

Capítulo 1: Marco teórico

Área de Soporte Técnico

Dentro de Oracle, como en cualquier compañía de software existe el área de Soporte Técnico, cuyo objetivo principal es el de brindar asistencia y proveer soluciones en los casos que exista algún resultado no esperado, dentro del producto, en su instalación o durante su operación normal.

Es importante aclarar que Soporte Técnico brinda apoyo en situaciones particulares y específicas, que pueden ir desde una pregunta simple, o un solo error, hasta problemas de mayor complejidad como puede ser un mal rendimiento de la Base de Datos, o que simplemente nadie en la organización puede acceder a la Base de Datos, a pesar de que esta operando en un cluster. Esto hace que el Ingeniero de Soporte Técnico tenga una participación limitada en tiempo durante la interacción con los clientes, debido a que una vez solucionado el problema, el reporte del cliente es cerrado.

En la mayoría de los casos, en los que el cliente ha reportado un problema, su reporte puede permanecer abierto por uno o varios días, siendo este tiempo

normalmente menos de una semana. Un reporte siempre debe de contener un solo problema, esto ayudará a evitar confusiones o malos entendidos con los clientes. Además que permitirá enfocarse en el problema que fué reportado originalmente.

Existen otras áreas de servicio como es Consultoría, las cuales pueden ser contratados para implantaciones de proyectos, que requieren un tiempo mayor. Sin embargo, dada la especialización que tienen los Ingenieros de Soporte Técnico es común, que otras áreas hagan uso de sus servicios, incluyendo el área de Consultoría.

Si bien es cierto, que existe en Soporte Técnico una base de datos que incluye las soluciones a problemas frecuentes, además de otra información técnica del producto, esto no significa que el trabajo del Ingeniero de Soporte Técnico se reduzca a hacer búsquedas de soluciones en esa base de datos, cada vez que se enfrenta a un problema reportado por algún cliente. De hecho ésta u otras prácticas que no sigan una metodología ordenada, serán poco fructíferas y costosas para el negocio, tanto de Oracle como de las organizaciones que han reportado el problema.

Sin lugar a dudas, el Ingeniero de Soporte Técnico debe de aplicar la Ingeniería identificando el problema, sus causas y de ser posible una solución, o alguna alternativa que permita al negocio seguir operando de manera aceptable. En cada una de éstas etapas, deben de usarse todas las herramientas de diagnóstico que ofrecen los productos incluyendo la documentación interna y externa. El esfuerzo tiene que ir más allá, en lo posible siempre basándose en los hechos que se tienen cuando se sufrió el problema y de ser necesario, crear nuevos métodos para diagnosticar alguna situación, cuya oportunidad se presenta, cuando es la primera vez que es reportada. En este contexto es claro que la aplicación de la creatividad y el ingenio, no deben limitarse porque son factores indispensables en la búsqueda de soluciones.

Sobre la comunicación con los clientes, en la actualidad con el desarrollo de

Internet, la gran mayoría de ellos reportan sus problemas usando este medio, pero aún todavía existe la posibilidad de usar el teléfono, como medio de comunicación con los Ingenieros de Soporte Técnico. Cualquiera que sea la forma establecida por el cliente, es de gran importancia ser claros en éste aspecto, durante todo el proceso, explicando las razones de cada uno de los pasos que se están tomando o recomendando seguir. Aún cuando no se tenga la información suficiente, para analizar la situación, cualquier plan de acción propuesto debe de tener un fundamento técnico.

Cuando se trabaja con los clientes, existe un sistema interno, donde se registra cada una de las interacciones con ellos. Desde el momento que el cliente reportó el problema, durante toda la investigación y planes de acción propuestos, hasta el punto que se encuentra una solución. Adicionalmente, al finalizar el reporte es posible crear un documento, en la base de datos de conocimiento, explicando el problema resuelto y su solución, el cual puede hacerse disponible a los clientes.

Grupo de Diagnóstico de Defectos

En un grupo separado, pero perteneciente a Soporte Técnico, se encuentra el grupo al que pertenezco. Cuyo objetivo es el de diagnosticar casos que se presumen son defectos del producto o *bugs*. El grupo se conoce como grupo de Diagnóstico de Defectos y Escalaciones.

Las personas de este grupo dedican su tiempo a analizar, diagnosticar y confirmar si la situación que ha sido reportada por el cliente, es un bug o no, lo cual requiere de un conocimiento más profundo del producto y análisis más detallado. Es común que en aquellas situaciones más complejas, donde se afecta la Base de Datos en varias máquinas del cluster, o donde estén involucrados varios productos, sean transferidas al grupo de Diagnóstico de Defectos para su análisis.

