Unit - III

Data Input and Topology

Scanner - Raster Data Input - Raster Data File Formats - Vector Data Input - Digitiser - Topology - Adjacency, Connectivity and Containment - Topological Consistency Rules - Attribute Data Linking - ODBC - GPS - Concept GPS based mapping.

Scanner:

Scanning converts paper maps into digital format by capturing features as individual cells or pixels, producing an automated image. Maps are generally considered the backbone of any GIS activity. But many a time paper maps are not easily available in a form that can be conventional surveys. New maps can be produced using improved technologies but this requires time as it increases the volume of work. Thus, we have to resort to the available maps. These paper maps have to be first converted into a digital format usable by the computer. This is a critical step as the quality of the analog document must be preserved in the translation to the computer domain.

The technology used for this kind of conversions is known as scanning and the instrument used for this kind of operation is known as a scanner. A scanner can be thought of as an electronic input device that converts analog information of a document like a map, photograph or an overlay into a digital format that can be used by the computer. Scanning automatically captures map features, text and symbols as individual cells, or pixels and produces an automated image.

Working of a Scanner:

The most important component inside a scanner is the scanner head which can move along the length of the scanner. The scanner head contains either a charged-couple device sensor or a contact image sensor. A CCD consists of a number of photosensitive cells or pixels packed together on a chip.

While scanning, a bright light from the scanner strikes the image to be scanned and is reflected onto the photosensitive surface of the sensor placed on the chip. The most advanced large format scanners use CCD's with 8000 pixels per chip for providing a very good image quality.

While scanning a color image is slightly different in which the scanner head has to scan image for three different colors ie) red, green, blue. In older color scanners, this was accomplished by scanning the same area three times over for the three different colors. This type of scanner is known as three-pass scanner. However, most of the colour scanners now scan in one pass scanning all the three colours in one go by using color filters. In principle, a color CCD works in the same way as a monochrome CCD. But this each colour is constructed by mixing red, green and blue. Thus, a 24-bit RGB CCD presents each pixel by 24 bits of information. Usually, a scanner using these three colors can create upto 16.8 million colours.

Types of scanners:

Hand-held scanners although portable, can only scan images upto about four inches wide. They require a very steady hand for moving the scan head over the document. They are useful for scanning small logos or signatures and are virtually of no use for scanning maps and photographs.

The most commonly used scanner is a flatbed scanner also known as desktop scanner. It has a glass plate on which the picture or the document is placed. The scanner head placed beneath the glass plate moves across the picture and the result is a good quality scanned image. For scanning large maps or top sheets wide format flatbed scanners are be used.

Then there are the drum scanners which are mostly used by the printing professionals. In this type of scanner, the image or the document is placed on a glass cylinder that rotates at very high speeds around a centrally located sensor containing photo-multiplier tube instead of a CCD to scan. Prior to the advances in the field of sheet-fed scanners, the drum scanners were extensively used for scanning maps and other documents.

Raster GIS File Formats:

Raster data is made up of pixels. They are usually regularly-spaced and square but they dont have to be. Rasters have pixels that are associated with a value or class.

i) ERDAS Imagine (.img)

ERDAS Imagine files are proprietary file format developed by Hexagon Geospatial. Img files are commonly used for raster data to store single and multiple bands of satellite data.

ii) American Standard Code for Information Interchange (ASCII Grid) (.asc)

ASCII uses a set of numbers between 0 and 255 for information storage and processing. They also contain header information with a set of keywords. In their native form, ASCII text files store GIS data in a delimited format. This could be comma, space or tab-delimited format. Going from non-spatial to spatial data, you can run a conversion process tool like ASCII to raster.

iii) GeoTiff (.tif, .tiff, .ovr)

The GeoTiff has become an industry image standard file for GIS and satellite remote sensing applications. GeoTiffs may be accompanied by other files.

* TFW is the world file that is required to give your raster geolocation.

* XML optionally accompany GeoTiffs and are your metadata.

* AUX auxilliary files store projections and other information.

