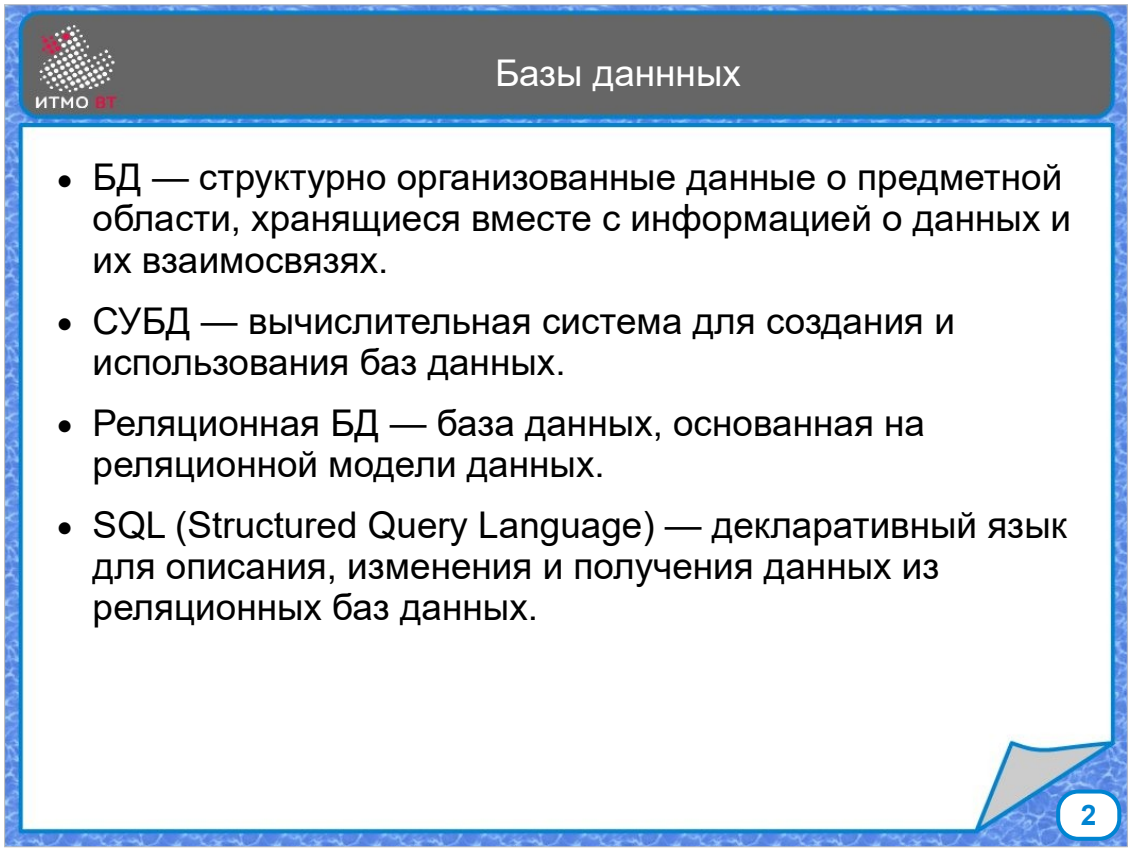
ИТМО ВТ

Взаимодействие с базами данных

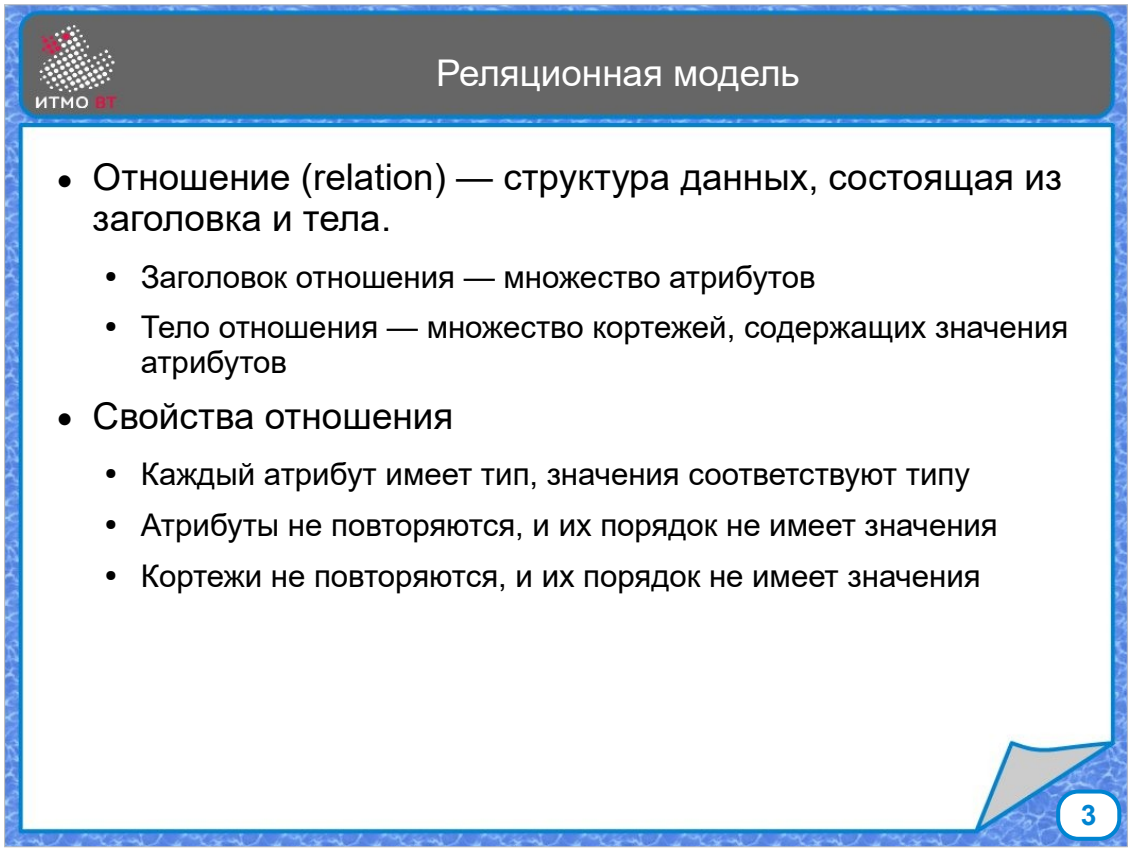


База данных — это совокупность структурно организованных данных, относящихся к некоторой предметной области, при этом вместе с данными хранится информация о характеристиках этих данных и об их взаимных связях.

Система управления базами данных — это приложение, которое предназначено для работы с базами данных.

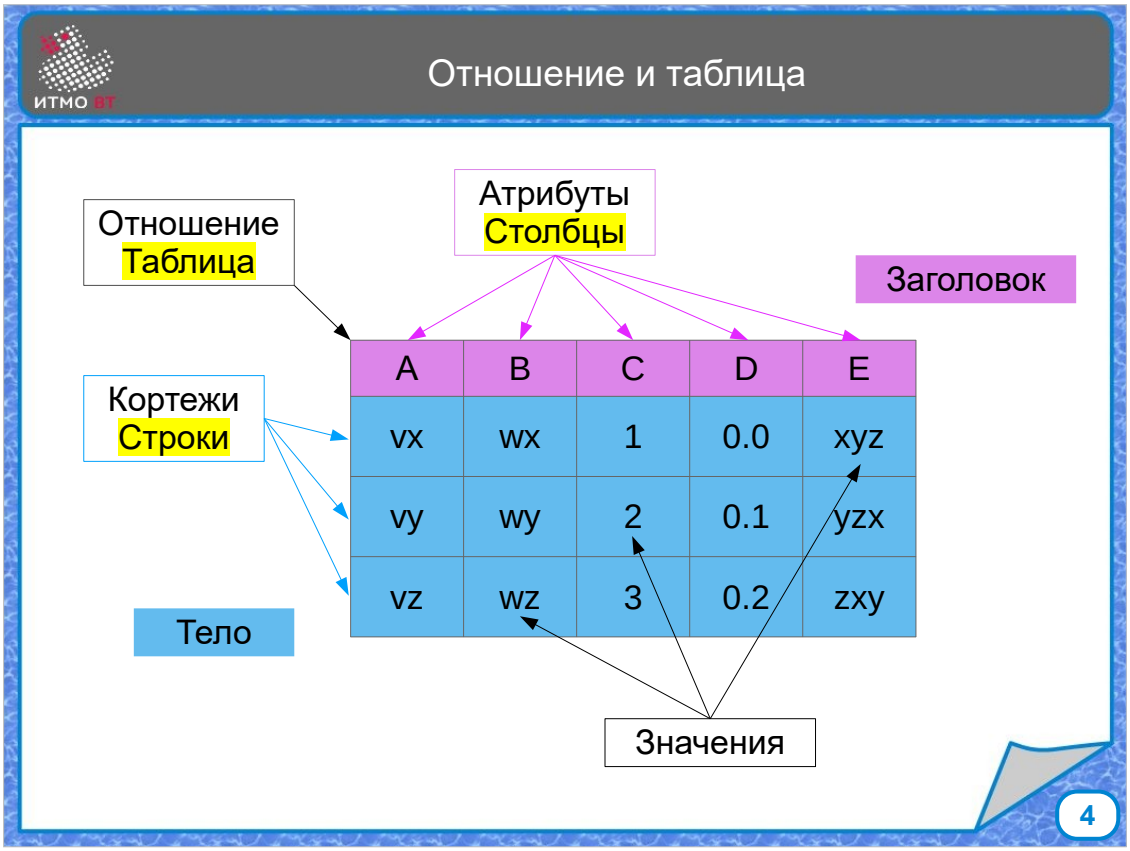
Базы данных бывают разных типов — иерархические, сетевые, объектно-ориентированные, объектно-реляционные. Наиболее распространенными на данный момент являются реляционные базы данных, основанные на реляционной модели.

Для работы с реляционными данными создан язык SQL (Structured Query Language). Это декларативный язык, он не говорит, как можно получить результат. Он указывает, какой нужен результат, а дальше СУБД сама решает, как его получить оптимальным способом.



Реляционная модель основана на понятии отношения. Отношение состоит из заголовка и тела. Заголовок представляет собой множество атрибутов, каждый из которых принадлежит к некоторому типу. Тело состоит из множества кортежей (tuples), представляющих собой значения для каждого атрибута.

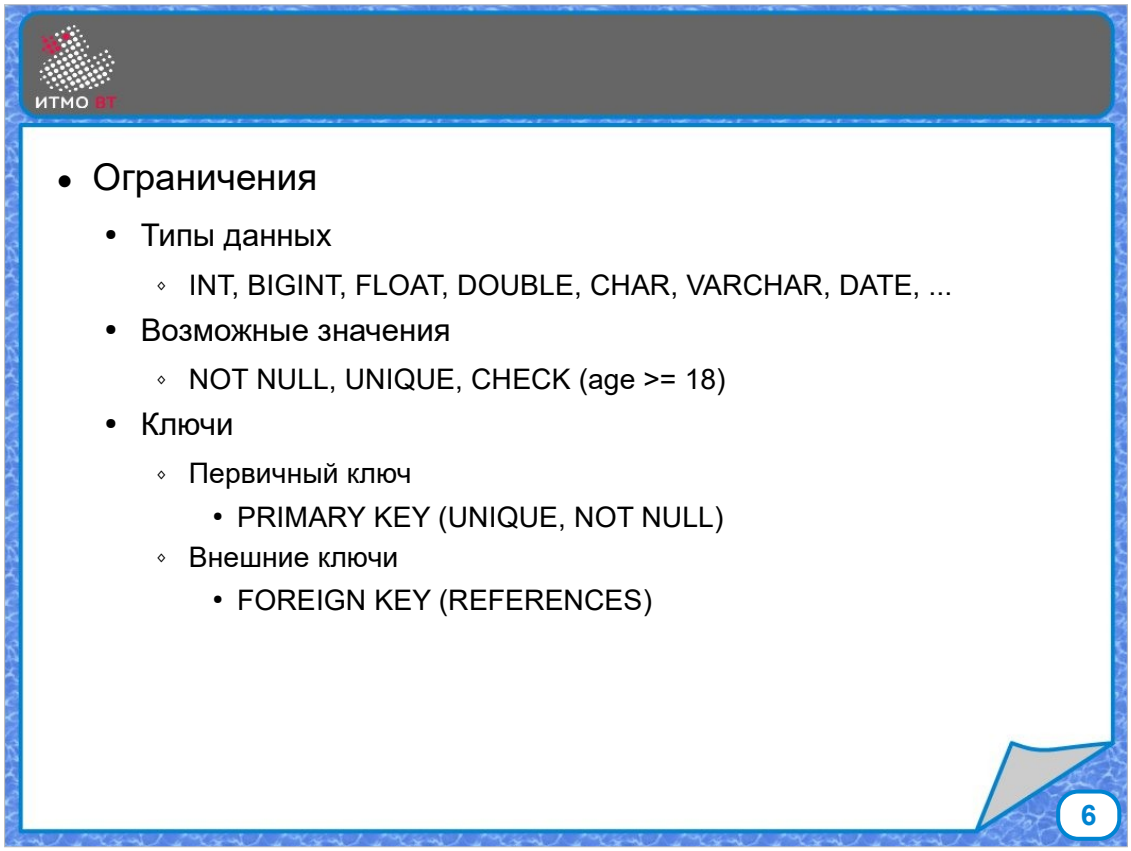
При этом в одном отношении не должно быть повторяющихся атрибутов, и не должно быть повторяющихся кортежей. Порядок следования атрибутов и кортежей не учитывается.



Отношение очень удобно представлять в виде таблицы, где столбцы соответствуют атрибутам, строки — кортежам, значения расположены в ячейках таблицы. При этом, хоть в таблице строки и столбцы расположены в некотором порядке, этот порядок не должен иметь значения.

В данном примере базы данных созданы 4 отношения, или 4 таблицы: Студенты, Группы, Курсы и Оценки. Каждая таблица имеет некоторое количество столбцов, и содержит данные в некотором количестве строк.

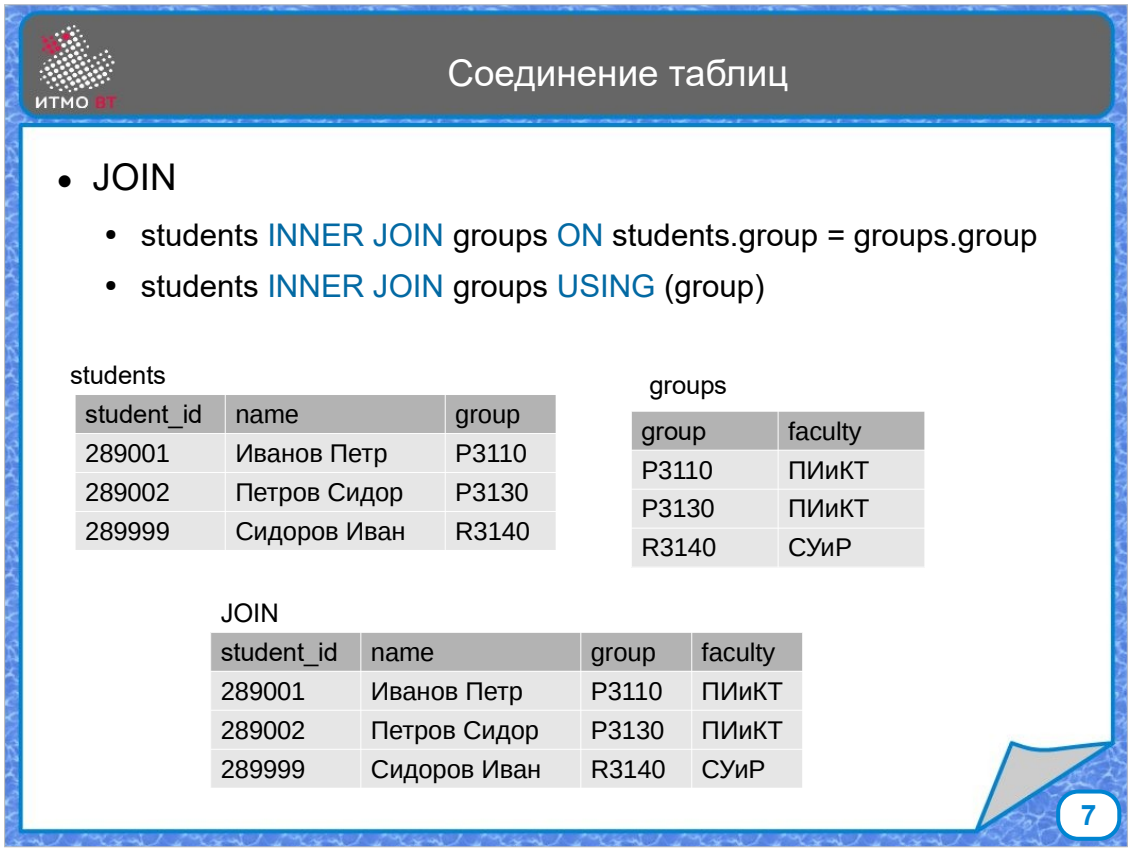
Хорошая модель данных не должна содержать повторяющиеся данные, каждый факт должен храниться в одном месте. Для этого выполняется процедура нормализации, в ходе которой таблицы могут быть декомпозированы на несколько таблиц.



