

**Oracle Secure Enterprise Search**  
Implementing **search.oracle.com**, a Case Study

*An Oracle Technical White Paper*  
*May 2006*

EXECUTIVE SUMMARY .....	2
1. Hardware Overview .....	2
2. Failover/High Availability.....	3
3. Installation Overview .....	4
4. Post-installation Tasks.....	5
4.1 Log into Admin UI .....	5
4.2 Logfile rotation and housekeeping cronjob.....	6
4.3 Set SGA/PGA parameters.....	6
4.4 Recommended patches.....	6
5. Monitoring .....	6
5.1 oc4j monitoring and restart cronjob.....	6
5.2 Enterprise Manager installation.....	6
5.3 Allow the Enterprise Manager console to connect.....	7
5.4 Enterprise Manager target configuration.....	7
6. Backups.....	9
7. Secure Crawls.....	10
Appendix A .....	11
Appendix B .....	11
Appendix C .....	12

## EXECUTIVE SUMMARY

Oracle Secure Enterprise Search 10g 10.1.6 is an out-of-the-box solution that provides search capabilities across multiple repositories - Oracle databases and Portals, websites, email systems, files on disk and many more.

With Secure Enterprise Search, it is possible to crawl Oracle's Portal across the content assets of a corporation, bringing to bear Oracle's core capabilities of platform scalability and reliability.

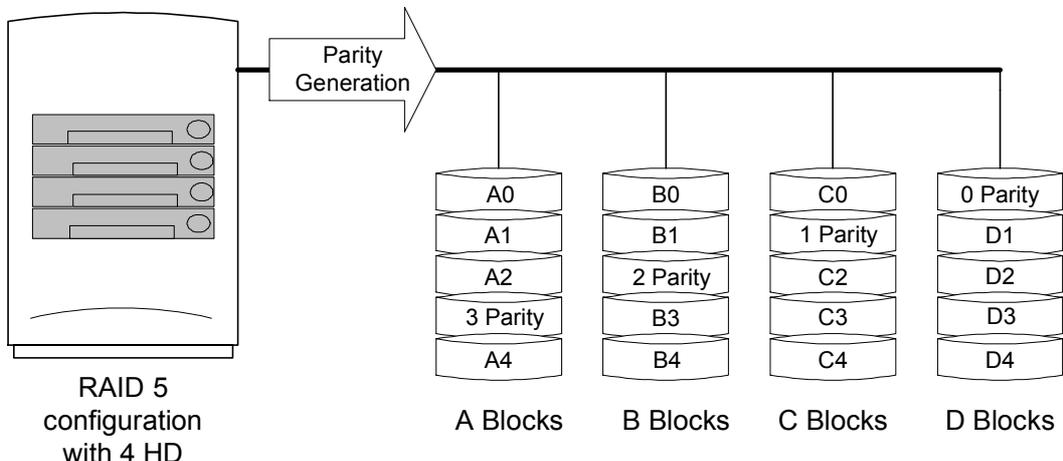
Oracle Secure Enterprise Search was configured to provide search-services for [www.oracle.com](http://www.oracle.com) (WWW) and [otn.oracle.com](http://otn.oracle.com) (OTN). This service can be accessed via the URL <http://search.oracle.com> and will be expanded over time to cover more of Oracle's content such as Documentation, Blogs, and Forums etc.

Following is an architectural overview and a look at best practices and lessons-learned gained from the experience of that implementation.

### 1. Hardware Overview

Oracle Secure Enterprise Search 10.1.6 was installed on 2 Dell Poweredge 2850 servers, each with 2 x 3GHz CPU's, but configured with hyper-threading to give the appearance of 4 CPU's. Each server has 8GB of memory and 12GB of swap configured.

A hardware RAID 5 configuration, over 4 300GB internal disks, was chosen for the implementation. It was considered the best overall balance of performance, fault tolerance, storage efficiency and price. RAID 5 stripes data and parity information across the drives, providing good performance especially for high number of reads. Further information on RAID configurations can be found at [http://en.wikipedia.org/wiki/Redundant\\_array\\_of\\_independent\\_disks](http://en.wikipedia.org/wiki/Redundant_array_of_independent_disks). The initial storage capacity of 700GB was considered adequate for the installation and additional storage could be added a later date. Create a cronjob to check the status of the RAID. See Appendix A for sample script. Set the cronjob to run every 15 minutes.



