

Performance Anforderungen

Als Teil der nicht-funktionalen Anforderungen beschreiben die Performance Anforderungen die gewünschte Leistungsfähigkeit einer Anwendung. Diese Anforderungen können grundsätzlich für die verschiedenen Teststufen existieren.

Wichtige Anforderungen müssen zum Projektstart schon initial festgelegt und frühzeitig getestet werden, denn gerade das nicht Einhalten von wesentlichen Performance Anforderungen kann strukturelle Änderungen, bis hin zum Architekturwechsel, notwendig machen. Die alte Weisheit, je später ein Fehler entdeckt wird, desto teurer wird die Fehlerbeseitigung trifft hier erst recht zu. Im weiteren Verlauf des Projektes, oder in der Wartung, müssen Anforderungen gegebenenfalls angepasst oder weitere hinzugenommen werden.

Zur Definition der Performance Anforderungen gehört auch eine Beschreibung der Randbedingungen für einen durchzuführenden Last- und Performance-Test, wie zum Beispiel des Testsystems, der Testdauer, der Testdaten oder wie Testfälle für das Szenario ermittelt werden sollen.

Entwicklung und Durchführung von Lasttests können sehr aufwändig sein hinsichtlich Kosten und Zeit. Das dafür vorgesehene Budget muss bei der Definition der Anforderungen mitberücksichtigt werden.

Bei der Definition einer Anforderung muss unbedingt überprüft werden ob diese überhaupt mit einem vertretbaren Aufwand unter Lastbedingungen Test- und Messbar ist. Die zu nutzende Messmethode sollte idealerweise mit festgelegt werden.

Performance Anforderungen können in folgende drei Bereiche unterteilt werden:

1) Belastbarkeit

Diese Anforderungen dienen zur Beschreibung der Last, unter die die Anwendung gesetzt werden können muss bezüglich Durchsatz, Kapazität und Konfiguration. Da unter unterschiedlicher Last auch unterschiedliches Verhalten der Anwendung bezüglich Benutzer Antwortzeiten und System Ressourcen Nutzung zu erwarten ist, wird zwischen verschiedenen Laststufen, wie Hoch Last und Maximaler Last, unterschieden.

Für jede Anforderung sollte ein Wertebereich (Minimum – Maximum) definiert werden: nur wenn die Werte aller dieser Anforderungen in ihrem Wertebereich liegen, liegt die entsprechende Laststufe vor. Ein guter Wert für das Minimum könnte zum Beispiel Maximum – 10% sein. Den korrekten Wertebereich für die einzelnen Laststufen zu finden ist meistens schwierig, da die unterschiedlichen Anforderungen

häufig voneinander abhängen. Hier müssen (erwartete) betrieblich Werte als Ausgangsbasis dienen.

Für alle Anforderungen zusammen wird eine einheitliche Zeiteinheit definiert, in der die jeweiligen Werte gelten. Diese Zeiteinheit darf nicht zu klein gewählt werden, damit etwaige Varianzen geglättet werden.

Es muss auch definiert werden, wie sich die Anwendung in Überlast, also nach Überschreitung der maximalen Last verhalten soll.

2) Benutzer Antwortzeiten

Die verschiedenen Eingaben werden, auch differenziert zwischen abgewiesenen und erfolgreichen, eindeutigen Antwortzeitentypen zugewiesen und somit gruppiert.

Da ja normalerweise nicht alle möglichen Eingaben in einem Lasttest ausgeführt werden können, darf eine entsprechende Vorgabe nicht fehlen.

Es kann auch sinnvoll sein ein Mischungsverhältnis zwischen den einzelnen Antwortzeitentypen vorzugeben, allerdings kompliziert das die Entwicklung eines Lasttest Szenarios und kann dazu führen, dass Anforderungen bezüglich Kapazität und Durchsatz anzupassen sind. Dabei muss unbedingt auf betriebsnähe geachtet werden.

Wichtig ist auch die genaue Definition der Messpunkte T0 (=Zeitpunkt der Eingabe) und Tend (=Zeitpunkt der Ausgabe), welche sich idealerweise in der eigenen, kontrollierbaren, Umgebung befinden. Die jeweilige Antwortzeit berechnet sich dann als $Tend - T0$.

Pro Antwortzeitentyp werden Quantil/Perzentile und Maximalwerte pro Laststufe definiert.

Tragen externe Services zur Antwortzeit bei, so muss definiert werden wie damit umgegangen wird. Das hat unter Umständen Einfluss auf die Entwicklung von Stubs oder der Service Virtualisierung.

3) System Ressourcen Nutzung

Mit diesen Anforderungen wird die Nutzung von Hardware Ressourcen wie CPU, Memory, Plattenplatz, Netzwerk-Bandbreite für jeden System-Typ definiert. Somit helfen sie die zur Anwendung passende Hardware auszuwählen.

Es werden Quantil/Perzentile und Maximalwerte, jeweils pro Zeiteinheit, definiert. Außerdem können unterschiedliche Werte für die einzelnen Laststufen definiert werden.

Um überhaupt aussagekräftige Daten zu bekommen sollten die Zeiteinheiten sinnvoll gewählt werden: Für CPU und Netzwerk ist es am besten sekundlich zu messen, während für Memory und erst Recht Plattenplatz größere Zeiteinheiten gewählt werden können. Neben Anforderungen für Gesamt-Werte können auch Werte für einzelne Anwendungs-Prozesse definiert werden.

Wichtig ist es zu erwähnen, dass die Ressourcen Nutzung der Messprozesse selbst aus den gemessenen Gesamt-Werten herauszurechnen ist.

Harald Klos