

An Oracle White Paper  
January 2010

# Information Management with Oracle Database 11g Release 2

Introduction.....	3
Advantages of Oracle Database Information Management Capabilities.....	3
Breaking the “Performance Barrier” .....	4
Oracle SecureFiles.....	5
Storage Optimization in SecureFiles.....	5
Files in the Database Reinvented – DBFS.....	6
Oracle Database File System (DBFS).....	6
DBFS Store API.....	6
Specialized Data Types and Data Structures.....	7
Oracle XML DB.....	7
Oracle Text.....	8
Oracle Secure Enterprise Search.....	9
Oracle Spatial.....	9
RDF, OWL and Semantic Database Management.....	11
Oracle Multimedia.....	11
Oracle Database 11g support for DICOM Medical Content Management.....	12
Conclusion.....	12

## Introduction

The operation of corporations, enterprises, and other organizations relies on the management, understanding and efficient use of vast amounts of information. This information may come from business applications – finance, order processing, manufacturing, and customer relationship management systems that easily conform to standard data structures (such as rows and columns with well defined schema). Increasingly, business value and operations depend on management, analysis and understanding of information that is not readily accessible without human or machine based interpretation. Common examples range from documents, XML, multimedia content, and web content to specialized information such as satellite and medical imagery, maps, and geographic information, sensor data, and semantic web structures. Oracle Database 11g Release 2 includes a wide range of capabilities that allow for intelligent management and analysis of these forms of information.

The ways in which these types of information are managed vary based on how the data are created and used:

- Huge volumes of data in desktop office systems (documents, spreadsheets and presentations) and specialized workstations and devices (geospatial analysis systems and medical capture and analysis systems)
- Multi-terabyte archives and digital libraries in government, academia and industry
- Image data banks and libraries used in life sciences and pharmaceutical research
- Public sector, telecommunications, utility and energy geospatial data warehouses
- Integrated operational systems including business or health records, location and project data, and related audio, video and image information in retail, insurance, healthcare, government and public safety systems
- Semantic data (triples) used in academic, pharmaceutical and intelligence research and discovery applications

## Advantages of Oracle Database Information Management Capabilities

Since the introduction of database management systems, database technology has been used to address the unique problems encountered when managing large volumes of all forms of

information. Databases are often used to catalog and reference documents, images and media content stored in files through “pointer-based” implementations. To store this “unstructured data” inside database tables, Binary Large Objects, or BLOBs have been available as containers for decades. Beyond simple BLOBs, for many years Oracle Database has incorporated intelligent data types and optimized data structures with operators to analyze and manipulate XML documents, multimedia content, text, and geospatial information. With Oracle Database 11g, Oracle is once again breaking new ground in the management of this information through dramatic improvements in the performance, security, and types of content managed by Oracle Database.

There are many reasons organizations store all forms of information with Oracle database management systems.

- Robust Administration, Tuning and Management:** Content stored in the database can be directly linked with associated data. Metadata and content are maintained in sync; they are managed under transactional control. The database also offers robust services for backup, recovery, physical and logical tuning
- Simplicity of Application Development:** Oracle’s support for a specific type of content includes SQL language extensions, PL/SQL and JAVA APIs, Xpath and Xquery (in the case of XML) and, in many cases, JSP Tag Libraries, as well as algorithms that perform common or valuable operations through built in operators.
- High Availability:** Oracle’s Maximum Availability Architecture makes “zero data-loss” configurations possible for all data. Unlike common configurations where attribute information is stored in the database with pointers to unstructured data in files, only a single recovery procedure is required in the event of failure.
- Scalable Architecture:** In many cases, the ability to index, partition, and perform operations through triggers, view processing, or table and database level parameters allows for dramatically larger datasets to be supported by applications that are built on the database rather than on file systems.
- Security:** Oracle Database allows for fine-grained (row level and column level) security. The same security mechanisms are used for all forms of information. When using many file systems, directory services do not allow fine-grained levels of access control. It may not be possible to restrict access to individual users; in many systems enabling a user to access any content in the directory gives access to all content in the directory.

## Breaking the “Performance Barrier”

Prior to Oracle Database 11g, these benefits came at a cost. Database features like domain indexes, partitioning, and parallelism can make geospatial applications and query and update-intensive XML applications perform better with content stored in the database than with content stored inside traditional file systems. However, in many other cases – multimedia applications, for example – managing and retrieving these types of data required additional processing power and memory to achieve performance equivalent to file systems.

