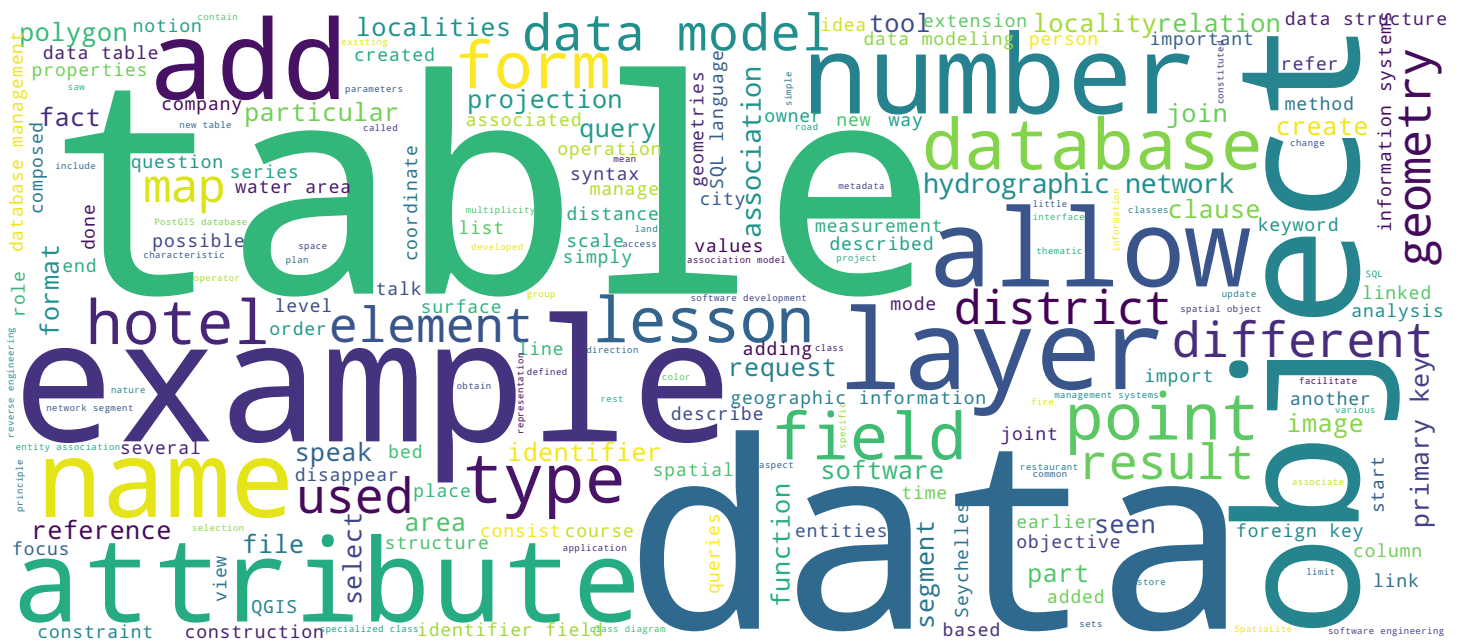


Data modeling

Stéphane Joost, Marc Soutter, Fernand Kouamé, Amadou Sall



Data modeling



Objectives of the lecture

- To highlight the added value and significance of data modeling,
- To describe the methods and tools used in data modeling

After this lecture you will be able

- To create or interpret a class diagram,
- To use data structure modeling and creation tools

An Introduction to Geographic Information Systems

Welcome to this lesson which will focus on data modeling, which is an essential step in the construction of relational databases. As it has just been suggested by my friend Amadou, the organization and structuring of data becomes something very important the moment a database expands and becomes more complex. It is therefore quite useful to be able to refer to a schematic representation of this organization, so to refer to a data model, the theme of this lesson. The objectives of this lesson consist in highlighting the inputs and the importance of data modeling and describing the tools which are used in this context, so that at the end, you are able to construct or interpret the class diagrams and to use the modeling and creation tools of data structures.

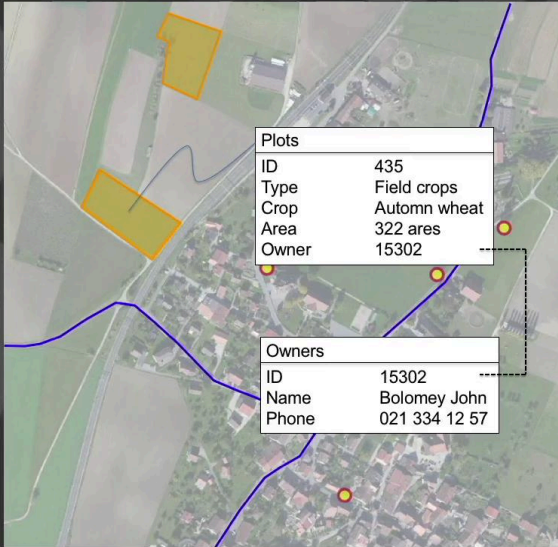
Notes

Summary



0m 23s

Role of modeling



Data organization

- Each plot has an owner
- Several plots >> same owner
- Repetition of the details related to the owner >> risk of errors during updates
- Separate plots' data and owners' data and create a relation through an identifier

➔ Need to organize data

Data modeling =
Schematization of data organization

An Introduction to Geographic Information Systems

In this lesson, we will address successively the role of modeling, then, more specifically, the entity-association model, before taking two examples of construction of a data model and finally, we will describe a bit the software engineering tools. We have seen in a previous lesson this example of agricultural plots that have an owner and we spoke about the need to bring out the specific information of the owner to place it in a separate table, so as to avoid data redundancy. And the general conclusion of this example was the necessity to structure the data in a file or in a database. So when we talk about data modeling, we talk about a schematization of the data structuring.

Notes

Summary



1m 22s

Role of modeling

- Modeling is not only about data but may also extend to other aspects of the design of a geographic information system

e.g. expected functionalities

- Not specific to GIS but belonging to the more general context of software engineering, from which it borrows methodologies and tools



Modeling is not only about data, but also about other aspects of geographic information systems particularly the expected functionalities of such a system. Likewise, the modeling approach is not specific to geographic information systems, but is part of the broader framework of software engineering, of software development, from which it borrows methods and tools.

