

Geo Spatial Analysis in Urban Planning
Prof. Saikat Kumar Paul
Department of Architecture and Regional Planning
Indian Institute of Technology, Kharagpur

Module - 01
Introduction to Geographic Information System and Geographic Distribution
Lecture - 04
Vector Data Model

Welcome dear students. Today, in this lecture we are going to see what are the different types of Vector Data Model. So, this particular lecture is for those students who for the first time they are learning about GIS systems. So, before we go into very involved analysis we should know what are the different basic data types and how and we can use this data types to code the different; I mean, the features that we see on earth surface or different events.

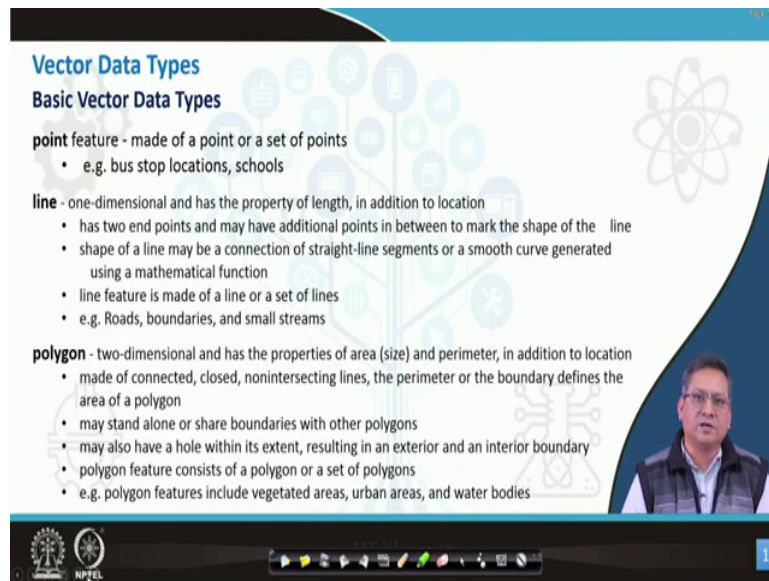
So, how we can code these particular events as data types? So, of that I mean there are primarily two types of data types that are being used in your GIS, one is the vector data type and second is the raster data type. So, in this lecture we are going to cover the vector data model.

(Refer Slide Time: 01:11)

The image shows a presentation slide with a dark blue header and a light blue footer. The header contains the text "CONCEPTS COVERED" in white. Below the header, there is a list of three items, each preceded by a right-pointing arrowhead: "Vector Data Types", "Topology", and "GeoRelational Data Model". In the bottom right corner, there is a small video inset showing a man with glasses and a dark vest over a light shirt. At the bottom of the slide, there is a navigation bar with several icons for navigation and a small "Page 1" indicator in the top right corner.

So, we are going to talk about the vector data types. We would see what is the topology and in the context of GIS I mean what is its use; we are going to look into the geo relational data models different types of geo relational data models.

(Refer Slide Time: 01:31)



Vector Data Types

Basic Vector Data Types

point feature - made of a point or a set of points

- e.g. bus stop locations, schools

line - one-dimensional and has the property of length, in addition to location

- has two end points and may have additional points in between to mark the shape of the line
- shape of a line may be a connection of straight-line segments or a smooth curve generated using a mathematical function
- line feature is made of a line or a set of lines
- e.g. Roads, boundaries, and small streams

polygon - two-dimensional and has the properties of area (size) and perimeter, in addition to location

- made of connected, closed, nonintersecting lines, the perimeter or the boundary defines the area of a polygon
- may stand alone or share boundaries with other polygons
- may also have a hole within its extent, resulting in an exterior and an interior boundary
- polygon feature consists of a polygon or a set of polygons
- e.g. polygon features include vegetated areas, urban areas, and water bodies

So, let us see what are the basic vector data types. So, the first one is a point feature and it is a layer or a point in a GIS where we have a point feature, it would comprise of a set of points. Now, I mean you can think of some examples like say location of wells, location of bus stops or schools or any other such feature.

Now, the next vector data type basic vector data type is a line. Now, it is a one-dimensional; it has one-dimensional property and it only has length. So, in addition to the location, it would also have two points the start point and the end points and in between we would have the link. So, that would basically by joining a line between this two start points, it would create a line segment.

Now, it may be a connected as line segment. So, it may have multiple segments or it could be a smooth curve; I mean, we can generate a smooth curve from point segments, I mean or line

segments these are these functions are generally known as slicing algorithms. So, we can generate smooth lines as well.