На значения в таблицах могут накладываться ограничения. Это могут быть ограничения по типу данных — каждый столбец имеет свой тип, и значения в этом столбце должны соответствовать типу. Может ограничиваться диапазон допустимых значений. Также можно разрешить только уникальные, либо только не пустые значения в столбце.

Есть специальные виды ограничений, Это, например, первичный ключ — один или множество столбцов, значения которого однозначно идентифицируют строку таблицы. Значения первичного ключа должны быть уникальны и не могут иметь значение NULL.

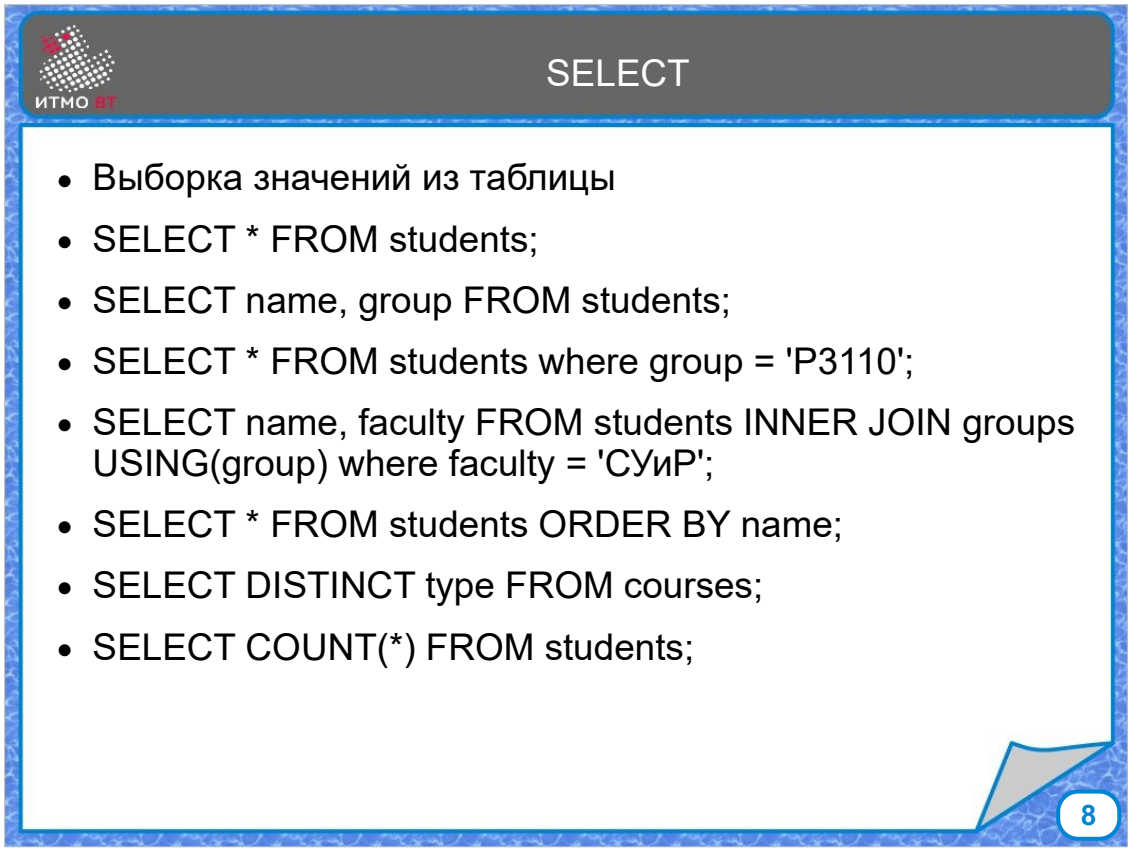
Внешний ключ задается, если значения столбца в таблице могут принимать только значения, которые содержатся в некотором столбце другой таблицы. Например, номер группы в таблице Students может принимать только значения, которые есть в соответствующем столбце таблицы Groups.



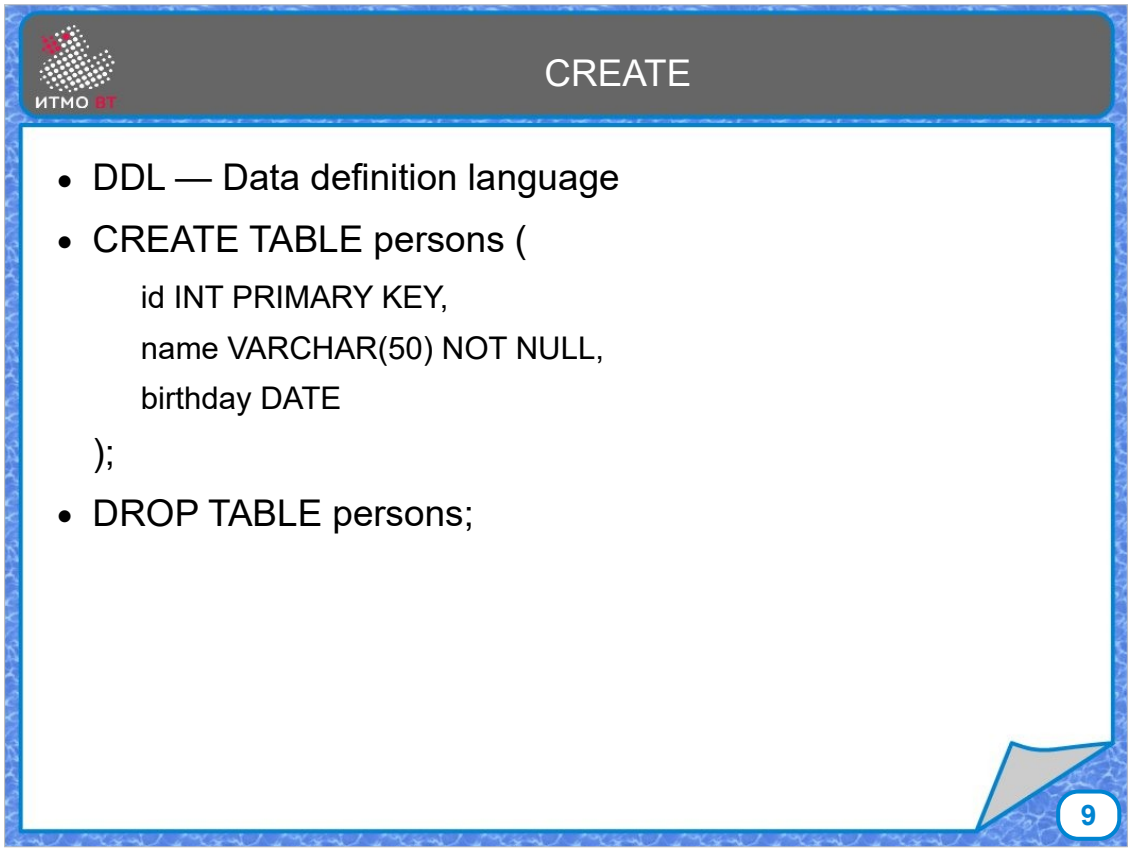
Таблицы можно соединять с помощью оператора JOIN.

Есть несколько типов соединений, самым распространенным их них является внутреннее соединение — INNER JOIN. При выполнении соединения двух таблиц в результирующий таблице набор столбцов представляет из себя объединение столбцов первой и второй таблиц. Значения формируются следующим образом: каждой строке первой таблицы ставится в соответствие каждая строка второй таблицы, и проверяется условие соединения (ON). Если оно истинно для данной строки, то данная строка попадает в итоговую таблицу.

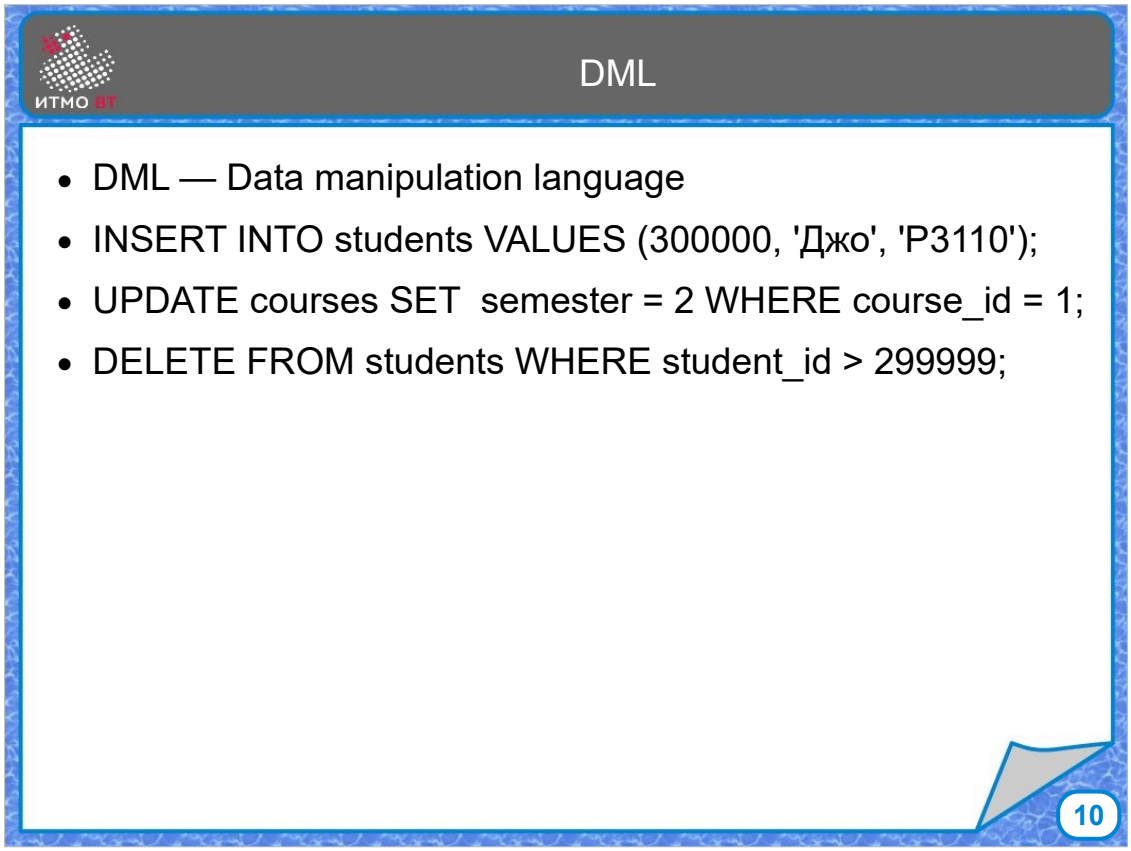
Если соединение производится по условию равенства значений столбцов (это наиболее частый вариант), и сравниваемые столбцы имеют одинаковое название, то можно использовать сокращенный синтаксис с использованием USING с именем столбца вместо ON с условием.



Для получения данных из таблиц используется SQL-команда выборки SELECT. С ее помощью можно выбрать все значения из таблицы, ограничить выборку только нужными столбцами, ограничить выборку условием, которому должны удовлетворять выводимые строки. Можно произвести выборку из соединения таблиц, также можно отсортировать выводимые данные по некоторому столбцу, выбрать только неповторяющиеся данные, а также применить агрегирующие функции, например, посчитать количество строк в таблице или сумму значений.



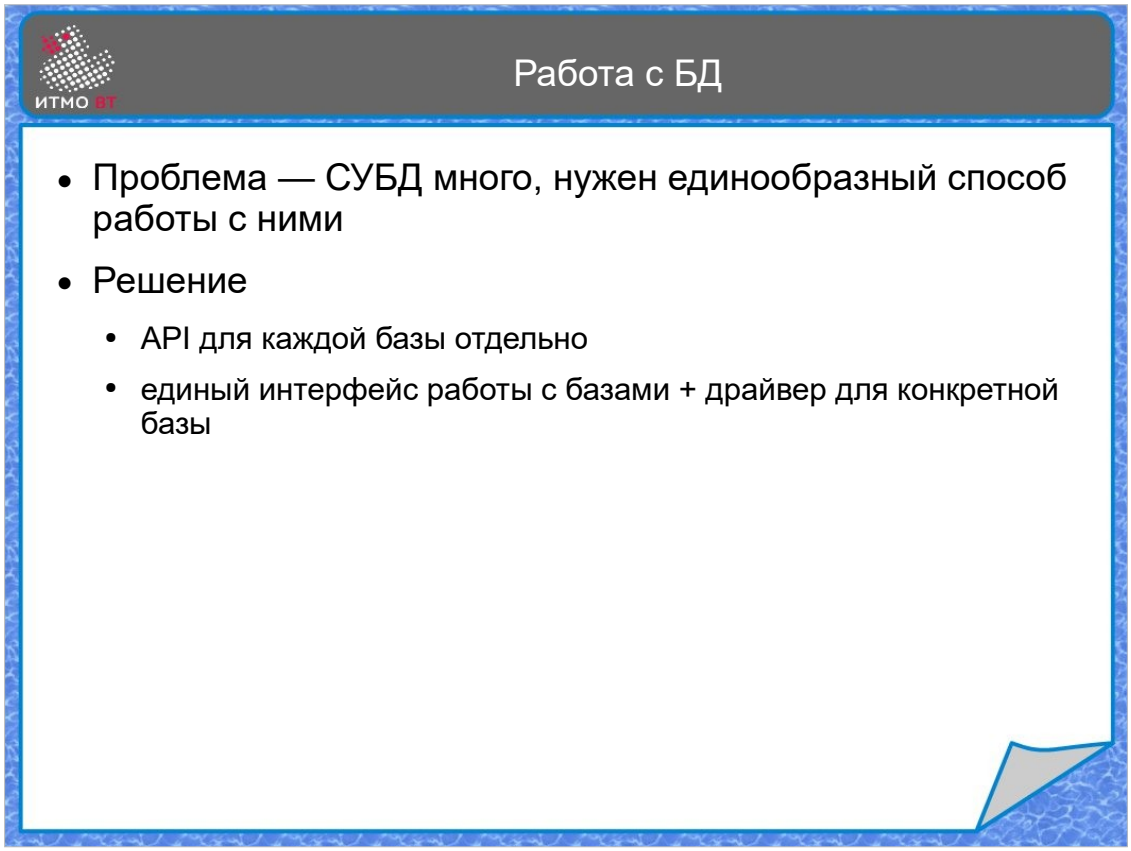
Создавать, изменять и удалять таблицы можно с помощью подмножества языка SQL — языка определения данных (DDL). К таким командам относятся — CREATE, предназначенная для создания таблиц, и DROP, предназначенная для удаления таблиц. При создании таблицы необходимо задать имена и типы данных всех столбцов, также при создании задаются ограничения.



Для работы с данными, когда таблицы уже созданы предназначен DML — язык манипулирования данными. К этим командам относятся:

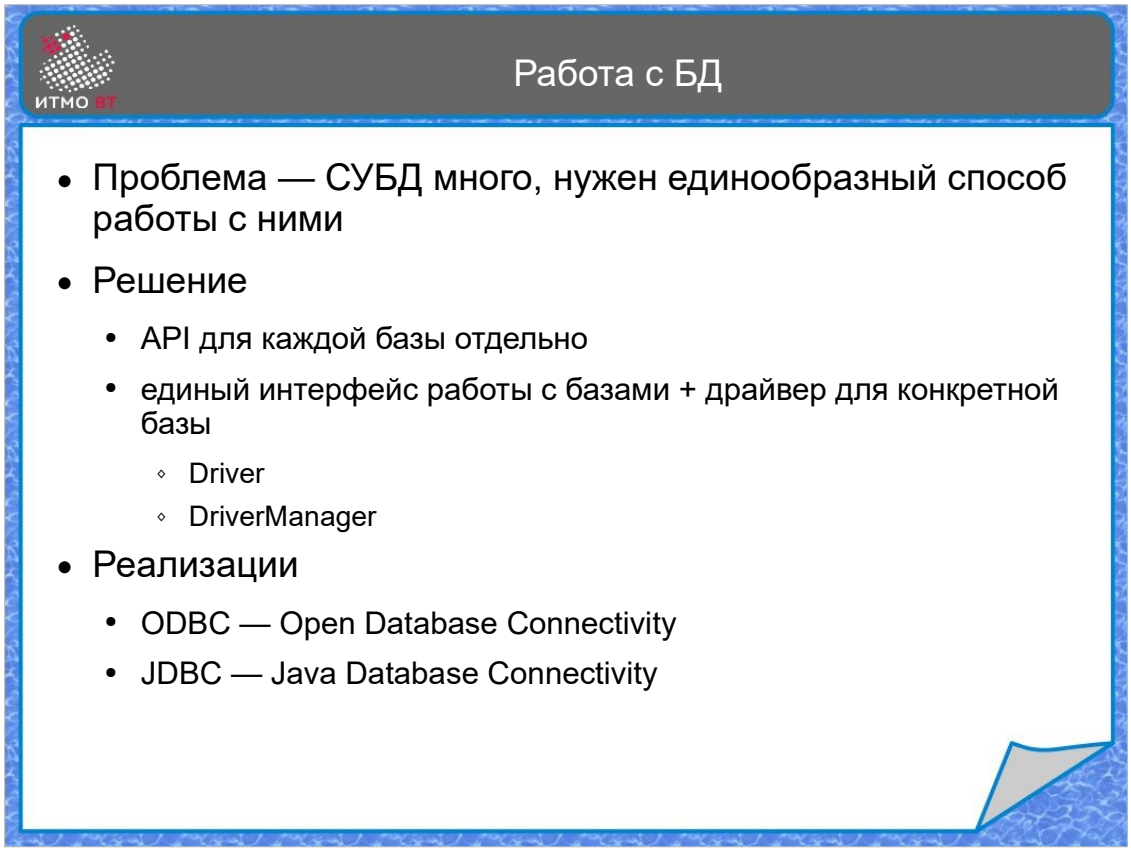
- INSERT — позволяет добавить в таблицы данные
- UPDATE — позволяет изменить значения данных в строках, удовлетворяющих некоторому условию.
- DELETE — позволяет удалить строки по заданному условию.