All that changed with the introduction of SecureFiles in Oracle Database 11g Release 1. SecureFiles is a new, high-performance storage feature that enables storage and retrieval of LOBs at speeds equal or superior to that of equivalent file system configurations. SecureFiles is a major re-architecture featuring entirely new disk formats, space and memory management techniques, and delivers drastically improved LOB performance along with optimized storage.

## Oracle SecureFiles

SecureFiles is designed with a completely new approach to how the database handles file data, and delivers file system-like performance for basic query and insert operations. The optimized algorithms in SecureFiles make it up to 10x faster than previous LOB support (now called BasicFiles). SecureFiles can take advantage of several advanced Oracle Database capabilities that are not possible with file systems:

- In an Oracle RAC environment, SecureFiles offers high levels of scalability that go far beyond what is offered in file systems
- SecureFiles allows for easy migration from older LOBs using Online Table Redefinition without affecting existing applications
- Applications no longer have to deal with multiple interfaces for manipulating relational and associated file data
- With SecureFiles, all information can be part of a database transaction, freeing the application from the complexity of guaranteeing atomicity, read consistency and other backup and recovery requirements

SecureFiles also extends Transparent Data Encryption (TDE) capability to LOB data. The Oracle database supports automatic key management for all LOB columns within a table and transparently encrypts/decrypts data, backups and redo/undo log files. Applications require no changes and can transparently take advantage of TDE capabilities with SecureFiles LOBs. SecureFiles supports the following encryption algorithms:

- 3DES168: Triple Data Encryption Standard with a 168-bit key size
- AES128: Advanced Encryption Standard with a 128 bit key size
- AES192: Advanced Encryption Standard with a 192-bit key size (Default)
- AES256: Advanced Encryption Standard with a 256-bit key size

## Storage Optimization in SecureFiles

Also available with SecureFiles are advanced file system features such as Deduplication and Compression. Deduplication eliminates multiple, redundant copies of SecureFiles data and is completely transparent to applications. Oracle Database automatically detects multiple, identical SecureFiles data and stores only one copy, thereby saving storage space. Deduplication not only simplifies storage management, but also results in significantly better performance, especially for copy operations.

SecureFiles data can be compressed using industry standard compression algorithms resulting in significant savings in storage and improved performance. Oracle Database automatically determines if the SecureFiles file is compressible or if compression savings are beneficial.

SecureFiles uses a server-wide default LOB compression algorithm and provides for varying levels of compression. Each compression level represents a tradeoff between compression factor and speed. Organizations can choose the compression level which best suits their needs based on storage and CPU usage constraints. SecureFiles files are compressed and uncompressed automatically, transparent to applications.

Both Deduplication and Compression are part of the Advanced Compression Option, a separately licensed database option available with Oracle Database 11g.

## Files in the Database Reinvented – DBFS

In Oracle Database 11g Release 2, Oracle introduced Database File System (DBFS) and reinvented files in the database. Through the combination of SecureFiles high performance file storage, and the implementation of simple file system interfaces to files stored in the database, Oracle has enabled existing file based tools to access database files. DBFS provides familiar file access through pathnames, directories, and links. These files are kept in a dedicated SecureFiles store, or may be stored as SecureFiles LOBs in existing application tables. With the introduction of DBFS, storing business data files inside the database is now simpler, faster, and more robust than storing them outside the database.

### Oracle Database File System (DBFS)

DBFS provides a client for Linux that allows the mounting of DBFS file systems on Linux hosts, similar to a standard Network File System (NFS) mount, allowing applications to make normal file system calls. The Linux FUSE module forwards file system calls to the DBFS client executable, which makes remote calls to DBFS “Stores” in the database.

Oracle Database 11g Release 2 comes with two built-in Store Providers: DBFS SecureFiles Store and DBFS Hierarchical Store (or HSM). The DBFS SecureFiles Store utilizes a table with a SecureFiles column to store the file system data. It implements POSIX-like capabilities. The DBFS Hierarchical Store allows files to be written to any tape storage units supported by Oracle Recovery Manager (RMAN) or to a cloud storage system. Currently, Amazon S3 is the only supported cloud storage system.

DBFS also has a client component called `dbfs_client`, which provides a command interface to allow files to be easily copied in and out of the database from any host on the network. It implements simple file system commands like `list` and `copy` in a manner that is similar to the linux shell utilities `ls` and `rcp`. The command interface creates a direct connection to the database without requiring an OS mount of DBFS.

One key advantage of DBFS is that it can use Oracle Real Application Clusters (RAC) to deliver high availability and scalability beyond what is available with traditional file systems. For seamless scalability, RAC allows for additional processing and storage resources to be added to the system without disrupting applications. In the event of a failure, Oracle RAC transparently redirects file access to alternate RAC instances.

