

Abstract

Providing primary health-care services remotely, in rural areas, particularly in developing countries, remains a significant challenge due to lack of inadequate infrastructure, high costs, lack of skilled personnel, and poor connectivity. Rural populations in these countries are mostly deprived of quality health-care services. To address this issue, it is required to implement innovative solutions made of modern technologies like mobile health (mHealth), cloud computing, sensor networks, and Big Data analytics. Several research have explored the use of these technologies in health-care domain to bridge the gaps between rural patients and health-care providers, ensuring that health-care is more accessible and affordable.

Kiosk-based remote health-care system is one important solution to overcome the issues. Kiosks are set up in rural villages, where local health-care providers use a touch-screen-based device and a mobile application to collect patients' vital signs and symptoms using a knowledge base which is created by doctors. The collected data is transmitted to cloud servers, allowing urban doctors to remotely access the information, prescribe medication, and recommend medical tests. In this research work, one such application is developed. Kiosks have been set up and the application has been deployed in multiple villages in West Bengal, India. A survey conducted after two years of operation showed that such systems could substantially improve health-care delivery in rural regions. The details of the implementation and the salient features of the application are presented in this thesis.

It has been observed that Internet connectivity is poor in the rural villages and in remote areas. In order to come up with a solution to this problem an SMS-based connectivity module is implemented to run the application at the time of problems with Internet connectivity. In addition to rural health-care services, the connectivity issue has also been dealt with for disaster sites where no network infrastructure exists. In such situations, the increasing proliferation of mobile devices has led to the exploration of Mobile Ad-hoc Networks (MANETs) as a framework for delivering real-time multimedia health-care services. MANETs allow transmission of multimedia data over a network of mobile devices without relying on a centralized infrastructured network. Despite challenges like network instability and packet loss, techniques such as cross-layered fragmentation have been used to improve real-time video and image transmission. This is particularly useful for carrying out rescue

operations in disaster sites and in remote health-care, where live video consultations between rural patients and urban doctors can enhance diagnosis and treatment.

The primary objective of this thesis is to explore the issues related to storage and management of health-care data. Considering that the health data has the properties of Big Data, the role of Big Data tools in health-care is crucial, particularly for storage and processing of the large amounts of data generated from various health-care devices along with the kiosks. Several Big Data solutions, including Hadoop, Cassandra, MongoDB, Neo4J, Hive, HBase have been examined for their applicability in health-care. Most of the tools designed for distributed storage and processing capabilities, to make them well-suited for handling massive health data such as Electronic Health Records (EHR). However, efficient mapping of health data onto these platforms requires careful consideration of data models and query structures. Comparative studies between Hadoop, Hive, HBase, Cassandra, MongoDB, Neo4j demonstrate their strengths and weaknesses in handling health-care queries, providing valuable insights into the selection of appropriate Big Data tools for health applications. While performing the experiment with the tools, some of the issues was recorded. To handle these issues, Node-Guided MapReduce concept was introduced.

The Node-Guided MapReduce (NGMR) framework represents a further advancement in Big Data processing through MapReduce. Traditional MapReduce implementations suffer from inefficiencies in execution of queries due to the non-selective involvement of nodes in the execution process, leading to increased execution time. NGMR addresses this issue through selective node selection and execution, thereby reducing query execution time and improves the overall performance of the system. The framework has been implemented and tested in small-scale environments using low-cost computational resources, offering a scalable solution for large health-care data systems.

The comparative analysis of various Big Data platforms such as Hadoop, HIVE, Cassandra, MongoDB, HBase, and NGMR has been done on health-care data, for health-care applications to determine the effectiveness of these tools in terms of performance in storing and query execution related to health data. The findings emphasize that, while NoSQL databases like Cassandra and MongoDB offer flexibility and scalability, the choice of platform should be driven by the specific needs of the health-care application, such as the volume of data and the type of queries being executed. On the other hand NGMR performs quite well in query processing.

Finally, an algorithm for automated node selection in NGMR has been proposed, which enables the system to efficiently guide queries to execute on an optimal set of nodes. To validate the performance of this node selection algorithm, a simulation set up using micro-services on AWS has been used.

The thesis concludes with the limitation of the research works presented in this thesis, outstanding issues and a direction towards the future work.