* OVR pyramid files improves performance for raster display.

iv) IDRISI Raster (.rst, .rdc):

IDRISI assigns RST extensions to all raster layers. They consist of numeric grid cell values as integers, real numbers, byte and RGB24.

The raster documentation file (RDC) is a companion text file for RST files. They assign the number of columns and rows to RST files. Farther to this, they record the file type, coordinate system, reference units and positional error.

v) Envi RAW Raster(.bil, .bip, .bsq):

Band Interleaved files are a raster storage extension for single/multi-band aerial and satellite imagery.

Band interleaved for line stores pixel information based on rows for all bands in an image.

Band interleaved by pixel assigns pixel values for each band by rows.

Band sequential format stores separate bands by rows.

Vector GIS file formats:

Vector data is not made up of grids of pixels. Instead, vector graphics are comprised of vertices and paths. The three basic types for vector data are points, lines and polygons(area).

i) Esri Shapefile(.shp, .dbf, .shx):

The shapefile is BYFAR the most common geospatial file type you'll encounter. All commercial and opensource accept shapefile as a GIS format.

ii) Geographic Java Script Object Notation (GeoJSON)

The GeoJSON format is mostly for web-based mapping. GeoJSON stores coordinates as text in JavaScript Object Notation form. This includes vector points, lines, polygons as well as tabular information.

iii) Geography Markup Language (.gml):

GML allows for the use of geographic coordinates extension of XML. And extensible Markup Language (XML) is both human-readable and machine-readable.

GML stores geographic entites in the form of text. Similar to GeoJSON, GML can be updated in any text editor. Each feature has a list of properties, geometry and spatial reference system.

There is generally more overhead when compare GML with GeoJSON. This is because GML results in more date for the same amount of information.

iv) Google Keyhole Markup Language (.KML, .KMZ)

KML stands for Keyhole Markup Language. This GIS format is XML-based and is primarily used for Google Earth. KML was developed by Keyhole Inc which was later acquired by Google.

KMZ replaced KML as being the default Google Earth geospatial format because it is a compressed version of the file. KML/KMZ became an international standard of the Open Geospatial Consortium in 2008.

Digitizing:

Digitizing in GIS is the process of converting geographic data either from a hardcopy or a scanned image into vector data by tracing the features. During the digitizing process, features from the traced map or image are captured as coordinates in either point, line or polygon.

Types of Digitizing in GIS:

The most common method of encoding spatial features from paper maps is manual digitizing. It is an appropriate technique when selected features are required from a paper map. Manual digitizing requires a digitizing table that is linked to a computer workstation. The digitizing table is essentially a large flat tablet, the surface of which is underlain by a very fine mesh of wires. Attached to the digitizer via a cable is a cursor that can be moved freely over the surface of the table. Buttons on the cursor allow the user to send instructions to the computer. The position of the cursor on the table is registered by reference to its position above the wire mesh.

Heads up digitizing is the method of tracing geographic features from other dataset directly on the computer screen. Automated digitizing involves using image processing software that contains pattern recognition technology to generate vectors. The procedure followed when digitizing a paper map using a manual digitizer has the following 5 stages.

i) <u>Registration</u>: The map to be digitized is fixed firmly to the table top with a sticky tape. Five or more control points are identified. The geographic co-ordinates of the control points are noted and their locations digitized by positioning the cross-hairs on the cursor exactly over them and pressing the digitize button on the cursor. This sends the co-ordinates of a point on the table to the computer and stores them in a file as 'digitizer co-ordinate'.

ii) <u>Digitizing point features</u>: Point features, for eg) spot heights, hotel locations or meteorological stations, are recorded as a single digitized point. A unique code number or identifier is added so that the attribute information may be attached later. For instance, the hotel with ID number '1' would later be identified as 'Mountain View'.

iii) Digitizing line features:

Line features are digitized as a series of points that the software will join with straight line segments. In some GIS packages lines are reffered to as arcs and their start and end points as nodes. This given rise to the term, arc-node topology, used to describe a method of structuring line features.

iv) Digitizing area features:

Area features on polygons for example forested areas or administrative boundaries, are digitized as a series of points linked together by line segments in the same way as line features. Here, it is important that the start point and end points join to form a complete area. Polygons can be digitized as a series of individual lines, which are later joined to form areas. In this case it is important that each line segment is digitized only once.

v) Adding attribute Information:

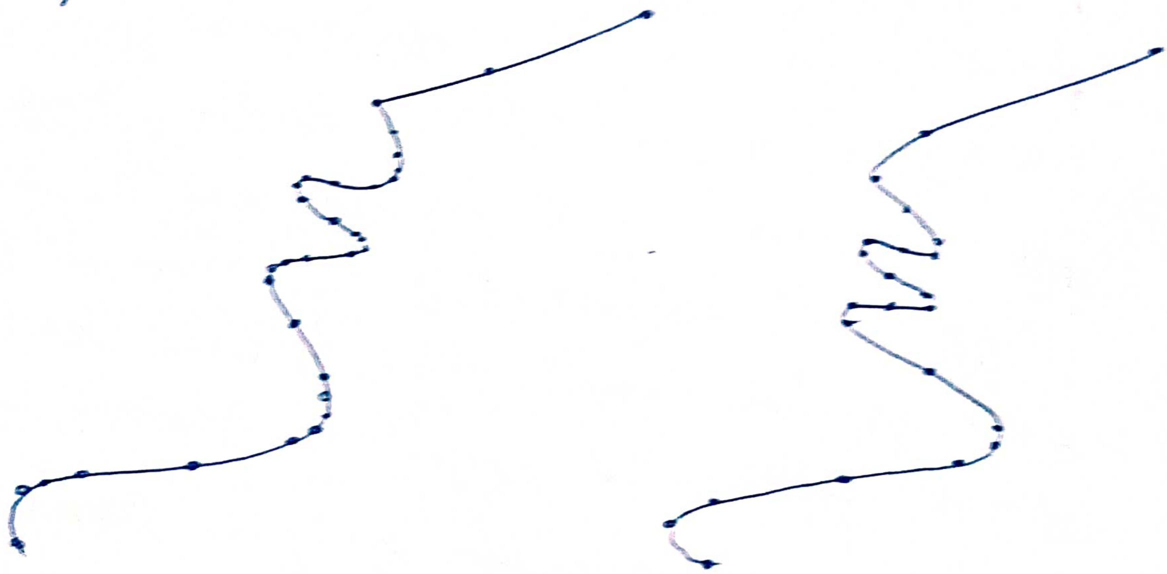
Attribute data may be added to digitize polygon features by linking them to a centroid in each polygon. These are either digitized manually or created automatically once the polygons have been encoded. Using a unique identifier or code number, attribute data can then be linked to the polygon centroids of appropriate polygons. In this way, the forest stand may have data related to tree species, tree ages, tree numbers and timber volume attached to a point within the polygon.

Manual digitizers may be used in two modes:

i) Point mode

ii) Stream mode.

In point mode the user begins digitizing each line segment with a start node, records each change in direction of the line with a digitized point and finishes the segment with an end node. Thus, a straight line can be digitized with just two points, the start and end nodes. For more complex lines, a greater number of points are required between the start and end nodes. Smooth curves are problematic since they require an infinite number of points to record their true shape.



Point Mode

Stream Mode

In stream mode the digitizer is set up to record points according to a stated time interval or on a distance basis. Once the user has recorded the start of a line the digitizer might be set to record a point automatically every 0.5 seconds and the user must move the cursor along the line to record its shape.

An end node is required to stop the digitizer recording further points. The speed at which the cursor is moved along the line determines the number of points recorded. Thus, where the line is more complex and the cursor needs to be moved slowly and with more care, a greater number of points will be recorded. Conversely, where the line is straight, the cursor can be moved more quickly and fewer points are recorded.

The choice between point mode and stream mode digitizing is largely a matter of personal preference. Stream mode digitizing requires more skill than point mode digitizing, and for an experienced user may be a faster method. Stream mode will usually generate more points and hence larger files, than point mode.

