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

A Framework for Execution's Robustness of a Blockchain-based Big data application using Container technology

**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 Doctor of Philosophy in
Computer Science**

By

Nawar Abdulghani Sultan Alyas

Supervised by

Asst. Professor Dr. Rawaa Putros Polos Qasha

2024 A.D. **1445 A.H.**

ABSTRACT

In recent years, big data has gained a lot of attention and served as a catalyst for improvements in data processing, storage, and management. Because of this, there was a bigger demand than ever for innovative techniques that can address the changing problems in the big data business. Besides that, blockchain technology emerged and gained widespread adoption, completely altering the landscape of data management and security. However, this technology has several drawbacks, such as complexity, size, scalability, and security flaws. Existing research suggests that container technology can help to address these drawbacks. This is because containerization opens the door for blockchain to be effectively used in big data applications, hence bringing in a new era of reliable, scalable, and efficient data management solutions. To overcome these weaknesses, this thesis introduced the design and development of a Containerized Blockchain-Based Big Data framework to improve the performance, robustness, and resiliency of big data systems based on blockchain.

The presented framework is an integrated outcome of three approaches to ensure service failure recovery, scalable querying techniques, and resource optimization. The first approach, “Big Data Framework for Monitoring Real-Time Vehicular Traffic Flow”, ensures fault tolerance and scalability, making it appropriate for handling massive data volumes generated by numerous connected vehicles. The second approach, “Blockchain-Based Framework for Secure Monitoring of Vehicles Traffic Flow”, ensures data immutability, gains high availability and throughput, offers full-text search, and reduces latency in big data applications based on blockchain. Finally, the Containerization: A Key to Robust Execution of Blockchain-Based Big Data Applications approach, has been developed to improve resource utilization by combining the two aforementioned approaches in a synergistic way in order to decrease latency and the computing power needed to process blocks of data.

The performance of this proposed framework was assessed through several real-world experiments concerning the management of vehicle flow data, focusing on key performance metrics. The results showcased a significant enhancement in system performance, notably in reducing block creation and search times while boosting transaction throughput. Specifically, as the number of nodes increased, transaction throughput rose,

achieving approximately 7500 transactions per second with 10 nodes. Moreover, the framework managed to validate 50,000 transactions in about 47 seconds. Additionally, it demonstrated an impressive capability to search through 20,000,000 transactions in merely 13 milliseconds. These findings underscore the substantial impact of integrating blockchain, big data processing, and containerization within the proposed framework, markedly improving processing speed, scalability, and reliability.

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

اطار عمل لمتانة تنفيذ تطبيق البيانات الضخمة المعتمد على سلسلة الكتل باستخدام تقانة الحاوية

اطروحة مقدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة دكتوراه فلسفة في
علوم الحاسوب

من قبل

نوار عبدالغني سلطان الياس

بإشراف
أ.م.د. رواء بطرس بولص قاشا

١٤٤٥ هـ ٢٠٢٤ م

الخلاصة

في السنوات الأخيرة، اكتسبت البيانات الضخمة الكثير من الاهتمام وكان بمثابة الحافز لاجراء العديد من التحسينات في معالجة البيانات وتخزينها وإدارتها. مما ادى الى زيادة الطلب، من أي وقت مضى، على التقنيات المبتكرة التي يمكنها معالجة المشكلات المتغيرة في الاعمال القائمة على البيانات الضخمة. بالإضافة إلى ذلك، ظهرت تقنية blockchain واكتسبت اعتماداً واسع النطاق، مما أدى إلى تغيير مشهد إدارة البيانات وأمنها بالكامل. ومع ذلك، فإن هذه التكنولوجيا لها عيوب عديدة، مثل التعقيد والحجم وقابلية التوسع والعيوب الأمنية. تشير الأبحاث الحالية إلى أن تكنولوجيا الحاويات يمكن أن تساعد في معالجة هذه العيوب. وذلك لأن النقل بالحاويات يفتح الباب أمام استخدام blockchain بشكل فعال في تطبيقات البيانات الضخمة، وبالتالي جلب حقبة جديدة من حلول إدارة البيانات الموثوقة والقابلة للتطوير والفعالة. للتغلب على هذه العيوب، قدمت هذه الأطروحة تصميم وتطوير لإطار عمل لتطبيقات البيانات الضخمة القائمة على تقنية سلسة الكتل لتحسين أداء وقوة ومرونة هذه التطبيقات ولتعزيز أداء بشكل عام.

يعد الإطار المقدم نتيجة متكاملة لثلاثة أساليب لضمان استئناف التشغيل عند حدوث الفشل او العطل، وتقنيات الاستعلام القابلة للتطوير، وكذلك تحسين استخدام الموارد. النهج الأول، "إطار البيانات الضخمة لرصد تدفق حركة مرور المركبات في الوقت الحقيقي"، يضمن التسامح مع الأخطاء وقابلية التوسع، مما يجعله مناسبًا للتعامل مع كميات هائلة من البيانات الناتجة عن العديد من المركبات المتصلة. أما النهج الثاني، وهو "الإطار القائم على سلسة الكتل للمراقبة الآمنة لتدفق حركة مرور المركبات"، فيضمن ثبات البيانات، ويكتسب توافرًا وإنتاجية عالية، ويوفر بحثًا عن النص الكامل، ويقلل من زمن الوصول في تطبيقات البيانات الضخمة القائمة على سلسة الكتل. أخيرًا، تم تطوير نهج الحاويات: مفتاح التنفيذ القوي لتطبيقات البيانات الضخمة القائمة على سلسة الكتل لتحسين استخدام الموارد من خلال الجمع بين النهجين المذكورين أعلاه بطريقة تآزريه من أجل تقليل زمن الوصول وقوة الحوسبة اللازمة لمعالجة كتل البيانات.

تم تقييم أداء الإطار المقترح من خلال إجراء عدد من التجارب الحقيقية لإدارة بيانات تدفق المركبات من خلال النظر في بعض المقاييس الهامة. كشفت النتائج التي تم الحصول عليها أن الإطار المطور عزز الأداء بشكل كبير من حيث عدد المعاملات، ووقت إنشاء الكتلة، ووقت البحث عن النص الكامل من خلال تحسين إنتاجية المعاملات مع تقليل أوقات إنشاء الكتلة والبحث. في الختام، أدى الإطار المقترح إلى تحسين سرعة المعالجة وقابلية التوسع والموثوقية وأثبت أنه حل قابل للتطبيق للتطبيقات في الوقت الحقيقي حتى مع تزايد حجم البيانات.