

Поддержка стандарта OpenGIS в СУБД ЛИНТЕР

Перспективы обработки ГИС-информации
ядром СУБД

ЛИНТЕР[®]



Поддержка пространственных типов данных в реляционной СУБД

- Представление пространственных данных в типах языка SQL.
- Выбор формата хранения данных в БД
- Методы доступа к пространственным данным, хранящимся в БД.
- Индексирование пространственных данных и использование индексов для поиска по ним.
- Единообразиие обработки пространственных и любых других типов данных: транзакционная работа, дискреционный и мандатный доступ, стандартные средства сохранения и восстановления данных

Поддержка пространственных типов данных в реляционной СУБД

- Поддержка множества специальных операций над пространственными данными: преобразования между различными форматами представления данных – текстовыми и бинарными; вычисление характеристик объектов — длины, площади и т.п; вычисление расстояний между объектами; определение взаимного расположения объектов
- Работа утилит СУБД с пространственными типами данных в смысле их визуального отображения
- Поддержка взаимодействия СУБД с существующими ГИС-приложениями.

Организация: Open Geospatial Consortium - международный некоммерческий консорциум разработчиков открытых ГИС-технологий

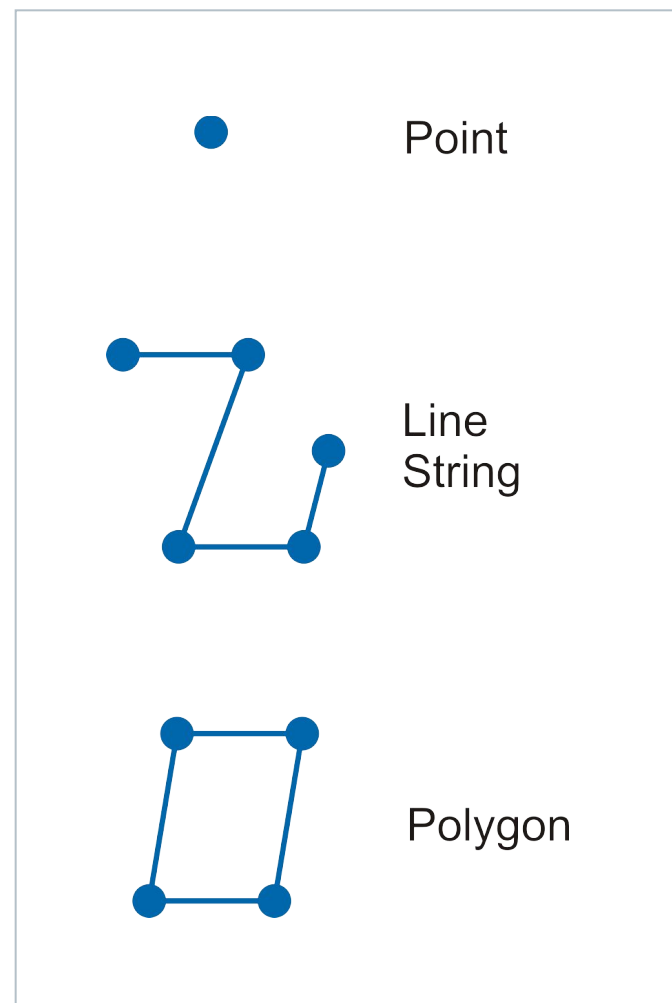
Стандарт: OpenGIS Implementation Specification for Geographic information - Simple feature access - Part 2: SQL option. Последняя версия этого стандарта (1.2.0) вышла в 2006 году

Организация: ISO – международная организация по стандартизации.

