

## Managing Very Large GeoRaster Databases with Oracle10g

Jim Farley, Principal Product Manager  
and  
Jeffrey Xie, Project Lead  
Oracle Spatial,  
Oracle New England Development Center  
One Oracle Drive,  
Nashua, NH 03062, U.S.A.  
{jim.farley, qingyun.xie}@oracle.com

### 1 Introduction

GeoRaster is a feature of Oracle Spatial that lets you store, index, query, analyze, and deliver GeoRaster data, that is, image and gridded raster data and its associated metadata. GeoRaster provides Oracle spatial data types and an object-relational schema. You can use these data types and schema objects to store multidimensional gridded data and raster layers that can be referenced to positions on the Earth's surface or in a local coordinate system. If the data is georeferenced, you can find the location on Earth for a cell in a raster; or given a location on Earth, you can find the cell in a raster layer associated with that location.

GeoRaster is designed to deliver enterprise-class data management capability to large image processing solutions. It is now possible for developers to integrate this powerful data management technology with the leading image processing and raster/grid analysis tools.

GeoRaster is designed to meet the general needs of broad application groups including:

- Traditional GIS and remote sensing applications – users manage their geographic raster and gridded data assets using a scaleable, secure, and robust RDBMS for defense, intelligence, agriculture, natural resource management.
- Business applications – leverage raster-based data in conjunction with other basic location data (address, etc.) to inventory and evaluate site locations and to track fixed and/or continuous assets. Opportunities here include Asset Management and Facilities Management particularly in energy and utilities.
- Image and Gridded Raster Data Repositories/Clearinghouses – support for clearinghouse servers that need to ingest, store, and disseminate very large volumes of geoimagery.

With GeoRaster, it is now possible to manage both georeferenced raster and spatial vector data in a single Oracle database. This means that the same storage, spatial referencing, indexing, and spatial operators can be used to store, query, and retrieve raster and vector data. Moreover, all of Oracle tools and utilities now support the management of raster data and related datasets in the data server. For the customer, this means that the cost of building and fielding applications that use this data will be reduced while the security, scalability and reliability of the application environment will improve dramatically.

## 2 *Benefits of Managing Raster Data in Oracle 10g*

By effectively managing raster, vector and attribute data in a single data management environment – with common storage, indexing, spatial referencing, query optimization, security, and user management – Oracle reduces the processing overhead and eliminates the complexity of coordinating and synchronizing disparate sets of spatial data. Other benefits include:

- Raster, gridded, vector, XML and various types of attribute data can be stored on single server.
- Seamless geographic coverage – no need for tiling.
- Better management of spatial data – SQL access.
- Support by leading 3<sup>rd</sup> party image processing, GIS and visualization tools.
- Consolidation of disparate data management environments (GIS, Remote sensing, business data).
- Greatly reduced complexity of systems management and better use of existing resources.
- Support for standard georaster and spatial types – proprietary data structures are avoided.
- Scalability, data security, replication, partitioning, bulk load utilities.
- Breaking size barriers – support for tens of terabytes of raster data.
- Internet deployment – Enables large number of users to access application at virtually no additional cost with 24 x 365 uptime.
- Support for both short and long transactions (version management).
- Open GIS conformance certification.
- Reduced training, software, support and application integration costs resulting from consolidating raster, vector and attribute storage.
- Risk reduction – Georaster information is integrated into Oracle Database 10g, resulting in scaleable, reliable, and fast performance.
- Market validation– In four separate surveys since 1999, IDC has found that Oracle holds 80-90% of the overall geospatial data management market.

## 3 *Geographic Raster Data Management*

The use of raster data has evolved from classified applications in defense and intelligence over the past fifty years to become a common tool in planning, security and surveillance, business intelligence, agriculture and in a wide range

of transportation and natural resource applications. In addition to these traditional consumers of raster data, the insurance industry, entertainment and media and real estate are increasing users of raster data.

There are two basic kinds of raster data types supported in Oracle Spatial 10g GeoRaster: grid-based data and raster data.

