Sagar Mozumder,
Part-time Faculty, ESDM, DIU
Department of Environmental Science and Disaster Management, DIU

Data storage in GIS

In contemporary geographic information systems, geospatial data are stored as digital data. This binary representation allows a high level of flexibility and accuracy in the representation and processing of data.

The most important feature of GIS is that spatial data are stored in a structured format referred to as a spatial database. The way spatial data are structured will determine how easy it is for the user to store, retrieve and analyze the information.

1. Individual / personal files (Example: shapefile)
2. Geodatabases

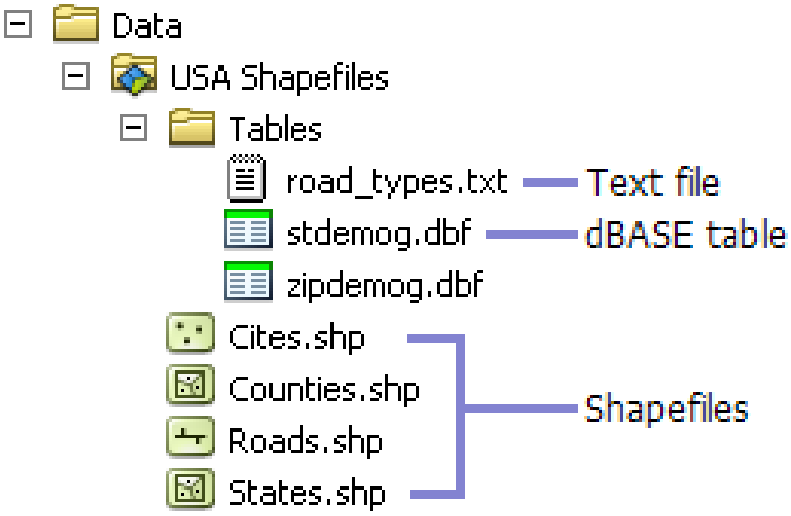
1. Individual / personal files (shapefile)

What is shapefile?

A shapefile is a simple, nontopological format for storing the geometric location and attribute information of geographic features. Geographic features in a shapefile can be represented by points, lines, or polygons (areas). The workspace containing shapefiles may also contain dBASE tables, which can store additional attributes that can be joined to a shapefile's features.

Below is an example of how shapefiles appear in ArcCatalog. You can also see the dBASE file (that may be associated with a shapefile).

Sagar Mozumder,
Part-time Faculty, ESDM, DIU
Department of Environmental Science and Disaster Management, DIU



All files that have the file extensions .txt, .asc, .csv, or .tab appear in ArcCatalog as text files by default. However, on the ***Options*** dialog box, you can choose which of these file types should be represented as text files and which should not be shown in the Catalog tree. When text files contain comma- and tab-delimited values, you can see their contents in the ArcCatalog table view and join them to geographic features. Text files can be deleted, but their contents are read-only in ArcCatalog.

You can join attributes stored in a dBASE table or text file to the features in a shapefile in a layer's Properties dialog box on the **Joins & Relates** tab. If a table contains information describing spatial locations, such as x,y,z coordinates or street addresses, you can create a shapefile representing those locations with tools available in ArcCatalog.

Shapefiles are simple storage formats that have been used in ArcMap since the 1990s when Esri created ArcView (the early version of ArcMap 10.3). Therefore, shapefiles have many limitations such as:

- Takes up more storage space on your computer than a geodatabase

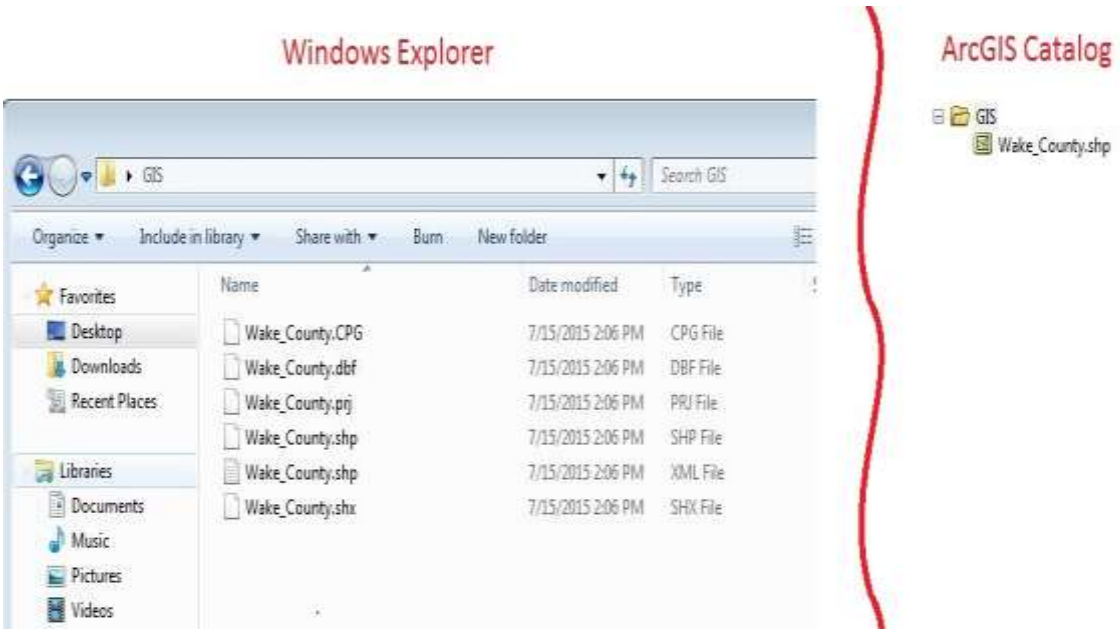
Sagar Mozumder,
Part-time Faculty, ESDM, DIU
Department of Environmental Science and Disaster Management, DIU

- Do not support names in fields longer than 10 characters
- Cannot store date and time in the same field
- Do not support raster files
- Do not store NULL values in a field; when a value is NULL, a shapefile will use 0 instead

Users are allowed to create points, lines, and polygons with a shapefile. One shapefile must have at least 3 files but most shapefiles have around 6 files. A shapefile must have:

- **.shp** – this file stores the geometry of the feature
- **.shx** – this file stores the index of the geometry
- **.dbf** – this file stores the attribute information for the feature

All files for the shapefile must be stored in the same location with the same name or else the shapefile will not load. When a shapefile is opened in Windows Explorer it will look different than when opened in ArcCatalog.



Sagar Mozumder,
Part-time Faculty, ESDM, DIU
Department of Environmental Science and Disaster Management, DIU

2. Geodatabases

The geodatabase is a "container" used to hold a collection of datasets. There are three types:

1. **File geodatabases**—Stored as folders in a file system. Each dataset is held as a file that can scale up to 1 TB in size. The file geodatabase is recommended over personal geodatabases.
2. **Personal geodatabases**—All datasets are stored within a Microsoft Access data file, which is limited in size to 2 GB.
3. **ArcSDE geodatabases**—Also known as multiuser geodatabases. Stored in a relational database using Oracle, Microsoft SQL Server, IBM DB2, IBM Informix, or PostgreSQL. These geodatabases require the use of ArcSDE and can be unlimited in size and numbers of users.

Comparing the three types of geodatabases

Key characteristics	ArcSDE geodatabase	File geodatabase	Personal geodatabase
Description	A collection of various types of GIS datasets held as tables in a relational database (This is the recommended native data format for ArcGIS stored and managed in a relational database.)	A collection of various types of GIS datasets held in a file system folder.(This is the recommended native data format for ArcGIS stored and managed in a file system folder.)	Original data format for ArcGIS geodatabases stored and managed in Microsoft Access data files.(This is limited in size and tied to the Windows operating system.)
Number of users	Multiuser: many readers and many writers	Single user and small workgroups:many readers or one writer per feature dataset, stand-alone feature class, or table. Concurrent use of any specific file eventually degrades for large numbers of readers.	Single user and small workgroups with smaller datasets: some readers and one writer. Concurrent use eventually degrades for large numbers of readers.

