

Proactive horizontal scaling of stateless applications based on historical data

Mit dem Aufkommen von Cloud-Anbietern und der Anpassung von Anwendungen an die durch ihren Einsatz gewonnene Elastizität werden Ansätze nötig, um diese neuen Möglichkeiten zu nutzen. Mit automatischer Skalierung können Administratoren diese Funktionen in ihre Anwendungen und Infrastruktur integrieren ohne die Skalierung einer Bereitstellung für Anwendungen, deren Anforderungen sich periodisch ändert, manuell anpassen zu müssen. Obwohl der traditionelle, reaktive Ansatz weit verbreitet ist, hat er Einschränkungen hinsichtlich der Ressourcennutzung und der Einhaltung von definierten Service-Level-Agreements.

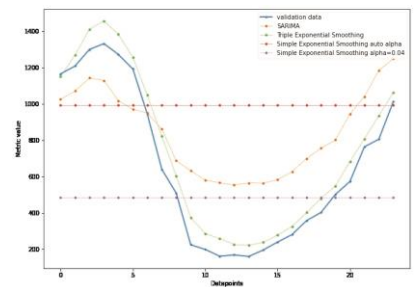
In dieser Arbeit werden die Unzulänglichkeiten des reaktiven Ansatzes aufgezeigt und verglichen. Es wird eine alternative Methode vorgestellt, um mögliche Verbesserungen bezüglich Ressourcennutzung und Performance zu untersuchen, wenn anstelle eines reaktiven Ansatzes ein proaktiver, vorausschauender Ansatz gewählt wird. Um die Fragestellung der Arbeit zu beantworten, werden geeignete Methoden zur Erstellung genauer Prognosen evaluiert und eine eigene Lösung implementiert. Die Lösung wird dann anhand von vier Testscenarien, die Zugriffe auf eine generischen Webanwendung simulieren, verifiziert.

Die Auswertungen zeigen, dass die implementierte Lösung die Gesamtressourcennutzung um 3.98 % und die durchschnittlichen Antwortzeiten um 8.42 % über einen kurzen Zeitraum und einen relativ kleinen Umfang des Testscenarios im Vergleich zu den herkömmlichen, reaktiven Ansätzen reduzieren kann. Diese Erkenntnisse können dabei helfen, bessere Skalierungsmethoden für verschiedene Anwendungsfälle zu entwickeln, z. B. für containerisierte Workloads oder Anwendungen, die auf PaaS-Lösungen von Hyperscalern bereitgestellt werden. Durch die Implementierung solcher Ansätze können die Gesamtressourcennutzung und der damit verbundene Fussabdruck der Anwendungen deutlich reduziert werden.

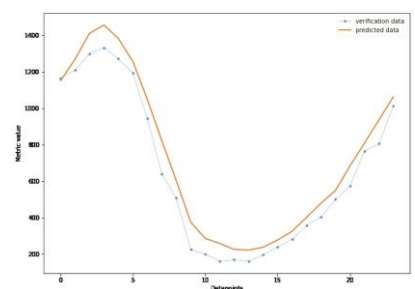


Diplomand
Lorenz Kopczynski

Dozent
Patrick Baumgartner



Die Vorhersagen der evaluierten, statistischen Modelle im Vergleich zu den tatsächlichen, gemessenen Werten.



Die Vorhersagen des gewählten Triple-Exponential-Smoothing-Modelles, welches in der Arbeit zur Beantwortung der Fragen zum Einsatz kommt.