So, I mean a feature a line feature would be made out of a set of lines as we said point feature would be made out of a set of points, line feature would be made out of a set of points. So, as an example; I mean, we can code the roads in a given city as line feature and record it as a in a layer as a road feature. We can have the boundaries of the villages, states and the different administrative areas in a very hierarchical manner. We can record the natural streams which basically converge into drains, which may again join to form river courses. So, these are some of the examples where in this features could be coded as line feature in the using basic vector data type in a GIS framework.

The next data type is a polygon data type which is a two-dimensional data it has not only the length, it also would have a property of the area as well as the perimeter. So, in addition to the location; I mean, when we are talking about GIS data types these data would always have its the latitude longitude coordinate or it would have the projected coordinate. In our earlier projection; earlier lectures on projections and transformations we have seen how the latitude longitude could be transformed into your cylindrical projections or chronic projections or orthographic projections. So, we have seen different context.

So, these data can either have a code of the latitude and longitude of the place; it is known as a geographic I mean units in a WGS framework WGS 84 framework or otherwise I mean, it can be in a projected coordinate system. So, this polygon it would have area and perimeter as I said in addition to the location aspects I mean, the location coordinates.

So, it is would be made by joining line segments which is closed I mean it is not open. I mean this line segments has to be closed and these line segments has to be non-intersecting line segments that is very important, because if you have intersecting line segments then that would form multiple polygons in that case.

So, when you are trying to create the topology create the relationships and the entities definitions within the GIS database you will see if there are intersecting polygons then it may

have a problem. But, we will see some special cases wherein even intersecting polygons or polygons with cavities inside are also possible I mean we can create such polygons in a GIS framework.

So, when we are looking into higher vector data types; I mean, these are the basic vector data types, when we look for advanced vector data types. We would see that how it could be made possible in a GIS framework to a code intersecting lines or polygons which are I mean which may have holes in it or which may have zones which should not have any area. So, those kind of polygons can also be coded, encoded in the GIS framework.

So, I mean the perimeter or the boundary would define the polygons. So, polygons it could be standalone polygon or it could share its boundary with other adjoining polygons and it may also have a hole within, I mean there could be a area which is not included. Like for an example, I mean if you are doing a land use and you do not want to include a pond comes within a pharmacy and you do not want to include the area of the pond if you are trying to calculate the productivity of that form; I mean, how what would be the output in terms of say tonnage of yield of rice or wheat. So, in that case you would discount the pond area which comes within the enclosure of the field of the agricultural field.

So, it is possible for us to code polygons which also has holes and we can remove that much of an area from the statistics. So, it would in this case where we would have these kind of polygons which are having a embedded holes, it would result in an interior and an exterior boundary. So, I mean this feature would consist of polygons or a set of polygon; I mean, polygon as a feature. And, examples of this what could come to my mind I mean, one such example I told you is a field agricultural field it could include vegetated areas, forested areas, urban areas, water bodies. So, I mean you can have different area based features and you can encode it as a polygon feature in GIS.

(Refer Slide Time: 08:31)

The slide illustrates vector data types in GIS, categorized into Points, Lines & Routes, and Polygon. Each category includes a visual representation on a grid and a corresponding attribute table.

Points

Visual: A grid with four points labeled 1, 2, 3, and 4. Point 1 is at (1, 2), 2 at (2, 4), 3 at (3, 2), and 4 at (4, 2).

Point ID	X	Y
1	1	2
2	2	4
3	3	2
4	4	2

Lines & Routes

Visual: A grid showing a route with segments 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100.

Arc #	From Node	To Node
1	100	101
2	100	102
3	100	103
4	102	101
5	103	102
6	103	104
7	103	104

Polygon

Visual: A grid showing a polygon with vertices 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100.

Arc #	From Node	To Node
1	100	101
2	101	102
3	102	103
4	103	104
5	104	105
6	105	106
7	106	107
8	107	108
9	108	109
10	109	110
11	110	111
12	111	112
13	112	113
14	113	114
15	114	115
16	115	116
17	116	117
18	117	118
19	118	119
20	119	120
21	120	121
22	121	122
23	122	123
24	123	124
25	124	125
26	125	126
27	126	127
28	127	128
29	128	129
30	129	130
31	130	131
32	131	132
33	132	133
34	133	134
35	134	135
36	135	136
37	136	137
38	137	138
39	138	139
40	139	140
41	140	141
42	141	142
43	142	143
44	143	144
45	144	145
46	145	146
47	146	147
48	147	148
49	148	149
50	149	150
51	150	151
52	151	152
53	152	153
54	153	154
55	154	155
56	155	156
57	156	157
58	157	158
59	158	159
60	159	160
61	160	161
62	161	162
63	162	163
64	163	164
65	164	165
66	165	166
67	166	167
68	167	168
69	168	169
70	169	170
71	170	171
72	171	172
73	172	173
74	173	174
75	174	175
76	175	176
77	176	177
78	177	178
79	178	179
80	179	180
81	180	181
82	181	182
83	182	183
84	183	184
85	184	185
86	185	186
87	186	187
88	187	188
89	188	189
90	189	190
91	190	191
92	191	192
93	192	193
94	193	194
95	194	195
96	195	196
97	196	197
98	197	198
99	198	199
100	199	200