Стандарт: ISO/IEC CD 13249-3:2006 Database languages – SQL. Multimedia and Application Packages — Part 3: Spatial. Последняя версия также вышла в 2006 году

Модель данных стандартов OpenGIS/ISO

- Каждый геометрический объект связан с некоторой системой координат (Spatial Reference System), которая описывает координатное пространство, в котором определен объект
- В основе классификации геометрических объектов стандарты OpenGIS и ISO лежит размерность этих объектов
- Возможно представление 2,5D-объектов с ограниченной функциональностью третьей координаты (высоты)
- Все геометрические объекты являются топологически замкнутыми
- Каждый объект задается множеством базовых точек
- Базовые точки могут соединяться отрезками прямых либо дугами окружностей



Общая организация подсистемы обработки геоданных в СУБД Линтер

1. Библиотека манипулирования геоданными LINGEO
2. поддержка компилятором языка SQL
3. поддержка ядром СУБД ЛИНТЕР
4. дополнения в системный словарь СУБД ЛИНТЕР

- модуль преобразования из текстового представления геодезии (WKT) в двоичное (WKB);
- модуль преобразование из двоичного представления геодезии (WKB) в текстовое (WKT);
- модуль базовых процедур работы с внутренним представлением объектов геодезии (которое основано на WKB);
- модуль проверок допустимости представлений объектов геодезии;
- модуль работы с минимальными ограничивающими прямоугольниками;
- модуль вычисления расстояний, длин, площадей и прочих функций, зависящих от системы координат;

- модуль процедур проверки равенства двух объектов геодезии;
- модуль вычисления отдельных функций от объектов геодезии;
- модуль процедур вычисления пересечений полигонов с полигонами или ломаными;
- модуль процедур определения взаимного расположения объектов геодезии;
- модуль процедур вычисления множественных операций над объектами геодезии (UNION, INTERSECTION, DIFFERENCE, SYMDIFFERENCE)

- средства распознавания лексем, специфичных для подсистемы обработки геоданных, в модуле лексического разбора
- средства проверки правильности синтаксических конструкций, использующие пространственные данные, и разбора их параметров в модуле синтаксического разбора
- средства формирования внутреннего представления запросов, содержащих конструкции, использующие пространственные данные

- в общий механизм сравнения значений включены геометрические типы данных; сравнение значений для них производится путем вызовов соответствующих функций библиотеки LINGEO;
- в общий механизм вычисления выражений включено вычисление функций, имеющих аргументы или возвращающих значения геометрических типов. Вычисление производится путем вызовов соответствующих функций библиотеки LINGEO;
- распознавание предикатов, содержащих данные геометрических типов, выбор и выполнение для этих предикатов соответствующих стратегий обработки

- Таблица GEOMETRY_COLUMNS, требуемая стандартом OpenGIS, реализована в СУБД ЛИНТЕР как VIEW над системными таблицами и создается при создании системного словаря базы путем выполнения специального SQL-скрипта
- VIEW GEOMETRY_COLUMNS в СУБД ЛИНТЕР является чисто информационным и для функционирования подсистемы геоданных его существование не критично.

- Поддержка типов данных модели OpenGIS: Point, LineString, Polygon, MultiPoint, MultiLineString, MultiPolygon, GeometryCollection
- Для совместимости с другими СУБД добавлены типы данных Box, Line и Circle
- С точки зрения хранения реализованы как разновидности типа данных VARBYTE (при максимальной длине значения до 4000 байт — по умолчанию) или BLOB (при максимальной длине значения более 4000 байт)
- В качестве внутреннего формата хранения используется модифицированный формат WKB – WKB, следом за которым (в конце) дописан SRID

