

Datenpersistenz in der Cloud

Moderne Cloud-Angebote erlauben Firmen bisher unbekannte Flexibilität für die Planung und Skalierung ihrer Datenzentren. Viele Firmen und Grossfirmen betreiben Applikationen, die auf Informationen, welche in Datenbanken gespeichert sind, angewiesen sind. Aus diesem Grund kommunizieren diese intensiv mit den Datenbanken. Die meisten dieser Applikationen werden täglich von hunderten Mitarbeitenden via Webbrowser verwendet oder laufen als Stapelverarbeitung für Berechnungen. Das Hybrid- und Public-Cloud-Modell ändert die Situation.

In dieser Bachelorarbeit werden die Möglichkeiten und Einschränkungen beim Integrieren von internen Ressourcen mit Externen aus einer Public Cloud analysiert. Verschiedene Cloud-Service-Anbieter und -Angebote werden evaluiert und verglichen. Dies wird mit Hilfe einer einfach zu verteilenden Cloud-lauffähigen Webapplikation, welche dafür entwickelt wurde, ermöglicht. Dies wird innerhalb der Cloud-Service-Anbieter, sowie mit externen Ressourcen getestet und verglichen.

Die entwickelte Applikation wurde in Java geschrieben und auf verschiedenen Cloud Umgebungen installiert. Zu diesen gehören die bekanntesten und grössten, die momentan auf dem Markt sind. Die Applikation kommuniziert mit definierten Datenbanken und testet diese mit einem standardisierten, sowie einem realen Szenario aus dem Geschäftsbereich der Swiss Re, welches man dynamisch anpassen kann.

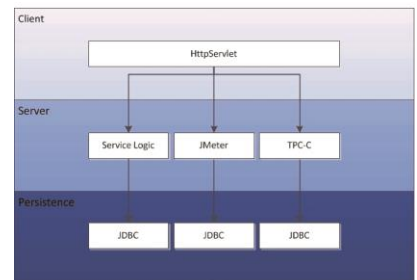
Verglichen mit traditionellen IT-Umgebungen sind die Server, welche man aus der Public Cloud bezieht, nicht lokal. Dies bedeutet, dass die Daten, welche für die Berechnungen benötigt werden, zuerst übertragen oder direkt beim Zugriff transportiert werden müssen. Dies führt zu starken Leistungseinbussen. Die zusätzlichen Ressourcen müssen früh geplant werden, da die Übertragung der Daten schon bei kleinen Mengen lange dauert. Bereits eine relativ kleine Datenbankbefüllung dauert mehr als einen halben Tag. Direkte Zugriffe auf die Datenbanken sind zwischen Faktor 2 und 100 langsamer. Bei datenintensiven Applikationen würden diese Leistungsunterschiede stark bemerkbar sein.

Für die Entwicklung und Tests gibt es einige Anwendungen der Public-Cloud-Angebote. Das hat vor allem den Grund, dass die Performanz in diesen Situationen meist nicht der wichtigste Faktor ist. Temporäre Umgebungen könnten in der Cloud abgebildet werden ohne neue Hardware zu erwerben. Sobald die Tests vorbei sind, können die Instanzen wieder freigegeben werden und es wird nur verrechnet, was man konsumiert hat.

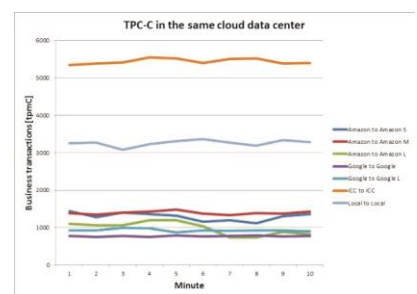


Diplomierende
Rafael Arizcorreta
Franz Faul

Dozierende
Christof Marti
Thomas Michael Bohnert



Architektur der Datenbank-Benchmark-Webapplikation, die auf Cloud-Servern installiert werden kann.



Resulte von TPC-C-Tests ausgeführt im selben Datacenter. Messzeitfenster von 10 Minuten. Die ICC Cloudverbindung ist bei weitem die schnellste.