Sagar Mozumder,
Part-time Faculty, ESDM, DIU
Department of Environmental Science and Disaster Management, DIU

Storage format	• Oracle • Microsoft SQL Server • IBM DB2 • IBM Informix • PostgreSQL	Each dataset is a separate file on disk. A file geodatabase is a file folder that holds its dataset files.	All the contents in each personal geodatabase are held in a single Microsoft Access file (.mdb).
Size limits	Up to DBMS limits	One TB for each dataset. Each file geodatabase can hold many datasets. The 1 TB limit can be raised to 256 TB for extremely large image datasets. Each feature class can scale up to hundreds of millions of vector features per dataset.	Two GB per Access database. The effective limit before performance degrades is typically between 250 and 500 MB per Access database file.
Versioning support	Fully supported across all DBMSs; includes cross-database replication, updates using checkout and check-in, and historical archiving	Only supported as a geodatabase for clients who post updates using checkout and check-in and as a client to which updates can be sent using one-way replication.	Only supported as a geodatabase for clients who post updates using checkout and check-in and as a client to which updates can be sent using one-way replication.
Platforms	Windows, UNIX, Linux, and direct connections to DBMSs that can potentially run on any platform on the user's local network	Cross-platform.	Windows only.
Security and permissions	Provided by DBMS	Operating file system security.	Windows file system security.
Database administration tools	Full DBMS functions for backup, recovery, replication, SQL support, security, and so on	File system management.	Windows file system management.
Notes	Requires the use of ArcSDE technology; ArcSDE for SQL Server Express included with • ArcGIS for Desktop Standard and Advanced • ArcGIS Engine • ArcGIS for Server Workgroup ArcSDE for all other DBMSs included with ArcGIS for Server	You can optionally store data in a read-only compressed format to reduce storage requirements.	Often used as an attribute table manager (via Microsoft Access). Users like the string handling for text attributes.

A comparison of the three types of geodatabases

Sagar Mozumder,
Part-time Faculty, ESDM, DIU
Department of Environmental Science and Disaster Management, DIU

File geodatabases and personal geodatabases

File and personal geodatabases, which are freely available to all users of ArcGIS for Desktop Basic, Standard, and Advanced, are designed to support the full information model of the geodatabase, which comprises topologies, raster catalogs, network datasets, terrain datasets, address locators, and so on. File and personal geodatabases are designed to be edited by a single user and do not support geodatabase versioning. With a file geodatabase, it is possible to have more than one editor at the same time provided they are editing in different feature datasets, stand-alone feature classes, or tables.

The file geodatabase was a new geodatabase type released in ArcGIS 9.2. Its goals are to do the following:

- Provide a widely available, simple, and scalable geodatabase solution for all users.
- Provide a portable geodatabase that works across operating systems.
- Scale up to handle very large datasets.
- Provide excellent performance and scalability, for example, to support individual datasets containing well over 300 million features and datasets that can scale beyond 500 GB per file with very fast performance.
- Use an efficient data structure that is optimized for performance and storage. File geodatabases use about one-third of the feature geometry storage required by shapefiles and personal geodatabases. File geodatabases also allow users to compress vector data to a read-only format to reduce storage requirements even further.

Sagar Mozumder,
Part-time Faculty, ESDM, DIU
Department of Environmental Science and Disaster Management, DIU

- Outperform shapefiles for operations involving attributes and scale the data size limits way beyond shapefile limits.

Personal geodatabases have been used in ArcGIS since their initial release in version 8.0 and have used the Microsoft Access data file structure (the .mdb file). They support geodatabases that are limited in size to 2 GB or less. However, the effective database size is smaller, somewhere between 250 and 500 MB, before the database performance starts to slow down. Personal geodatabases are also only supported on the Microsoft Windows operating system. Users like the table operations they can perform using Microsoft Access on personal geodatabases. Many users really like the text-handling capabilities in Microsoft Access for working with attribute values. ArcGIS will continue to support personal geodatabases for numerous purposes. However, in most cases, Esri recommends using file geodatabases for their scalability in size, significantly faster performance, and cross-platform use. The file geodatabase is ideal for working with file-based datasets for GIS projects, personal use, and in small workgroups. It has strong performance and scales well to hold extremely large data volumes without requiring the use of a DBMS. Plus, it is portable across operating systems.

Typically, users will employ multiple file or personal geodatabases for their data collections and access these simultaneously for their GIS work.

