УДК 004.62 DOI: 10.30748/zhups.2018.56.14

В.В. Федько

Харківський національний економічний університет ім. С. Кузнеця, Харків

АНАЛІЗ ДАНИХ В SQL SERVER ЗАСОБАМИ PYTHON

Розглянуто можливості інструментальних засобів аналізу даних. Викладено способи зберігання даних, які адаптовані до ефективного виконання запитів аналізу даних, а також мовні засоби, що представлені в компоненті Microsoft SQL Server як Machine Learning Services (in-database). Проведено порівняння операційних баз даних (OLTP-систем) і сховищ даних, які орієнтовані на аналіз даних (OLAP-систем). Наведено приклади обох систем, а також розглянуто систему їхньої взаємодії (ETL-система). Описано інструментальні засоби аналізу даних, які в найпростіших випадках застосовуються до OLAP-кубів. Подано мовні засоби виконання аналізу даних в більш складних випадках. Проведено порівняння мов R і Python, з якого випливає, що мова Python дозволяє будувати завершені програми обробки даних, а бібліотеки в ній майже такі самі, як і в мові R. Показано, що, з огляду на велику популярність мовних засобів аналізу даних в останні випуски SQL Server включено компонент SQL Server R Services, в результаті чого нові можливості в SQL Server дозволили обійти обмеження, яке полягає в тому, що всі дані повинні зберігатися в пам'яті. Описано основні переваги компонента Machine Learning Services, а також особливості його установки. Продемонстровано на конкретних прикладах можливості виконання розрахунків і графічного подання результатів мовою Python у середовищі SQL Server для проведення аналізу даних.

Ключові слова: *Business intelligence, Data mining, Data Scientist, Data Engineer, SQL Server, Machine Learning Services, операційна база даних, сховище даних, мови R та Python, аналіз даних, візуалізація даних.*

Вступ

Постійний моніторинг процесів, що відбуваються в бізнесі, став важливою рисою функціонування абсолютної більшості компаній [1–2]. Його якість в значній мірі забезпечує конкурентоспроможність фірми. Для успішного ведення бізнесу необхідні знання щодо стану загальноприйнятих показників, прояв яких є явним, а також процесів, що тільки починають виявлятися, а в цілому мають поки що латентний характер [3–4]. Для дослідження процесів першого типу в аналізі даних сформувався напрямок, який прийнято називати бізнес-аналізом. У закордонних виданнях він отримав скорочену назву BI (від англ. Business intelligence) [5–6]. Другий тип аналізу даних ще не сформувався остаточно і в наш час з'являється багато нових публікацій у цьому напрямку. У закордонних публікаціях його називають Data mining, [7]. Для комп'ютерної реалізації задач аналізу даних ряд провідних компаній створили відповідні інструментальні засоби [8]. В даній роботі розглядаються засоби, що мають загальну назву Machine Learning Services (in-database), наводяться приклади їхнього використання.

Основна частина

Аналіз даних на підприємстві виконують бізнес-аналітики. Вони дають економічне трактування результатам, що отримані на основі бізнес-моделей. У зв'язку із значним ускладненням процесів та зростанням обсягів даних для розробки таких моделей з'явилася нова IT-спеціальність Data Scientist, яка ще не має українського відповідника, але спеціалісти у цій галузі вже активно працюють. У США їхня середня заробітна платня становить 91 тис. дол. на рік. Для побудови і використання моделей інші фахівці, а саме інженери даних (Data Engineer), готують дані, які придатні для обробітку, а також створюють відповідні застосування. У США їхня середня заробітна платня майже співпадає з тією, що отримують Data Scientist, і становить 91 тис. дол. на рік [9–10].

Перехід бізнесу на комп'ютерну обробку даних відзначився збереженням даних в базах даних. Були створені системи, що орієнтувалися на операційну (транзакційну) обробку даних , так звані OLTP-системи (On-Line Transaction Processing – оперативна транзакційна обробка). Їхньою основною функцією є одночасне виконання великої кількості коротких транзакцій від значного числа користувачів. Системам OLTP притаманні такі риси:

велика кількість інформації;

часто автономні бази даних для різних підрозділів;

нормалізована схема, відсутність дублювання інформації;

інтенсивна зміна даних;

транзакційний режим роботи;

у транзакціях бере участь невелика кількість даних;

переважна обробка поточних даних;

значна кількість клієнтів;

малий час відгуку (не більше кількох секунд).

