ORACLE SECURE ENTERPRISE SEARCH 11g
VERSION 11.1.2.2

KEY FEATURES

RELEASE 11.1.2.2
- Stores 10x more content per SES server for the same license cost. Search performance on large scale content radically improved via new search engine index layout, internal parallelism & index partitioning
- Near real-time search support. Significantly reduces impact of frequent index updates on query performance. New feature automates index fragmentation management for the benefit of high churn applications like email or news content
- SES is now internally using WebLogic App. Server 10.3 and its JRockit VM & JDK
- Improved Manageability via new Administration APIs and Command line Interface
- Autovue 2D & 3D CAD file support. Open CAD files directly inside Autovue
- Customize lexing and stop list behavior
- Connectivity to E-Business Suite and Siebel. Many ERP modules offered out of the box. Search Modeler tools allow customers to create searchable objects for modules in EBS 11 or 12 which are not yet supported
- Connector Updates: Documentum 6.5, Lotus Notes 8.5, Siebel activity/attachment support. Higher incremental crawl performance for NTFS- and Oracle ContentDB sources, Image document services (ability to extract meta data from image file types)
- Oracle Access Manager (OAM) SSO support for Web-, database-, and RSS crawler driven sources (e.g. Content Server)

Oracle Secure Enterprise Search 11g (SES), a standalone product from Oracle, enables a high quality, secure search across all enterprise information assets. Key SES features include:

- The ability to search and locate public, private and shared content across intranet web content, databases, files on local disk or file-servers, IMAP email, document repositories, applications, and portals
- Excellent search quality, with the most relevant items for a query spanning diverse sources being shown first
- Sub-second query performance
- Highly secure crawling, indexing, and searching
- Integration with Desktop Search tools
- Ease of administration and maintenance – a ‘no-DBA’ approach to Search.

Information Uplift for the Intranet
As a result of search engines on the Internet, the power of effective search technologies has become clear to everyone. Using the World Wide Web, consumers have become their own information retrieval experts. But search within enterprises differs radically from public Internet search. The information that businesses store and use for decision making is scattered across millions of documents and data fragments. It resides on a wide range of, often incompatible, IT systems. Many documents are confidential or private -- access to these must be controlled to ensure that only documents that the user is authorized to see are returned by the search engine.

Internet search engines such as Google use the links that URLs provide between web pages to deduce the importance or relevance of a document in a given search. This mechanism, called Page Rank, is akin to a citation count, or a ‘vote’ for one page by another. Unfortunately, Intranet resources do not vote for each other in the same way: a document authored in PDF may not URL-link to the database record of
HIGHLIGHTS

- Japanese data search improvements
- Internal parallelization to leverage multiple cores and disk spindles. Exploits today’s high RPM disk drives.
- Supports high churn applications/very low latency of index updates by automating de-fragmentation of search engine index.
- Very flexible presentation of search GUI. Incorporates the Freemarker templating engine, a Java library which makes customization of the default UI very easy without the need for any coding.
- Supports management of large-scale deployments. Includes web service- and command line interfaces to administrative tasks & additional functionality not available in the Admin GUI.

BETTER INTRANET SEARCH QUALITY

- Web-page link analysis.
- Duplicate and near-duplicate document detection.

SUB-SECOND QUERY PERFORMANCE

- New internal index design techniques for high-performance, high-throughput queries over millions of documents.

SEARCHABLE REPOSITORIES

- HTML pages served up by a Web Server.
- Database Tables - Search. Oracle databases and any other databases that support the ODBC standard. Database tables can reside in Enterprise Search’s own database instance, or they can be part of a remote database accessed over a network. Both full text columns and “fielded columns” can be crawled.
- Files - Local or remote files can be made searchable.

Oracle’s Database and Application Server technologies are widely used by businesses to store and access both data and content. For a number of years, Oracle has invested in building out technologies for secure access to data and content, text search and indexing, as well as connectors to different data sources and IT systems.

