**Инвариантная самостоятельная работа 4**

**Что такое база данных?**

**База данных** — совокупность данных, хранимых в соответствии со схемой данных, манипулирование которыми выполняют в соответствии с правилами средств моделирования данных.

**Что такое банк данных?**

**Банк данных (БнД)** — это автоматизированная система специ­альным образом организованных данных — баз данных, программ­ных, технических, языковых, организационно-методических средств и персонала, предназначенных для обеспечения централизованного накопления и коллективного многоцелевого использования данных.

**В чем отличие базы данных от банка данных?**

База данных – это совокупность информации, довольно часто не систематизированная каким-то образом. Данные могут относиться к любой сфере (люди, финансы, контакты, отчеты) быть представлены в разном виде (текстовом, табличном, числовом и др.). Банк данных – это более широкое понятие. Сюда включается база данных + система управления этими данными. То есть информация в банке уже дается не хаотично, а структурировано. Она упорядочена по определенным критериям. К банкам данных предъявляются определенные требования – упорядоченность, удобство использования и возможность пополнения новыми базами данных.

**Что такое Big Data?**

**Большие данные** (англ. big data) — обозначение структурированных и неструктурированных данных огромных объёмов и значительного многообразия, эффективно обрабатываемых горизонтально масштабируемыми программными инструментами, появившимися в конце 2000-х годов и альтернативных традиционным системам управления базами данных и решениям класса Business Intelligence.

**Что такое СУБД?**

**Система управления базами данных, сокр. СУБД** (англ. Database Management System, сокр. DBMS) — совокупность программных и лингвистических средств общего или специального назначения, обеспечивающих управление созданием и использованием баз данных.

**Что такое аналитика больших данных?**

**Аналитика больших данных** - это разработка различных подходов к управлению бизнесом, осуществляемая на основе анализа больших данных, например, по базам данных о банковских транзакциях, телефонных номерах в колл-центрах, по уровню посещаемости клиентами интернет-магазинов и сделанных ими покупок.

Аналитика больших данных включает:

* сбор данных;
* выборку и сортировку информации для проведения аналитики;
* поиск закономерностей в наборах данных;
* виртуализацию данных для лучшего выявления тенденций и представления результатов анализа;
* формулировку прогнозов и предположений по оптимизации бизнес-процессов за счет изменений разных показателей (метрик);
* разработку и тестирование различных моделей машинного обучения;
* настройку стимулирующих решений для клиентских сервисов.

**Поставщики Big Data**

[Socrata OpenData](https://dev.socrata.com/data/)

Одна из крупнейших и наиболее мощных поисковых систем, в которой размещены тысячи наборов данных по таким темам, как финансы, инфраструктура, транспорт, окружающая среда, экономика и общественная безопасность. Все наборы данных категоризированы с помощью алгоритмов машинного обучения.

[Kaggle](https://www.kaggle.com/)

[Kaggle](https://www.kaggle.com/) является самой большой платформой и сообществом по науке о данных, которая поражает разнообразием наборов данных, соревнований и примеров проектов. Ресурс содержит в основном чистые данные, особенно если они является частью соревнования. Наборы данных можно искать по различным тегам.

[Репозиторий UCI](http://mlr.cs.umass.edu/ml/)

Обширная платформа, которая содержит наборы данных для решения задач машинного обучения. Для каждого набора данных указано его задание по умолчанию, типы атрибутов, типы данных и другие функции. Несмотря на небольшие размеры многих наборов данных, они отлично подходят для выполнения образовательных проектов.

[ProPublica](https://www.propublica.org/)

Эта американская некоммерческая организация известна тем, что предлагает универсальное [хранилище данных](https://www.propublica.org/datastore/datasets), которое охватывает вопросы здравоохранения, уголовного правосудия, образования, политики, бизнеса, транспорта и финансов. Помимо этого, данные часто обновляются.

Коллекция содержит как платные, так и бесплатные наборы данных. Платные наборы данных, в свою очередь, доступны по академическим, коммерческим, студенческим и журналистским лицензиям.

[InsideAirbnb](http://insideairbnb.com/get-the-data.html)

Сервис данных, созданный и поддерживаемый компанией [Airbnb](https://www.airbnb.com/). Он содержит уникальную коллекцию данных Airbnb, категоризированных по регионам и странам. Здесь можно просматривать данные по определенному городу и исследовать информативные отчеты с креативными визуализациями.

**Технические характеристики баз данных**

1. Производительность - определяется временем выполнения запросов, скоростью поиска информации, временем импортирования базы данных из других форматов, временем генерации отчета и др.
2. Обеспечение целостности данных на уровне базы данных - наличие средств, позволяющих удостовериться, что информация в базе корректная и полная.
3. Обеспечение безопасности - шифрование прикладных программ, шифрование данных, защита паролем, ограничение уровня доступа (к базе, к таблице и т.д.).
4. Работа в многопользовательских средах - предполагает возможность блокировки базы данных, файла, записи, поля, а также идентификацию станции, установившей блокировку, обработку транзакций - последовательности операций пользователя над базой данных, сохраняющей ее логическую целостность, работу с сетевыми операционными системами.
5. Возможность импорта - экспорта информации, подготовленной другими программными средствами.
6. Инструментальные средства разработки прикладных программ (языки программирования, средства генерации приложений, реализации меню, форм, отчетов). Реализация языковых средств осуществляется по-разному - для разработчиков программных продуктов язык представляется в явной синтаксической форме (например, в Access - язык Access Basic, в FoxPro и dBase - язык xBase). Для пользователей функции языка часто доступны косвенным образом - через меню, диалоговые сценарии и т.д. - при этом синтаксические конструкции языка формируются автоматически и передаются на исполнение.