Development of national health data warehouse framework for Bangladesh

Shahidul Islam Khan
Department of Computer Science and Engineering
International Islamic University Chittagong (IIUC), Bangladesh

Abstract
Data Warehouse (DW) integrates data from two or more sources into a repository for reporting, analysis, and knowledge discovery. This technology can be used in healthcare to develop National Health Data Warehouse (NHDW) that combines patient information from clinics, hospitals, laboratories, diagnostic centers, etc. In this research, we propose a framework for NHDW based on studying different global frameworks and conducting surveys among the potential stakeholders of the framework. Here, patient data will be collected in two phases: physical data collection and data collection over the internet. We did surveys on different healthcare organizations among patients, medical staff, and doctors. The survey questionnaire was divided into four sections: personal information, information privacy and security, performance, and usability to know the requirement of the stakeholders. As healthcare data is very sensitive and can be misused, providing privacy for these patients’ information is properly addressed in this framework. The NHDW framework may significantly improve our healthcare services in Bangladesh.

Keywords Healthcare data warehouse, Knowledge discovery, Medical records, Privacy

Paper type Research paper

1. Introduction
The goal of a Data Warehouse (DW) is to combine data from the heterogeneous source and later use it for analytics and decision-making purpose. Health data from a Health Data Warehouse (HDW) can be used to identify patients, and analysis of disease trends, decisions making, and knowledge discovery (Mawelmada, 2011; Razi & Samani, 2014). During data integration from different sources, DW faces some issues and challenges (Razi & Samani, 2014; Islam, & Biswas, 2014). The component of the data warehouses is source data, data storage, OLAP (Online analytical processing) server, and decision-making tools. To present the multidimensional view of the data star schema has been presented in the data warehouse architecture (Pubudika, 2011; Khan & Hoque, 2016).

ETL (Extraction-Transformation-Load) process is
applied to extract data from different sources, transform the data into a standard model for analysis and mining, identify data quality problems, clean data that remove noise from data, and finally load into the DW. For the development of the DW, the ETL process is very important in the medical field to make the data for further analysis and decision-making that will provide complete patient information to the users. ETL process consists of four steps: Extraction, Transformation, Loading, and Refresh (Khan & Hoque, 2015; Ralph & Joe, 2004; Sheta & Eldeen, 2012; Mia, Hoque, Khan, & Ahmed 2022). According to Lu, Wu, Liu, Chen, and Guo (2013) in medical information ensuring the security and privacy of patient data is very important. Different types of encryption techniques and protection schemes are used to secure patient information. The development of the DW framework depends on the complexity of data warehouse architecture, Security, Privacy, healthcare quality, and cost (AlJarullah & El-Masri, 2013; Rainardi, 2008).

In this paper, we have presented a framework for the National Health Data Warehouse (NHDW) for Bangladesh. We have studied the different healthcare frameworks of the USA, UK, Canada, etc. We have also studied the different healthcare service-providing systems, currently used in Bangladesh such as Village Vaccine Center, Nagor Shastho Kendro, General Hospitals, and Diagnostic Centers. We have also conducted a survey among 84 patients, 62 doctors, and 48 medical staff (a total of 194) to find out their expectations from National Health Data Warehouse (NHDW), Bangladesh. We have analyzed this survey data to fine-tune different components of our proposed framework.

This paper is organized as follows. In chapter II, we have presented a summary of different healthcare frameworks proposed and used in modern countries. We have depicted the current healthcare service scenario in Bangladesh in Section III. Then in Section IV, we presented the survey questionnaire and different analyses from the survey we conducted. Section V presents our proposed NHDW framework. The discussion and limitation section is presented in Section VI. Finally, we presented the conclusions in Section VII.