Notes

Summary

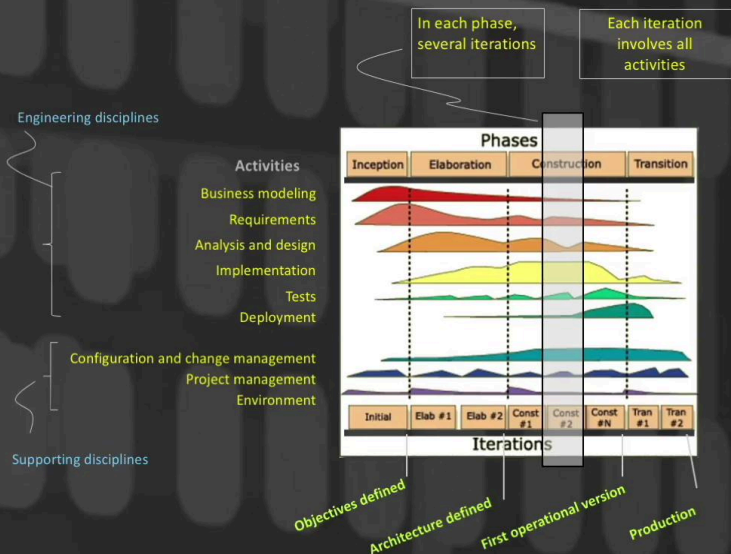


2m 21s

Role of modeling

Software engineering methodologies

- Merise
 - RUP – Rational Unified Process
- Iterative method



An Introduction to Geographic Information Systems

From a methodological point of view, it is possible to refer to the Merise method which is a sequential method which dates from the 70's-80's where we address successively the different phases of software development starting from strategic planning, preliminary, detailed study, development of the solution, implementation, and maintenance to finish. The RUP method, for Rational Unified Process, is a more recent alternative method which has the particularity of being an iterative method. This method is based on the realization of a number of activities in the various phases of a software development which goes from inception, to elaboration, to construction, to end with a transition phase. Each of these activities is treated with a varying intensity in the different phases, and in each phase, in an iterative manner, that is to say that typically, the modeling that interests us here will be performed at the start of the project but reviewed, corrected, adapted constantly, during the later phases.

Notes

Summary



2m 49s

Role of modeling

Software engineering methodologies

- Merise
- RUP – Rational Unified Process
- « Agile » methods
- Sequential approach and/or stiff frameworks are poorly adapted to the constraints and requirements of fast evolving systems
- **Adaptative development**
- **Focussed on production, continuous testing and validation of an integrated solution**
- 12 «agile» principles
- https://en.wikipedia.org/wiki/Agile_software_development

An Introduction to Geographic Information Systems

These two methodologies, Merise and RUP, present the default of providing a relatively rigid framework, which is often perceived as poorly adapted to constraints and to system requirements which we would like them to evolve rapidly and it is for this reason that the methods called "agile methods" are increasingly taking over the methods that are oriented towards an adaptive development, where the focus is on production, validation and continuous testing, of an integrated application. The agile methods are based on a dozen principles which are described in the agile manifesto accessible through Wikipedia.

Notes

Summary



3m 59s

Role of modeling

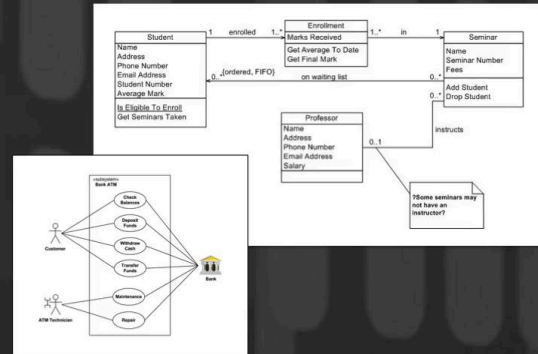
Software engineering methodologies

- Merise
- RUP – Rational Unified Process
- « Agile » methods

- User identification
- Missions or tasks
- Data and functionalities

Common issues

- Needs assessment
- Formalization in a schematic representation or a model



An Introduction to Geographic Information Systems

All these methods have in common a number of elements, in particular to rest on the analysis of needs, analysis of needs which consist in identifying future users of the application that is being developed. For these different users, identifying the missions they must perform using the software in question, and for these different missions, make an inventory of the data used and the functionalities that will be needed. The results of the analysis of needs are generally formalized in the form of a schematic representation or in the form of a model.

Notes

Summary



4m 37s

Role of modeling

The data model

- Conceptual
 - ➔ Drawing, more or less approximate schema
- Logical
 - ➔ Tables, attributes, domains, formats
- Physical
 - ➔ Effective implementation in a DBMS



on Systems

This data model must allow to master the thematic complexity, it is therefore a tool for reflection, it should also facilitate the exchanges of opinions and ideas, it is therefore also a communication tool. It must allow to improve the software development process itself by pointing out errors or obstacles, it is therefore also a technical tool. And finally, it must allow to set up a follow-up, a maintenance, in the long term, of the application that is being developed, it is also a documentation tool. The data model often appears at the start in the form of a simple sketch sketched on a table corner, or a small drawing scribbled on the blackboard, before being presented in a more elaborate form. In this first case, we speak of a conceptual model since the objective, is to bring together the concepts and ideas into one synthetic vision. In a later phase, this conceptual model must be translated in database language, in the form of tables, with attributes in these tables, validity domains for these attributes, and formats, is it a text, numbers and similar things, and we speak in this case of the logic model. And finally, when this logic model is implemented in a real database management system, we talk of a physical model.

Notes

Summary



5m 17s

The entity-relationship model

Terminology

- An **entity** is an object



George's blue car

An Introduction to Geographic Information Systems

In the entity-association model, objects are entities, so that in the case of Georges' blue car, we have two entities: the automobile entity and Georges' entity.

Notes

Summary

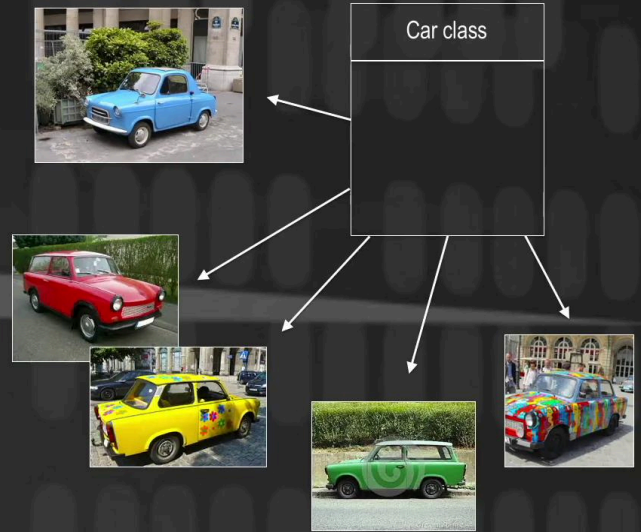


6m 55s

The entity-relationship model

Terminology

- An **entity** is an object
- A set of entities of similar nature, i.e. a type of entities, form a **class**
e.g. the class «car»
- Entities are thus **instances** or realizations of an abstract «parent» class



An Introduction to Geographic Information Systems

A set of entities of the same nature, so a type of entity, constitutes a class. For example here the class of automobiles which Georges' blue car belongs to. In other words, the entities are instances or achievements of an abstract parent class.