ИТМО ВТ
Основы JDBC



При взаимодействии с базами данных, желательно обращаться к различным базам унифицированным способом. Есть 2 варианта решения проблемы:

1. Для каждой базы пишется своя собственная библиотека, набор функций или методов, позволяющих любому приложению пользоваться основными функциями базы данных. Основной недостаток такого способа — необходимость переписывать приложения при изменении базы данных на другую (а иногда и при обновлении базы данных до новой версии).
2. Разработать единый интерфейс взаимодействия, а для каждой конкретной базы написать свой драйвер.



В настоящий момент используются 2 основных реализации варианта единого интерфейса работы с базами данных: ODBC и JDBC. При этом приложению предоставляется единый интерфейс взаимодействия, работа осуществляется с помощью менеджера драйверов, который подключает драйвер для определенной СУБД.

ODBC — Open Database Connectivity, стандарт разработан Microsoft в 1992 г., реализует процедурную парадигму, имеются реализации для многих языков программирования, Поддерживает большинство СУБД.

JDBC — появился в 1997 г., реализация для языка Java. Также поддерживает большинство СУБД.

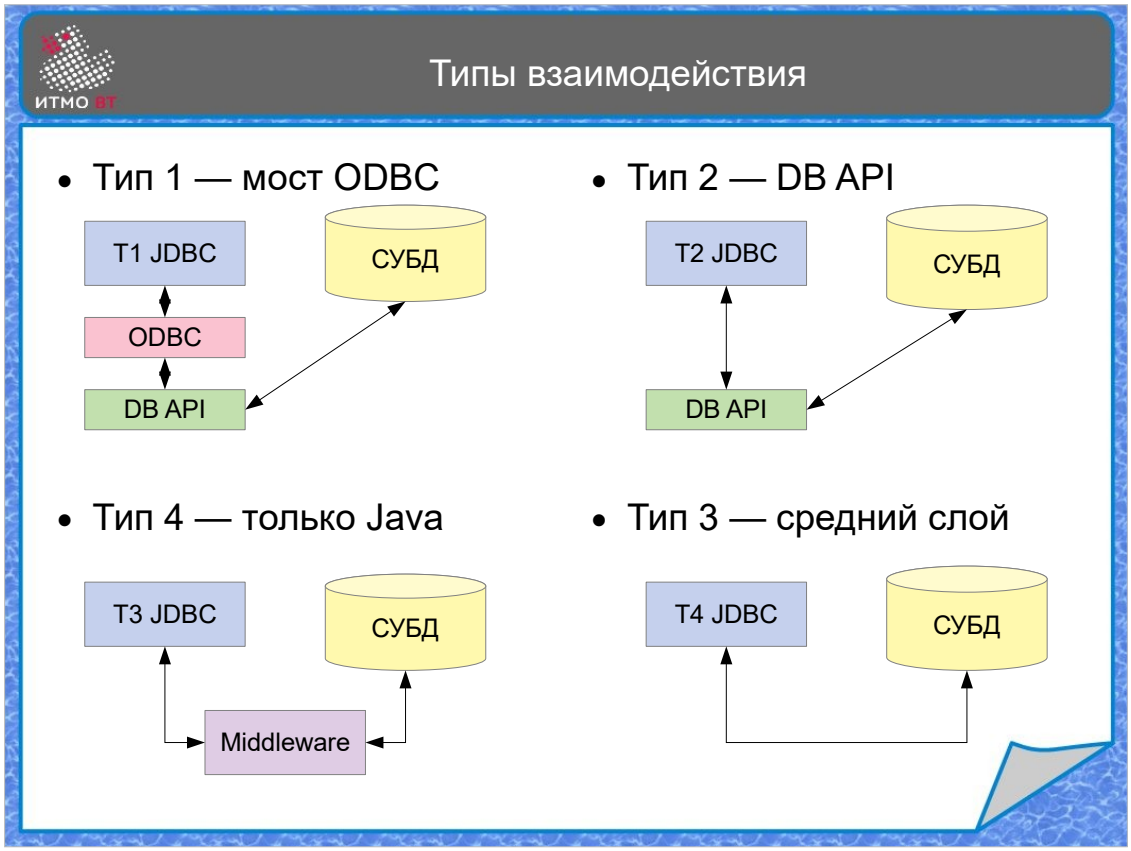
JDBC - Java Database Connectivity, описывает 2 основных интерфейса: JDBC API - это интерфейс для доступа к данным со стороны пользователя базы, его обычно используют разработчики приложений, работающих с базой данных.

JDBC Driver API - это интерфейс драйверов, его используют разработчики самих драйверов, обычно это производители конкретных СУБД.

Все основные классы JDBC API находятся в 2 пакетах:

java.sql - базовая библиотека, которая содержит основной набор интерфейсов и классов, позволяющих выполнять все основные функции взаимодействия с базами данных

javax.sql - расширенная библиотека, содержащая классы и интерфейсы, позволяющие обеспечить дополнительные возможности работы с базами данных.



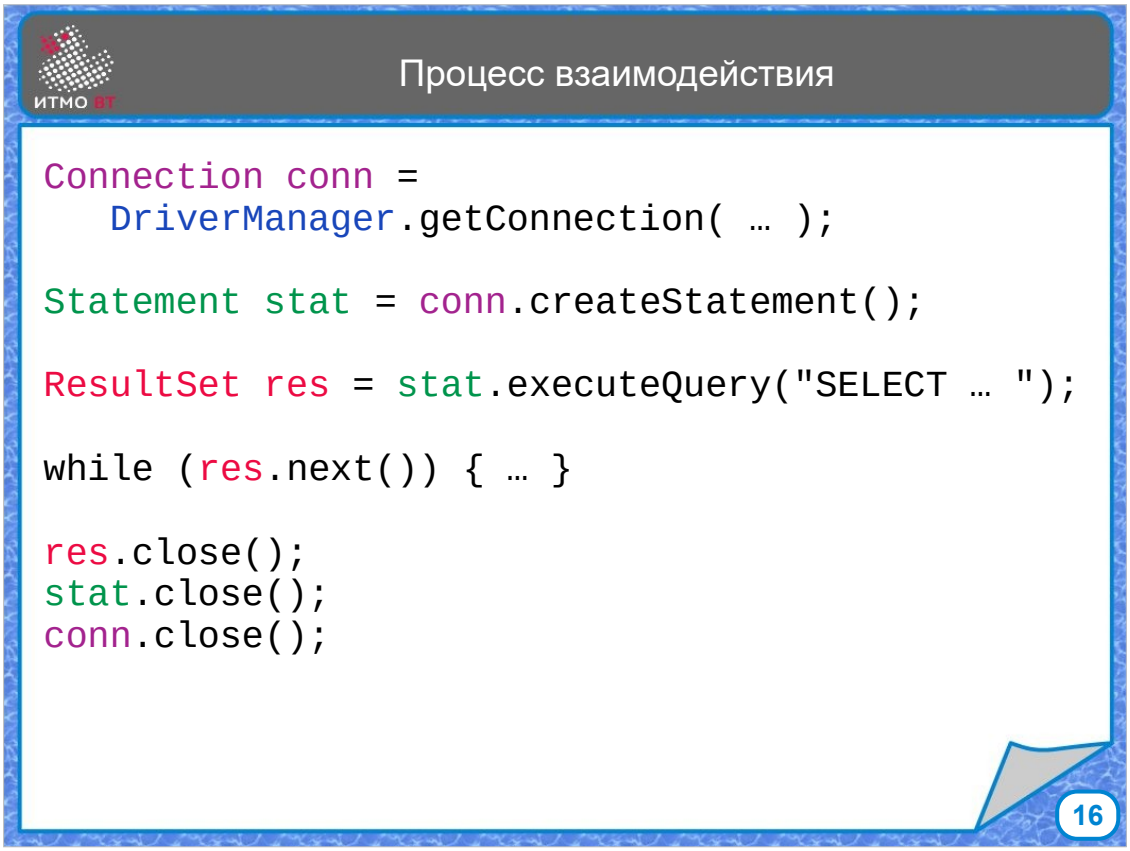
Типы драйверов JDBC:

1 тип — драйвер, работающий через ODBC-драйвер. Он появился самым первым, в то время для большинства СУБД уже существовали ODBC-драйверы, поэтому проще всего было реализовать драйвер, преобразующий вызовы JDBC в вызовы ODBC, а уже драйвер ODBC непосредственно работал с базой данных.

2 тип — драйвер, использующий для работы нативный интерфейс вызовов базы данных — DB API.

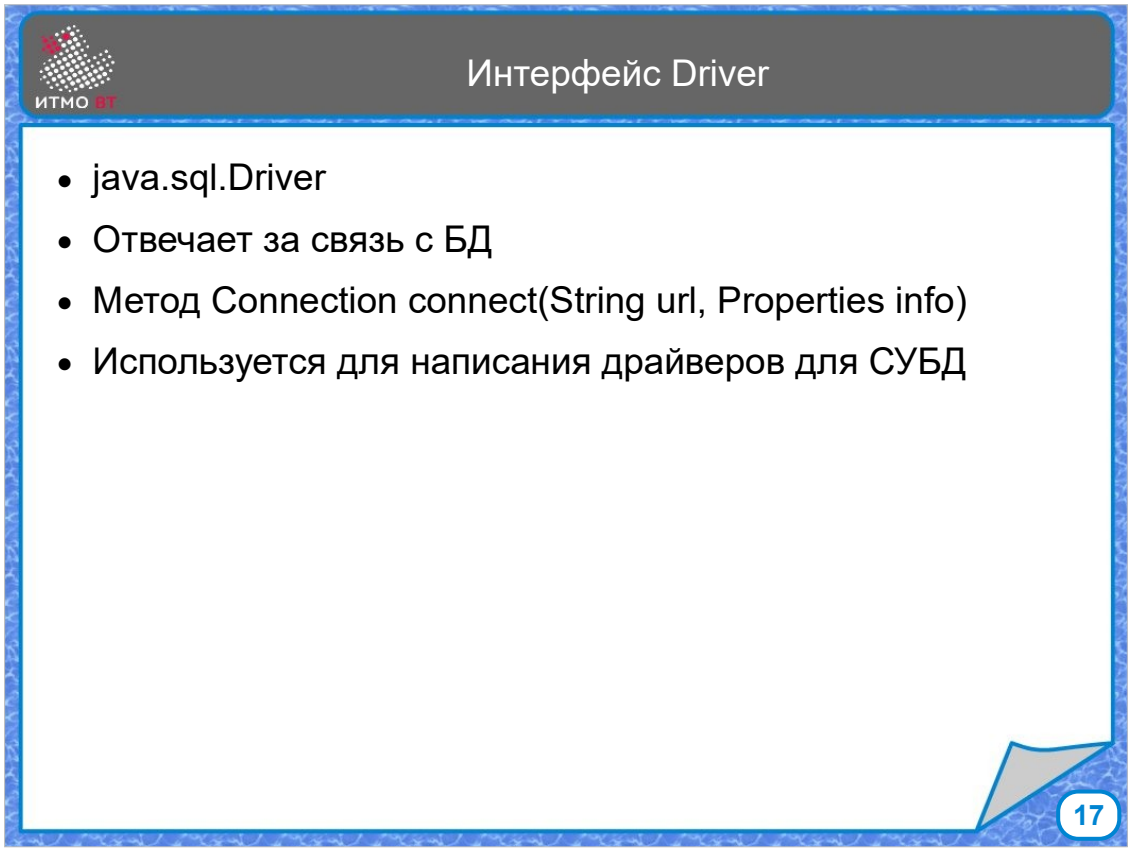
3 тип — драйвер, использующий промежуточный компонент — сервер приложений, взаимодействующий с JDBC по сети, и перенаправляющий запросы в СУБД.

4 тип (сейчас наиболее распространенный) — драйвер для СУБД, написанный на Java, напрямую реализующий спецификацию JDBC.

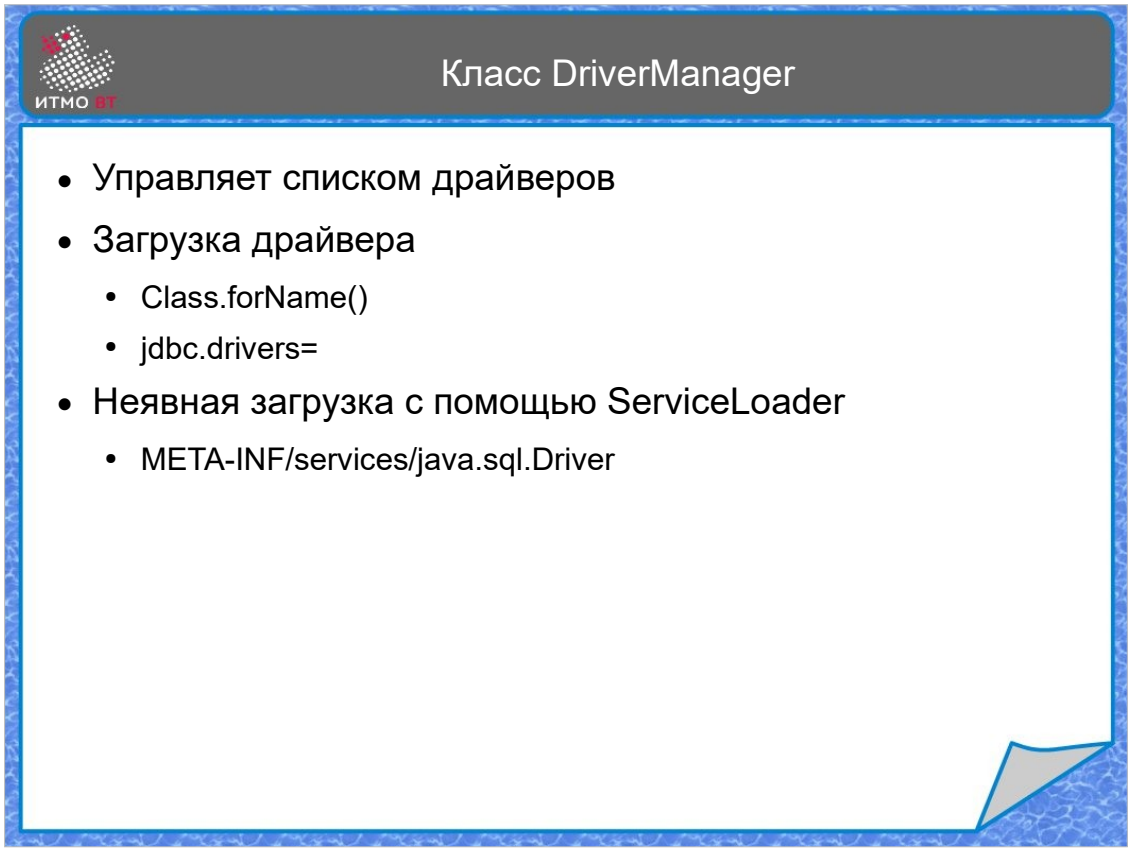


Рассмотрим вкратце пример взаимодействия.

