ИСР 4. Крупнейшие базы данных мира

Оглавление

[1. Что такое база данных?](#_pdwbiuttcika)

[2. Что такое банк данных?](#_iitki4cdzdjz)

[3. В чем отличие базы данных от банка данных?](#_lgizzjz9f2ay)

[4. Что такое Big Data?](#_4ddot3hooms)

[5. Что такое СУБД?](#_3tkpuz79b5nn)

[6. Что такое аналитика больших данных?](#_3fj2e66f7kgn)

[7. Какие компании являются поставщиками Big Data в России?](#_kros3u37dng9)

[8. Какие компании являются поставщиками Big Data в мире?](#_nszao0o3hdmr)

[9. Технические характеристики баз данных.](#_gmdva0qaqv5h)

[10. Используемые ресурсы (по ГОСТ).](#_eog8cqxt7xbp)

#### Что такое база данных?

База данных - это организованная структура данных, которая обеспечивает хранение и доступ к информации. Она обычно используется для хранения и обработки данных в приложениях и системах. Базы данных могут быть разных типов, таких как реляционные, иерархические, сетевые и объектно-ориентированные. Они могут быть реализованы на различных технологиях, таких как SQL, NoSQL, NewSQL и другие.

Одним из наиболее распространенных типов баз данных является реляционная база данных. Реляционные базы данных используют таблицы для хранения данных и связи между таблицами для организации информации. Они основаны на модели данных, которая состоит из отношений (таблиц) и атрибутов (столбцов).

Реляционные базы данных обычно управляются системами управления базами данных (СУБД), которые предоставляют инструменты для создания, изменения и управления базами данных. СУБД, такие как MySQL, PostgreSQL, Oracle, Microsoft SQL Server и другие, предоставляют функции для работы с данными, такие как запросы, обновления, вставки, удаления и др.

Базы данных также могут использовать различные модели данных, такие как объектно-реляционные модели, которые сочетают объекты и отношения, или многомерные модели, которые используются для анализа данных.

В целом, базы данных являются важным компонентом многих современных приложений и систем, обеспечивая надежное и эффективное хранение и обработку данных.

#### Что такое банк данных?

Банк данных - это система управления базами данных, которая предоставляет пользователям возможность хранения, обработки и доступа к информации. Банки данных обычно используются для хранения больших объемов информации, такой как данные научных исследований, медицинские данные, финансовые данные и т.д. Банки данных могут быть централизованными или распределенными, в зависимости от потребностей пользователей.

Основными функциями банков данных являются:

Хранение данных: Банк данных предоставляет безопасное и надежное место для хранения информации, защищая ее от потери, повреждения и несанкционированного доступа.

Управление данными: Банк данных позволяет управлять данными, такими как их поиск, сортировка, фильтрация и анализ.

Доступ к данным: Банк данных обеспечивает пользователям быстрый и удобный доступ к информации, хранящейся в нем.

Безопасность данных: Банки данных обеспечивают высокий уровень безопасности данных, защищая их от несанкционированного доступа, изменения и удаления.

Интеграция данных: Банки данных позволяют объединять данные из различных источников и представлять их в удобном для анализа виде.

Управление метаданными: Банк данных хранит информацию о данных, такую как их тип, источник, формат и другие характеристики, что позволяет быстро находить и обрабатывать нужные данные.

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

Банк данных - это совокупность одной или нескольких баз данных, которые организованы и управляются вместе для достижения определенных целей. Банки данных могут включать в себя базы данных различных типов, таких как реляционные, объектно-ориентированные, графовые и другие.

Отличие базы данных от банка данных заключается в том, что база данных обычно содержит информацию одного типа, такую как записи о клиентах, товарах или заказах. Банк данных же может включать в себя несколько баз данных, каждая из которых содержит информацию разного типа.

Банки данных обычно используются для хранения и обработки больших объемов данных, а также для обеспечения доступа к ним различным пользователям. Они могут быть использованы для научных исследований, бизнес-аналитики, медицинских исследований и других целей.

Кроме того, банки данных часто имеют более сложную структуру и функциональность, чем базы данных. Они могут включать в себя системы управления базами данных, инструменты для анализа данных, системы безопасности и другие компоненты.

В целом, отличие базы данных от банка данных заключается в масштабе и функциональности. Банк данных может включать в себя множество баз данных и предоставлять более широкий спектр возможностей для работы с данными.

#### Что такое Big Data?

Big Data - это термин, который описывает большие объемы данных, которые невозможно обработать традиционными методами. Это может быть связано с различными источниками, такими как социальные сети, интернет, мобильные устройства и многое другое. Big Data используется в различных областях, таких как маркетинг, аналитика, медицина и другие.

Big Data на сегодняшний момент являются одним из ключевых драйверов развития информационных технологий. Это направление относительно новое для российского бизнеса, и оно получило широкое распространение в западных странах. Связано это с тем, что в эпоху информационных технологий, особенно после бума социальных сетей, по каждому пользователю интернета стало накапливаться значительное количество информации, что в конечном счете дало развитие направлению Big Data.