Notes

Summary



7m 17s

The entity-relationship model

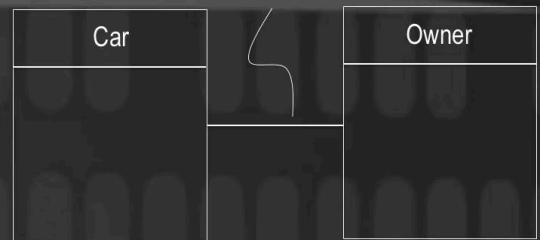
Terminology

- An **relationship** (or relation) is initially a **link between two entities**
- By extension, the word relationship refers in fact to **a class of relationships**, this is a set of links between pairs of entities, instances of two **entity classes**
- **Data modeling** is about **classes of objects** (entities) and their **relations** (relationships)
 - ➔ **Class diagrams**

Relationship : George owns the blue car



Relationship class «own»



An Introduction to Geographic Information Systems

An association is basically a link between two entities. "Georges is the owner of the blue car" is an association between Georges and the blue car. The association or the class of association "to be the owner of" is therefore a type of association which connects an owner to his automobile. Data modeling is about classes of objects, the entities, and their relations, in other words, their associations. This is why we often speak of class diagrams.

Notes

Summary



7m 34s

The entity-relationship model

Characteristics of a relationship

- Name



An Introduction to Geographic Information Systems

A class is characterized by its name, "automobile, owner, wheels, repair" and by its properties, in the case of the automobile, the type, the make, the color, etc. Operations or responsibilities can be sometimes added to these main characteristics. In the case of associations, the characteristics are the name, for example "working for" or "being recruited by" in the case of a relation between a person and his business, or "voting for" when a citizen votes for another citizen in the case of an election.

Notes

Summary



8m 06s

The entity-relationship model

Characteristics of a relationship

- **Name**
- **Termination**
- **Role**
- **Multiplicities or cardinalities**
- **Navigability**



Number of objects that may be in the position defined by the relationship termination

1..1 = exactly one
 0..* = many
 1..* = at least one
 1..6 = from one to six

1,1 = exactly one
 0,n = exactly one
 1,n = at least one
 0,1 = possibly one

An Introduction to Geographic Information Systems

The endings, so the extremities of the association, and in particular the notion of role, the person who is employed in a company and the company as the employer of the person, the point as the summit of a polygon which is a geometry, etc. The multiplicity or the cardinality, so the number of objects susceptible to occupy the position defined by the endings of the association. In the case of a polygon defined by its summits, we see that the multiplicity allows to express the fact that at least 3 points are required to define a polygon. Multiplicity and cardinality express more or less the same thing, as we can see from these examples, which are the most classic examples, they are simply used alternatively in different data modeling languages. The navigability expresses the fact that an association can be crossed in one direction and not in the other, in this case an elector votes for a candidate and not the reverse but the association "being elected by" would be valid and navigable in the other direction.

Notes

Summary



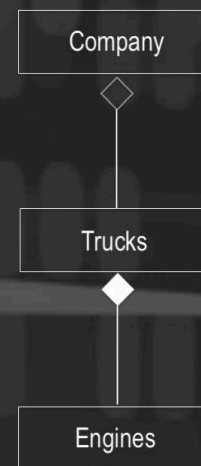
8m 43s

The entity-relationship model

Relationship special case – aggregation

A **whole vs. parts** relationship where one class is a larger element – **whole** – made up of smaller elements – **parts**

- Simple aggregation
The parts remain if the whole disappears
- Composite aggregation or composition
The parts disappear when the whole disappears



An Introduction to Geographic Information Systems

There are various particular cases of associations, in particular, the relations between a larger element, the "all", which is composed of smaller elements which would be the "parts", and we speak of simple aggregation when the parts subsist if the "all" disappears, which would be the case for trucks belonging to a company, the company can disappear, the trucks remain and we talk of composite aggregation or composition if the parts disappear when the "all" disappears, which would be the case of a relation between engines that are part of a series of trucks, when the trucks disappear, the engines also disappear.

Notes

Summary



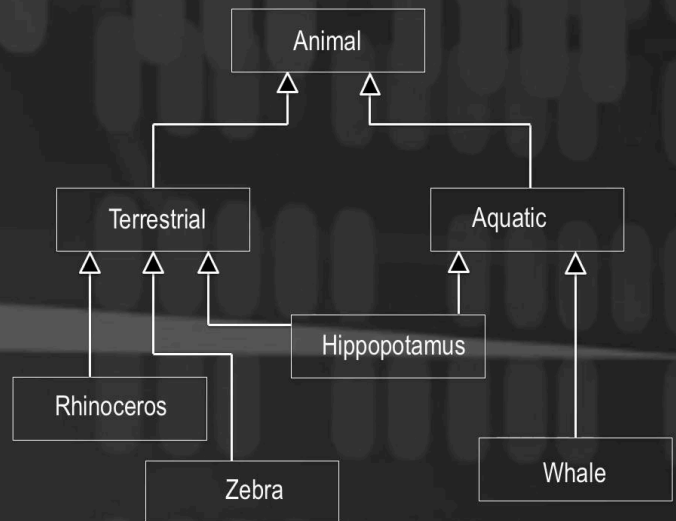
10m 01s

The entity-relationship model

Relationship special case – generalization-specialization

A relation between a **general** (base or parent) class and a **specialized** class (sub-class).

- The specialized class extends the base class with **additional** properties
- An object of the specialized class may be used wherever an object of the base class is valid
- A class may have several parents, which is referred to by **multiple inheritance**



An Introduction to Geographic Information Systems

The generalization-specialization relation, which associates a general class, basic or parent class, and a specialized class or subclass, for example the "animal" general class which includes the "terrestrial animals" and "aquatic animals", in the "terrestrial animals" rhinoceros, zebras, hippopotamuses, etc. The specialized class takes up the basic class by adding additional attributes. An object of the specialized class can be used everywhere where an object of the basic class is possible. And finally, a class can have several parents then we speak of multiple inheritance.

