

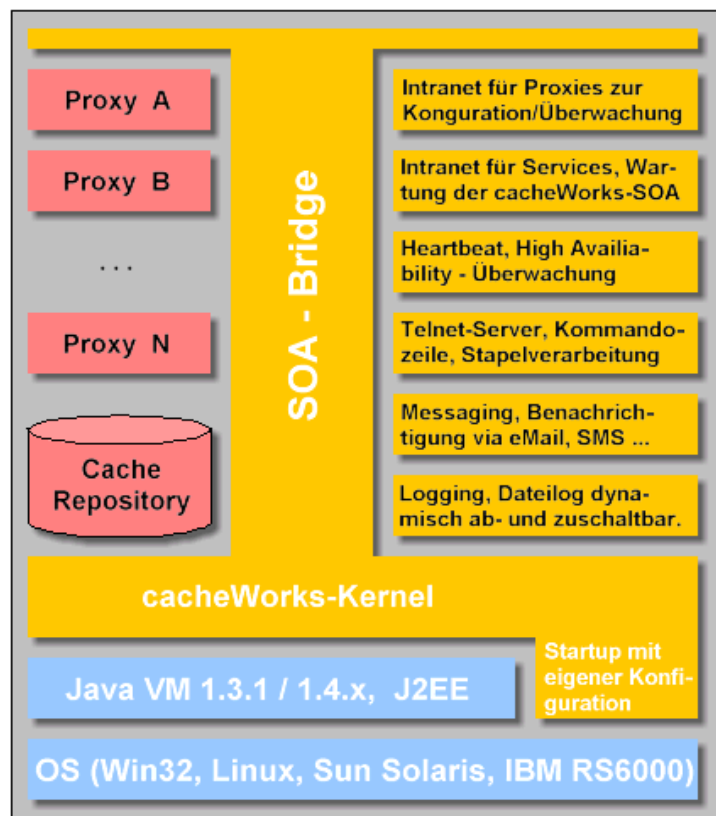
Architektur und Proxysystem von cacheWorks

cacheWorks ist eine frei konfigurierbare Zwischenschicht zwischen den contentausliefernden Instanzen (Apache Server, Microsoft IIS, ...) und den Content - Generatoren (CMS - Systeme, CGI-Bin, JSP/ASP - Container, ...). Sie wurde entwickelt, um eine schnelle, sichere Auslieferung des Contents zu garantieren.

Architektur von cacheWorks – Service Oriented Architecture (SOA)

cacheWorks wurde streng nach der SOA (Service – Oriented - Architecture) entwickelt, das heißt, dass die Funktionalitäten von cacheWorks über eigene, voneinander unabhängige Teilbereiche („Services“) angesprochen werden. So ist es möglich, Konfigurationen im System zu ändern, ohne dass der cacheWorks – Server gestoppt werden muss (echtes 24 x 7). Die wichtigsten Schnittstellen dabei sind das eigene Intranet zur Steuerung der Services und Online - Konfiguration der Proxies, die Kommandozeile von cacheWorks, und die beliebig zu- und abschaltbaren Proxy - Instanzen, die jeweils eine Verbindung zwischen der Auslieferung und einem Generator darstellen. Die Proxy – Instanzen sind ebenfalls nach dem SOA – Prinzip als eigene Services entwickelt und können so im lebenden System zu- und abgeschaltet sowie umkonfiguriert werden.

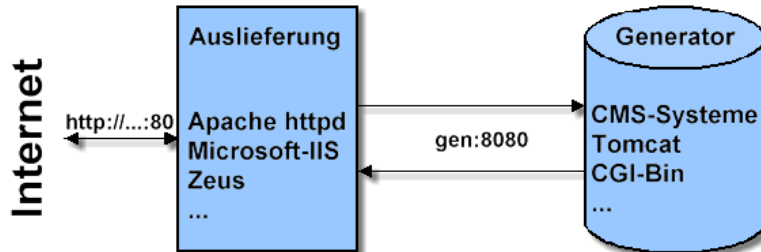
cacheWorks - schematischer Aufbau



Klassische Auslieferung im Internet

Im Gegensatz zur klassischen Auslieferung, die den Content nur "on request" erzeugt, fragt bei cacheWorks die ausliefernde Instanz die Content - Generatoren nicht direkt ab, sondern wird über die Proxies in cacheWorks geroutet. cacheWorks speichert dabei den Content in einem eigenen Cache zwischen, so dass er auch bei einem Generatorausfall noch zur Verfügung steht.