The servers are running Red Hat Linux Advanced Server version 3.2.3-49 (32-bit) with the 2.4.21-37.ELhugemem kernel.

Two filesystems were created on locally attached disks and allocated specifically for the SES installation:

/u01 - size 50GB (configured on a simple disk partition) - holding the binaries for the application

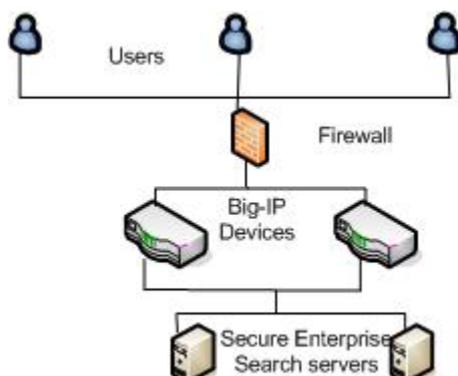
/u02 - size 700GB (configured on the underlying hardware RAID 5) - holding the logfiles, cache and database which, in turn, holds the crawled and indexed data

The minimum disk space required by the install is specified in the installation guide. The maximum amount of disk space required will be dependant on the amount of data to be crawled and indexed. Using the disk layout above, the /u01 filesystem is unlikely to expand much further, but the database on /u02 will continue to dynamically expand itself as necessary as more data is crawled.

With just over 350,000 documents indexed so far, the back-end database has dynamically extended to approx 15GB in size. The cache files are occupying 17GB and the logfiles 1GB of disk space so approx 33GB of disk space has been consumed on the /u02 filesystem. The cache and log files can be sent to a different filesystem after the installation, if required, by changing the Crawler configuration on the Global Settings tab via the Admin UI.

## 2. Failover/High Availability architecture

For high availability we use an active-active configuration across two SES servers. Each server contains a freestanding, but symmetrically configured, setup of SES running it's own set of crawls. In order to provide a single entry point to the `search.oracle.com` application, a pair of load balancing routers (active-passive) are located in front of these two servers. The type of LBR used is F5's BigIP. The LBRs are used for load balancing and failover. The LBR virtual server for `search.oracle.com` has been configured to use a round-robin method of load-balancing and sticky routing – this means that subsequent requests from the same client will continue to be routed to the same server for a set period. We are using a sticky routing period of 12 hours.



A monitor is defined on the LBR and assigned to the SES LBR nodes – at intervals, the http monitor performs a health check using the specified send string and checks for the existence of the receive rule to validate that the SES service is responding as expected. If the test to either node fails, that node will automatically become disabled in the LBR so ensuring that client requests are only directed to correctly functioning SES services.

**ECV Monitor oes\_monitor**

<b>Gets Defaults From (uses):</b>	http
<b>Interval:</b>	20
<b>Timeout:</b>	61
<b>Send String:</b>	GET /monitor/check.jsp HTTP/1.0\n
<b>Receive Rule:</b>	Enterprise Search
<b>User Name:</b>	
<b>Password:</b>	
<b>Reverse:</b>	<input type="checkbox"/>
<b>Transparent:</b>	no
<b>Destination IP:Service :</b>	*:*

This LBR configuration also enables one of the SES servers to be taken down, for maintenance or other reasons, without disrupting the service to the users.

Further information on LBR's in general can be found at [http://en.wikipedia.org/wiki/Load\\_balancing\\_\(computing\)](http://en.wikipedia.org/wiki/Load_balancing_(computing))

Details on F5's BigIP concepts and configuration in an Oracle environment are at [http://www.oracle.com/technology/products/ias/hi\\_av/bigip.pdf](http://www.oracle.com/technology/products/ias/hi_av/bigip.pdf)

### 3. Installation Overview

The Oracle Secure Enterprise Search (SES) architecture includes an Oracle database and a mid-tier standalone oc4j installation. Due to the fact that the SES product has been designed as an all-inclusive appliance, installation is a one-step process, making use of Oracle's Universal Installer and no specific dba or Oracle product skills are required to install and configure the application.

