

**МИНОБРНАУКИ РОССИИ**

Федеральное государственное бюджетное образовательное учреждение высшего образования

**«МИРЭА – Российский технологический университет»**

|  |
| --- |
| **РТУ МИРЭА** |
|  |
| **Институт кибербезопасности и цифровых технологий (ИКБ)** |
|  |
| КБ-2 «Информационно-аналитические системы кибербезопасности» |

**ОТЧЕТ О ВЫПОЛНЕНИИ ПРАКТИЧЕСКОЙ РАБОТЫ №1**

# **В РАМКАХ ДИСЦИПЛИНЫ «МЕТОДЫ И СРЕДСТВА ПРОЕКТИРОВАНИЯ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ СИСТЕМ»**

Выполнил:

Студент 3-ого курса

Учебной группы БИСО-02-22

Зубарев В.С.

Работа с терминологической базой и источниками

Цель работы: сформировать понятийный аппарат в ходе самостоятельной работы студента с источниками, в том числе представленными на сайтах вендоров проектных средств Используя имеющиеся у Вас знания и различные источники, в том числе справочные интернет-ресурсы, ответьте на поставленные вопросы, исключая не относящиеся к предмету ИАС понятия. Ответы дополните ссылочным материалом

1. **Дайте определение понятий извлечение, преобразование, загрузка данных.**

Извлечение — это процесс получения данных из различных источников и, зачастую, в различных формах представления.

Преобразование — это процесс приведения данных к общему, заранее оговоренному, пригодному для загрузки в систему хранения виду.

Загрузка данных — это процесс загрузки данных в систему хранения.

1. **Поясните понятие витрина данных.**

Витрина данных — это организованное хранилище данных, в которой собраны необходимые материалы по конкретной теме.

https://www.adventum.ru/articles/upravlenie-dannymi/chto-takoe-vitrina-dannyh/#chto-takoe-vitrina-dannyh

1. **Укажите различие между операционной базой данных и оперативной базой данных.**

Операционная база данных – это система, настроенная на хранение большого объема данных на энергозависимых носителях и поддержку не требовательных операций, необходимых для работы системы.

Оперативная база данных – это система, настроенная на быструю обработку данных, в которой основной объем помещен в энергозависимой памяти (ОЗУ), что позволяет выполнять гораздо более требовательные операции, за более короткое время

1. **Дайте определение понятию хранилище данных.**

Хранилище данных — это система которая объединяет в себе множество данных по разным направлениям и из разных источников.

https://ru.wikipedia.org/wiki/Хранилище\_данных

1. **Поясните назначение OLTP информационных систем.**

OLTP системы нужны для оперативной записи данных в базу данных (бронирование билетов, прием ставок)

1. **Поясните назначение OLAP информационных систем.**

OLAP системы нужны для анализа данных в реальном времени (СППР, прогнозирование)

1. **Назовите методы интеграции (объединения) данных организации.**

Консолидация  
Федерализация  
Распространение данных

Сервисный подход

https://ru.wikipedia.org/wiki/Интеграция\_данных

1. **Поясните аббревиатуру FASMI.**

Fast Analysis of Shared Multidimensional Information**.** FASMI – тест который определяет соответствует ли OLAP система установленным правилам.  
**Fast (быстрый).** Система должна обеспечивать выдачу большинства ответов пользователям в сжатые сроки.

**Analysis (анализ).** Система может справляться с любым логическим и статистическим анализом, характерным для данного приложения, и обеспечивает его сохранение в виде, доступном для конечного пользователя.

**Shared (разделяемый).** Система осуществляет все требования защиты конфиденциальности (до уровня ячейки) и, если множественный доступ к записи необходим, обеспечивает блокировку модификаций на соответствующем уровне. **Multidimensional (многомерный).** Система должна обеспечить многомерное концептуальное представление данных, включая полную поддержку для иерархий и множественных иерархий.<https://en.wikipedia.org/wiki/FASMI>

https://wiki.loginom.ru/articles/fasmi.html

1. **Дайте определение понятий Knowledge Discovery in Databases**

Knowledge Discovery in Databases — это процесс получения из данных знаний в виде закономерностей, правил, моделей.

Data Mining — это обнаружение в «сырых» данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний.