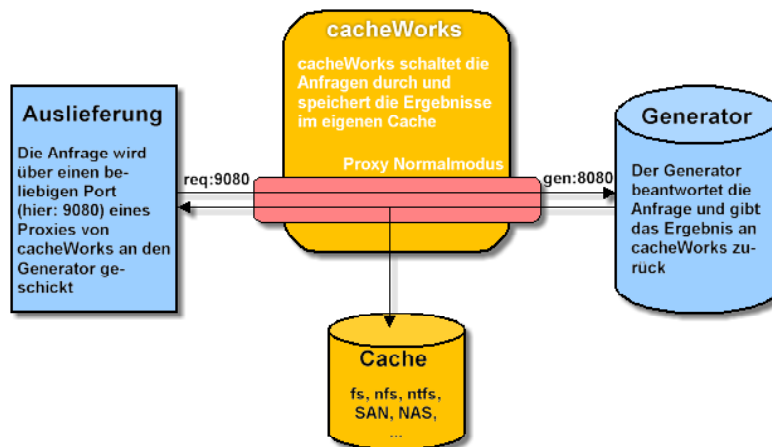
Klassische Auslieferung von dynamischem Content im Internet



„Normale“ Auslieferung mit cacheWorks

Wie die Anfragen behandelt und die Auslieferung vonstatten geht wird dabei durch die verschiedenen Auslieferungsarten bestimmt. Dazu werden die oben erwähnten, frei definierbaren „Proxies“ in cacheWorks konfiguriert, die die jeweilige Auslieferungsart bestimmen. Der einfachste Fall einer solchen Konfiguration ist ein Proxy im Normalmodus. Hier werden die Anfragen durch cacheWorks hindurch direkt an die Generatoren gestellt und das erhaltene Ergebnis an die Auslieferungsinstanzen zurückgegeben. Die Anzahl der gleichzeitigen Anfragen an einen Generator kann dabei durch cacheWorks skaliert werden, so dass es nicht vorkommen kann, dass zu viele Anfragen den Generator „überhitzen“ und zum Absturz bringen. Das Ergebnis wird dabei ebenfalls im Cache zur späteren Nutzung im Ausfallsfall oder bei einer Modusänderung abgelegt.

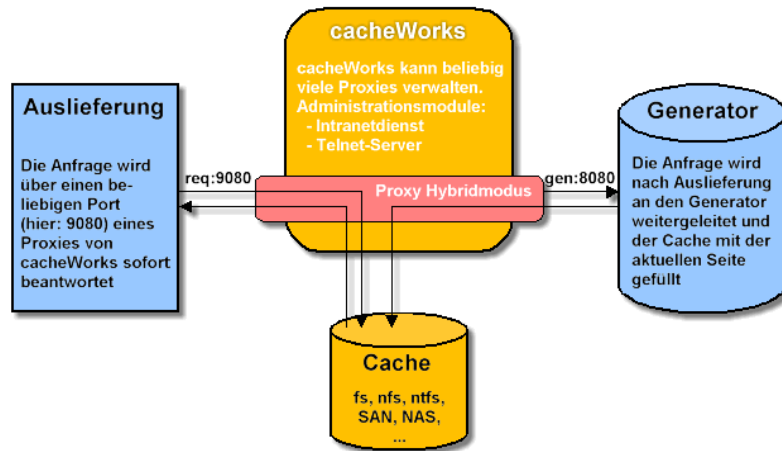
Auslieferung mit cacheWorks im Modus "normal"



Performanzsteigerung durch den Modus „hybrid“

Ein Merkmal von cacheWorks ist die Skalierung der Anfragen an den Generator mit einer pro Proxy frei konfigurierbaren Maximalanzahl an offenen Verbindungen zum Generator. Dadurch kann es allerdings zu Verzögerungen bei der Auslieferung kommen, gerade wenn sehr viele Anfragen gestellt werden oder die Generierung von Content im Generator viele Ressourcen benötigt.