Сфера Больших Данных характеризуется следующими признаками:

1. Volume – объем, накопленная база данных представляет собой большой объем информации, который трудоемко обрабатывать и хранить традиционными способами, для них требуются новый подход и усовершенствованные инструменты.
2. Velocity – скорость, данный признак указывает как на увеличивающуюся скорость накопления данных (90% информации было собрано за последние 2 года), так и на скорость обработки данных, в последнее время стали более востребованы технологии обработки данных в реальном времени.
3. Variety – многообразие, т.е. возможность одновременной обработки структурированной и неструктурированной разноформатной информации. Главное отличие структурированной информации – это то, что она может быть классифицирована. Примером такой информации может служить информация о клиентских транзакциях.
4. Неструктурированная информация включает в себя видео, аудио файлы, свободный текст, информацию, поступающую из социальных сетей. На сегодняшний день 80% информации входит в группу неструктурированной. Данная информация нуждается в комплексном анализе, чтобы сделать ее полезной для дальнейшей обработки.
5. Veracity – достоверность данных, все большее значение пользователи стали придавать значимость достоверности имеющихся данных. Так, у интернет-компаний есть проблема по разделению действий, проводимых роботом и человеком на сайте компании, что приводит в конечном счете к затруднению анализа данных.
6. Value – ценность накопленной информации. Большие Данные должны быть полезны компании и приносить определенную ценность для нее. К примеру, помогать в усовершенствовании бизнес-процессов, составлении отчетности или оптимизации расходов.

При соблюдении указанных выше 5 условий, накопленные объемы данных можно относить к числу больших.

#### Что такое СУБД?

СУБД - это система управления базами данных, которая позволяет создавать, изменять и удалять базы данных, а также работать с содержащейся в них информацией. СУБД используются для управления данными в различных приложениях, таких как базы данных клиентов, каталоги продукции, системы учета и другие.

#### Что такое аналитика больших данных?

Аналитика больших данных - это процесс анализа больших объемов данных с целью извлечения полезной информации. Большие данные могут поступать из различных источников, таких как социальные сети, веб-сайты, мобильные устройства, датчики и другие. Аналитика больших данных используется для различных целей, таких как улучшение принятия решений, прогнозирование тенденций, обнаружение аномалий и многое другое. Для анализа больших данных используются различные методы, такие как машинное обучение, статистический анализ, визуализация данных и другие.

#### Какие компании являются поставщиками Big Data в России?

На сегодняшний день Big Data используется в следующих российских компаниях: Mail.ru, Яндекс, Вымпелком (Билайн), Мегафон, МТС, Пенсионный фонд России, ФНС России, Сбербанк, ВТБ24, Альфа-Банк, Юлмарт

#### Какие компании являются поставщиками Big Data в мире?

На сегодняшний день Big Data активно внедряются в зарубежных компаниях. Такие компании, как Nasdaq, Facebook, Google, IBM, VISA, Master Card, Bank of America, HSBC, AT&T, Coca Cola, Starbucks и Netflix уже используют ресурсы Больших Данных.

#### Технические характеристики баз данных.

Основные характеристики БД - объём памяти и необходимая производительность. В зависимости от этого для различных проектов выбираются 2-3 СУБД, которые соответствуют выявленным требованиям. Например, если объём БД не превысит 100М, большинство запросов выбирает от 1 до 20 записей и время реакции системы не должно превышать 10 секунд, то следует остановить выбор на системах среднего класса, таких как Firebird, PostgreSQL, FoxPro. Для меньших по объёму БД можно выбрать Access или MySQL, а такие серьёзные СУБД как Oracle, DB/2 или Informix следует рассматривать в тех случаях, когда велик объём данных или имеются высокие требования к производительности системы. Выбранные СУБД оцениваются по степени соответствия выявленным требованиям к БД, и выбирается та система, которая лучше им соответствует.

#### Используемые ресурсы (по ГОСТ).

1. А. Б. Градусов Базы данных. Введение в технологию баз данных. / А. Б. Градусов [Электронный ресурс] // : [сайт]. — URL: https://dspace.www1.vlsu.ru/bitstream/123456789/8790/2/02100.pdf
2. Е. И. Чирагина Базы данных / Е. И. Чирагина [Электронный ресурс] // : [сайт]. — URL: http://repo.ssau.ru/bitstream/Uchebnye-posobiya/Bazy-dannyh-Elektronnyi-resurs-ucheb-posobie-po-napravleniu-podgot-bakalavrov-23010062-Informatika-i-vychisl-tehnika-54858/1/Чигарина%20Е.И.%20Базы.pdf
3. И. П. Карпова Базы данных / И. П. Карпова [Электронный ресурс] // : [сайт]. — URL: https://publications.hse.ru/mirror/pubs/share/direct/259052819
4. Дружинин, Н. Ф. Практический анализ данных в Microsoft SQL Server / Н. Ф. Дружинин. – М.: Издательский дом КОДЕКС, 2018.