

ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ
Федеральное государственное образовательное бюджетное
учреждение высшего профессионального образования
«САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ
им. проф. М. А. БОНЧ-БРУЕВИЧА»

М. Б. Вольфсон

ХРАНИЛИЩА ДАННЫХ

МЕТОДИЧЕСКИЕ УКАЗАНИЯ
К ПРАКТИЧЕСКИМ РАБОТАМ

СПбГУТ)))

САНКТ-ПЕТЕРБУРГ
2020

Задание №1. Анализ рынка ETL средств

1. Используя открытые источники данных в сети Интернет составить перечень имеющихся на рынке ETL-средств (не менее пяти).
2. Разработать критерии сравнения ETL-средств по технико-экономическим характеристикам (не менее пяти).
3. Заполнить таблицу 1.

Таблица 1

Критерий сравнения	Система 1	Система 2	Система N

Задание №2. Анализ данных в реляционной модели

Исходные данные

Имеется компания, занимающаяся продажей продуктов питания. Для хранения и последующего анализа выбираются данные, описывающие бизнес-процесс продажи товаров оптовым клиентам.

Объектами предметной области являются клиенты, товары, сотрудники отдела продаж, поставщики.

Задание

1. В Microsoft Access 2007 создать реляционную базу данных (рис. 1.1). Имя файла должно соответствовать фамилии студента.

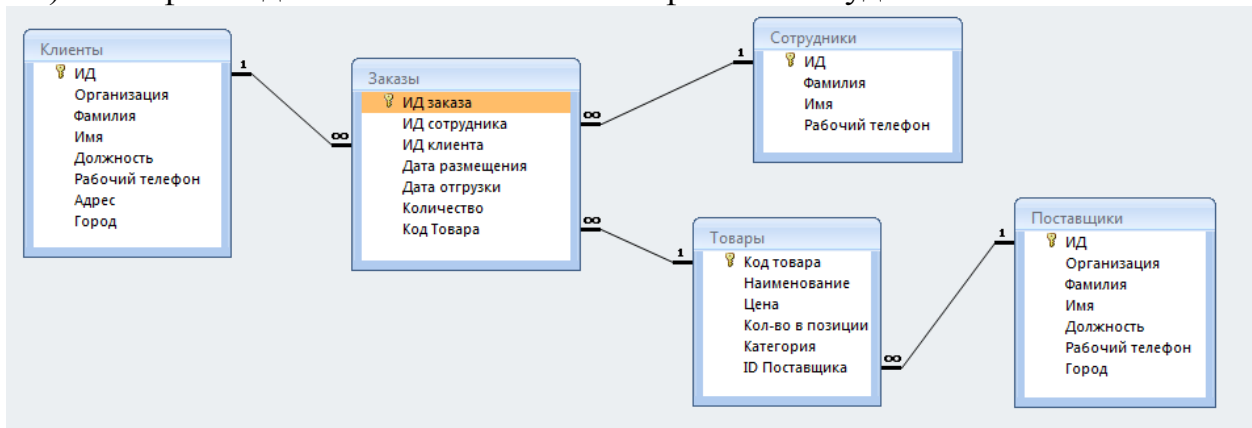


Рис. 1.1

Для более быстрого заполнения таблиц рекомендуется создать маски ввода и использовать операции копирования и вставки.

Таблица *Заказы*, не менее 30 записей (рис. 1.2).

ИД заказа	Сотрудник	Клиент	Дата размещения	Дата отгрузки	Количество	Код Товара
30	9	1	15.01.2010	22.01.2010	100	NWTB-1
31	3	2	20.01.2010	22.01.2010	200	NWTB-10
32	4	3	22.01.2010	22.01.2010	300	NWTB-11
33	6	4	30.01.2010	31.01.2010	400	NWTB-12
34	9	5	06.02.2010	07.02.2010	500	NWTB-13
35	3	6	10.02.2010	12.02.2010	600	NWTB-14
36	4	7	23.02.2010	25.02.2010	700	NWTB-15
37	8	8	06.03.2010	09.03.2010	800	NWTB-16
38	9	9	10.03.2010	11.03.2010	900	NWTB-17
39	3	10	22.03.2010	24.03.2010	100	NWTB-19
40	4	11	24.03.2010	24.03.2010	200	NWTB-20

Рис. 1.2

Таблица *Клиенты*, не менее 10 записей (рис. 1.3).

