**Требования к содержанию и структуре курсовых работ по дисциплине**

**«Анализ данных»**

Курсовая работа по дисциплине «Анализ данных» должна включать следующие компоненты:

1. Титульный лист
2. Содержание
3. Введение
4. Информативную часть
5. Заключение
6. Список литературы
7. Приложения

***Титульный лист***оформляется в соответствии с образцом на сервере в папке кафедры «Прикладная информатика».

***Содержание*** включает наименование разделов, подразделов и пунктов с указанием номера начальной страницы.

***Введение***содержит общую характеристику решаемой проблемы, обоснование ее решения и сведения о связях данной проблемы с другими, а также обоснование выбора средств анализа данных. Объем введения 1-2 страницы.

***Информативная часть***состоит из *разделов* и *подразделов,* раскрывающих суть проблемы. Объем этой части курсовой работы 50 - 60% от общего объема работы. Каждый раздел курсовой работы должен заканчиваться выводами.

*Разделы информативной части* должны быть следующими:

1. Область применения
2. Теоретическое обоснование модели анализа данных
3. Проектная часть
4. Реализация модели анализа в Deductor

При необходимости структура информационной части курсовой работы может быть изменена с учетом особенностей решаемой задачи.

Раздел *Область применения* должен содержать постановку задачи, включая описание проблемной области, актуальности проекта, цель создания модели, решаемые с помощью нее задачи, ожидаемый эффект от разработки и внедрения проекта.

Раздел *«Теоретическое обоснование модели анализа данных»* включает подбор методов анализа данных, необходимых для решения задач исследования:

1. Обзор существующих моделей для решения задачи анализа.
2. Формальное Описание моделей, включая формализацию расчетов и алгоритмы.

*Раздел* «*Проектная часть*» должна включать подразделы:

* 1. Описание входных данных
	2. Разработка структуры хранилища данных
	3. Описание процесса ETL

В подразделе *Описание входных данных* приводится перечень источников данных с указанием их типов, особенностей, описываются структура, формат и содержимое каждого источника данных.

В подразделе *Разработка структуры хранилища данных* описывается структура хранилища проекта и приводится обоснование принятых решений по выбору типа архитектуры хранилища, схемы реляционного хранилища данных (звезда или снежинка). Особое внимание следует уделить структуре компонентов хранилища данных, с уточнением типа данных, наименования, метки и размера каждого атрибута. Если в модели присутствуют кубы, то их так же следует описать с указанием условий среза. Для описания структуры таблиц хранилища данных и связей между ними рекомендуется использовать модель ERD в нотации DM (Dimensional Modeling).

В подразделе *Описание процесса ETL* описываются алгоритмы, в соответствии с которыми производится извлечение данных из источников, их очистка, преобразование и загрузка в хранилище. Особое внимание следует уделить последовательности заполнения таблиц в хранилище данных.

Раздел *“*Реализация модели анализа в Deductor» описывает

* 1. Описание компонентов проекта
	2. Описание сценариев проекта Deductor
	3. Описание выходных отчетов

Подраздел *Описание компонентов проекта* содержит описание файлов-сценариев, файлов - источников, хранилищ данных и прочих физических объектов проекта, а также зависимости между ними.

В подразделе *Описание сценариев* *проекта Deductor* приводится пошаговое описание построения сценариев в Deductor с приведением соответствующих интерфейсов.

В подразделе *Описание выходных отчетов* описываются отчеты, получаемые на выходе системы.

***Заключение*** должно содержать выводы о проделанной работе и рекомендации по ее использованию и совершенствованию. Объем заключения 5 - 10% от общего объема работы.

***Список литературы*** располагается с новой страницы в алфавитном порядке, должен содержать полные названия литературных источников с указанием фамилий и инициалов авторов, полное название издательства и года издания. Литература должна быть издана не более чем за пять лет до защиты курсовой работы.

***Приложения****.* Каждое приложение должно быть размещено с новой страницы с указанием в правом верхнем углу слова «Приложение», его номера и иметь тематический заголовок. Нумерация приложений выполняется в порядке ссылок на них. В приложении приводится вспомогательная информация, включая макеты входных и выходных данных, диаграммы.

**Тематика и примерные варианты курсовых работ по дисциплине** **«Анализ данных»**

**Вариант 1. Приложение для анализа отгрузки товаров со складов**

Необходимо создать OLAP-приложение «Анализ отгрузки со склада» с отчетами:

1. OLAP-отчет «Анализ отгрузки со складов».

2. Таблица «Количественный анализ отгрузки различных товаров».