Спроби виконання аналізу даних на таких системах зустрілися з необхідністю очікування виконання запитів протягом значних проміжків часу. Це спричинялося в першу чергу нормалізацією баз

даних, оскільки часто в запитах брала участь велика кількість таблиць. Таблиці пов'язувалися операціями JOIN, на виконання яких потрібно багато часу, особливо при значній кількості записів у таблицях.

Для вирішення зазначеної вище проблеми для розв'язання задач аналізу даних запропоновано так звані OLAP-системи (On-Line Analytical Processing – оперативна аналітична обробка даних). Їхніми функціями є аналіз даних, моделювання процесів предметної області, прогнозування, виявлення залежностей між даними тощо. Системи OLAP мають такі риси:

велика кількість інформації;

синхронізована інформація із різних баз даних;

ненормалізована схема бази даних з дублюванням;

дані змінюються рідко;

виконуються складні нерегламентовані запити над великим обсягом даних з використанням груп та агрегатних функцій;

аналіз часових залежностей;

невелика кількість користувачів (аналітики та менеджери);

великий час відгуку (але прийнятний – кілька хвилин).

Основою кожної системи є база даних і в OLTP-системі це нормалізована база даних. При наявності у ній ста і більше таблиць часто запити для розв'язання задач аналізу стають надто складними і довготривалими. Тому в основу OLAP-систем закладено сховище даних, схема якого є ненормалізованою. Найчастіше його проектують за схемою зірки, тобто воно має одну таблицю, у якій зберігаються дані, що аналізуються (таблицю фактів), і таблиці-довідники, у яких пояснюється призначення параметрів аналізу (таблиці вимірів). Це дозволяє замінити у запитах довгі ланцюжки із операцій JOIN на одноланкові.

Розглянемо відмінності між операційними базами даних і сховищами даних на прикладі модельної задачі про облік продажів хлібобулочних товарів у кіоску, де не видаються чеки. Він ведеться за результатами продажів у кінці дня. Таку операційну база даних Хліб у спрощеному вигляді подано на рис. 1.

На відміну від неї сховище даних ХлібСД подається схемою, що зображена на рис. 2.

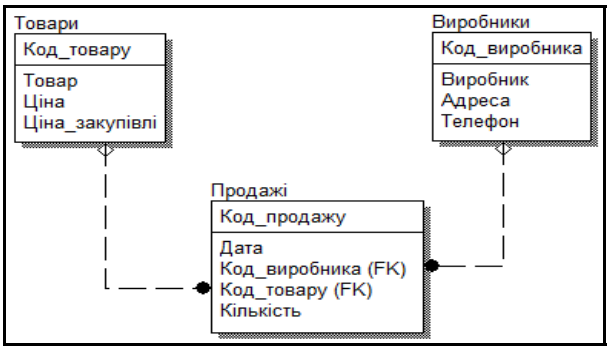

Рис. 1. Схема операційної бази даних Хліб

У цьому найпростішому випадку на перший погляд може здаватися, що сховище даних має більш складну схему (у ній на одну таблицю більше, ніж в операційній базі даних). Але завдяки тому, що дати винесено в окрему таблицю, значно поліпшується побудова запитів за виміром часу з різною глибиною. При використанні сховища даних немає потреби у виконанні функцій над датами при вилученні з них значень року, номера місяця чи дня місяця. Вартість проданого за день товару, що зберігається у таблиці фактів сховища даних, також обчислена заздалегідь і вона вже готова для використання у запитах аналізу. Якщо б операційна база даних ускладнилася, наприклад, облік продажів вівся за чеками, то схема операційної бази даних ускладняється, що призводить до збільшення часу на виконання запитів аналізу до неї. Схема сховища даних у цьому разі не змінюється.