Es importante mencionar que en la mayoría de los reportes de *bugs*, realmente no son defectos reales de los productos y entonces es posible dar una solución, o alternativa para que el negocio del cliente pueda seguir operando de manera normal. En otros casos es posible encontrar reportes, que ya han sido abiertos anteriormente, donde el área de Desarrollo ha proveído una solución, en lo que conocemos como *parches*, que son las modificaciones al código para prevenir el problema.

Cuando se llega a confirmar que tenemos un defecto en el producto, es nuestro trabajo transferir el caso a el área de Desarrollo, quien es el grupo que se encarga de continuar con el análisis del problema, y de ser necesario, hacer las correcciones a nivel del código del producto, para entonces entregar un parche al cliente. Normalmente a la organización de Desarrollo se le entrega una síntesis de todo el trabajo que se ha desarrollado para identificar el problema, la causa y solución. Además se les indica si el problema se ha podido reproducir, por el cliente o internamente y los detalles de la configuración del ambiente donde se ha producido el problema.

Por su operación, el grupo de Diagnóstico de Defectos es visto como la interface entre las organizaciones de Soporte Técnico y Desarrollo. Y por las funciones que realiza tiende por un lado, a hacer uso del código del producto, para aclarar o tener un mejor entendimiento de la operación del mismo en alguna situación en particular. Y por otro lado tiene una visión más genérica de la situación misma a la que se esta enfrentando, lo cual es útil para el área de Desarrollo.

En mi grupo en particular nos encargamos de revisar, los reportes de defectos relacionados con los productos de RAC, ASM y Oracle Clusterware. Estos productos han sido creados, pensando en la operación de la Base de Datos en un cluster, por lo que están íntimamente relacionados. En la figura 1-1, se muestra la interacción de dichos productos, que son instalados en cada una de las máquinas que forman parte del cluster. Debe hacerse notar que el sistema operativo no se muestra, sin embargo este debe de existir antes de instalar cualquier producto de Oracle. RAC es una opción del manejador de la Base de

Datos para que pueda operar en el cluster. La Base de Datos a su vez usa ASM para el manejo de los discos que proveen el medio de almacenamiento. Ambos usan servicios del Clusterware para su operación y sobre todo para su mantenimiento interno. Es responsabilidad del Clusterware de mantener la integridad del cluster. Además de estos componentes, en la actualidad ASM provee un sistema de archivos conocido como *ASM Cluster File System (ACFS)*, usado para alojar archivos que no se soportan en ASM directamente. Es importante mencionar, que éstos productos trabajan de manera distribuida, en cada uno de los nodos del cluster, por lo que muchas veces es necesario revisar su operación, en cada una de las máquinas. Sobre las aplicaciones que se muestran en la figura 1-1 normalmente ellas se instalan en máquinas distintas a las del cluster, conocidas como servidores de aplicaciones.