Secure Enterprise Search, a new standalone product from Oracle combining the above technologies, provides a comprehensive search over all intranet content including databases, files on local disk or file-servers, IMAP email, document repositories, applications, and portals. The search is secure and fully multilingual. Advanced search, including meta-data search, is supported. The product has been designed to be simple to use and administer. This Data Sheet summarizes the key features.

A Document Service API provides developers with a hook into the SES crawler pipeline. They can obtain documents found by the crawler and manipulate them before they are passed to the indexer. This is useful if you would like to generate and insert your own meta data tags into your documents before they are being indexed. Or, use it to extract entities like addresses or phone numbers from your content for compliance or auditing purposes. Some of our customers even use this service to build their own customize search engine, while taking advantage of the
through the file:// protocol

- Emails - Emails and mailing lists can be crawled via the IMAP protocol.
- Oracle Application Server Portal repositories – including private and public Portal pages, folders, subfolders and text items.
- Oracle Collaboration Suite – files in Content Services and Calendar entries.
- Oracle ContentDB
- Desktop content via integration with Google Desktop for Enterprise, which can be optionally configured to search locally downloaded email, files on local storage, previously viewed web pages etc.

CONNECTORS

Pre-built connectors for legacy applications, shipped with the product. Support document-level access control/ utilize document ACLs (user and group-membership):

- Microsoft Windows NT Filesystems (NTFS)
- Oracle Content Server / UCM
- EMC Documentum Content Server DocBases
- IBM Lotus Notes/Domino databases and Email
- Microsoft Exchange 2003 Servers
- Cognos, Business Objects, Microstrategy (via partner)
- Selected modules of Oracle E-Business Suite, Oracle Siebel, Peoplesoft (planned)

SECURE SEARCH

- Search non-public sources, but see only what you are authorized to see.
- Security Plug-in API allows for directly accessing groups and users in non-Oracle LDAP-based identity management software.
- Support for secure crawling, including HTTPS (secure socket layer/encrypted data existing SES crawler, legacy software connectors, and infrastructure technology.

Developers can directly influence relevancy rankings by changing how document attributes like title and keywords are factored into ranking scores.

Through the file:// protocol

- Emails - Emails and mailing lists can be crawled via the IMAP protocol.
- Oracle Application Server Portal repositories – including private and public Portal pages, folders, subfolders and text items.
- Oracle Collaboration Suite – files in Content Services and Calendar entries.
- Oracle ContentDB
- Desktop content via integration with Google Desktop for Enterprise, which can be optionally configured to search locally downloaded email, files on local storage, previously viewed web pages etc.

CONNECTORS

Pre-built connectors for legacy applications, shipped with the product. Support document-level access control/ utilize document ACLs (user and group-membership):

- Microsoft Windows NT Filesystems (NTFS)
- Oracle Content Server / UCM
- EMC Documentum Content Server DocBases
- IBM Lotus Notes/Domino databases and Email
- Microsoft Exchange 2003 Servers
- Cognos, Business Objects, Microstrategy (via partner)
- Selected modules of Oracle E-Business Suite, Oracle Siebel, Peoplesoft (planned)

SECURE SEARCH

- Search non-public sources, but see only what you are authorized to see.
- Security Plug-in API allows for directly accessing groups and users in non-Oracle LDAP-based identity management software.
- Support for secure crawling, including HTTPS (secure socket layer/encrypted data

existing SES crawler, legacy software connectors, and infrastructure technology.

Developers can directly influence relevancy rankings by changing how document attributes like title and keywords are factored into ranking scores.

The 11.1.2 Release

Oracle SES is now internally using WebLogic Server 10.3 and its JRockit VM & JDK (both formerly BEA). The earlier used Standalone OC4J web server has been replaced. Oracle also replaced Document filters from Autonomy/Verity and we now ship Outside In HTML Export filters (formerly Stellent). The new filters add support for (among other formats):

- Microsoft Office 2007 .docx/.xlsx/.pptx formats
- Oracle Sun Star Office
- Haansoft Hangul documents

Product Upgrade from SES 10g is supported. A special post-upgrade script is provided to accommodate for optimizations and changes in the structure of the search engine index due to scale improvements in SES 11 (see sections below).

