

Performanzanalyse von Key-Value-Datenspeichern

Performance Analysis of Key-Value Stores

Motivation

Neben den relationalen Datenbanksystemen gewinnen auch NoSQL-Datenbanksysteme [1] immer mehr an Bedeutung. Diesen NoSQL-Datenbanksystemen liegen verschiedene Datenbankmodelle zugrunde. Eines dieser Datenbankmodelle ist das Key-Value-Datenmodell, das aufgrund des sehr einfach gehaltenen Schemas schnelle Zugriffe auf die Werte bieten, wenn die Schlüssel bekannt sind.

Die Art der Implementierungen von Key-Value-Datenspeichern reichen von Bibliotheken, die die **Cache**¹- oder **Map**-Schnittstelle (z. B. JCache-API) implementieren und die Daten persistieren (z. B. MapDB²) bis hin zu eigenständigen clusterfähigen Key-Value-Datenbanksystemen (z. B. Riak KV³).

Je nach Anwendungsfall ist die eine Implementierungsstrategie besser geeignet als die andere. Da clusterfähige Key-Value-Datenbanksysteme die Komplexität der Testinfrastruktur erhöhen, wird beim Testen oft auf einen eingebetteten Key-Value-Datenspeicher (z. B. eine persistente Map) zurückgegriffen. Hierbei wäre es wissenswert, ob durch diese Substitution auch Annahmen bei der Performanz gemacht werden können.

Als Anwendungsfall kann von der Speicherung von Seiten eines Datenbankpuffers [2] ausgegangen werden. Als Schlüssel wird ein Hashwert (z. B. *sha256*) und als Wert Byte-Arrays mit festen Längen von typischen Seitengrößen in Datenbankpuffern wie beispielsweise 4 oder 16 KB angenommen.

Aufgabenstellung

In diesem Forschungsprojekt soll die Performanz von verschiedenen persistenten Key-Value-Datenspeichern anhand von synthetischen Testdaten analysiert werden. Hierfür müssen zuerst nach verschiedenen Key-Value-Datenspeichern recherchiert und in die jeweilige Implementierungsstrategie eingeordnet sowie nach Eigenschaften wie beispielsweise Clusterfähigkeit, Integrationskomplexität in eine Anwendung, Replizierbarkeit und Skalierbarkeit kategorisiert werden. Danach sollen synthetische Testdaten generiert und ausgesuchte⁴ Key-Value-Datenspeicher je Implementierungsstrategie auf ihre Performanz evaluiert werden.

Daher umfasst die Arbeit folgende Aufgaben:

¹<https://javadoc.io/doc/javafx.cache/cache-api/latest/index.html>

²<https://mapdb.org>

³<https://riak.com/products/riak-kv/index.html>

⁴Nach Rücksprache mit dem Betreuer

- 1) Einarbeitung in das Themengebiet der Key-Value-Datenspeicher und Recherche zu bestehenden Key-Value-Datenspeichern
- 2) Eingruppierung der Key-Value-Datenspeicher in die jeweiligen Implementierungsstrategien
- 3) Kategorisierung der Key-Value-Datenspeicher anhand ihrer Eigenschaften wie beispielsweise Clusterfähigkeit, Integrationskomplexität in eine Anwendung, Replizierbarkeit und Skalierbarkeit
- 4) Generierung von synthetischen Testdaten mit Hashwerten als Schlüssel (z. B. *sha256*) und Byte-Arrays mit Festlänge als Wert
- 5) Durchführung der Performanzanalysen für ausgesuchte Key-Value-Datenspeicher je Implementierungsstrategie mithilfe der in Punkt 4 generierten synthetischen Testdaten

Die Ergebnisse müssen in einem Zwischen- und Abschlussvortrag präsentiert werden. Es wird erwartet, dass die Arbeitsergebnisse mit einer uneingeschränkten Nutzungslizenz (z. B. Apache 2.0) zur Verfügung gestellt werden.

Voraussetzungen

- ◇ Vorkenntnisse im Bereich der Datenbanksysteme und des Key-Value-Datenmodells
- ◇ Mindestens gute Programmierkenntnisse

Literatur

- [1] G. Vaish, *Getting started with NoSQL*. Packt Publishing Ltd, 2013.
- [2] A. Kemper and A. Eickler, *Datenbanksysteme*. Oldenbourg Wissenschaftsverlag, 2009.

Organisatorisches

Art der Arbeit	Bachelor-Forschungsprojekt Informatik
Titel der Arbeit (de)	Performanzanalyse von Key-Value-Datenspeichern
Titel der Arbeit (en)	Performance Analysis of Key-Value Stores
Betreuer	Dennis Przytarski, M.Sc.
Prüfer	Prof. Dr.-Ing. habil. Bernhard Mitschang
Beginn	ab sofort