

Information Management Technologies for Big Data: A Case of Oracle

Kalinka Kaloyanova

Sofia University, FMI 5, J. Bourchier Blvd. Sofia Bulgaria, kkaloyanova@fmi.uni-sofia.bg

Tsvetomir Hristov

Sofia University, FMI 5, J. Bourchier Blvd. Sofia Bulgaria, injekciona@gmail.com

Ina Naydenova

Technologica 3, Sofiysko pole Str. Sofia Bulgaria, naydenova@gmail.com

Zlatinka Kovacheva

MEC, KOM, P.B. No. 79, Muscat, Oman, PC: 124 Sultanate of Oman zkovacheva@hotmail.com

ABSTRACT

During the last decade the volume, the rate of accumulation and the diversity of data in general have been steadily increasing, which leads to the rapid development of Big data and technological enhancements associated with it. New hardware and software technologies are introduced for information management. This paper addresses a set of Big data solutions that are part of the fundamental technologies offered by Oracle for data storage and analysis. In several examples, a number of opportunities for the deployment of Big data solutions with Hadoop framework and Oracle NoSQL Database have been demonstrated. The main methods for integration of data between Big data platforms and Oracle RDBMS have been presented and a comparison of the performance of the different options has been provided.

The study reveals that Oracle has a large range of solutions to integrate data from Hadoop framework in Oracle Database. Each of the discussed solutions has some positive aspects and drawbacks. Hadoop framework inherently provides solutions for simultaneous processing of large amounts of data that are generally used to data warehouse tasks and build data warehouse of next generation. This partly explains the many different solutions to integrate data from Hadoop framework in an Oracle Database. Unlike the wide range of solutions for Hadoop framework Oracle does not offer many opportunities for NoSQL Database integration with Oracle Database. The only solution that is available for now is Oracle NoSQL Publish Database Utility. The test results show that it is not optimized for operation over a part of data. There is no general rule that defines the best technology for the integration of Big data. The users could choose a solution that is most closely to the business needs.

Технологии за управление на информацията за големи данни на базата на Oracle

Калинка Калоянова

Софийски университет, ФМИ, бул. „Дж. Баучер“ 5, София България, kcaloyanova@fmi.uni-sofia.bg

Цветомир Христов

Софийски университет, ФМИ, бул. „Дж. Баучер“ 5, София България, injekciona@gmail.com

Ина Найденова

Технологика, ул. Софийско поле 3 София България, naydenova@gmail.com

Златинка Ковачева

МЕС, КОМ, Р.В. № 79, Маскат, Оман, ПК: 124 Султанат Оман zkovacheva@hotmail.com

РЕЗЮМЕ

През последното десетилетие обемът, степента на натрупване и разнообразието от данни като цяло непрекъснато нарастват, което води до бързото развитие на големите данни и технологичните подобрения, свързани с това. Въведени са нови хардуерни и софтуерни технологии за управление на информацията. Тази статия разглежда множество от решения за обработката на големи данни, които са част от основните технологии, предлагани от Oracle за съхранение и анализ на данни. В няколко примера са демонстрирани редица възможности за прилагане на решения за големи данни с Hadoop Framework и Oracle NoSQL база данни. Разгледани са основните методи за интегриране на данни между Big Data платформи и Oracle RDBMS и е направено сравнение на представянето на различните опции.

Изследването показва, че Oracle има голям набор от решения за интегриране на данни от Hadoop Framework в Oracle Database. Всяко от разгледаните решения има някои положителни страни и недостатъци. Hadoop Framework предоставя решения за едновременно обработка на големи обеми данни, които обикновено се използват за задачи за съхранение на данни и за изграждане на складове на данни от следващо поколение. Това отчасти обяснява многото различни решения за интегриране на данни от Hadoop Framework в база данни на Oracle. За разлика от широкия набор от решения за Hadoop Framework, Oracle не предлага много възможности за интеграция на NoSQL Database с Oracle Database. Единственото решение, което се предлага засега, е Oracle NoSQL Publish Database Utility. Резултатите от проведените тестове показват, че той не е оптимизиран за работа с част от данните. Няма общо правило, което да определя най-добрата технология за интегриране на големите данни. Потребителите могат да изберат решение, което е най-близко до потребностите на техните конкретни задачи.