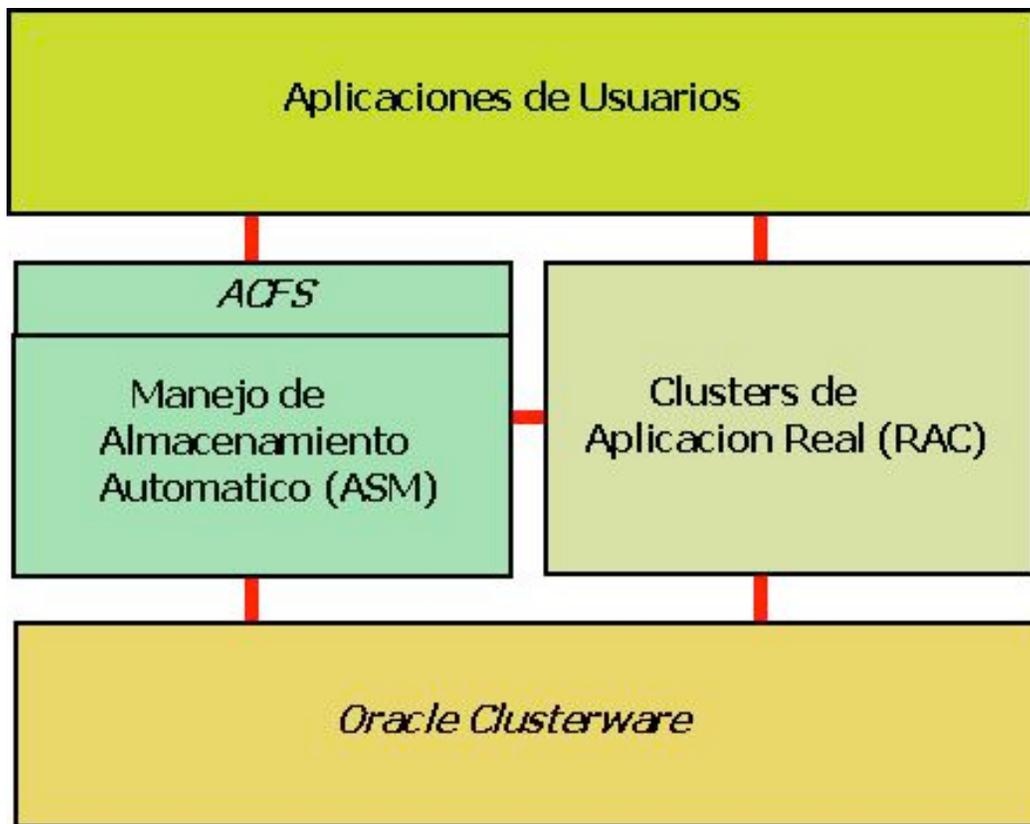


Fig. 1-1. Productos de Oracle para un cluster

Quiero aclarar que el término RAC se usa para describir una Base de Datos Oracle operando en un cluster, aunque como mencioné anteriormente es una opción adicional al manejador de Base de Datos.

Clusterware de Oracle

El primer componente que debe de instalarse, después del sistema operativo, es el Clusterware cuya misión principal es la de mantener al cluster como una sola entidad, y proveer servicios a RAC y ASM para su operación. De manera breve presento algunas de sus funciones más importantes:

- Existen procesos que se encargan de monitorear a los nodos pertenecientes al cluster, los cuales deben de seguir protocolos de comunicación, desde el momento que se unen al cluster hasta el momento que lo abandonan, considerando si fueron apagados de manera anormal o no. Si alguno de los nodos, no esta funcionando correctamente debe de eliminarse lo antes posible para evitar corrupciones en la Base de Datos, o algún otro problema que pueda afectar al cluster en su totalidad. Para poder detectar algún nodo que no esta operando de manera adecuada, el Clusterware usa la red de interconexión del cluster, así como un archivo que todos los nodos tienen acceso conocido como Archivo de Votación. Parte de los servicios que estos procesos proveen a los otros productos de Oracle, es la lista de nodos que se encuentran en el cluster. Esta información es usada por RAC y ASM, para su configuración y operaciones de recuperación.
- Todos los procesos de Oracle que operan dentro del cluster, inclusive aquellos que no pertenecen al Clusterware son representados con lo que se conoce como recursos. Los recursos son entidades que tienen propiedades, que determinan como son manejados, y describen las dependencias entre unos y otros. Es responsabilidad del Clusterware manejar dichos recursos, esto significa que son iniciados, monitoreados, detenidos y de ser necesario también recuperados por el Clusterware. Un ejemplo son las instancias mismas

de las Bases de Datos, o procesos internos como el *Cluster Time Synchronization Services Daemon (CTSSD)* que se encarga de sincronizar el tiempo de las maquinas pertenecientes al cluster. El manejo de los recursos, se hace a través de procesos llamados agentes. Cabe mencionar que es posible registrar programas, que no son de Oracle, para que sean administrados a través del Clusterware.

- Para la propagación de eventos, internos al Clusterware o hacia afuera del cluster, existen procesos que reciben la subscripción y envío de los mismos. Sobre eventos internos, un ejemplo claro puede ser cuando uno de los nodos es terminado de manera anormal, en este caso el evento se propaga entre los procesos que se encargan de monitorear el cluster, para ejecutar lo que se conoce como una reconfiguración. Sobre la propagación de eventos que se hacen hacia el exterior del cluster, un ejemplo es una funcionalidad conocida como Notificación de Aplicaciones Rápida, que consiste en que la aplicación recibe eventos del cluster, cuando algún componente necesario para acceder la Base de Datos de manera remota no esta disponible. Gracias a ésta funcionalidad, cuando un nodo es reiniciado de manera no esperada, la aplicación al ser notificada puede direccionar de manera inteligente, los nuevos requerimientos hacia el resto de los nodos que no tienen ningún problema.
- En la última versión de Clusterware se han incluido otros componentes, como el proceso *Grid Plug and Play Daemon (gpnpd)*, que permiten añadir y borrar nodos del cluster de manera más simple, simplificando el procedimiento que se necesita ejecutar. Pero quizá una de las funcionalidades más importantes, es la de permitir guardar los archivos necesarios para la operación del Clusterware dentro de ASM. Dichos archivos son el *Oracle Cluster Registry (OCR)* y el Archivo de Votación. Hasta antes de la versión 11.2 no era posible tener estos archivos en ASM. El archivo de OCR es muy importante, ya que se guarda en él la información de configuración de los recursos administrados por el Clusterware.

