

Praktikum 5: Skalierbarkeit mit Hadoop

Ziel des Praktikums

Ziel ist es, die Skalierbarkeit von MapReduce-Jobs bei Nutzung mehrerer Server des Clusters nachzuweisen.

Aufgabe 1 (Nachweis der Skalierbarkeit)

In diesem Praktikum erstellen Sie ein Dokument, mit dem Sie Ihren Chef überzeugen wollen in Zukunft auf skalierbare Datenbankarchitekturen zu setzen. Ihr Chef ist ein Informatiker alter Schule, der glaubt, dass man ein Ansteigen der Antwortzeiten bei zunehmender Last nur durch immer leistungsfähigere Server verhindern kann. Sie wollen ihm nun beweisen, dass man das Ziel auch durch Nutzung mehrerer Server erreichen kann, ohne dass der einzelne Server leistungsfähiger werden muss.

Ihr Dokument können Sie völlig frei gestalten. Einzige technische Rahmenbedingung: Sie müssen Ihre Aussagen durch Messungen auf unserem Cluster belegen. Um Ihr Dokument so überzeugend wie möglich zu gestalten, sollten Sie folgende Qualitätsanforderungen einhalten:

1. Beschreiben Sie im Dokument eindeutig, welches Ziel Sie verfolgen.
2. Stellen Sie eine klare Hypothese, d.h. eine Erwartung für die Ergebnisse Ihrer Messungen auf.
3. Erklären Sie, warum Ihre Messergebnisse geeignet sein werden, die Hypothese zu stützen oder zu widerlegen.
4. Definieren Sie ihre Messumgebung und Ihre Messungen vollständig.
5. Führen Sie aussagekräftige, statistisch belastbare Messungen durch.
6. Geben sie eine klare Interpretation Ihrer Messergebnisse an.
7. Beschreiben Sie, ob weiterführende Messungen sinnvoll wären und geben Sie an, welche Messungen Sie dazu in einer weiterführenden Arbeit durchführen könnten.
8. Fassen Sie ihre Ergebnisse zu einer klaren Aussage und Empfehlung zusammen.

Technische und organisatorische Rahmenbedingungen:

1. Es steht Ihnen die Hadoop-Installation unseres Clusters zur Verfügung.
2. Der Cluster wird jeweils für ein bzw. mehrere Tage als Cluster mit acht, vier und zwei Servern konfiguriert. Genaue Zeitangaben finden Sie ab Juni in Moodle
3. Nutzen Sie einen Datenbestand von mindestens 10 GB¹

¹Eine Möglichkeit dazu sind die Hadoop Map Reduce Examples `hadoop-mapreduce-examples-2.7.5.jar` mit der Option `randomtextwriter`

Praktikumsbericht

Abgabe des Praktikumsberichts eine Woche nach dem jeweiligen Praktikumstermin.