Store 10X More Data

Oracle SES 11g improves its search engine index structure to scale better for very large content sources and introduces internal parallelization to leverage multiple CPU cores and disk spindles.
stream), cookies to maintain a session, HTTP digest authentication, and HTML form login.

- Multiple ways of getting document visibility Access Control Lists (ACLs) from data sources: Applied by the crawler based on administrator input (“ACL stamping”), supplied by the sources (e.g., Document ACLs), or supplied by the search user (“self-service”)
- Integrated with Oracle Internet Directory (OID) LDAP-based enterprise security. Share LDAP-users and groups between your data sources and Enterprise Search. Restrict search results based on LDAP-group membership or user id

**Generic Security Infrastructure with plugs into:**

- Oracle Internet Directory (OID)
- Microsoft’s Active Directory
- Other plug-ins for Novell eDirectory, SUN ONE/iPlanet, OpenLDAP, Oracle Virtual Directory included (need to be activated)

**150 SUPPORTED DOCUMENT FORMATS**

- Sun StarOffice
- Other popular desktop formats – Lotus 1-2-3®, Lotus Freethand®, Corel Word Perfect®, etc.
- Adobe Acrobat® PDF.
- All other common document formats.
- Documents in these formats can be searched whether they are files on disk, email attachments, served by web servers, stored in a database or part of another supported repository.

**WEB QUERY LANGUAGE**

- Basic keyword search.
- Advanced, parametric search and Query Syntax, including

SES now fetches and caches index blocks from disk in much larger, contiguous chunks and buffers than before, minimizing the number of times the engine has to go to disk when serving search requests. Further changes in the layout of the index improve the performance of single word searches.

New parallel querying and index partitioning algorithms significantly improve search query performance and facilitate searches of very large data sources. The new query architecture is based on Oracle Database partitioning and on enhancements in Oracle Text.

Parallel query is automatically used for querying large, partitioned data sources. Each partition may reside on a separate disk.

To make best use of this feature, you should run Oracle SES 11g on a server with a 64-bit multi core CPU. The server should have at least 8GB of RAM and multiple fast disk drives, either directly attached or via SAN with sufficient available bandwidth.

**Fresher Search Results**

SES 11 also offers a significant reduction in query performance degradation caused by very frequent index updates as seen in, for example, near real-time indexing use cases or when a given schedule is executed very often (for example, in high churn applications like email- or news content-indexing).

The ideal goal for any search engine is to auto-manage index fragmentation. With semi-automatic index fragmentation management, Oracle SES comes very close to this goal. Some garbage collection will still be needed on a very infrequent basis, perhaps once a month. For this reason, the Oracle SES administrator still has the ability to schedule index optimizations to run during non-peak hours.

The new index fragmentation management feature is implemented on top of an enhancement in Oracle Text, STAGE_ITAB, which allows the search engine index to change and be updated while Oracle SES is executing searches. It does this by gathering changes to the in-memory index and periodically merging them with the larger disk-based search engine index.

You can turn index fragmentation management on and off, and you can specify the frequency of index merges.

**Presentation Enhancements**

SES incorporates the Freemarker templating engine, a Java library which makes customization of the default UI very easy. The idea behind Freemarker is that you separate UI design from the actual program code. This allows for changing the appearance of a UI page without the need for changing or recompiling code, because the application logic (the SES Java programs) and the query page design (Freemarker templates) are separated.

See figure 2 for an example how Freemarker can be used to customize and
Thesaurus (synonym terms, narrower term, broader term), fuzzy spellings, wildcard matching, NEAR, and nested search expressions like \((x \text{ AND } y) \text{ OR } z\).

- Dynamic page summaries. Search Keywords shown bold within title and excerpt of search-result page.
- Cached pages.
- Highlight query term on cached pages.
- Customize search result page look-and-feel using Freemarker templates

SEARCH ANALYTICS & METRICS
- Reports: Most popular queries, documents not found, click-throughs, and many other reports.
- Document Relevancy boosting. Administrator can boost document relevancy and customize result sets.