Sagar Mozumder,
Part-time Faculty, ESDM, DIU
Department of Environmental Science and Disaster Management, DIU

ArcSDE geodatabases

When you need a large, multiuser geodatabase that can be edited and used simultaneously by many users, the ArcSDE geodatabase provides a good solution. It adds the ability to manage a shared, multiuser geodatabase as well as support for a number of critical version-based GIS workflows. The ability to leverage your organization's enterprise relational databases is a key advantage of the ArcSDE geodatabase.

ArcSDE geodatabases work with a variety of DBMS storage models (IBM DB2, Informix, Oracle, PostgreSQL, and SQL Server). ArcSDE geodatabases are primarily used in a wide range of individual, workgroup, department, and enterprise settings. **They take full advantage of underlying DBMS architectures to support the following:**

- Extremely large, continuous GIS databases
- Many simultaneous users
- Long transactions and versioned workflows
- Relational database support for GIS data management (providing the benefits of a relational database for scalability, reliability, security, backup, integrity, and so forth)
- SQL types for Spatial in all supported DBMSs (Oracle, SQL Server, PostgreSQL, Informix, and DB2)
- High performance that can scale to a very large number of users

Through many large geodatabase implementations, it has been found that DBMSs are efficient at moving in and out of tables the type of large binary objects required for GIS data. In addition, GIS database sizes can be much

Sagar Mozumder,
Part-time Faculty, ESDM, DIU
Department of Environmental Science and Disaster Management, DIU

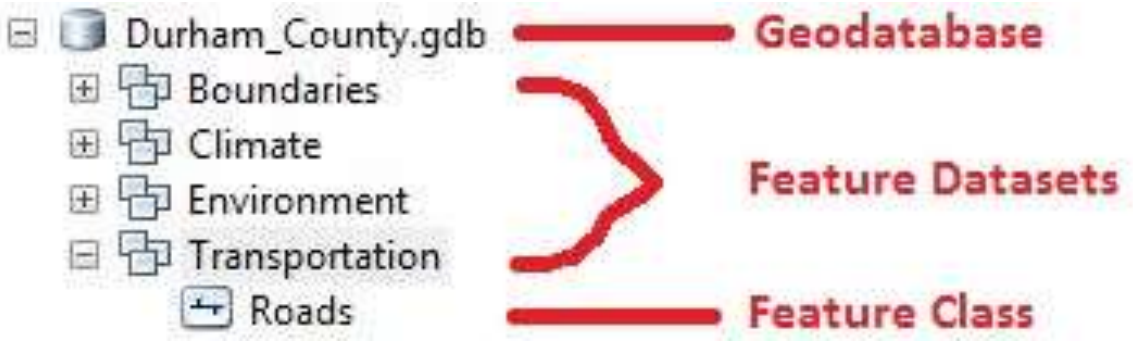
larger and the number of supported users greater than with file-based GIS datasets.

Geodatabases allow users to thematically organize their data and store spatial databases, tables, and raster datasets. There are two types of single user geodatabases: File Geodatabase and Personal Geodatabase. File geodatabases have many benefits including:

- 1 TB of storage limits of each dataset
- Better performance capabilities than Personal Geodatabase
- Many users can view data inside the File Geodatabase while the geodatabase is being edited by another user
- The geodatabase can be compressed which helps reduce the geodatabases' size on the disk

On the other hand, Personal Geodatabases were originally designed to be used in conjunction with Microsoft Access and the Geodatabase is stored as an Access file (.mdb). Therefore Personal Geodatabases can be opened directly in Microsoft Access, but the entire geodatabase can only have 2 GB of storage. To organize your data into themes you can create Feature Datasets within a geodatabase. Feature datasets store Feature Classes (which are the equivalent to shapefiles) with the same coordinate system. Like shapefiles, users can create points, lines, and polygons with feature classes; feature classes also have the ability to create annotation, and dimension features.

Sagar Mozumder,
Part-time Faculty, ESDM, DIU
Department of Environmental Science and Disaster Management, DIU



In order to create advanced datasets (such as add a network dataset, a geometric network, a terrain dataset, a parcel fabric, or run topology on an existing layer) in ArcGIS, you will need to create a Feature Dataset. You will not be able to access any files of a File geodatabase in Windows Explorer. When you do, the Durham_County geodatabase shown above will look like this:

Sagar Mozumder,
Part-time Faculty, ESDM, DIU
Department of Environmental Science and Disaster Management, DIU