Now, let us see some of the examples. So, here you can see an example wherein you see we have an example of points, we have few points whose I mean points are given, whose locations are given. So, we have some number of points out here whose attributes are given this is the idea of the point and we have the location coordinates in terms of x and y.

So, you can see for the first point the location coordinates are given, for the second coordinate point again the location coordinate are given, for third I mean point and the for the fourth point we have the location coordinates. So, this is how the data is basically stored in reference to the points in the GIS data.

So, now, we have an example of the lines wherein you can see we have the from node and we have the to node. So, if you see the segment one. So, in this case, this line segment that is 11 to 12 is your arc ID number 1 which is given over here which is the link ID and then you have

the node ID that is 11 and 12 which is given as these points. So, we have the node wherein we have the from node and we have the to node and then we have the link IDs as arcs.

So, similarly, if we pick up say arc number your arc number 5. So, in this case if we go to this we see the starting point is 14; so, the from node should be your 14. In this case, it is 15 starting node is 15 and to node is 14. So, the reason being this is that the line would have been digitized in this direction that when somebody would have been digitize this line, they would have first clicked on this point 15 and then on point 14. So, this arc 5 has from node is 15 and the to node is 14 in this case.

So, we can have the length of these arcs, we can have the adjacency matrix if we see the polygon coverage. So, in this case we see the polygon coverage. So, you can see the polygon IDs are given over here 100, 101, 103 and 104, but you also see a default ID which is 100 which is beyond the lines that we see in this particular area, this ID is I mean comes as default. So, any field which has a hash after it is the default field created by the software.

So, you can see for the ID 101 that is the polygon which is listed as 101 has the line segments that is one starting from here then line segment 4 which is this to this then we have the line segment 6 so, this line segment is line segment 6. So, this completes the polygon 101 and similarly we have polygon say 103 or 104 and we can find out what are the list of arcs which is enclosing or making this particular polygon.

Now, we can also find out from this that your x-y coordinate of arc 1 is given over here we were talking about line segments. So, you can see that coordinate 1 and 3, 1 and 9 and 4 and 9 are basically creating this particular arc segment 1. So, I mean we you can go to the data and you can refer to it and see how I mean the data is arranged I mean you can create a shape file or arc file in arc GIS or q GIS or other softwares.

So, we will have one lecture wherein I will try to tell you to install the software's and we can do some basic exercises on these softwares. So, we will have a dedicated lecture for that and we will try to work on the QGS platform. So, apart from this you can also see there are two

matrix which is the adjacency and incidence matrix which we are going to talk about in our later slides.

So, in the adjacency matrix you can see that the adjacency of these points are given as a matrix. So, if your point 12; this is point 12 whether it is adjacent to point 11, so, if it is adjacent then its score is given as 1. Now, if you see the point 13; 13 is not adjacent to 11. So, in that case it is coded as 0 wherein the I mean for 14 the adjacency is 1 since you can reach to 12 you can reach to 14 from point 11. But, if you see I mean the adjacency from point 13 you can only reach the point 11 from point 13. So, for adjacency matrix of 13 you can see it is only having the adjacency to point 11 which is this point.

So, now what we can do is if we see the incidence matrix for this I mean the point segments and the line segments. So, the link codes which are given out here that is for point 11 you can see that it is connected to the first segment, but it has a negative connection. I mean, the line of flow direction of flow is negative in this case wherein it is connected to the second arc and it is in the positive direction, it radiates from this point 11. So, it radiates from this point 11 so, it is given as 1.

So, similarly you can reach this point 14 and this is through this particular line segment that is 4; so, again that is a positive value. So, you can likewise I mean, if you are unable to reach it directly if the connection is negative it would be represented as negative one in the incidence matrix. So, we have gone through the different types of data sets that is the polygon data sets, we have the line data sets, we have the polygon data sets and the point data sets.

using lines or say I mean it would result in a polygon, so that geometry should remain invariant. Invariant means when we are doing geo-referencing in our last class we had done the transformation. We had done the transformation and we have seen the polygon or polynomial or affine transformation in which we are basically trying to change the geometric attributes the values of the points.