*Refer to the Oracle Secure Enterprise Search Installation Guide for full instructions and pre-requisites.*

Preplanning the Server Name during the installation must be taken, as the requirement for secure crawls is to have unique names for each of the instances being registered. Secure Enterprise Search is currently configured to register into Oracle Internet Directory (LDAP) only. During the directory setup configuration for secure crawls, the name of the application to be registered is derived from Server Name (aka ORACLE\_SID) specified during the install. There is a limitation of eight characters of this name.

If installing multiple Oracle homes on the same server, the default database port of 1521 is taken for the first installation and subsequent installations will increment by the database port by 1.

#### **4. Post-installation Tasks**

As soon as the Universal Installer has completed, the Oracle Secure Enterprise Search application is up and running and ready for use. However, it is beneficial to perform a few post-installation tasks to enhance the availability and functionality of the application.

##### **4.1 Login into the admin UI to set some defaults**

There is one account, eqsys, used to get into the Admin UI for SES. The Admin UI is the place to administer the activities of the application like creating the data sources to be crawled, scheduling and monitoring the crawls, viewing statistics of crawlers and queries, configuring directory and proxy settings to list just a few of the configurations maintained and monitored in the Admin UI. The “Help” link on the upper right corner is context-sensitive and will give definitions of the nomenclature found on the page.

Global settings can be used to set defaults for such items as number of threads or default character set. Although defaults are set in the Crawler configuration, these are also specified per data source and can be changed for each data source. Any changes made to a data source while a crawl is in progress will not be read until the next time a crawl is initiated. If a change needs to be applied sooner, the running crawl can be stopped and resumed. Most of the defaults are well set out-of-the-box for Crawler configuration, though changing here the number of threads, character set, or language to match the environment will help to avoid mis-configurations in the data sources.

### Configuration

Number of Crawler Threads

Number of Processors

Crawling Depth  No limit  
 Limit

Enable Language Detection  No  
 Yes

Default Language

Crawler Timeout Threshold (seconds)

Maximum Document Size (megabytes)

Default Character Set

Cache Directory Location (Absolute Path)

Indexing Batch Size (megabytes)

Clear Cache After Indexing  Yes  No

Under the Search section on the Global Settings tab is a link for Index Optimization. Enable the schedule to run at a time where the load on the server will likely be light.

### Index Optimization

To ensure fast query results, the crawler maintains an active index of all documents crawled over all sources. You can schedule when to optimize the index.

**Note: It is important that the index be optimized during hours of low usage. This ensures minimal disruption to users.**

Estimated Fragmentation Level (%): 0

Frequency: [ [Every week on Saturday at 02:00 o'clock](#) ]

Status: **Scheduled**

Start Time: **none**

Finish Time: **none**

Specify a maximum duration for the index optimization process. The actual time taken for optimization will not exceed this limit, but it can be shorter. Specifying a longer optimization time results in a more optimized index.

- Do Not Run Optimization Longer Than:  hours
- Until the Optimization is Finished

**4.2 Create a cronjob to rotate logfiles to avoid disk space being consumed excessively.** See Appendix C for source code. Set the cronjob to run once a day:

E.g.  

```
50 23 * * * /usr/local/bin/rotate_logs.sh > /dev/null 2>&1
```

### 4.3 Change SGA / PGA parameters

Follow the steps to increase these parameters as outlined in Quick Installation Guide. We increased the SGA on our servers to 2500M.

### 4.4 Install recommended patches (if any)

check the latest Readme files for Secure Enterprise Search at <http://www.oracle.com/technology/documentation/ses1016.html>

## 5. Monitoring

Oracle Enterprise Manager Grid Control Release 10.2 has been implemented along with some OS-level scripts in order to provide real-time monitoring, notification and statistical data.

