RAC-Installation und Bedienung unter Windows – Tipps und Tricks

Susanne Jahr
Herrmann & Lenz Services GmbH
Burscheid

Schlüsselworte
RAC, Grid Infrastructure, Clusterware, Windows

Einleitung

Die Installation einer Oracle Clusterware bzw. einer Grid Infrastructure ist ganz einfach… WENN man alle Betriebssystem-, Netzwerk-, Hardware- und benutzerseitigen Voraussetzungen erfüllt. Diese Aussage ist unabhängig vom Betriebssystem gültig – egal, ob man eine Oracle-Clusterware 10gR2, 11gR1 oder eine Grid Infrastructure 11gR2 installiert. Die Installation unter MS Windows stellt den DBA hier nicht unbedingt vor größere, aber teilweise ganz andere Herausforderungen als es unter einem UNIX- oder Linux-System der Fall wäre. Die Anforderungen sind zahlreich, und man ist gut beraten, keine zu vernachlässigen, denn: dem Installer entgeht nichts! Doch auch wenn die Installation geschafft ist, gilt es noch vieles zu beachten.

Installation Clusterware / Grid Infrastructure - Voraussetzungen

1. Plattenplatz

1a). Plattenplatz – für Installationen

Es ist zu bedenken, dass ab der Version 11.2 für die Grid Infrastructure ein Out of Place-Upgrade zwingend vorgeschrieben ist. Auch für die Datenbank-Software ist es "strongly recommended". Es sollte also bereits vor Beginn der Installation darauf geachtet werden, vorausschauend auf die nächsten Patchsets bereits hinreichend Plattenplatz zur Verfügung zu haben – zumindest, bis die Patch-Installation abgeschlossen ist. Danach kann die alte Software-Version auch deinstalliert werden.

Für alle Oracle-Versionen gilt: Auch Interim-Patches wie die Bundle-Patches (Windows-Version der PSU) oder andere empfohlene Patche (OPatch) können sehr klein, aber auch leicht diverse 100MB groß sein. Dieser Platz wird dann sowohl im GI-Home als auch in dem oder den DB-Home(s) benötigt. Die neuesten OPatch-Versionen für 10.2 und 11.2 enthalten den Befehl "util", mit dem u.a. das mitunter mehrere GB große Verzeichnis %ORACLE_HOME%\.patch_storage aufräumen kann.

1.b) Plattenplatz – für den Shared Storage

Entgegen der Intuition, die Partitionen für AMS-Disks im Shared Storage als primäre Partitionen zu definieren – schließlich soll es pro LUN immer nur eine sein – müssen dies erweiterte Partitionen mit logischen Laufwerken sein. Diese können in Windows 2008 nicht mehr mit dem Festplattenmanager angelegt werden. Hier kommt das Kommandozeilen-Tool `diskpart` zum Einsatz:

```
DISKPART> select disk 1
Disk 1 is now the selected disk.
DISKPART> create partition extended
DiskPart succeeded in creating the specified partition.
DISKPART> create partition logical
DiskPart succeeded in creating the specified partition.
DISKPART> list disk

Disk ###  Status         Size     Free     Dyn  Gpt
--------  -------------  -------  -------  ---  ---
 Disk 0    Online           68 GB      0 B
*Disk 1    Online         1479 MB      0 B
```

Im Anschluss können die ASM-Label mit dem Tool asmtoolg (grafisch) oder asmtool (Kommandozeile) angebracht werden.

2. Benutzer

2. a) Berechtigungen

Die Anforderungen an den Installationsbenutzer sind relativ simpel: Er muss lokaler Administrator sein, und zwar direkt durch Gruppenzugehörigkeit in der Administratoren-Gruppe auf jedem Knoten (nicht indirekt über andere Gruppen). Der Installationsbenutzer benötigt das Recht, auf entfernten Rechnern die Registry zu bearbeiten sowie das Benutzerrecht "Manage Auditing and Security Log". Wie in der UNIX-/Linux-Welt ist es auch unter Windows vorgeschrieben, das ORACLE_HOME der Grid Infrastructure nicht innerhalb des ORACLE_BASE-Pfades anzulegen.

