

Oracle Clusterware / Grid Infrastructure am Limit

Jens-Christian Pokolm
Postbank Systems AG
Bonn

Schlüsselworte

ASM, Cluster Services, CRS, Performance, Probleme, Monitoring, Einsparung, Kosten, Administration

Einleitung

Neben dem Betrieb von Oracle Real Application Cluster Datenbanken kann man die Oracle Clusterware (neu: Grid Infrastructure) 11g R2 auch zur Absicherung weiterer betrieblicher Prozesse einsetzen. Im Rahmen des Vortrags werden die Möglichkeiten hierzu aus der täglichen Praxis im Betrieb eines großen und komplexen Bankenkernsystems und vieler Umsysteme beleuchtet. Es werden eigene Erweiterungsmöglichkeiten und Automatisierungen der Überwachung dokumentiert sowie auch die noch vorhandenen Limitierungen aufgezeigt (und wie man diese ein Stück weit, trotz fehlender Dokumentation, umgehen kann).

Rahmenbedingungen

Die Postbank setzt seit einigen Jahren strategisch die Oracle Clusterware zum Betrieb aller mit Oracle betriebenen Datenbanken ein. Dies brachte und bringt zum einen eine deutliche Kosteneinsparung externer Lizenzen mit sich und vereinheitlicht darüber hinaus auch die Umgebungen über unterschiedliche Plattformen hinweg.

Bei der Überwachung und Steuerung der Datenbanken zeigen sich jedoch noch Limitierungen auf, die für einen 7*24-Stunden Betrieb teilweise problematisch sind – vor allem wenn man damit ein Kernbankensystem einer Großbank steuert.

Das Cluster Filesystem ACFS

Unter Linux stellt sich dieses neue und vielfach ersehnte Filesystem recht unproblematisch dar. Jedoch gibt es vor allem auf der Plattform AIX ein leider vorhandenes Stabilitätsproblem welches zum Totalausfall des Clusters führen kann. Diese werden aktuell bei Oracle unter den Bug-Nummern 9892060 und 12979721 geführt und sollten zeitnah mittels PSU oder One-Off Patches gefixed werden.

Überwachung von Prozessen

Eigentlich geht der geneigte Administrator davon aus, über sämtliche Events im Cluster umfänglich informiert zu werden. Dies wird man laut Oracle Dokumentation mit FAN-Callouts implementieren. Dazu legt man unter das Verzeichnis `$CRS_HOME/racg/usrco/` das eigene Script z.B. wie nachfolgen dargestellt ab:

```
#!/usr/bin/sh
HOST=`hostname`
FAN_LOGFILE=/var/log/oracle/fan_log_${HOST}.log
echo $* "reported="`date`>>$FAN_LOGFILE
```

Bei jedem auftretenden Schaltvorgang der Clusterware wird dieses Programm/Script aufgerufen und als Parametersatz die entsprechenden Objekte mitgegeben. Das hier dargestellt Script protokolliert diese nur in das hinterlegte Logfile.

Einziges Manko hierbei ist: Wenn z.B. der lokale Listener oder auch der Scan-Listener terminiert, wird kein FAN-Event erzeugt !

In der Dokumentation findet man auch keinerlei Hinweise wie man einen solchen Event überwachen könnte um z.B. eine Störungskonsole wie MasterCell oder Tivoli Enterprise Console bzw. HP Service Desk damit zu versorgen.

Eine Lösung für diese Problem bietet die leider undokumentierte Funktion der Event Monitor Daemon (EVM) Konfiguration. Speziell diese Art der Benachrichtigungsmöglichkeiten werden detailliert im Folienvortrag erläutert sowie auch Verweise auf entsprechende Informationsquellen dargelegt.

Bitte achten Sie hierbei auch auf die Rechte der Dateien im Verzeichnis

`§CRS_HOME/evm/admin/conf/` - diese sind je nach Patchstand leider nicht korrekt. Die Files müssen für den „Grid-User“ zumindest lesbar sein.

Abhängigkeiten von Prozessen untereinander

Beim normalen Start von Datenbanken, Listnern und Services auf einem RAC-Cluster tritt der Effekt niemals auf – bei der Installation eines großen SAP-Systems sieht das ganze bereits anders aus.

Hier gibt es die Notwendigkeit Prozesse wie z.B. den Enqueue-Server und den Replicated-Enqueue garantiert zu jeder Zeit auf unterschiedlichen Servern im Cluster laufen zu lassen.