2. Related work
This section gives an overview of the preceding work attempts on the Healthcare Data Warehouse framework that are pertinent to the work in this study. Healthcare organizations contain lots of patients’ health data but it is very important to keep these data in a central repository to ensure the needs of all clinical users and support medical decision-making (Khan & Hoque,
Development of national health data warehouse 55

2016; Razi & Samani, 2014). The semi-centralized architecture stores a
summarized copy of the patient’s EHR (Electronic Health Records) and a
reference link of each healthcare organization in a nationwide central
repository (AlJarullah & El-Masri, 2013). National Health Data Warehouse
(NHDW) architecture presents a suitable model to integrate health data from
different healthcare sources. The Manitoba Centre for Health Policy
(MCHP) manages 25 years of health data from the University of Manitoba
using the population-based registry and available data for research purposes
(Smith, et al. 2016).

The Data cube is used to show the level of summarization and a star
schema is developed to represent the multidimensional view of data
(Pubudika, 2011). Patient identification or linking patients to their health data
across different healthcare organizations is completed with NPID (National
Patient Identifiers). In USA Europe, and Canada, they use SSN (Social
Security Number) and Birth Certificate Number respectively to identify their
patient (Fair, Cyr, Allen, Wen, Guyon, & MacDonald, 2000; Riplinger,

In medical fields, patients’ health information contains a huge amount of
sensitive data so it is very important to ensure the security and privacy of
these data but traditional privacy algorithms cannot prevent data from
internal staff from being disclosed. A patient privacy protection scheme for
medical information systems proposed an effective scheme that can prevent
data from internal staff (Lu, Wu, Liu, Chen, & Guo, 2013). Privacy-preserving
index for encrypted electronic medical records presents a
scheme called P-index which secures against the adaptively chosen keyword
attack. Protecting privacy during the peer-to-peer exchange of documents is
a peer-to-peer interoperable system that controls the exchange of the EMR
(Electronic Medical Records) in an SHR (Shared Health Records) between
healthcare organizations and clinicians (Fair, Cyr, Allen, Wen, Guyon, &
MacDonald, 2000; Weber-Jahnke & Obry, 2012). Mia, Hoque, Khan, and
Ahamed (2022) presented architecture and analysis for privacy-preserved
national clinical data warehouse (NCDW). They also elaborate some
important use-cases of NCDW.

3. Current healthcare service scenario in Bangladesh
For the development of the NHDW framework for Bangladesh, we have
studied the current healthcare system in Bangladesh. We visited different
healthcare providing centers in Bangladesh from village to city level. During
the visit, we mainly focused on the present healthcare system of these
healthcare organizations including the Village vaccine center, Nagor Shastho
Kendra, General hospital, Government hospital, and Diagnostic center. The findings are summarized below:

**Village vaccine center**
In the village, the health care system is not well-equipped and only some vaccine centers or NGO health organizations are available here. So, the village people can go there for their primary treatment. Figure 1 shows the healthcare system for the village vaccine center.

![Figure 1](image1.png)
*Healthcare system of village vaccine center*

**Nagor shastho kendro**

![Figure 2](image2.png)
*Healthcare system of Nagor shastho kendro*
Development of national health data warehouse

In Bangladesh, every city has some health clinics where women or babies can take treatment like a vaccine or primary treatment shown in Figure 2.

**General hospital**
The General Hospital of any city gives better healthcare than the health clinic or NGO health. A patient can get many kinds of healthcare and test from this kind of hospital. The healthcare system of the General Hospital is shown in Figure 3.

**Government hospital**

[Diagram of Government Hospital]

**Figure 4**
Healthcare system of Government Hospital
In Government hospitals and medical colleges, a patient can take every kind of treatment and most importantly the emergency care shown in Figure 4.

4. **Survey data analysis and result**

We did surveys among patients, medical staff, and doctors from different hospitals in Bangladesh using an individually printed questionnaire. The total number of participants was 194. Table 1 shows the total attendance of patients, medical staff, and doctors with gender information.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Doctor</th>
<th>Medical Staff</th>
<th>Doctor</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>84</td>
<td>48</td>
<td></td>
<td>62</td>
<td>194</td>
</tr>
<tr>
<td>Male (54)</td>
<td>Male (22)</td>
<td>Male (38)</td>
<td></td>
<td>114 (Male)</td>
</tr>
<tr>
<td>Female (30)</td>
<td>Female (26)</td>
<td>Female (24)</td>
<td></td>
<td>80 (Female)</td>
</tr>
</tbody>
</table>

We assigned the questionnaire corresponding to five numeric values (1-5) for our analysis purposes. These survey questionnaires for the participant are divided into four sections: Personal information, Information privacy and security, Performance, and Usability. In this report first of all we are going to present the question of the survey then the details with used options correspondent numeric values and finally the analysis table of four sections of the participant response. Table 2 shows the options with correspondent values and the short term of the survey questionnaire.