**5.1 Create a cronjob that will automatically restart processes if any part of the component stack is found to be not responding as expected.** See Appendix B for source code. Set the cronjob to run every 2 minutes:

E.g.  
\*/2 \* \* \* \* /home/oracle/scripts/check\_oc4j.sh >>  
/home/oracle/scripts/check\_oc4j.log 2>&1

### 5.2 Ready the environment for monitoring by Enterprise Manager.

- Create a file called oratab in /etc, owned by the same account that installed the Enterprise Search code.
- Make an entry in /etc/oratab to reflect the ORACLE\_SID and ORACLE\_HOME of the Enterprise Search installation:

```
ses:/disk/oracle_home:N
```

- Install the Oracle Enterprise Manager agent software in a separate ORACLE\_HOME to the Enterprise Search software. See the Oracle Enterprise Manager Installation Guide for full details.

### 5.3 Allow the Enterprise Manager console to connect to the sqlnet listener.

- `cd $ORACLE_HOME/network/admin`
- Backup the existing copy of sqlnet.ora –  
`cp sqlnet.ora sqlnet.ora.orig`
- Edit the sqlnet.ora file – add the EM console hostname(s) to the list of tcp.invited\_nodes:

E.g.  
`tcp.invited_nodes=(rmpsmt02.us.oracle.com,127.0.0.1,rgmem1.us.oracle.com,rgmem2.us.oracle.com)`

- Reload the listener configuration:  
`lsnrctl reload`

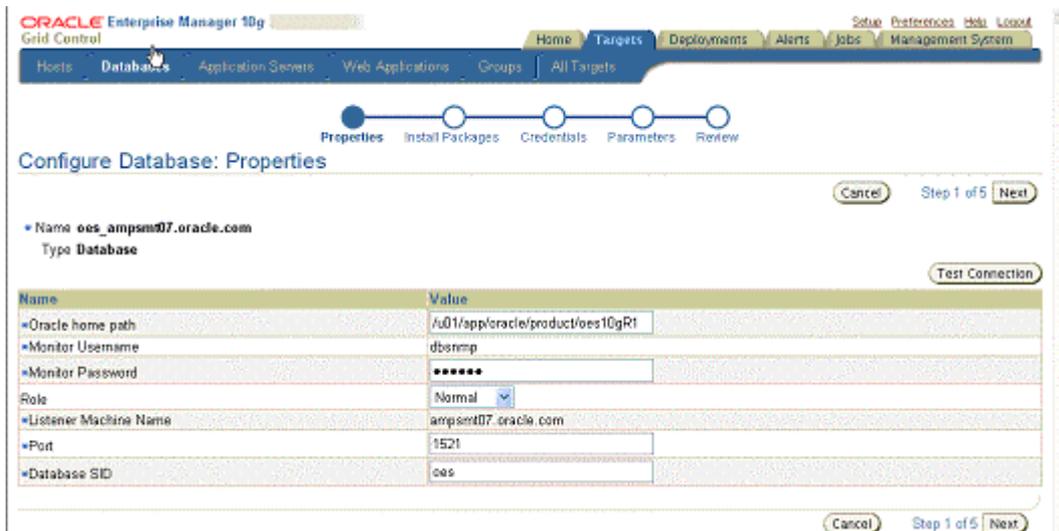
### 5.4 Create Enterprise Manager targets to monitor the Enterprise Search application

- The dbsnmp database account will need to be unlocked and the password reset to satisfy security requirement – this should be done from the command line on the SES server:

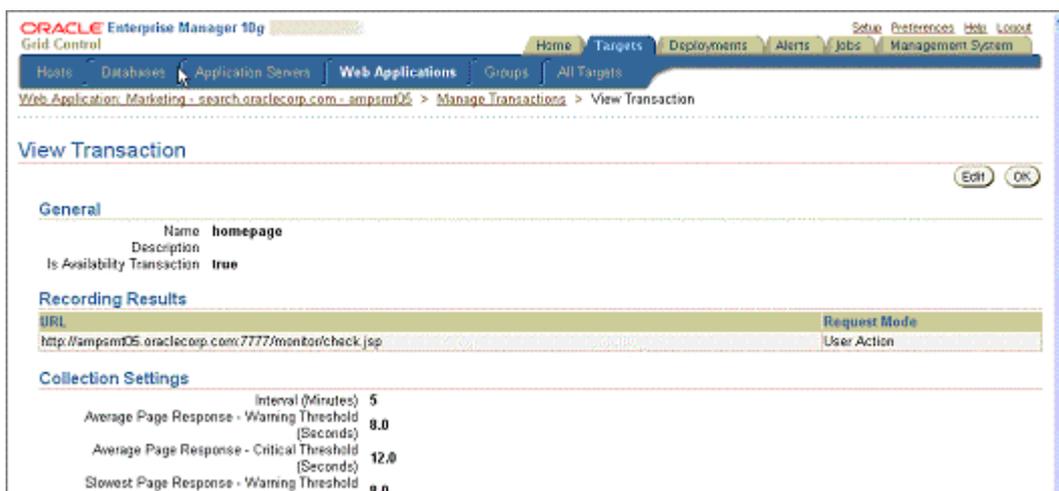
```
sqlplus sys/<password> as sysdba
alter user dbnmp account unlock identified by newpassword;
```

When unlocking any accounts that are delivered with the database installation, like dbnmp, the account will be affected when changing the password for eqsys. Changing the password for eqsys can easily be done through the Admin UI and does not require a restart of services. However, you will have to reset the password for any other accounts you may have unlocked as all unlocked accounts are reset to the new password specified through the Admin UI.

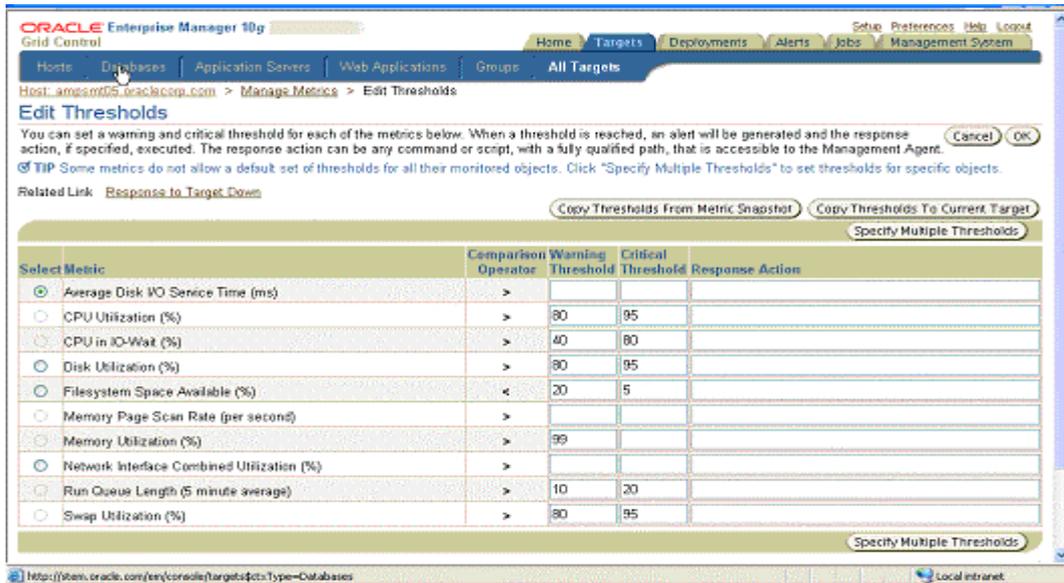
- Then the new dbnmp password needs to be specified in the EM Database configuration page for the SES database:



- Create web application monitors within EM with corresponding notification rules.

