Building an Enterprise Geodatabase Raster Data Management System

Gillian Silvertand – ESRI Educational Services
Qian Liu – ESRI Software Products

Related Instructor-led courses:

- **ArcSDE Administration for <DBMS>**
- **Building Geodatabase 1**

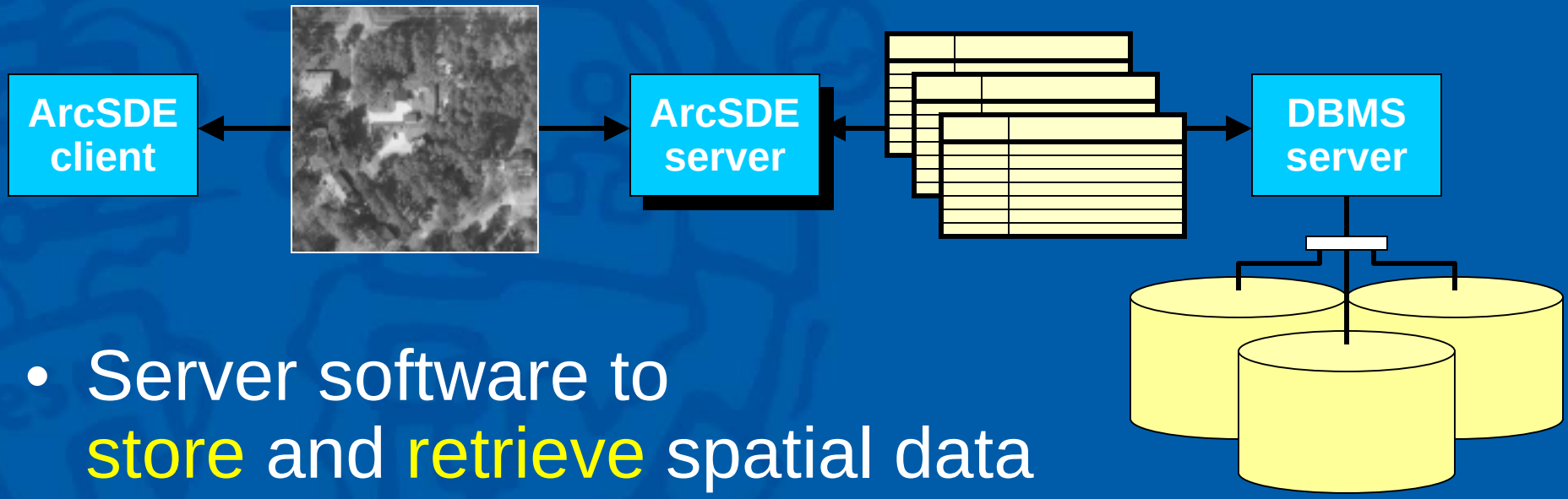
Presentation goals and audience

- Goals:
  - Understand raster data storage and properties
  - Understand raster data models
- Audience:
  - DBAs
  - GIS Administrators
  - Data providers
  - Managers

Presentation outline

- Address the role of ArcSDE in raster data management
  - Alternative to files
- Detail the architecture of ArcSDE rasters
  - Database schema
  - Storage properties
- Working with geodatabase raster data
  - Loading options
  - Loading recommendations

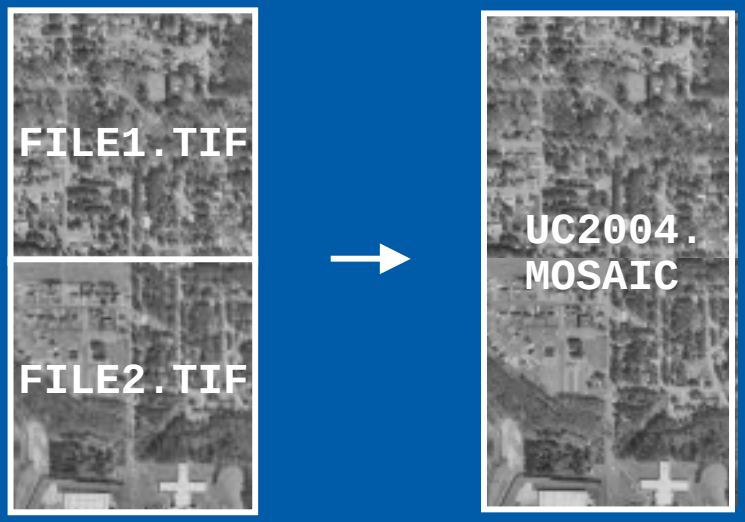
ArcSDE

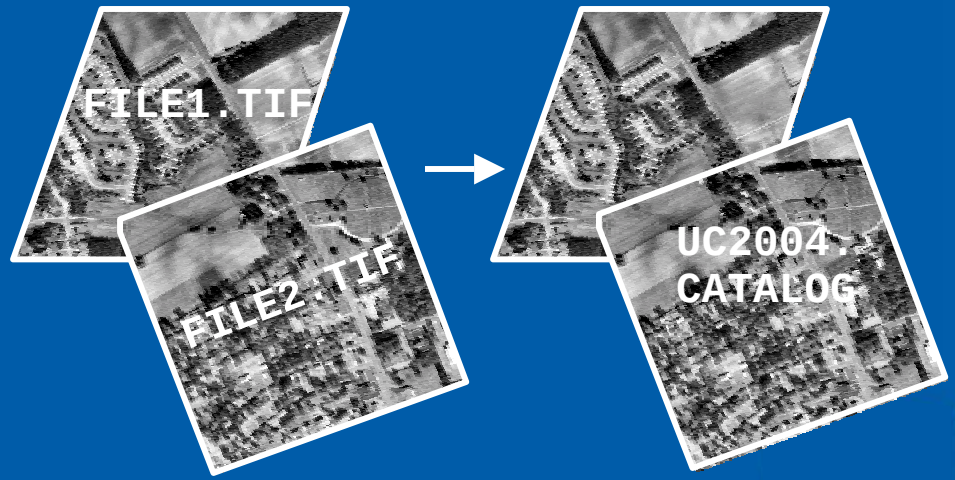


- Server software to store and retrieve spatial data
  - Stored in underlying, industry-standard database management system (DBMS)
  - For GIS clients
  - Part of ArcGIS software family
  - Supports raster and vector data

ArcSDE offers: Support for large data

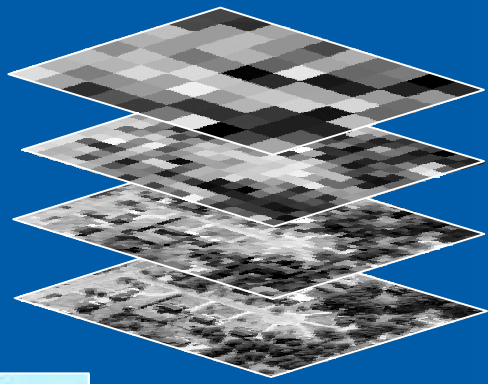
- Avoids manually tiling files
  - Raster size is virtually unlimited
  - Exploits DBMS indexing and partitioning
- Raster dataset
  - Append multiple files together
- Store collections in raster catalogs
  - Related rasters are separate rows in same tables