Различие между KDD и Data Mining в том, что KDD — это общий процесс извлечения знаний из данных, а Data Mining — это шаг внутри этого процесса, который занимается выявлением закономерностей в данных.

1. **Перечислите этапы анализа данных.**

Определение проблемы

Сбор данных

Очистка данных

Данные анализа

Предварительный просмотр данных

Представление данных

https://www.geeksforgeeks.org/six-steps-of-data-analysis-process/

1. **Какие основные задачи решает информационно-аналитические системы, перечислите их?**

Информационно - аналитические системы решают задачи обработки, хранения и представления данных в необходимом для пользователя виде**.**

1. **Перечислите известные Вам информационно-аналитические системы.**

Tableau

Apache

Power BIhttps://www.geeksforgeeks.org/top-10-best-analytics-platforms/

1. **Какую задачу решает классификация данных, назовите известные Вам методы (алгоритмы) классификации?**

Классификация решает задачу декомпозиции множества объектов на более меньшие группы, обладающие общими свойствами.  
Алгоритмы  
Метод ближайших соседей  
Дерево решений

1. **Какую задачу решает кластеризации данных, назовите известные Вам методы (алгоритмы) кластеризации?**

Кластеризация решает задачу нахождения неявных структур в объектах.

Алгоритмы  
К-средних  
DBSCAN

1. **Какую задачу решает ассоциации данных, назовите известные Вам методы (алгоритмы) ассоциации?**

Задача ассоциаций данных — выявление закономерностей между связанными событиями в базах данных.

Алгоритмы

AIS

SETM

https://spravochnick.ru/informatika/algoritmy\_poiska\_associativnyh\_pravil/

1. **Какую задачу решает выявление последовательностей, назовите известные Вам методы (алгоритмы) выявления последовательностей?**

Выявление последовательностей позволяет решать задачу прогнозирования ситуации при схожих условиях, основываясь на выявленных ранее последовательностях событий.

Алгоритмы

Последовательный поиск

1. **Какую задачу решает прогнозирование, назовите известные Вам методы (алгоритмы) прогнозирования?**

Задача прогнозирования – это на основе известных данных, построить вариант\варианты будущего, на основе которого

1. **Поясните, что такое анализ отклонений?**

Анализ отклонений – это процесс, при котором выявляются отклонения от средних значений. Это позволяет находить экстремумы поведения и анализировать качество проведенных измерений.

1. **Какие уровни информационно-аналитических систем Вам известны?**

Сбор данных

Извлечение данных

Складирование данных

Представление данных

Анализ данных

1. **Какие средства разработки ИАС Вам известны (если не знакомы ни с одним так и пишите)?**

Не знаком

1. **В каких средствах разработки ИАС в Вас есть опыт работы, в каком качестве/какой роли?**

Нет опыта

1. **Раскройте аббревиатуру ETL.**

Extract Transform Load – общий термин, описывающий путь данных от источника до базы данных, с приведением данных к стандартизированному виду.

1. **Какие инструменты ETL Вам известны.**

Apache Kafka

1. **Назовите известные Вам высокоуровневые средства описания модели ИАС.**

Диаграммы

Графические бизнес-модели

1. **Какие уровни информационно-аналитических систем Вам известны?**

Сбор данных

Извлечение данных

Складирование данных

Представление данных

Анализ данных

1. **Опишите кратко понятие факт таблица, таблица размерности, консольная таблица.**

Таблица факта — это центральная таблица в схеме «звезда». Она может состоять из миллионов строк и содержать суммирующие или фактические данные, которые могут помочь ответить на требуемые вопросы.

Таблицы размерности (иногда их называют таблицами измерений) окружают таблицу факта и содержат данные по признакам. Они являются родительскими по отношению к таблице факта, которая является дочерней.

Консольные таблицы присоединяются к таблицам размерности и детализируют отдельные атрибуты. Они являются родительскими по отношению к таблицам размерности.

1. **Что означает понятие Data Mart?**

Data Mart (витрина данных) — база данных, предназначенная для решения специализированной задачи или набора задач из одной предметной области. Например, по поиску наименьшей цены товара, расчёту загрузки производственных мощностей предприятия, организации тематических рассылок и т. п.