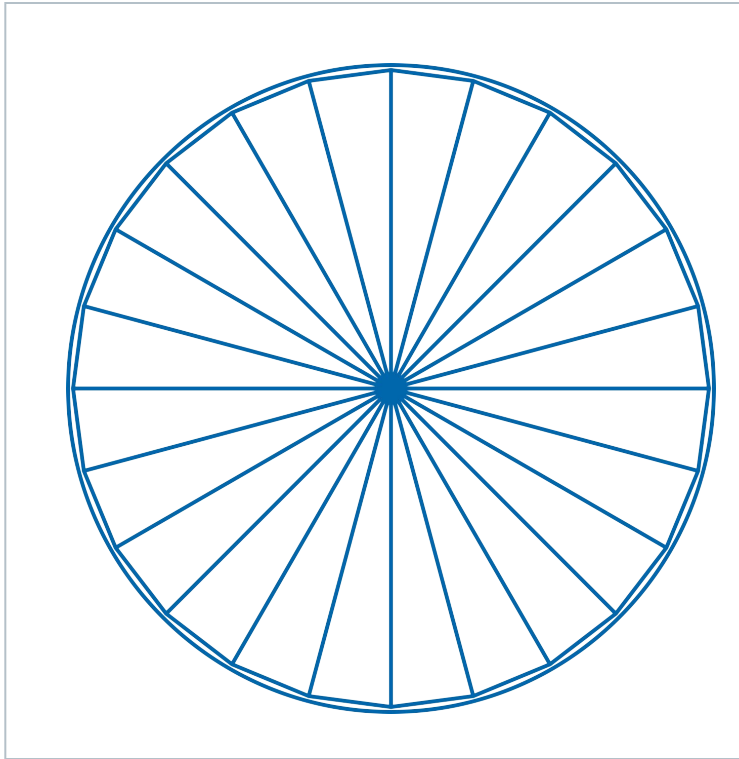
Столбцы геометрического типа создаются в СУБД ЛИНТЕР стандартным способом:

- либо при создании таблицы (CREATE TABLE), например:

```
CREATE TABLE geo_test (p POINT);
```
- либо при добавлении новых столбцов в существующую таблицу (ALTER TABLE ADD COLUMN), например:

```
ALTER TABLE geo_test ADD COLUMN ls LINESTRING;
```
- Никакой специальной регистрации столбца геометрического типа в БД процедурами типа AddGeometryColumn не требуется.

Особенности реализации стандартных функций над пространственными данными



- Функция Centroid – возвращает координаты "центра масс" многоугольника при условии, что вся его "масса" равномерно распределена между вершинами внешней границы многоугольника
- Функция PointOnSurface — возвращает первую точку внешней границы аргумента
- Пересечение значения типа Circle и значения, включающего Polygon - окружность аппроксимируется вписанным правильным многоугольником с 32 вершинами

Индексация пространственных данных

- Индексирование геометрических данных (как и других типов данных) в настоящее время выполняется с помощью B-дерева
- Синтаксис команды создания индекса тот же, что и для любых других типов данных:

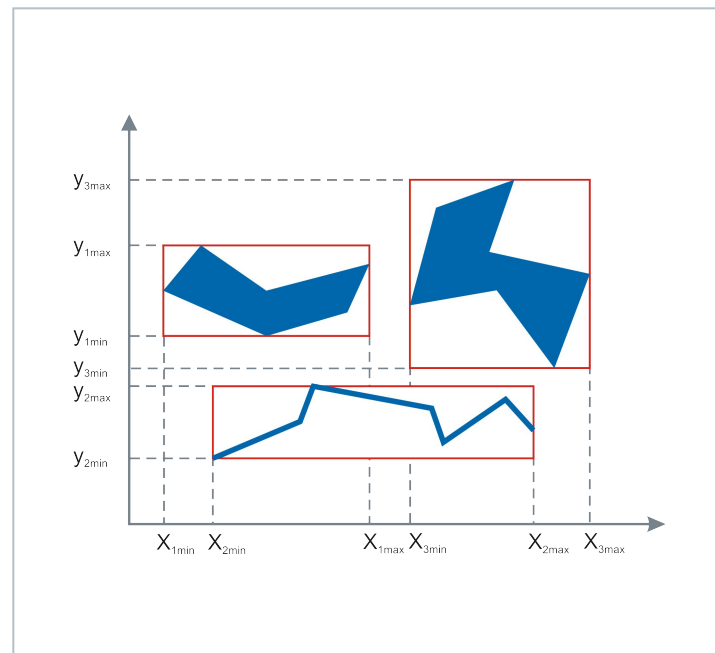
```
CREATE OR REPLACE TABLE point_test( p0 POINT );  
CREATE INDEX p0 ON point_test;
```