<table>
<thead>
<tr>
<th>Options</th>
<th>Short term</th>
<th>values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree/poor/Not at all/ No/ 1-2 minutes</td>
<td>SD/P/NA/NO/1-2M</td>
<td>1</td>
</tr>
<tr>
<td>Disagree/Less than satisfactory/Little/Not sure/3-4minutes</td>
<td>D/LTS/L/NS/3-4M</td>
<td>2</td>
</tr>
<tr>
<td>Neutral/Satisfactory/Less satisfactory/Average/Maybe/5minutes/Medium</td>
<td>N/S/LS/AVG/MB/5M/MD</td>
<td>3</td>
</tr>
<tr>
<td>Agree/Good/Yes/Highly/6-10minutes</td>
<td>A/G/Y/H/6-10M</td>
<td>4</td>
</tr>
<tr>
<td>Strongly agree/Very good/Definitely/Extremely</td>
<td>SA/VG/DF/E/10+M</td>
<td>5</td>
</tr>
</tbody>
</table>

**Questionnaire for patient/medical staff/doctor**

Few questions are common for all of them. Few are only applicable to patients, and few are for doctors and staff. The questions are numbered as follows: Patient's Section A, B, C, and D questions are numbered as PA, PB, PC, and PD respectively. Doctor's Section A, B, C, and D questions are
numbered as DA, DB, DC, and DD respectively. Medical Staffs Section A, B, C, and D questions are numbered as MA, MB, MC, and MD respectively. Please refer to Table II for other abbreviations.

Section A - Personal information
PA 1. Your gender ?
PA 2. How old are you ?
PA 3. How long have you been taking health care from this hospital ?
PA 4. What about your last educational qualification ?

Section B – Information Privacy and Security
PB 1. Do you think that your health information should be kept secret ?
PB 2. Who should use your health information to be controlled with your permission ?
PB 3. Do you think the current safety measures for storing your health information are adequate in the hospital ?
PB 4. For better privacy and security, the patient information needs to store on computers (Excel, Word, Database...).

Section C - Performance
PC 1. How do you feel about the current way doctors give handwritten prescriptions ?
PC 2 & MC 2 & DC 1. In your opinion, current patient data storage is used for the same patient for further treatment in current health care ?
PC 3. Do you feel comfortable using a computer to get health care ?
PC 4 & MC 4 & DC 4. Will the quality of health care provided by hospitals increase if an online and computer-based system is implemented through the Internet ?
MC 2. Do you feel comfortable using a computer for storing patient data ?
DC 2. Do you feel comfortable using a computer for patient care ?

Section D - Usability
PD 1. Will it be more understandable for you if a doctor provides the prescription and information on the computer ?
PD 2 & MD 4 & DD 4. Do you think it will be good to open a health service document for each patient using the National ID / Birth Certificate Number ?
PD 3 & MD 5 & DD 5. To make health information more accessible, the government should invest more in health care.
PD 4 & MD 2 & DD 2. Do you think the patient's health information needs to store on a computer for further treatment and research purposes ?
PD 5 & MD 8 & DD 8. Is it important to develop a national healthcare data
warehouse for all the patients in Bangladesh?
MD 1 & DD 1. Do you think, currently the amount of time spent on storing patient information is less?
MD3 & DD 3. Do you think that storing patient information on a computer can take a longer time than the current system?
MD6 & DD 6. If the time spent on the online system is more than the current system, then how much online system would be acceptable to the physicians?
MD7 & DD 7. In your hospital, how much time, on average, a doctor can spend for a patient to store his/her demographic and health information on a computer?