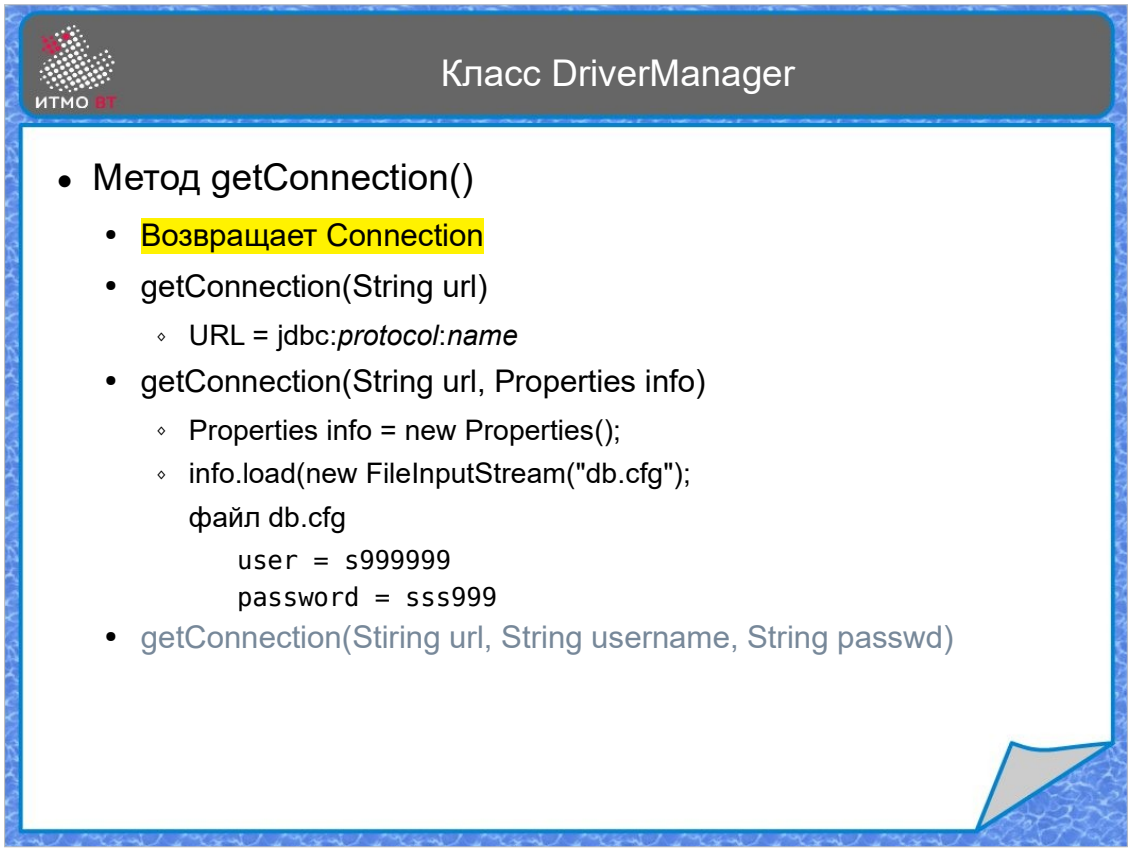
Вызываем метод getConnection у класса DriverManager, которому передаем параметры соединения с базой данных. Этот метод вернет объект типа Connection, представляющий собой абстракцию соединения. У соединения вызывается метод createStatement, который создает запрос - объект типа Statement. Исполнение запроса производится с помощью метода executeQuery. Он возвращает результат - объект типа ResultSet. Результат представляет из себя таблицу и реализует шаблон Iterator, то есть позволяет перебирать строки результата с помощью метода next(). После окончания работы необходимо закрыть все объекты: ResultSet, Statement и Connection. Для этого можно применить блок try с ресурсами.



Интерфейс Driver отвечает за связь с базой данных. При использовании JDBC в прикладных задачах этот интерфейс напрямую не используется. У него есть метод connect(), предназначенный для установки соединения, возвращающий объект типа Connection. Также имеется метод acceptsURL, принимающий строку, и возвращающий true, если данный драйвер способен обработать переданный URL. С его помощью можно найти драйвер, способный работать с конкретной базой данных.

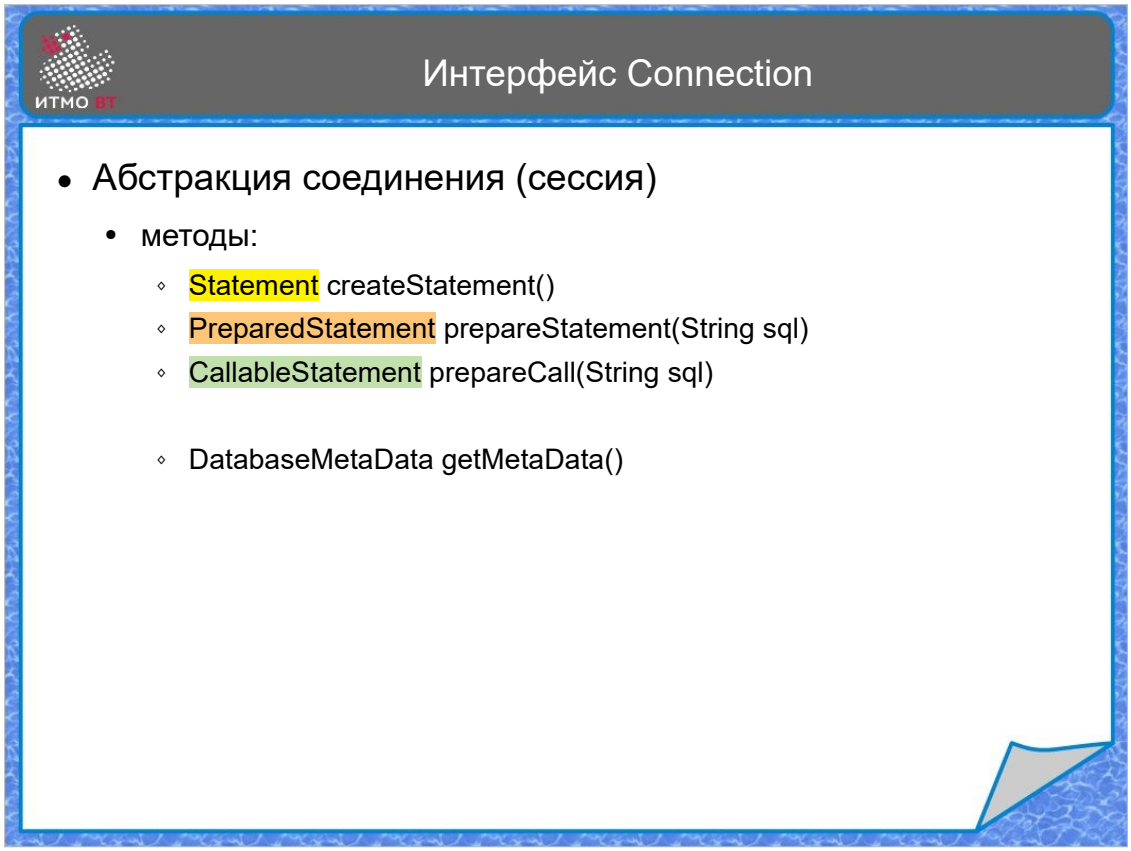


Для организации работы с базой данных используется класс DriverManager, который содержит список драйверов. В прошлых версиях JDBC драйверы надо было загружать явно, сейчас это делается автоматически. Загрузить драйвер явно можно, вызвав метод Class.forName() и передав ему имя класса драйвера. Также можно перечислить нужные классы драйверов в системном свойстве jdbc.drivers. Для автоматической загрузки драйверов применяется механизм сервис-провайдеров, для этого в jar-архиве с драйверами в файле java.sql.Driver указывается список классов, реализующих интерфейс Driver, которые должны быть загружены для работы с базой данных.

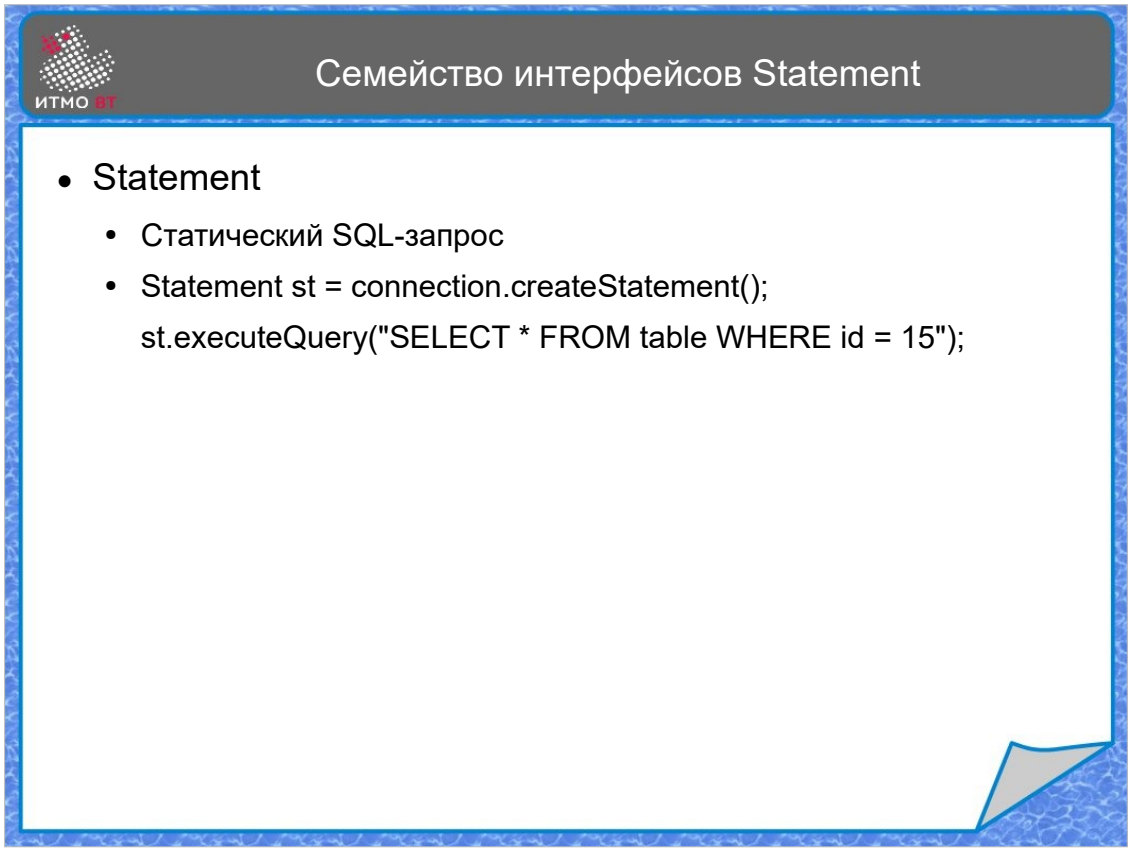


DriverManager содержит метод getConnection(). Ему передаются параметры соединения, один из вариантов - передать URL, который имеет вид jdbc:protocol:name, компоненты разделяются точкой с запятой. Протокол обычно зависит от типа базы данных, а поле name содержит название конкретной базы данных. Есть еще второй формат метода, где указывается объект Properties, в котором задаются имя пользователя и пароль для установления соединения. Значения можно прочитать из файла. Также можно указать имя пользователя и пароль прямо в аргументах метода, но так делать не рекомендуется.

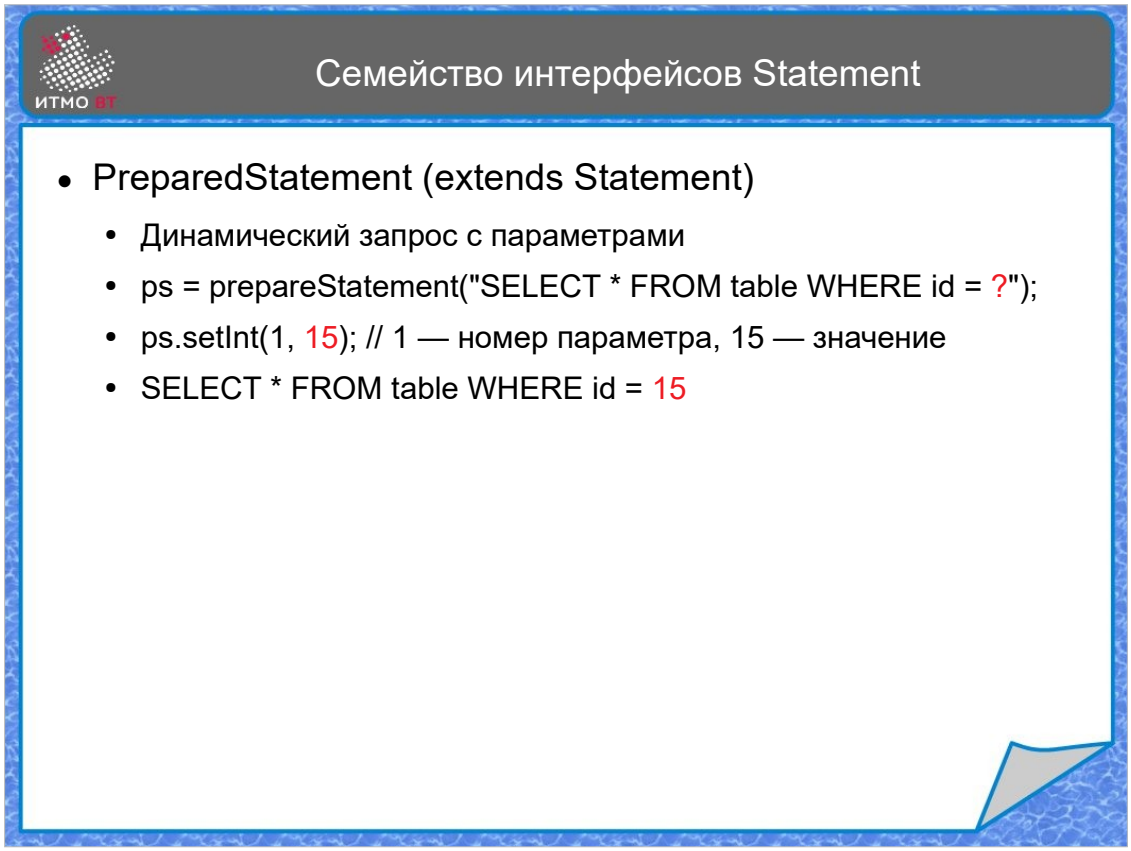
Метод getConnection() возвращает объект типа Connection.



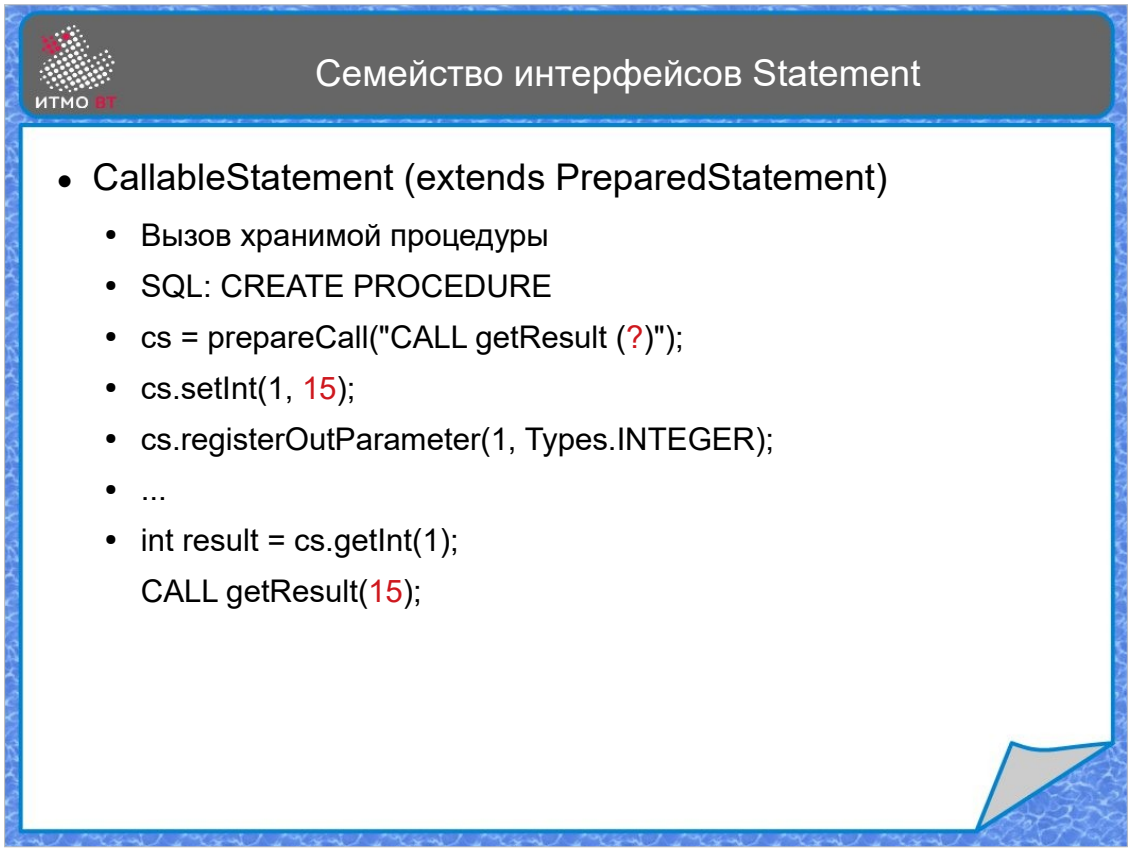
Connection это абстракция соединения с базой данных, или сессии. Интерфейс Connection имеет методы для получения объекта запроса - createStatement, prepareStatement и prepareCall. Кроме этого, Connection можно использовать для получения метаданных о базе данных, с которой установлено соединение.