- Индекс по геометрическим типам данных может быть и составным:

```
CREATE OR REPLACE TABLE lspoint_test( p POINT,  
ls LINESTRING );  
CREATE INDEX test ON lspoint_test(p,ls);
```

Индексация пространственных данных

- Каждому значению ставится в соответствие его MBR — минимальный ограничивающий прямоугольник
- Для поиска по геометрическим типам данных используется предикат BETWEEN
- Нижняя и верхняя граница предиката BETWEEN для геометрических данных задают левый нижний и правый верхний угол прямоугольника, при полном или частичном попадании в который значение предиката BETWEEN считается истинным (TRUE)



- Для обработки пространственных предикатов других типов (например, проверки расстояние между объектом и фиксированной точкой или двумя объектами) индексы в настоящее время не используются. Стратегия вычисления таких предикаты организована таким образом, чтобы перенести проверку их истинности на самый поздний этап, когда вычислены все остальные предикаты

Проблемы совместимости и производительности

Поскольку подсистема работы с пространственными данными определяет большое количество дополнительных ключевых слов (в основном, названий функций над данными геометрических типов), которые могут совпасть с именами таблиц и столбцов, уже существующих в базах данных, в СУБД ЛИНТЕР предусмотрен режим совместимости с предыдущими версиями. Для включения этого режима ядро СУБД должно быть запущено с ключом "/COMPATIBILITY=GEOPREFIX. При работе в таком режиме для использования имен геометрических функций к ним необходимо добавлять префикс "LIN_".

Ввод пространственных данных сложной структуры (т.е. преобразование их из текстового представления в двоичное) предполагает значительное количество проверок правильности вводимой информации. Если данные берутся из надежного источника (т.е. уже были ранее проверены), то такие проверки можно отключить с помощью выбора специального режима работы:

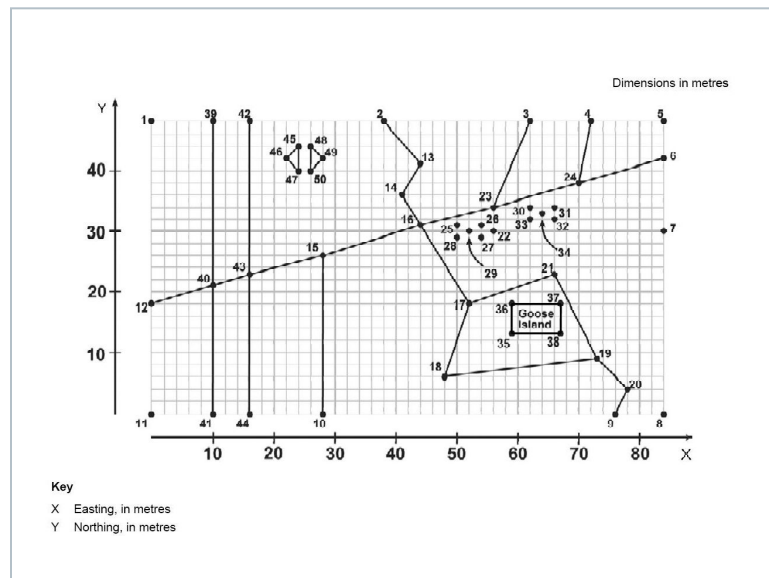
- либо для отдельного соединения:
`SET CONNECTION GEODATA VALIDITY CHECKING OFF;`
- либо для ядра СУБД в целом:
`SET DATABASE GEODATA VALIDITY CHECKING OFF;`

Сравнение значений геометрических типов

- Сравнение значений геометрических типов данных сводится в конечном счете к сравнению вещественных чисел (тип DOUBLE), а эти числа можно сравнивать лишь с определенной точностью
- В настоящее время в СУБД Линтер эта точность жестко задана и равна $+1e-10$. То есть числа, отличающиеся на меньшую величину, считаются равными
- В дальнейшем планируется реализовать возможность задания этой величины пользователем.