WEB BASED ADMINISTRATION CONSOLE
- Administer crawl activities
- Configure secure search
- Insight into user search behavior: Monitor crawl and search query statistics; most popular end user searches, searches with no results.

HIGH PERFORMANCE, SCALEABLE CRAWLING
- Multi-threaded Java crawler.
- Gather from multiple Web sites and other data sources, each on a different schedule.
- Crawler can be adapted to crawl new repositories through the Secure Crawler SDK
- Page link relationships are mapped and analyzed for better relevancy.
- Limit crawling to specific sections of your Intranet by setting 'inclusion' and 'exclusion' domains.

FLEXIBLE APIS
- Integrate Oracle Secure reconfigure the SES search page without the need for any Java coding.

SES provides Freemarker elements – templates, skins, macros – to customize all GUI elements of the search application.

Extensive Connectivity to Leading Portals-, CMS, ERP/CRM Applications
- E-Business Suite and Oracle Siebel. Peoplesoft planned. Many out of the box searchable objects in EBS R11 and R12 (HRMS, Procurement, CRM, Items, Assets, WF, iRep, Search Modeler available to enable other modules.) and Siebel
- SES powers Oracle Enterprise Crawl and Search Framework (ECSF), Oracle’s Search Framework across the entire Fusion Middleware Suite
- Essbase- and BI Search

New Administration API
A new Administration API supports the management of large-scale deployments by providing a command-line interface to administrative tasks previously only available through the SES Admin GUI:
- Create, change, or delete sources or schedules
- Start and stop schedules
- Configure SES crawlers
- Failed operations are automatically rolled back

Use the Administration API within an interactive session, or by executing commands from operating system prompt.

Other Noteworthy New SES 11 Features
- Windows Native Authentication via Kerberos. A type of Single Sign-on where Users who are already logged into their Windows workstation are automatically signed into the SES search page. Reduces time and password fatigue
- Oracle Access Manager (OAM) Support. Allows for Oracle Access Manager 10.1.4 or later to SSO-protect the Oracle SES default search application running on WebLogic Server. Crawling support for OAM protected sources is planned for a future release of SES.
- Reduced impact on crawled sources via a new User Authorization Cache
- Incremental Crawl Performance Improvements for Windows fileshares
and Oracle’s Content DB.

Figure 2: Freemarker templating greatly eases the process of customizing the look and feel of the default SES interface. Here, changing the word “left” (highlighted) to “right” will have the effect of moving the “Filter Results By” sidebar of Figure 1 from the left side of the page to the right side.

**Content Connectors**

SES features a family of built-in ‘connectors’ (Connectors are Java classes implementing the SES plug-in API) for unlocking stored content in many legacy repositories, including EMC Documentum, IBM Lotus Notes, FileNet and many others. All connectors are shipped with the SES product, easily activated, and protect your documents by providing access control integration.

**Comprehensive Security Support**

Secure Enterprise Search comprehensively addresses secure search. All your resources – intranet web pages, rows in database tables, Oracle Application Server Portal pages, emails, and documents in files or special repositories – can be protected. Search results are filtered so that the results page only shows links for documents to which the user has access. Oracle provides three ways through which this secure results filtering can be accomplished:

- Secure Enterprise Search can work from a centralized authentication scheme, like a Unix-, Microsoft-, or Oracle Internet Directory LDAP login, to identify which sources a given user can access. If you can’t access Oracle Portal per your LDAP privileges, for example, Secure Enterprise Search will not show you any results coming from your Oracle Portal system.