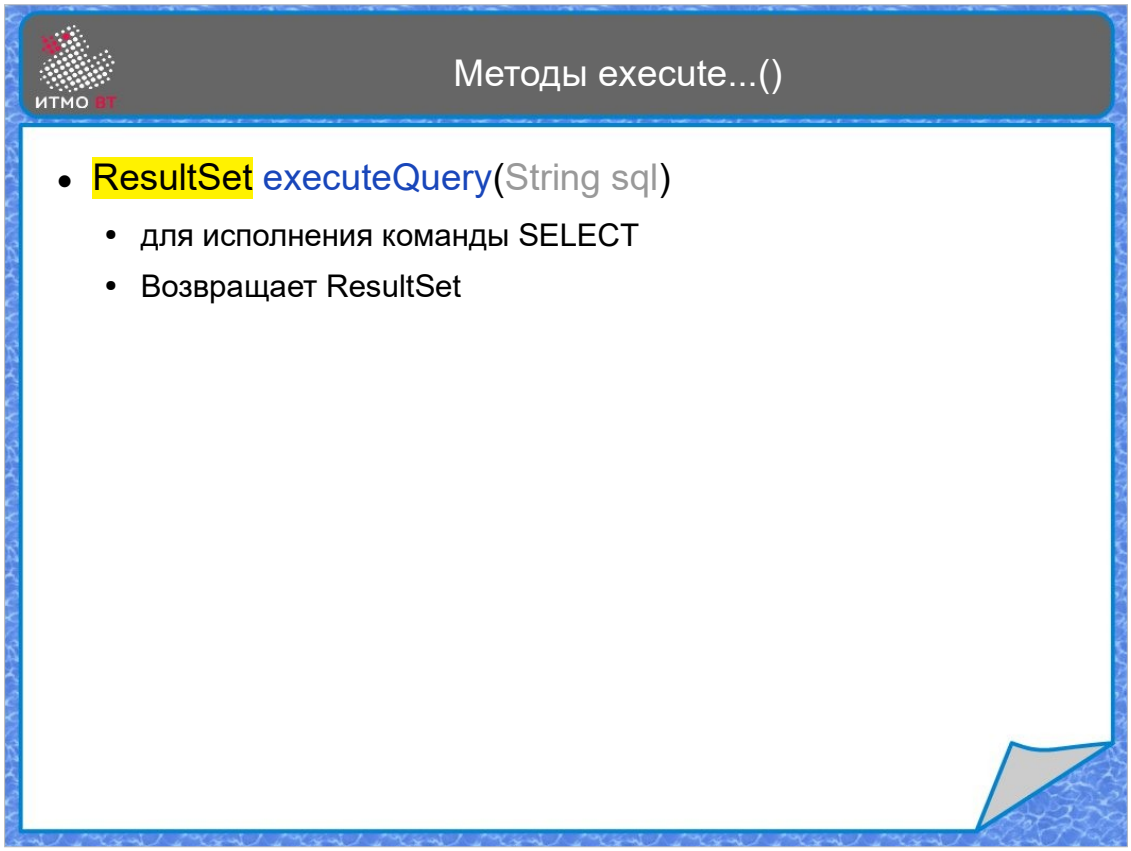
Рассмотрим подробнее интерфейс Statement и его потомков. Объект интерфейса Statement сам запрос не содержит. Запрос передается в качестве параметра при вызове метода executeQuery(). При вызове данного метода нужно быть внимательным при формировании запроса, если он составляется из данных, предоставленных пользователем. Никогда нельзя доверять полученным от пользователя данным, нужно контролировать их тип и содержимое, чтобы не допустить применение SQL-инъекций.



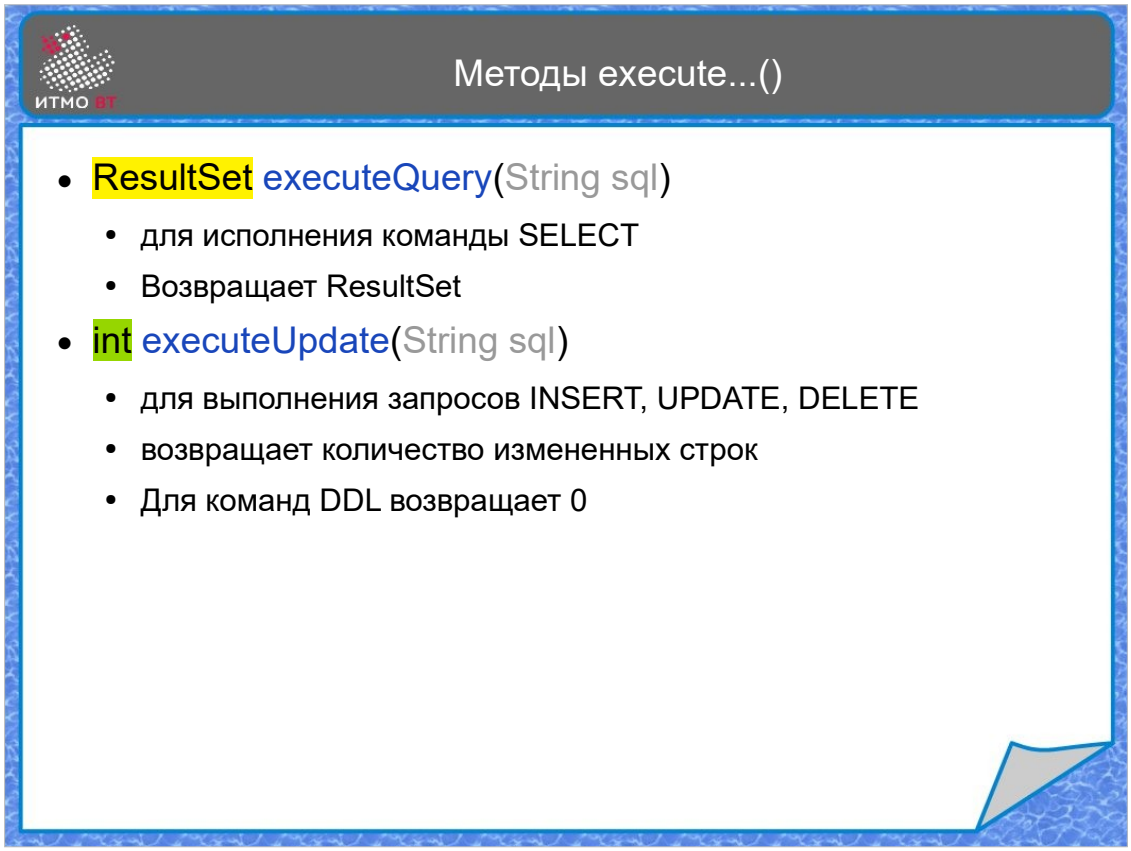
Объект интерфейса PreparedStatement представляет подготовленный запрос с параметрами. Он применяется для выполнения идентичных запросов, отличающиеся одним или несколькими параметрами, например, ID. В таких случаях запрос подготавливается заранее, а на месте параметров указываются знаки вопроса. Затем можно задать значения параметров с помощью одного из методов setInt, setString, или подобных (выбор метода зависит от типа параметра). Запрос выполняется при вызове метода executeQuery(). Далее можно установить новые значения параметров и опять выполнить запрос. Этот способ имеет несколько преимуществ. Во-первых, запросы исполняются быстрее, так как при подготовке запроса СУБД формирует оптимальный план исполнения запроса, остается только подставить параметры и выполнить его. Во-вторых, такие запросы защищены от SQL-инъекций, так как все специальные символы экранируются автоматически.



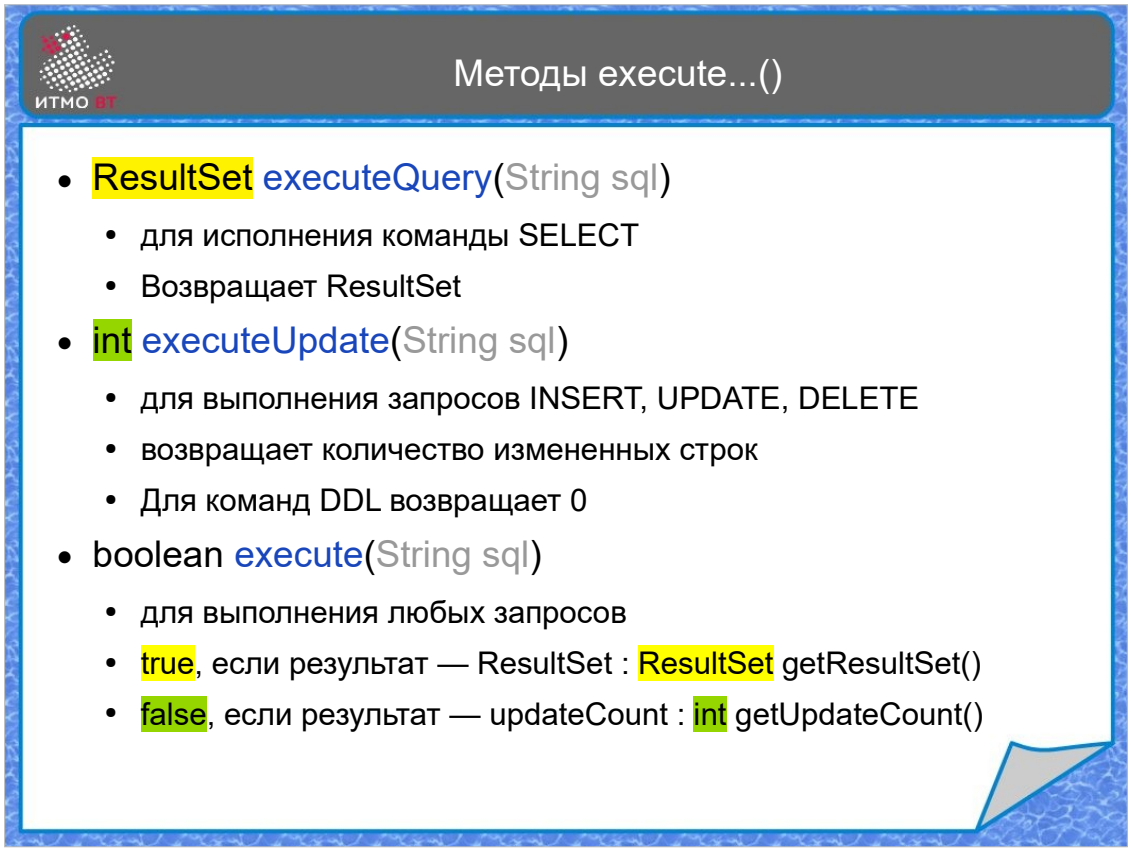
Интерфейс CallableStatement расширяет PreparedStatement, и еще умеет вызывать хранимые процедуры, то есть процедуры, которые хранятся в базе данных. В языке SQL они вызываются с помощью команды CALL. Для этого в объекте типа CallableStatement, задаются входные параметры, и регистрируются выходные параметры. При регистрации выходного параметра его номеру ставится в соответствие тип параметра. После выполнения запроса можно получить значения выходных параметров.



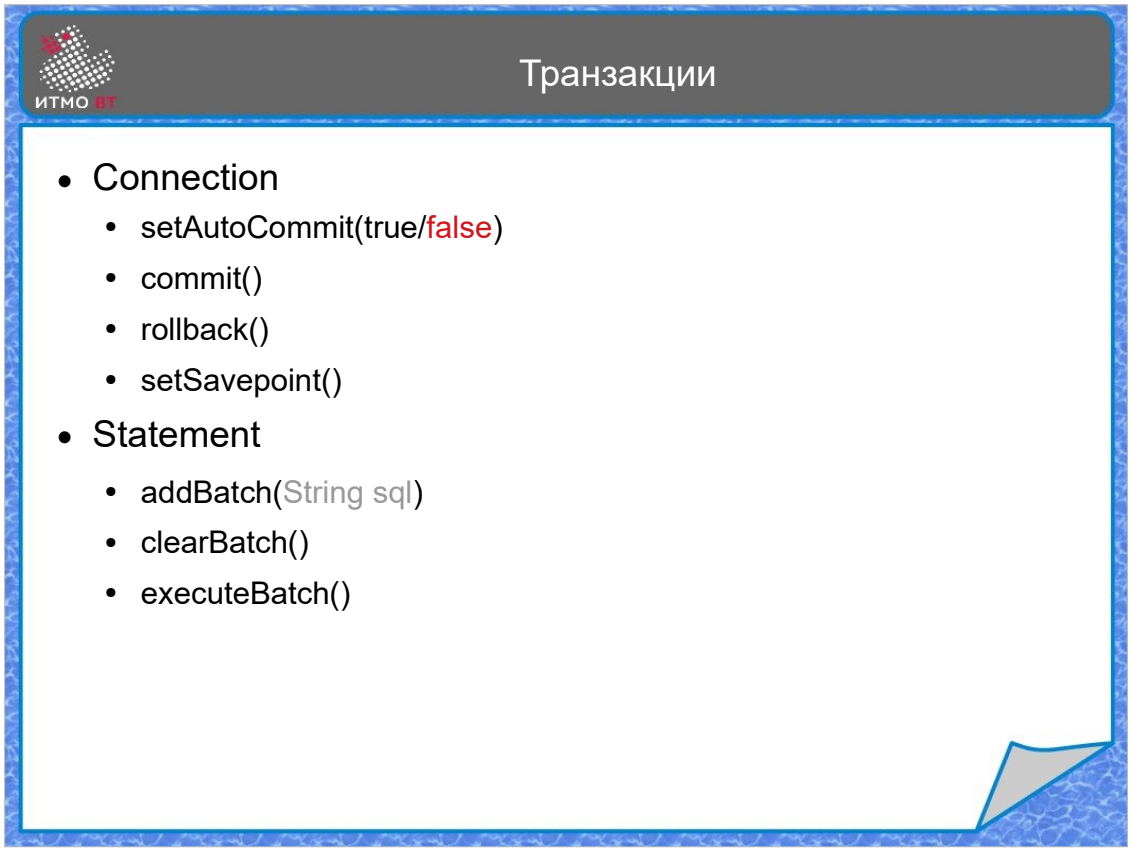
Для объектов Statement запрос указывается в параметрах методов исполнения запроса. Для объектов PreparedStatement и CallableStatement параметр не нужен, так как запрос уже задан. Метод exeeceuteQuery выполняет запрос (обычно это запрос типа SELECT) и возвращает в качестве результата ResultSet, из которого потом можно будет получить данные.



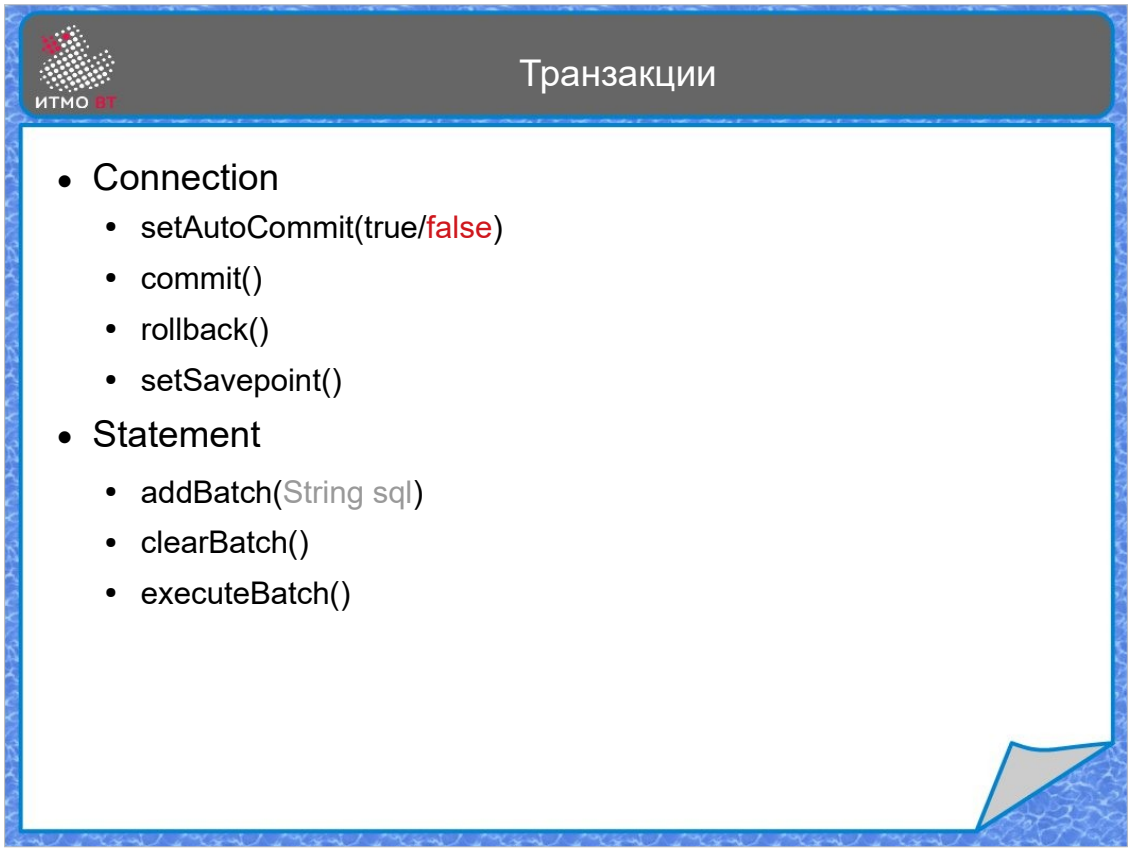
Запросы типа INSERT, UPDATE или DELETE, выполняются методом executeUpdate(), который возвращает количество изменившихся в результате запроса строк. Кроме команд DML, можно также выполнять команды DDL, при этом метод executeUpdate будет возвращать 0, так как команды DDL не работают со строками.



