

Тема 6. Системы управления базами данных.

Базы данных и их типы. Проектирование баз данных. Основные объекты баз данных. Реляционные, иерархические и объектно-ориентированные модели. Методы организации баз данных. Централизованные и распределенные базы данных. Создание, редактирование и управление базами данных в Microsoft Access. Обсуждение проектирования, разработки и управления базами данных. Обсуждение вопросов, связанных с управлением транзакциями, хранением данных и другими. Создание отчетов в Microsoft Access.

План

База данных и ее типы.

Архитектура базы данных.

Классификация баз данных.

Системы управления базами данных.

Система управления базами данных Microsoft Access (MBBS). Методы работы с типами данных.

Базовые объекты Microsoft Access.

Технологии больших данных

База данных (МБ)— это совокупность взаимосвязанных данных, организованных, хранящихся и управляемых в соответствии с определенной структурой.

Она обеспечивает эффективное хранение, извлечение, изменение и удаление данных. Базы данных широко используются в различных областях, включая бизнес, образование, здравоохранение и государственное управление.

Основные возможности базы данных:

Порядок: данные организованы в соответствии с определенной структурой (например, таблицы, записи, поля).

Связь: между данными могут существовать связи (например, данные в одной таблице могут быть связаны с данными в другой таблице).

Хранение: данные хранятся на постоянном устройстве хранения данных (например, жестком диске, SSD).

Управление: Для управления данными используется система управления базами данных (СУБД).

Безопасность: существуют механизмы контроля доступа к базе данных и защиты данных.

Типы баз данных:

Базы данных можно классифицировать по различным критериям. Одним из наиболее распространённых методов классификации является классификация на основе модели данных.

1) Реляционная база данных: наиболее распространенный тип; данные хранятся в таблицах; таблицы состоят из строк (записей) и столбцов (полей); между таблицами могут устанавливаться связи (например, внешний ключ в одной таблице может указывать на первичный ключ в другой таблице); SQL (язык структурированных запросов) — стандартный язык для работы с реляционными базами данных.

Примеры: MySQL, PostgreSQL, Oracle, Microsoft SQL Server.

2) База данных NoSQL: базы данных, не основанные на реляционной модели; предназначены для хранения и обработки больших объемов неструктурированных или полуструктурированных данных; поддерживают различные модели данных (например, документную, ключ-значение, графовую, столбчатую); обеспечивают более высокую масштабируемость и скорость по сравнению с реляционными базами данных.

Примеры: MongoDB (документальный), Redis (ключ-значение), Neo4j (граф), Cassandra (столбчатый).

3) Объектно-ориентированная база данных: база данных, основанная на принципах объектно-ориентированного программирования; данные хранятся в виде объектов; объекты могут иметь атрибуты (поля) и методы (функции);

Примеры: ObjectDB, db4o.

4) Графовая база данных: данные хранятся в виде графа; граф состоит из узлов (объектов) и ребер (связей); удобен для анализа связей и решения сложных вопросов.

Примеры: Neo4j, Amazon Neptune.

5) База данных документов: данные хранятся в виде документов (например, JSON, XML); документы могут иметь иерархическую структуру; удобна для хранения полуструктурированных данных.

Примеры: MongoDB, Couchbase.

6) База данных «ключ-значение»: данные хранятся в виде пар «ключ-значение»; предназначена для простого и быстрого хранения и извлечения данных.

Примеры: Redis, Memcached.

Выбор базы данных. Выбор базы данных зависит от многих факторов, включая: размер и структуру данных; требования к скорости обработки данных; требования к масштабируемости; требования к безопасности; а также расходы на поддержку и управление.

Архитектура базы данных относится к компонентам системы базы данных, их взаимосвязям и общей структуре системы.

Эта концепция охватывает методы хранения, обработки и доступа к данным, а также такие важные аспекты базы данных, как безопасность, надежность и масштабируемость. Архитектура влияет на общую производительность и эффективность базы данных, поэтому важно тщательно подходить к её проектированию.

Архитектура базы данных— это концептуальная и физическая конструкция, определяющая структуру системы базы данных, её основные компоненты, их взаимосвязи и общие принципы работы системы. Она определяет методы хранения, обработки, управления и защиты данных.

Ключевые аспекты архитектуры:

1) Компоненты: Система базы данных состоит из следующих основных компонентов:

Ядро (ядро базы данных): основная часть, отвечающая за хранение, обработку и управление данными.

Словарь данных: хранит информацию о структуре базы данных, таблицах, полях, связях и других объектах.

Процессор запросов: анализирует, оптимизирует и выполняет SQL-запросы.

Менеджер транзакций: управляет транзакциями для обеспечения целостности данных.