2.b) Benutzer-Äquivalenz

Als Installationsbenutzer kann ein Domänen- oder ein lokaler Benutzer verwendet werden. Wird ein lokaler Benutzer verwendet, so muss dieser auf allen Knoten identisch heißen und auch das identische Passwort haben. Diese Maßnahmen sind zur Erfüllung der Anforderung "Benutzer-Äquivalenz" erforderlich. Der Installationsbenutzer muss auf jedem Knoten Mitglied in der Gruppe ORA_DBA sein, die während der Installation automatisch angelegt wird

3. Groß- und Kleinschreibung: egal bei Windows...?

Man hat sich bei der Administration und Benutzung von Windows-Systemen daran gewöhnt, dass der Groß- und Kleinschreibung weitgehend keine Beachtung geschenkt werden muss. Aber Vorsicht: in der RAC-Welt gilt dies nicht uneingeschränkt!

Dies gilt auch schon für Versionen < 11gR2; konkret aufgetreten ist Fall a) bei einer Installation 11gR1.

3.a) Hostnamen

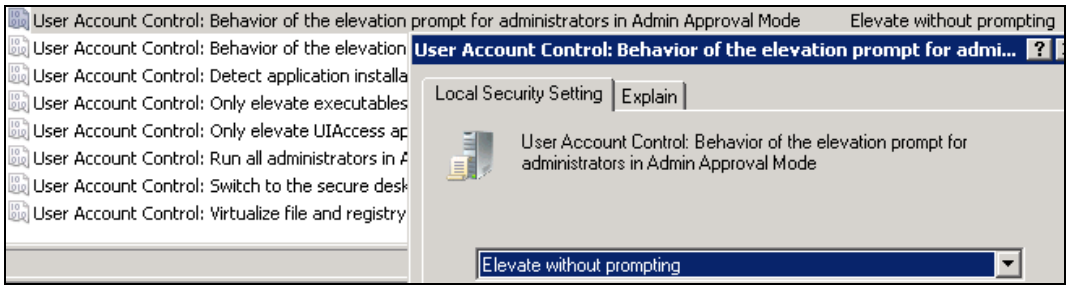
Die Schreibweise der Hostnamen für die Cluster-Knoten muss für alle beteiligten Knoten identisch sein – also nicht `NODE1` für den ersten und `node2` für den zweiten Knoten! Dies ist sowohl für die Namensgebung auf den Servern selbst im Betriebssystem als auch für die DNS-Registrierung zu beachten. Werden die Hostnamen auf Betriebssystem-Ebene unterschiedlich geschrieben, kann es passieren, dass zwar die Installation funktioniert, anschließend jedoch Teile der Clusterware (ONS) nicht starten. Ist dies der Fall, so muss der ganze Knoten aus dem Cluster auskonfiguriert werden, dann die Namensgebung auf Betriebssystem-Ebene geändert und dann der Knoten wieder hinzugefügt werden.

3.b) Schnittstellennamen für das öffentliche und das private Netzwerk

Es entspricht der Best Practice, die recht umständlichen Standard-Namen für Netzwerkschnittstellen (LAN Connection 1 etc.) in kürzere und benutzerfreundlichere Namen zu ändern, damit die Schnittstellen auf den ersten Blick dem einen oder anderen Netzwerk zugeordnet werden können. Aufgrund eines Bugs sollten hier jedoch nicht die Namen PUBLIC und PRIVATE (alles in Großbuchstaben) verwendet werden! Die Namen public und private oder irgendeine andere Schreibweise hingegen sind ok. Auch hier gilt: Namen müssen auf allen Knoten identisch sein! Es ist z.B. nicht erlaubt, den Hostnamen, die Knotennummer oder ähnliches in den Schnittstellennamen mit einzubauen, da sich dadurch unterschiedliche Namen auf den Knoten ergeben würden.