### DBFS Store API

DBFS uses an API referred to as the DBFS Store API. This API includes Create and Delete file operators for directories and links, Get/Put to read and write LOBs and attributes of existing file paths, Directory operations, Locking operations, and Snapshot creation. The DBFS Store API offers strong support for storage of metadata associated with files and transactional file system operations.

Developers can build file system implementations in the database by writing a PL/SQL package matching the DBFS Store API. This approach is conceptually similar to Linux FUSE user mode file system interface, and it allows many kinds of DBFS Store Providers to be created. Some examples include:

- A provider to allow file system access to LOBs in an application table
- A filter file system provider that passes operations to an underlying file system, but adds additional logic, e.g. A virus check filter, or filter that enforces application rules on access
- A provider that translates relational data into file data, or vice-versa

## Specialized Data Types and Data Structures

In the same way that database management systems include data types, storage and index structures, and operators to allow for meaningful query and analysis of structured data, they require these elements to add value when managing unstructured data. These features of Oracle Database 11g offer unique advantages specific to the management of XML, Text, Spatial, Semantic, and Multimedia and DICOM data.

### Oracle XML DB

XML has been widely adopted in just about every industry. XML based standards can be found in the Health-care, Manufacturing Financial Services, Government and Publishing sectors. The introduction of XML-based standards, such as XBRL, has led to XML becoming the de-facto mechanism for exchanging information among application systems. This has led to a growth in the use of XML as a persistence model for mission critical data.

To meet this need, Oracle developed Oracle XML DB. Oracle XML DB is a high-performance, native XML storage and retrieval technology that is delivered with all versions of Oracle Database. It provides full support for all of the key XML standards, including XML, Namespaces, DOM, XQuery, SQL/XML and XSLT. Oracle XML DB is the first platform to deliver true hybrid relational / XML capabilities, making it possible to bring the full power of the SQL language to bear on XML content and the full power of the XML paradigm to relational data.

Oracle Database 11g extends its industry leading XML support ensuring that Oracle Database remains the best platform for storing, managing and querying all possible types of XML content. Features in Oracle Database 11g offer improved performance and scalability and enable

complete support for the flexibility that makes the XML data model so attractive to so many different organizations.

Oracle Database 11g offers a number of improvements for users of Oracle XMLSchema-optimized XML storage:

- In-place evolution of XML Schemas
- Oracle Partitioning of XML Schema optimized storage
- Intelligent defaults for XMLSchema-optimized for an optimal storage model
- XQuery operations on Schema-Optimized storage improvements
- Support for replication of text-based XMLType storage via Oracle Streams

To address non-schema based XML in an optimal manner, Oracle Database 11g introduced a new Binary XML storage option and new XML Indexing capabilities that deliver high performance insert, update and query operations. This Binary XML format allows very efficient path based indexing of XML content. The format provides optimization of both XQuery execution and fragment extraction. The new XML Indexing capabilities of Oracle Database 11g take full advantage of this.

Oracle Database 11g simplifies the implementation of light-weight Service Oriented Applications by exposing PL/SQL packages, procedures and functions directly as web services. Oracle XML DB repository now includes an event model to support workflow type applications where the simple task of creating or modifying a file automatically initiates the appropriate process.

In Oracle Database 11g Release 2, Oracle XML DB includes significant improvements in performance and scalability and enhanced support for extremely complex XML schema like those used by industry standards such as the United States - Generally Accepted Accounting Principles (US-GAAP) and National Information Exchange Model (NIEM). This release also improves support for Oracle Partitioning, Real Application Clusters resulting in improved Xquery, Binary XML, XML generation and XML DB repository operation performance.

## Oracle Text

Oracle Text is the leading text searching, retrieval and management system to be integrated into a database environment. With Oracle Database 11g, Oracle Text introduced new features that aim to keep it in the leading position. These new features include:

- Improved Performance
- Minimization of application downtime
- Internationalization
- Ease of Maintenance

The performance of “mixed queries” – queries that have a text search part and a structured part – has been improved through the introduction of SDATA Sections and Compound Domain Indexes. The number of supported partitions has been dramatically increased; in Oracle Database 10g, the maximum number of partitions that could be used was 9999; in Oracle



Database 11g the limit for text index partitions is now the same as the limit for table partitions –  $2^{20} - 1$ , or 1,048,575.