3. Кластерный анализ «Анализ клиентов по типу договорных отношений».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Выявить склад, обеспечивший наибольшую отгрузку товара в ноябре 2000 года. Определить тройку клиентов, которым отгрузили товара на этом складе на большую сумму. Проанализировать динамику объемов отгрузки свитеров в первой половине ноября 2000 года с этого склада.

2. Провести анализ отгруженного товара по различным складам за первые 5 дней ноября 2000 года. Определить, на какую сумму в этот период времени были закуплены юбки. С помощью круговой диаграммы определить, какую долю составляют юбки от общего объема отгруженных товаров в этот период.

3. Распределить всех клиентов на группы по типу договорных отношений. Определить, на какую сумму со складов было опушено товаров по предоплате в ноябре 2000 года.

**Вариант 2. Приложение для анализа транспортных затрат**

Необходимо создать OLAP-приложение «Анализ транспортных затрат» с отчетами:

1. OLAP-отчет «Анализ транспортных затрат».

2. Таблица «Количественно-стоимостной анализ поставки товаров».

3. Кластерный анализ «Анализ средних цен на товары».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Выяснить, какой из типов транспортировки товара более дорогой для поставщиков. На какую сумму в ноябре 2000 года автомобильным транспортом было поставлено на склады различных видов обуви. С помощью линейной диаграммы проанализировать изменение объемов поставок для 1-го склада железнодорожным транспортом.

2. Провести анализ количества поставленного товара ценой не более 200 рублей за единицу продукции. Проанализировать динамику изменения стоимости поставок этой продукции на склады. Кто является основным поставщиком такой продукции. Какая доля продукции этого производителя поступила на 1-ый склад.

3. Сравнить средние цены на поставляемые товары и определить 3 наиболее дорогих товара.

**Вариант 3. Приложение для анализа бухгалтерского баланса**

Необходимо создать OLAP-приложение «Анализ бухгалтерского баланса» с отчетами:

1. OLAP-отчет «Бухгалтерский баланс».

2. Таблица «Постатейный анализ активов и пассивов за 1998 год».

3. Тренд «Анализ динамики изменения активов».

4. Кластерный анализ «Постатейный анализ структуры активов».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Проверить сходимость балансовых данных за весь отчетный период. Отобразить на круговой диаграмме структуру активов за октябрь 2000 года с указанием доли каждой статьи в процентах. Проанализировать динамику изменения пассивов, прошедших по статье «Акционерный капитал» за 2000 год по месяцам.

2. Отобразить в таблице балансовые данные за первый квартал 1998 года. Определить, какая часть активов за этот период прошла по статье «Касса».

3. Исследовать динамику изменений активов по статье «Дебиторская задолженность» за первые 5 месяцев 2000 года.

4. Выявить статьи активов, суммарный показатель которых превышает 2 млн. рублей.

**Вариант 4. Приложение для анализа клиентской базы кредитного учреждения**

Необходимо создать OLAP-приложение «Анализ клиентской базы кредитного учреждения» с отчетами:

1. OLAP-отчет «Анализ остатков и оборотов по счетам контрагента». При этом для измерения контрагента должно выбираться единственное значение.

2. Таблица «Анализ объемов кредитования различных отраслей по филиалам».

3. Тренд «Анализ динамики остатков по счетам контрагента «Инкар».

4. Кластерный анализ «Анализ средних кредитов по счетам в разрезе индустрии».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. С помощью линейной диаграммы проанализировать изменение остатков по счетам контрагента «Тарос», начинающихся с 0003, за отчетный период по месяцам.

2. Определить, какой отрасли выдается большее количество кредитов в каждом из 4 филиалах банка.

3. Проанализировать динамику изменения остатков по счетам контрагента «Инкар» за период с мая по октябрь 2002 года.

4. Определить, в какой отрасли размер среднего кредита по счету меньше.

**Вариант 5. Приложение для анализа закупочной деятельности магазинов спортивных товаров**

Необходимо создать OLAP-приложение «Анализ закупки спортивных товаров» с отчетами:

1. OLAP-отчет «Анализ закупки спортивных товаров».

2. Таблица «Количество закупленных менеджерами товаров».

3. Тренд «Динамика закупки спортивных товаров по кварталам».