Менеджер хранилища: управляет хранением и доступом к данным на диске.

Менеджер безопасности: управляет доступом к базе данных и защищает данные.

2) Модель данных: определяет, как организованы и хранятся данные (например, реляционная, объектно-ориентированная, NoSQL).

3) Схема данных: определяет логическую структуру базы данных, определяя таблицы, поля, связи и ограничения.

4) Язык запросов: язык, используемый для взаимодействия с базой данных (например, SQL).

5) Интерфейсы: определяют, как пользователи и другие приложения получают доступ к базе данных (например, API, ODBC, JDBC).

6) Масштабируемость и надежность: определяют, насколько хорошо база данных может справляться с нагрузкой и насколько она устойчива к сбоям.

7) Безопасность: механизмы защиты данных от несанкционированного доступа, изменения и удаления.

Важность архитектуры:

Эффективность: правильная архитектура обеспечивает быструю и эффективную обработку данных.

Масштабируемость: архитектура гарантирует, что база данных может адаптироваться к росту и увеличению рабочей нагрузки.

Надежность: Архитектура обеспечивает целостность данных и их долгосрочное сохранение.

Безопасность: Архитектура обеспечивает защиту данных и контроль доступа.

Управление: правильная архитектура упрощает управление базой данных и ее обслуживание.

При проектировании архитектуры базы данных следует учитывать бизнес-требования, характеристики данных, прогнозы рабочей нагрузки и доступные ресурсы.

Классификация базы данных Под «классификацией баз данных» понимается разделение баз данных на группы в соответствии с их характеристиками, структурой, областями применения и другими критериями. Эта классификация играет важную роль в выборе, проектировании и управлении базами данных.

Классификация баз данных — это процесс разделения баз данных на различные группы на основе их характеристик, структуры, областей применения и других критериев. Эта классификация важна для выбора, проектирования и управления базами данных. Основные классификации баз данных подробно описаны ниже:

1) По модели данных: Реляционные базы данных: данные хранятся в таблицах; связи между таблицами устанавливаются с помощью ключей; используется язык запросов SQL (язык структурированных запросов).

Примеры: MySQL, PostgreSQL, Oracle, Microsoft SQL Server.

Объектно-ориентированные базы данных: данные хранятся в виде объектов; объекты организованы по принципам классов и наследования;

Примеры: GemStone, ObjectDB.

Объектно-реляционные базы данных: сочетают реляционный и объектно-ориентированный подходы; позволяют хранить объекты в реляционных таблицах.

Примеры: Oracle, PostgreSQL.

Базы данных NoSQL: базы данных, не основанные на реляционной модели; предназначенные для хранения больших объемов неструктурированных или полуструктурированных данных; поддерживают различные модели данных (например, ключ-значение, документ, граф).

Примеры: MongoDB, Cassandra, Redis, Neo4j.

2) По способу хранения данных: централизованные базы данных. Данные хранятся на одном центральном сервере. Пользователи подключаются к центральному серверу и получают доступ к данным.

Распределенные базы данных: данные хранятся на нескольких компьютерах; данные могут находиться в разных местах; доступ к данным осуществляется по сети.

Облачные базы данных: базы данных, размещенные на облачных платформах; обеспечивают масштабируемость, гибкость и скорость;

Примеры: Amazon RDS, Google Cloud SQL, база данных Azure SQL.

3) По области применения: Операционные базы данных: предназначены для обработки транзакций в режиме реального времени. Примеры: банковские системы, платформы электронной коммерции.

Хранилища данных: предназначены для аналитических запросов и отчетов; объединяют данные из разных источников; Примеры: Teradata, Amazon Redshift.

Озера данных: предназначены для хранения данных в любом формате; включают неструктурированные, полуструктурированные и структурированные данные.

Примеры: Hadoop, Amazon S3.

4) По количеству пользователей: Однопользовательские базы данных: используются только одним пользователем. Примеры: Microsoft Access.

Многопользовательские базы данных:

Используется несколькими пользователями одновременно.

Примеры: Oracle, MySQL.

5) По изменчивости данных: Статические базы данных: данные изменяются редко. Примеры: архивные данные. Динамические базы данных: данные изменяются часто. Примеры: социальные сети, интернет-магазины.

Системы управления базами данных СУБД (СУБД) — это программное обеспечение, облегчающее создание, управление, доступ и использование баз данных. Оно выполняет такие базовые функции, как хранение, обработка, защита данных и управление доступом к ним.

MBVT обеспечивает целостность данных, обеспечивает эффективный доступ к ним и создает удобный интерфейс работы с данными для приложений.

Microsoft Access — удобный инструмент для создания и управления базами данных, поддерживающий работу с различными типами данных. Типы данных определяют тип данных, хранящихся в базе данных, и определяют способы работы с ними.