En la figura 1-2, se incluyen todos los recursos que son manejados por el

Clusterware. Todos ellos representan realmente procesos a nivel sistema operativo, algunos de ellos forman parte del mismo Clusterware y otros representan entidades propias de ASM y de las instancias de Base de Datos. Debo aclarar que el diagrama lista el nombre de los recursos pero puede corresponder a un nombre distinto, en el momento que esta siendo ejecutado a nivel sistema operativo.

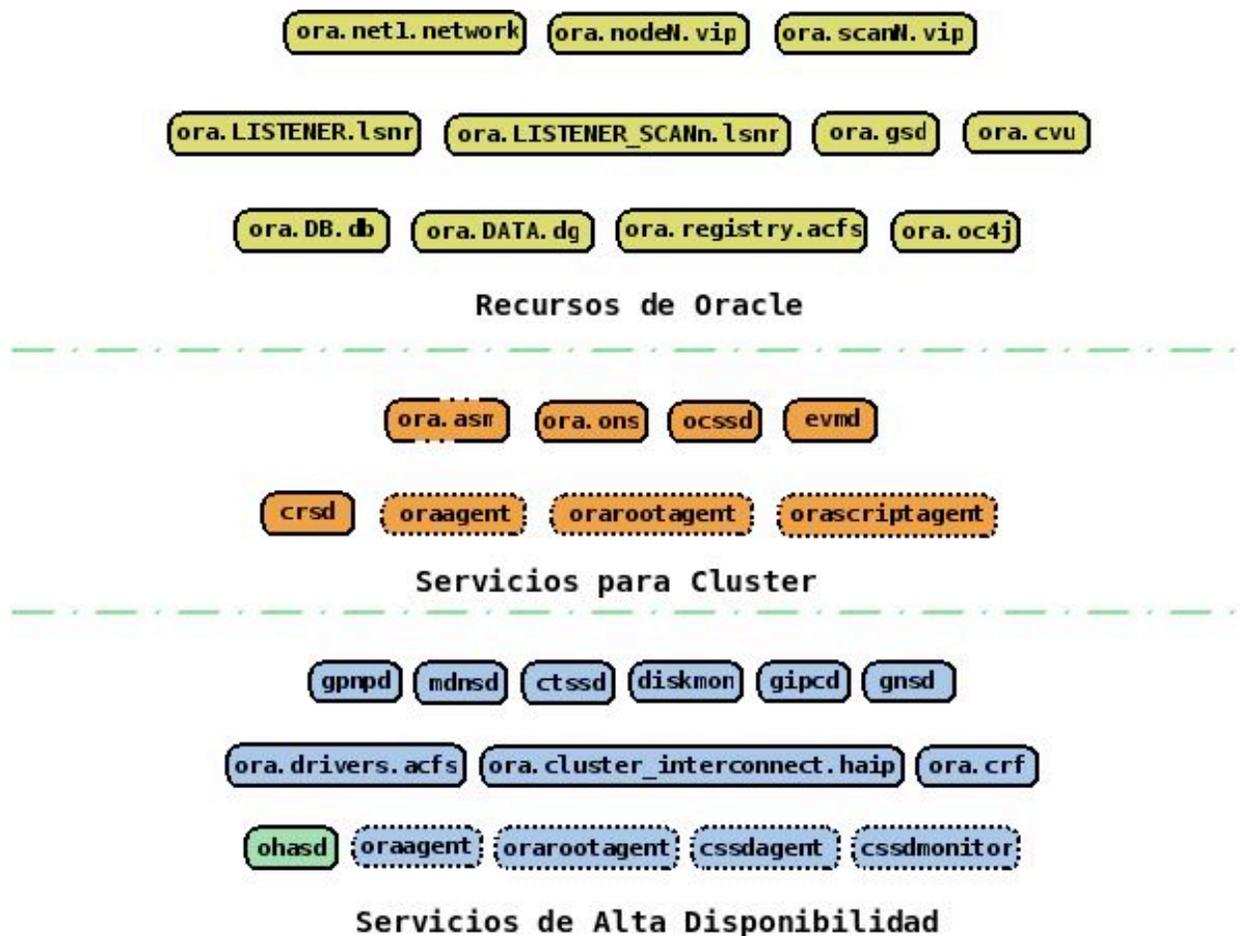


Fig. 1-2. Recursos existentes en el Clusterware de Oracle

Algo que es importante mencionar, es que el Oracle Clusterware se complementa con los mecanismos de protección de información de la Base de Datos para mantener la consistencia e integridad de la misma, en el caso de que se tenga problemas y se haga necesario reiniciar un nodo.