Notes

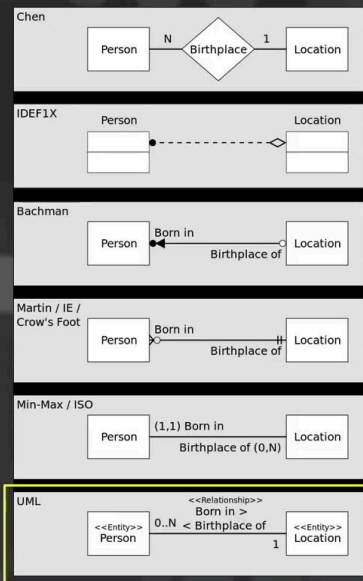
Summary



10m 43s

The entity-relationship model

- Variety of languages



UML – Unified Modeling Language

An Introduction to Geographic Information Systems

The entity-association model has undergone many changes and several variants of languages with syntaxes, different schematisms allowing to always expose the same idea. Here, the example of the relation between a person and the place of birth. One of the most used languages today is the UML language, for Unified Modeling Language, which is a description language of databases in the form of a class diagram as we have seen so far, but which are associated with a whole series of other diagrams in particular use case diagrams, activity diagrams, sequential diagrams, to mention only the four most important ones.

Notes

Summary

11m 22s



The entity-relationship model

- Several ways to model relationships



An Introduction to Geographic Information Systems

The association can be presented, as we have seen, in a very synthetic way by linking two classes with a certain number of properties, of characteristics, but it can also take the form of a closer representation of the implementation in a database where we would have for example an attribute describing the company in which the person works and reciprocally, at the enterprise level, an attribute that would refer to a vector, to a list of people who would be employees.

Notes

Summary



12m 10s

Some examples

Geometry class hierarchy according to OGC (1999)



Building a data model consists of defining and schematically representing the classes, the objects which will be used to describe a thematic area and showing how these different objects, these different classes, are linked to one another. As a first example, We will interest ourselves to the hierarchy of the geometry classes which are used to describe the spatial objects in geographical information systems.

Notes

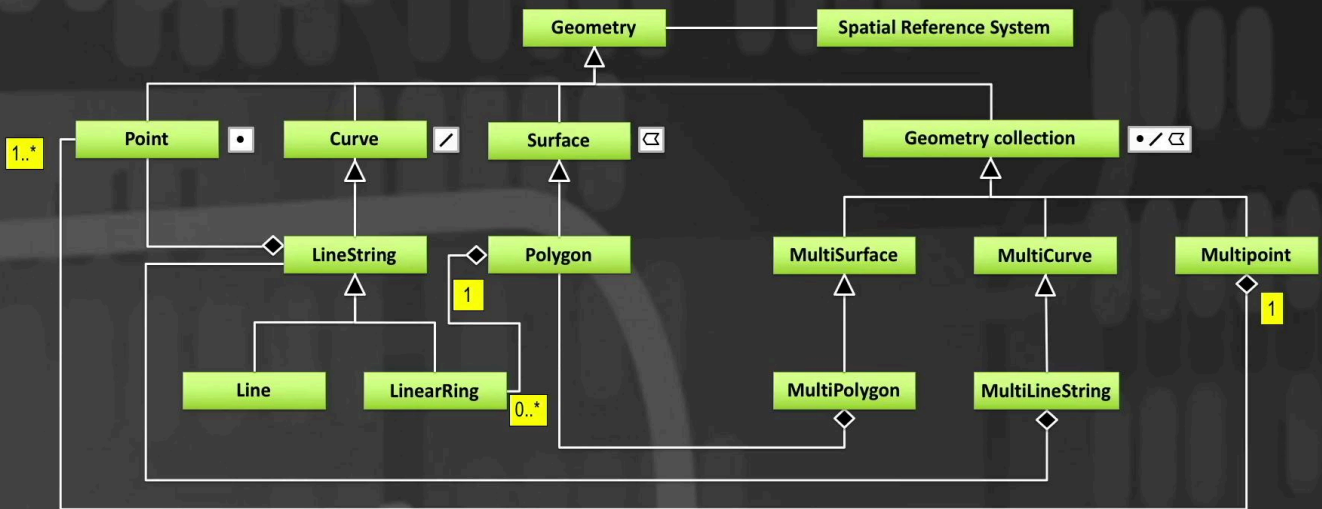
Summary



12m 46s

Some examples

Geometry class hierarchy according to OGC (1999)



An Introduction to Geographic Information Systems

The starting object here is the geometry which is associated with a spatial reference system. This geometry can be specialized in the form of points, curves or surfaces, or even in the form of geometry collections. These objects can be described by pictograms, which facilitates the subsequent reading of the class diagrams. The curves can be described as polylines which are themselves constituted by a set of points. The surfaces can specialize in polygons. The geometry collections can specialize in set of points, curves or surfaces and in the latter two cases, in sets of polylines or polygons. Multipoints, multipolylines and multipolygons being sets of points, lines and polygons. The polyline still has two types of specializations: the simple polyline and the closed polyline, the latter being a constituent element of the polygon. By adding the multiplicity, we can still express the fact that a multipoint for example is constituted by at least one point, or that a polygon is constituted of a border.

Notes

Summary



Some examples

- Water resources management GIS



An Introduction to Geographic Information Systems

The second example is about the construction of the data model of a geographical information system allowing the management of water resources.

Notes

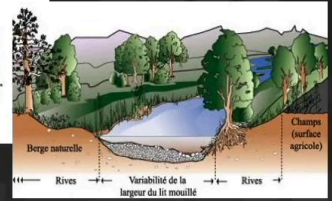
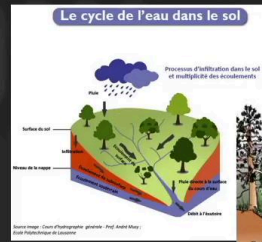
Summary



14m 39s

Some examples

• Water resources management GIS



An Introduction to Geographic Information Systems

Amongst the different data concerned there are, of course, rivers, lakes, the catchment basins for the hydrological dimension, the stream cross-sections for the hydraulic dimension, with the idea of comparing the flows produced by the rain and the flows that streams are capable of conveying. All that is sample must be added to this through technical works for the irrigation of crops, for the production of hydroelectricity, the measurement sites where water flows will be measured or the environmental aspects such as ecomorphology or biodiversity.