4. Кластерный анализ «Стоимостной анализ закупок товаров в разрезе поставщиков».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Определить, какой вид товара был самым закупаемым весной 2003 года. Ранжировать производителей этого товара по их популярности среди закупщиков.

2. Выявить самого активного менеджера по закупке товаров за второе полугодие 2003 года. Проанализировать, благодаря закупке какого товара он вышел на первое место в этот период времени.

3. Исследовать динамику изменения объемов закупки велосипедов за три первые квартала 2003 года.

4. Провести анализ поставщиков по их доле в товарообороте. Определить самого крупного среди них.

**Вариант 6. Приложение для контроля количества бытовых приборов на складах**

Необходимо создать OLAP-приложение «Контроль количества товаров на складах» с отчетами:

1. OLAP-отчет «Анализ номенклатуры товаров на складах».

2. Таблица «Анализ общей стоимости хранимых товаров».

3. Тренд «Динамика поступлений товаров на склады».

4. Кластерный анализ «Сравнение стоимости принятых на хранение товаров».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Сравнить загруженность складов по кварталам. Выявить склад, через который за третий квартал 2003 года прошел наименьший объем товаров. Проанализировать номенклатуру товаров, прошедших за третий квартал на этом складе, и определить тройку самых популярных электротоваров.

2. Проанализировать доли различных категорий товаров в общей стоимости. Определить какой товар занимал большую долю общей стоимости в первом полугодии 2003 года и как эта сумма разделена по различным регионам.

3. Сравнить динамику u1087 поступлений товаров на склады за летние месяцы 2003 года.

4. Сравнить стоимость принятых на хранение товаров. Определить пятерку самых дорогостоящих товаров.

**Вариант 7. Приложение для анализа соотношения цена-качество компьютеров отечественного производства**

Необходимо создать OLAP-приложение «Качественный анализ отечественных компьютеров» с отчетами:

1. OLAP-отчет «Анализ качества выпускаемой продукции».

2. Таблица «Анализ цен на компьютерную технику».

3. Тренд «Динамика изменения объема выпуска компьютеров в различных регионах».

4. Кластерный анализ «Стоимостной анализ отечественной компьютерной техники».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Проанализировать изменение объема производства компьютерной техники в 2000 году по сравнению с 1999 в различных регионах. Определить, повысилось ли качество выпускаемой продукции в 2000 году. Какой процент от общего объема продукции составляет брака у Московских производителей.

2. Проанализировать разброс цен на все мониторы. Определить самый дорогой отечественный монитор. И с помощью круговой диаграммы определить, какую долю в выпуске этих мониторов занимают производители Санкт-Петербурга.

3. Сравнить динамику изменения объема выпуска компьютеров в Москве и Санкт-Петербурге за 1999-2000 года.

4. Определить перечень товаров, сумма от продажи которых составляет 80% от общей стоимости всей продукции.

**Вариант 8. Приложение для анализа потребления электроэнергии на промышленных предприятиях**

Необходимо создать OLAP-приложение «Анализ потребления электроэнергии» с отчетами:

1. OLAP-отчет «Анализ потребления электроэнергии промышленными объектами».

2. Таблица «Расход электроэнергии в регионах».

3. Кластерный анализ «Потребление электроэнергии различными промышленными объектами».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Рассчитать тариф потребляемой электроэнергии. Проанализировать объемыпотребления электроэнергии различными промышленными объектами. Определитьдесятку самых энергоемких промышленных объектов. Определить какую долю срединих занимают объекты города Москвы. Сравнить динамику потребления электроэнергии в Москве и Санкт-Петербурге в первую декаду января 2000 года.

2. Сравнить объем потребления электроэнергии в различных регионах. Определить, как распределилась потребляемая электроэнергия по промышленным объектам Санкт-Петербурга. С помощью линейной диаграммы проанализировать динамику потребления электроэнергии промышленными объектами Москвы в первую декаду января 2000 года.

3. Сравнить количество расходуемой электроэнергии различными промышленными объектами. Определить десятку наименее энергоемких промышленных объектов.

**Вариант 9. Приложение для анализа потребления расходных материалов сотрудниками супермаркета Маркет +**

Необходимо создать OLAP-приложение «Анализ потребления расходных материалов» с отчетами:

1. OLAP-отчет «Анализ потребления расходных материалов».

2. Таблица «Анализ потребления расходных материалов различными категориями пользователей».

3. Тренд «Динамика изменения объема потребляемых расходных материалов различных категорий».

