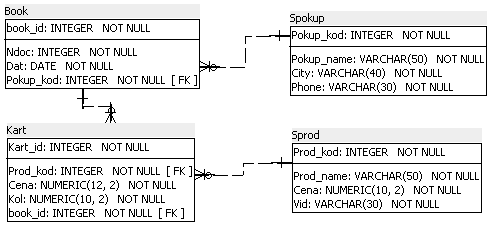
**Практическое занятие № 5**

**Загрузка и выгрузка данных БД PostgreSQL**

**Цель занятия**. Научиться подготавливать и загружать данные в базы данных PostgreSQL, выгружать данные из баз данных PostgreSQL.

1. **Загрузка данных из текстовых файлов**

Загрузку и выгрузку данных будем рассматривать на примере БД продаж, созданной на предыдущих занятиях (рис. 1).



**Рис. 1 База данных Продажи**

Последовательность загрузки данных должна быть следующая. Вначале загружаются данные в таблицы, не имеющие внешних ключей, затем в таблицы, которые ссылаются на таблицы с уже загруженными данными и т.д.

Для рассматриваемой базы данных Продаж последовательность загрузки должна быть следующей:

* Продукция (таблица Sprod);
* Покупатель (таблица Spokup);
* Книга продаж (таблица Book);
* Ассортимент продаж (таблица Kart).

Данные могут быть подготовлены в текстовых файлах и файлах формата CSV.

Файлы CSV содержат данные в текстовом формате, разделенные запятыми. Одна строка — это одна запись об одном объекте. Однотипные свойства в каждой строке занимают фиксированные места. Некоторые значения могут быть пропущены. В месте пропуска будут подряд следовать символы разделители. В первой строке файла могут быть указаны имена свойств объекта, разделенные запятыми. Имена свойств обычно соответствуют именам колонок таблицы БД. Файлы формата удобно подготавливать и просматривать в Microsoft Excel.

**Подготовка данных в Excel**

Откройте программу Microsoft Excel и введите данные для таблиц (рис.2-5). Обратите внимание на правильность заполнения внешних ключей.



**Рис. 2 Данные для таблицы sprod (Продукция)**



**Рис. 3 Данные для таблицы spokup (Покупатели)**

****

**Рис. 4 Данные для таблицы book (Книга продаж)**

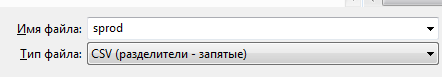


**Рис. 5 Данные для таблицы kart (Ассортимент продаж)**

Введенные данные необходимо сохранить в формате CSV (текстовые данные с разделителями между колонками значений). Для этого необходимо выполнить команду меню

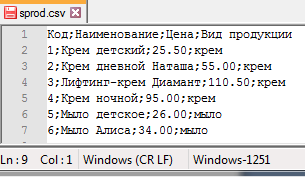
*Сохранить как - Другие форматы*

В открывшемся окне необходимо выбрать формат сохранения CSV, а имя файла на английском (рис.6). Для удобства выберем имя, совпадающее с именем таблицы, для которой данные предназначены.



**Рис. 6 Сохранение данных в формате CSV**

Откроем один из сохраненных файлов в текстовом редакторе (рис.7). Первая строка содержит название атрибутов. Это надо учитывать при загрузке данных в таблицы.



**Рис. 7 Данные в формате CSV**

**Загрузка данных в таблицы БД**

Для загрузки данных из подготовленных файлов в таблицы базы данных используется команда COPY, имеющая следующую структуру (в квадратные скобки заключается элемент, который может быть опущен):

COPY *имя\_таблицы* [ ( *имя\_столбца* , ... ) ]

FROM '*имя\_файла*'

[*параметр* , ...]

В команде могут быть использованы *параметры (ниже указаны основные)*:

*имя\_формата* данных (по умолчанию txt). Часто используется формат CSV (нужно указывать явно);

DELIMITER '*символ\_разделитель*' – символ-разделитель значений полей (для CSV по умолчанию запятая).

Если текстовые строки содержат запятые, то в качестве разделителя можно использовать символы ';' или '|';

NULL - '*маркер\_NULL*' определяет строку, задающую значение NULL. По умолчанию в текстовом формате это \N (обратная косая черта и N), а в формате CSV — пустая строка без кавычек;

HEADER [ *boolean* ] - Указывает, что файл в 1-й строке содержит строку заголовка с именами столбцов. Этот параметр допускается только для формата CSV. При отсутствии значения *boolean* подразумевается *true;*

QUOTE '*символ\_кавычек*' - Указывает символ кавычек, используемый для заключения значений в кавычки. Этот параметр поддерживается только для формата CSV;

ENCODING '*имя\_кодировки*' - Указывает, что файл имеет заданную*, например* WIN1251.

