

Evaluierung von NoSQL-Datenbanksystemen



SP Consulting GmbH

Udo Flory (Geschäftsführer)
Mundenheimer Straße 141
67061 Ludwigshafen
Tel.: +49 (0)621 9534 957-0
E-Mail: info@sp-con.de
www.sp-con.de

hochschule mannheim



Hochschule Mannheim

Prof. Dr. Michael Gröschel
Prof. Dr. Sachar Paulus
Fakultät für Informatik
Paul-Wittsack-Straße 10
D-68163 Mannheim
Tel.: +49 (0) 621 292 6764
E-Mail:
m.groeschel@hs-mannheim.de
s.paulus@hs-mannheim.de
www.informatik.hs-mannheim.de



Das Projekt wurde im Wintersemester 2016/17 von folgenden Studierenden des Studienganges Unternehmens- und Wirtschaftsinformatik in 4 Teams bearbeitet (in alphabetischer Reihenfolge): Marcel Eberling, Christian Garske, Marcel Geyer, Maximilian Groß, Johannes Heiler, Svenja Hering, Tobias Jünemann, Lucas Kneis, Elbasan Krasniqi, Henrik Lawall, Qingyao Liu, David Lupa, Jonathan Markgraf, Daud Nasir, Marvin Ostermaier, Natalina Radosevic, Andrej Russkov, Philipp Scholz, Katharina Spinner, Ilker Turan, Steffen Wagner, Regina Wajant, Büsra Yildiz, Mei Yuan

Gröschel, Stand: 15.08.2017

Der stetige Anstieg der Datenmenge in der heutigen Zeit stellt in der Praxis eine technisch hohe Herausforderung dar, und nicht alle Datenbanksysteme sind dafür gleich gut geeignet. Dies wurde an der Hochschule Mannheim in einem Unternehmens-Informatik-Projekt untersucht. Ausgangspunkt war eine Anwendung, die beim Aggregieren und Pivotieren der Datensätze nicht die gewünschte Performance von Antwortzeiten im Sekundenbereich lieferte. Durch die Evaluation von zwölf alternativen Datenbanktechnologien wurde zunächst eine Vorauswahl von vier geeigneten Systemen getroffen: Die spaltenbasierten Datenbanken HBase und Cassandra, sowie die dokumentenbasierten Datenbanken MongoDB sowie Couchbase. Im Detail wurden die Kriterien Installationsaufwand, Performance, Skalierbarkeit und Benutzungsfreundlichkeit untersucht. HBase wird nicht für den Einsatz empfohlen; die übrigen drei Produkte und Technologien sind vielversprechend.

Ausgangssituation und Anforderung

Der Ausgangspunkt des Projekts war die Herausforderung, große Datenmengen performant aggregieren und zu pivotieren zu können. Die aktuelle Lösung des Auftraggebers bestand aus einer Anwendung zur Analyse von Vorhersagen auf Basis einer relationalen Oracle-Datenbank, welche den geforderten Performancekriterien nicht entsprach. Konkret ging es um eine Reihe von KPIs, die im Zuge einer Reporterzeugung, wie sie in Managementpositionen zum Alltagsgeschäft gehört, berechnet werden sollten. Diese Berechnung, inklusive dem Anzeigen der abgefragten Daten, sollte eine Abfragezeit von fünf Sekunden nicht überschreiten, dauerte aber oft mehrere Minuten.

Zielsetzung

Ziel des Projektes war es, eine NoSQL-Datenbank zu finden, die sich für den beschriebenen Anwendungsfall eignet. Um eine solche Empfehlung aussprechen zu können, war es notwendig, eine entsprechende Evaluation durchzuführen. Im Zuge dieser galt es, ausgewählte Datenbanktechnologien im produktiven Einsatz zu testen. Dies beinhaltete das Einspielen von Testdaten, sowie das Übertragen und Implementieren einer Abfragelogik entsprechend der Vorgaben des Auftraggebers. Im Vordergrund der Evaluation stand die Vergleichbarkeit der jeweiligen Gruppenergebnisse. Hierzu wurden alle Systeme einheitlichen Performance-Tests (mit JMeter) unterzogen, um anhand dessen eine Empfehlung für den Auftraggeber aussprechen zu können.