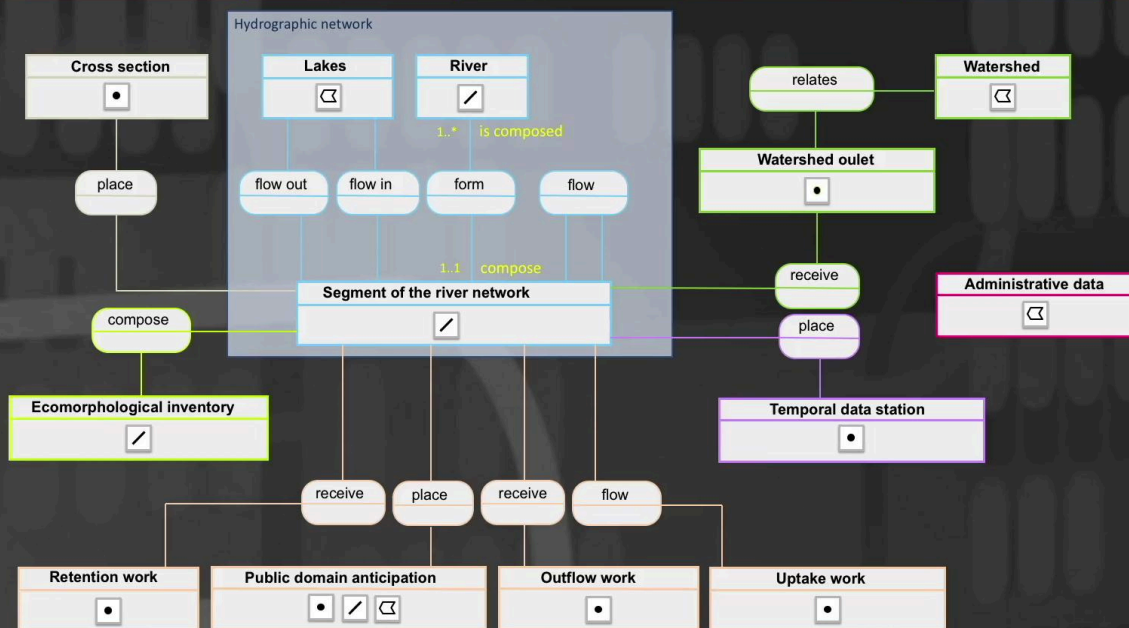
Notes

Summary



14m 56s

Some examples



An Introduction to Geographic Information Systems

The basic element here is the hydrographic network segment. Each segment flows into another segment and a set of segments composes a watercourse. By adding the multiplicity, we see that a water course is composed of at least one segment of the hydrographic network and that a segment of the hydrographic network is part of one, and only one stream. The water areas are linked to the hydrographic network in the sense that a segment of the hydrographic network flows into a water area and that a water area, in turn, flows into a segment of the hydrographic network. The catchment basin has a punctual outlet and this outlet point is in turn linked to the hydrographic network. The cross-sections are points positioned on the hydrographic network segments while the ecomorphological inventory is composed of segments with specific characteristics which overlap with the hydrographic network segments. To this we add the technical structures, dams, water rejections, extraction structures, temporal value measuring stations, administrative data like the limits of districts, management areas, etc. to finally get to a relatively complex model in which we find the basic hydrographic network, streams, water areas and segments of this network.

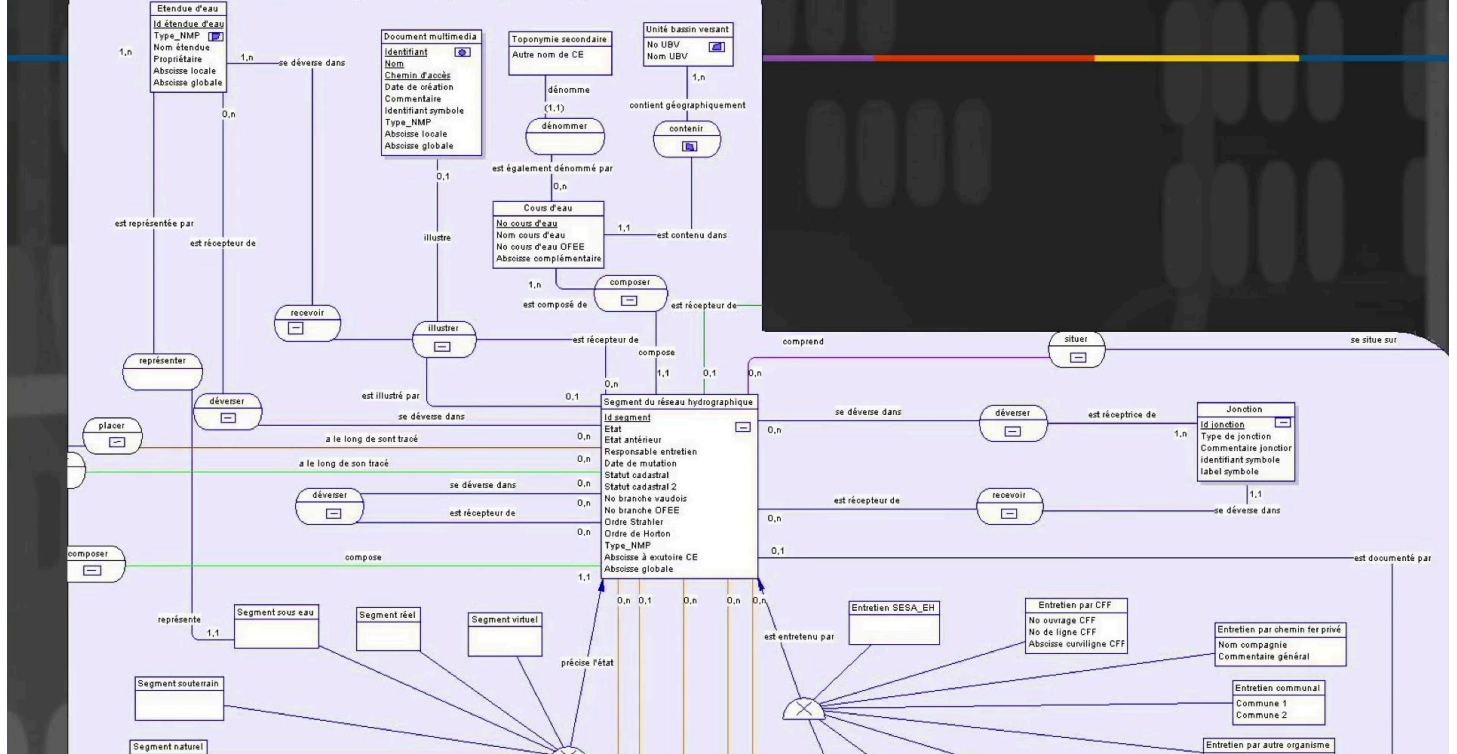
Notes

Summary



15m 35s

Réseau hydrographique



This schematic data model is then refined by describing the attributes of the various objects of the system, here the attributes of the hydrographic network segments, the attributes of streams, the attributes of water areas.