4. Кластерный анализ «Анализ денежных средств, затрачиваемых на расходные материалы».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Проанализировать суммы, потраченные на различные категории расходных материалов за 4 первых месяца 2000 года. Определить на какую категорию расходных материалов тратятся большие суммы. Выявить какие материалы входят в эту категорию и количество какого из них требуется меньше всего. Определить, сотрудникам какой категории требуется этот расходный материал.

2. Проанализировать объемы потребления расходных материалов различными категориями пользователей. Определить категорию пользователей, потребляющую наибольше количество расходных материалов и выявить самый потребляемый ими товар.

3. Проанализировать динамику изменения расхода униформы за первый квартал 2000 года.

4. Сравнить суммы, затрачиваемые на приобретение различных расходных материалов. Определить пятерку материалов, на которые расходуется большее количество средств.

**Вариант 10. Приложение для анализа работы риелторской конторы**

Необходимо создать OLAP-приложение «Анализ аренды помещений» с отчетами:

1. OLAP-отчет «Анализ аренды помещений различными категориями пользователей».

2. Таблица «Анализ целевой аренды помещений».

3. Тренд «Динамика изменения стоимости аренды на различные категории помещений».

4. Кластерный анализ «Средний срок аренды на различные категории помещений».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Проанализировать аренду помещений различными категориями пользователей в зависимости от площади снимаемых объектов. Определить какие типы помещений пользуются большей популярностью у юридических лиц. Выявить, для каких целей в первом полугодии 2003 года в основном снимались помещения юридическими лицами.

2. Проанализировать назначение арендуемых помещений в зависимости от площади. С помощью долевого графика определить долю, полученную от сдачи складов, во втором полугодии 2003 года.

3. Проанализировать динамику изменения сумм, полученных в результате сдачи в аренду различных категорий помещений, за 3 квартала 2003 года. Определить категорию помещений, сдача в аренду которых приносит наиболее стабильную прибыль.

4. Проанализировать средние сроки аренды различных категорий помещений. Определить категорию помещений, которая сдается на наиболее длительный срок.

**Вариант 11. Приложение для анализа расходов на заработную плату сотрудникам сети кафе г. Москвы**

Необходимо создать OLAP-приложение «Анализ расходов на заработную плату» с отчетами:

1. OLAP-отчет «Заработная плата сотрудников сети кафе».

2. Таблица «Анализ сумм выделенных на выплату зарплаты в филиалах».

3. Тренд «Динамика расходов на зарплату различным категориям сотрудников».

4. Кластерный анализ «Заработная плата сотрудников за 2004 год».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Определить месяц, в который была выделена наименьшая сумма для выплаты заработной платы сотрудникам сети кафе. Проанализировать, как она распределилась между сотрудниками различных категорий. Определить, какая часть из этой суммы пошла на заработную плату менеджерам и как она распределилась между менеджерами различных филиалов.

2. С помощью круговой диаграммы определить долю заработной платы сотрудников главного филиала во втором полугодии 2004 года занимала. Определить, как эта сумма распределилась между сотрудниками разных профессий. С помощью линейной диаграммы проанализировать динамику изменения суммы, выделяемой на заработную плату сотрудникам главного филиала в 2004 году.

3. Проанализировать динамику изменения средств, выделяемых на выплату заработной платы главному бухгалтеру за 2004. Определить насколько процентов изменилась сумма, выделяемая на заработную плату главному бухгалтеру, во втором квартале.

4. Определить десятку самых высокооплачиваемых сотрудников.

**Вариант 12. Приложение для анализа оптовых закупок сумок сетями спортивных магазинов**

Необходимо создать OLAP-приложение «Анализ оптовых продаж сумок» с отчетами:

1. OLAP-отчет «Анализ оптовых продаж сумок».

2. Таблица «Количественный анализ оптовых продаж».

3. Тренд «Динамика изменения объема продаж различными менеджерами».

4. Кластерный анализ «Сравнение объема закупок различными сетями спортивных магазинов».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Определить тройку самых дорогих закупаемых товаров. Определить месяц 2000 года, в который этих товаров было закуплено на большую сумму. С помощью круговой диаграммы проанализировать, как объем закупленных в январе тройки самых дорогих товаров распределился между сетями магазинов. Просмотреть детальные данные, расшифровывающие значение суммы, уплаченной за эти товары в январе.

