

Praxis Knowhow: Skalierung von SOA Suite 11g Cluster

Gabriel Lobstein
OPITZ CONSULTING München GmbH
München

Schlüsselworte

Oracle SOA Suite, Administration, Hochverfügbarkeit, Skalierbarkeit, Scale Up, Scale Out.

Einleitung

In vielen Projekten werden inzwischen Oracle SOA Suite 11g Cluster erfolgreich betrieben. Kommt dann ein SOA Konzept in der eigenen Organisation erst mal richtig ins Laufen, steigt auch schnell die Last, da sich die Nutzung der bereitgestellten Services immer weiter intensiviert. In diesem Vortrag werden wichtige Erwägungen bei der Skalierung der SOA Suite dargestellt und zudem praktisch beschrieben, welche Schritte vorzunehmen sind um die Leistung der SOA Plattform zu verbessern.

Steigende Last – ein Zeichen für Erfolg

Eine erfolgreiche SOA Strategie führt zu Services mit einem hohen Wiederverwendungsgrad. Zudem kommen im Laufe der Zeit weitere Services hinzu. Dies führt zu steigender Last auf der SOA Plattform. Um steigende Last auch über einen langen Lifecycle verarbeiten zu können, ist eine skalierbare Infrastruktur erforderlich.

Der Hauptdarsteller: Oracle SOA Suite Cluster

Als Single Node Installation ist die Oracle SOA Suite nur eingeschränkt skalierbar. Zwar können moderne Server mit sehr viel physischem Hauptspeicher ausgestattet werden und auch mit einer Vielzahl hochperformanter Multicore CPUs, aber ein einzelner JVM Prozess kann diese Ressourcen nur unzureichend auslasten.

Um mehrere JVM Prozesse zum Betrieb der Oracle SOA Suite nutzen zu können, können mehrere SOA Suite Instanzen zu einem Cluster verbunden werden. Die Oracle SOA Suite nutzt dazu die Clusterfunktionalität des Applicationsservers auf dem sie betrieben wird. Aktuell unterstützt werden hierbei der IBM Websphere Applicationserver (WAS) und der Oracle WebLogic Server (WLS). Dieser Vortrag behandelt ausschließlich die Installation auf Basis des WebLogic Servers.

Der WebLogic Cluster ist im Enterprise Umfeld erprobt und bewährt und zeichnet sich durch hohe Skalierbarkeit aus. Im Folgenden findet sich eine schematische Darstellung des WebLogic Clusters.

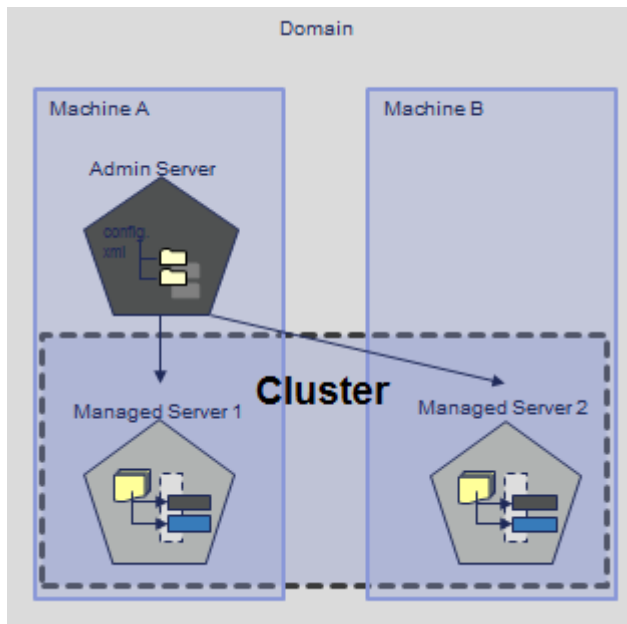


Abb. 1: Oracle WebLogic Cluster

Der Cluster besteht aus einer beliebigen Zahl von sogenannten WebLogic Managed Servern. Auf jedem dieser Managed Server ist der vollständige Satz von Java EE Anwendungen deployed welche zusammen die SOA Suite bilden. Diese Anwendungen sind bewusst stateless gehalten. Dadurch wird das Handling des Clusters wesentlich einfacher.

Dies beginnt beim Frontend, dem Loadbalancer. Hier kann jeder beliebige Verteilmechanismus verwendet werden, es muss keine Session Stickiness implementiert werden, jeder Aufruf kann an einen beliebigen SOA Suite Knoten weitergeleitet werden. Und im Backend ist zudem auch keine Session Replikation erforderlich.

Nicht geclustered werden kann der Admin Server. Auf diesem laufen allerdings keine zum Betrieb der SOA Suite notwendigen Anwendungen. Durch den sogenannten Managed Server Independent Mode können die SOA Suite Managed Server unabhängig vom AdminServer gestoppt und gestartet werden. Die SOA Suite steht damit in vollständiger Funktionalität zur Verfügung, auch wenn der AdminServer nicht läuft.

