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

Cloud-Based Big Data Framework for Cultural Heritage Preservation in Nineveh

**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

Shaimaa Fahad Rashed Abd Albari

Supervised by

Asst. Prof. Dr. Rawaa Putros Qasha

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

Abstract

During the last decades, various aspects of Nineveh's cultural heritage have been destroyed during wars or natural causes; therefore, preserving these valuable heritages has become necessary. With the increased use of the internet, social media networks have become part of people's daily life for publicly sharing information, including their feelings, opinion expression, and knowledge, as well as sharing images, videos, audio and even their locations. With the growth of digitisation and data gathering, digital information processing and archiving for cultural heritage keeps expanding and becoming critical.

This thesis presents a new approach based on cloud computing for managing the multi types of data related to the cultural heritage of Nineveh's areas. It involves using different virtualisation techniques in the cloud, such as virtual machines and Docker containers.

The main aim of this thesis is to design and implement a new framework to support digital archiving, structuring and analysis the data of Nineveh's cultural heritage from different heterogeneous data sources such as social media networks, multimedia collections, digital repositories and archives, and web data services. The implemented framework is characterised to support the preservation of the cultural heritage process, including both tangible and intangible heritage. Different approaches have been implemented to extract data from heterogeneous sources depending on several techniques. Once the data are collected, a number of pre-processing operations are implemented to clean, structure and archive the resulting data in the NoSQL databases. Finally, the archived cleaned information has been queried to be analysed and classified using machine learning algorithms.

This thesis is designed to be implemented using a container-based cloud framework for data collecting and archiving data to support portable and lightweight run-time environments.

The obtained results showed the efficiency of the implemented approaches for the collection of diverse and valuable data about the cultural heritage of Nineveh governorate, where a set of various queries could be performed on the resulting data based on heritage-related keywords to extract the beneficial and personalised information to users based on their preferences and contexts. The developed framework provides a formative reference for the preserving techniques for sustainable protection and development of cultural heritage, which can be utilised by heritage experts and researchers.

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

أطار عمل للبيانات الضخمة المستند الى السحابة الالكترونية للحفاظ على التراث الثقافي في نينوى

رسالة مقدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة ماجستير علوم في
علوم الحاسوب

من قبل
شيماء فهد راشد عبد الباري

بإشراف
أ. م. د. رواء بطرس بولص

١٤٤٤ هـ ٢٠٢٣م

الملخص

خلال العقود الماضية ، تعرض التراث الثقافي في نينوى للتدمير من جوانب مختلفة وذلك لأسباب طبيعية أو في أثناء الحروب ،لذلك أصبح الحفاظ على هذه الموروثات القيمة أمرًا ضروريًا. ومع زيادة مستخدمي الانترنت أصبحت شبكات التواصل الاجتماعي جزءًا اساسياً للحياة اليومية للناس فهي تُستخدم لتبادل المعلومات بصورة عامة ,مثلا تُستخدم للتعبير عن مشاعر الناس وعن آرائهم في قضايا معينة ومعرفتهم بما حولهم، فضلاً عن مشاركة الصور ومقاطع الفيديو والصوت وحتى مواقعهم الجغرافية. مع ظهور الرقمنة وجمع البيانات يتم استخدام المعلومات الرقمية المتوفر عبر الانترنت لأرشفة التراث الثقافي وأصبحت شيئا مهماً. تقدم هذه الرسالة نهجًا جديدًا يعتمد على الحوسبة السحابية لإدارة أنواع متعددة من البيانات المتعلقة بالتراث الثقافي لمناطق نينوى. حيث تتضمن استخدام تقنيات المحاكاة الافتراضية المختلفة في السحابة ، مثل الأجهزة الافتراضية وحاويات دوكر. الهدف الرئيسي من هذا المشروع هو تصميم وتنفيذ إطار عمل جديد للمساهمة في الأرشيف الرقمي, حيث يقوم بجمع وهيكله وتحليل البيانات الخاصة بالتراث الثقافي لنينوى بالاعتماد على مصادر مختلفة للبيانات مثل شبكات التواصل الاجتماعي ومجموعات الوسائط المتعددة والمستودعات الرقمية والأرشيف الإلكتروني وخدمات بيانات الويب. إن الإطار الذي تم تصميمه مصمم للمساهمة في عملية حفظ التراث الثقافي بنوعيه المادي و الغير المادي. تم تنفيذ طرق مختلفة لاستخراج البيانات من المصادر المتنوعة اعتمادًا على عدة تقنيات. بمجرد جمع البيانات ، يتم تنفيذ عدد من عمليات المعالجة لتنظيف وهيكله وأرشفه البيانات الناتجة في قواعد بيانات غير علائقية. وأخيرًا ، يتم الاستعلام عن المعلومات النظيفة المؤرشفة ليتم تحليلها وتصنيفها باستخدام خوارزميات التعلم الآلي. تم تصميم هذا المشروع ليتم تنفيذه باستخدام إطار عمل سحابي قائم على الحاوية لجمع البيانات وأرشفتها , الأطار المقترح يدعم العمل في بيئات لأنظمة تشغيل مختلفة بحيث يكون محمولاً و خفيف الوزن. أظهرت نتائج العمل كفاءة الأساليب المنفذة في جمع البيانات المتنوعة والقيمة حول التراث الثقافي لمحافظة نينوى ، حيث يمكن إجراء استعلامات مختلفة على البيانات بناءً على الكلمات المفتاحية المتعلقة بالتراث لاستخراج معلومات قيمة ومفيدة للمستخدمين بناءً على تفضيلاتهم سياقاتهم. يوفر الإطار الجديد مرجعًا للحماية الرقمية وللمساهمة في الاستدامة الرقمية للتراث الثقافي في نينوى والتي يمكن استخدامها من قبل خبراء التراث والباحثين.