Por otro lado, el Clusterware de Oracle tiene la habilidad de integrarse con otros Clusterwares de terceros, o de otras compañías, para los casos en que por alguna razón la organización requiera de su uso para mantener las aplicaciones del negocio. En estos casos, el Clusterware de Oracle modifica su operación, para crear una dependencia con el Clusterware de terceros. Sin embargo el Clusterware de Oracle se requiere siempre, para poder tener RAC y ASM funcionando en el cluster.

ASM (Manejo de Almacenamiento Automático)

Antes de que ASM existiera, era común que en las organizaciones que usaban Bases de Datos de Oracle, no tuvieran una distribución adecuada de la información en los discos de la máquina, lo cual era causa de problemas de rendimiento, en la operación del día a día afectando así, no solo a la Base de Datos, si no también a las aplicaciones usadas por los usuarios.

ASM puede definirse como un manejador de volúmenes de almacenamiento, en el que básicamente, se registran todos los discos que serán usados para la Base de Datos con ASM, creando lo que se conoce como Grupos de Discos. Después cuando la Base de Datos es creada se indica el Grupo de Discos, donde los archivos de la Base de Datos son creados.

ASM se encarga entonces, de organizar y mantener la localidad donde la información es escrita o leída en cada uno de los discos. Pero son los procesos de la Base de Datos, los responsables de ejecutar dichas operaciones. Esto es que ASM, como cualquier otro manejador de volúmenes, le indica a la Base de Datos el lugar donde se lee o se escribe la información, pero los procesos de la Base de Datos, son los encargados de ejecutar las operaciones de escritura y lectura.

ASM como tal, se asegura de usar todos los discos asignados al Grupo de

Discos en la misma proporción, es por ello que se recomienda usar discos con la misma capacidad y velocidad. De lo contrario terminaremos con discos conteniendo más información que otros, afectando seguramente los tiempos de respuesta del sistema.

Otro beneficio que provee ASM, es la de hacer un balanceo automático cuando se añaden o se borran discos del Grupo de Discos, más sin embargo se tiene la habilidad adicional de ejecutar esta operación a través de un comando que permite además incluir paralelismo, si la situación lo requiere.

Otra propiedad de ASM, tendiente a proteger la información y mantener el sistema operando en el caso de existir una falla en algún disco o más de uno, es usando lo que se conoce como redundancia, que consiste en tener copias adicionales de información en distintos discos. ASM tiene tres formas de declarar la redundancia a nivel del Grupo de Discos:

Externa. Esto significa que ASM no hará copias adicionales, porque el arreglo de discos que se esta usando incluye dicha capacidad. En estos casos, si existiera una perdida de información se tendrían que usar los mecanismos propios del arreglo de discos, para hacer la recuperación de información.

Normal. En este caso ASM hará una copia adicional de los datos distribuyendo tanto la copia primaria y la secundaria entre grupos de Discos. Es decir que por ejemplo podríamos tener 8 discos, en los que se distribuirían las copias primarias y secundarias. En este esquema si se se perdiera un disco, la Base de Datos seguiría operando. Es importante aclarar que las copias primarias y secundarias se distribuyen entre todos los discos, es decir que no hay un disco que tenga solo copias primarias o solo secundarias. Así también, la misma copia primaria y secundaria de los mismos datos no pueden estar en el mismo disco.

Alta. En esta redundancia se tendrían tres copias de los datos, distribuidos entre los distintos discos que se han definido en el Grupo de Discos.

Es importante aclarar, que para la definición de la redundancia de tipo normal y alta deben de crearse dentro del Grupo de Discos, lo que se conoce como Grupos de Fallas. Un Grupo de Fallas es el conjunto de discos que comparten el mismo hardware, que puede ser la tarjeta controladora de discos de los nodos en el cluster. Al existir una falla en alguna parte del hardware, potencialmente todos los discos del Grupo de Fallas dejarán de estar disponibles. Cuando se crea un Grupo de Discos con redundancia de tipo normal, se requieren por lo menos dos Grupos de Fallas. Para el caso de redundancia alta, se necesitan por lo menos tres Grupos de Fallas.