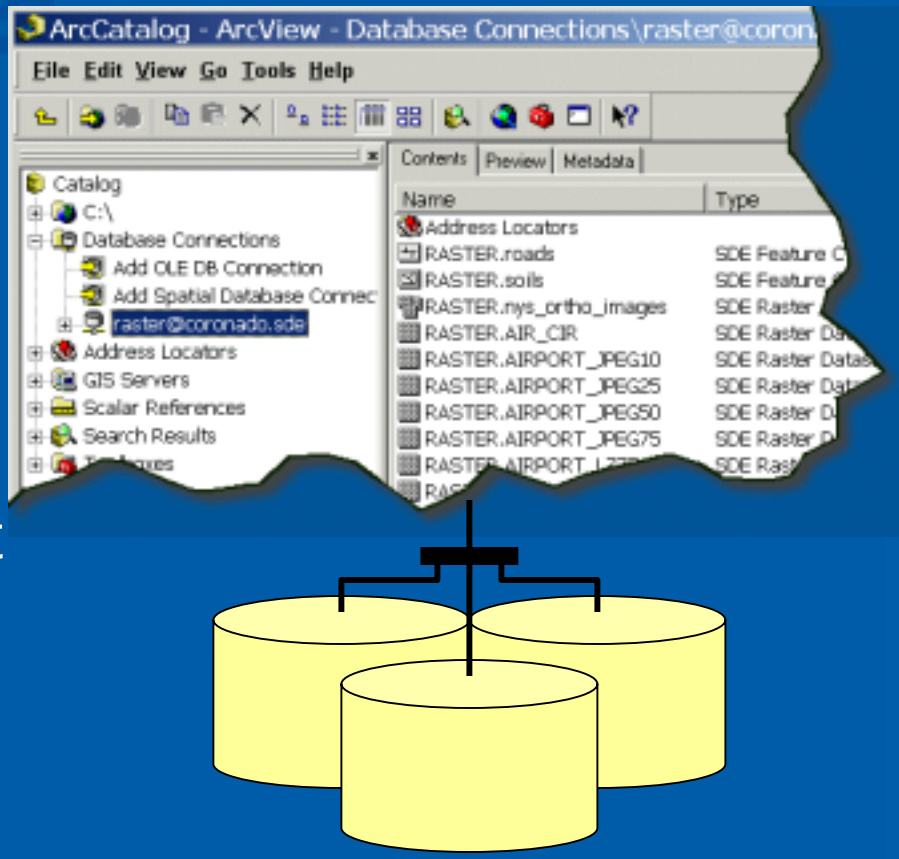
ArcSDE offers: High-performance queries

- Groups pixels into raster tiles
  - Automatic, internal tiling structure
- Pre-computes pyramids
  - Reduced-resolution copies of source data
- Supports three types of compression
  - Loss-less LZ77
  - Lossy JPEG
  - Lossy JPEG 2000



ArcSDE offers: Centralized geodatabase management

- Single storage location
  - Store rasters with related vector data
  - Accessible through same client connection
- Common storage format
  - Converted to ArcSDE format
    - Pixel values are not changed
- Robust data protection
  - DBMS authentication and privileges
  - DBMS reliability / recoverability features



Alternative option

New at 9

- Personal geodatabase raster storage
- Stores data as files
  - Managed
    - Stored in folder at same directory level as the personal GDB
    - Raster data is converted to IMG format
    - Data over 2GB will use IMG and IGE files for storage
  - Not managed
    - Raster data is referenced through **NAME** field
    - No loading of raster data
    - Pathname is used to find a raster file

Presentation goals and outline

- Address the role of ArcSDE in raster data management
  - Alternative to files
- Detail the architecture of ArcSDE rasters
  - Database schema
  - Storage properties
- Working with geodatabase raster data
  - Loading options
  - Loading recommendations

ArcSDE raster data

- Raster data stored as raster datasets
  - Can be grouped into a raster catalog
- Geodatabase container for raster data
  - Stores all pixel values
- Logically one table with spatial and raster column
  - Physically stored as many tables

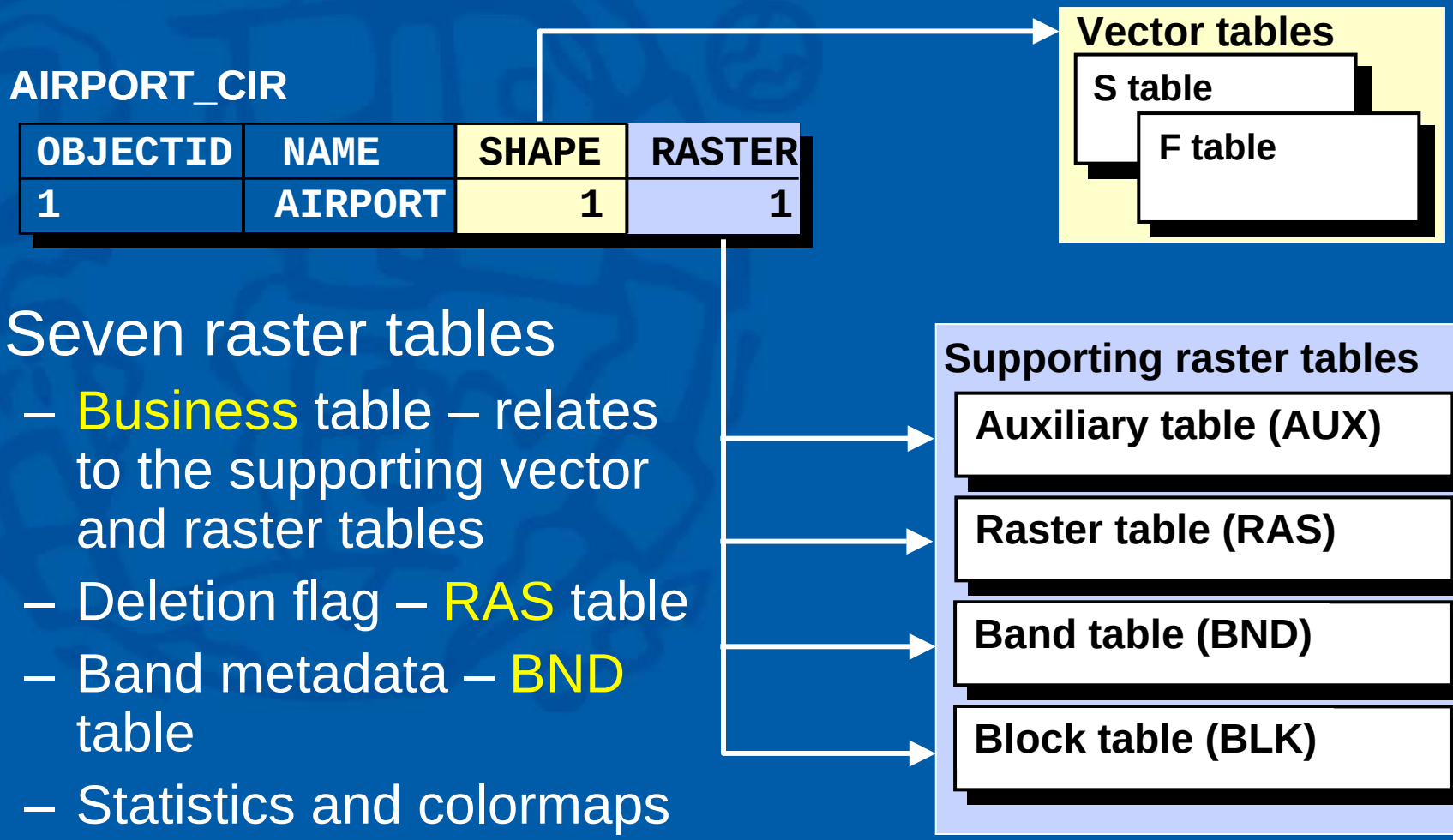
End user experience

UC2004.NYS (business table)

RASTER	NAME	SHAPE
	e_07411428	Polygon
	e_07381430	Polygon
	e_0734126	Polygon
	e_07351428	Polygon