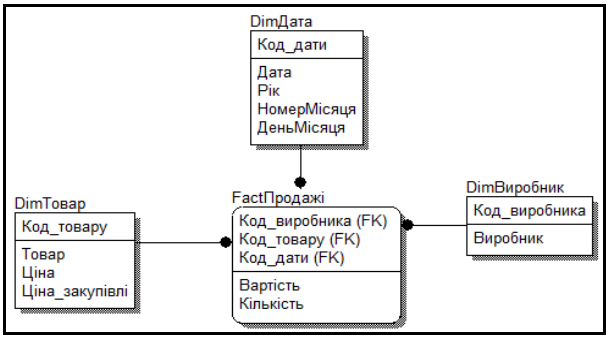

Рис. 2. Схема сховища даних ХлібСД

Сховище даних заповнюється даними, що зберігаються в операційній базі даних. У найпростішому випадку це може бути такий скрипт:

```
INSERT INTO [Шлях до сховища даних].dbo.FactПродажі
   (Код_дати, Код_виробника,Код_товару, Кількість, Вартість)
SELECT CONVERT(char(8),п.Дата,112) AS Код_дати,
п.Код_виробника,
   п.Код_товару, п.Кількість,
   т.Ціна*п.Кількість AS Вартість
FROM Товари т
   JOIN Продажі п ON т.Код_товару=п.Код_товару
WHERE п.Дата Between  @ПочатковаДата And
@КінцеваДата;
```

У промислових базах даних пересилання даних з операційної бази даних до сховища здійснюється так званими системами ETL (Extract, Transform, Load – вибірка, перетворення, завантаження). Зокрема для виконання цих процесів Microsoft створив службу SSIS (SQL Server Integration Services).

Після заповнення сховища даних за певним графіком виконують аналіз даних. У найпростіших випадках аналітики використовують загально розповсюджені інструментальні засоби такі як Excel, Power BI, Tableau тощо. Вони під'єднуються безпосередньо до сховища даних як джерела даних або у

більш складних випадках до кубів OLAP. Останні створюються спеціальними засобами, зокрема для цього Microsoft створив службу SSAS (SQL Server Analysis Services).

У більш складних випадках, коли для проведення аналізу даних недостатньо готових інструментальних засобів, використовують спеціальні пакети, у яких зібрана велика кількість модулів, що реалізують різноманітні алгоритми. Для «склеювання» окремих модулів з метою побудови ланцюжка процесу обробки даних такі пакети мають спеціальні мови. Найбільш поширеною мовою у цьому напрямку є мова R. На даний час для неї розроблено понад 9000 бібліотек, в яких реалізовано різноманітні статистичні алгоритми обробки даних та візуального подання результатів [11–12].

Останнім часом у сферу обробки даних активно увійшла мова Python [13]. На відміну від мови R вона не тільки може подавати фрагменти коду у застосуваннях, що реалізуються іншою мовою, наприклад, C#. Python дозволяє будувати завершені застосування з обробки даних, а бібліотеки в ній майже ті самі, що й в мові R. За кількістю користувачів Python лише на рік відстає від мови R, але за темпами поширення перевищує її у кілька разів. Тому очікується, що в найближчі роки мова Python опередить R.

Враховуючи значну популярність мовних засобів аналізу даних до випуску SQL Server 2016 включено компонент SQL Server R Services. Нові можливості у SQL Server дозволяють обходити обмеження, яке полягає в тому, що всі дані мають зберігатися у пам'яті. SQL Server 2016 надає можливість аналізувати більші обсяги інформації, які можуть виходити за рамки оперативної пам'яті. Розроблено нові пакети, що дозволяють обробляти дані у декілька потоків та отримувати дані за необхідністю. Усі розрахунки можна проводити на сервері, де розгорнута база даних, завдяки чому зникає необхідність використовувати пам'ять та інші ресурси клієнтських комп'ютерів. R Services дозволяють виконувати будь-який код мовою R та встановлювати додаткові пакети для різноманітних цілей, які розширюють можливості цієї мови.

З точки зору безпеки даних використання R Services ліквідує необхідність передавати усі дані у мережі. А також скрипти, що виконуються у R Services, не мають доступу до мережі, чи інших процесів на сервері і тому не можуть зашкодити.

Оскільки мова R має широкі графічні можливості, можна будувати графіки безпосередньо у SQL Server і відображати у Reporting Services або десктопному чи веб застосуванні, що розробляється на замовлення аналітиків.