With Oracle Database 11g, Oracle Text also supports Incremental Indexing and Online Index Recreation to improve system availability. Incremental Indexing allows you to create an index gradually at quiet times for your system. Online Index Recreation lets you create a “shadow” text index that can be built while the original index is still in use. When the index build is complete, the original index can be exchanged for the newly built shadow index. As soon as this is done, queries will automatically transition to the new index.

In addition, with Oracle Database 11g, additional internationalization support enables automatic language identification, stemming and segmentation operations for many more languages.

### Oracle Secure Enterprise Search

Oracle Secure Enterprise Search (SES), a standalone Oracle product, enables a high quality, secure search across all enterprise information assets. With Oracle SES 10.1.8.4, the key SES features include:

- The ability to search and locate public, private and shared content across intranet web content, databases including Oracle Database 11g, files on local disk or file-servers, IMAP email, document repositories, applications, and portals
- An excellent search quality, with the most relevant items for a query spanning diverse sources being shown first and easy navigation to drill down the search result with the Result Hit Clustering
- 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

Oracle SES allows you to save 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.

### Oracle Spatial

In repeated studies by IDC, Oracle is the most widely used enterprise spatial database server with over 80% of the enterprise spatial database market. Every Oracle Database includes the Locator capability, built-in location features that enable any business application to directly incorporate location information and realize competitive advantages.

Oracle Spatial 11g is a comprehensive spatial database offering, including native support for vector and raster data, topology and network models, 3D data, geocoding, routing, and OGC-standard Web Services. It is designed to meet the advanced geospatial requirements of business

and government applications such as business intelligence, land management, utilities, defense, and homeland security. With open native spatial support, Oracle Spatial eliminates the cost and complexity of separate, proprietary systems while enabling the use of all leading GIS tools. This extends Oracle's industry-leading security, performance, scalability, and manageability to mission critical spatial assets.

With Oracle Spatial 11g, Oracle introduces:

- Integration with Oracle Business Intelligence Suite Enterprise Edition, Oracle Fusion Middleware WebCenter, and Oracle Business Activity Monitoring, through Oracle Fusion Middleware MapViewer.
- Support for storage and management of 3-dimensional data, point clouds, and terrain models.
- OpenGeospatial Consortium (OGC) Web Services standards: Web Map Service (WMS), Web Feature Service – Transactions (WFS-T), Web Catalog Services (CS-W), and Open Location Services (OpenLS).
- GeoRaster data type and network data model now handles significantly larger data sets with greater simplicity.
- Routing engine, geocoding, topology data model, and network data model enhancements.

Oracle Spatial 11g provides significant new functionality that makes it the complete data management platform for any geospatial or enterprise location-enabled application. The spatial geometry data type has been enhanced with support for 3-dimensional data and new data types have been added with support for applications in domains such as urban planning, homeland security, or Lidar-based map production. These applications require storage and management of urban models, point clouds, and terrain models. Oracle Spatial now supports geospatial web services standards, to provide a secure, scalable service-oriented architecture platform. The GeoRaster data type and network data model have been enhanced to handle data sets larger by orders of magnitude with high performance and with greater ease of use.

Combined with the performance, scalability, and security of Oracle Database, Oracle Spatial 11g is the most advanced spatial database platform available for enterprise class deployments.

Oracle Spatial 11g Release 2 includes numerous enhancements for performance, management, network applications, routing and geocoding, GeoRaster, 3-dimensional data and operations and standards-based web services.

Oracle Spatial now has faster coordinate system transformation operations and supports cross-endian operations for transportable tablespaces and spatial indexes. The Oracle Spatial geocoder now includes point geocoding; the routing engine includes support for truck routing applications. Oracle Spatial 11g Release 2 support for GeoRaster has a new Java API and is now supported by the GDAL open source Extract-Transform-Load tool. GeoRaster enhancements include support for reprojections, Ground Control Point-based georeferencing, clipping, and interpolations. A modeling, visualization and simulation infrastructure has been added to the support for 3-dimension data, which was introduced in Oracle Spatial 11g Release 1.

With Oracle Spatial 11g Release 2, conformance to Open Geospatial Consortium (OGC) and ISO TC211 standards has been enhanced. This release supports OGC Web Feature Server (WFS) 1.1 and has full support for database transactions on WFS feature tables.

## RDF, OWL and Semantic Database Management

New software and data models are emerging to help in sharing of knowledge among multiple applications in areas such as data/content integration and enterprise application integration. This software will be based on semantic data modeling standards, such as RDF and OWL from the W3C.