- A raster dataset is a row in a table
  - Spatially-enabled table is business table
  - Raster column stores link to pixel data
  - Additional columns store user-defined attributes
- Performance is good and consistent at multiple scales

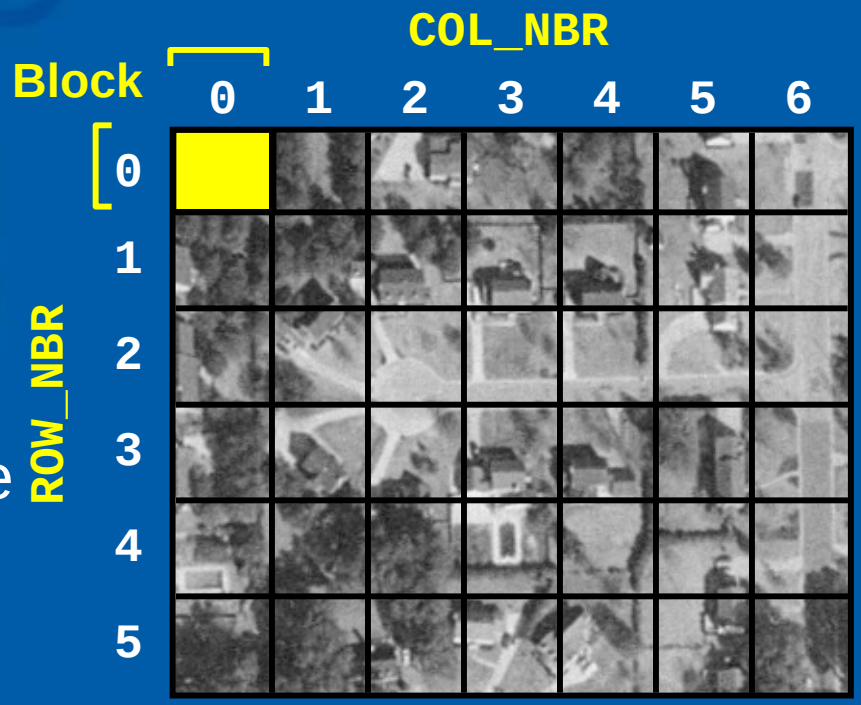
Raster dataset schema overview

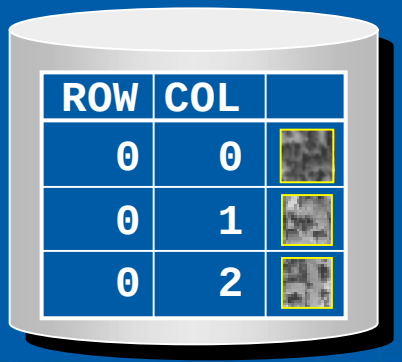


- Seven raster tables
  - Business table – relates to the supporting vector and raster tables
  - Deletion flag – RAS table
  - Band metadata – BND table
  - Statistics and colormaps – AUX table
  - Pixel data – BLK table

Raster properties: Raster tiles

- ArcSDE subdivides raster into tiles for storage
  - Automatic
  - Required
  - Invisible to end users
- Access to data in small pieces
  - For fast display performance
- Tile IDs are indexed
- Tile size defined by user
  - Defined at creation time
  - ESRI recommends 128x128 pixels





Compression

- Reduces storage size
- Improves performance
  - Less disk I/O
  - Less network I/O
- Set at creation time
- Settings:

LZ77 — **No user defined settings**

JPEG 6 — 94

JPEG 2000 1 — 255

User defined quality setting

High compression
***very lossy**

Low compression
***not very lossy**

Compression options

	Loss-less	Lossy
Type:	▪ LZ77	▪ JPEG (quality 6 – 94) ▪ JPEG 2000 (quality 1- 255)
Applies to:	▪ All data types	▪ JPEG – 8 bit unsigned ▪ JP2 – 8 + 16 bit unsigned
Amount of compression:	▪ Moderate ▪ Higher for more homogenous data	▪ High ▪ User controls compression / quality ratio
Best for:	▪ Data analyzed by computer ▪ Data with sharp edges	▪ Data analyzed visually ▪ Smooth, continuous images
Examples:	▪ Analytical surfaces ▪ Analytical satellite / aerial images ▪ Scanned linework and colormapped images	▪ Digital photos of assets ▪ Imagery for basemapping cartographic output ▪ Scanned true-color maps

Compression recommendations

- Load uncompressed or loss-less compressed source data, if available
  - ArcSDE uncompresses / re-compresses data in its own format
  - Reduces excessive image degradation
- When loading lossy-compressed files:
  - Load and view sample to verify visual quality
- General
  - Optional, but always recommended
  - Lossy compression may affect scale of use
  - Must re-create raster to alter
  - ★ Store lossy, serve/use lossy
  - ★ Store loss-less, serve/use either way

Compression example

Source data

Pixel depth: 8-bit

Number of bands: 3

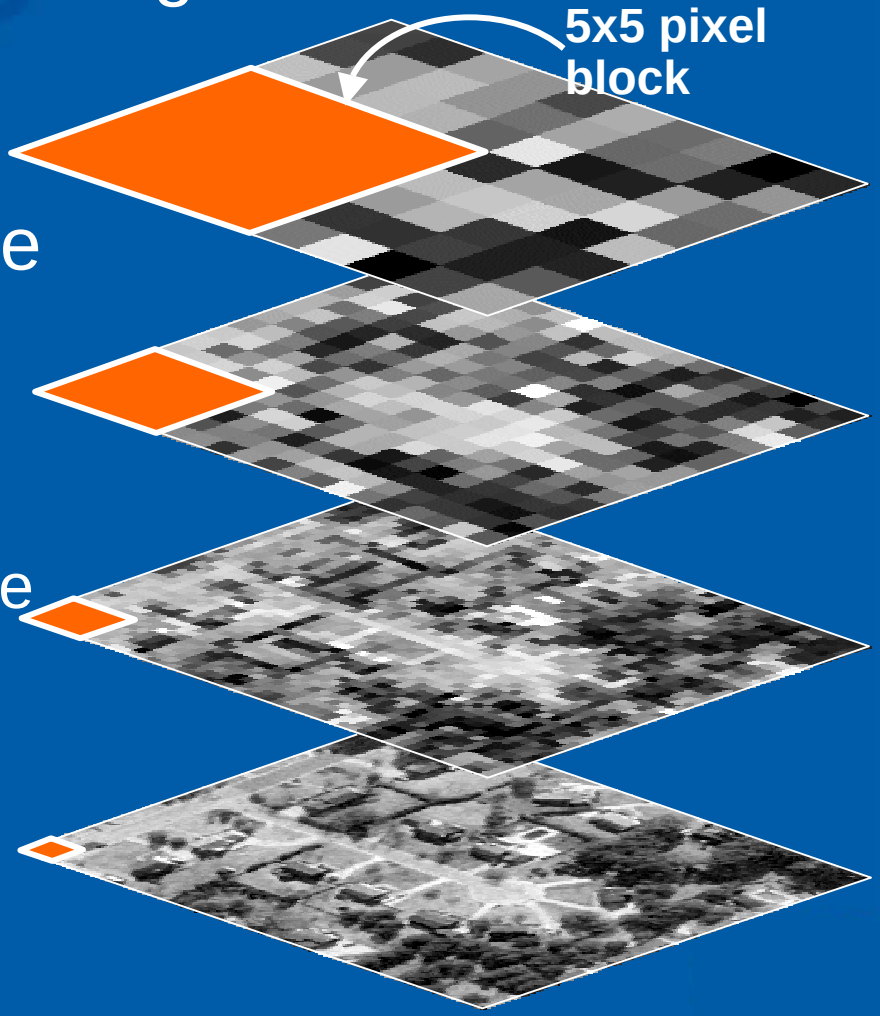
File size: 1157MB

Compression type	Loading time (seconds)	Size (MB)
None	711	2330
LZ77	836	1161.2
JPEG (25)	347	165.6
JPEG (50)	369	238.6
JPEG (75)	427	340.5
JP2 (75)	684	46.3
JP2 (100)	737	80.7
JP2 (125)	827	141.3
JP2 (150)	976	260.1
JP2 (175)	1176	455.9
JP2 (200)	1367	749.4
JP2 (225)	1475	1114.7
JP2 (250)	1529	1147.4