Роботу з R Services можна організувати за двома сценаріями. За першим сценарієм власна програма може виконувати збережені процедури на SQL Server, які містять код мовою R, і отримувати з них результати у вигляді таблиці чи графіків. За другим сценарієм аналітик може відправляти код мовою R у SQL Server на виконання і отримати готовий результат. Таким чином зникає необхідність завантажувати дані на локальний комп'ютер.

У випуску SQL Server 2017 розробники компанії Microsoft пішли ще далі. Вони додали мову Python. Тепер обидві мови R та Python у SQL Server утворюють компонент Machine Learning Services (in-database). Він встановлюється під час інсталяції ядра SQL Server чи додається пізніше. Під час встановлення компонента потрібно зважати на те, що з двох мов R та Python на одній сутності сервера можна встановити лише одну мову. Якщо вибрати обидві мови, процес інсталяції завершиться нормально, але жодна з них не буде встановлена. У разі, коли потрібні обидві мови, встановлюють додаткову сутність SQL Server.

Після встановлення мови Python слід увімкнути можливість виконання скриптів Python, виконавши в SQL Server Management Studio такий SQL-скрипт:

```
EXEC sp_configure  'external scripts enabled', 1
RECONFIGURE WITH OVERRIDE
```

Розробники баз даних та програмного забезпечення отримують доступ до великих бібліотек Python з екосистеми з відкритим кодом, включаючи популярні стеки технологій scikit-learn, TensorFlow, Caffe та Theano / Keras.

Запуск на виконання Python-скриптів у середовищі SQL Server Management Studio здійснюється за допомогою системної збереженої процедури sp_execute_external_script. Вона має такий синтаксис:

```
sp_execute_external_script
   @language = N'language',
   @script = N'script'
   [ , @input_data_1 = N'input_data_1' ]
   [ , @input_data_1_name = N'input_data_1_name' ]
   [ , @output_data_1_name = N'output_data_1_name' ]
   [ , @parallel = 0 | 1 ]
   [ , @params = N'@parameter_name data_type [ OUT |
OUTPUT ] [ ,...n ]' ]
   [ , @parameter1 = 'value1' [ OUT | OUTPUT ] [ ,...n ] ]
   [ WITH RESULT SETS { UNDEFINED | NONE | ((
OutputColumnName [,...n ] )) } ]
```

Основні параметри процедури мають таке призначення:

@language – мова програмування, на якій подається скрипт (R або Python);

@script – текст самого Python-скрипта. При цьому потрібно дотримуватися усіх правил мови Python, у тім числі і відступи у вигляді чотирьох пробілів;

@input_data_1 – SQL-скрипт, що визначає вхідні дані, які обробляються поточним Python-скриптом;

@output_data_1_name – ім'я змінної в Python-скрипті, який містить дані, що повертаються для SQL Server після завершення виклику збереженої процедури. Значення за замовчанням – OutputDataSet;

WITH RESULT SETS – задає імена стовпців, що виводяться в результаті.

Більш детально про ці та інші параметри процедури sp_execute_external_script можна дізнатися в [14].

Як приклад виконання Python-скриптів у середовищі SQL Server Management Studio розглянемо побудову зведеної таблиці про продаж хлібобулочних товарів за 2017 рік. У модельному сховищі даних ХлібСД зберігаються дані про продаж трьох товарів від двох виробників за 2010–2017 р.р. (усього 17500 записів). Дані таблиці фактів сформовано випадковим чином. Для побудови зведеної таблиці можна використати такий скрипт:

```
execute sp_execute_external_script
@language = N'Python',
@script = N'
import pandas as pd
from pandas import DataFrame
import numpy as np
df=pd.DataFrame(InputDataSet)
df1=pd.pivot_table(df,index=["Виробник"],
columns=["Товар"],values=["Вартість проданих товарів"], aggfunc=np.sum)
print(df1)
'
,
@input_data_1 = N'SELECT
            в.Виробник as Виробник,
            т.Товар as Товар,
            cast(п.Вартість as float) as [Вартість проданих товарів]
FROM DimТовар т
            JOIN FactПродажі п ON т.Код_товару=п.Код_товару
            JOIN DimВиробник в ON п.Код_виробника=в.Код_виробника
            JOIN DimДата д ON  д.Код_дати=п.Код_дати
WHERE д.Рік=2017
'
```

Отриманий результат подано на рис. 3.