- Grid-based data or gridded data is a general term used for raster data. A rectangular grid of cells that are aligned to the X and Y-axes overlying an area. Each cell in the grid is the same size; this size is the *resolution* of the grid. Given the dimensions of the raster data, and the number of rows/columns, the cell size can be determined. Grid data typically has associated attribute values for each cell in the grid (measuring elevation, frequency, concentration, etc.). Given information about the bounding coordinates of the grid, the location of each cell can be calculated. Therefore, no explicit location value is needed for each cell. Each cell can store one or more attribute values. Examples of attributes stored for each cell include digital terrain elevation data, land use information, pollution concentration, land cover information, geological information, rainfall information, and many others.
- Digital imagery is a specialized type of raster data. It is a two dimensional array (a matrix or grid) of regularly spaced picture elements (pixels). Typically, digital images do not require a separate table to hold value/attribute information. An image is created from optical or other sensor data, and is collected using a variety of technologies including satellite remote sensing, airborne photogrammetry, and sonar. The smallest meaningful pixel in the image is determined by the optics of the camera lens or sensor that captured the image. The size of the pixel is referred to as the resolution of the image. The higher the resolution, the smaller the pixel, and the better the quality of the image to the human eye. Digital images can be composed of one or more bands. Each band often represents an interval of wavelengths along the electromagnetic spectrum. Multiple bands of an image can be simultaneously recorded, as is the case with the popular LANDSAT 7 band platform. Imaging platforms that capture data in a series of bands are referred to as multispectral and hyperspectral.

One key feature associated with remotely sensed raster data is the frequency with which the data is acquired. Typically a satellite will orbit the earth several times a day. This means that the satellite will photograph the same portion of the earth at regular, predictable intervals making it the most cost-effective way to digitally track changes on the ground. Raster datasets tend to be large, even by GIS standards and the lack of effective lossless compression technology (compression schemes that *retain all of the information in the original pixels*) makes compression unrealistic for many applications. A single LANDSAT

multispectral image is more than 250 megabytes. The base size of individual images plus the fact that many applications look for changing patterns on the ground requiring multiple images of the same area as a time series for comparison ensure that applications will accumulate data at a rapid rate. As an example of the rate of accumulation consider the IKONOS platform from Space Imaging. Once every 98 minutes, 14 times a day, Space Imaging's IKONOS® satellite circles the globe collecting images of the Earth. IKONOS has produced over 100 million square kilometers of customer-ready imagery alone since it achieved orbit. Clearly, image-based sensor devices and applications consume large quantities of data and have significant storage requirements. The Oracle Spatial 10g GeoRaster feature is specifically engineered to meet these needs.

## 4 The Application of GeoRaster

There is a wide range of use cases for GeoRaster. The section below highlights use cases for deploying the Oracle Spatial 10g GeoRaster feature with 3<sup>rd</sup> party analysis and visualization tools.

### 4.1 Defense and Security

- Problem – To detect threats to security and to ascertain threat levels while managing an appropriate response matrix both friendly and unfriendly assets need to be tracked.
- Context – Detecting changes over time on the ground is a prerequisite to effective planning and response.
- Solution – Aerial and satellite remote sensing platforms are tasked to generate images from an area of interest on a regular basis. Images are collected, georeferenced, and loaded into a database as raster or gridded layers. Client tools are used to examine current images in association with historic images for the same surface coverage archived in the database. The current raster is mosaicked with surrounding rasters to create seamless coverage of area of interest.
- Critical Fact – Image data is the only geographic information that can be acquired in a scheduled/timely manner (tasked) for a specific local providing rapid access to current, accurate geodata.

### 4.2 Emergency Response

- Problem – Assess on-the-ground damage and develop suitable response scenario given damage to critical infrastructure.
- Context – Making a rapid assessment of damage to infrastructure in the aftermath of a disaster event is critical to a timely response and mitigation.
- Solution – Raster data from aerial and satellite remote sensing platforms are used to compare “*before and after*” conditions on the ground. GeoRaster data is used in conjunction with associated vector data for continuous asset infrastructure (e.g., road, rail, power grid, gas, telco) to determine a) damage b) response scenario c) viable corridors for first responders and d) routing to appropriate facilities.

