



Performance Analyse der Optimierung von Datenbankabfragen in der HANA Calculation Engine

Projektarbeit 2

im Rahmen der Prüfung zum
Bachelor of Science (B.Sc.)

des Studienganges Wirtschaftsinformatik
an der Dualen Hochschule Baden-Württemberg Mannheim

von

Jared Heinrich

Abgabedatum:	26. August 2024
Bearbeitungszeitraum:	06.05.2024 - 25.08.2024
Matrikelnummer, Kurs:	5101479, WWI22SEA
Ausbildungsfirma:	SAP SE Dietmar-Hopp-Allee 16 69190 Walldorf, Deutschland
Unternehmensbetreuer:	Rainer Agelek
Wissenschaftlicher Betreuer:	Prof. Dr. Hans-Henning Pagnia

Ehrenwörtliche Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit mit dem Thema:

*Performance Analyse der Optimierung von Datenbankabfragen in der HANA
Calculation Engine*

selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Ich versichere zudem, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

Mannheim, den 10. Juni 2024

Heinrich, Jared

Inhaltsverzeichnis

Abkürzungsverzeichnis	III
Abbildungsverzeichnis	IV
Tabellenverzeichnis	V
Quellcodeverzeichnis	VI
1 Einleitung	1
2 Grundlagen	2
2.1 Calculation Engine	2
2.2 Ausführung von Datenbankabfragen	2
2.3 Verschiedene Arten von Performance Analyse	2
3 Aktueller Stand der Performance Analyse in der Calculation Engine	3
3.1 Google Benchmark	3
3.2 HANA Profiler	4
4 Benchmarking-Konzept	5
4.1 Messung	5
4.2 Testdaten	5
5 Umsetzung	6
5.1 Implementation der neuen Methode	6
6 Verifikation	7
7 Fazit und Ausblick	8
Literaturverzeichnis	VII

Abkürzungsverzeichnis

CE Calculation Engine

Abbildungsverzeichnis

3.1 Google Benchmark Ausgabe	4
--	---

Tabellenverzeichnis

Quellcodeverzeichnis

1 Einleitung

2 Grundlagen

2.1 Calculation Engine

2.2 Ausführung von Datenbankabfragen

Datenbankabfragen werden in der Calculation Engine (CE) in drei Schritten ausgeführt. Zuerst wird die Calculation View instanziiert. Anschließend wird dieses mithilfe verschiedener Methoden optimiert. Dieses optimierte Modell wird dann auf der Datenbank ausgeführt.

Wo genau werden die Abfragen ausgeführt? Was gehört alles zur CalcEngine?

2.3 Verschiedene Arten von Performance Analyse

Benchmarking

Profiling

3 Aktueller Stand der Performance Analyse in der Calculation Engine

Die CE nutzt bereits verschiedene Wege um die Performance und das Verhalten der HANA Datenbank zu analysieren.

3.1 Google Benchmark

Google Benchmark ist ein Open Source Benchmarking Tool von Google, welches es einem ermöglicht einzelne Funktionen zu benchmarken.

Dazu kann man ähnlich zu den meisten Test Frameworks drei verschiedene Funktionen definieren.

Die Hauptfunktion definiert was genau im Benchmark untersucht werden soll. Diese legt die für die Messung relevante Logik fest.

Die Setup Funktion wird einmal vor jeder Ausführung der Hauptfunktion aufgerufen. In dieser werden die Voraussetzungen für die Ausführung der Hauptfunktion geschaffen. Man könnte sie beispielsweise nutzen, um Testdaten für den Benchmark zu generieren oder zu laden, da sie nicht bei den Messungen beachtet wird.

Die Teardown Funktion wird nach jeder Ausführung der Hauptfunktion aufgerufen. Diese beeinflusst, wie die Setup Funktion, die Messung nicht.

Des Weiteren bietet Google Benchmark die Option, einen Benchmark mehrmals mit unterschiedlichen Parametern durchzuführen. Um beispielsweise die Auswirkung der Größe des Testdatensatzes auf das Ergebnis zu beobachten.

Abbildung 3.1 zeigt eine Beispielhafte Google Benchmark Ausgabe. Benchmark ist dabei der Name des Benchmarks und hinter dem Schrägstrich die Größe des Inputs. Time ist die durchschnittliche Dauer einer Ausführung der Hauptfunktion über alle Iterationen hinweg. CPU ist die durchschnittliche CPU Zeit einer Ausführung der Hauptfunktion

über alle Iterationen hinweg. Iterations ist die Anzahl der durchgeführten Iterationen, bis ein stabiler Durchschnittswert erreicht wurde. (Vgl. Google 2024)

Benchmark	Time	CPU	Iterations
BM_RemoveDetachedNodes/2	4 us	4 us	189823
BM_RemoveDetachedNodes/8	9 us	9 us	78165
BM_RemoveDetachedNodes/64	58 us	58 us	12106
BM_RemoveDetachedNodes/512	469 us	468 us	1515
BM_RemoveDetachedNodes/4096	4237 us	4234 us	166
BM_RemoveDetachedNodes/8192	10026 us	10019 us	60

Abbildung 3.1: Google Benchmark Ausgabe

Google Benchmark wird in der CE meistens genutzt, um das Verhalten der Laufzeit bestimmter Funktionen in künstlich generierten Testfällen zu vergleichen. Hierbei wird jedoch keine HANA Instanz benötigt, da keine Operationen auf einer Datenbank ausgeführt werden. Es werden folglich auch keine Testdatensätze für die Datenbank benötigt, sondern nur Daten, auf welchen man die zu analysierende Methode aufrufen kann. Solche Performance Tests sind also im Vergleich zu Tests, welche auf eine HANA Instanz mit Testdatensätzen zugreifen, relativ einfach. Da sie sich weniger Voraussetzungen haben, und nur auf einen kleinen Teil der Logik fokussiert sind.

3.2 HANA Profiler

4 Benchmarking-Konzept

4.1 Messung

Um Informationen über die benötigte CPU-Zeit in verschiedenen Schritten der Optimierung herauszufinden, eignet sich der in Abschnitt 3.2 angesprochene Booss-Profiler sehr gut.

4.2 Testdaten

5 Umsetzung

5.1 Implementation der neuen Methode

6 Verifikation

7 Fazit und Ausblick

Literaturverzeichnis

Google (2024). *benchmark A microbenchmark support library*. URL: <https://google.github.io/benchmark/> (Einsichtnahme: 10.06.2024).