4. Zu beachten bei Windows Server 2008

4.a) UACL: In der lokalen Sicherheitsrichtlinie sollte für Administratoren eingestellt werden, dass Programme ohne Nachfrage als Administrator ausgeführt werden können:



4.b) "Run as Administrator": Einige Tools müssen mit Administrator-Privilegien ausgeführt werden, z.B. netca, asmtool etc.. Werden diese über die Icons im Startmenü ausgeführt, so ist hier das Privileg "Run as Administrator" bereits implementiert. Werden sie hingegen von der Kommandozeile oder vom Windows-Explorer aus aufgerufen, so sollten sie extra mit der Option "Run as Administrator" ausgeführt werden. Es hat sich als hilfreich erwiesen, eine Administrator-Kommandozeilen-Verknüpfung an die Taskleiste zu heften und die erforderlichen Tools aus dieser heraus auszuführen.

4.c) ***(Windows allgemein)*** IPv6 ist zurzeit nicht für RAC unterstützt (auch nicht in 11.2.0.2). Es sollte daher nicht konfiguriert werden. Wenn doch, sollte die Priorität von IPv4 in den erweiterten Netzwerkeinstellungen über diejenige von IPv6 gesetzt werden.

5. Registry-Parameter

a) Der net use-Befehl auf die administrative Freigaben (c$ ...) muss zwischen den Knoten ohne Angabe von Passworten möglich sein. Ist dies nicht der Fall, so fehlt wahrscheinlich der Registry-**`Parameter LocalAccountTokenFilterPolicy`** (REG_DWORD), oder er hat nicht den erforderlichen Wert 1. Zu finden ist er in der Registry unter **`HKLM\SOFTWARE\Microsoft\Windows\CurrentVersion\Policies\System`**.

b) DHCP Media Sensing erlaubt es dem Betriebssystem, eine IP-Adresse vom Netzwerkadapter zu entkoppeln, wenn die Verbindung zum Switch abbricht. Dies ist für eine RAC-Konfiguration jedoch nicht erwünscht und muss daher deaktiviert werden. Hierzu dient der Parameter **`DisableDHCPMediaSense`** (REG_DWORD, Wert=1) unter **`HKLM\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters`**.

c) Zeitserver: Die Nutzung eines Zeitservers ist erforderlich, um Zeitsynchronisation zwischen den Cluster-Knoten zu gewährleisten. Hierzu dient der Windows-Zeitserver-Dienst (>= 11.2 wahlweise auch der Cluster-eigene Zeitserver-Dienst ctssd). Wird der Windows-Zeitserver verwendet, so muss gewährleistet sein, dass die Zeit nicht in die Vergangenheit angepasst wird, sondern immer nur in die Zukunft. Dies stellt der Parameter MaxNegPhaseCorrection sicher, der den Wert 0 erhält. Er befindet sich unter **`HKLM\SYSTEM\CurrentControlSet\Services\W32Time\Config`**. Anschließend muss die Änderung noch mit dem Befehl **`W32tm /config /update`** an der Kommandozeile aktiviert werden.

6. Installation 10gR2 auf Windows Server 2008R2

Für Microsoft-64bit-Betriebssysteme ab Server 2008 sind nur noch signierte Treiber zugelassen. Die Signaturen der OCFS- und FENCE-Treiber für Oracle 10gR2 sind jedoch Ende 2009 abgelaufen, so dass hierfür nicht nur ein Patch-Download, sondern auch der längste Installer-Aufruf überhaupt erforderlich ist – einzugeben im DOS-Fenster in einer Zeile:

```
setup.exe New_Driver_Loc="C:\setup\OraFence\7320726" oracle.has.cfs:s_newOcfspath="C:\setup\OraFence\7320726"
oracle.has.crs:b_isWIN2k8="TRUE" oracle.has.crs:s_newOcfspath="C:\setup\OraFence\7320726" -ignoreSysprereqs
```