Without pyramids

Raster properties: Pyramids

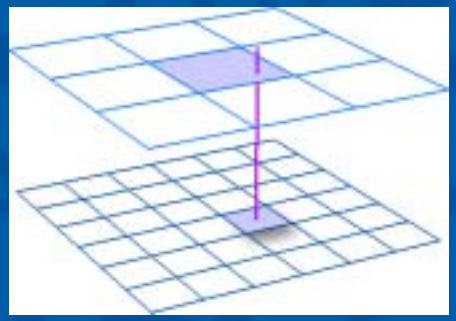
- Reduced resolution copies of original raster
- Stored re-sampling vs. display re-sampling
- Improves query performance
  - Clients only request required resolution
- Set at creation time
  - Adds about 33% more storage
- Pyramids are not used during analysis
- When built with load
  - Uses same compression



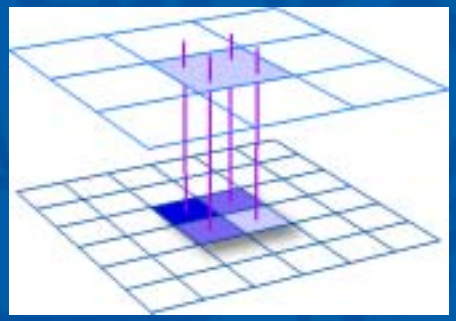
Pyramid resampling methods

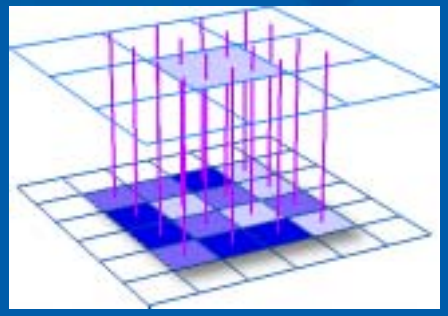
- Interpolates pyramid values from original raster values

For discrete data (e.g. colormap)



For continuous data (does not support colormap)



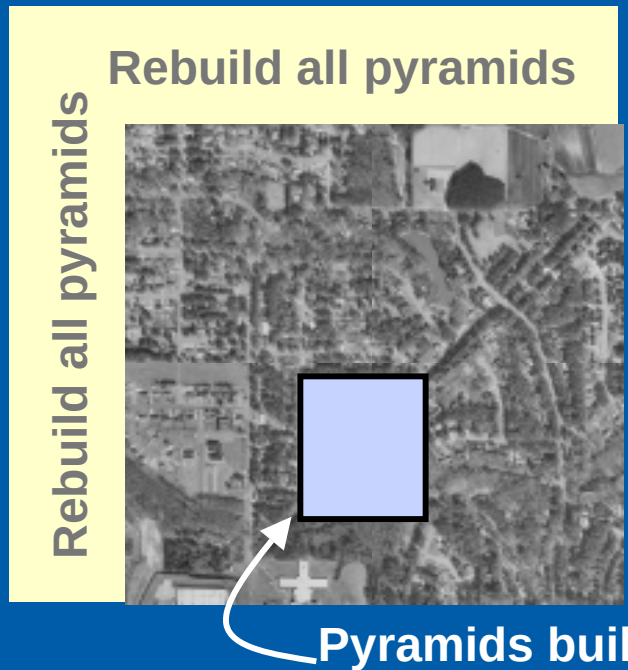


- **Nearest neighbor**
  - **Value of nearest pixel**
  - **Does not create new values**
  - **Fastest overall**
  - **Must use with colormap**
- **Bilinear interpolation**
  - **Weighted average of four nearest pixels**
  - **Fastest creation**
- **Cubic convolution**
  - **Smooth curve fitted to sixteen nearest pixels**
  - **Typically produces sharper output**

Pyramid enhancements

New at 9

- ArcSDE 8.3
  - Mosaicking images recreates pyramids for entire raster dataset
  - During pyramid creation, users have no access
- ArcSDE 9
  - Partial pyramids created for area that is loaded / updated
  - Use data while loading / updating other areas
  - Must specify same re-sampling method or all pyramids will be re-created
  - Set offset origin coordinate
    - Use to avoid rebuilding pyramids for entire raster dataset



Pyramid example

Source data

Pixel depth: 8-bit

Number of bands: 3

File size: 1157MB

Compression type	Loading time (seconds)	Size MB (no pyramids)	Size MB (pyramids)
None	2216	2330	2123.1
LZ77	1759	1161.2	1563.7
JPEG (25)	654	165.6	238.3
JPEG (50)	808	238.6	345.8
JPEG (75)	916	340.5	505.2
JP2 (75)	1290	46.3	66.9
JP2 (100)	1476	80.7	126.7
JP2 (125)	1720	141.3	234.3
JP2 (150)	2118	260.1	421.6
JP2 (175)	2618	455.9	722.4
JP2 (200)	3147	749.4	1131.1
JP2 (225)	2656	1114.7	1514.7
JP2 (250)	3766	1147.4	1546.7

Raster statistics

Unstretched:

Min: 24

Max: 241

Stretched:

Min: 0

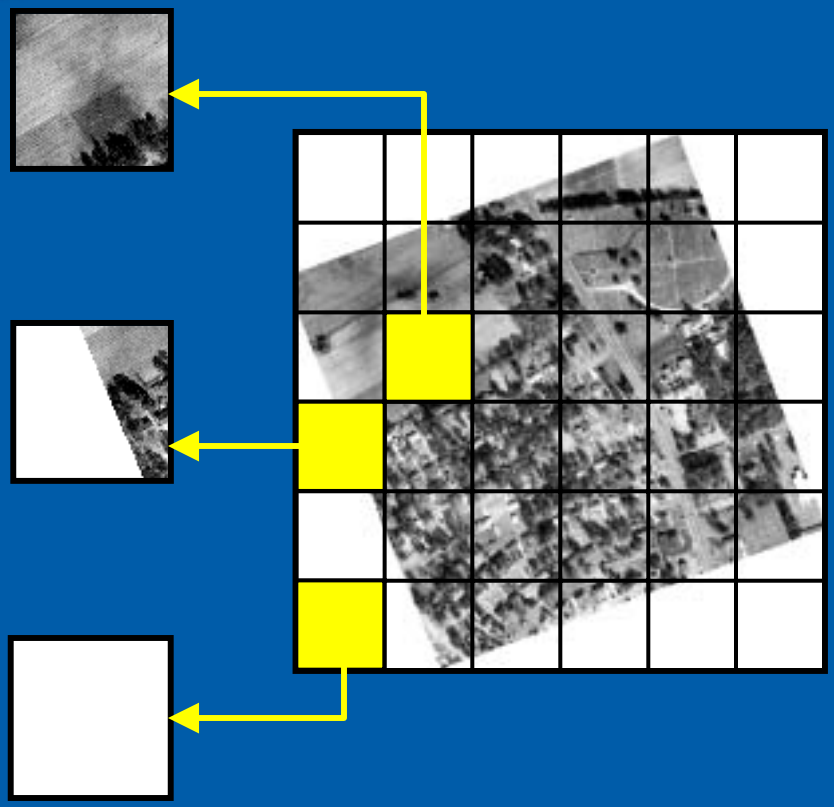
Max: 255

Unstretched **Stretched**

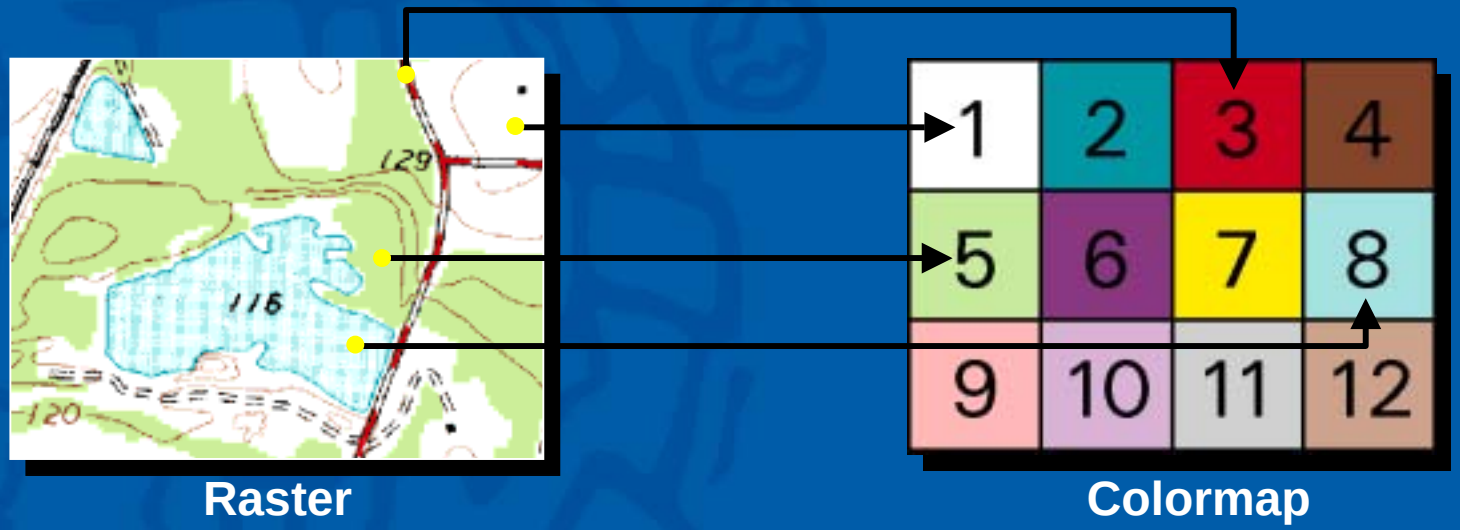
- Basic description of pixel value distribution
- New at 9: Build for defined pyramid level
  - Faster to create
    - Check accuracy for min/max values
- Used by clients for display
  - Apply stretch for continuous data
  - Default rendering depends on statistics and format
  - Data may look different after loading, but values are the same

NODATA storage

- Pixels with unknown values
- Stored as bitmask
  - One bit-per-pixel
  - Only stored if block contains NODATA
- Convert data to NODATA with ArcToolbox



Colormaps



- "Table" linking pixel values to colors
  - For discrete data
  - Manage colormaps with ArcCatalog
- Properties:
  - Nearest neighbor
  - LZ77

```
1       255,255,255
2       17,110,173
3       233,37,37
4       116,58,0
. . .
```

Presentation goals and outline

- Address the role of ArcSDE in raster data management
  - Alternative to files
- Detail the architecture of ArcSDE rasters
  - Storage properties
  - Database schema
- Working with geodatabase raster data
  - Loading options
    - Raster dataset
    - Raster catalog
    - Raster attribute
  - Loading recommendations

Loading Options

1) 

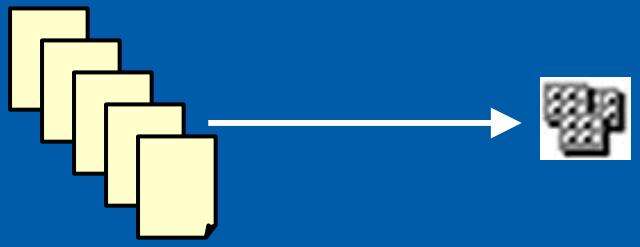
Mosaic images into a raster dataset

2) 

Multiple raster datasets

3) 

Raster catalog with multiple raster datasets

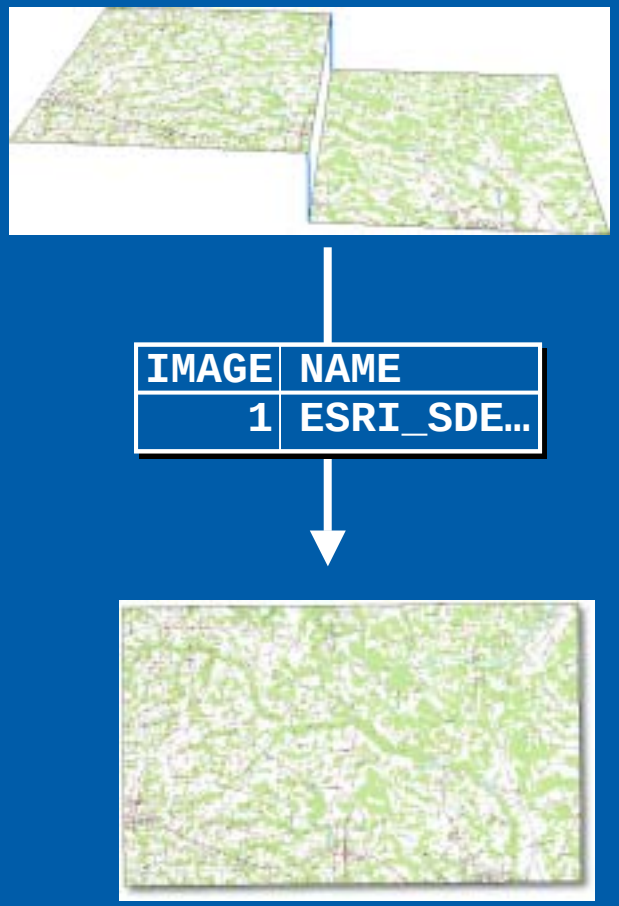
4) 