Below we analyze the results of the survey data in three sections information privacy and security, performance, and usability. The next table 3 shows the results of the information privacy and security-related survey question that the participants' responses of present privacy and security and what they want regarding privacy and security for the patient health information of NHDW, Bangladesh.

Table 3
Information privacy and security

<table>
<thead>
<tr>
<th>Survey Q. No</th>
<th>PA(%)</th>
<th>MS(%)</th>
<th>DO(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB1 &amp; MB1 &amp; DB1</td>
<td>SD(28.6)</td>
<td>SA(33.3)</td>
<td>SA(61.3)</td>
</tr>
<tr>
<td>PB2</td>
<td>A(26.2)</td>
<td>A(41.7)</td>
<td>A(58)</td>
</tr>
<tr>
<td>PB3</td>
<td>D(31.0)</td>
<td>A(45.8)</td>
<td>NU(38.7)</td>
</tr>
<tr>
<td>PB4</td>
<td>A(35.7)</td>
<td>A(45.8)</td>
<td>A(51.6)</td>
</tr>
</tbody>
</table>

Information privacy and security
Privacy and security are very important issues in the medical sector. Table 3 shows most of the participants (medical staff and doctors) strongly agree that patient health information should be kept secret. But only 28.6% of patients strongly disagreed about it, they think that their health information shouldn’t keep secret. In our further analysis of patients’ personal information, we come to know that those who responded strongly disagree that most of the patients age is above 40 years old and they lack knowledge about the security and privacy of health data. So, this is the reason they don’t think patient information should keep secret.

Although, patients thought that health information shouldn’t keep secret other participants (medical staff and doctors) agree that patient information sharing should be controlled by the patient.
Of the medical staff, almost 45.8% agree that the current safety measures for storing patients’ health information are adequate in the hospital but patient and doctor responses are disagreed and neutral because they are not used to storing the health information of the patient.

For better privacy and security, almost every participant agreed that patient information needs to store on the computer (Excel, Word, Database, etc.).

**Performance**

The results of performance-related survey questions of recent healthcare organizations and what should be the future performance of the healthcare organizations are shown in Table 4

<table>
<thead>
<tr>
<th>Survey Q. No</th>
<th>PA(%)</th>
<th>MS(%)</th>
<th>DO(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>P(35.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC2 &amp; MC1 &amp; DC1</td>
<td>S(35.7)</td>
<td>S(29.2)</td>
<td>S(32.3)</td>
</tr>
<tr>
<td>PC3</td>
<td>VG(45.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC4 &amp; MC3 &amp; DC3</td>
<td>DF(47.6)</td>
<td>Y(45.8)</td>
<td>DF(41.9)</td>
</tr>
<tr>
<td>MC2</td>
<td></td>
<td>VG(45.8)</td>
<td></td>
</tr>
<tr>
<td>DC2</td>
<td></td>
<td></td>
<td>G(38.7)</td>
</tr>
</tbody>
</table>

In Table 4, we see that 35.7% of the patient feels uncomfortable about the current way doctors give handwritten prescription so it is important to develop an e-prescription so that doctor can give computer-typed prescription.

All the participants are sure that patient data storage is used for the same patient for further treatment in current healthcare.

Almost every participant is positive about using a computer for patient care where 45.2% of patients will feel very good about using a computer to get healthcare, 45.8% of medical staff will feel comfortable using a computer to store patient data and 38.7% of doctors will feel comfortable using a computer for patient care.

The response of participants is above 45% (definitely and yes) that the quality of healthcare provided by the hospitals will increase if an online and computer-based system is implemented through the internet. The results of the usability-related survey questions on the recent health data of the patient and what would be the future usability of these data according to the NHDW framework are shown in Table 5.
Usability