З метою демонстрації графічних можливостей Python-скриптів побудуємо горизонтальну гістограму, на якій подаються вартості проданих товарів за 2017 р.

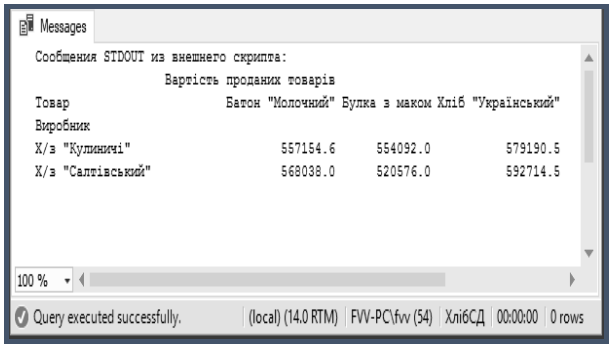

Рис. 3. Зведена таблиця, що отримана Python-скриптом

Для побудови горизонтальної гістограми можна використати такий скрипт:

```
execute sp_execute_external_script
@language = N'Python',
@script = N'
import pandas as pd
from pandas import DataFrame
import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt
df=pd.DataFrame(InputDataSet)
df=df.groupby("Товар").sum()
pt=df.plot.barh()
pt.set_title(label="Вартість проданих товарів за 2017 р.", y=1.04,family="Calibri", fontsize=20, color="black")
pt.legend().set_visible(False)
pt.grid(color="slategray", alpha=.5, linestyle="dotted", linewidth=.5)
pt.set_xlabel("грн")
plt.savefig("F:\\pyPic\\GoodSales2017.png",bbox_inches="tight", pad_inches=.5)
print("Діаграму побудовано. Перегляньте її у папці" + " F:\\pyPic\\")
'
,
@input_data_1 = N'SELECT
            т.Товар as Товар,
            cast(п.Вартість as float) as [Вартість проданих товарів]
FROM DimТовар т
            JOIN FactПродажі п ON т.Код_товару=п.Код_товару
            JOIN DimДата д ON  д.Код_дати=п.Код_дати
WHERE д.Рік=2017'
```

Отриману гістограму збережено у вигляді png-файла і подано на рис. 4.

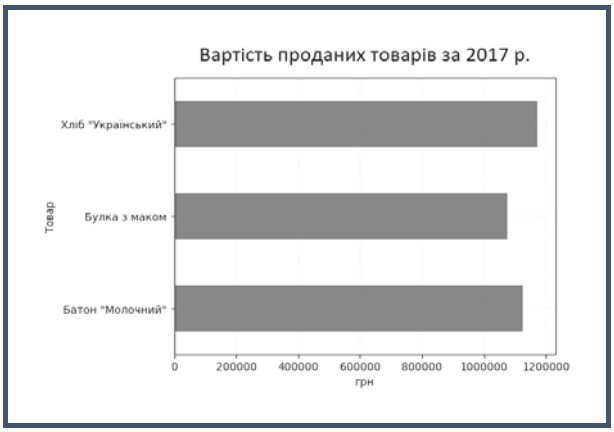

Рис. 4. Горизонтальна гістограма, що отримана Python-скриптом

Висновки

Додавання до SQL Server 2017 компоненти Machine Learning Services (in-database) надало можливість аналізувати значні обсяги інформації, використовуючи тисячі бібліотек, що розроблені у передових дослідницьких центрах і реалізовані мовами C та C++ для більш швидкого виконання. Усі розрахунки проводяться на сервері, де розгорнуто сховище даних, завдяки чому зникає необхідність використовувати пам’ять та інші ресурси клієнтських комп’ютерів. Використання мови Python дозволяє

будувати замкнуті застосування на відміну від мови R, у якій можна подавати лише фрагменти коду у застосуваннях, що реалізуються іншою мовою.

Продемонстровані можливості виконання розрахунків і графічного подання результатів свідчать про достатні можливості мови Python в середовищі SQL Server для проведення аналізу даних і простоту його реалізації.