Cluster-Betrieb

1. Windows-Dienste

Auffällig ist nach Fertigstellung der Grid Infrastructure-Installation zunächst das Vorhandensein anderer Dienste als aus früheren Releases gewohnt. Der Clusterware-Stack begnügt sich nunmehr mit einem einzigen Dienst (OracleOHService) anstelle der einzelnen Dienste für die verschiedenen Prozesse crsd, cssd und evmd. Dafür gibt es mehrerer Listener-Dienste. Jeder SCAN-Listener besitzt einen eigenen Dienst auf jedem Cluster-Knoten, wobei die Clusterware dafür sorgt, dass immer diejenigen Dienste gestartet sind, deren SCAN-Listener momentan auf dem lokalen Knoten laufen – in der Abbildung ist dies der LISTENER_SCAN1 (es gibt hier lediglich zwei SCAN-Listener). Ein nicht gestarteter Listener ist also an dieser Stelle kein Problem, sondern Teil des Designs.

Oracle Object Service	Started	Automatic
Oracle VMRACDB1 VSS Writer Service		Manual
OracleASMService+ASM1	Started	Manual
OracleDBConsoleVMRACDB	Started	Automatic
OracleJobSchedulerVMRACDB1		Disabled
OracleMTSRecoveryService	Started	Automatic
OracleOHService	Started	Automatic
OracleOraCrs11g_home1TNSListener	Started	Manual
OracleOraCrs11g_home1TNSListenerLISTENER_SCAN1	Started	Manual
OracleOraCrs11g_home1TNSListenerLISTENER_SCAN2		Manual

Abb. 1: Windows-Services Grid Infrastructure

2. Parameterdateien

In Oracle RAC 11gR2 gehört der Listener zur Grid Infrastructure, nicht zur Datenbank-Software. Die Parameterdateien listener.ora und sqlnet.ora befinden sich daher auch im ORACLE_HOME der Grid Infrastructure. Zu beachten ist, dass in der sqlnet.ora sowohl TNSNAMES als auch EZCONNECT im NAMES.DIRECTORY_PATH definiert sein sollten, da es sonst Probleme bei der Registrierung der Datenbank-Instanzen bei den SCAN-Listenern geben kann. Der Parameter SQLNET.AUTHENTICATION_SERVICES muss den Wert NTS haben, da ansonsten keine Cluster-Ressourcen starten können. Diese verwenden die Anmeldung / as sysdba, was den Wert NTS für den genannten Parameter erfordert.

2.a) tnsnames.ora-Einträge

Für Oracle >=11.2.0.1 gilt: die Clients sollen sich zur Verbindungsaufnahme an den SCAN-Listener wenden. Die hier für einen einzigen Namen hinterlegten drei IP-Adressen können jedoch von älteren Clients nicht verarbeitet werden. Diese sollten daher nach wie vor eine Host-Adressliste in der tnsnames.ora verwenden, wobei jedoch nicht die VIPs der einzelnen Knoten, sondern die SCAN-IPs eingetragen werden sollten.

3. Probleme beim Failover durch geändertes Gratuitous ARP-Verhalten unter Windows 2008

Es wurde bei Kunden beobachtet, dass nach einem Failover Clients erst nach mehreren Minuten die VIP des ausgefallenen Servers auf dem überlebenden Knoten erreichen können, wenn die Clients sich in einem anderen Netzwerksegment befinden als die RAC-Knoten. Grund hierfür ist jedoch nicht der Failover-Mechanismus des RAC (die VIP erscheint prompt auf dem überlebenden Knoten und ist innerhalb des RAC-Segments auch verfügbar), sondern die Tatsache, dass im ARP-Cache nicht sofort die Information über den Schwenk der VIP an den Switch (Cisco; wahrscheinlich aber auch andere Fabrikate) veröffentlicht bzw. von diesem in andere Segmente weitergegeben wird. Erst nach Ablauf

eines definierten Intervalls auf dem Switch (in unserem Fall 8 Minuten) wird der ARP-Cache aktualisiert und die VIP den dahinter liegenden Clients verfügbar gemacht. Dieses Verhalten ist laut Microsoft Support so beabsichtigt; der MS-eigene Failover-Cluster umgeht dieses Problem "durch einen internen Call, die API ist jedoch nicht veröffentlicht"... Da sowohl eine Reduzierung des Intervalls auf dem Switch als auch eine Verschiebung des RAC in das Netzwerksegment der Clients keine Option war, schaffte hier folgender Workaround Abhilfe:

4. Erstellung von benutzerdefinierten Cluster-Ressourcen

Es wurden eigene Cluster-Ressourcen erstellt. Deren Action-Skripte gewährleisten, dass bei ONLINE-Stellung einer VIP (SCAN- oder Knoten-VIP) nach Reboot oder Schwenk auf einem beliebigen Knoten hier der lokale ARP-Cache geleert und ein Ping von der geschwenkten VIP aus auf das Default Gateway abgesetzt wird. Der lokale ARP-Cache ist somit aktualisiert und die VIP für Clients aus anderen Netzwerksegmenten erreichbar. Bei manuellen Schwenks dauerte die Aktualisierung 2 Pings; nach einem Reboot ca. 10 Pings. Die Ressourcen wurden wie im folgenden Beispiel angelegt:

```
crsctl add resource ora.scan1_vip.app -type cluster_resource -attr
"ACTION_SCRIPT='C:\script\refresh_scan1_vip.cmd', PLACEMENT='restricted',
SERVER_POOLS='Generic', CHECK_INTERVAL='60',
START_DEPENDENCIES='hard(ora.scan1.vip) pullup(ora.scan1.vip)',
STOP_DEPENDENCIES='hard(ora.scan1.vip)', RESTART_ATTEMPTS='2', ACL='owner:nt
authority\system:rwx,pgrp::r-x,other::---'"
```

Das erforderliche Action-Skript wird leider in der Dokumentation und auch in My Oracle Support nirgendwo als Beispiel für Windows angeboten. Man findet jedoch nach einigem Suchen ein Beispiel innerhalb der Grid-Infrastructure-Installation im Ordner GI_HOME\crs\demo. Die Datei action_scr.bat ist recht umständlich, kann aber als Basis für den eigenen Bedarf genutzt werden.

Fazit

Oracle und insbesondere RAC auf der MS Windows-Plattform hat traditionell mit vielen Vorurteilen (vor allem aus dem UNIX-Lager) zu kämpfen. Diese können nach unseren Erfahrungen so nicht bestätigt werden. Die Eigenheiten unter Windows sind andere, jedoch nicht unbedingt schwieriger oder umfangreicher als anderswo – allerdings zeigt die Erfahrung, dass Oracle selbst den Windows-Kunden an vielen Stellen etwas "stiefmütterlich" behandelt. Insbesondere in der Dokumentation und in MOS-Artikeln fällt immer wieder auf, dass Beispiele eben nicht für das Microsoft-Betriebssystem geeignet sind. Auch schleicht sich manchmal der Verdacht ein, dass der auf UNIX/Linux-Plattformen entwickelte Code anschließend etwas lieblos auf Windows portiert wurde. Dennoch sind unsere Erfahrungen bei Kunden mit RAC auf Windows (und zwar sowohl 10gR2 als auch 11gR1 / 11gR2, Windows Server 2003 oder 2008) überwiegend positiv. Einmal korrekt installiert, läuft das System ebenso stabil wie auf anderen Plattformen. Die Empfehlung für neue Systeme ist also wie immer, bei dem Betriebssystem zu bleiben, über das das meiste Wissen im Hause vorhanden ist.

Kontaktadresse:

Susanne Jahr
Herrmann & Lenz Services GmbH
Höhestr. 79
D-51399 Burscheid

Telefon: +49 (0) 2174-6712-0
Fax: +49 (0) 2174-6712-22
E-Mail susanne.jahr@hl-services.de
Internet: www.hl-services.de