Универсальный метод execute может использоваться для любого типа запросов. Он возвращает результат типа boolean. Значение true обозначает, что после выполнения запроса есть результат в виде таблицы, то есть ResultSet, который можно получитьт с помощью метода getResultSet(). А значение false обозначает, что в результате выполнения запроса имеются модифицированые строки, количество которых можно узнать с помощью метода getUpdateCount().

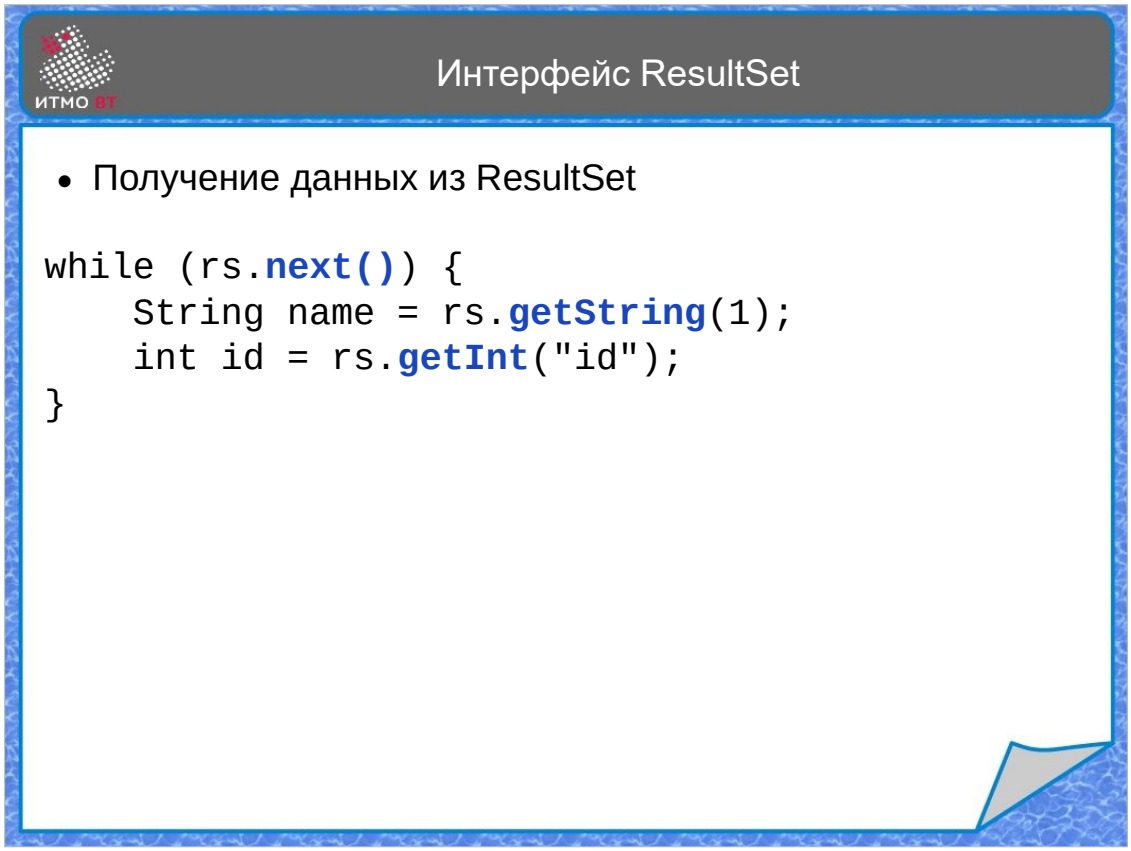


Некоторые запросы должны выполняться вместе, то есть за одну транзакцию. Например, мы управляем средствами на банковских счетах. Допустим, что выполняется перевод с одного счета на другой. При этом важно обеспечить атомарность перевода. Операция снятия денег с одного счета и операция зачисления этих денег на другой счет зависимы, и должны либо вместе выполниться, либо вместе не выполниться. Если транзакция началась, то либо она успешно завершится целиком, либо она будет отменена с возвратом базы данных в состояние до начала транзакции. Транзакции в базе обычно реализуются с помощью механизма фиксаций (commit) и откатов (rollback). При старте транзакция, реальное состояние данных в базе не изменяется до момента выполнения команды COMMIT, фиксирующей состояние базы данных. До выполнения команды COMMIT есть возможность отменить транзакцию командой ROLLBACK.

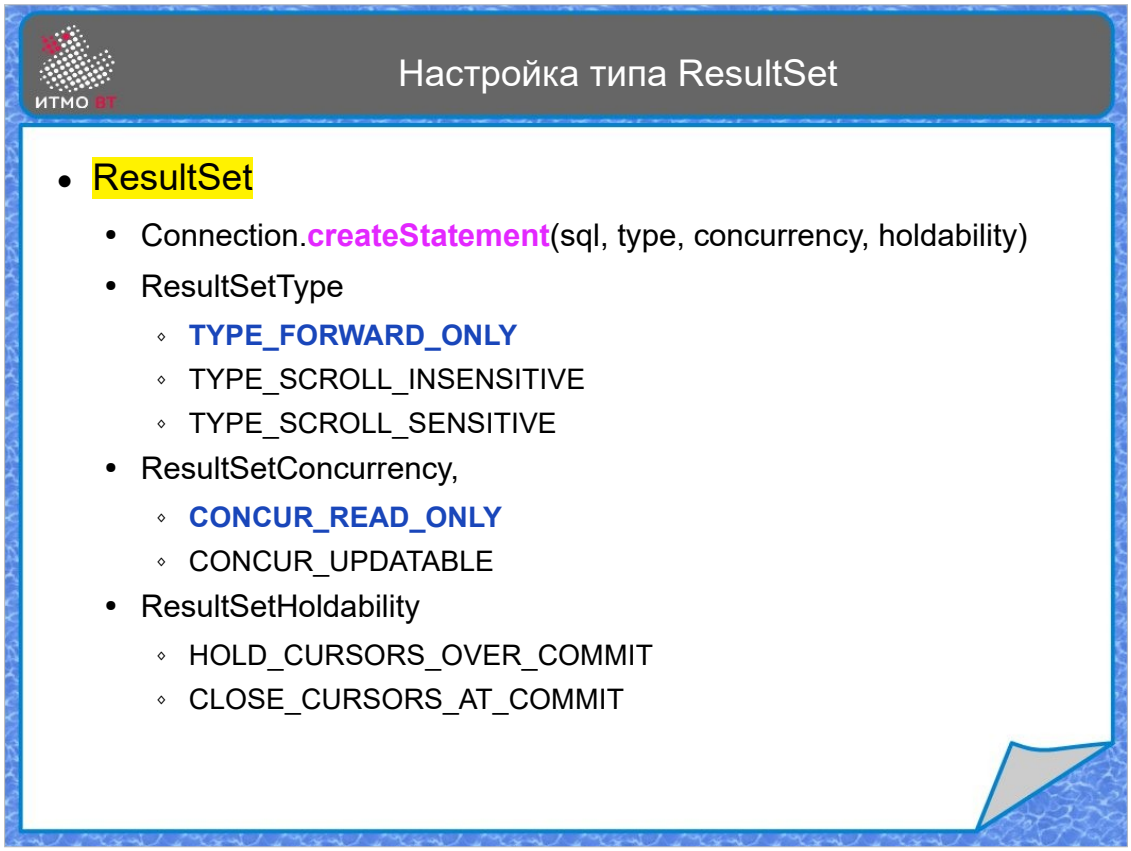


При обычной работе с JDBC состояние базы фиксируется после каждого запроса (Autocommit). Для перевода в режим транзакций, нужно вызвать метод setAutoCommit с параметром false. Метод commit() фиксирует состояние базы данных, метод rollback() позволяет откатить все изменения, сделанные после последней фиксации. Метод setSavePoint() позволяет создать точку сохранения.

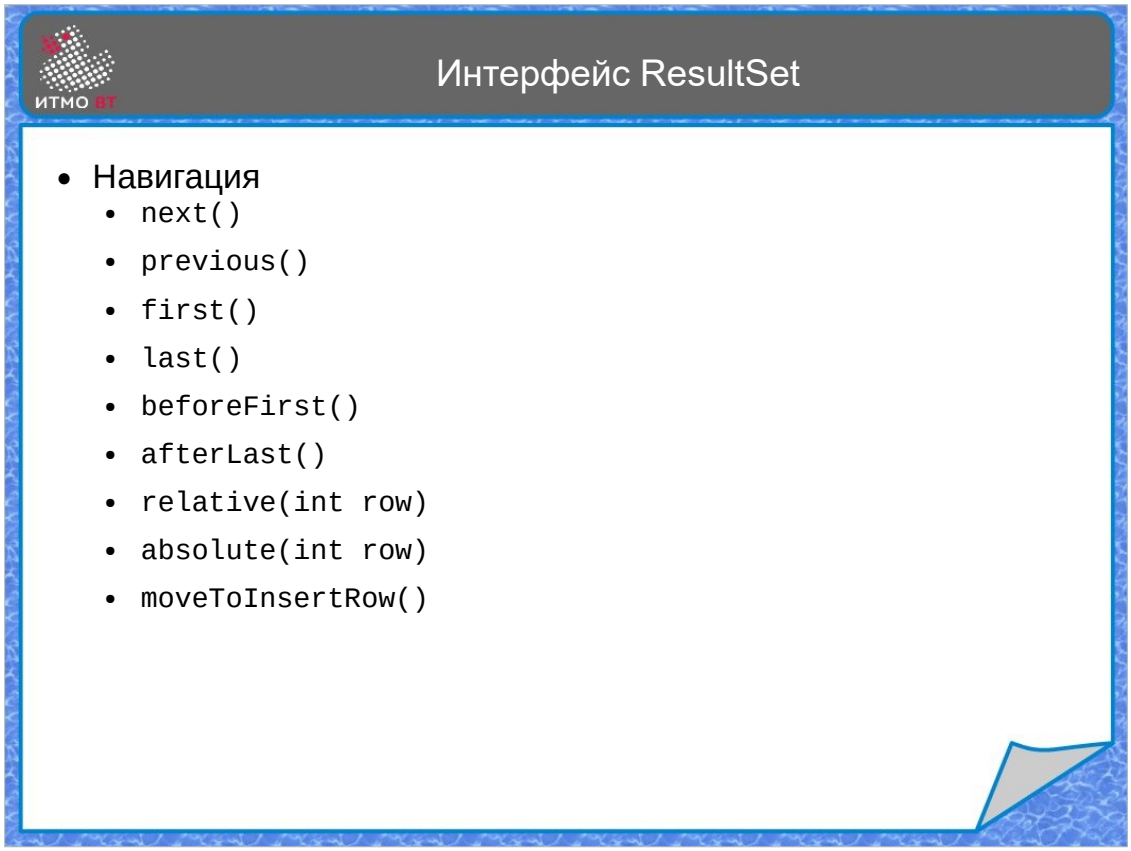
Запросы можно выполнять пакетами. Для этого можно вызвать метод addBatch() который добавляет запрос в пакет. После формирования пакета можно выполнить все запросы за один раз методом executeBatch(). Метод clearBatch() очищает пакет.



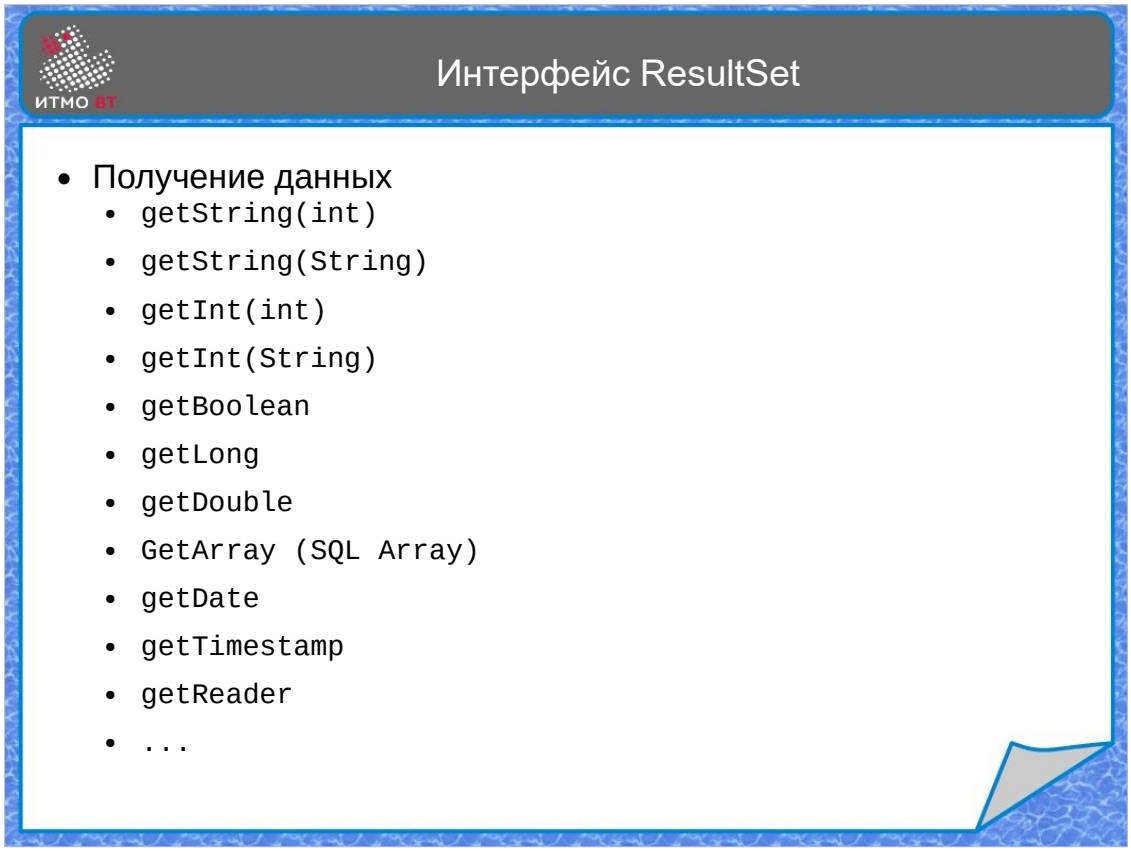
Переходим к рассмотрению интерфейса ResultSet, объект которого возвращают методы executeQuery и getResultSet. Получить результат из ResultSet можно как из обычного итератора, в цикле while вызывается метод next для установки курсора на очередную строку. С помощью методов getString(), getInt() и т. д. можно получать значения столбцов для текущей строки.



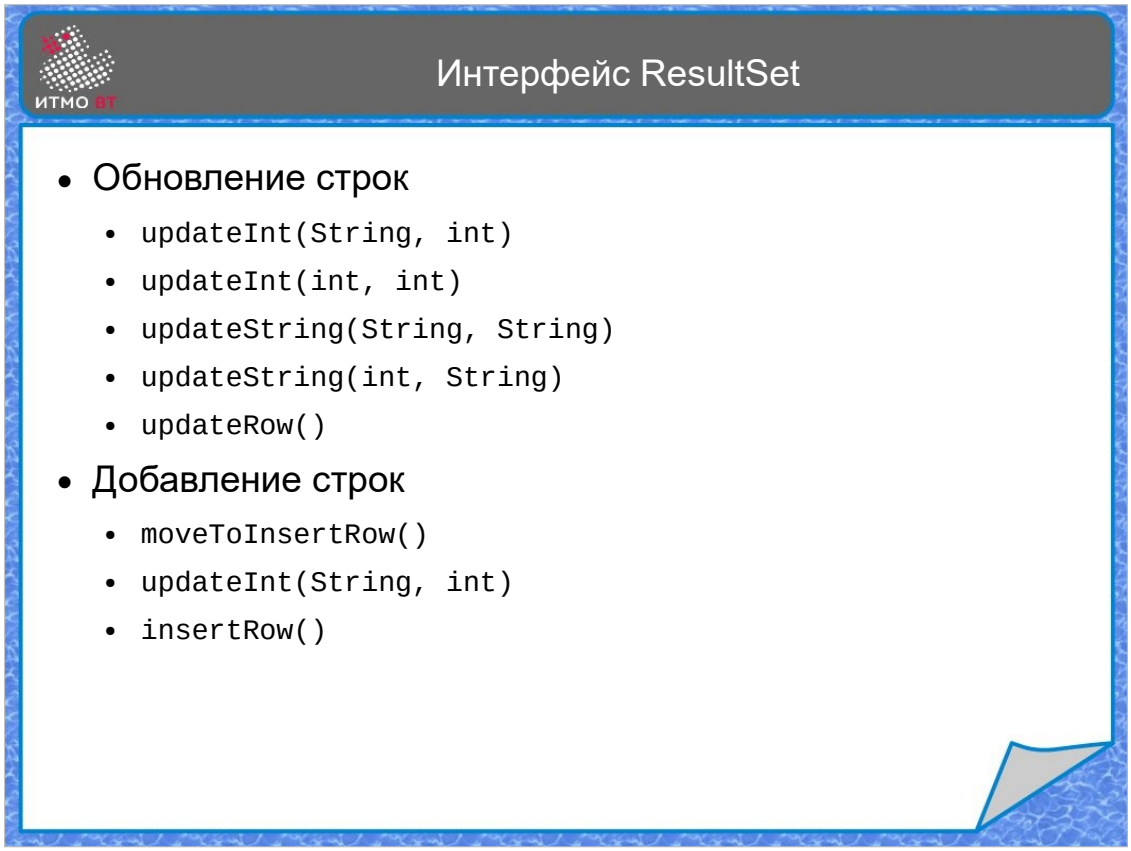
При создании запроса можно указать, какие характеристики будет иметь выдаваемый запросом ResultSet. По умолчанию заданы FORWARD_ONLY и CONCUR_READ_ONLY, при этом ResultSet позволяет перемещать курсор только вперед и работает только на чтение. Если задать любой из типов SCROLL, то перемещать курсор можно будет в обоих направлениях. При этом в INSENSITIVE не видны изменения данных, произошедшие после получения результата, а в SENSITIVE – видны. Характеристика CONCUR_UPDATABLE позволяет вносить в ResultSet изменения, передающиеся в базу данных. Еще 2 характеристики управляют состоянием курсора после команды COMMIT.