Для выполнения команды необходимо открыть окно ввода SQL-запросов Пример записи команды загрузки в таблицу sprod из файла sprod.csv, находящегося в каталоге C:/data:

SET SEARCH\_PATH TO prod; // Установка текущей схемы

COPY sprod

FROM 'C:/data/sprod.csv' CSV

DELIMITER ';' // разделитель «точка с запятой»

HEADER; // первая строка содержит имена атрибутов

Если в файле использована кодировка 1251, а в БД используется кодировка UTF8, то при экспорте будет выведено сообщение об ошибке:

*ОШИБКА: неверная последовательность байт для кодировки "UTF8"*

В этом случае надо указать кодировку исходного файла в параметре:

ENCODING 'WIN1251'

И команда загрузки будет такой:

COPY sprod

FROM 'c:/data/sprod.csv' CSV

DELIMITER ';'

HEADER

ENCODING 'WIN1251';

Если все данные соответствуют тем типам, которые определены в таблице БД, то не будет ошибок загрузки данных. Если же ошибки появляются, необходимо скорректировать подготовленные данные.

Полный пакет команд загрузки данных в базу данных Продажи (последовательность команд должна быть именно такой):

SET SEARCH\_PATH TO prod; -- Установка текущей схемы

COPY sprod FROM 'c:/data/sprod.csv' CSV

DELIMITER ';' HEADER ENCODING 'WIN1251';

COPY spokup FROM 'c:/data/spokup.csv' CSV

DELIMITER ';' HEADER ENCODING 'WIN1251';

COPY book FROM 'c:/data/book.csv' CSV

DELIMITER ';' HEADER ENCODING 'WIN1251';

COPY kart FROM 'c:/data/kart.csv' CSV

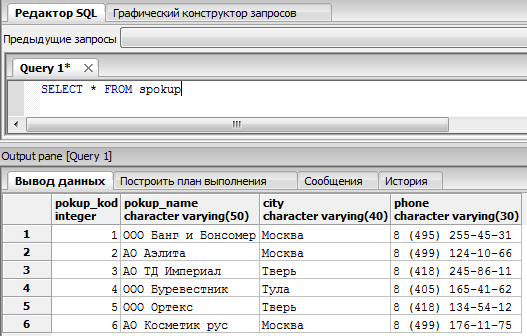
DELIMITER ';' HEADER ENCODING 'WIN1251';

***Замечание.*** Путь к файлам рассматривается относительно сервера. Поэтому при подключении к удаленному серверу необходимо использовать сетевое имя. Для сервера БД на локальном компьютере задаем обычные пути к файлу.

Можно проверить загрузку данных выполнив запрос к таблице spokup (Покупатели):

SELECT \* FROM prod.spokup

Результат запроса приведен на рис.9



**Рис. 9 Результат запроса выборки данных**

При загрузке данных мы явно задавали значения суррогатным первичным ключам (идентификаторам), которые при добавлении данных должны заполняться с использованием последовательностей. Чтобы механизм автоувеличения далее работал корректно необходимо изменить текущие значения последовательностей на последнее значение, заданное в файлах данных. Это выполняется оператором:

setval (имя\_последовательности, последнее\_значение, true)

Максимальные значения первичного ключа в таблицах (по файлам данных):

book: book\_id=5

kart: kart\_id=10

Поэтому для изменения текущего значения последовательностей ждя этих колонок необходимо выполнить операторы:

SELECT setval('prod.book\_seq', 5, true);

SELECT setval('prod.kart\_seq', 10, true);

**Выгрузка данных в файл**

Можно не только загружать данные в таблице, но также и выгружать во внешние файлы. Это также выполняется командой COPY:

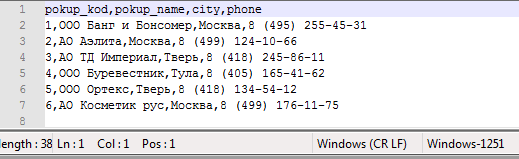
COPY имя\_таблицы TO 'путь и имя файла' [параметры];

Необязательные параметры такие же, как и при загрузке данных. Для примера выведем данные из таблицы Покупатель (spokup) в формате CSV с названием колонок в первой строке, с разделителем «запятая», в кодировки WIN-1251 в файл c:\data\spokup.csv

COPY prod.spokup TO 'c:/data/spokup.csv' DELIMITER ','

CSV HEADER ENCODING 'WIN1251';

Содержимое файла в текстовом редакторе приведено на рис. 10.



**Рис. 10 Результат выгрузки данных**

**Задание.** Для базы данных созданной на предыдущем практическом занятии:

1) Подготовить файлы данных

2) Выполнить загрузку данных в таблицы БД из внешних файлов

3) Выполнить выгрузку данных из таблиц БД во внешний файл