Raster catalogs with mosaicked raster datasets

Raster datasets

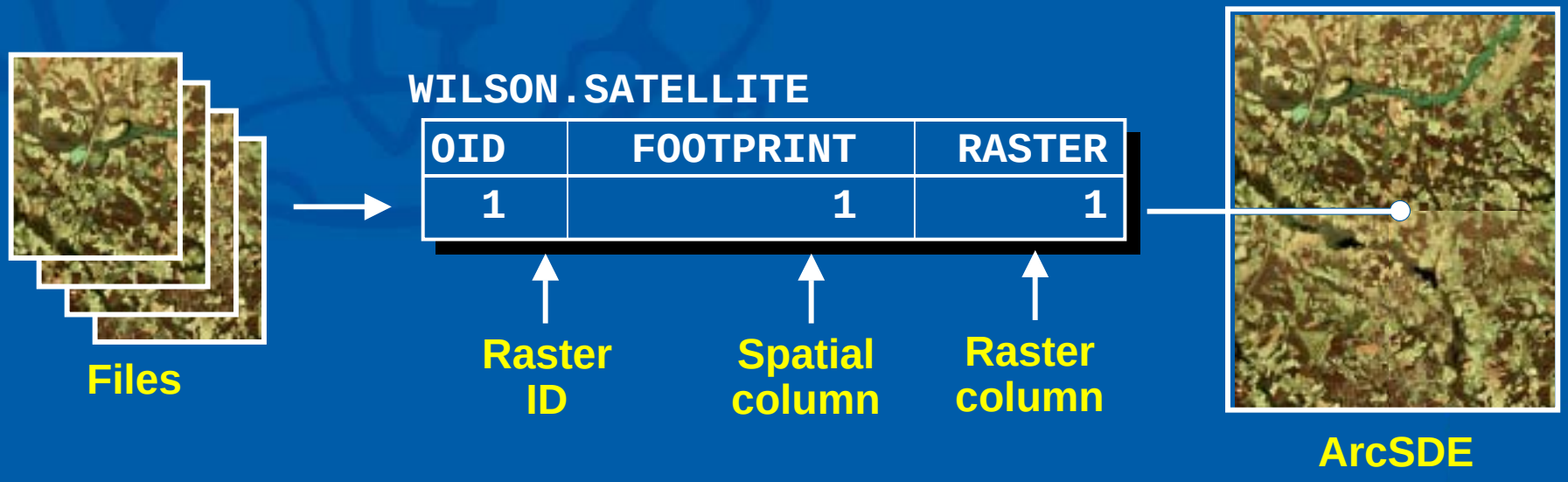
- Container of pixels
  - One or more images
  - One or more bands
- Seamless
- Fast display at any scale
- Can mosaic multiple inputs together
  - Append into one image in the geodatabase
  - Partial pyramid creation

New at 9



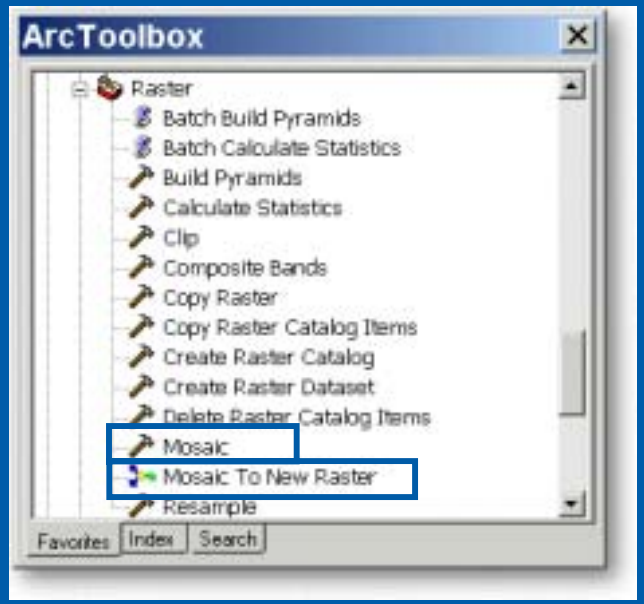
Mosaicking

- Mosaic multiple files into one raster dataset
  - Many files → one row in business table
  - Data properties
    - Inputs will be reprojected and resampled if necessary
    - Inputs are aligned, or can be shifted during loading
  - Seamless query access



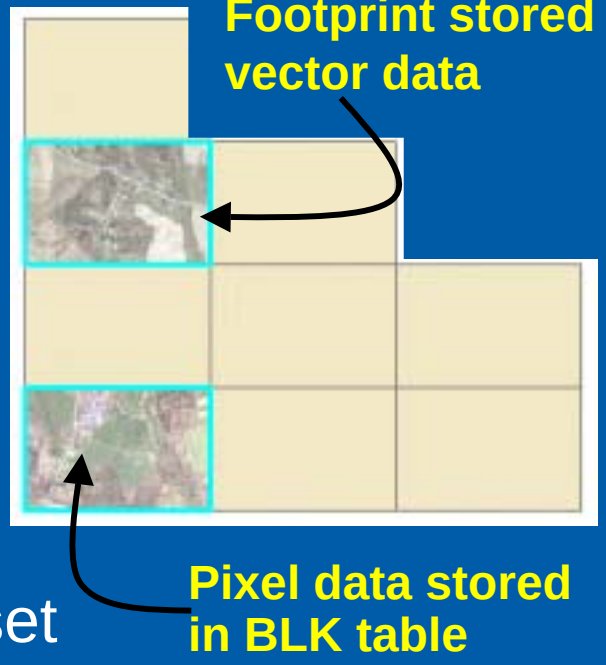
Mosaicking options / requirements

- Data
  - Exact pixel alignment
    - ArcCatalog resamples as needed
  - Inputs have same number of bands / pixel depth
  - Inputs will use same projection / coordinate system
    - Images will be reprojected to match
- ArcCatalog manages overlap:
  - First, last, blend, mean, min, max
- ArcCatalog manages colormap:
  - First, last, reject, match

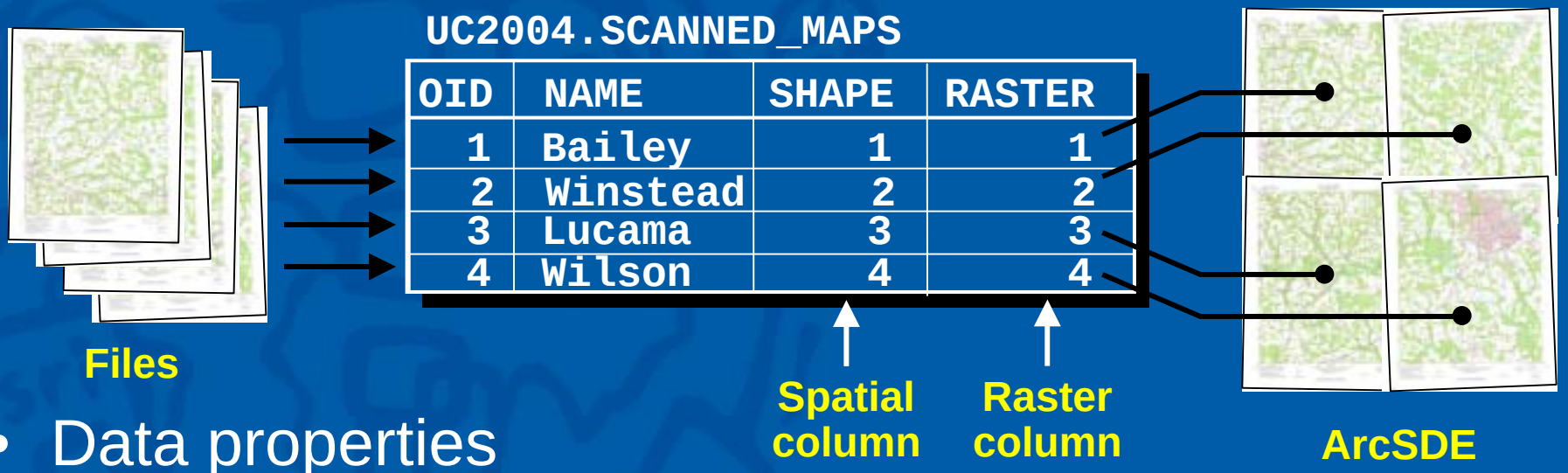


Raster catalogs

- Collection of raster datasets
  - Each raster dataset has own properties
    - e.g., Colormap, pyramids,pixel size…
  - If projection defined all will share the spatial reference
- Storage properties:
  - Multiple rows in the business table
  - One envelope for each raster dataset
- Footprint used for searching and wire frames
  - Dynamic (8.3)
  - Stored (9.0)
- All pixel data stored in the underlying tables