Notes

Summary

17m 09s



CASE tools

= Computer Aided Software Engineering

- Definition of data structures
- Creation of physical databases from a logical model
 - ➔ Forward engineering
- (Re-) creation of the data model from a physical database
 - ➔ Reverse engineering



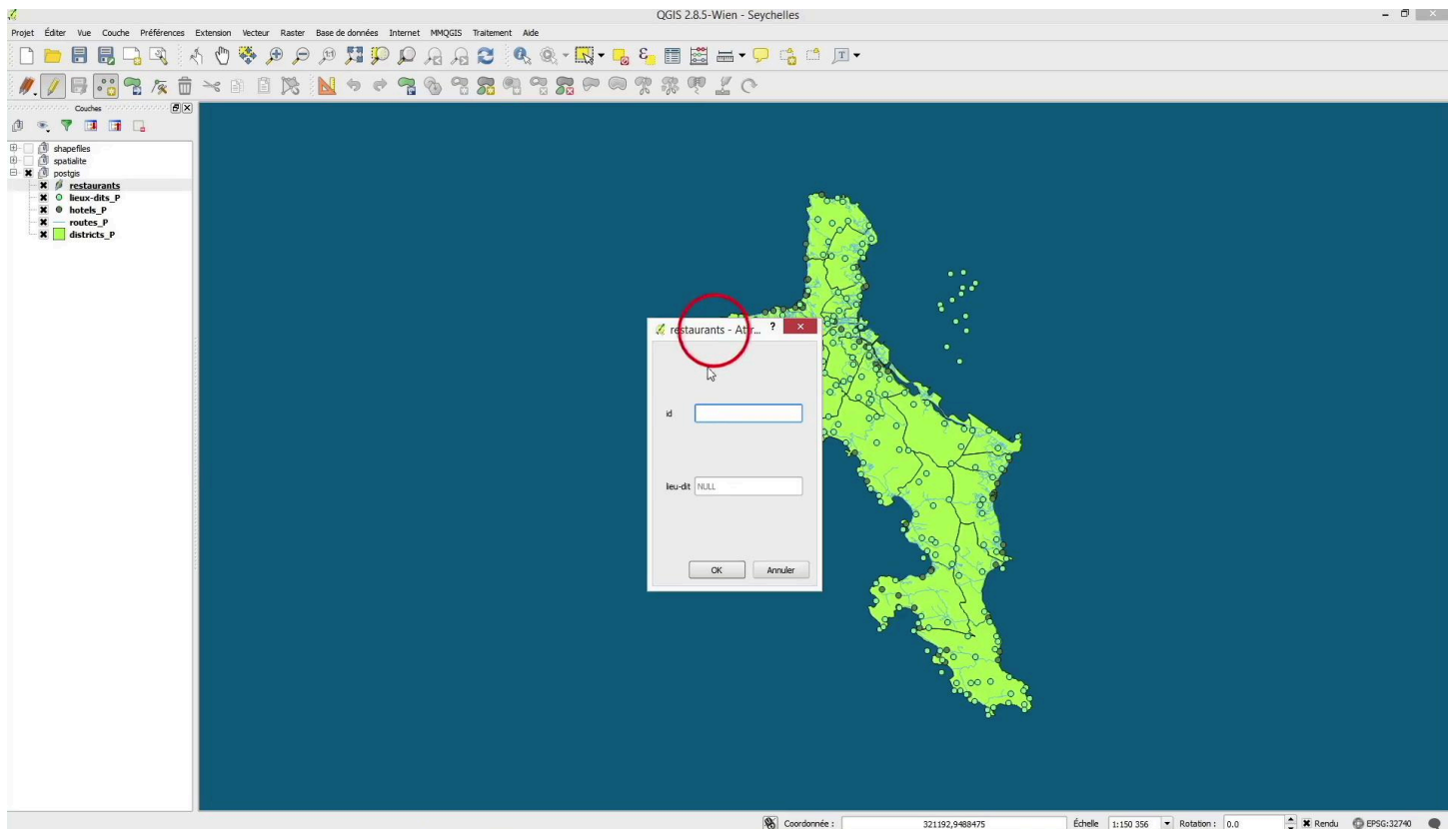
And by repeating this operation for all the objects of the system, a set of tables of the future database, we obtain a relatively complex model. The CASE tools, or CASE, we can pronounce it as we like, are software engineering tools, CASE meaning Computer Aided Software Engineering. These tools have essentially three functions: first, the definition of data structures then the creation, from a logical data model, from a physical database and we then speak of forward engineering and finally, the re-creation from a physical data model, of an existing physical database, the re-creation of a logical data model and we then speak of reverse engineering.

