

ВЗАИМОДЕЙСТВИЕ R И SQL SERVER

Рувим БОЗАЖИ

Департамент Программной Инженерии и Автоматики, группа TI-195, Факультет Вычислительной Техники, Информатики и Микроэлектроники, Технический Университет Молдовы, Кишинев, Республика Молдова

Автор корреспондент: Рувим БОЗАЖИ, e-mail: bozaji.ruvim@isa.utm.md

Научный руководитель: Дориан САРАНЧУК, DISA, FCIM, UTM

Аннотация: в данной работе проводится описание взаимодействия R и SQL server. В статье представлены история интеграции, а также описаны базовые возможности, принцип работы и основные инструменты.

Ключевые слова: SQL Server, R Services, T-SQL, база данных

Введение

R — это интерпретируемый язык программирования с минимальным набором примитивных типов данных для статистической обработки данных с открытым исходным кодом. Основной способ работы с R – командный интерпретатор. Используется как ПО для анализа данных, статистической графики и является одним из самых популярных языков науки о данных [1].

Для удобной работы языка, были созданы графические интерфейсы, такие как RStudio, а также в некоторых редакторах кодов предусмотрены специальные режимы для R. Со временем этот язык, с огромным количеством библиотек для машинного обучения, практически стал стандартом для статистических программ. Большие возможности R стали средством расширения функционала SQL.

1. Интеграция R в SQL

В 2016 году, преследуя цель обеспечить расширенную аналитику на локальных платформах данных, Microsoft приобрела R, после чего интегрировала этот язык с SQL Server и с другими программными обеспечениями [2].

Таким образом Revolution R Enterprise, самое быстрое и экономичное решение для масштабной аналитики и доступа к информации, скрытой в данных, был переименован в SQL Server R Services.

R Services - это функция, которая интегрирует язык R с SQL Server дает возможность запускать скрипты R с реляционными данными, выполнять аналитику и устранять риски безопасности, связанные с перемещением данных. Архитектура спроектирована таким образом, что внешние скрипты, использующие R, выполняются в отдельном процессе от SQL Server [3].

2. Взаимодействие с SQL

R, как язык программирования с открытым исходным кодом, имеет все программные конструкции, такие как переменные, циклы, операторы. Однако его ключевым отличием является то, что он состоит из огромного количества библиотек для статистических вычислений, визуализации, обработки естественного языка, машинного обучения и т. Д.

Когда возникает необходимость вычислить определенную статистику в обычном коде T-SQL, появляется проблема с тем что это требует очень сложной логики в своей реализации. Поэтому используется R для этой цели, поскольку есть готовые к использованию библиотеки, которые могут вычислять эти значения без необходимости изучения или реализации каких-либо сложных статистических формул.

Так как для выполнения скриптов в SQL используется SQL Server Management Studio, то его же можно использовать чтобы вписать R в экосистему SQL.

Запуск команд на R внутри SQL Server создаётся с помощью отдельной хранимой процедуры *sp_execute_external_script*, которая принимает несколько параметров. Основными обязательными параметрами являются *@language* и *@script*. *@language* принимает значение – R. А в *@script* передается код на языке R, который необходимо выполнить. При запуске R из T-SQL с помощью служб машинного обучения, все скрипты R выполняются вне процесса SQL Server, это обеспечивает безопасность и большую управляемость. А пользователь получает к себе результаты в виде таблиц, графиков [4].

Второй способ применить R в SQL - с помощью специального API посылать команды на SQL Server, где выполняются все вычисления, а возвращаются только результаты. Таким образом есть возможность избежать долгого процесса загрузки данных на свою машину.

Одна из главных особенностей — это возможности R по графическому представлению данных. R позволяет генерировать эти графики внутри SQL Server 2016 и отображать в Reporting Services.

3. Проблемы

R обрабатывает только те данные, которые на данный момент расположены в оперативной памяти что создает проблемы, когда данные выходят за пределы оперативной памяти. Поэтому подобные задачи возможно решить если разделить их по частям. R Server содержит расширенные пакеты и функции, которые позволяют обрабатывать данные многопоточно, загружая в память необходимые отрывки по мере необходимости. Все эти технологии были встроены и в SQL Server.

Т.к. T-SQL и R разные языки, каждый со своими типами данных, то сразу возникает много вопросов о совместимости между ними. R поддерживает только ограниченный набор типов: *numeric*, *integer*, *complex*, *logical*, *character*, *date/time* и *raw*. Все типы T-SQL будут неявно преобразованы в их соответствующие аналоги в R.

Заключение

R Services значительно расширяет возможности SQL Server особенно в области анализа данных и предоставляет те функции, которые невозможно выполнить в T-SQL. SQL Server R Services предоставляют большой дополнительный функционал и ряд библиотек которые значительно упрощают работу. Задачи обработки больших объемов данных обычно предпочтительнее выполнять в T-SQL, а с помощью R возможно выполнять специальные задачи, такие как статистические вычисления, машинное обучение, обработка больших объемов данных.

Библиография

1. Getting started with R in SQL Server. - [online]. – [дата обращения 02.03.2023], Доступно: <https://www.mssqltips.com/sqlservertutorial/9040/getting-started-with-r-in-sql-server/>
2. R development on SQL Server - [online]. – [дата обращения 02.03.2023], Доступно: <https://docs.microsoft.com/en-us/sql/machine-learning/r/sql-server-r-services?view=sql-server-2016>
3. SQL Server 2016: R Services. - [online]. – [дата обращения 02.03.2023], Доступно: <https://olontsev.ru/tag/r-services/>
4. Служба SQL Server R – для чего? - [online]. – [дата обращения 02.03.2023], Доступно: <http://wisedata.ru/2017/01/12/rec-sql-server-r-service/#respond>