ИД	Организация	Фамилия	Имя	Рабочий телефон	Адрес	Должность	Город
1	Defa	Костерина	Ольга	(123)555-0111	1-я улица, д. 12	Ответственный	Сочи
2	ООО "Эдельвейс"	Верный	Григорий	(123)555-0112	2-я улица, д. 12	Ответственный	Иваново
3	ООО "Алкон"	Егоров	Владимир	(123) 555-0113	3-я улица, д. 12	Сотрудник отдела снабжения	Уфа
4	ООО "Альтаир"	Омельченко	Светлана	(123) 555-0114	4-я улица, д. 12	Начальник отдела снабжения	Москва
5	ОАО "Арго"	Песоцкий	Станислав	(123) 555-0115	5-я улица, д. 12	Ответственный	Иркутск
6	ООО "Балт-Конд"	Шашков	Руслан	(123) 555-0116	6-я улица, д. 12	Начальник отдела снабжения	Казань
7	ЗАО "Глобар"	Вронский	Юрий	(123) 555-0117	7-я улица, д. 12	Ответственный	Воронеж
8	ЗАО "ЛЭНД"	Подколзина	Екатерина	(123) 555-0118	8-я улица, д. 12	Сотрудник отдела снабжения	Омск
9	ЗАО "Камея"	Ерёменко	Алексей	(123) 555-0119	9-я улица, д. 12	Начальник отдела снабжения	Пермь
10	ООО "Клин"	Грачев	Николай	(123) 555-0120	10-я улица, д. 12	Начальник отдела снабжения	Саратов
11	ООО "Локис"	Орехов	Алексей	(123) 555-0121	11-я улица, д. 12	Начальник отдела снабжения	Орел
12	ООО "Риф"	Володин	Виктор	(123) 555-0122	11-я улица, д. 12	Начальник отдела снабжения	Тюмень
13	ЧП "Синицын"	Туманов	Александр	(123) 555-0123	13-я улица, д.4	Сотрудник отдела снабжения	Курск

Рис. 1.3

Таблица *Сотрудники*, не менее 10 записей (рис. 1.4).

ИД	Имя	Рабочий телефон	Фамилия
1	Юлия	(123)555-0101	Ильина
2	Андрей	(123) 555-0102	Гладких
3	Евгений	(123) 555-0103	Куликов
4	Мария	(123) 555-0104	Сергиенко
5	Николай	(123) 555-0105	Новиков
6	Вадим	(123) 555-0106	Корепин
7	Сергей	(123) 555-0107	Климов
8	Инна	(123) 555-0108	Ожогина
9	Дарья	(123) 555-0109	Попкова

Рис. 1.4

Таблица *Поставщики*, не менее 10 записей (рис. 1.5).

ИД	Организация	Фамилия	Имя	Рабочий телефон	Город	Должность
1	ООО "Аврора"	Андреева	Елизавета	1234567	Санкт-Петербург	Начальник отдела сбыта
2	ООО "Балтика"	Волкова	Марина	2345678	Санкт-Петербург	Начальник отдела сбыта
3	ИП "Волга"	Котова	Маргарита	3456789	Москва	Сотрудник отдела сбыта
4	ИП "Горизонт"	Орлов	Николай	4567890	Москва	Маркетолог
5	ОАО "Диета"	Зорин	Антон	5678901	Москва	Начальник отдела сбыта
6	ОАО "Питер"	Хромов	Евгений	6789012	Санкт-Петербург	Помощник по маркетингу
7	ЗАО "Нева"	Гласнов	Олег	7890123	Санкт-Петербург	Маркетолог
8	ЗАО "Белые ночи"	Сидоров	Борис	8901234	Санкт-Петербург	Сотрудник отдела сбыта
9	ООО "Ладога"	Соколова	Лариса	9012345	Санкт-Петербург	Начальник отдела сбыта

Рис. 1.5

Таблица *Товары*, не менее 30 записей, не более 10 категорий (рис. 1.6).