**Federated Search APIs**

- Federate queries to other Enterprise Search servers, or internet search engines, combining results in one display
- Suggested content feature lets you index and display real time content together with the search results. SES retrieves data from your applications and applies a stylesheet to create an HTML fragment
- Google OneBox providers can be configured as Oracle SES content providers

**Simple to Administer**

- Simple one-touch Install
- No-DBA approach to administration
- Browser-based interface to manage search configuration.
For more granular security, Secure Enterprise Search can store Access Control List (ACL) information associated with each document as part of the search engine index. Both ‘ACL Crawling’ and ‘ACL Stamping’ are supported. In ACL Crawling, Secure Enterprise Search obtains ACLs for each document directly from your repositories. This is not always possible, as when the source does not have a document model – e.g. an application generating dynamic content. In ACL Stamping, the search administrator specifies authorization roles directly in the Enterprise Search administrative console by entering a “grant” list of LDAP users, and groups, which are allowed to search a particular source. For example, all documents retrieved during a crawl can be marked searchable by anyone belonging to groups G1, G2 and G3. Or, all users U1, U2 and U3 can be granted permission to search all documents of a source.

Lastly, SES can filter search requests through the authorization mechanism of the source system. This is called Query time authentication. After the search engine index has returned a hit list, for each item in the list the search engine re-accesses each store, and passes to the store the user’s credentials to check if the user is (still) authorized to see the item. Thus, even if the access privileges have changed for the user since the last crawl, security is not compromised.

Security Plug-In Architecture
To avoid duplication of access control information, Secure Enterprise Search is integrated with Oracle’s own corporate identity management solution, Oracle Internet Directory (OID), and it can be synchronized with other Identity Management solutions like Microsoft’s Active Directory and Novell’s eDirectory products. SES can directly access Active Directory (no extra coding required) through a new authorization API and identity ‘plug-in’ architecture. SES ships plug-ins for Oracle’s Internet Directory and Microsoft’s Active Directory, among others. The new architecture even allows customers to build their own ‘identity plug-ins’ (supplies user and group information) for crawling sources with proprietary (non-LDAP) security schemes.

The Web Service Query API and query samples allow for specifying a search user. Only users known to your corporate LDAP server can submit searches.

A number of optimizations have been engineered to provide good performance. ACLs are cached in memory and evaluated at run time during query invocation. Authorization automatically turned off if there are no secure documents in an Enterprise Search instance.

Higher Search Quality
In a typical Internet search, hundreds of thousands of hits are returned. Similarly, with a comprehensive intranet search, the number of repositories increase, the ‘noise’ of poor matches in the search hit list can quickly overwhelm the user. As we have discussed, the Intranet is also handicapped by not being able to use URL
references to determine the popularity of pages. New approaches are needed to be able to find the needle in the intranet haystack.

Secure Enterprise Search has built a new relevance model optimized for the Intranet.

- It incorporates award-winning relevance ranking of Oracle Text to ensure that users consistently find very precise information, with its algorithms re-calibrated for typical intranet loads
- It builds separate internal indexes for data and metadata, and uses a unique weighting mechanism to prioritize metadata over data.
- It performs URL link analysis, where this is possible and useful.
- It includes de-aliasing and disambiguation. A significant amount of the ‘noise’ in an Intranet search comes from nearly identical documents. If you created a presentation, mailed it to a colleague as an attachment, who copied a dozen other people who stored it in different places, should the search engine return dozens of identical hits simply because the location of each is different? Secure Enterprise Search introduces detection of identical content available at multiple URLs or locations.

**Crawling and Filtering**

Secure Enterprise Search searches a variety of repositories. It gathers information by crawling your corporate intranet and looking through all the information that exists in the various repositories. Secure Enterprise Search features:

- Multi-threaded Java crawlers
- Ability to dispatch crawler on multiple machines
- Security in the crawl framework, including HTTP basic and digest authentication, as well as ACL gathering as described in the Comprehensive Security Support section above.
- For binary document formats – Office, PDF etc. -- filters are needed to extract plain text from the binary document. Filtering technology that automatically identifies document type and invokes the correct filter to produce textual data and metadata suitable for indexing. Filters are provided for most (150+) popular file types.

By using the crawl-index-search model, SES insures data gathering is non-intrusive - documents are analyzed, but are physically left in their original location under their own name. Crawls can be scheduled for low-load times.

**Metadata**

Metadata is one of the key enablers of effective enterprise search.

- Secure Enterprise Search extracts metadata fields from documents, database tables, email and other repositories.
It provides a flexible metadata mapping methodology, mapping the extracted metadata information to query page attributes for a powerful combination of full text retrieval and “fielded” text retrieval.