Список літератури

1. Hammer, Michael & Champy, James. Reengineering the Corporation: Manifesto for Business Revolution. HarperCollins e-books. – 2009. – 272 p.

2. Dumas, Marlon & La Rosa, Marcello & Mendling, Jan & Reijers, Hajo A. Fundamentals of Business Process Management. Springer. – 2013. – 399 p.

3. What is business activity monitoring? [Online]. Available: https://www.ibm.com/support/ knowledgecenter/en/SSBN76_7.0.1/com.ibm.btools.help.monitor.intro.doc/intro/keyconcepts.html.

4. Debra, Paul & Turner, Paul & Cadle, James. Business Analysis Techniques, revised Edition. BCS, The Chartered Institute for IT. – 2014. – 356 p.

5. Паклин Н. Бизнес-аналитика. От данных к знаниям / Н. Паклин, В. Орешков. – СПб.: Питер, 2013. – 704 с.

6. Парамонов С. Что такое Business Intelligence. [Электронный ресурс] Режим доступа: https://habrahabr.ru/company/navicon/blog/250875/.

7. Witten, Ian H. & Frank, Eibe. (). Data Mining: Practical Machine Learning Tools and Techniques, Second Edition. Elsevier Inc. – 2005. – 558 p.

8. Howson, Cindi & Sallam, Rita L. & Richardson, James Laurence & Tapadinhas, Joao & Idoine, Carlie J. & Woodward, Alys. Magic Quadrant for Analytics and Business Intelligence Platforms. [Online]. Available: https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb.

9. Data Scientist, IT Salary. [Online]. Available: https://www.payscale.com/research/ US/Job=Data_Scientist%2c_IT/Salary.

10. Data Engineer Salary. [Online]. Available: https://www.payscale.com/research/US/Job=Data_Engineer/Salary.

11. LeaRning Path on R – Step by Step Guide to Learn Data Science on R. [Online]. Available: https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/learning-path-r-data-science/.

12. Gulalkari, Nupur. Implementing Apriori Algorithm in R. [Online]. Available: https://datascienceplus.com /implementing-apriori-algorithm-in-r /.

13. McKinney W. Python for Data Analysis. O'Reilly. – 2013. – 454 p.

14. sp_execute_external_script (Transact-SQL). [Online]. Available: https://docs.microsoft.com/en-us/sql/relational-databases/system-stored-procedures/sp-execute-external-script-transact-sql.

References

1. Hammer, Michael & Champy, James (2009), *Reengineering the Corporation: Manifesto for Business Revolution*. HarperCollins e-books, 272 p.

2. Dumas, Marlon & La Rosa, Marcello & Mendling, Jan & Reijers, Hajo, A. (2013), *Fundamentals of Business Process Management*, Springer, 399 p.

3. What is business activity monitoring? [Online]. Available: https://www.ibm.com/support/ knowledgecenter/en/SSBN76_7.0.1/com.ibm.btools.help.monitor.intro.doc/intro/keyconcepts.html.

4. Debra, Paul & Turner, Paul & Cadle, James (2014), *Business Analysis Techniques*, revised Edition. BCS, The Chartered Institute for IT, 356 p.

5. Paklin, N. and Oreshkov, N. (2013), *Biznes-analitika. Ot dannykh k znaniyam*, Piter, SPb, 704 p.

6. Paramonov S. Chto takoye Business Intelligence. [Online]. Available: https://habrahabr.ru/ company/navicon/blog/250875/.

7. Witten, Ian H. & Frank, Eibe (2005), Data Mining: Practical Machine Learning Tools and Techniques, Second Edition. Elsevier Inc, 558 p.

8. Howson, Cindi & Sallam, Rita L. & Richardson, James Laurence & Tapadinhas, Joao & Idoine, Carlie J. & Woodward, Alys. Magic Quadrant for Analytics and Business Intelligence Platforms. [Online]. Available: https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb.

9. Data Scientist, IT Salary. [Online]. Available: https://www.payscale.com/research/ US/Job=Data_Scientist%2c_IT/Salary.