So, in that case it may undergo warping, it may undergo a deformation in terms of the change of shape. So, we had seen shearing, rotation, translation these are different things that may happen to a database. So, it could be a image, but today we are talking about vector data. So, when we are doing a transformation on the vector layer what happens is the, there may be a deformation in terms of the geometry, but the relationship the properties of the geometric objects they remain same. I mean the road the name of the road would not get changed if you are doing a geometric transformation that needs to be preserved.

So, in a topology when we have a topology, this properties of the geometric objects it would remain invariant. It would not change under any such transformation such as bending or stretching, you can see the example of a rubber band; I mean, we use it for time.

So, if you stretch the rubber band I mean imagine it to be made out of lines so, you would see that the entities would be remaining the same. The line segments would be the same may be the length increases or it gets deformed or if you want to if you make a polygon out of it may become deformed, but the entities remain. The geometric objects would remain invariant even if you apply some kind of a transformation such as stretching or I mean skewing or any such transformation.

So, this I mean relationships this property's could be explained through directed graphs, it is also known as digraphs. So, this is aspect of graph theory. So, I mean, the directional nature can be expressed through the digraphs that is the directed graphs which shows the arrangement of the geometric objects.

In our earlier slide, we had seen the lines in which the directions were there. As you start digitizing it, it also takes the direction which is encoded in the to point, from point and the to

point. So, those relationships could be encoded and I mean these relationships could be embedded as a table in a table in different columns. So, this basically is the topology or the relationships that we build with the points and the lines.

Next, we had talked about the adjacency and indices incidences. So, these are also fundamental relationships that are extensively used in GIS, wherein it would establish a relationship between the nodes and arcs. So, in digraph topology it would require additional files to store the spatial relationships. When we have this spatial relationship we have seen that adjacency and incidence matrix where two separate matrix, two separate data sets apart from the basic data sets of the location, length or area of the line point or the polygon.

So, these are two additional data sets which basically stores your adjacency and incidence data sets incidence information. So, we have already seen this example of the polygon topology and we had explained what is adjacency matrix; adjacency matrix and a incidence matrix in our earlier slide and how they are connected and how they are encoded in this particular table.

Now, let us see this topological relationships is also useful if we want to do some kind of a spatial query. If we you are doing some kind of a spatial data query say suppose you want to search for a road and you see if you are travelling somewhere you open the Google maps and you can see how which segment of the road is congested and what is the travel time. So, it is shown through colors indicated through colors red indicates that there is a traffic jam, the traffic speed is very less in a particular segment, if it is green or blue then it indicates there is a smooth flow of traffic in that particular line segment.

So, I mean we need to do some queries, I mean these are queries which are self generated, but you can also give a specific query say suppose what is the length of a particular road. I mean you may have a particular road which may be a express way or I mean a highway part of the highway running through a city. So, you may want to know its length. So, you can find out using a spatial query.

So, when you have these topological relationships it helps or it aids in creating these spatial queries like say suppose, I mean there is a case of say kind of pollution. So, how much or how many a buildings are I mean, affected or how much population is affected because of this pollution. So, you can try to use measures of containment or intersection. So, these are very important two topological relationships important for the spatial data query.

So, we shall subsequently see in our decodes of time in our next lectures series of lectures that how we can generate a spatial data query and how it is useful.

(Refer Slide Time: 23:55)

Vector Data Structure

GeoRelational Data Model

- ❑ Stores geometries and attributes separately in a split system: geometries ("Geo") in graphic files in a spatial subsystem and attributes ("Relational") in a relational database
- ❑ Uses the feature identification number (ID) to link the two components
- ❑ Two components are synchronized so that they can be queried, analyzed, and displayed in unison

Examples - coverage and shapefile

- ❑ COVERAGE is TOPOLOGICAL, and the SHAPEFILE is NONTOPOLOGICAL.

Graphic Files

Polygon/arc list
Arc-coordinate list
Left/right list
...

INFO File

PolygonID	Field 1	...
1		
2		
3		

ArcInfo coverage has two components: graphic files for spatial data and INFO files for attribute data. The label connects the two components

So, we would look into the georelational data model. So, it stores geometry I mean as geo I mean which signifies the graphic aspects I mean, how the data is coded as lines or polygons and the relation that we build into this model which is the database part that is the attribute part. So, it uses feature identification IDs to look the to basically link this two components

that is the geometric component and the attribute component. We have the attribute component as well as the geometric component.