Cuando tenemos una redundancia, diferente a externa, los procesos de las instancias de Base de Datos pueden leer de cualquier copia, primaria o secundaria, pero cuando se escriben los datos se hace una escritura simultánea en todas las copias. Si por alguna razón una copia no esta disponible, los procesos de las Bases de Datos pueden acceder alguna otra copia que si lo esté. Si no hay ninguna, entonces se da un error y ASM pondrá el disco involucrado en el error, en un estado fuera de línea.

El código de ASM se encuentra integrado con el del manejador de Base de Datos, para su funcionamiento existe como una instancia dentro de la misma máquina. En la figura 1-3 se muestran tres nodos, los cuales incluyen dos Bases de Datos operando en el cluster. Debe hacerse notar, que existe solo una instancia de ASM en cada nodo dando servicio a ambas instancias de las Bases de Datos. Los discos que se incluyen en el mismo diagrama son administrados por ASM:

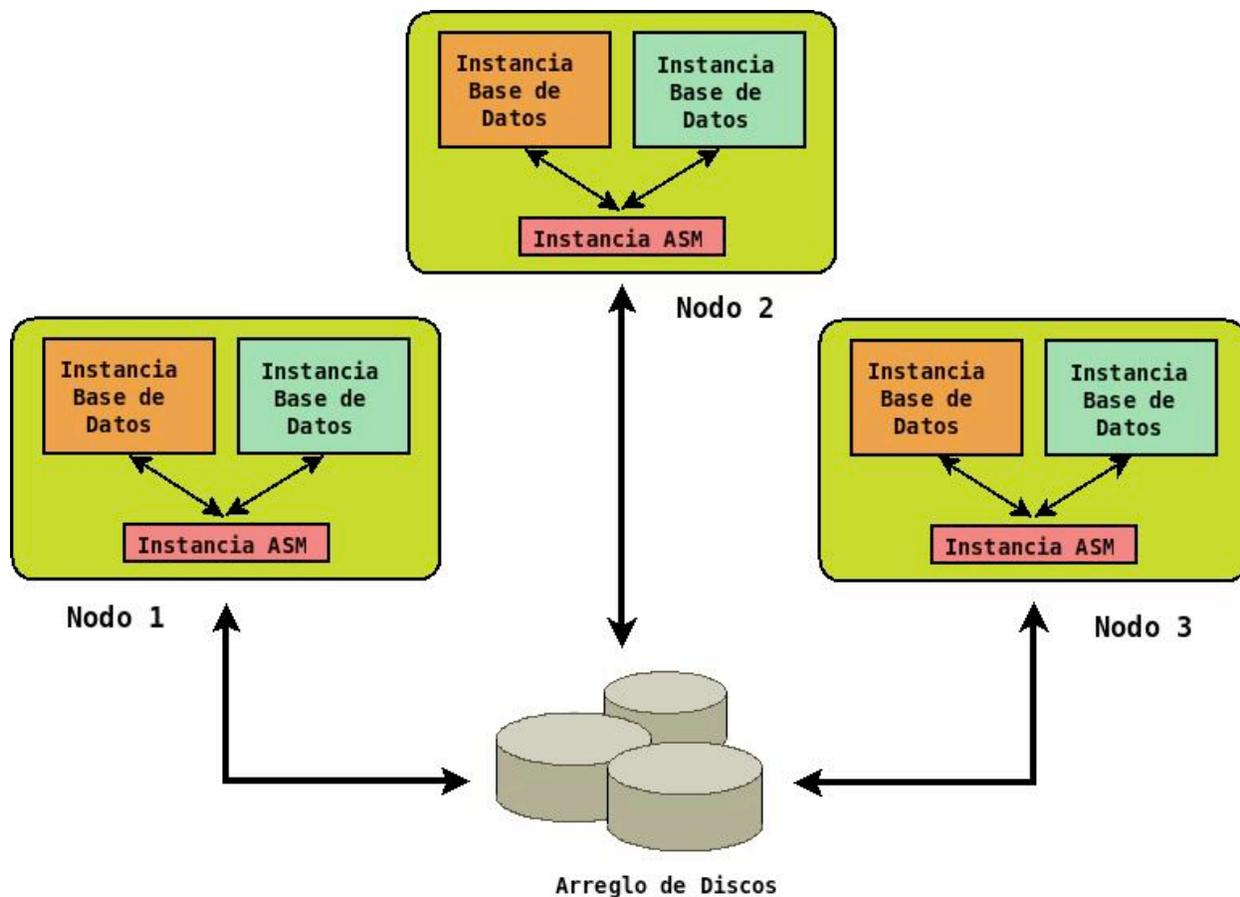


Fig. 1-3 ASM sirviendo a dos Bases de Datos en un cluster.