Код товара	Наименование	Цена	Кол-во в позиции	Категория	ID Поставщика
NWTB-1	Ананас	1,80р.	15,25 унций	Фруктовые и овощные консервы	9
NWTB-10	Зеленый горошек	1,50р.	14,5 унций	Фруктовые и овощные консервы	4
NWTB-11	Зеленый чай	2,99р.	20 пакетиков в коробке	Напитки	1
NWTB-12	Индийский чай	4,00р.	100 штук в коробке	Напитки	6
NWTB-13	Карри	40,00р.	12 банок по 400 г	Соусы	7
NWTB-14	Конфеты	11,80р.	5 коробок	Кондитерские изделия	3
NWTB-15	Копченый лосось	4,00р.	5 унций	Мясные консервы	6
NWTB-16	Кофе	46,00р.	16 банок по 500 г	Напитки	7
NWTB-17	Кукуруза	1,20р.	14,5 унций	Фруктовые и овощные консервы	3
NWTB-19	Лаваш	10,00р.	24 пакета по 4 штуки	Хлебобулочные изделия	3
NWTB-20	Луизианский соус	21,05р.	32 бутылки по 8 унций	Соусы	4
NWTB-21	Мармелад	81,00р.	30 коробок	Варенье и джемы	2
NWTB-22	Миндаль	10,00р.	пакет 5 кг	Сухофрукты и орехи	7
NWTB-28	Персики	1,50р.	15,25 унций	Фруктовые и овощные консервы	8
NWTB-29	Пиво	14,00р.	24 бутылки по 12 унций	Напитки	4
NWTB-3	Вишневый пирог	2,00р.	15,25 унций	Фруктовые и овощные консервы	1
NWTB-30	Пирожное с орехами	12,49р.	3 коробки	Хлебобулочные изделия	4

Рис. 1.6

2. В созданной базе данных создать запросы на выборку данных (при необходимости осуществив группировку данных) и проанализировать результаты.

2.1. Вывести фамилию самого успешного сотрудника по критерию количества продаж (*Запрос 2_1*)

2.2. Вывести фамилию самого успешного сотрудника по критерию объема выручки (*Запрос 2_2*).

2.3. Вывести самый продаваемый товар по количеству заказов (*Запрос 2_3*).

2.4. Вывести поставщика, продукция которого оказалась самой востребованной (*Запрос 2_4*).

2.5. Вывести самого выгодного клиента по критерию количества закупаемого товара (*Запрос 2_5*).

2.6. Вывести средний объем заказа в количественном и денежном выражении для каждого клиента (*Запрос 2_6*).

2.7. Вывести минимальное и максимальное количество дней между размещением и отгрузкой заказов (*Запрос 2_7*).

3. Эмулируя создание многомерной таблицы создать перекрестные запросы (использовать Мастер запросов) и проанализировать полученные результаты.

Для вывода полей из разных таблиц предварительно создайте запрос (*Запрос_темп*), содержащий все необходимые для запросов п.3 поля. Обратите внимание на многократное повторение данных в полученной таблице.

3.1. Вывести распределение принятых заказов каждого сотрудника по месяцам и определить количество заказов за каждый месяц (*Запрос 3_1*).

3.2. Вывести количество купленных товаров каждым из клиентов по товарным категориям и выявить количество заказов по самой популярной категории (*Запрос 3_2*).

3.3. Вывести распределение количества поставок по товарным категориям и городам фирм-поставщиков (*Запрос 3_3*).

Задание №3. Анализ данных в многомерной модели

Целью лабораторной работы является изучение и анализ возможностей реляционной модели хранения данных для последующего бизнес-анализа.

Исходные данные

Имеется компания, занимающаяся продажей продуктов питания. Для хранения и последующего анализа выбираются данные, описывающие бизнес-процесс продажи товаров оптовым клиентам.

Объектами предметной области являются клиенты, товары, сотрудники отдела продаж, поставщики.

Задание

1. Создать новую книгу Excel. Имя файла должно соответствовать фамилии студента.

2. Первый лист назвать «*Справочники*». Выбрав пункт меню «Получить данные из Access» последовательно импортировать на этот лист содержание таблиц «*Товары*», «*Клиенты*», «*Поставщики*», «*Товары*», «*Сотрудники*», снабдив таблицы соответствующими заголовками.