So, we can link these two components. So, these two components are against synchronized; I mean, these are two separate sets. Like you if you are some of you have worked on auto CAD you would see that you can draw lines, you can draw polygons, you can draw shapes, you can draw points. So, these are only a drawing entities I mean it does not have any attribute and on the other hand you can have a excel table wherein you may have attributes. So, these two information sets could be having a unique tag or a unique ID which would link these two components that is I mean, geographic or the geometric component and the attribute component.

So, these are synchronized so that they can be queried, analyzed or displayed together. So, the examples of such georelational data models are COVERAGE and SHAPEFILE. So, COVERAGE is a very popular Esri I mean format and SHAPEFILE is a open source format, we will have a look at it. This coverage is a topological I mean model wherein a shape file is a non-topological model. We will again discuss what is a, I mean difference between coverage and shape file when we look into this two aspects we shall see.

So, you can see the coverage has two component arc coverage has two component. So, you can see it has it is linked the geo that we had talked the geometry that we had talked this is the geometry part and then we have the info file which is basically the attribute. So, they are linked with two sets of I mean entities, I mean you can see a common field having the entity numbers which link this geometry as well as the attribute.

So, whenever we said we are doing some kind of a query or we are doing some kind of I mean analysis they are these two components. These two components that is the geometry and the attribute of this particular system, that is why we are calling it a GIS, Geographical Information System which has these components. So, these are synchronized and they can be queried analyzed or displayed together displayed in unison.

(Refer Slide Time: 27:19)

GeoRelational Data Model

Esri Coverage

Coverage supports three basic topological relationships

- Connectivity: Arcs connect to each other at nodes.
- Area definition: An area is defined by a series of connected arcs.
- Contiguity: Arcs have directions and left and right polygons.

Starting point of an arc is the from node, and the end point is the to-node

data structure of a point coverage

ID	X	Y
1	2.0	9.0
2	4.0	9.0
3	2.0	7.0
4	4.0	7.0

data structure of a line coverage

Arc ID	From Node	To Node
1	1	2
2	2	4
3	1	3
4	3	4

Arc ID	From Node	To Node
1	1	2
2	2	4
3	1	3
4	3	4

data structure of a polygon coverage

Polygon ID	Arc ID
1	1, 2, 4, 3
2	3, 4, 2, 1

Polygon ID	Arc ID
1	1, 2, 4, 3
2	3, 4, 2, 1

The slide also includes a grid diagram with points 1, 2, 3, 4 and their respective coordinates, and a diagram showing arcs connecting these points to form two adjacent polygons.

So, let us see the Esri coverage. This is a proprietary format given by Esri and the connectivity is through arcs and nodes. So, there is a area which is defined for this polygons when the arcs are connected together to enclose area and they are referred to as polygons. And, we also can identify the contiguity; I mean, we can identify whether area is lying close by. So, we can find out which are the left polygon and the right polygon that is there in the database.

Now, you can see the data structure for a point we had already studied this. So, you can know the coordinates from this I mean graph. So, for point 1; for point 1 we have coordinates 2 and 9. So, in the x direction we travel two grids and in the y direction we travel two grids to have this point 1 which has the coordinates 2 and 9. So, this is how the point data structure is

linked, then we have the data for the line coverage in which again you can see that we have the arc ID. So, you would have arc ID and you would have the from node and the to node.

So, we have already covered this. So, we have the from node and the to node. So, this we have the from node as 11 and to node as 12. So, similarly we can also have your polygon coverage wherein we have the left polygon and we have the right polygon. So, we had talked about the contiguity, in this case we had talked about the left and the right polygons. So, you can see that is encoded in this particular data set the left and the right polygon.

So, for arc 1 you have the left polygon which is 100 which is the outer area and the code for the outer area and for the right polygon you have the right polygon as 101. So, this is your arc 1. So, similarly you may also have a polygon arc table polygon table in which you can find out which are the arcs which enclose to make this polygonal area. So, it has arc 1, 4 and 6. So, you have 1, arc 4 and you have arc 6 to make your polygon 101. So, this is how the Esri arc coverage would store the data.