Table 5

<table>
<thead>
<tr>
<th>Survey Q. No</th>
<th>PA(%)</th>
<th>MS(%)</th>
<th>DO(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD1 A(40.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD2 &amp; MD4 &amp; DD4</td>
<td>A(38.1)</td>
<td>SA(50)</td>
<td>A(41.9)</td>
</tr>
<tr>
<td>PD3 &amp; MD5 &amp; DD5</td>
<td>SA(59.5)</td>
<td>SA(75)</td>
<td>SA(51.6)</td>
</tr>
<tr>
<td>PD4 &amp; MD2 &amp; DD2</td>
<td>SA(45.2)</td>
<td>SA(45.8)</td>
<td>SA(96.8)</td>
</tr>
<tr>
<td>PD5 &amp; MD8 &amp; DD8</td>
<td>SA(54.8)</td>
<td>E(54.2)</td>
<td>E(58.1)</td>
</tr>
<tr>
<td>MD1 &amp; DD1</td>
<td>D(45.8)</td>
<td></td>
<td>A(38.7)</td>
</tr>
<tr>
<td>MD3 &amp; DD3</td>
<td>SD(33.3)</td>
<td>D(45.2)</td>
<td></td>
</tr>
<tr>
<td>MD6 &amp; DD6</td>
<td>AVG(29.2)</td>
<td>AVG(45.2)</td>
<td></td>
</tr>
<tr>
<td>MD7 &amp; DD7</td>
<td>5M(29.2)</td>
<td>6-10M(29)</td>
<td></td>
</tr>
</tbody>
</table>

In Table 5, we can see participants agrees that it will be more understandable for a patient if the doctor provides the prescription and information on a computer.

For this purpose, most of the participants strongly agree to open a health service document for each patient using the National ID/ Birth Certificate number and health information needed to store in the computer for further treatment and research purposes. So, participants want an NHDW framework for Bangladesh and the government should invest more in the health sector.

Medical staffs 45.8% think that current the amount of time spend on storing patient information is high so they disagree that storing patient information on a computer can take longer time than the current system. If the time spent on an online system is more than the current system, then the acceptance of the online system will be challenging for medical staff and physicians. Finally, the medical staff and doctor will spend 5-10 minutes on average for a patient to store his/her demographic and health information on a computer.

5. Proposed architecture of NHDW framework

The architecture of the NHDW framework is shown in Figure 5 In this framework data collection will be held in two sections: Physically data collection and data collection over the internet.

Physically data collection for health data from the Village health center, NGO and Union health, etc. to the Upazilla health complex because the village and union level of our country has fewer facilities for internet, electricity, and technology. So, data will be collected by a health employee for
sading data to the Upazila health complex. Then the clinician or staff will upload these patient data into the NHDW server because in our country Upazilla has more facilities of internet, electricity, and technology than the village or union level.

In the city healthcare organizations like government medical, private hospitals, diagnostic centers, laboratories, etc. will collect data through the internet where technologies (internet, electricity) are available.

After the data collection, health data will be integrated into an NHDW repository using the ETL (Extraction-Transform-Load) process. The data are prepared to use in healthcare, health data analysis, and knowledge recovery.

Finally, from our survey, we know that participants feel important to start a National Healthcare Data Warehouse (NHDW) framework for all patients in Bangladesh.

Patient data entry for Upazilla health complex
In recent health care, the healthcare organization doesn’t maintain any format to enter patient data. So, in Table 6 we propose a format for patient data entry in the Upazilla health complex that every healthcare provider will maintain this format. In this result, no complexity will arise during data upload to the NHDW server, and filling up this format will take less than 2 minutes which will reduce time.
Table 6
Patient Data entry for Upazilla health complex

<table>
<thead>
<tr>
<th>No</th>
<th>NID/BC no</th>
<th>NAME</th>
<th>AGE</th>
<th>DISEASE</th>
<th>MEDICINE</th>
<th>DOSES</th>
<th>DATE</th>
<th>SIGNATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>XXX</td>
<td>Mr.X</td>
<td>23</td>
<td>Fever</td>
<td>Napa</td>
<td>1+0+1</td>
<td>17-1-1</td>
<td></td>
</tr>
</tbody>
</table>

**Patient identification**

Patient identification is very important in the medical sector because patient health information contains a lot of private data of the patient. So, it is essential to have a unique identification of the patient with any unique number.