Таким образом, мы получаем следующие ключевые измерения данных: *название и цена товара, организация покупателя, фамилия сотрудника и организация поставщика*. Кроме того, имеются несколько дополнительных измерений данных: *категория продукта, город поставщика* и др.

3. Второй лист назвать «*Заказы*». Импортировать на этот лист содержание таблицы «*Заказы*» (при желании, впоследствии подключения к базе данных можно удалить через меню «*Данные-Подключения*»).

4. Заменить внешние ключи реляционной базы даны (*коды сотрудников, клиентов и товаров*) на названия, взятые из таблиц листа «*Справочник*». Для этого создать новые поля и воспользоваться функцией ПРОСМОТР (кодовые поля на листе «*Справочник*» должны быть отсортированы по возрастанию).

5. Аналогично п. 4. добавить в таблицу на лист «*Заказы*» измерения *Категория товара, Цена, Название поставщика и Город поставщика*. Кодовые поля на листе «*Заказы*» можно скрыть.

Таким образом, мы получаем исходную таблицу для последующего многомерного анализа, имеющую измерения *Сотрудник, Клиент, Товар, Категория, Цена, Поставщик, Город* и факты *Дата размещения, Дата отгрузки, Количество*.

6. Построить сводную таблицу, соответствующую кубу данных с измерениями *Клиент, Товар, Месяц размещения заказа* и с набором фактов о

выручке с продаж. В качестве дополнительных измерений (фильтров сводной таблицы) выбрать *Категорию товара*, *Поставщика* и *Сотрудника*.

7. Оформить таблицу по образцу рис. 2.3.

8. Проанализировать полученный результат и получить ответы на следующие вопросы (результаты сохранить на отдельных листах Excel).

8.1. Каков объем выручки за зимние месяцы от продажи фруктов и продуктов из фруктов?

8.2. Каков итоговый объем выручки от продажи товаров поставщиков – акционерных обществ, реализованных первыми тремя сотрудниками?

	A	B	C	D	E	F	G	H
1	Категория	(Все)						
2	Поставщик	(Все)						
3	Сотрудник	(Все)						
4								
5	Сумма по полю Оборот	Названия столбцов						
6	Названия строк	январь	фев	мар	апр	май	июн	Общий итог
7	Defa							
8	Ананас	180,00р.	-	-	-	-	-	180,00р.
9	Карри	-	-	-	-	-	20 000,00р.	20 000,00р.
10	Персики	-	-	750,00р.	-	-	-	750,00р.
11	Фасоль	-	-	-	120,00р.	-	-	120,00р.
12	Defa Итого	180,00р.	-	750,00р.	120,00р.	-	20 000,00р.	53 400,00р.
13	ЗАО "Глобар"							
14	Грецкие орехи	-	-	-	-	16 275,00р.	-	16 275,00р.
15	Копченый лосось	-	2 800,00р.	-	-	-	-	2 800,00р.
16	Луизианский соус	-	-	-	-	-	4 210,00р.	4 210,00р.
17	Сушеные сливы	-	-	-	700,00р.	-	-	700,00р.
18	ЗАО "Глобар" Итого	-	2 800,00р.	-	700,00р.	16 275,00р.	4 210,00р.	93 240,00р.
19	ЗАО "Каменя"							
20	Ежевичный джем	-	-	-	-	22 500,00р.	-	22 500,00р.
21	Кукуруза	-	-	1 080,00р.	-	-	-	1 080,00р.
22	Миндаль	-	-	-	4 000,00р.	-	-	4 000,00р.
23	Тихоокеанские крабы	-	-	-	7 360,00р.	-	-	7 360,00р.
24	ЗАО "Каменя" Итого	-	-	1 080,00р.	22 720,00р.	22 500,00р.	-	141 960,00р.
25	ЗАО "ЛЭНД"							
26	Груши	-	-	-	-	1 040,00р.	-	1 040,00р.
27	Кофе	-	-	36 800,00р.	-	-	-	36 800,00р.
28	Мармелад	-	-	-	24 300,00р.	-	-	24 300,00р.
29	Сушеные яблоки	-	-	-	15 900,00р.	-	-	15 900,00р.
30	ЗАО "ЛЭНД" Итого	-	-	36 800,00р.	80 400,00р.	1 040,00р.	-	398 860,00р.
31	ОАО "Арго"							
32	Карри	-	20 000,00р.	-	-	-	-	20 000,00р.
33	Кукуруза	-	-	-	-	-	1 080,00р.	1 080,00р.