Allerdings liegen die WebLogic Administrations Konsole sowie der Oracle Enterprise Manager Fusion Middleware Control zur Konfiguration der SOA Suite auf dem Admin Server.

Wartungsarbeiten wie Konfiguration und Deployments können also erst wieder durchgeführt werden, wenn der Admin Server wieder gestartet ist. Im Sinne einer Verfügbarkeitslösung wird der Admin Server im Allgemeinen nicht als Single Point of Failure angesehen.

Weitere Mitspieler

Neben dem WebLogic Cluster werden noch weitere HA Elemente verwendet.

Oracle Coherence ist ein verteilter In-Memory Objekt Cache. Ein SOA Suite Cluster verwendet Coherence um das Clusterweite Deployment von Composite Anwendung zu koordinieren. Eine fehlerhafte Coherence Konfiguration kann verhindern, dass der SOA Suite Cluster startet. In der

Standardkonfiguration erkennen sich die Coherence Clusterknoten beim Start gegenseitig via Multicast. Steht Multicast im Netzwerk nicht zur Verfügung oder ist von der Netzwerkabteilung nicht erwünscht, so kann auch eine Gegenseitige Erkennung via Unicast konfiguriert werden, was in diesem Vortrag anhand eines Beispiels dargestellt wird.

Außerdem verwendet die SOA Suite JMS Topics. Bei der Initialen Implementierung eines SOA Suite Clusters werden die notwendigen JMS Objekte automatisch angelegt. Erweitert man einen Cluster sind hier manuelle Anpassungen erforderlich die im Vortrag detailliert beschrieben werden.

Skalierbarkeit ist nicht automatisch Hochverfügbarkeit

Dieser Vortrag beschäftigt sich explizit nicht mit dem Thema Hochverfügbarkeit. Zwar bietet es sich an, für eine ausfallsichere Umgebung Techniken zu verwenden, die auch für die Skalierbarkeit eingesetzt werden; Z.B. den hier beschriebenen SOA Suite Cluster. Allerdings kommt man damit nicht automatisch zu einer Hochverfügbarkeitslösung.

Betrachten wir einen SOA Suite Cluster mit vorgeschaltetem Hardware Loadbalancer, der für die aktuelle Last ausreichend dimensioniert ist. Ist der Loadbalancer nicht durch eine Failoverlösung abgesichert, sind zwar die Skalierbarkeitsanforderungen erfüllt, aber nicht die Anforderungen an eine Hochverfügbarkeitslösung, da das System bei Ausfall des Loadbalancers nicht mehr verfügbar ist. Zur Erreichung einer HA-Lösung sind also Maßnahmen erforderlich die über das Thema Skalierbarkeit hinausreichen.

Skalierbarkeit liegt in Ihrem (Er)messen

Welche Ansprüche an die Skalierbarkeit angelegt werden ist im Vorfeld zu planen. Hier werden meistens Ansprüche an Durchsatz und Antwortzeit für bestimmte Anwendungsfälle festgelegt. Mit Lasttests wird dann das Sizing ermittelt um die festgelegte Leistung zu erreichen. Das initiale Sizing ist aber nur der Startpunkt für ein System, dessen Last mit der Zeit steigt.

Es muss also durch kontinuierliche Messungen der Wachstumstrend ermittelt werden. Bei Erreichung zuvor festgelegter Schwellen muss dann eine Skalierung der Umgebung erfolgen.

Gemessen wird hierbei zum einen der Ressourcenverbrauch:

- CPU
- Hauptspeicher
- Festplatten Speicher
- Festplatten IO
- Netzwerk IO
- Logfilemonitoring
 - Timeout Fehler: HTTP, Database, ...
 - Out of Memory Exception

Zum anderen volumetrische Daten:

- Gleichzeitig angemeldete Benutzer (Concurrent User)
- BPEL Prozesse
- Mediator Transformationen
- JMS Messages

Außerdem die Antwortzeiten.

Up or out

Je nachdem wie sich diese Werte entwickeln kann eine Skalierung auf der vorhandenen Hardware (Scale Up) oder unter Hinzunahme neuer Hardware erfolgen (Scale Out).

Überschreiten die volumetrischen Daten oder Antwortzeiten die festgelegten Schwellwerte während die Ressourcen sowie typische Bottlenecks (Datenbank, Netzwerk, ...) noch im grünen Bereich liegen, können die vorhandenen Hardwareressourcen möglicherweise durch ein Scale up besser ausgelastet werden.

Ist hingegen auch die Ressourcenauslastung schon über den gewünschten Schwellwerten, wird weitere Hardware benötigt und ein Scale Out vorgenommen.

Der Vortrag demonstriert die notwendigen Schritte um den Oracle SOA Suite auf der bestehenden Hardware und auf zusätzliche Hardware zu erweitern.

Kontaktadresse:

Gabriel Lobstein
OPITZ CONSULTING München GmbH
Weltenburger Straße 4
81677 München

Telefon: +49 (0) 89 680098-0
Fax: +49 (0) 680098-4400
E-Mail: gabriel.lobstein@opitz-consulting.com
Internet: www.opitz-consulting.com