Como mencioné anteriormente, ASM existe como una instancia de tipo ASM, es decir que tendremos las mismas áreas de memoria y algunos procesos de las instancias de las Bases de Datos, pero su propósito será distinto por el simple hecho que ASM no maneja la información de la Bases de Datos, sino que es responsable de manejar los discos que se usarán en las Bases de Datos.

Una limitante que existía originalmente en ASM, es que no soportaba archivos que no forman parte de la Base de Datos (archivos de control, de datos, etc). Sin embargo, es común que los clientes usen archivos de texto, también llamados planos, que son copiados en el cluster y contienen información que es cargada a la Base de Datos. Para solucionar esta limitante, existe un sistema de archivos conocido como ACFS, el cual se configura sobre un Grupo de Discos de ASM. A nivel del núcleo del sistema operativo Unix, se crean los objetos necesarios,

como módulos en Linux, que permiten presentar al sistema de archivos ACFS como cualquier otro en la máquina, teniendo la capacidad de ejecutar cualquier comando nativo de Unix para el manejo de archivos como pueden ser mount, ls, cat, etc. Dado que ACFS se crea sobre ASM, también tiene la habilidad de funcionar en el cluster.

Debo mencionar que ACFS, a pesar de ser un sistema de archivos genérico no soporta archivos que sean manejados directamente por ASM, esto incluye por supuesto los archivos de las Bases de Datos. Además de archivos planos, ACFS puede ser usado para alojar la instalación del mismo manejador de Base de Datos en ambientes de cluster o de una sola máquina.

RAC (Clusters de Aplicación Real)

RAC es realmente una opción adicional, del mismo manejador relacional de Base de Datos, necesaria para la operación en un cluster. De hecho ASM usa la misma opción de RAC, para poder sincronizar su operación en las distintas máquinas. Sin embargo cuando se usa el término RAC, es para referirse a una Base de Datos operando en más de una máquina que forma parte de un cluster.

Al instalar RAC, quizá la parte más importante de su arquitectura, es la de proveer los mecanismos para hacer el uso de el área de memoria de *buffers*, que existe en cada una de las maquinas individuales y mostrarlas en su conjunto como un solo recurso existente en el cluster. Esto es que cuando un usuario necesita leer datos de alguna tabla, se hace el requerimiento local en una de las máquinas. Sin embargo si el bloque de datos no se encuentra en la memoria de buffers de ese nodo, entonces es posible localizar el buffer en alguna de las máquinas remotas. Si el bloque de datos existe, ya cargado en memoria, se envía a través de la red de interconexión del cluster. En el peor de los casos, si el buffer requerido no se encuentra ya cargado en alguna área de memoria de buffers, de cualquier máquina en el cluster, entonces el proceso que hizo el requerimiento recibirá el permiso para leer del disco el bloque de información

que se necesita.

Como puede verse, en RAC se trata de evitar el uso de la operación de lectura del disco, para satisfacer algún requerimiento de datos, y en su lugar tomar ventaja de la red de interconexión. Es por esta razón, que cada máquina debe de tener por lo menos dos interfaces de red, una para la comunicación privada dentro del cluster, que será usada por los productos de Oracle instalados. Y una para la comunicación pública, usada para el acceso a la Base de Datos de usuarios y aplicaciones. Otro requerimiento de instalación de RAC, es la de evitar usar la red de interconexión de los nodos, para algún otro objetivo que no sea el propio de RAC, ASM y el Clusterware. Cabe mencionar que para la red de interconexión, se recomienda que se use tarjetas controladoras de red de una velocidad de 1 Gb/seg, aunque poco a poco las de 10 Gb/seg están empezando a ser más utilizadas.

Cuando RAC ha sido instalado, existen procesos adicionales en cada una de las instancias de Base de Datos, si lo comparamos a una instalación en una sola máquina, lo cuales permiten coordinar el acceso de las diferentes áreas de memoria y datos, además de ejecutar operaciones de mantenimiento, como es el monitoreo constante de las instancias a nivel de la Base de Datos. Si por alguna razón existiera una instancia que no está funcionando adecuadamente, alguna otra puede eliminarla y de ser necesario escalar este requerimiento al Clusterware para que el nodo sea reiniciado, en el caso de que exista algún problema a nivel sistema operativo que impida que la instancia de Base de Datos sea terminada. Todas estas acciones son tendientes a evitar una pérdida del acceso total al cluster y mantener la alta disponibilidad.