Oracle Database 11g incorporates native RDF/RDFS/OWL support, enabling application developers to benefit from a scalable, secure, integrated, and efficient platform for semantic data management. These features are part of Oracle Spatial 11g. Application developers can add meaning to data and metadata by defining a set of terms and the relationships between them. These sets of terms (“ontologies”) enable query, analysis and actions based on semantic content, rather than simply data values. Ontologies are increasingly used to build applications that utilize domain-specific knowledge. Ontological data sets, often containing 100s of millions of data items and relationships, can be stored in groups of three, or "triples" using the new RDF data model. Oracle enables scaling to billions of triples to meet the needs of the most demanding applications.

The semantic features have been significantly enhanced in Oracle Database 11g Release 2. With support for Virtual Private Database and Oracle Label Security, the semantic database can be secured at the triple level. New support for popular natural language tools and services enables semantic indexing of documents. This release extends support for the SPARQL query language, the Web Ontology Language (OWL) with by supporting union, intersection, OWL2 property chains and disjoint properties. Plug-ins and SDK integration for open source products like Pellet OWL DL reasoner, Jena version 2.5, and Sesame simplify the use of Oracle Database 11g for enterprise semantic applications. In addition, this release includes support for widely-used standard ontologies like the W3C Simple Knowledge Organization System (SKOS) and Systematized Nomenclature of Medicine (SNOMED).

## Oracle Multimedia

Oracle Multimedia (formerly Oracle *interMedia*) is a feature that enables Oracle Database to store, manage, and retrieve images, audio, video, or other media data in an integrated fashion with other enterprise information. Oracle Multimedia extends Oracle Database reliability, availability, and data management to multimedia content in traditional, Internet, electronic commerce, and media-rich applications.

With Oracle Database 11g, Oracle Multimedia includes significant performance and scalability improvements. Multimedia supports SecureFiles, to dramatically improve performance and significantly strengthen the native content management capabilities of Oracle Database. In addition, the size limit for individual media objects that can be stored and retrieved within database storage structures (BLOB) is increased to the BLOB size limit, which is between 8 terabytes and 128 terabytes.

In addition to storing and retrieving large images, Multimedia can also extract image attributes

including height, width, and compression format for images that contain up to two billion pixels, or with a resolution of up to 46000x46000.

Oracle Database 11g Release 2 includes database-enforced image watermarking, improvement to thumbnail generation, and updated support Microsoft Windows Media Services and RealNetworks streaming servers.

## Oracle Database 11g support for DICOM Medical Content Management

Oracle Database 11g includes features and delivers the performance necessary to build large-scale repositories and archives of DICOM format medical images. By extending Multimedia to store image, audio and video using SecureFiles in Oracle databases, all the security, performance and management tools that have made Oracle Database the standard for enterprise class databases are now available for huge archives of media objects as well.

Specifically for medical image applications, Oracle Database provides methods to:

- Convert images to formats useful in web applications to simplify development of visually oriented applications.
- Extract both standard and private metadata for indexing.
- Validate that the metadata conforms to the DICOM standard and/or user -and vendor-specific extensions of the standard.
- Remove all private patient data to create anonymous images for research or training.
- Create new images with corrected metadata.
- Create DICOM format images from non-DICOM images.

Oracle Database 11g Release 2 delivers up to three times faster metadata extraction of DICOM tags. It includes a new mid-tier and client metadata extraction tool to enable preprocessing operations and enable metadata-based partitioning of DICOM data in the database. This release supports DICOM video including conversion of DICOM video to AVI and MPEG for viewing in any browser and generation of DICOM video from MPEG video and metadata.

All of these features are built for easy customization to support local requirements using a powerful model driven programming methodology. A secure Data Model Repository is used to support changes in the DICOM standard and for local requirements, without upgrading to a new version of the database and without downtime.

## Conclusion

The dramatic performance and functional improvements in Oracle Database 11g make the two essential elements for better information management: the ability to manage, secure, query, and administer information with the highest levels of performance, and the ability to derive understanding and knowledge in an open, standard manner from data which had previously been dependent upon proprietary application or device logic. Over a decade of development, research, and close collaboration with customers and application providers have resulted in these unique capabilities found only in Oracle Database 11g.



Information Management with Oracle Database 11g  
Release 2

February 2010

Author: James Steiner

Contributing Authors: Melliyal Annamalai, Mark  
Drake, Roger Ford, Jean Ihm, Kevin Jernigan,  
Xavier Lopez, MaryBeth Pierantoni, Jinyu Wang  
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

oracle.com



| Oracle is committed to developing practices and products that help protect the environment

Copyright © 2009, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

0110