- Critical Fact – Only raster data can provide the near real-time data acquisition needed to accomplish damage assessment and meet the time critical requirements of first responders. A raster data layer provides an ideal backdrop to display infrastructure data (e.g., pipelines, transmission lines) in a readily comprehensible form.

### 4.3 Enterprise Asset Management

- Problem – Optimize preventative maintenance, field service and operations across a network of continuous and fixed assets (e.g., stations, substations and pipeline).
- Context – Accomplish ongoing monitoring and management across a variety of assets to support normal preventative maintenance and operations.
- Solution – Incorporate raster data from aerial photography into the existing geospatial data used to map assets under one single enterprise database management system. Raster data used in conjunction with vector information representing property boundaries, lease zones and easements enable field service personnel to save time and wear and tear on equipment, zeroing in on problem areas on the ground.
- Critical Fact – Raster data increases the efficiency of resources in the field and drives down costs making operations more efficient.

### 4.4 State and Local Government

- Problem – State and local government have limited resources to manage zoning, tax assessment, etc.
- Context – State and local government budgets are stretched like no time in recent memory. With tax base often shrinking and public service expenses increasing, local officials are compelled to improve efficiency and to maximize coordination and communication between local departments in government.
- Solution – a base map of digital raster data (digital orthophoto quads or contracted photogrammetric product) when stored in the enterprise database in state government provides a common frame of reference that can be used across multiple departments (Transportation, Tax Assessment, Zoning, School Administration, etc.) to support decision making and the formulation and implementation of policy.
- Critical Fact – Raster data provides a common frame of reference that can be used across departments and state agencies.

### 4.5 Agriculture Monitoring

- Problem – National agricultural agencies need to document and verify agricultural utilization of the land.
- Context – Farmers often are required to report the land use practices they employ and the crops they seed allowing officials to project earnings

based on yield. In addition, it is common under some circumstances to compensate a farmer for land that is left fallow or unseeded.

- Solution – Use of aerial image data acquired on an annual basis in conjunction with vector information from the land management agency enable officials and local agriculturalists to create an accurate record of acres in crop and acres fallow.
- Critical Fact – Only image-based information acquired on an annual basis can provide the baseline information needed to make the assessments needed to support equitable taxation and remuneration.

## 5 Working with Raster Data

Raster data is different from the vector data type traditionally associated with Oracle Spatial technologies (Oracle Locator and Oracle Spatial). These differences stem from the structure of the data, the ways in which the data is captured, and the ways in which the data is stored and used to achieve optimum benefit.

Digital raster data is highly structured and comprised of specific elements that are readily managed in GeoRaster.

- Cells or pixels hold the actual raster data measured either as an absolute value (elevation above mean sea level, concentration in part per billion, saturation per cubic meter of soil matrix) or as a spectral signature of some sort measuring reflectance across the visible, near visible and invisible portions of the spectrum.
- Spatial domain (footprint), reflects the on-the-ground extent of the raster data on the surface of the earth. This is typically represented as a spatial bounding box or envelope for the raster data.
- Spatial, temporal, and spectral information is the What, Where and How for any given digital image or satellite scene. This is a form of metadata that documents the characteristics of the raster data including: imaging platform, frequency and schedule for imaging specific portions of the earth, number of bands or channels of data, attributes specific to a given image, and documentation describing the manner in which the digital image data is *packed* to form bands.

In addition to these key data elements there is often a range of descriptive data associated with the pixels/cells (attributes), some ancillary metadata, and information describing processing.