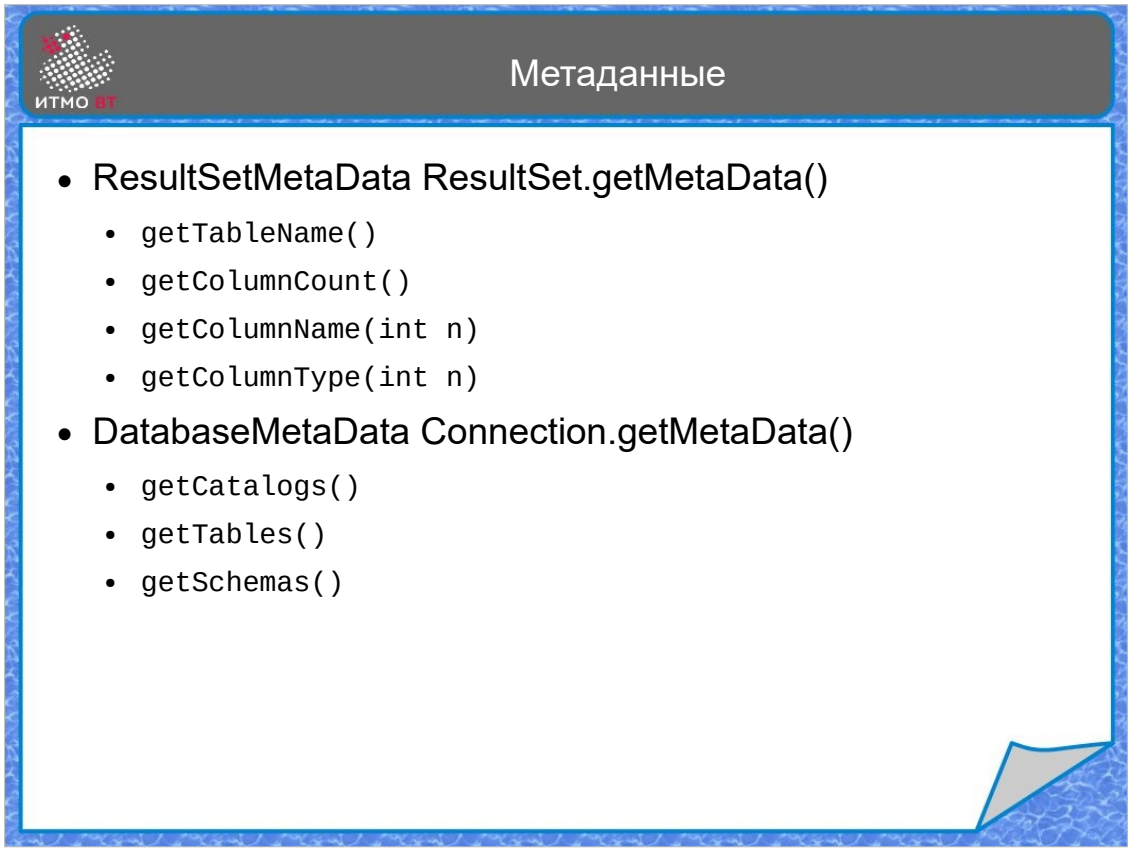
Часть методов интерфейса ResultSet предназначена для перемещения курсора на следующую, предыдущую, первую или последнюю строку, Также курсор можно установить перед первой или после последней строки, переместить его абсолютно или относительно, либо установить на специальную строку для вставки.



Другие методы нужны для получения данных - это методы get с разными типами данных SQL. Все они принимают в качестве параметра либо порядковый номер элемента в строке, либо имя столбца. Второй способ предпочтительнее, так как при изменении запроса не придется менять код для получения результата.

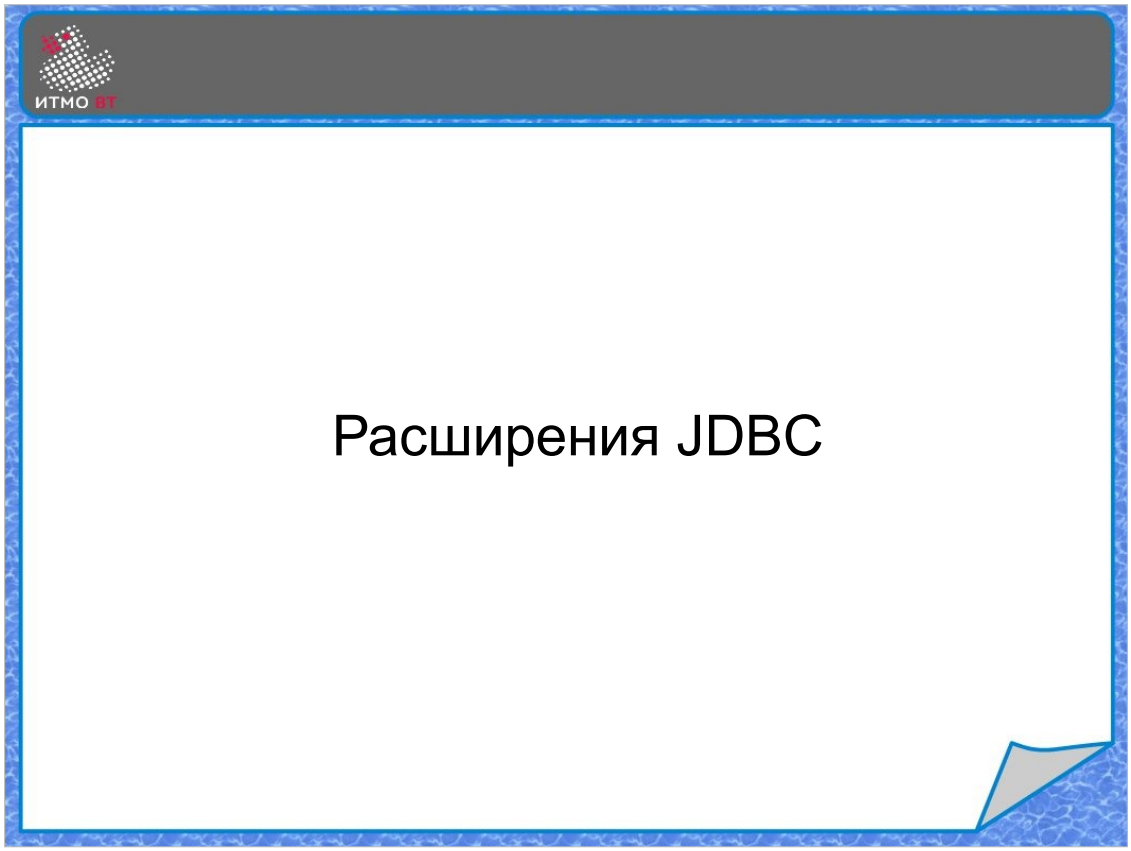


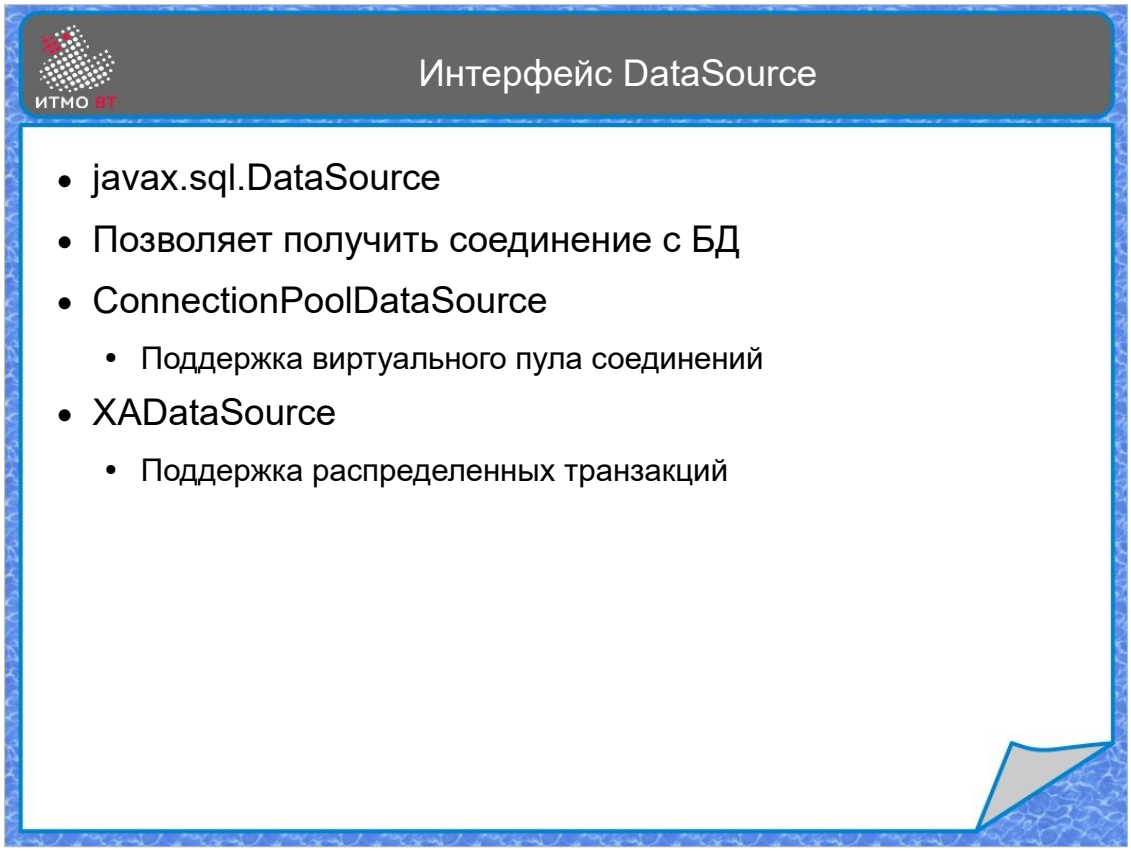
Методы обновления данных позволяют изменить данные Сначала вызываются методы updateInt, updateString и т. д., указывающие, какие столбцы в строке должны быть обновлены. Этим методам передается либо порядковый номер, либо имя столбца, и новое значение. Затем вызывается метод updateRow для обновления данных в базе. Для добавления новой строки нужно переместить курсор на специальную строку методом moveToInsertRow, затем установить нужные значения и потом обновить таблицу базы данных с помощью метода insertRow.



С помощью метода getMetaData у объекта ResultSet можно получить метаданные результата, например, количество, имена и типы данных столбцов, имя таблицы, из которой получен столбец, и т. д.

С помощью метода getMetaData у объекта Connection можно получить метаданные о самой базе данных, например, какие каталоги есть в данной базе, какие схемы в ней определены, какие таблицы хранятся в базе и т. д.

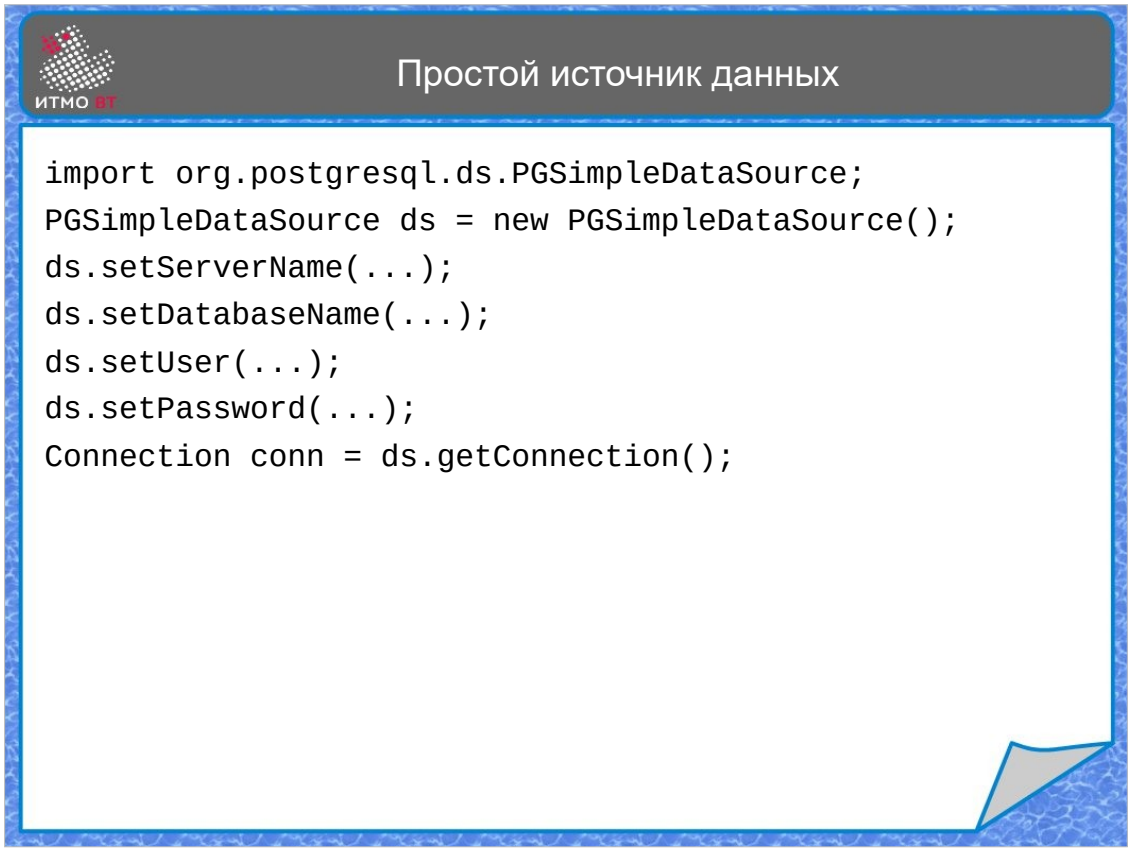

ИТМО ВТ
Расширения JDBC



Кроме классов, находящихся в пакете java.sql и реализующих базовые возможности взаимодействия с базой данных, имеется библиотека дополнительных классов и интерфейсов в пакете javax.sql. Начнем их рассмотрение с интерфейса DataSource. Этот интерфейс представляет источник данных и позволяет получить соединение с базой данных. В JDBC определены 3 типа источников данных: простой, с поддержкой пула соединений и с поддержкой распределенных транзакций.

ConnectionPoolDataSource позволяет снизить затраты на создание соединений. Он работает аналогично пулу потоков, когда после использования соединение не закрывается, а возвращается в пул.

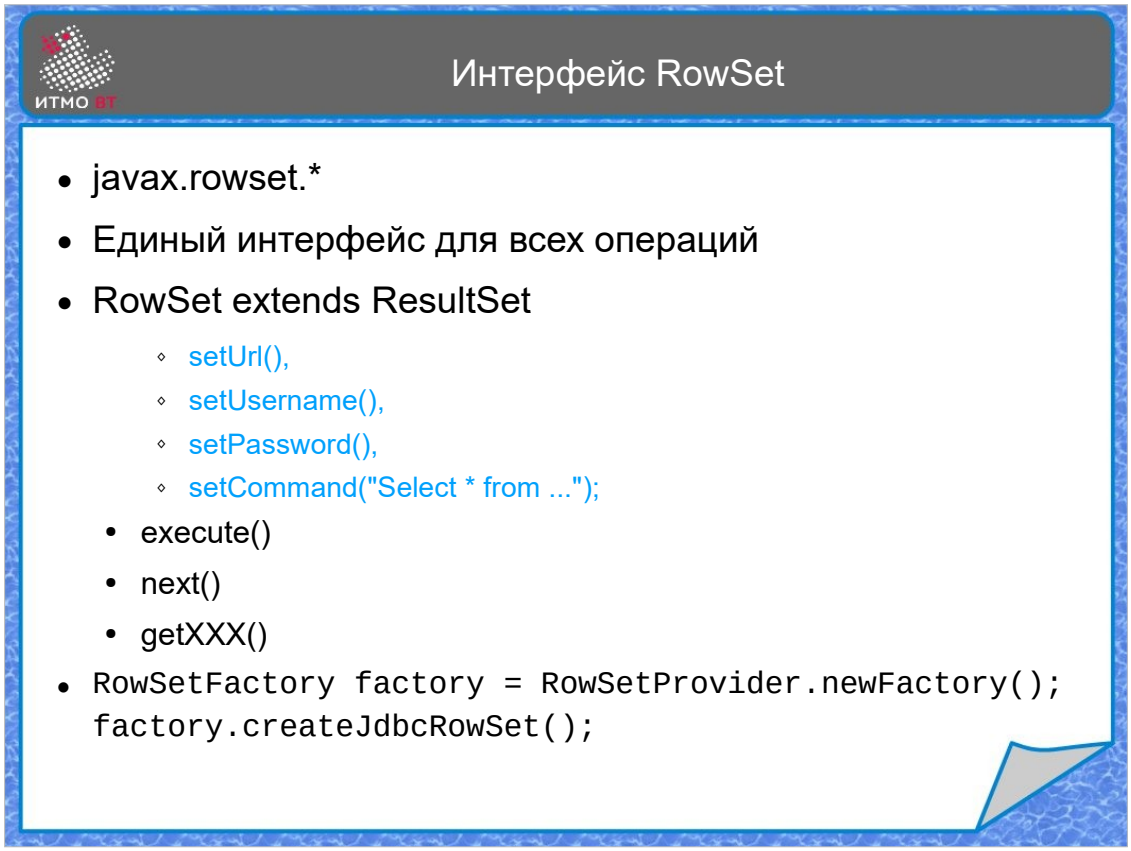
XADataSource, позволяет работать с таблицами, находящимися в разных базах данных, при этом менеджер транзакций обеспечивает непротиворечивость данных.