Рис. 2.3.

9. На отдельном листе построить новую сводную таблицу, показывающую распределение объема продаж по городам закупок по месяцам. На том же листе построить гистограмму, показывающую итоги продаж по городам (рис. 2.4).

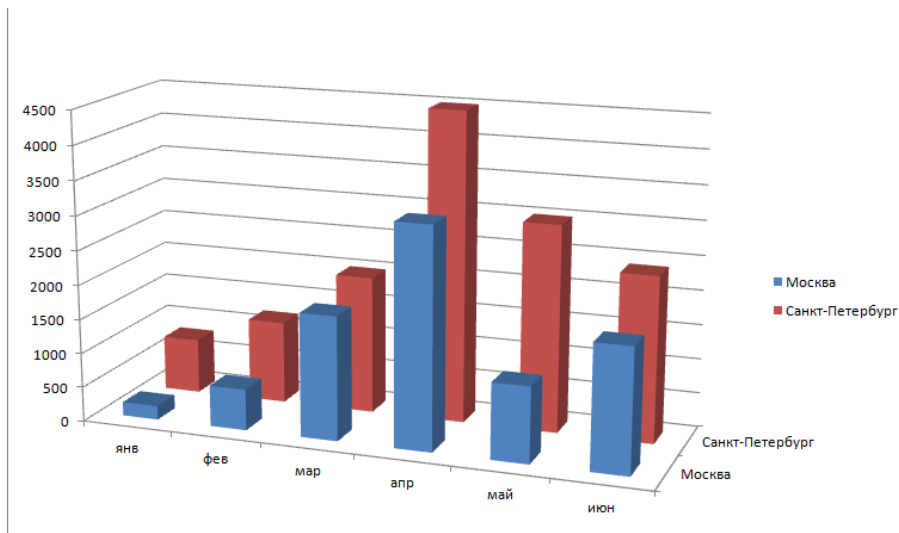


Рис. 2.4

Задание №4. Работа с OLAP кубом

Исходные данные

Имеется компания, занимающаяся продажей продуктов питания. Для хранения и последующего анализа выбираются данные, описывающие бизнес-процесс продажи товаров.

Объектами предметной области являются продавцы и продукты.

Задание

1. В MS Excel создать новую книгу “*OLAP*”.
2. В ту же папку сохранить с сайта файл *nwdata_cube.cub*.
3. Через пункт меню «*Вставка-Сводная*» таблица запустить Мастер, выбрать пункт '*использовать внешний источник данных*', нажать кнопку '*выбрать подключение*' и указать в качестве источника файл *nwdata_cube.cub*.
4. В открывшемся справа окне изучить структуру OLAP куба.
5. Используя операцию **Сечения**, построить сводную таблицу с измерениями *Страна* (по столбцам) и *Год продажи* (по строкам). В качестве фактов взять количество проданных товаров. Агрегировать данные по странам и по годам.
6. Используя операцию **Транспонирования**, поменять местами измерения Страна и Год продажи.
7. Используя операцию **Сечения**, добавить в качестве измерения Категорию продукта.
8. Используя операцию **Детализации**, отобразить в таблице названия компаний и названия продуктов.
9. Используя операцию **Свертки**, скрыть названия продуктов по всей таблице.

10. Добавить агрегированное значение по сумме продаж по странам Аргентина и Австрия. Названную группу назвать «Страны на А». Установить показ промежуточных итогов в нижней части группы.

11. Сравнить полученный результат с рис. 3.1. Убедиться в невозможности изменять заданную в кубе иерархию уровней.

12. Скопировать полученную сводную таблицу на новые листы и построить на них диаграммы, позволяющие получить ответы на следующие вопросы (предварительно при необходимости внести изменения в структуру скопированных таблиц):

а) Доли пяти самых крупных стран продавцов по итоговым продажам (рис. 3.2.).