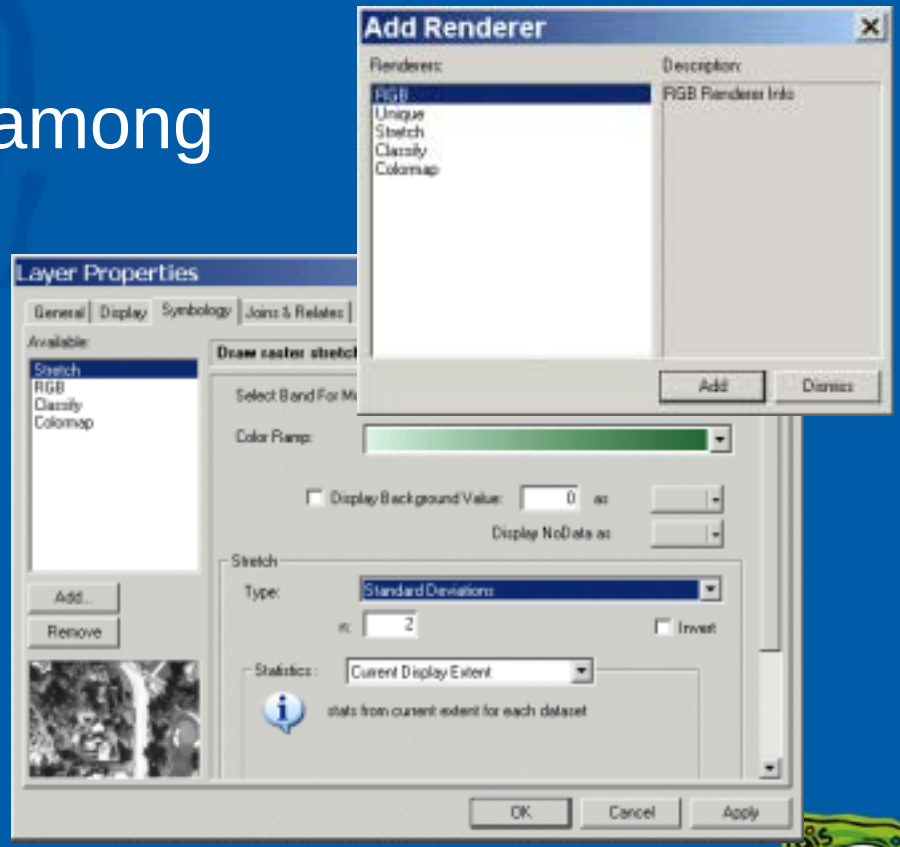
When to use raster catalogs



- Data properties
  - Overlapping inputs
  - Not aligned and cannot be shifted
- Query access
  - By individual raster dataset
  - By user-defined attribute (date, percent cloud cover, etc.)
- Can add user defined attributes to business table

Rendering raster catalogs

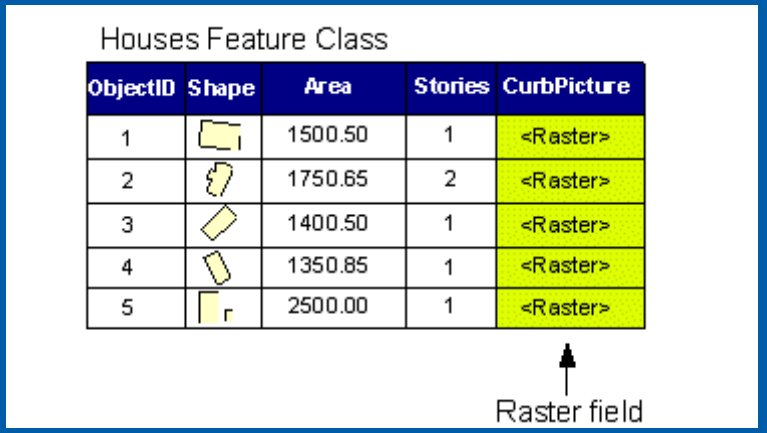
- Each raster dataset in a raster catalog can use different rendering properties
  - Add multiple renderers
- Client will choose best among available options
  - Save map document or layer file
- Display statistics
  - Current extent
  - Individual raster dataset
  - Custom



Raster datasets vs. raster catalogs

	Raster Dataset	Raster Catalog
Description:	▪ Single raster dataset for multiple inputs ** can mosaic	▪ A collection of independent raster datasets
Properties:	▪ Properties must be the same: ▪ Reprojected to same spatial reference ▪ Data type ▪ Colormap ▪ Pyramid resampling ▪ Tile size	▪ Properties can be different ▪ Will share the defined spatial reference
Impact:	▪ Only reads one set of properties	▪ Must read each set of properties so rendering can be slightly slower
Overlap:	▪ Determine how to merge overlapping pixels	▪ Overlap preserved; therefore can have overlapping data

Raster attributes



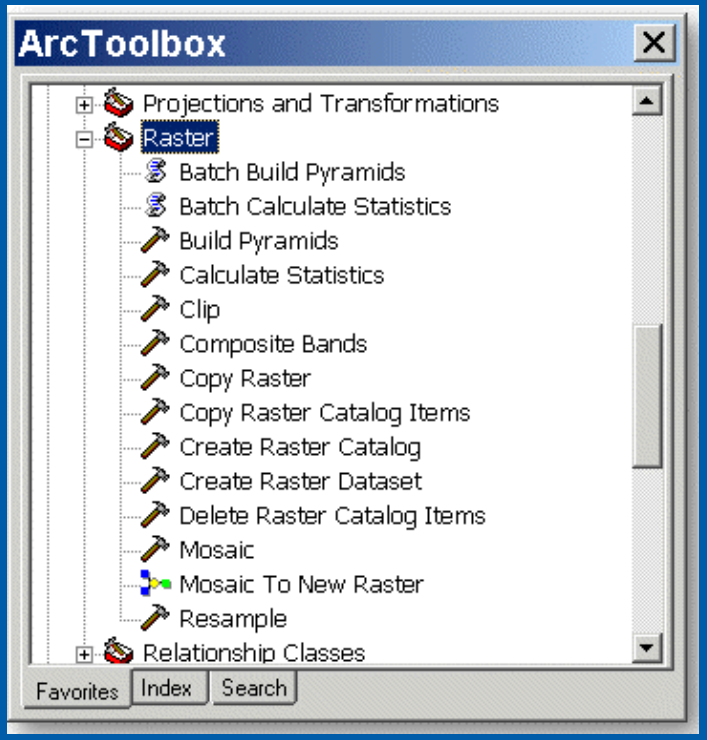
- New column type
  - `Raster` column
- Raster column on any table object
  - Raster stored with the feature class in the geodatabase
- Images may not have a spatial reference
  - Image is a property of another object
- Used to reference additional source data for rows in your table/feature class
  - Analogous to a hyperlink
- Multiple images supported through related tables

Editing raster attributes

- Register feature class as *versioned*
- Add raster data during ArcMap edit session
  - Raster dataset inserted to the supporting raster tables (BLK etc.)
- One feature can only have one raster attribute
  - Updates will replace the BLK table rows for the edited raster dataset
  - Can not preserve alternate version of raster for a feature

Loading tools

- ArcCatalog / Geoprocessing tools
  - Graphic interface
  - Loads all supported ArcGIS raster formats
  - Aligns inputs for mosaic
  - Customize loading
  - Can access through
    - Tools
    - Scripts
    - Model builder



- **`sderaster`**
  - Not geodatabase aware (more information later)
  - Utilize if no ArcGIS client available

