وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم علوم الحاسوب

تأثير مفهوم الانجراف على قواعد البيانات المتدفقة لأساتذة قسم علوم الحاسوب

رسالة مقدَّمة
إلى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة ماجستير في
علوم الحاسوب

من قبل

منذر يونس شيت ذنون

إشراف
م.د. عمار ظاهر ياسين طه

١٤٤٤ هـ ٢٠٢٣ م

المُلَخَص

حياتنا لا تتوقف عن التطور والتغيير لذا يجب أن تتكيف أنظمتنا الحاسوبية مع هذا التغيير. فقد لوُحِظ مؤخراً نموٌ وتدفق للبيانات بشكل هائل وغير محدود من مصادر مختلفة ومتنوعة. وأغلب الأنظمة الحالية تتطلب معالجة الأحداث بوقت آني لما للبيانات من أهمية في وقت حدوثها ويجب معالجتها بأسرع وقت وإلا تفقد البيانات أهميتها. ومع كثرة التغيرات التي تحدث على قاعدة البيانات التقليدية من حذف وإضافة وتحديثات مستمرة. نشأة الحاجة إلى أداة تستطيع إدارة تدفق البيانات ومعالجتها في أقصر وقت ممكن.

تهدف الرسالة إلى إنتاج أداة من خلالها يتم إدارة تدفق البيانات ومعالجة الأحداث في الوقت الآني وذلك من خلال استخدام لغة بايثون V3.8 وSQL Sever ، فعند تصميم قاعدة بيانات متدفقة وتنفيذها تتم معالجة تدفق البيانات بالوقت الآني لاتخاذ قرارات مهمة على أساس تدفق تلك البيانات، فضلاً عن كشف الانجراف الحاصل على قاعدة البيانات المتدفقة. تتمتع معظم الأنظمة الحاسوبية الحالية بمعدلات أداء جيدة وقابلية للتوسع، ولكن في المقابل، يتطلب التثبيت والإعداد لتلك الأنظمة مستوىً عالياً من الخبرة، فعند وجود أنظمة تنفذ مفاهيم تدفق البيانات في لغة SQL يعزز جميع مزاياها التقليدية مع إضافة ميزات التدفق.

تسهم الرسالة من خلال النظام المقترح إمكانية التعامل مع تدفق البيانات Data Stream، وتوفير نظام لتحليل البيانات من خلال استعمال خوارزميات كشف الانجراف Concept Drift. كان من الضروري اختيار الأدوات البرمجية بدقة لتحقيق تكامل قاعدة البيانات التقليدية مع معالجة البيانات في الوقت الآني للبيانات المتدفقة. حالة الدراسة لقسم علوم الحاسوب يتم من خلالها بناء نظام قاعدة بيانات متدفقة وذلك باستعمال الأداة المقترحة لاستخراج بيانات أنشطة الكادر التدريسي في القسم.

كشف الانجراف الحاصل على قاعدة البيانات المتدفقة باستخدام ثلاث خوارزميات DDM، EDDM، ADWIN لتحليل البيانات المتدفقة. تم اختبار الخوارزميات لكشف الانجراف الحاصل على قاعدة البيانات المتدفقة، على وفق نتائج المخرجات، تحقق خوارزمية EDDM الأكفأ زمنياً في تحليل البيانات. وتشير النتائج إلى أن الأداة حققت نجاحاً في أداء وظائفها بشكل جيد في إدارة ومعالجة تدفق البيانات بالوقت الآني وتحقيق التكامل مع قاعدة البيانات العلائقية. مع إمكانية تطبيق قاعدة البيانات المتدفقة في كثير من أقسام وكليات جامعة الموصل.

Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Computer Science

Concept Drift Effect in Streaming Database for Computer Science Department Staff

A Thesis Submitted to the Council of the College of

Computer Science and Mathematics

University of Mosul

as a Partial Fulfillment of Requirements

for the Degree of Master of Science

in

Computer Science

By

Munther Yonis Sheet Thanoon

Supervised by

Dr. Ammar Thaher Yaseen Taha

2023 A.D. **1444 A.H.**

Abstract

Our life never stops evolving and changing, and our computer systems must adapt this change. It has been noted recently that there has been an enormous and unlimited flow of data from various and varied sources. Most of the current systems require processing events in real time, because the data are important at the time they occur, and it must be processed as soon as possible, or else the data will lose their importance. With the many changes that occur on the traditional database as deletion, addition and continuous updates, the need arises for a tool that can manage the data stream and process them in the shortest possible time.

This thesis aims to produce a new tool through which the data stream is managed and events are processed in real time in order to achieve an integration between Python V 3.8 and SQL Server, and designing and implementing a streaming database, through which the data stream is processed in real time to make important decisions based on the stream of those data, as well as drift detection obtained on the streaming database. Most computer systems have great performance and scalability, but the installation and the setup require a high level of expertise and involve a very difficult learning curve. Having systems that implement stream concepts in SQL enhances all of its benefits with the addition of stream features.

This thesis contributes to providing a digital archive for the Department of Computer Science and a general study on data stream, the design and implementation of a streaming database. A case study was made for the Department of Computer Science to build a streaming database system, and the proposed tool was contacted to extract data on the activities of the teaching staff in the department. The drift on the streaming database was detected by using three algorithms: DDM, EDDM, and ADWIN to analyze the streaming data. The algorithms were tested to detect drift on the

streaming database, according to the output results, the EDDM algorithm achieves the most efficient execution time in data analysis.

The results show that the tool has achieved a success in performing its functions in a good way in managing and processing data stream in real time and achieving integration with the relational database, with the possibility of applying the streaming database in many departments of the University of Mosul.