Тестирование работы подсистемы геоданных с помощью тестов OpenGIS

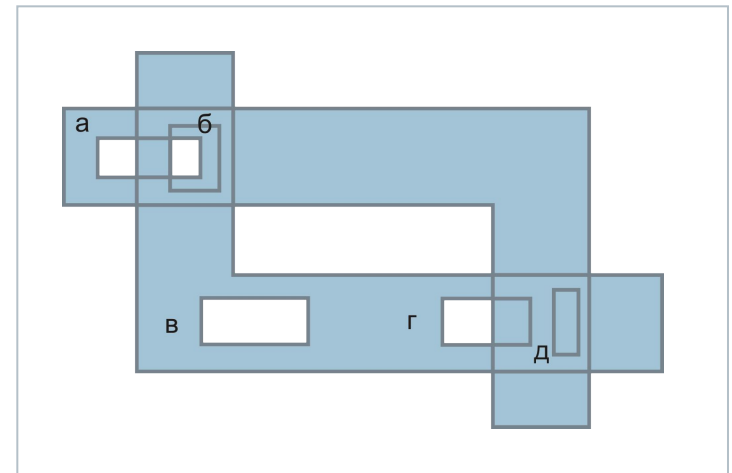
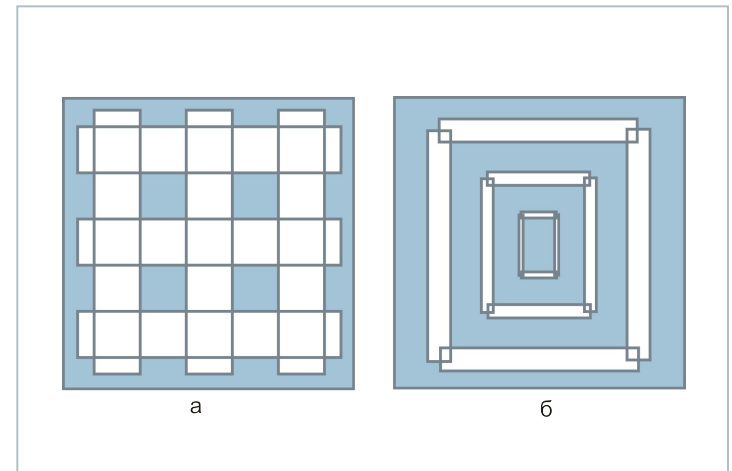
Для проверки функционирования подсистемы геоданных СУБД ЛИНТЕР используется набор тестов, созданный на основе приложения С3 к стандарту OpenGIS Simple Features for SQL («Conformance Tests»). В настоящее время СУБД ЛИНТЕР корректно выполняет все тесты этого набора, за следующими исключениями:



- поддерживается только одна система координат — декартова прямоугольная; ведутся работы по поддержке других систем координат
- не поддерживается обобщенная операция Relate, хотя специфические операции проверки взаимного расположения объектов (Equals, Disjoint, Touches, Within, Overlaps, Crosses, Intersects, Contains) поддерживаются для всех типов пространственных данных.

Тестирование работы подсистемы геоданных с помощью собственных наборов тестов

- Тесты преобразования всех типов геометрических объектов в текстовый формат (WKT) и внутренний формат (WKV) из текстового
- Тесты вычисления характеристик геометрических объектов различных типов
- Тесты вычисления расстояний между различными типами геометрических объектов для всех возможных вариантов их взаимного расположения;
- Тесты объединения, пересечения, разности и симметрической разности фигур сложной формы (многоугольники, содержащие исключенные области, в том числе множественные, вложенные друг в друга, имеющие форму «решётки», «расчёски» и т.д.)



Спасибо за внимание!

Вопросы?