Notes

Summary



17m 23s



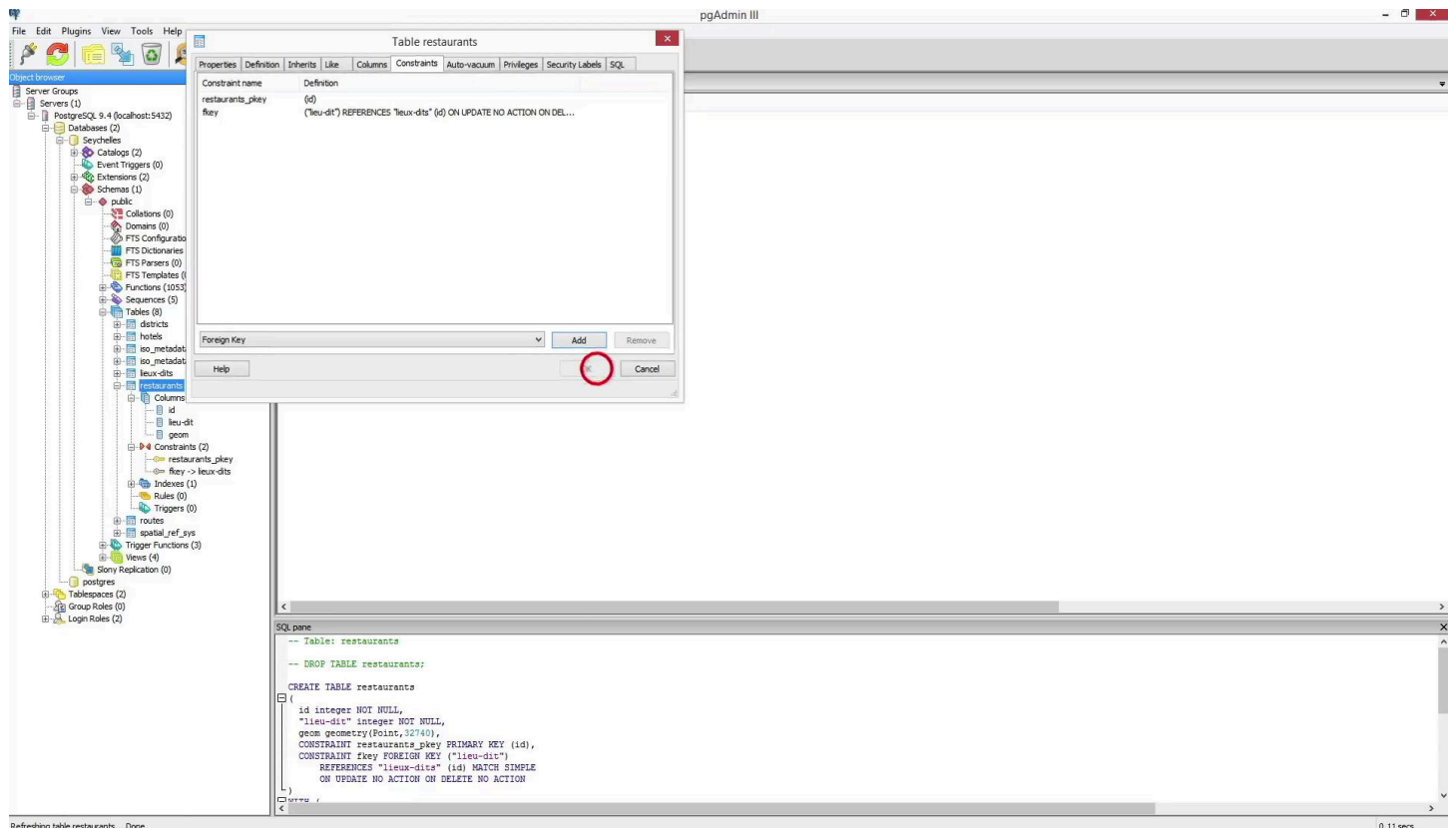
Most geographic information systems include database management systems which enable, if not truly modeling the data, to build or define the structure of these data. We see an example here with QGIS where we have the layers of the Seychelles for the PostGIS database version and we can see that, by means of the adequate extension, we can add a new data table in the PostGIS database for example, to describe restaurants. We add an identifier field of non-zero integer type and we define this field as the primary key of the table. We add a second field which we will call "locality" and in which we will store, in integer form, a reference to the layer of existing localities. And we then add a column for geometry in the form of points and then the reference to the local coordinate system, so UTM South 40. We can see that this new layer of restaurants was created, that it has a primary key but on the other hand the link towards the layer of the localities cannot be set up in the QGIS interface. However, this connection can be done, so if we add here now this layer of restaurants and we see that if we want to add a new restaurant, we will associate a locality to it, so we will spot a locality here towards Anse Major which bears the number 136 so we will create now a new restaurant on the Anse Major side, and then...

Notes

Summary

18m 29s





the identifier of the object is number 1 and we will identify the locality with the number 136 and now for that number 136 to be linked to the denomination of the locality, we need to create a joint between the restaurant data table and the locality data table, which is done in this interface of definition of vectorial joints. So we define the field and the attribute that should serve as a joint. And we notice the actual realization of this joint by consulting the attributes of the restaurant object created which now in addition to the identifier and the locality, has the name of the locality that is associated. This link, this joint, is effective at the level of the QGIS project but is not at all reflected at the level of the structure of the database. If we look into the administration of this database, with pgAdmin, we see that in the constraints we have the primary key under the identifier of the restaurants, but no foreign key which allows to link the restaurant table and the locality table. This foreign key can be created in the software, the DBMS of Postgres. We add a new foreign key that we will call "fkkey" for foreign key, we define this key with the reference layer which is the layer of the localities and its identifying field and which will be linked to the local locality field.

Notes

Summary

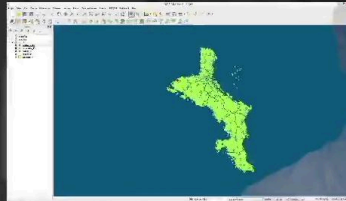
20m 30s



CASE tools

CASE software

- Database management systems, ev. Integrated in GIS



An Introduction to Geographic Information Systems

We validate these different things and we see that this foreign key was added to the constraints in the database, so in the data structure itself.

Notes

Summary



22m 22s

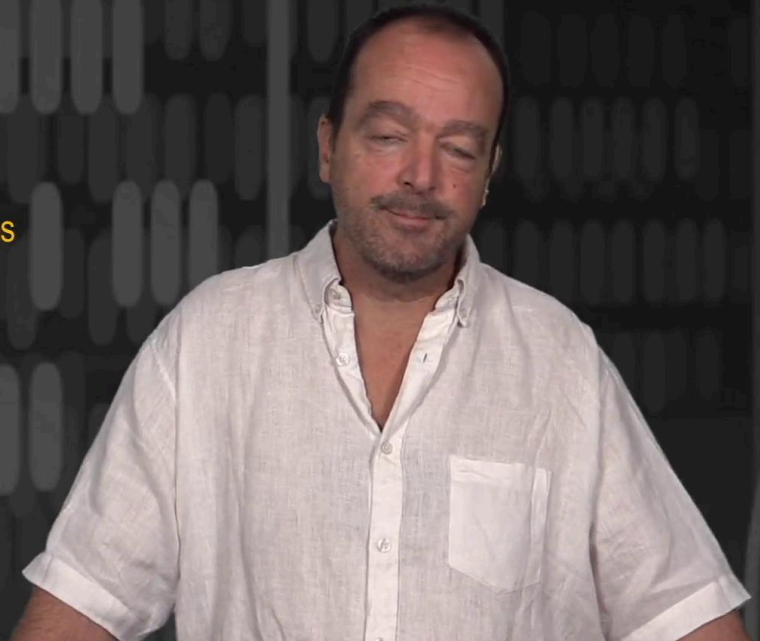
CASE tools

CASE software

- Database management systems, ev. Integrated in GIS
- Lots of specialized softwares
 - ➔ Modeling of various types of diagrams
 - ➔ Modeling data, including forward & reverse engineering