(Refer Slide Time: 30:15)

GeoRelational Data Model

Shape file

Standard nontopological data format used in Esri products

- ❑ Shapefile treats a point as a pair of x-, y-coordinates, a line as a series of points.
- ❑ Polygon as a series of line segments.
- ❑ Spatial relationships among these geometric objects are not described in any file
- ❑ Polygons actually have duplicate arcs for the shared boundaries and can overlap one another

Advantage of using Shape files

- ❑ They can display more rapidly on the computer monitor
- ❑ Nonproprietary and interoperable
- ❑ Interoperability - Open Geospatial Consortium Inc. 1994 (<http://www.opengeospatial.org/>)

❑ **.shp** file stores the feature geometry, and

❑ **.shx** file maintains the spatial index of the feature geometry

NPTEL

6

Now, next let us see how the shape file is different from the Esri arc coverage. Now, shape file it would it is a non-topological format, it is open source format and it is extensively used in a multi in a range of multitude of GIS softwares, GIS platforms. So, this shape file it treats a point as the x-y coordinates; as the x-y coordinate that we had seen in arc in the Esri coverage file. So, it has a line would have series of points, a polygon would have series of line segments and the difference is that this polygons would have duplicate arcs when they have sheared boundaries which is not so in case of the arc coverage that we had seen earlier.

So, we can see that there are two types of files in this. One is your shape file and it is coded as shx file, shp file and another is a shx file. So, shp file it I mean records the geometry wherein shx maintains the spatial index. We had talked about the spatial index in the last slide; so, it maintains the spatial index of the feature geometry.

So, there are few advantages of using shape files they are I mean, they basically can be very rapidly displayed on a computer monitor. So, it is very useful when a user is trying to only view some data sets. So, in that case it is easier if we use shape file as they are they can be displayed quickly. These are non-proprietary formats actually there was a initiative in 1990 and there was a demand for having non-proprietary open source GIS data file.

So, this format shape file is the result of that initiative and this consortium is known as the open geo spatial consortium which came up in 1994. So, it basically I mean is very; I mean, it presses on interoperability of the data sets, so that I mean you can use it in multitudes of platforms. So, if you go to this particular website you can find out the details regarding open geo spatial consortium.

(Refer Slide Time: 33:03)

GeoRelational Data Model

GIS operations are affected by

- PROPERTY** describing an attribute or characteristic of an object – shape, extent
- METHOD** performing a specific action – copy, delete

Object Based Data Model

Standard nontopological data format used in Esri products

- ❑ Stores geometries and attributes in a single system.
- ❑ Geometries are stored as a collection of binary data in a special field with the data type BLOB (binary large object).
- ❑ Spatial feature (object) to be associated with a set of properties and methods

Example of a Landuse

Object ID	Shape	Landuse_ID	Category	Shape_Length	Shape_Area
1	Polygon	1	5	14,6077	5,959,800
2	Polygon	2	8	16,979.3	5,421,216
3	Polygon	3	5	42,654.2	21,021,728

geometry of each land-use polygon

NPTEL

7

Now, we have the object based data model. So, this is again a standard non-topographical data format it is also used in Esri products. It is format from Esri, but it is I mean it is a non-topological proprietary format. So, it stores geometries and attributes in a single system. Unlike earlier data bases like shape file wherein the geometries and attributes are stored in different systems with the unique ID relating both of them this would store the geometries and attributes in a single system.

So, the geometry is a stored as collection of binary data and it is in a specific field which is known as a BLOB; I mean, this is known as binary large object. So, it is abbreviated as BLOB. So, I mean there are spatial feature or objects which are associated with the set of properties and methods. So, you can see here that the geometry of each land use is coded here and the shape is given here it is given as polygon for the land use ID. So, this is how your object based data model is stored.

Now, they are affected by the property which describes the attribute or characteristics of the object, I mean your GIS object it could be the shape or the area that is the extent or the method like I mean, performing a specific action such as copy or delete. So, these GIS operations are affected by the property and method which are encoded in the object based model.

(Refer Slide Time: 35:01)

The slide is titled "GeoRelational Data Model" and features a list of characteristics for a Geodatabase. The background includes a stylized tree with various icons (gear, lightbulb, Wi-Fi, smartphone, etc.) and a blue atom symbol. A small video inset of a presenter is visible in the bottom right corner. The NPTEL logo is in the bottom left, and a navigation bar with a slide number '8' is at the bottom.

GeoRelational Data Model

Geodatabase

- ❑ Uses points, polylines, and polygons to represent vector based
- ❑ Similar to arc coverage in simple features
- ❑ Differ in the composite features of regions and routes
- ❑ Geodatabase can also store raster data, triangulated irregular networks, location data, and attribute tables

Now, talking about the next geo relational data model which is the geo database this is again by Esri, it uses points, lines and polygons to represent the vector data. It is very similar to arc coverage in terms of its simple features, but it differs from arc coverage in terms of composite features as you can code roots or regions which we shall discuss in our next few slides. So, it can also store raster data, it can also store triangulated irregular networks, it can also store location data.