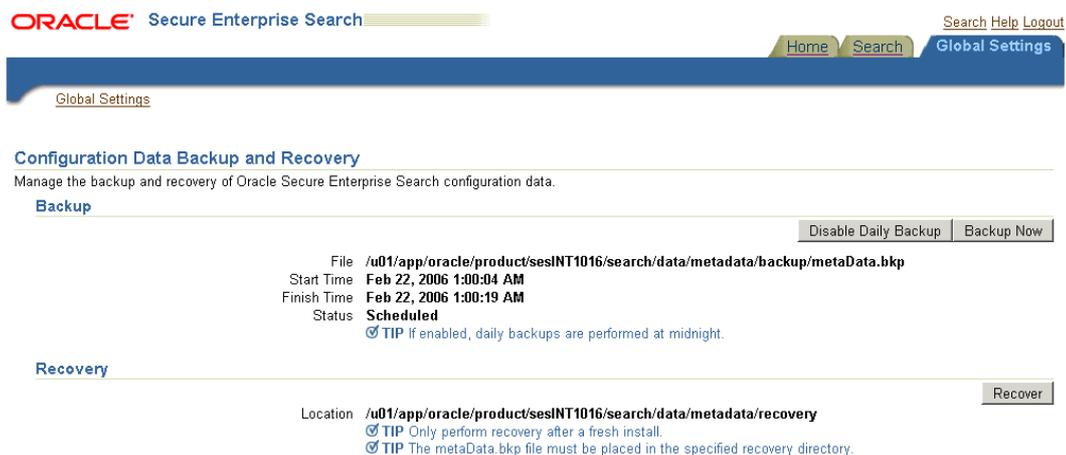


- Configure host target and metrics within EM to ensure host/hardware availability is monitored



## 6. Backups

If recovery of the application is required, this can be done by restoring the metadata and recrawling the data, after having re-installed SES. Nightly exports of the crawler metadata are enabled via the Admin UI:



The metadata is exported to a sub-directory of the ORACLE\_HOME and is subsequently backed up as part of a nightly /u01 filesystem image backup. If the SES needs to be re-installed, or if a new installation on another server is performed, the metadata can then be imported into the new database. This will recreate the previously defined Data Sources, crawl schedules etc. The crawls can then be re-performed to repopulate the database.

## 7. Setting up secure crawls

For secure crawls, an account with “viewall” privilege needs to be created. This account has to have access to crawl the data source with access to all items. For OracleAS Portal crawls, the crawler captures the guid of the page and this is used to determine which results to give the end user. User will see different results depending on whether they are logged in or not.

Registering the SES application into Oracle Internet Directory is straightforward process. The Directory Setup page can be found on the Global Settings tab in the System Section. Before running this configuration for the first time, it is a good idea to turn debug on as specified in Administration Guide. The log file that is written to for all error and success messages can be found in `$ORACLE_HOME/oc4j/j2ee/OC4J_SEARCH/log/oc4j.log`.

Enter all of the values specified on the page.

ORACLE Secure Enterprise Search [Search](#) [Help](#) [Logout](#)

[Home](#) [Search](#) [Global Settings](#)

[Global Settings](#)

### Directory Setup

#### Directory Connection Status

Enterprise Search is not connected to a directory.

#### Connect Enterprise Search to Directory

Enter the Directory parameters:

Host	<input type="text" value="ldap.yourcompany.com"/>
Port	<input type="text" value="389"/>
SSL Port	<input type="text" value="636"/>
Realm Distinguished Name	<input type="text" value="dc=company,dc=com"/>
Administrator User Name	<input type="text" value="cn=orcladmin"/>
Administrator Password	<input type="password" value="*****"/>

Enter Enterprise Search parameters:

Administrator User Name	<input type="text" value="eqsys"/>
Administrator Password	<input type="password" value="*****"/>

**TIP** Directory administrator user name should be in the format: cn=username  
 **TIP** Enterprise Search can be connected to a directory with the above parameters regardless of its current directory connection status

## Appendix A – cronjob to monitor the RAID disks

```
#!/bin/ksh

for FILE in /proc/megaraid/*/raiddrives*
do
    if grep -is degraded $FILE > /dev/null 2>&1
    then
        logger "WARNING: Error in hardware RAID. Review $FILE and the
front panel LCD."
    fi
done
```

