

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ

УНИВЕРСИТЕТ им. Р.Е.АЛЕКСЕЕВА



Институт радиоэлектроники и информационных технологий

## Конспект

по теме

«Технология доступа к серверу БД. BIG DATA»

по дисциплине

«Базы данных»

РУКОВОДИТЕЛЬ:

\_\_\_\_\_

(подпись)

Мисевич П. В.

(фамилия, и.,о.)

СТУДЕНТ:

\_\_\_\_\_

(подпись)

Сухоруков В.А.

(фамилия, и.,о.)

19-B-2

(шифр группы)

Работа защищена «\_\_» \_\_\_\_\_

С оценкой \_\_\_\_\_

Нижний Новгород 2021

## Технология доступа к серверу БД.

Эволюция структур данных породила системы упорядочивания, структуризации и управления данными – **серверы баз данных**.

Сетевое многопользовательское приложение строится по принципу **файл-серверной архитектуры**. Данные в виде одного или нескольких файлов размещаются на файловом сервере. Файловый сервер принимает запросы, поступающие по сети от компьютеров-клиентов, и передает им требуемые данные.

В архитектуре "клиент-сервер" сервер базы данных не только обеспечивает доступ к общим данным, но и берет на себя всю обработку этих данных. Клиент посылает на сервер запросы на чтение или изменение данных, которые формулируются на языке SQL. Сервер сам выполняет все необходимые изменения или выборки, контролируя при этом целостность и согласованность данных, и результаты в виде набора записей или кода возврата посылает на компьютер клиента.

### Недостатки архитектуры с файловым сервером:

1. Данные хранятся в одном месте, а обрабатываются в другом. Это означает, что их нужно передавать по сети, что приводит к очень высоким нагрузкам на сеть и, вследствие этого, резкому снижению производительности приложения при увеличении числа одновременно работающих клиентов.
2. Децентрализованное решение проблем целостности и согласованности данных и одновременного доступа к данным. Такое решение снижает надежность приложения.

**ADO** – технология стандартного обращения к реляционным данным от Microsoft. ADO представляет собой высокоуровневый программный интерфейс для доступа к OLE DB-интерфейсам. Он позволяет манипулировать данными с помощью любых OLE DB-провайдеров, как входящих в состав MDAC некоторых других продуктов Microsoft, так и произведенных сторонними производителями. ADO содержит набор объектов, используемых для соединения с источником данных, для чтения, добавления, удаления и модификации данных.

Microsoft разработала ADO.NET на базе уже зарекомендовавшей себя объектной технологии ADO. Но ADO.NET ориентируется на три важные возможности, которые не поддерживаются ADO:

1. Поддержка модели доступа к несвязанным данным, что является ключевым элементом для работы в Web;
2. Поддержка тесной интеграции с XML;
3. Интеграция с .NET Framework (например, совместимость с базовой библиотекой классов типичной системы).

Поставщик данных .NET состоит из следующих основных компонентов:

1. **Connection** – объект для связи с источником данных;
2. **Command** – объект выполняет команды над источником данных;
3. **DataReader** – читает данные из источника данных в однонаправленном режиме «только чтение».
4. **DataAdapter** – читает данные из источника данных и использует их для заполнения объекта DataSet.
5. **DataSet** – поддерживает копии записей из базы данных без соединения.

## BIG DATA и особенности технологии при проектировании АС.

**BIG DATA** – серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объемов и значительного многообразия для получения воспринимаемых человеком результатов.

В качестве основных характеристик для BIG DATA отмечают:

1. **Объем** - величина физического объема
2. **Скорость** - высокоскоростная обработка и получение результатов
3. **Многообразие** - возможность одновременной обработки различных типов, структурированных и полу структурированных данных.

Процесс обработки BIG DATA включает этапы очистки данных путем последовательного моделирования: выдвигается гипотеза, строится статистическая, визуальная или семантическая модель, проверяется верность выдвинутой гипотезы, выдвигается следующая гипотеза.

Существует множество инструментов и технологий анализа больших данных, в основе которых лежит инструментарий, заимствованный из статистики и информатики: машинное обучение, распознавание образов, имитационное моделирование, включающее интерпретацию данных и составление интерактивных запросов.

Чем более объемный и дифференцируемый массив подвергается анализу, тем более точные и релевантные данные удастся получить на выходе.

### Жизненный цикл проекта BIG DATA:

1. **Обследование** - определение наличия, полноты и точности необходимых данных
2. **Подготовка данных** - определение и обеспечение необходимого уровня качества данных
3. **Планирование модели данных** - определение типа и разработка плана построения модели данных
4. **Построение модели данных** - применение аналитических методов, метрик и структур данных для построения модели данных
5. **Представление результатов** - определение и презентация ключевых результатов, измерение их ценности для бизнеса, оценка успешности проекта спонсором и заказчиком
6. **Использование результатов** - подготовка финального отчета и технической документации, запуск пилотного проекта по реализации модели в реальной производственной среде заказчика