Microsoft Access 2016– система управления базами данных (СУБД), входящая в пакет Microsoft Office, предоставляющая удобные и эффективные инструменты для создания, управления и анализа данных. Предназначена для малого и среднего бизнеса, а также для личного использования.

1. Основные функции и возможности Microsoft Access 2016

Удобный интерфейс: ленточный интерфейс упрощает поиск и использование функций. Панель навигации упрощает управление объектами базы данных (таблицами, запросами, формами, отчётами).

Столы: основной объект для хранения данных; Поддерживает различные типы данных (текст, число, дата, время, валюта и т. д.); Возможность настройки свойств полей (размер, формат, правила проверки).

Запросы (Queries): используется для сортировки, фильтрации и объединения данных. Поддерживает язык запросов SQL (язык структурированных запросов). Различные типы запросов: выборка, обновление, удаление, вставка.

Формы: Позволяет создать удобный интерфейс для ввода и редактирования данных. Поддерживает различные элементы управления (текстовые поля, списки, кнопки и т.д.). Возможность настройки и оформления форм.

Отчеты: Позволяет представить данные в удобном формате для печати или анализа. Поддерживает различные функции группировки, сортировки и суммирования. Возможность настройки и создания отчётов.

Макросы: Используется для создания автоматизированных задач. Возможность последовательного выполнения различных действий. Возможность создания макросов на основе событий (например, нажатия кнопки, открытия формы).

VBA (Visual Basic для приложений): Язык программирования для решения сложных задач. Возможность создания пользовательских функций и расширения функциональности базы данных.

2) Новые функции и улучшения (Access 2016):

Новые шаблоны: готовые шаблоны для различных областей (например, управление проектами, инвентаризация, управление студентами). Шаблоны позволяют быстро и легко создать базу данных.

Веб-базы данных: Возможность создавать и совместно использовать веб-базы данных на платформе SharePoint. Возможность доступа к данным и их редактирования через веб-браузер.

Ссылки на внешние данные: Возможность подключения к различным источникам данных (например, Excel, SQL Server, Oracle). Возможность импорта данных или прямого подключения к ним.

Улучшенный конструктор запросов: Удобный и наглядный интерфейс для создания и редактирования запросов. Возможность применения различных условий и функций.

Новые типы диаграмм: Новые типы диаграмм (например, древовидная диаграмма, диаграмма «солнечные лучи») для визуального представления данных. Возможность настраивать и создавать диаграммы.

3) Возможности:

Управление данными: хранение, редактирование и удаление данных; обеспечение целостности данных; управление доступом к данным.

Анализ данных: Сортируйте, фильтруйте и группируйте данные с помощью запросов; Обобщайте и представляйте данные с помощью отчетов; Визуально анализируйте данные с помощью диаграмм.

Автоматизация: Автоматизируйте задачи с помощью макросов; реализуйте сложные сценарии с помощью VBA.

Интеграция: Интеграция с приложениями Microsoft Office (Excel, Word, Outlook); Работа с платформой SharePoint

Технологии больших данных) понимается как совокупность специального программного обеспечения, оборудования и методов, используемых для сбора, хранения, обработки, анализа и извлечения полезной информации из очень больших, быстро меняющихся и разнообразных форматов данных.

Благодаря этим технологиям компании могут принимать важные решения для улучшения своего бизнеса, удовлетворения потребностей клиентов и получения конкурентного преимущества.

Большие данные (Big Data) — это термин, обозначающий огромные объёмы структурированных и неструктурированных данных, ежедневно генерируемых из различных источников, для извлечения полезной информации из которых требуются специализированные технологии и методы обработки. Эти технологии позволяют организациям анализировать большие объёмы данных быстрее и эффективнее традиционных методов, выявляя скрытые закономерности, тенденции и взаимосвязи, что помогает организациям принимать обоснованные решения, улучшать качество продуктов и услуг, оптимизировать бизнес-процессы и повышать конкурентоспособность.

Основные характеристики больших данных объем, скорость ввода данных, разнообразие форматов данных, достоверность данных и ценность полученных на их основе выводов.

Большие данные – термин, обозначающий огромные объемы структурированных и неструктурированных данных, которые невозможно эффективно обработать традиционными методами анализа.

Большие данные характеризуются тремя ключевыми аспектами, называемыми «три V»: объем (количество данных), скорость (скорость, с которой они поступают и обрабатываются) и разнообразие (разнообразие форматов).

Для хранения, обработки и анализа этих данных требуются специализированные технологии и инструменты, такие как базы данных Hadoop, Spark и NoSQL.