б) Объемы продаж самых продаваемых товаров из каждой товарной категории по годам (рис. 3.3.).

с) Динамика продаж морепродуктов в США по месяцам (рис. 3.4.).

	A	B	C	D	E
1	Sum Of Quantity	Названия столбцов			
2	Названия строк	1996	1997	1998	Общий итог
3	Страны на А				
4	Argentina				
5	Beverages		3	79	82,
6	Condiments		10	35	45,
7	Confections		29	28	57,
8	Dairy Products		3	51	54,
9	Grains/Cereals			20	20,
10	Produce		19	14	33,
11	Seafood		30	18	48,
12	Argentina Итого		94	245	339,
13	Austria				
14	Beverages	188	335	459	982,
15	Condiments	184	410	126	720,
16	Confections	65	393	117	575,
17	Dairy Products	212	430	385	1027,
18	Grains/Cereals	60	220	300	580,
19	Meat/Poultry	14	283	65	362,
20	Produce	99	201	88	388,
21	Seafood	127	75	331	533,
22	Austria Итого	949	2347	1871	5167,
23	Страны на А Итого	949,	2441,	2116,	5506,
24	Другие				
25	Belgium				
26	Beverages	12	92	168	272,
27	Condiments		60	87	147,
28	Confections	40	123	107	270,
29	Dairy Products	65	110	120	295,
30	Grains/Cereals		67	78	145,
31	Meat/Poultry	40	14	35	89,
32	Produce	28	50	20	98,
33	Seafood			76	76,