Database types

Handling spatial geometries



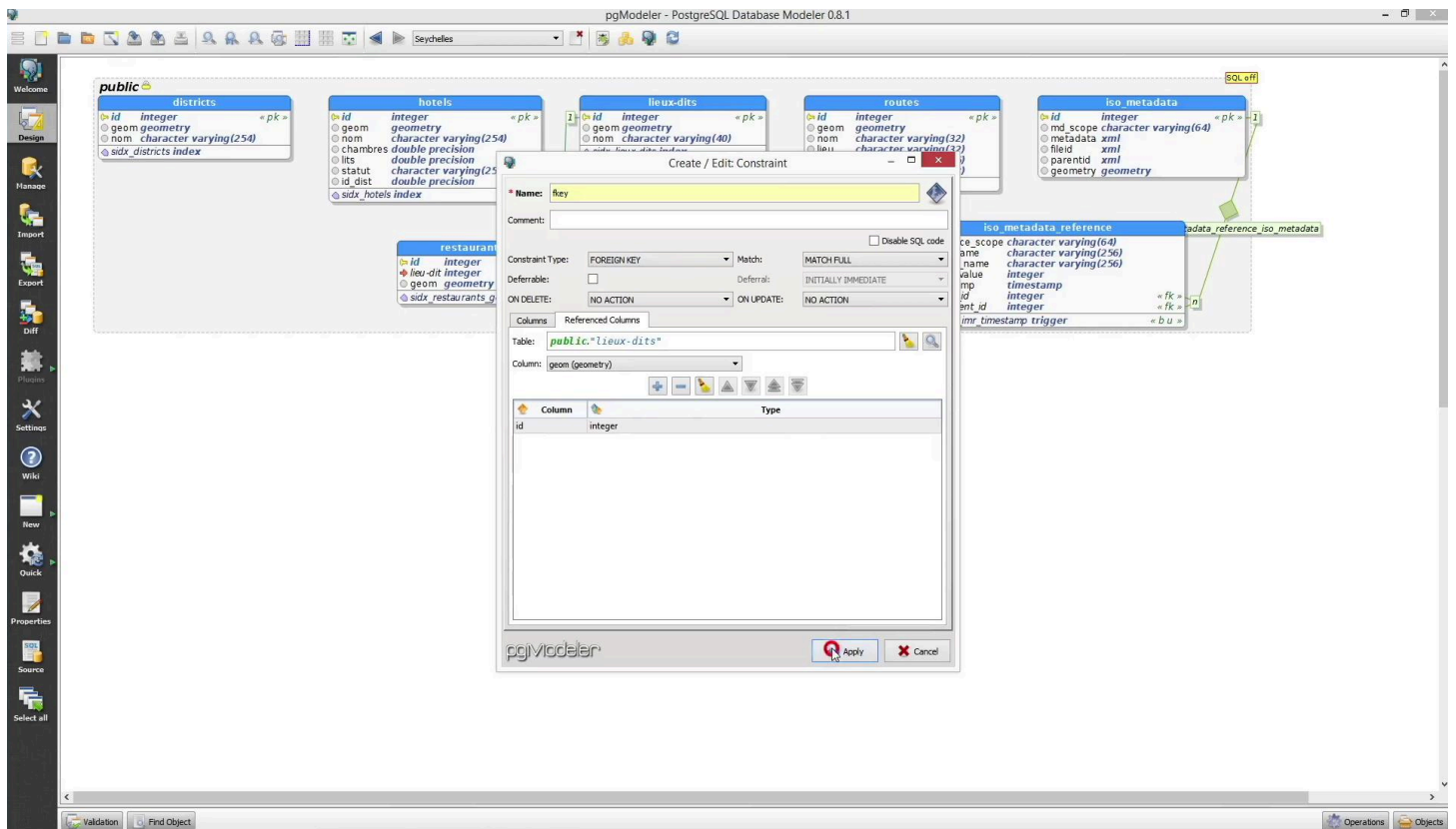
Beyond the data structuring tools integrated into geographic information systems or into database management systems, there are many specialized softwares. These softwares can be grouped into two main families. A first family which concerns softwares specialized in modelling, modelling of UML diagrams in general and not just data diagrams, and softwares specialized in data modelling with forward engineering and reverse engineering functions. In the latter family, the different softwares are distinguished by the type of database they support, there are specialized softwares on the PostgreSQL databases, others specialize on MySQL and whether or not they manage the geometries of spatial objects. We will also see in the rest of this module that the SQL language can be used to construct and manipulate the data structures and there are a number of softwares which, without going so far as the graphic representation of the data model, enable to manage the data models using the SQL language. But we will come back to these questions in the rest of the course.

Notes

Summary



22m 32s



For now, we will focus on one of these softwares, specifically designed for Postgres PostGIS databases. So we start by importing a database, in this case the well-known Seychelles database, we will not import the roles and in the diagram, we will not import the extensions and functions, we will just import the data tables. And we see, it is reverse engineering, the software allows to rebuild the data model and we find here the restaurant table we created with its relation to localities so its foreign key which we created earlier in the PostGIS administration system. As earlier we may want to create a new table for the beaches in which we will add an identifier field in the form of a non-null integer, we will also add a second field as earlier to host the locality reference, so here also, we will take an integer and as earlier, in the constraints we will add a primary key that we will call "pkey" and this primary key, we will use the identifier field which we have defined for this new table and we will also define a secondary key, as before we will call it "fkey", so a constraint of secondary type, foreign type, sorry, with as a reference the locality column and as a reference table the table of the localities and in particular in this table the field that will make the joint is the "identifier" field.

Notes

Summary



23m 57s

CASE tools

CASE software

- Database management systems, ev. Integrated in GIS
- Lots of specialized softwares
 - ➔ Modeling of various types of diagrams
 - ➔ Modeling data, including forward & reverse engineering

Database types

Handling spatial geometries

An Introduction to Geographic Information Systems

We validate all of this and we see that this new table is created with its association, the table of localities. This structure can be exported, after validation, to update the data model.

Notes

Summary



26m 12s