Topology:

Topology is the mathematical representation of the physical relationships that exists between the geographical elements. Topology has long been a key GIS requirement for data management and integrity. In general, a topological data model manages spatial relationships by representing spatial objects as an underlying graph of topological primitives - nodes, faces and edges. These primitives, together with their relationships to one another and to the features whose boundaries they represent, are defined by representing the feature geometries in a planar graph of topological elements.

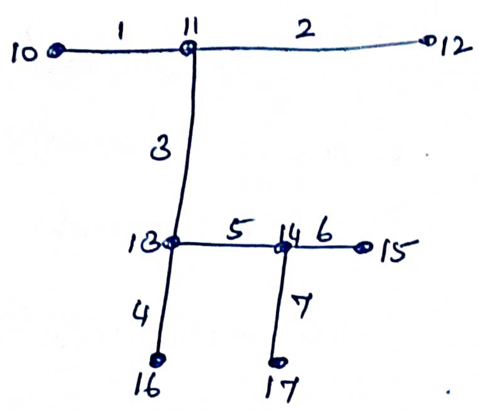
Topology is useful in GIS because many spatial modeling operations don't require co-ordinates, only topological information. For eg, to find an optimal path between two points requires a list of the arcs that connect to each other and the cost to traverse each arc in each direction. Coordinates are only needed for drawing the path after it is calculated.

The topological structure supports three major topological concepts:

- Connectivity: Arcs connect to each other at nodes.
- Area Definition: Arcs that connect to surround an area define a polygon.
- Contiguity: Arcs have direction and left and right sides.

Connectivity:

Connectivity is defined through arc-node topology. This is the basis for many network tracing and path finding operations. Connectivity allows you to identify a route to the airport, connect streams to rivers or follow a path from the water treatment plant to a house.



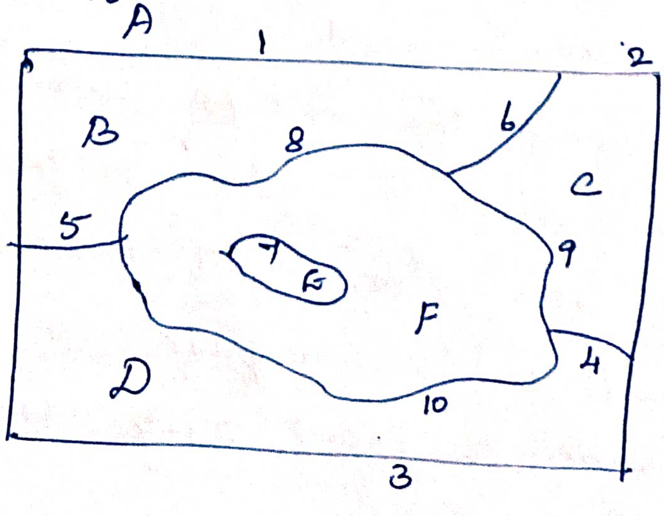
Arc	From Node	To Node.
1	10	11
2	11	12
3	11	13
4	13	16
5	13	14
6	14	15
7	14	17

Arc-node topology is supported through an arc-node list. The list identifies the from- and to-nodes for each arc. Connected arcs are determined by searching through the list for common node numbers. In the above example, it is possible to determine that arcs 1, 2 and 3 all intersect because they share node 11. The computer can determine that it is possible to travel along arc 1 and turn

onto arc 3 because they share a common node (11)
But its not possible to turn directly from arc 1 onto arc 5 because they don't share a common node.

Containment:

Many of the geographic features that may be represented cover a distinguishable area on the surface of the earth, such as lakes, parcels of land, and census tracts. An area is represented in the vector model by one or more boundaries defining a polygon. Although this sounds counterintuitive, consider a lake with an island in the middle. The lake actually has two boundaries: one that defines its outer edge and the island that defines its inner edge. In the terminology of the vector model, an island defines an inner boundary of a polygon.