10. Data Engineer Salary. [Online]. Available: https://www.payscale.com/research/US/Job=Data_Engineer/Salary.

11. LeaRning Path on R – Step by Step Guide to Learn Data Science on R. [Online]. Available: https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/learning-path-r-data-science/.

12. Gulalkari, Nupur. Implementing Apriori Algorithm in R. [Online]. Available: https://datascienceplus.com /implementing-apriori-algorithm-in-r/.

13. McKinney, W. (2013), Python for Data Analysis. O'Reilly, 454 p.

14. sp_execute_external_script (Transact-SQL). [Online]. Available: https://docs.microsoft.com/en-us/sql/relational-databases/system-stored-procedures/sp-execute-external-script-transact-sql.

Надійшла до редколегії 12.03.2018
Схвалена до друку 17.04.2018

Відомості про автора:

Федько Віктор Васильович
кандидат фізико-математичних наук доцент
доцент кафедри
Харківського національного
економічного університету ім. С. Кузнеця,
Харків, Україна
https://orcid.org/0000-0002-4146-5272

Information about the author:

Viktor Fedko
PhD Associate Professor
Senior Lecturer of Department
of Simon Kuznets Kharkiv National
University of Economics,
Kharkiv, Ukraine
https://orcid.org/0000-0002-4146-5272

АНАЛИЗ ДАННЫХ В SQL SERVER СРЕДСТВАМИ PYTHON

В.В. Федько

Рассмотрены возможности инструментальных средств анализа данных. Изложены способы хранения данных, адаптированные к эффективному выполнению запросов анализа данных, а также языковые средства, представленные в компоненте Microsoft SQL Server как Machine Learning Services (in-database). Проведено сравнение операционных баз данных (OLTP-систем) и хранилищ данных, которые ориентированы на анализ данных (OLAP-систем). Даны примеры обеих систем, а также рассмотрена система их взаимодействия (ETL-система). Описаны инструментальные средства анализа данных, которые в простейших случаях применяются к OLAP-кубам. Представлены языковые средства выполнения анализа данных в более сложных случаях. Проведено сравнение языков R и Python, из которого следует, что язык Python позволяет строить завершенные приложения обработки данных, а библиотеки в нем почти такие же, как и в языке R. Показано, что, учитывая большую популярность языковых средств анализа данных в последние выпуски SQL Server включены компонент SQL Server R Services, в результате чего новые возможности в SQL Server позволили обойти ограничения, которое заключается в том, что все данные должны храниться в памяти. Описаны основные преимущества компонента Machine Learning Services, а также особенности его установки. Продемонстрированы на конкретных примерах возможности выполнения расчетов и графического представления результатов на языке Python в среде SQL Server для проведения анализа данных.

Ключевые слова: *Business intelligence, Data mining, Data Scientist, Data Engineer, SQL Server, Machine Learning Services, операционная база данных, хранилище данных, языки R и Python, анализ данных, визуализация данных.*

DATA ANALYSIS IN SQL SERVER MEANS OF PYTHON

V. Fedko

The possibilities of data analysis tools are considered. Methods of data storage adapted to the effective execution of data analysis requests are described, as well as the language tools presented in the Microsoft SQL Server component as Machine Learning Services (in-database). Comparison of operational databases (OLTP-systems) and data warehouses, which are focused on data analysis (OLAP-systems) are compared. Examples of both systems are given, and the system of their interaction (ETL-system) is considered. Describes data analysis tools, which in the simplest cases are applied to OLAP-cube. Presented are the language tools for performing data analysis in more complex cases. Comparison of the R and Python languages is performed, from which it follows that the Python language allows you to build complete data processing applications, and the libraries in it are almost the same as in the R language. It is shown that, given the great popularity of language analysis tools in the latest issues of SQL Server included the SQL Server R Services component, resulting in new features in SQL Server that circumvented the restriction that all data must be stored in memory. Describes the main advantages of the Machine Learning Services component, as well as the features of its installation. Demonstrated on specific examples of the possibility of performing calculations and graphical representation of results in Python in a SQL Server environment for data analysis.

Keywords: *Business intelligence, Data mining, Data Scientist, Data Engineer, SQL Server, Machine Learning Services, operating database, data warehouse, R and Python languages, data analysis, data visualization.*