RAC provee además mecanismos adicionales, cuando se tiene que hacer recuperación para el caso en que alguna instancia sea terminada de manera anormal.

En el área de diagnóstico, poco a poco se han creado procesos y mecanismos que permiten monitorear el sistema y generar información de diagnóstico de

manera global, en cada una de las máquinas, en el caso de que exista una situación anormal.

En la figura 1-4 se muestra el diagrama de un cluster con dos nodos, incluyendo algunos de los componentes de las instancias de Bases de Datos, como son los procesos que existen para mantener el sistema y servir los requerimientos de información hechos por los usuarios. Como se mencionó anteriormente, en RAC se tienen procesos adicionales que son LMS, LMD0, LMON, LMHB y DIAG. También se muestran las áreas de memoria que son compartidas entre los procesos que están operando en la misma máquina. Gracias a RAC es posible compartir los recursos alojados en esas áreas de memoria en todo el cluster. En la figura 1-4 también se muestra el arreglo de discos donde se alojan los archivos de Base de Datos:

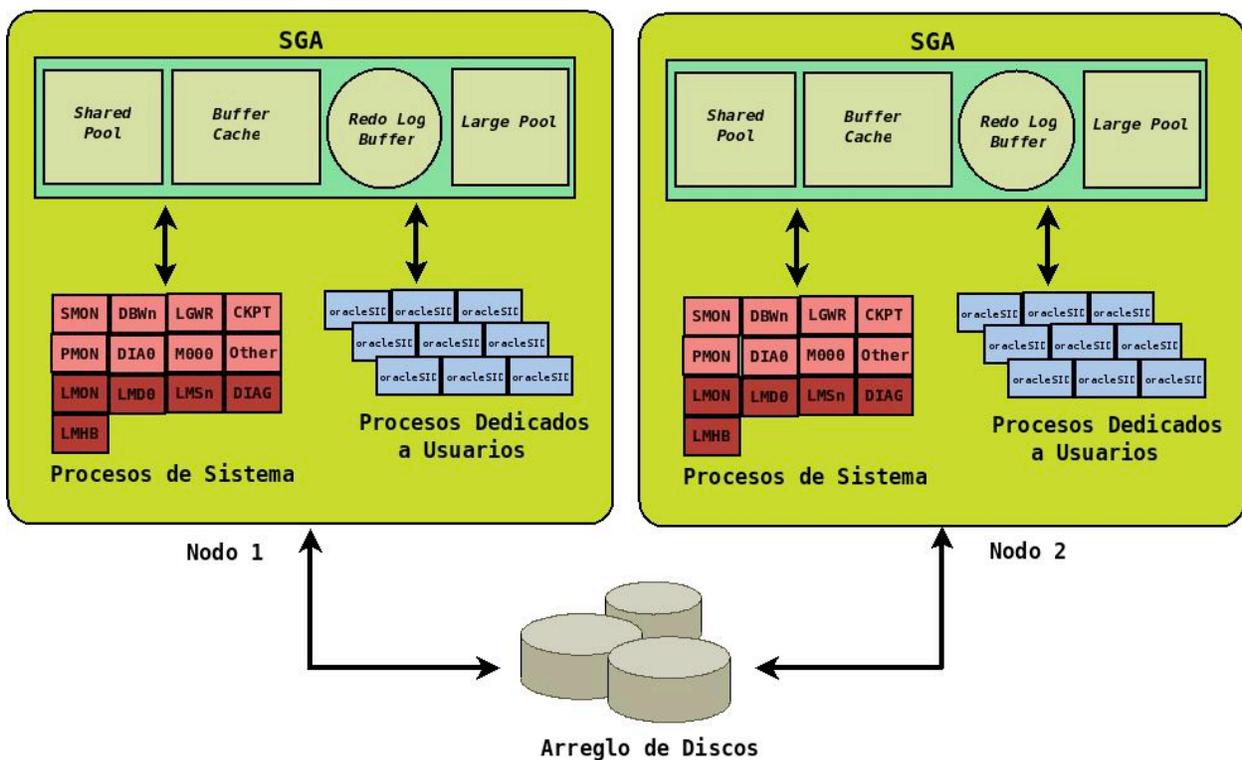


Fig. 1-4 Procesos de las Instancias cuando se tiene RAC instalado

Es importante mencionar que las estructuras de memoria y disco en RAC, son iguales a las de una instancia, que opera en un servidor que trabaja en un ambiente de una sola máquina. Esta propiedad permite hacer la conversión de cualquier Base de Datos a un ambiente de cluster fácilmente o desactivar la opción de RAC, de ser necesario.