Целью анализа больших данных является выявление скрытых закономерностей, тенденций и корреляций, которые помогают организациям принимать обоснованные решения, улучшать обслуживание клиентов, оптимизировать бизнес-процессы и повышать конкурентоспособность.

К основным характеристикам больших данных относятся три важных аспекта, называемых «три V»:

1) Объем: Большие объемы данных, превосходящие возможности традиционных систем управления базами данных.

2) Скорость: Высокая скорость сбора, обработки и анализа данных в режиме реального времени.

3) Разнообразие: различные типы данных, включая структурированные, полуструктурированные и неструктурированные формы (тексты, изображения, аудио, видео).

Кроме того, часто добавляются дополнительные функции:

4) Достоверность: надежность и точность информации, влияющие на качество принимаемых решений.

5) Ценность: способность извлекать важную информацию и полезные выводы из наборов данных.

Эти характеристики определяют сложность и необходимость специализированных подходов и технологий для эффективного управления большими данными.

Какие технологии используются для обработки больших данных?

Для обработки больших данных используется ряд технологий и инструментов, каждый из которых решает определенные задачи. Ниже приведены некоторые из наиболее популярных:

Платформы распределенной обработки данных:

Hadoop: открытая экосистема, включающая инструменты для распределенного хранения и обработки больших объемов данных. Включает

HDFS (распределённую файловую систему), MapReduce (алгоритм параллельной обработки данных) и YARN (менеджер ресурсов).

Spark: распределенная вычислительная среда, которая обеспечивает быструю обработку данных в памяти, поддерживает потоковую обработку, SQL-запросы, машинное обучение и графовые вычисления.

Базы данных и репозитории: Базы данных NoSQL: предназначены для хранения больших объемов разнообразных данных, такие как MongoDB (документно-ориентированная база данных), Cassandra (столбчатая база данных), Neo4j (графовая база данных).

Столбчатые базы данных: например, Vertica, ClickHouse, предназначены для быстрого анализа больших объемов данных путем оптимизации запросов по столбцам таблиц.

Обработка потоков данных:

Kafka: система потокового ввода-вывода, позволяющая передавать большие потоки данных в режиме реального времени.

Flink, Storm: Инструменты потоковой обработки данных, позволяющие проводить практически мгновенный анализ входящих событий.

Аналитические инструменты и библиотеки:

Pandas: библиотека Python для анализа и обработки данных, особенно полезная для первоначальной обработки перед применением сложных алгоритмов.

Dask: расширяемая структура для параллельных вычислений, совместимая с Pandas и NumPy.

Apache Arrow: стандартизированный формат данных для эффективной передачи и обработки больших объемов данных в памяти.

Машинное обучение и искусственный интеллект:

TensorFlow, PyTorch: фреймворки глубокого обучения, используемые для создания моделей машинного обучения на больших объемах данных.

Scikit-Learn: популярная библиотека Python для классического машинного обучения, которая поддерживает широкий спектр методов классификации, регрессии и кластеризации.

Облачные сервисы:

Amazon Web Services (AWS): предлагает масштабируемые облачные решения для обработки больших данных, такие как Amazon Redshift, Amazon Kinesis, AWS Lambda.

Google Cloud Platform (GCP): предоставляет Google BigQuery, Google Dataproc, Google Pub/Sub для быстрой аналитики и обработки данных.

Microsoft Azure: поддерживает Azure HDInsight, Azure Databricks, Azure Synapse Analytics для аналитики больших данных и разработки приложений.

Каждая технология имеет свою область применения и подходит для разных задач, таких как хранение, обработка, анализ или визуализация больших данных.

Литература

1. Аминов С.М., Мухамадиев С.И., Расулов С.Ш. Методическое пособие для проведения практических и лабораторных занятий по предмету информационно-коммуникационные технологии. –Т.: ТошДАУ, 2020. – 248 с.

2. Урдушев Х., Мавлянов М., Эшанкулов С. Информационно-коммуникационные технологии в сфере. Часть I. Учебное пособие. – Самарканд: Издательско-полиграфический центр Самаркандского государственного университета ветеринарной медицины, животноводства и биотехнологии, 2024. 188 с.

3. Урдушев Х., Мавлянов М., Эшанкулов С. Информационно-коммуникационные технологии в сфере. Часть II. Учебное пособие. – Самарканд: Издательско-полиграфический центр Самаркандского государственного университета ветеринарной медицины, животноводства и биотехнологии, 2025. 200 с.

4. Д. Уотсон и Х. Уильямс «Информатика». Hodder Education, 2-е издание, 2023 г. — 404 стр.

5. Г. Браун и Д. Уотсон. Cambridge IGCSE ICT. Hodder Education, 3-е издание, 2023 г. — 571 стр.