2. Проанализировать каким образом распределился между сетями магазинов Москвы общий объем закупаемых товаров. С помощью диаграммы типа «Площадь» определить самый закупаемый товар сетями магазинов Москвы.

3. Проанализировать динамику изменения объемов продаж менеджером Сидоровым А. за 4 квартала 2000 года.

4. Выявить сети магазинов, объем закупок которых превышает 10 тыс. единиц. С помощью круговой диаграммы определить долю сети магазинов «Спорт товары» в общем объеме закупок.

**Вариант 13. Приложение для анализа грузовых перевозок**

Необходимо создать OLAP-приложение «Анализ доходности грузовых перевозок» с отчетами:

1. OLAP-отчет «Анализ доходности грузовых перевозок».

2. Таблица «Количественный анализ грузовых перевозок».

3. Тренд «Динамика \_\_\_\_изменения доходности от перевозок в различных типах вагонов».

4. Кластерный анализ «Сравнение доходности перевозки различных грузов».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Определить тройку самых доходных типов груза. Определить какой из этих грузов в большем объеме перевозили во втором и четвертом кварталах 2003 года. Определить в каких вагонах в основном осуществлялась перевозка автомобилей во втором полугодии 2003 года, и какой доход был получен от их перевозки в сентябре.

2. Определить в каком направлении (получатель) отправляют грузы на платформах и какие это грузы. Определить какой груз в основном отправляется в Москву, из каких городов и в каком объеме он поступает.

3. Проанализировать изменение доходности от перевозки грузов в крытых вагонах за 4 квартала 2004 года. Определить насколько процентов возросла доходность от перевозки грузов в крытых вагонах в 4-ом квартале по сравнению с 3-им.

4. Определить десятку самых доходных грузов. Выявить сколько из них входят в список грузов, обеспечивающих 80 % дохода от перевозок в целом.

**Вариант 14. Приложение для анализа использования автотранспорта для доставки товаров в магазины**

Необходимо создать OLAP-приложение «Анализ использования автотранспорта для доставки товаров в магазины» с отчетами:

1. OLAP-отчет «Анализ использования автотранспорта для грузоперевозок».

2. Таблица «Анализ работы водителей».

3. Тренд «Динамика изменения километража».

4. Кластерный анализ «Частота поездок водителей за второе полугодие».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Определить два месяца 2004 года, за которые было сделано наибольшее количество поездок. Какой товар чаще всего перевозили в эти месяцы. Определить фамилию водителя и марку автомобиля, лидирующего по количеству рейсов по итогам двух месяцев.

2. Определить водителя, выполнившего за 2004 год наибольшее количество поездок, тот ли это водитель, который лидирует по километражу. Какой груз в основном перевозил водитель, наездивший наибольшее количество километров, в летние месяцы 2004 года.

3. Проанализировать динамику изменения километража автомобиля «Renault» за период с мая по ноябрь 2004 года. Определить суммарный пробег всех автомобилей за декабрь 2004 года.

4. Определить список водителей, которые за второе полугодие 2004 года выполнили меньше 25 поездок с грузом.

**Вариант 15. Приложение для учета кадров на предприятии**

Необходимо создать OLAP-приложение «Учет кадров на предприятии» с отчетами:

1. OLAP-отчет «Уровень образования работников филиалов предприятия». При этом для измерения «Филиал» должно выбираться единственное значение.

2. Таблица «Учет кадров предприятия».

3. Тренд «Динамика изменения количества работников в филиалах».

4. Кластерный анализ «Минимальный возраст работников предприятия различных профессий».

С помощью созданных отчетов приложения необходимо решить следующие практические задачи:

1. Проанализировать текучесть кадров в главном филиале предприятия за четыре квартала 2003 года. Выяснить, как изменился уровень образования сотрудников главного филиала в четвертом квартале 2003 года по сравнению с третьим. Сотрудники каких профессий с высшем образованием в четвертом квартале уволились из главного филиала предприятия.

2. Какую долю от общего количества сотрудников занимают сотрудники с высшим образованием, сотрудники Пермского филиала, водители\_\_\_\_\_. Определить, сотрудники какой профессии, получившие только обязательное образование, преобладают на предприятии.

3. Проанализировать динамику изменения количества работников в главном филиале предприятия в различные кварталы 2003 года. Определить, как изменилось количество работников этого филиала во втором квартале по сравнению с первым.

4. Определить какую должность занимает самый молодой сотрудник предприятия.