Auslieferung mit cacheWorks im Modus "hybrid"

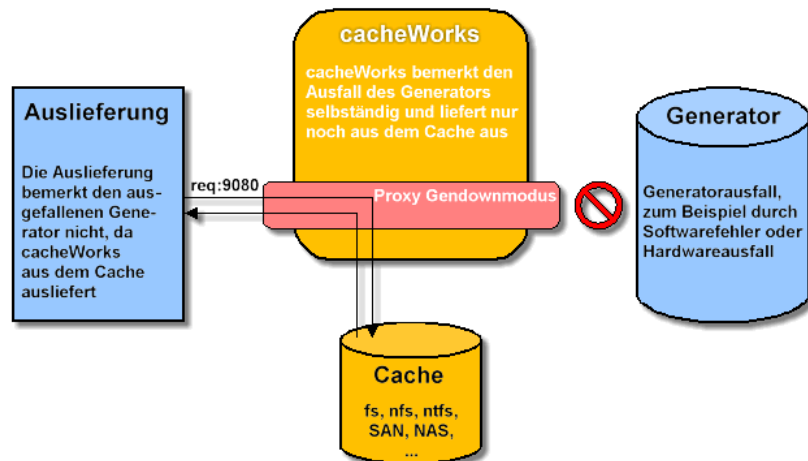


Um trotzdem jederzeit schnell auszuliefern kann man in cacheWorks Proxies in den Modus „hybrid“ schalten. In diesem Modus wird zuerst versucht, jede Anfrage aus dem Cache von cacheWorks heraus zu beantworten. Wird kein gecachtes Fragment gefunden, so wird es wie im Modus „normal“ vom Generator abgeholt, ansonsten wird das gecachte Fragment sofort an die ausliefernde Instanz zurückgegeben. Gleichzeitig wird in diesem Fall die Generierung erneut angestoßen und die neu erzeugte Version im Cache gespeichert. So wird gewährleistet, dass der Cache selbst immer in einem Zustand gehalten wird, der so aktuell wie möglich ist („sekundenaktuell“). Dieser Modus entlastet zum Beispiel aktuelle CMS – Systeme extrem und garantiert eine Auslieferung, die nahezu den Merkmalen einer statischen Auslieferung entspricht.

(Geplanten) Ausfall von Generatoren abfangen: Modus „gendown“

Durch das intelligente Caching in cacheWorks und das separierte Vorhalten des Contents kann cacheWorks einen Ausfall von einem/mehreren Generatoren überbrücken. Dabei bemerkt cacheWorks den Ausfall selbständig dadurch, dass die Generatoren nicht mehr innerhalb einer vorgegebenen Zeit reagieren und spricht sie mit einem definierten Request an. Ist der Generator ausgefallen,

Auslieferung mit cacheWorks im Modus "gendown"



so schaltet cacheWorks automatisch in den „gendown“ und benachrichtigt den Administrator vom Ausfall. Der Modus „gendown“ ermöglicht aber auch einen „geplanten“ Ausfall des Generators. Da zu jeder Zeit bereits im Cache vorhandener Content ausgeliefert wird ist es möglich, cacheWorks per Intranet oder Kommandozeile definiert in den Modus „gendown“ zu schalten, um so Wartungsarbeiten an den Generatoren oder Generatormaschinen vorzunehmen (z.B. bei Softwareupdates, Umstellung der DV-Landschaft etc.). Nach außen hin liefert cacheWorks trotzdem den zuletzt online gesetzten und abgerufenen Content aus.