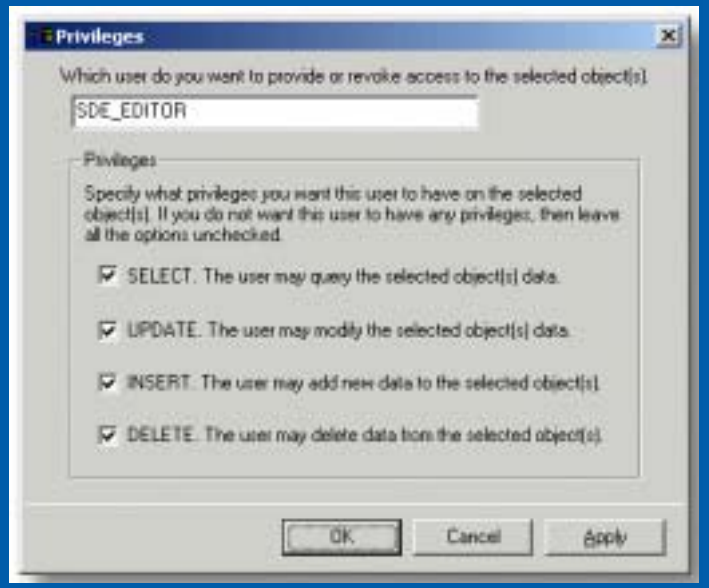
Recommendation: Use the ArcCatalog / Geoprocessing tools

Pyramid building recommendations

- 8.3 – Wait until all data is loaded
  - Changing pixels invalidates existing pyramids
  - Building pyramids is time-consuming
- 9.0 – Load top-left image first if possible or set offset origin coordinate
  - Partial pyramid building for subsequent additions
  - If mosaicking use same interpolation method
    - Different method will result in complete rebuild of pyramids
- General
  - Use computer with fastest CPU
    - Use direct connect, if appropriate

Additional loading tasks

- Pre loading tasks
  - Pre-processing
    - If necessary
  - **`sdedbtune`**
    - Manage storage parameters in DBTUNE table



- Post loading tasks
  - Manage privileges
    - `sdetable`
    - ArcCatalog
  - Calculate DBMS statistics
    - `sdetable`
    - ArcCatalog

```
sdetable -o grant
sdetable -o revoke
```

```
sdetable -o update_dbms_stats
```

DBMS and storage requirements

- Many factors influence raster size
  - Raster block size
  - Pixel depth
  - Compression
  - NODATA
  - Pyramids
  - Row overhead
- Load a sample dataset
  - Use DBMS tools to determine exact table / index sizes
  - Scale up to full data set size
  - Requires good representative sample
- Technical workshop:

  Loading Large Volumes of Raster Data Into a Geodatabase

  Wed 3:30 Room 29-A

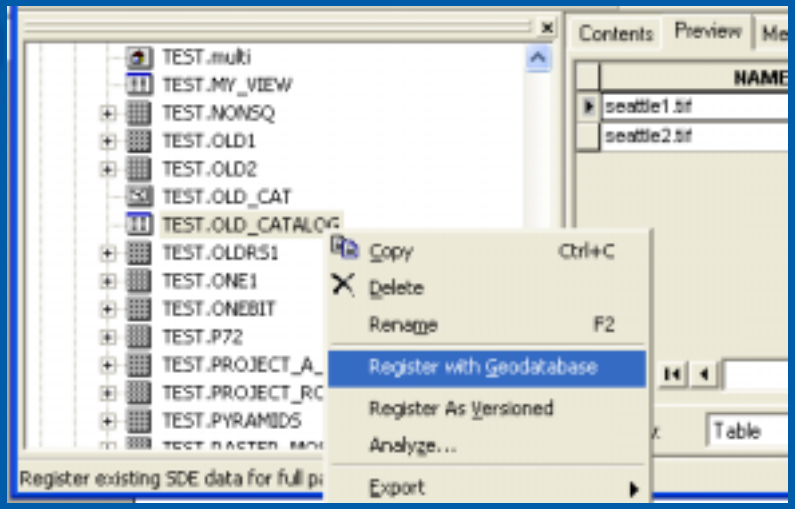
  Thurs 3:30 Room14-B

Data loading summary

- Prototyping is vital to success!
- Must evaluate:
  - Loading tools
  - Compression method and quality
  - Pyramid resampling method
  - Mosaic into raster dataset or group in raster catalog
  - Display speed
  - Multi-user access
  - Hardware
- Use the data in applications
  - Printing / mapping
  - Analysis
  - Scale of use (zoom factor)

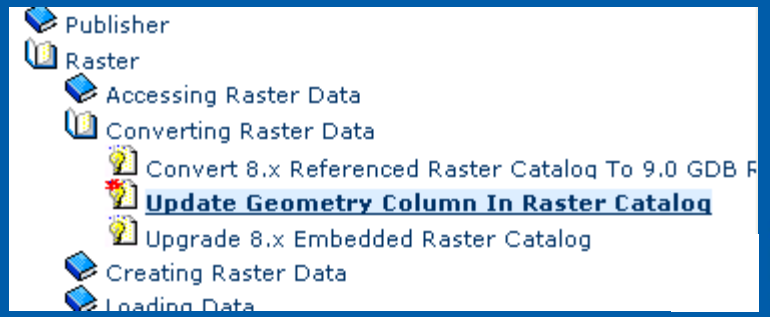
Upgrade path

- Raster datasets
  - No migration path necessary
    - Geometry column not used by ArcGIS
- Raster catalogs
  - Register with Geodatabase
    - Adds geometry column
    - Updates ArcSDE repository tables
    - Takes advantage of ArcGIS 9 rendering capabilities
- Referenced raster catalogs
  - Not supported with ArcSDE 9
    - Option: use not-managed personal geodatabase raster catalog
    - Option: Load data into ArcSDE 9 raster catalog



Upgrade path continued

- Register with the Geodatabase
  - Adds:
    - SHAPE column
    - Creates F and S table (Oracle and SQL Server)
- sderaster
  - Is not geodatabase aware
    - Creates 8.3 raster schema (no geometry column)
    - Can not use ArcGIS 9 enhancements
  - Registered with the Geodatabase:
    - `sderaster` will not update footprint column
      - Use ArcObjects script to update footprint



What's Next

- Prior to our next big release
  - Build a patch for users with Leica's LPS 8.7
  - Release HDF reader
  - Find and fix bugs

- Future releases of ArcGIS
  - Work on our Raster Data Services
    - Part of ArcGIS Server
  - Finish our Raster Transformation Pipeline
  - Build support for associated GDB raster tables
  - Build support for more raster formats

Additional resources at UC

- Raster technical workshops
  - Introduction to ArcGIS Raster – Thurs 10:30, Room 9
  - Raster Data Model – Wed 10:30, Room 5B
  - Loading Large Volumes of Raster Data Into a Geodatabase – Wed 3:30 Room 29A / Thurs 3:30 Room14B
- ArcSDE administration technical workshops
  - DB2 – Tue 10:30, Room 15A
  - Informix – Tues 8:30 and Wed 1:30, Room 15A
  - Oracle – Tues 1:30 and Wed 1:30, Room 15A
  - SQL Server – Wed 8:30 Room 15A, Thurs 10:30 Room 14B
- Doctor's Office
- Product Islands

Additional resources after UC

- Instructor-led classes:
  - ArcSDE Administration for DB2
  - ArcSDE Administration for Oracle
  - ArcSDE Administration for SQL Server
  - Building Geodatabases 1
- Virtual Campus:
  - Working with Rasters in ArcGIS
  - Storing Raster Data in an ArcSDE Geodatabase
- http://support.esri.com
  - Technical papers
  - User forum
- Systems Integration group

Thank you!

Open to Questions

Reminder: Workshop evaluation

Reminder: Ed Services survey – free Web Workshop