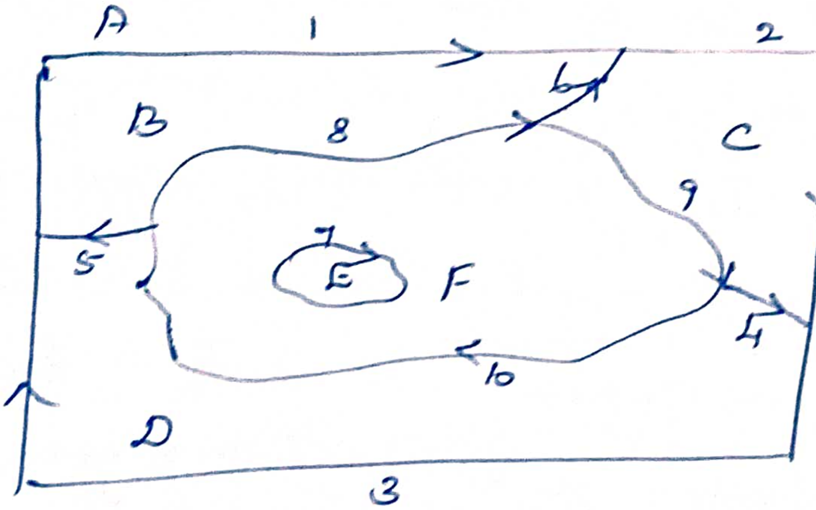
Polygon	Arc List
B	1,6,8,5
C	2,4,9,6
D	3,5,10,4
E	7
F	8,9,10,0,7

The arc-node structure represents polygons as an ordered list of arcs rather a closed loop of x,y coordinates. This is called polygon-arc topology. In the illustration above, polygon F is made up of arcs 8,9,10 and 7. The 0 before the 7 indicates that this arc creates an island in the polygon.

Each arc appears in two polygons. Since the polygon is simply the list of arcs defining its boundary, arc co-ordinates are stored only once, thereby reducing the amount of data and ensuring that the boundaries of adjacent polygons don't overlap.

Contiguity:

Two geographic features that share a boundary are called adjacent. Contiguity is the topological concept that allows the vector data model to determine adjacency. Polygons topology defines contiguity. Polygons are contigous to each other if they share a common arc. This is the basis for many neighbor and overlay operation



Arc	Left Polygon	Right Polygon
1	A	B
2	A	C
3	A	D
4	C	D
5	D	B
6	B	C
7	F	E
8	B	F
9	C	F
10	D	F

Recall that the from-node and to-node define an arc. This indicates an arc's direction so the polygons on its left and right sides can be determined. Left-right topology refers to the polygons on the left side and right side of an arc. In the given example, polygon B is on the left of arc 6, and polygon C is on the right. Thus, we know that polygons B and C are adjacent.

Notice that the label for polygon A is outside the boundary of the area. This polygon is called the external or universe, polygon and represents the world outside the study area. The universe polygon ensures that each arc always has a left and right side defined.

Topological Consistency Rules:

Many topology rules can be imposed on features in a geodatabase. A well-designed geodatabase will have only those topology rules that define key spatial relationships needed by an organization. Most topology, violations have fixes that you can use to correct errors.

Topology Rules Based on Points:

i) Must Concide with

ii) Must be disjoint

iii) Must be covered by endpoints of

iv) Points must be covered by line

v) Must be properly inside polygons

vi) Must be covered by boundary of

Topology Rules Based on Lines:

i) Must Not have dangles

ii) Must not overlap

iii) Must not self overlap

iv) Must not self intersect

v) Must not intersect

vi) Must not have pseudo nodes

vii) Must be larger than ~~psa~~ closure tolerance

Topology Rules Based on Polygons:

i) Must not overlap

ii) Must not have gaps

iii) Contains point

iv) Contains one point

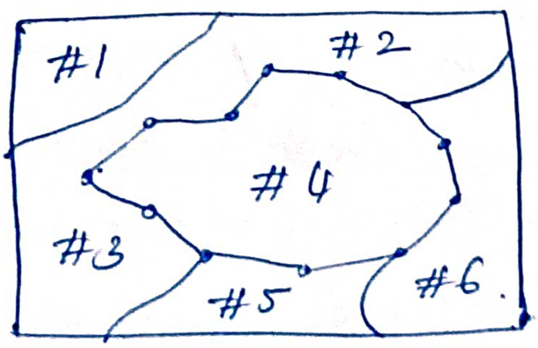
v) Must not overlap with

vi) Must cover each other

vii) Area boundary must be covered by boundary of.