![Diagram](image)

**Figure 6**
Patient identification

In Semi-centralized, NPID (National Patient Identifier) is used for patient identification because NPID is a unique number for each patient (AlJarullah & El-Masri, 2013). USA use SSN (Social security number) for patient identification and to get the SSN they need a Birth Certificate so after a new baby is born a Birth certificate provides for the baby (Riplinger, Piera-Jiménez, & Dooling, 2020). Canada uses Birth Certificates to identify the patient and they provide a Health Insurance Card Number which use to get the Birth Certificate (Fair, Cyr, Allen, Wen, Guyon, & MacDonald, 2000).

In our country, patient identification can be done with National Identification (NID) no or Birth Certificate (BC) no. Because in our survey we come to know that most of the participants think that it will be acceptable to open a health service document for each patient using National ID/ Birth certificate no. So, for the identification we have to use one of them as shown in Figure 6, if a patient is above 18 years they will use National Identification No and if the age is under 18 years they will use the Birth certificate no.
NHDW data security technique

In the security and privacy section, we showed how patients or users will keep secure their individual health data but in the medical field, there are a huge amount of health data will be present in a central repository so it’s very important to provide proper security of the data stored in NHDW (Khan & Hoque, 2019). For this reason, we used public-key cryptography (asymmetrical cryptography) to keep secure the data during the communication between sender and receiver.

![Key generation program](image)

**Figure 7**

*Key generation program*

Figure 7 shows that public-key cryptography uses public and private keys generated by the key generation program (Chen, Horng, Lin, & Chen, 2013). Generally, the public key is accessible by all, and the private key is known only by the recipient.

![NHDW data security technique](image)

**Figure 8**

*NHDW data security technique*
From the healthcare organization, the sender will send encrypted data where the original data is encrypted using the receiver’s public key and data can be decrypted only by a private key that is known to the receiver. Although anybody can encrypt the data nobody can decrypt the original data without the private key.

6. Discussions and limitations
In this paper we have presented a framework for the National Health Data Warehouse (NHDW) for Bangladesh based on three things: i) studying the available frameworks abroad ii) studying current healthcare systems in Bangladesh and iii) conducting a survey among the stakeholders. We have analyzed this survey data to fine-tune different components of the NHDW framework. The main challenge we have faced is the reluctance of the participants, the majority of whom were patients. We approached thousands of people for a survey but most of them did not participate. Few medical staff did not participate in the survey thinking it might affect their job. Many doctors were not able to participate because of their extreme work pressure. If we could collect more data, then more insight could be found. Another challenge was the ignorance and illiteracy of some patients regarding the new terms such as Network, Privacy, Encryption, etc. which make our task time-consuming. As the field is vast and very important, more studies regarding the technical aspects of the framework such as warehouse schema architecture, data storage, and standardizations are also required. We hope, our framework will contribute to improving the quality of the healthcare service in Bangladesh if the Government policymakers take the necessary steps to implement it.

7. Conclusion
In this paper, we have proposed the National Healthcare Data Warehouse (NHDW) framework for Bangladesh. The aim is to integrate the patient data from the different types of healthcare organizations and provide an easily accessible, complete patient history all over the country. If we consider the current infrastructure of healthcare in Bangladesh, the situation is not too well in the rural area or union level. Most of the village people, as well as the healthcare providers, are not well educated they lack internet facilities, incapability of operating a computer, and use traditional medical information systems in healthcare organizations. We have studied different healthcare frameworks of modern countries that provide better healthcare. We have summarized the services of different health centers in Bangladesh such as Village Vaccine Center, Nagor Shastho Kendro, General Hospitals, and
Diagnosis. We have also conducted a survey among 84 patients, 62 doctors, and 48 medical staff (a total of 194) to find out their experience in current health systems in Bangladesh and their expectation from National Health Data Warehouse (NHDW), Bangladesh. We have analyzed this survey data to fine-tune different components of our proposed framework. The government should invest more and appoint experts in the medical field and make opportunities for people to learn about technology and improve healthcare. National identification (NID) and Birth certificate (BC) numbers should be introduced in storing health data to identify patients with uniqueness. Privacy and security techniques need to be ensured to protect patient information from unauthorized access.

References


Corresponding author
Shahidul Islam Khan can be contacted at: nayeemkh@gmail.com, shahid@iiuc.ac.bd