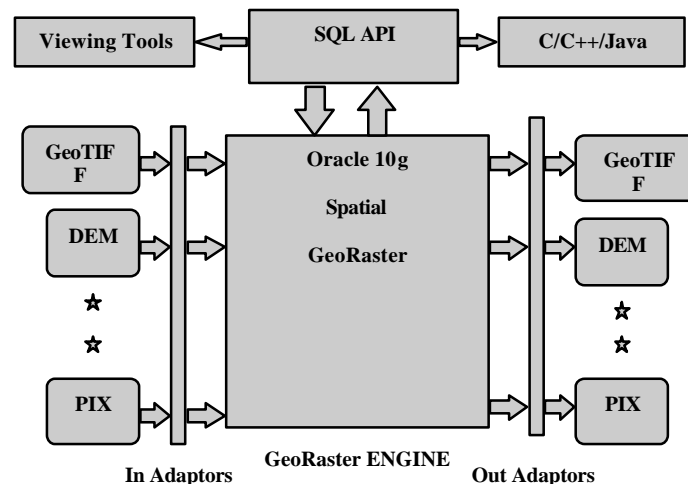
Prior to Oracle 10g most location-based data handled by Oracle Spatial was vector-based (point, line, polygon). As a result, some of the concepts and processes associated managing raster and/or grid-based data may be new. Important concepts germane to the handling of raster data include:

- Georeferencing – establishes the relationship between cell coordinates of GeoRaster data and real-world ground coordinates (or some local coordinates). Georeferencing assigns ground coordinates to cell coordinates, and vice versa.
- Pyramids – used to accelerate data retrieval by storing a series of “generalized” representations of the original image at a range of resolutions. Pyramids typically range from the original data to fairly generalized representations in which many of the original pixels have been resampled or “thinned” to reduce the size of the image while maintaining as much of the original image composition as possible. If an application requires the exact image, the element of the pyramid that contains the original data at full resolution would be returned. If an application required a less detailed representation of the image than the original, an alternate element of the pyramid (containing much less original data) would be returned. Pyramids are used to enhance performance and to minimize the throughput impacts resulting from constrained bandwidth.
- Blocking/Tiling – the process of storing raster data in blocks (on disk) for optimized storage, retrieval and processing.
- Resampling – the process of extrapolating new cell values from neighboring source cell values, that may have a different resolution and/or may be in a different coordinate system. Resampling is used in image visualization, raster data re-projection, rectification and generalization.
- Generalization – a resampling process that results in an image containing less information and when closely examined, has fewer pixels. The purpose of generalization is to achieve maximum reduction in image data volume while minimizing impact on image quality and degradation to individual features in the image.
- Interleaving – the way in which multi-channel (multi-band) data is captured by a sensor and stored. Interleaving may be band sequential (BSQ) in which all the information for each band is stored as a contiguous unit on disk (e.g., all of band 1 followed by all of band 2, etc.). Interleaving can also be band interleaved by line (BIL) and band interleaved by pixel (BIP). The selection of ideal interleaving approach is determined by the analysis to be performed on the raster data.

## 6 Architecture

The GeoRaster architecture provides the core functionality needed to support the use of raster data in the Oracle 10g database. At a very high level of abstraction the Oracle 10g GeoRaster architecture includes six basic components:

1. GeoRaster Engine – Core GeoRaster functionality includes data, metadata, methods and indexing.
2. SQL API – SQL access to the raster data in GeoRaster.
3. C/C++/Java – OCI, OCCl, and Java access to the raster data in GeoRaster with or without calling the GeoRaster API
4. Viewing Tools: A variety of third party viewing and analysis tools now support GeoRaster. In addition, a free downloadable viewer is available from Oracle.
5. Input [data] adapters – Facilitate loading raster data from well-known file formats into GeoRaster.
6. Output [data] adapters – Facilitate unloading raster data into well-known file formats from GeoRaster.



## 7 Summary

Oracle Spatial 10g is uniquely positioned to deliver enterprise-class support for storing and managing raster data in the context of a high performance, scalable, secure environment. Oracle is the only commercial provider that takes this valuable raster-based spatial data out of the file system, where it exists in an insecure, transient state, storing it instead as named types in the secure environment of the world's leading database management system.