Рис. 3.1

Продажи в %

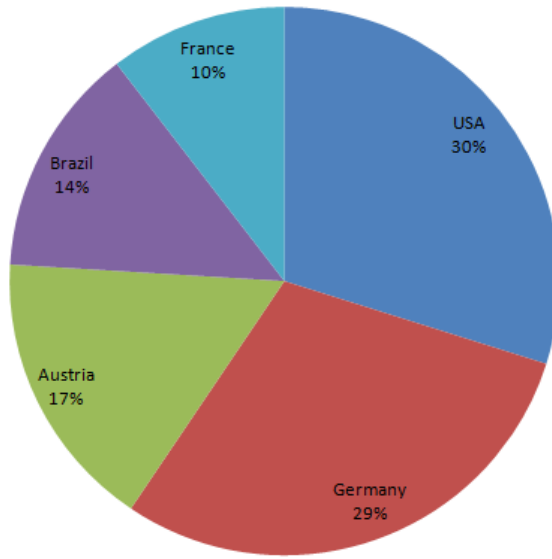


Рис. 3.2

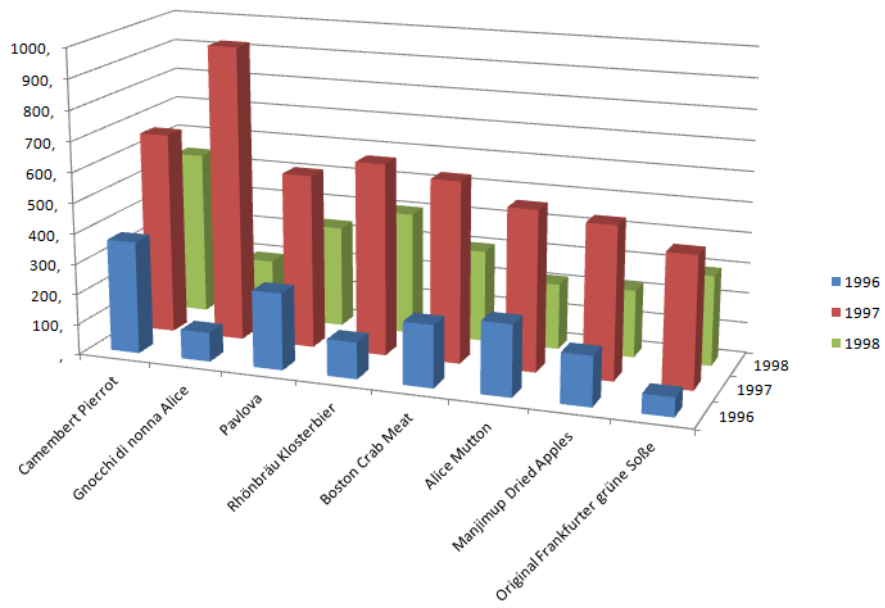


Рис. 3.3

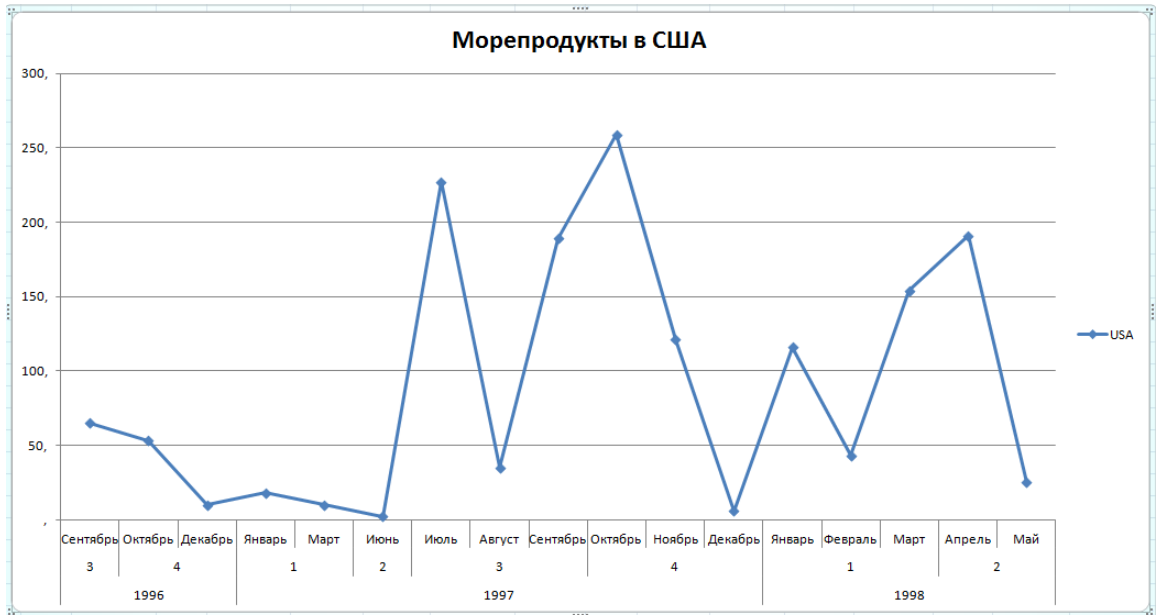


Рис. 3.4

Задание №5. Многоуровневое хранение данных.

На основе выданного задания составить:

1. Разбить данные на классы.
2. Составить таблицу с описанием классов.
3. Разработать регламент миграции данных по классам.

Пример таблицы приведен на рис. 4.1.

Класс данных	Идентификация класса по атрибутам						Регламент			
	приложение	владелец	файлы	путь	объем	дата создания	дата последнего доступа	скорость доступа	доступность	Политика жизненного цикла
критичные данные для бизнеса	SAPWebAS	*	*	/sap		*	*	15ms	99.99%	
	DB2	db2admin	*.dat	/db2		*	*	15ms	99.99%	
	-	domain\accounting*	*.xls	/home			<6 месяцев	40ms	99%	Перевести в класс важных файлов, если не было доступа в течение 6 месяцев
Важные файлы	-	domain\accounting*	*.xls	/home			>6 месяцев	120ms	99%	
Электронная Почта	LotusDomino	logistics*	*	*	>20MB	<6мес	>30дней			перевести в класс архив почты, если не было доступа в течение 30 дней и размер сообщения больше 20MB
Архив почты								3min	98%	
временные файлы	-	-	*.tmp, *.log, *.dmp, *_* tmp*	*		-	>7 дней	-	-	удалить
	-	-	*	/tmp		-	>7 дней	-	-	удалить
дублированные файлы	-	-					>30дней			удалить
ненужные файлы			*.mp3	*						удалить

Рис. 4.1.