Projektaufgaben und -phasen

- *Meilenstein 1 – Vorschlag:* In dieser Projektphase war es die Aufgabe der Gesamtgruppe, eine Reihe von Datenbanksystemen auf ihre Anwendbarkeit im gegebenen Kontext zu untersuchen.

Dabei wurde eine Vorauswahl von zwölf Technologien auf 4 reduziert, welche sich zu diesem Zeitpunkt als Favoriten hervorgetan hatten. Jedes der Projektteams erhielt eine Technologie zur näheren Evaluierung.

- *Meilenstein 2 – Entwurf:* In der zweiten Phase des Projektes wurden die verschiedenen NoSQL-Lösungen installiert und die Grundzüge der späteren Architektur festgelegt.
- *Meilenstein 3 – Prototyp:* In dieser Phase sollte eine erste Version der Lösung erstellt werden, welche die Grundfunktionen erfüllt. Dazu gehörte auch, erste Abfragen an die Datenbank senden und verarbeiten zu können.
- *Meilenstein 4 – Optimierungen:* Die erste Version sollte nun erweitert und verbessert werden. Hierzu floss beispielsweise Feedback des Auftraggebers ein. Am Ende der Phase stand ein fertiger Prototyp der Lösung, der alle Anforderungen erfüllte.
- *Meilenstein 5 – Abgabe:* Das fertige Produkt war bereit zur Abnahme und besitzt alle Funktionen, die den Anforderungen des Kunden entsprechen.

Projektergebnisse

Ein projektbegleitendes Anforderungsdokument spiegelt die Zielsetzung des Projekts wider. Die getroffene Technologieauswahl stützt sich auf das Dokument [1]. Hierin wird eine Vorauswahl für den Anwendungsfall möglicherweise geeigneter Technologien getroffen. Im Ergebnisdokument „Evaluation von NoSQL Datenbanksystemen in Bezug auf Performance eines Forecast-Systems“ wird eine Empfehlung zu diesen ausgewählten Technologien gegeben. Diese stützt sich auf definierte, einheitliche Bewertungskriterien. Zu jeder eingesetzten Technologie gibt es eine Implementierung zur Berechnung der für den Anwendungsfall relevanten Kennzahlen.

Projektmanagement

Für das Projekt und zum Erreichen des geforderten Ergebnisses, stand den Studierenden die Wahl der Projektmanagementmethode frei. Da sich einige Anforderungen erst über den Projektzeitraum entwickelten, entschied sich jedes Team für die agile Methodik KANBAN. Diese wurde von manchen Teams mit der Projektmethodik SCRUM kombiniert. Es wurden wöchentliche Statusmeetings abgehalten sowie tägliche Stand-Up-Meetings. Jedes Team arbeitete mit einem KANBAN-Board, welches der Visualisierung der Aufgaben und Zuständigkeiten diente. Um den Kundenkontakt stets aufrecht zu erhalten, stellten die Studierenden ein virtuelles Team zusammen, welches für die Kommunikation zuständig war.

Fazit

Alle vier behandelten Datenbanken haben ihre Vor- und Nachteile. Jedoch kann man mit Sicherheit HBase als geeignete Lösung ausschließen, denn der manuelle Aufwand ist zu hoch und bei der betrachteten Konkurrenz nicht empfehlenswert. Im Gegensatz dazu bieten sich die übrigen Technologien sehr gut für die Erfüllung der Anforderungen an.

Weitere Informationen

[1] Faculty of Computer Science, Mannheim University of Applied Sciences (Hrsg.): Betrachtung von 12 Datenbanktechnologien und deren Konzepte der Datenverarbeitung im Kontext Big Data, <https://doi.org/10.13140/RG.2.2.32598.24643>