## Appendix B -- Cronjob to automatically restart processes

```
#!/bin/bash
# simple shell script to do basic checking for SES still running
# enter appropriate values in the following 2 environment variables
export ORACLE_HOME=/disk/oracle_home
export SERVER=hostname.company.com # fully qualified hostname
# If there is only 1 installation on this server, the following values
should be correct
export ORACLE_SID=ses
export PORT=7777
export HOST=`hostname`
export DT=`date +%m%d%y:%T`

#-----
# Check if this script is already running
#-----
if [ -f /tmp/check_oc4j.txt ] ; then
    echo "Monitor still Running - exiting"
    /bin/mailx -s "WARNING: SES restart job is still running on ${HOST} :
${DT}. Please investigate" alerts@company.com < /tmp/check_oc4j.txt
    exit
fi
echo "oc4j check started at ${DT}" > /tmp/check_oc4j.txt

# first check for correct OC4J running
p=`ps -deafw | grep -e "-out
${ORACLE_HOME}/oc4j/j2ee/OC4J_SEARCH/log/oc4j.log" | grep -v grep | wc -l`

if [[ p -ge 1 ]]
then echo > /dev/null #echo "oc4j is running OK"
else echo `date`
    echo "OC4J is dead - restarting"
    ${ORACLE_HOME}/bin/searchctl restart
fi

rm -rf /tmp/check_SES_up_$$
mkdir /tmp/check_SES_up_$$
pushd /tmp/check_SES_up_$$ > /dev/null

wget --timeout 30 --no-host-directories --proxy=off
http://$SERVER:$PORT/search/query/search.jsp 2> /dev/null

if [[ -f search.jsp ]]
then echo > /dev/null #echo "Fetched search.jsp OK"
else echo `date`
    echo "No search.jsp fetched - restarting SES"
    ${ORACLE_HOME}/bin/searchctl restart
    wget --timeout 30 --no-host-directories --proxy=off
http://$SERVER:$PORT/search/query/search.jsp
fi

q=`grep "searchQbox" search.jsp | wc -l`
if [[ q -ge 1 ]]
then echo > /dev/null # echo "Query Page appears OK"
else echo `date`
    echo "Problem with query page - restarting SES"
```

```

        $ORACLE_HOME/bin/searchctl restart
fi

popd > /dev/null
rm -rf /tmp/check_SES_up_$$
rm /tmp/check_oc4j.txt
### End ###

```

## Appendix C – cronjob to rotate logfiles

```

#!/bin/sh
####
set -x
#
#   Archive current logfiles and remove logs older than 4 days
#### Environment - set correct value for ORACLE_HOME

ORACLE_HOME=/disk/oracle_home ; export ORACLE_HOME
ALERT_HOME=${ORACLE_HOME}/admin/ses/bdump ; export ALERT_HOME
NETWORK_HOME=${ORACLE_HOME}/network/log ; export NETWORK_HOME
J2EE_HOME=${ORACLE_HOME}/oc4j/j2ee/ ; export J2EE_HOME

NEWLIST=/tmp/loglist.txt
DATE=`date +%m%d%y:%T` ; export DATE

#### Start Clean
cat /dev/null > $NEWLIST

### find log files

    find $ALERT_HOME -type f -name \*log > $NEWLIST
    find $NETWORK_HOME -type f -name \*log >> $NEWLIST
    find $J2EE_HOME -type f -name \*log >> $NEWLIST

### For each line in NEWLIST, dev null the file
    for FILE in `cat $NEWLIST`
    do
        cp -p $FILE $FILE.archive.$DATE
        echo " " > $FILE
    done

### Remove old archived logs older than four days

find $ORACLE_HOME -type f -name \*.archive.\* -mtime +4 -exec rm
-f {} \; > /dev/null 2>&1

### End ###

```



**Oracle Secure Enterprise Search**

**May 2006**

**Authors/contributors: Julie Wilson  
Barbara Jones**

**Oracle Corporation  
World Headquarters  
500 Oracle Parkway  
Redwood Shores, CA 94065  
U.S.A.**

**Worldwide Inquiries:  
Phone: +1.650.506.7000  
Fax: +1.650.506.7200  
[www.oracle.com](http://www.oracle.com)**

**Oracle Corporation provides the software  
that powers the Internet.**

**Oracle is a registered trademark of Oracle Corporation. Various  
product and service names referenced herein may be trademarks  
of Oracle Corporation. All other product and service names  
mentioned may be trademarks of their respective owners.**

**Copyright © 2006 Oracle Corporation  
All rights reserved.**