Zwar gibt es den Parameter

```
START_DEPENDENCIES=dispersion:active(ENQUEUE_RES)
```

Dies bedeutet jedoch im Ernstfall, wenn nicht genügend Server verfügbar sind um diese Anwendungen zwangsweise zu trennen, das die Clusterware dennoch die Applikationen auf dem gleichen Node startet.

In einem 4 Knoten-Cluster in 2 Rechenzentren hat man damit auch keine Kontrolle, das der ENQUEUE im Rechenzentrum A läuft und der REP_ENQUEUE in Rechenzentrum B. Dies müssen in jedem Fall Wrapper-Scripte um die Ressourcen herum lösen – ansonsten ist im Fall eines Ausfalls ein Problem vorprogrammiert.

An dieser Stelle auch der Hinweis an alle Benutzer des „SAP CTL“ – diesen unbedingt auf einen aktuellen Stand bringen zu lassen um bestehende Probleme zu beheben und vor allem sehr ausgiebig zu testen.

Dynamik von Usern und Gruppen

Wenn eine Ressource (z.B.) der ENQUEUE-Server dem User „oraSID“ gehört und der Unix-Gruppe sapadmin dann kann man dies in der Berechtigung der CRS-Ressource sehr einfach hinterlegen.

Das was im Unix Umfeld unproblematisch ist – einem zuvor bestehenden User zur Laufzeit des CRS eine weitere Gruppe hinzufügen (`useradd -G gruppenname Benutzer`) und sofort mit dem „Benutzer“ nutzen zu können – geht bei Oracle Cluster Infrastructure leider nicht.

Denn als Gruppe eines Users akzeptiert die Clusterware nur die Gruppen die zum Zeitpunkt des Starts der Clusterware bereits vorhanden und zugewiesen waren.

Ein Stop und Start der Clusterware lädt dann den zu diesem Zeitpunkt gültigen Zustand neu. Ein völlig unerwartetes Verhalten – bedeutet jedoch eine genaue Planung der Rechte sowie eine Knotenweise Downtime.

Interne Events an Managementkonsole

Beliebige Events der Clusterware kann man sehr einfach und Simple überwachen indem eine Konfigurationsdatei unter `$CRS_HOME/evm/admin/logger/` angelegt wird.

Die im nachfolgenden Beispiel genannte Datei `Postbank.conf` sollte dann wie folgt aussehen:

```
forward {
  name Postbank
  filter "[prio >= 200]"
  command "/grid/control/ctrl-Alerts.sh"
  maxqueue 100
}
```

Damit werden alle Events ab der Prioritätsklasse 200 als Parameter an das Script `/grid/control/ctrl-Alerts.sh` übergeben und können dort geordnet an die Störungskonsole gesendet werden. Dies funktioniert absolut tadellos – ist nur leider in keiner einzigen Dokumentationszeile der Oracle Clusterware erwähnt.

Remote Steuerung

Auch die Applikationsserver im SAP-Kontext werden mittels Oracle Clusterware gesteuert. Jedoch ist auf den Systemen keine Clusterware installiert. Lediglich die zentrale Datenbank ist damit ausgestattet. Auf dem RAC-Verbund laufen lokal die Enqueues und Replicated Enqueues. Mittels „ssh remote command execution“ werden die SAP-Applikationsserver vom RAC mit verwaltet. Dies hat insofern einen schönen Nebeneffekt als das der Verkabelungsaufwand für die Cluster-Interconnects entfällt und natürlich auch der Installations/Upgrade-Vorgang. Details zu diesem Vorgehen und der Überwachung werden während der Doag-Konferenz anschaulich erläutert.

Dennoch

Neben allen „Schwachstellen“ - die Oracle Clusterware ist sowohl vom Funktionsumfang als auch der Stabilität her selbst für den hochkritischen Einsatz im Kern einer Bank vollauf geeignet.

Probieren Sie es selber aus – Sie werden nach einer Einarbeitung sicher auch positiv überrascht sein.

Kontaktadresse:

Jens-Christian Pokolm
Postbank Systems AG
Baunscheidtstrasse 8
D-53113 Bonn

Telefon: +49 (0) 228-920 63155
Fax: +49 (0) 228-920 63100
E-Mail jens-christian.pokolm@postbank.de
Internet: www.postbank.de