Чтобы получить объект соединения, нужно создать базовый источник данных для соответствующей СУБД. Затем с помощью соответствующих методов установить параметры соединения. И вызвать у объекта DataSource метод getConnection.

В корпоративных приложениях источники данных обычно получают через службу имен. Доступ к ней производится через интерфейс JNDI (Java Naming Directory Interface). При использовании JNDI системный администратор создает контекст и источник данных, устанавливает необходимые параметры соединения с базой данных, затем привязывает источник данных к элементу каталога службы имен.

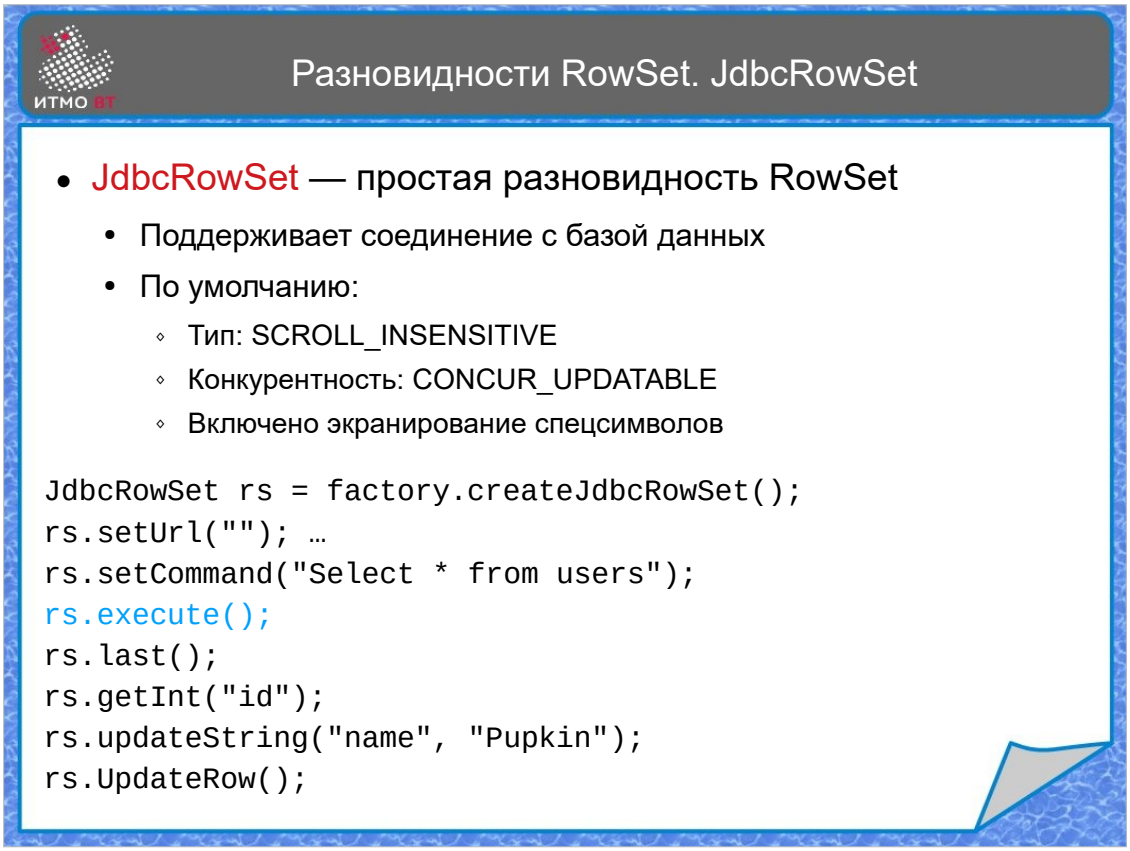
Далее разработчик приложения получает из контекста по имени источник данных и может пользоваться им для установления соединения с базой данных, без явного задания параметров соединения (в том числе логина и пароля) Таким образом можно отделить управление базами данных от приложения.



В расширенный пакет входит интерфейс RowSet. Он расширяет ResultSet. Кроме работы с результатом запроса RowSet может осуществить соединение с базой данных и выполнить сам запрос. В этом интерфейсе есть методы setURL, setUsername и setPassword для задания параметров соединения. Метод setCommand() позволяет задать запрос, метод execute выполняет запрос и получает результат.

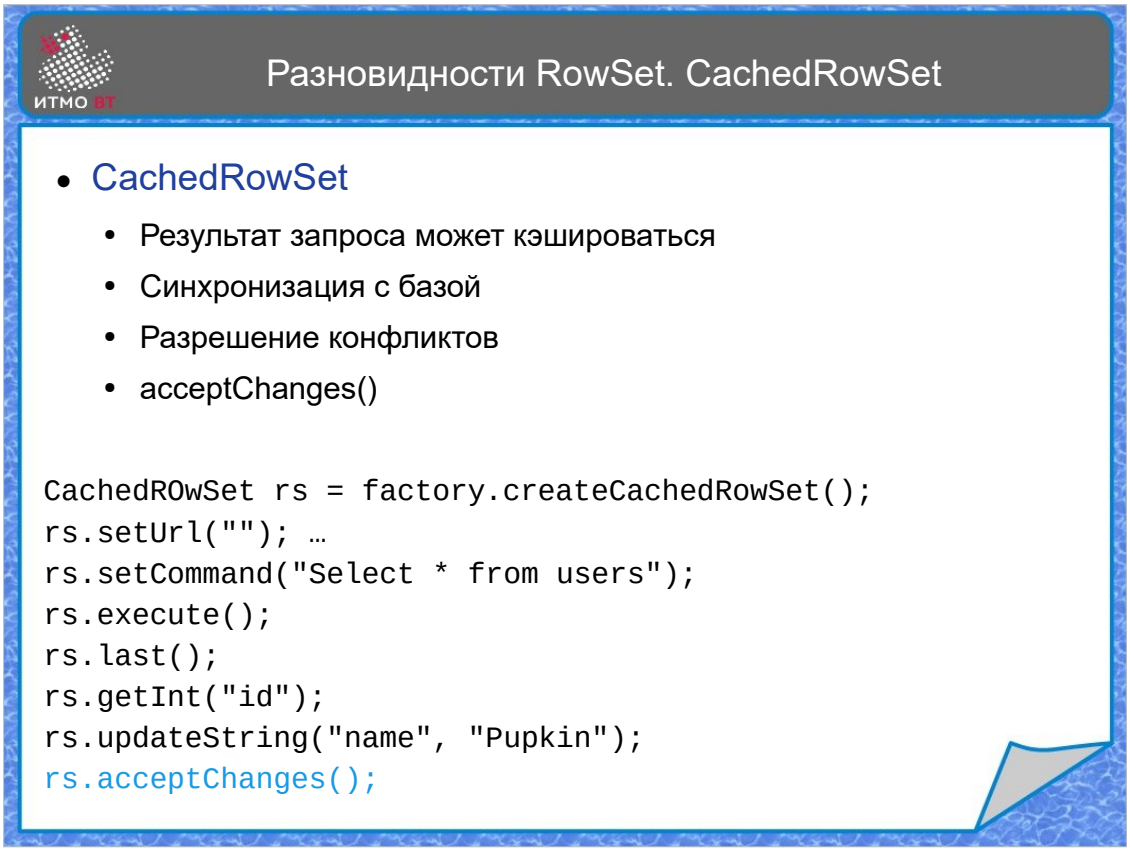
RowSet может поддерживать постоянное соединение с базой данных, а может соединяться с базой только в случае необходимости.

В пакете javax.sql имеется класс RowSetProvider, который возвращает фабрику, способную создавать объекты разных типов, реализующих интерфейс RowSet.



JdbcRowSet — это самый простой из интерфейсов-наследников RowSet. Он единственный из них поддерживает соединение активным. Его удобно использовать вместо обычного ResultSet, когда не нужна дополнительная функциональность. Созданный по умолчанию JdbcRowSet — скроллируемый и поддерживает обновления.

Краткое описание использования — получаем объект типа JdbcRowSet, устанавливаем параметры соединения, задаем запрос с помощью setCommand. Далее выполняем запрос методом execute, после чего RowSet готов к работе, можно получать и обновлять данные.



CachedRowSet — это наследник RowSet с поддержкой кэширования, который может не поддерживать постоянное соединение с базой. Оно устанавливается только в нужные моменты, при этом происходит синхронизация с базой данных изменений, сделанных в RowSet. Иногда при этом может потребоваться разрешение конфликтов.

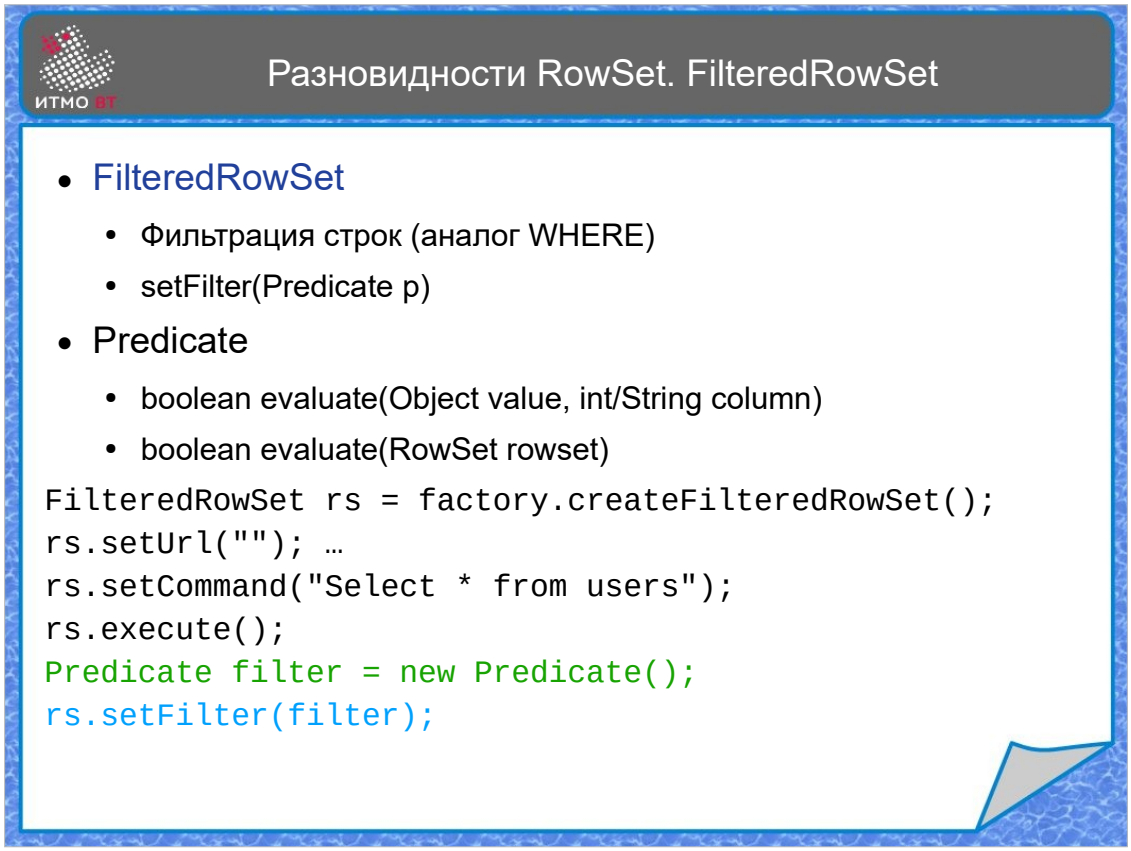
В отличие от JdbcRowSet изменения не сразу отражаются в базе, так как нет постоянного соединения, изменения переносятся в базу при вызове метода acceptChanges().

Разновидности RowSet. WebRowSet

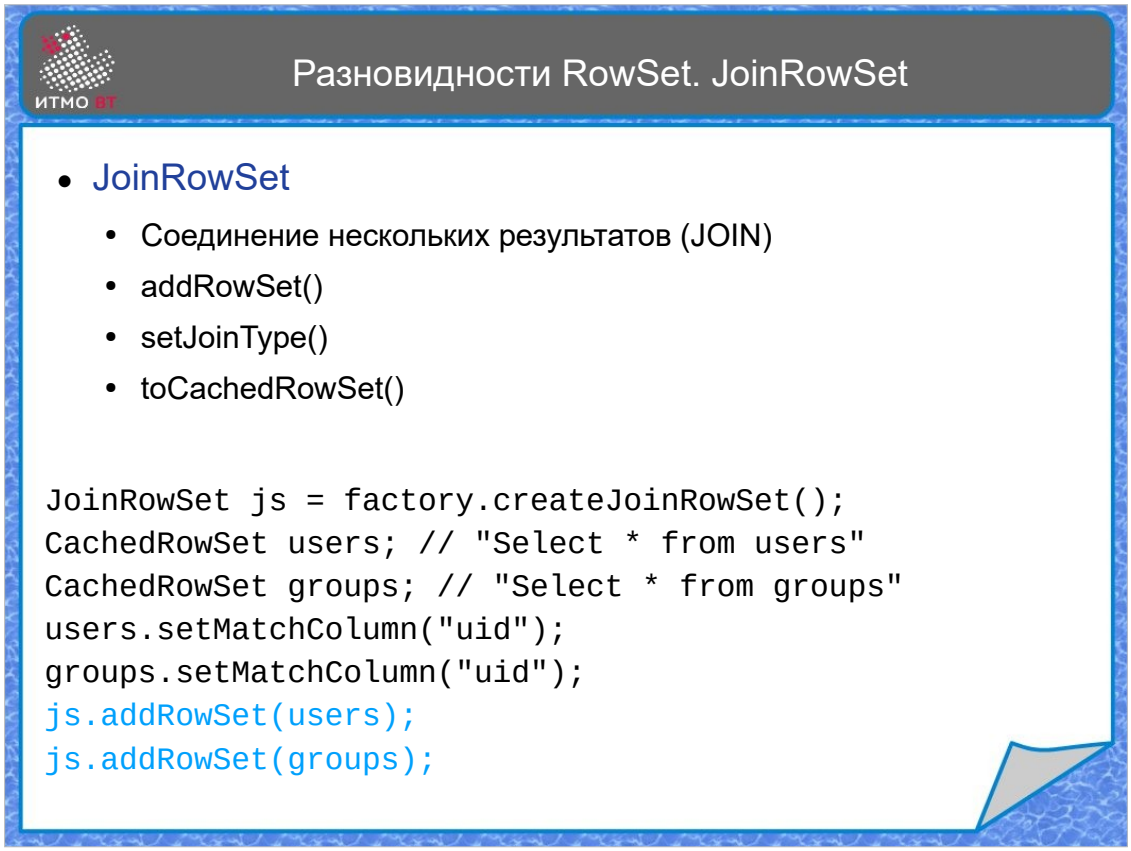
- WebRowSet
  - Может записывать и читать результат в виде XML
  - writeXML()
  - readXML()

```
WebRowSet rs = factory.createWebRowSet();
rs.setUrl(""); …
rs.setCommand("Select * from users");
rs.execute();
rs.writeXML(new FileWriter("data.xml"));
```

WebRowSet — это наследник CachedRowSet, который дополнительно умеет сохранять результат в виде XML. Кроме сохранения в виде XML, WebRowSet может прочитать XML данные либо в новый объект WebRowSet, либо в тот, который использовался для записи XML.



FIlteredRowSet — это расширение интерфейса WebRowSet с добавлением поддержки фильтрации результата по какому-нибудь условию, аналогично фразе WHERE. Условие можно задать с помощью метода setFilter, которому передается объект типа Predicate. Тогда в RowSet будут доступны только строки, удовлетворяющие данному условию. Отменить фильтр можно установкой пустого фильтра.



JoinRowSet — это также расширение WebRowSet с возможностью соединения нескольких объектов RowSet, аналогично оператору JOIN в SQL. Соединять можно наборы строк, реализующие интерфейс Joinable. Сначала задаются столбцы, по которым делается соединение, а затем объекты RowSet добавляются в JoinRowSet. Методом setJoinType можно задать тип соединения.