(Refer Slide Time: 35:41)

The slide, titled "GeoRelational Data Model", is presented by NPTEL. It defines a Geodatabase as a system that organizes vector data sets into feature classes and feature datasets. Feature classes are described as storing spatial features of the same geometry type and participating in topological relationships with one another, such as coincident boundaries between different levels of census data (village, taluka/block, district, state, country). Feature datasets are defined as storing feature classes that share the same coordinate system and area extent. The slide includes a small video inset of the presenter in the bottom right corner and a navigation bar at the bottom.

So, the geo database, vector data bases is organized into feature classes and feature data sets. So, feature classes would store the spatial features of a similar type of geometry, if you have point. So, for different points it would store the spatial features in the feature class. And, it would participate in the topological relationship with one another, that is, for an example I mean if you have say coincident boundaries like, if you have census data which is very hierarchal in nature wherein you may have village, you may have taluka blocks or districts, state and the country boundaries. So, there could be I mean coincident boundaries between this different scales of data.

So, this geo database I mean participates in the topological relationship which is wherein you have coincident boundaries. So, the feature data set, I mean we had talked about the feature

data set. It stores the feature classes and that would share the same coordinate system and the area extent.

(Refer Slide Time: 36:57)

Vector Data Structure

Triangulated Irregular Network (TIN)

- Approximates the terrain with a set of nonoverlapping triangles
- Each triangle in the TIN assumes a constant gradient
- Each node of a triangle is a point, and each edge of a triangle is a line
- Data structure therefore includes the triangle number, the number of each adjacent triangle, and data files showing the lists of points, edges, as well as the x, y, and z values of each elevation point

data structure of a TIN

Triangle	Neighbors
101	102, 103
102	101, 103
103	101, 102

Triangle	Neighbors
101	102, 103
102	101, 103
103	101, 102

So, next we had talked about the triangulated area network in our earlier slide in the geo database. So, we can see what is the triangulated irregular network is; it is used to code the undulating nature of a terrain. If you have a hilly area you can code the slope, you can code the height, you can code the terrain information using this particular type of vector data structure which is known as TIN; Triangulated Irregular Network.

So, these are basically a set of non-overlapping triangles and each triangle would have a constant gradient. The slope in each of this triangle would remain same throughout that triangle, it would not change. So, you can see, you can say that the triangle would be of a

nature of equiplanar triangle for the points lying in a triangle, they would be equiplanar. They would be lying in the same plane; it would have a constant gradient.

Now, the node of the triangle is a point and each edge would be a line and the triangle itself would be a polygon. So, that is how the data is structure. So, it has a triangle number, the number of each adjacent triangle, the data would have the list of points, edges that is the points the lines as well as the x, y and z value of the elevation points. So, this you can see how triangulated irregular network looks like.

So, if you have low lying area where there are very less undulations in terms of your height information, you can see the triangles would be much bigger than areas where you would have frequent change in the elevation values, in the height values. So, there you can see the triangulated facets would be much dense compared to the plain areas where the size of the triangles are much bigger.

So, you can see the structure of a triangulated irregular network. So, we have the nodes that is node 11 and it would have the elevation value that is the I mean, height value and it would contain the x and y reference with respect to the coordinate frame that is 2 points along the x-axis and 9 units along the y-axis. So, this is where your node 11 is stored. Now, this triangle 101 is comprised of these three nodes that is node 11, node 12 and node 13.

So, you can see these three nodes which are listed over here in the node list and you can see the adjacency also is given over here. So, we had talked about the adjacent triangle in the data structure. So, you can see the neighbors the in the list of neighbors you have 102 as the adjacent neighbors, and if we see the entity 102 it would have two adjacent members that is 101 and the other member being 103. So, here you can see the neighbors are 103 and 101. So, this is how we would be storing data in a GIS system with respect to the triangulated irregular network system.

(Refer Slide Time: 40:33)

Vector Data Structure
Regions

- ❑ Set of geographic areas with similar Characteristics
- ❑ Two or more regions can cover or overlap the same area
- ❑ Can include areas that are spatially disjoint
- ❑ Regions are organized as subclasses in a polygon coverage
- ❑ Additional data files, regions are related to the underlying polygons and arcs

Region-polygon list

Region #	Polygon #
101	11
101	12
102	12
102	13
102	14

Region-arc list

Region #	Ring #	Arc #
101	1	1
101	1	2
102	1	3
102	1	4
102	2	5

Data Structure of a Region Subclass

(a)

(b)

Hierarchy of administrative units

11

Now, let us see how the regions are encoded. They are geographic areas and they have similar characteristics, they may cover or they may overlap the same area. You can see this particular example wherein they are overlapping; this polygons are overlapping each other. So, they have similar characteristics, but they are overlapping or they may cover each other. So, they can also include areas that are specially disjoint.

