Environmental Data Management Best Practices

Geospatial Data Management
GIS Hardware

This document presents common considerations regardless of the specific geographic information system (GIS) solution selected for the information technology (IT) professional to reference. The topics discussed in this document include GIS hardware; data storage; and scalability and deployment considerations.



Overview

A geographic information system (GIS) is a computer system that analyzes and displays geographically referenced information (USGS Undated). Environmental data often have a spatial component making GIS a useful application for storing, managing, and analyzing environmental data. Modern GISs often have multiple components, including desktop applications, server(s), data stores, and system redundancies. Each component has hardware requirements that need to be assessed when planning GIS deployment. Figure 1 depicts the physical components of a typical GIS deployment.

Hardware

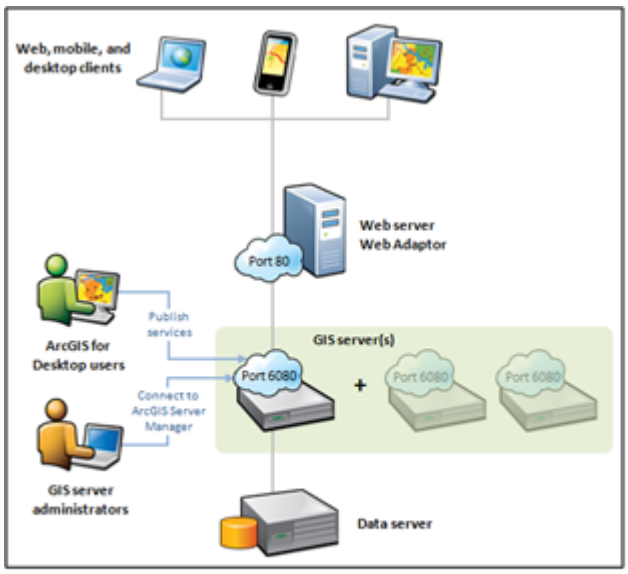


Figure 1. ArcGIS server site architecture.
Source: ESRI Undated

When deciding on appropriate GIS hardware, it is best to engage IT staff to understand the available resources and security requirements for the GIS. Refer to the needs assessment developed during GIS planning and any organizational standards (see Organization Standards for Geospatial Environmental Data Management for additional information) that may be in place for the organization. GIS systems rarely involve a single computer in modern deployments. In general, the GIS is divided into four primary components:

- end-user computer
- GIS server
- one or more data stores
- network server

Each component may reside on physical or virtual hardware, with small or project-specific deployments sometimes installed on a single computer.

GIS applications have requirements for computer processor units (CPUs), graphics processing units (GPUs), operating systems, and network communications. In general, desktop applications currently require hardware that has multiple CPU cores capable of parallel processing in a 64-bit architecture. The GPU is most used in three-dimensional visualizations and analyses but can also be used during intensive geospatial analysis applications. Three-dimensional visualizations, including contaminant plumes and geologic models, are common in environmental data analysis, so a GPU capable of supporting planned analysis and visualizations should be selected.

When selecting the hardware for GIS servers the key considerations should include:

- number of concurrent users accessing or performing analysis of the data.
- operating system (Linux and Windows are currently the primary operating systems for GIS. Apple's operating system can be used if a virtual environment is set up.)
- data visualizations in two or three dimensions.
- data security and duplication.
- anticipated network load for public-facing services.

Data Storage

Data storage represents the location of the data presented in GIS, which may include:

- spatial vector data. Vectors are presented as points (no dimension), lines (1 dimensional), and polygons (2 dimensional).
- spatial raster data. A grid of cells with fixed dimensions in which each cell stores a single value and the origin corner has a defined coordinate location.
- nonspatial structured data. Data without spatial geometry that have a structure to the data that allows for parsing, ordering, and querying by a computer program.
- nonspatial unstructured data. Data without spatial geometry that is not computer-readable without preprocessing (for example, a photograph of handwritten notes).

This data may be stored in varying file formats, including flat or text files, spreadsheets, shapefiles, feature classes within geodatabases, image files, and relational database tables. Spatial data can be mapped in GIS, and nonspatial data can be joined and linked to spatial data as attributes. For example, a point can represent the location of a well, and a nonspatial well construction log can be linked and displayed at the well point. The options for data visualization, analysis, change management, archiving, and transferring are dependent on the format the information is stored in. In some GIS implementations, data storage is referred to as a data store. For an enterprise, multi-user data storage deployment, a data store should include components for controlled access and user rights, file auditing to record creation, modification, and potentially versioning to allow workflows for data verification prior to publishing and interval recovery to allow revocation of edits to the data records. For best practices on verification and quality assurance/quality control (QA/QC), see the Geospatial Data Standards subtopic sheet.

Most data stores are relational database instances that contain multiple tables that are structured to hold the spatial geometry of GIS features and the attributes describing the GIS features. The data stores respond to requests by the GIS server or user for data within the boundaries of a geographic area. An example of an environmental data store is USEPA EnviroMapper, which is managed by USEPA personnel and contain data from various USEPA programs. Each USEPA regulated facility is mapped and attribute information from different sources is linked to the location. Data can be accessed by the public with simple searches or custom queries that access the Oracle relational database management system.

The computer housing the data store should be physically and digitally secured, a thorough backup model should be implemented to prevent data loss or service downtime, and a disaster recovery plan should be in place. Additional information on disaster recovery is available in the Data Disaster Recovery document.

Scalability and Deployment Considerations

Scalability refers to the incremental growth of a computer system, including not only adding additional end-user computers, but also increasing storage capacity, adding memory for analysis capacity, and possibly purchasing additional field collection devices. GIS can accommodate scaling when appropriately scoped, designed, and implemented. During the needs assessment, the potential scale of data storage, number of users, server processor load, and system redundancy should be described. Some aspects of GIS, such as versioning of project data, can be incorporated after the GIS server has been set up. If selecting GIS hardware for a particular environmental project, consider how the project will expand either with additional data, advanced analysis and visualization, or a larger geographic area. Also consider incorporating additional environmental data, projects, and regions in the future and make sure hardware selection will support that expansion.

Resources

Related documents:

- GIS Design Strategies – System Design Strategies Preface – GIS Wiki | The GIS Encyclopedia
- Enterprise GIS Architecture – CP-29 – Enterprise GIS | GIS&T Body of Knowledge (ucgis.org)
- GIS Wiki http://wiki.gis.com/wiki/index.php/Main_Page