Many different attribute types, including date ranges and List of values (LOV) are supported.

Metadata search is automatically incorporated in Basic Search to determine which documents are the most relevant. It can also be explicitly invoked from Advanced Search.

Web-based Deployment & Administration Environment
Secure Enterprise Search is simple to use and deploy because it is based on the open standards of the Internet and a ‘no-DBA’ philosophy. Users can install the product, initiate crawling, and get some search results in a very short period of time. All the components needed internally by the search engine – text index, web server, metadata store, crawlers – are all bundled and need no separate installation or configuration. For secure search, in addition to installing SES you have to point it to the appropriate LDAP installation that manages user identity.

Flexible, Easy-to-Integrate Query Framework
Secure Enterprise Search lets you create custom search applications that work with any type of information by means of a set of query APIs. These APIs can be used from Web applications to retrieve and display query results. A Web service query API provides for an easy embedding of Enterprise Search into your pages. Even Search Administrator functions, such as data source creation and scheduling can be driven from Web Services.

Web-style Query Interface
To shorten your development cycles, Secure Enterprise Search includes a fully functional query application for users to query and display search results. The query application comes in Java (JSP).

Intuitive, Web-based Administration Environment
Secure Enterprise Search provides a simple browser-based administration tool that centralizes its maintenance. Wizard-like pages and dialog boxes simplify the tasks of identifying data sources, specifying security rules, scheduling crawls, monitoring search effectiveness, and tuning relevance.

Administrative functions: Search engine administrators can quickly and easily define information sources to be crawled, schedule maintenance crawling, and define user accounts for administrative users. Users simply identify themselves as an
authorized search administrator to the Secure Enterprise Search administration login Web page. Once the administrator logs in, the administration environment displays Wizard-like pages and dialog boxes for managing the search engine.

![Figure 3: Part of the SES Administration Interface.]

**Search Once, Find Everything with Federated Search**

Federated Search allows for searches both over Secure Enterprise Search crawled repositories and other heterogeneous data sources which do their own crawling and indexing. For several reasons a repository may not be suitable for crawling. The rate of updates may be too rapid for scheduled crawling. Opening a door for a crawler may lead to vulnerability for secure data sources, or crawling a database-backed website as a set of unstructured documents may be redundant if the data is already contained and indexed in structured form inside a database.

The ability to federate searches complements Secure Enterprise Search centralized crawling:

- Data sources keep their own full-text index or do their own crawling.
- SES accepts search terms from end users and brokers them to other data sources. The sources implement a web service using the SES query API to be able to send results back to SES for display. Samples are available for download.
Integration with Google Desktop for Enterprise

Secure Enterprise Search enables search of users’ local desktop content via an integration with the Google Desktop for Enterprise (GDfE), which can be configured to search locally downloaded email, files on local storage, previously viewed web pages etc. The use of this feature is optional, and meant only for organizations and employees who have separately chosen the Google’s Desktop for Enterprise for searching local desktop content.

Figure 4: Integration between Oracle Secure Enterprise Search and Google Desktop for Enterprise.

The integration is bidirectional; users are able to include results from GDfE on the SES search screens, and also configure SES as a data sources so that results from Oracle’s search show up when searching via GDfE.

Oracle is also working with other desktop search interfaces to provide multiple desktop search integrations.

Pricing

Oracle Secure Enterprise Search is licensed both by processor and by named users. Please consult [http://www.oracle.com](http://www.oracle.com) for the latest pricing information.

Summary

Oracle’s new standalone Secure Enterprise Search product allows you to reduce the
time spent finding relevant documents on your company's information repositories. It crawls, indexes and makes searchable your corporate intranet through a Web-style search. It eliminates the need for coding against hard-to-use low-level APIs. It organizes content from multiple repositories by extracting valuable metadata that can be used in portal applications. It provides effective search by returning more relevant hits - the best relevance ranking in the industry - and finds what you want. And it provides the best database integration and secure searching in the industry.