Attribute Data Linking:

There are two types of GIS data: spatial data and attribute data. Attribute data is additional information appended in tabular format linked with spatial features. The attribute data is linked with spatial data through unique id. The spatial data contains information about where and attribute data can contain information about what, where and why. Attribute data provides characteristics about spatial data.



Id	Size	Cover	Soil	Age	Etc.
#1	65	SP	ALV	>	—
#2	160	H	COH	60	—
#3	206	H	SRP	55	—
#4	119	DF	COH	5	—
#5	67	H	SRP	55	—
#6.	—	—	—	—	—

Joins:

When our data was all in a single table we could easily retrieve a particular row from that table. But if the data we are looking for is available in two or more tables then join can be used to retrieve those data. Join is used to fetch data from two or more tables, which is joined to appear as single set of data. It is used for combining column from two or more tables by using values common to both tables.

There are several types of JOINs.

* INNER
* LEFT OUTER
* RIGHT OUTER

They all do slightly different things, but the basic theory behind them all is same.

Inner Join:

An inner join returns a result set that contains the common elements of the tables. ie) the intersection where they match on the joined condition. An inner join focuses on the commonality between two tables. When using an inner join, there must be at least some matching data between two tables that are being compared. Inner joins are the most frequently used join operation.

Location.

fid	city	state	country
101	Chennai	TN	Ind
102	Coimbatore	TN	Ind
103	Nellore	AP	Ind
104	vellore	TN	Ind
105	Tirupathi	AP	Ind.

+

City Details.

City	Population	Area	Rainfall
Chennai	12239	9236	55.32
Coimbatore	11234	10255	120.20
Vellore	12345	9555	55.20
Kancheepuram	11235	7895	102.3
Nellore	11105	6000	90.3

fid	City	State	Country	City	population	area	rainfall.
101	Chennai	TN	Ind	Chennai	12234	9236	55.32
102	Coimbatore	TN	Ind	Coimbatore	11234	10355	120.20
103	Nellore	AP	Ind	Nellore	11105	6000	90.3
104	Vellore	TN	Ind	Vellore	12345	9555	55.20

Left Outer Join:

A left join or left outer join takes all the rows from one table, defined as the left table and joins it with a second table. A left join will always include the rows from the left table, even if there are no matching rows in the table joined with.

fid	city	State	Country	City	Population	Area	R...
101	Chennai	TN	Ind	Chennai	12234	9236	55
102	Coimbatore	TN	Ind	Coimbatore	11234	10355	120.2
103	Nellore	AP	Ind	Nellore	11105	6000	90.3
104	Vellore	TN	Ind	Vellore	12345	9555	55.20
105	Tirupati	AP	Ind	–	–	–	–

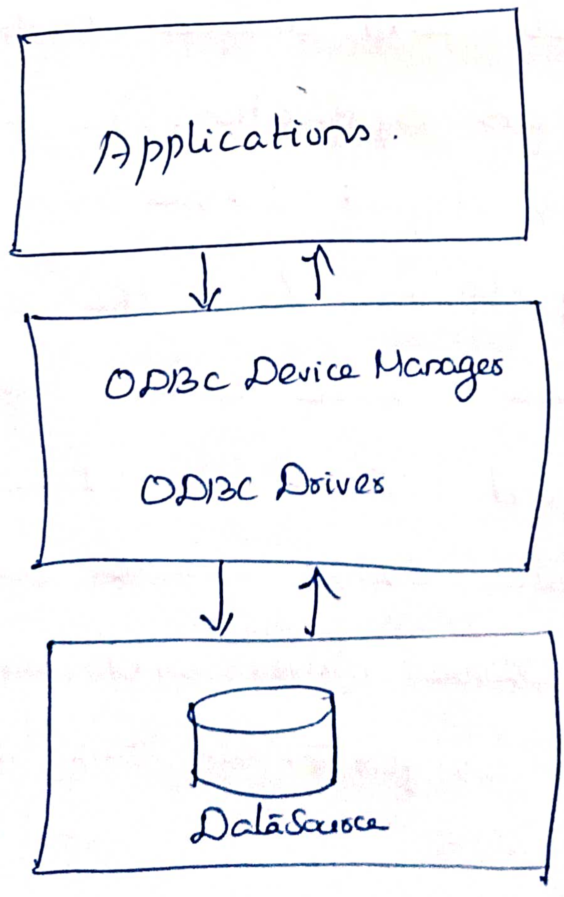
Right Outer Join:

A right outer join is similar to a Left outer join except that the roles between the two tables are reversed, and all the rows on the second table are included along with any matching rows from the right table, even if there are no matching rows in the table it is joined with.

fid	city	State	Country	City	Population	Area	Rainfall
101	Chennai	TN	Ind	Chennai	12234	9236	55.32
102	Coimbatore	TN	Ind	Coimbatore	11234	10355	120.20
104	Vellore	TN	Ind	Vellore	12345	9555	55.20
103	Nellore	AP	Ind	Nellore	11105	6000	90.3
NIL	NIL	NIL	NIL	Kanchipuram	11235	7895	102.3

Open Database Connectivity (ODBC):

An Open Database Connectivity is an interface that allows applications to access data in database management systems using SQL as a standard for accessing the data. ODBC permits maximum interoperability, which means a single application can access different DBMS. Application end users can then add ODBC database drivers to link the application to their choice of DBMS. Application end users can then add ODBC database ~~dev~~ drivers to link the application to their choice of DBMS.



The ODBC solution for accessing data led to ODBC database drivers, which are dynamic-link libraries on Windows and shared objects on Linux/Unix. These drivers allow an application to gain access to one or more data sources. ODBC provides a standard interface to allow application developers and vendors of database drivers to exchange data between applications and data sources.

ODBC Driver Manager:

The ODBC driver processes ODBC function calls. It loads and unloads ODBC drivers on behalf of an application. The Windows platform comes with a default Device Manager, while non-windows platforms have the choice to use an open source ODBC Driver Manager like unix ODBC and iODBC. The ODBC driver manager processes ODBC function calls or passes them to an ODBC driver and resolves ODBC version conflicts.

ODBC Driver:

The ODBC driver processes ODBC function calls, submits SQL requests to a specific data source and return results to the application. The ODBC driver may also modify an application's request so that the request confirms to syntax supported by the associated database. A framework to easily build an ODBC drivers is available from Simba Technologies.

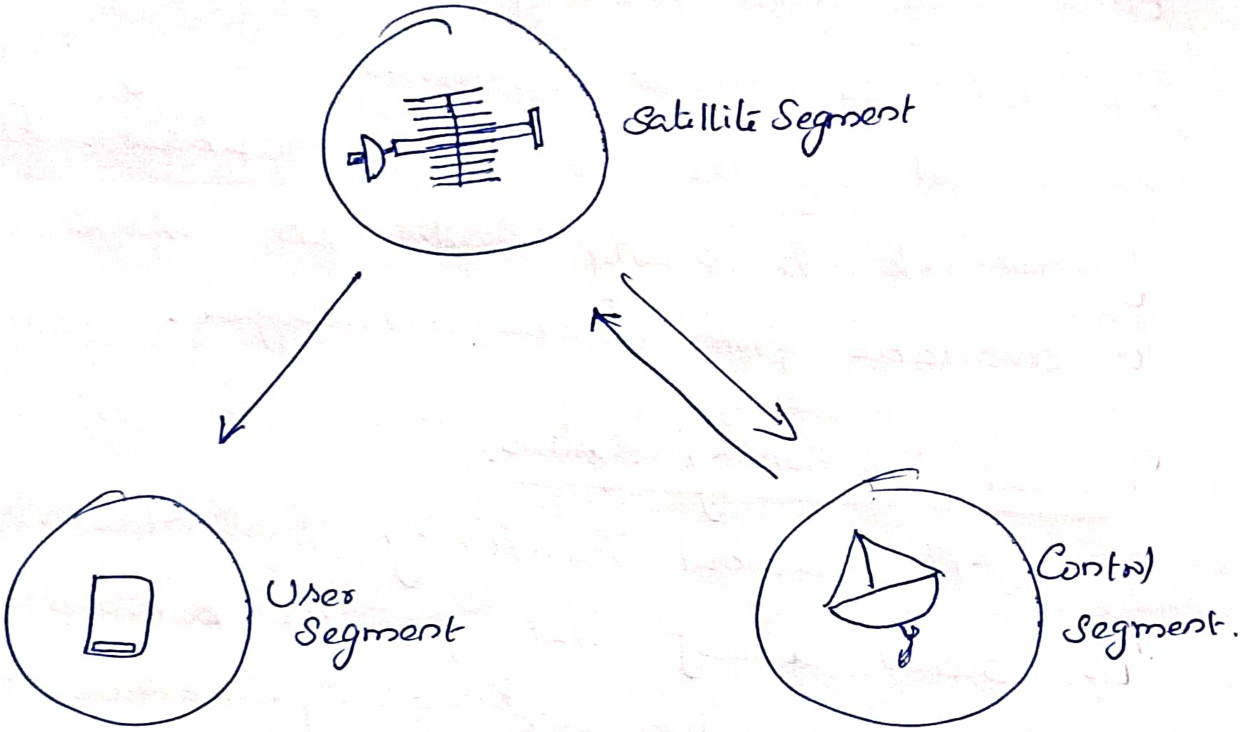
Global Positioning System:

The Global Positioning System (GPS) is an US-owned utility that provides users with positioning, navigation and timing services. This system consists of three segments.

- The Space Segment
- The Control Segment
- The User Segment.

GPS technology was first used by the United States military in the 1960's and expanded into civilian use over the next few decades.

Today GPS receivers are included in many commercial products such as automobiles, smart phones, exercise watches and GIS devices.



Space Segment:

The GPS space segment consists of a constellation of satellites transmitting radio signals to users. The United State is committed to maintaining the availability of at least 24 operational GPS satellites, 95% of the time. To ensure this commitment, the Air Force has been flying 31 operational GPS satellites for the past few years. GPS satellites fly in medium Earth Orbit (MEO) at an altitude of approximately 20,200 km. Each satellite circles the earth twice a day. The satellites in the GPS constellation are arranged into 6 equally placed orbital planes surrounding the earth.

Control Segment:

The GPS Control segment consists of a global network of ground facilities that track the GPS satellites, monitor their transmissions, performance analyses and send commands and data to the constellation. The current operational control segment includes a master control station, an alternate master control station, 11 command and control antennas and 16 monitoring sites.

User Segments:

Like the internet, GPS is an essential element of the global information infrastructure. The free, open and dependable nature of GPS has led to the development of hundreds of applications affecting every aspect of modern life. GPS technology is now in everything from cell phones and wrist watches to bulldozers, shipping containers and ATM's.

GPS ~~Based~~ based Mapping:

The surveying and mapping community was one of the first to take advantage of GPS because it dramatically increased productivity and resulted in more accurate and reliable data. Today, GPS is a vital part of surveying and mapping activities around the world. When used by skilled professionals, GPS provides surveying and mapping data of the highest accuracy. GPS based data collection is much faster than conventional surveying and mapping techniques, reducing the amount of equipment and labor required. A single surveyor can now accomplish in one day what once took an entire team weeks to do.

GPS supports the accurate mapping and modeling of the physical world - from mountains and rivers to streets and buildings to utility lines and other resources. Features measured with GPS can be displayed on maps and in GIS that store, manipulate and display geographically referenced data.

Governments, scientific organizations and commercial operations throughout the world use GPS and GIS technology to facilitate timely decisions and wise use of resources. Any organization or agency that requires accurate location information about its assets can benifit from the efficiency and productivity provided by GPS positioning. Unlike the conventional techniques, GPS surveying is not bound by constraints such as line-of-sight visibility between survey stations. The stations can be deployed at greater distances from each other and can operate anywhere with a good view of the sky, rather than being confined to remote hilltops as previously required.