So, in this case you can see this polygon 1 which is encoded as 1, this value as 1. So, your polygon these three polygons that we have in the second case that is b are not I mean spatially joint, I mean they are separate, they are spatially disjoint. So, but they still have the same entity. So, this is what a region is wherein they include areas which are spatially disjoint.

So, regions are organized as sub classes in polygon coverage. These regions you would have a polygon classes and you would have a separate region class where they would be sub classes

in the polygon coverage. So, if you are making say a land use data set so, you may have different types of land use say forest, water body, urban built up areas. So, you may have disjointed polygons, but they may have same data value like a polygon which belongs to say residential, it may be disjointed. So, it can be organized as sub classes in the polygon coverage and it would be a region data structure. It is encoded as a vector data structure which is region.

Your additional data files would be generated and these regions are related to the polygons and the arcs. So, you can see that how they are we were talking about the hierarchical data; so, how the regions are related. So, you can see these two examples out here, the case of you are I mean disjointed polygons having similar value sets or overlapping polygons which are having different values of classes that is 11 and 13, but they may be overlapping. So, they are encoded in the list of table in the region arc table or the region polygon table. So, they are encoded in the data set. So, this is how we store the region data set.

(Refer Slide Time: 43:27)

Vector Data Structure

Routes

- Linear feature
- Stored as subclasses in a line coverage
- Collection of sections
- It has a from position (F-POS) and a to-position (T-POS)
- It has a from measure (F-MEAS) (the beginning point of the route) and a to-measure (T-MEAS)

Both the shapefile and the geodatabase use polylines with *m* (measure) values to replace route subclasses for GIS applications

Route ID	Section ID	Arc ID	F-MEAS	T-MEAS	F-POS	T-POS
1	1	7	0	40	0	100
1	2	8	40	170	0	100
1	3	9	170	210	0	80

Data structure of a route subclass

Route is shown here as a thicker, gray line, is built on a polyline with linear measures in a geodatabase

Next, data type is the roots wherein we have the linear feature we which stores the subclasses as line coverage. I mean we have talked about line coverage. So, in that when we have similar classes in line coverage, we can store it as a subclass and call it as root. So, we have collection of sections. So, it has from position so, which is the F-POS and the T-POS or the to position. So, it also has a measure value which is the F-measure and the T-measure T MEAS so, which give the measure of the I mean these routes.

So, your shape file and the geodatabase, they use polylines with measure values and this can be use to replace the route subclasses in the GIS application. So, in the here you can see an example of a route wherein this buffer area that is the grey. I mean, the route is shown as a thicker grey line in this case this particular thicker gray line which incorporates this polylines

or lines which is one the line segment 1, the line segment 2; the line segment 2 and the line segment 3.

So, we see that there are this the link numbers are given out here and the distance values are also given. So, you can see up to this it measures 40, up to this it measures 40 to 170. So, you can see the from measure and the to measure are given out here 40 to 170 and for the third component it is from 170 to 210. So, we have the section IDs; we have the section IDs that we have I mean, highlighted over here so, which gives the route. So, new data is created out of the line data set as a subclass to the line coverage which is stored as a separate feature and which is known as root.

So, this is extensively helpful when we are doing transportation analysis in urban area, if we are running say some of the transportation models trying to find out the shortest route between the different road segments in a given area, if you want to reach a point from your origin, you can find out the shortest route. So, in that case the route would be a subset or a subclass of the line coverage, the different line segments that you have in the data set.

(Refer Slide Time: 46:29)

Recapitulation

- Vector Data Types
- Topology
- GeoRelational Data Model

NPTEL Online Certification Course

So, for recapitulation to end this particular lecture, we have talked about the different vector data types, the basic vector data types then we have seen how we can build or encode a relationship to link. This basic vector type of point to create lines and from lines, to create polygons enclosed polygons and even higher data types such as triangulated irregular network or we can create roots. So, we have also seen what are the different types of geo relational data models; so, I mean, this is what we have covered today.

So, in our next class we would be looking into the raster data type which is another data modal which is extensively used for modeling in the GIS environment.

Thanks for your patient hearing.

