benutzt sowohl den Webserver als auch die Datenbank. Wir beobachten das System 30 Minuten lang in der Hauptzeit, in dieser Zeit werden 9000 Anfragen bedient. An den Servern wurden folgende Gesamtabfertigungszeiten gemessen: WS – 1350 Sec Prozessorbenutzungszeit; DB1 – 810 Sec, DB2 – 1320 Sec Plattenbenutzungszeit

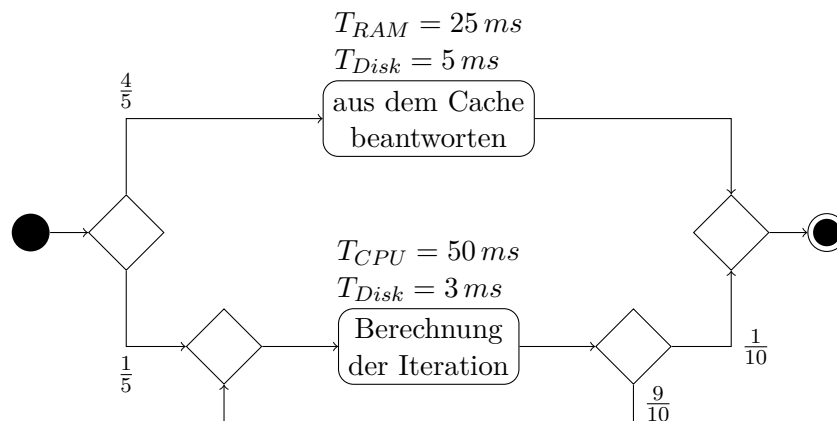
- Entwickeln Sie ein Prozessmodell über die Bedienung der Anfragen anhand der Beschreibung!
- Wie hoch ist der aktueller Durchsatz der einzelnen Server?
- Wie viele Zeit verbrauchen die einzelnen Server mit der Bedienung der einzelnen Anfragen?
- Wie hoch ist der Grenzdurchsatz des Gesamtsystems?
- Warum haben wir verschiedene Zeitparameter für die verschiedene Ressourcetypen in Betracht genommen?
- In welchem Hinsicht betrügt noch unser Modell?

4 Mikro-Service (* frühere Klausuraufgabe)

Unsere Mikro-Service implementiert eine Annäherungsalgorithmus, die von anderen Dienstleistungen als Bauelement über den Netzwerk verwendet wird, um ihre eigene Funktionen zu entwickeln. Weil viele gleichförmige Abfragen einkommen, die Service platziert den Antworten in einem Cache um die ausgerechnete Daten sofort zurückschicken zu können. Die Rechnung selbst ist ein sehr teurer, iterativer Prozess der vielmal wiederholt werden muss bis der Antwort genug pünktlich wird.

Das untene Model fasst die Erfahrungen der experimentalen Ablauf zusammen. Anhand unseren Beobachtungen 80% der Abfragen sind aus dem Cache beantwortbar, und während eine Iteration der Antwort wird genug pünktlich mit einer Wahrscheinlichkeit von 10% sein. Für die zwei verschiedene Aktivitätä haben wir auch die durchschnittliche Belegung der einzigen Ressourcen gemesst, diese Daten sind auf dem Bild unten dargestellt (nur die nicht unerhebliche Belegungen).

Für die Messung wurde ein Server mit 2 CPU Kerne, 1 RAM Modul und 1 Disk benutzt.



- Falls in jedem Schritt nur die am längsten belegte Ressource wird in Betracht genommen, wie viel kann der Durchsatz des Systems sein?
- Falls nur die in beiden Schritten benutzte, einsame gemeinsame Disk-Ressource wird in Betracht genommen, wie viel kann der Durchsatz des Systems sein?
- Wie hoch ist die wahre Durchsatz des Systems anhand diese Daten? Mit der Skalierung welcher Ressource kann es erhöht werden?
- Falls die Belegung 50% ist, und von der Empfang der ersten Abfrage bis der Übertragung des Antwortes die Prozess durchschnittlich 100ms dauert (mit dem Warten inklusiv), wie viele solche Abfragen befinden sich in dem System, die Verarbeitung von welchen noch nicht fertig ist?

5 Verkehrsnetz einer Insel (* frühere Klausuraufgabe)

Die Bewohner einer Insel durchqueren den die Insel umliegenden See jeden Morgen, wenn sie zu ihren Arbeitsplätzen unterwegs sind. An der Nordseite der Insel gibt es eine Brücke, an der Südseite eine Fähre. Die zweispurige Brücke (ein Spur pro Fahrtrichtung) ist 200 m lang, die erlaubte Höchstgeschwindigkeit ist 60 km/h, und eine Sicherheitsabstand von 30 m (von Wagenende zu Wagenende) ist

zu halten. Es gibt vier Fähren, die die Insel-Festland-Insel Rundfahrt in 15 Minuten schaffen können, und so können sie höchstens 800 Wagen pro Stunde in einer Richtung befördern.

- a. Wie hoch ist der Grenzdurchsatz der Brücke in Richtung Norden?
- b. Wie viele Wagen finden Platz in einer Fähre?
- c. Wie hoch ist der gemeinsame Grenzdurchsatz der zwei Inselausfahrten in der Hauptverkehrszeit?
- d. Um Mittag wurde die Hauptstraße am Festland wegen eines Unfalles gesperrt, und der Verkehr wurde durch die Insel (über der Brücke und den Fähren) umgeleitet. Wie hoch ist der Grenzdurchsatz der Umleitung?
- e. An einem Morgen zwischen 7:00 und 8:30 verließen 900 Wagen die Insel an Bord der Fähren. Wie hoch war zu dieser Zeit der Durchsatz und die Auslastung der Fähren?
- f. Wie viele Wagen mussten in dem vorigen Fall am Ufer gleichzeitig Schlange stehen, wenn sie gut geplant durchschnittlich eine halbe Minute vor ihrem Einstieg im Hafen angekommen sind?

6 Wissensbasis (*)

Die öffentliche Fachwissensdatenbank unserer Firma bietet einander referierende Einträge für unseren Kunden weltweit an. Die Bedienung eines einzigen Eintrages besetzt den Server 60 ms lang. Nach Lesen des Eintrages verlässt der Kunde unsere Seite in 30% der Fälle, in anderen Fällen ruft er auch weitere Einträge ab.

- a. Durchschnittlich wie viel Serverzeit braucht das Erlöschen des Gesamtwissensgiers eines Kunden?
- b. Nehmen Sie an, dass die Bedienung der Anfragen durch den Server sich nicht parallelisieren lässt. Wie viele Kunden pro Stunde können von dem Server bedient werden?