Journal of Surgery and Trauma 2025; 13(2):63-74.
jsurgery.bums.ac.ir

ORIGINAL ARTICLE

Registry Event Log Warehouse: Toward Designing A Care Process Digital Twin

Mohammad Hossein Roozbahani[1]✉, Mansoureh Yari Eili[2], Mahdi Sharif-Alhoseini[3], Fatemeh HajiAliAsgari[4]

[1] *Department of Nanotechnology, Faculty of Advanced Technology, Iran University of Science and Technology, Tehran, Iran*

[2] *Department of Computer Engineering and IT, Faculty of Technology and Engineering, University of Qom, Iran*

[3] *Department of Neurotrauma, Sina Trauma and Surgery Research Center, Tehran University of Medical Sciences, Tehran, Iran*

[4] *Department of E-health, Tehran university of medical science, Tehran, Iran*

Received: January 09, 2025 ***Revised: April 09, 2025*** ***Accepted: April 20, 2025***

Abstract

Introduction: Many data warehouses (DWs) developed for the healthcare field only represent subject-oriented data without addressing the control flow. Each process includes many ordered specific activities, which cannot be represented by DWs. This limitation, next to the need to find correlations between case and activity attributes in our health data, motivates the extension of a Registry Event Log Warehouse (REWH). Since the disease registry is saturated with valuable, critical data on human life and supports clinicians and professionals in their decision-making and patient monitoring, the present study aimed to develop a twin digital for trauma care as part of our overall work. Because the process warehouse is the first step in designing the trauma warehouse digital twin, the focus is developing a registry process warehouse.

Methods: The methodology comprises the following steps: 1) data gathering, 2) data preprocessing, 3) data entry to cube, and 4) conducting queries. The primary contribution of this work is to demonstrate how to gather heterogeneous data extracted from different sources into a single process-oriented repository and express online analytical processing (OLAP) queries through SQL to promote data filtering and aggregation.

Results: A top-bottom dimensional schemas of a process warehouse that support detailed process analysis, with each level represented by a specific granularity level.

Conclusion: Due to its ability to provide data of interest for all components of the digital twins, the DW is one of the primary components of healthcare digital twins. The multilevel analysis presented in this study, utilizing OLAP tools, can assist easy access to specific information from process warehouse dimensions almost instantly.

Key words: Data Warehousing, Digital Health, Disease registries, Health Information Systems, Machine Learning

Introduction

Trauma stands as a major contributor to mortality rates and ED admissions, with over 90% of injury-related cases disproportionately affecting low- and middle-income countries, posing significant challenges to healthcare systems (1). Although several studies report on the development of machine learning algorithms for trauma outcome evaluation, none have comprehensively described the temporally-sequenced care pattern of trauma treatment by multidimensional modeling through

@2025Journal of Surgery and Trauma
Tel: +985632381214
Fax: +985632440488
Po Box 97175-379
Email: jsurgery@bums.ac.ir

✉**Correspondence to:**
Mohammad Hossein Roozbahani, Department of Nanotechnology, Faculty of Advanced Technology, Iran University of Science and Technology, Tehran, Iran
Telephone Number: +982173225849
Email: Roozbahani@iust.ac.ir

[Downloaded from jsurgery.bums.ac.ir on 2026-11-25]

[DOI: 10.61186/jsurgtrauma.13.2.63]

enriching the contextual information and event data warehousing. This analytical tool is critical for informing clinical decision-making.

The need to support organizations in information analysis and decision-making led to the initiation of the data warehousing concept in the mid-1980s (1). According to (2), a data warehouse (DW) consists of a subject-oriented, integrated, time-variant, and non-volatile component that supports the management decision-making process. DW consolidates information from different sources in a multi-dimensional structure, allowing analyses from different perspectives at different levels of granularity (3). Nevertheless, when the integration of process data and operational data is of concern, DWs are often less active (4). To optimize care processes, hospital management can follow procedures suggested by business analysts, such as eliminating unnecessary medical imaging activities, thereby reducing workload and treatment costs. Addressing this issue requires an integrated global analysis tool that ensures both process data and additional patient information are considered.

According to (2), DW consists of a subject-oriented, integrated, time-variant, and non-volatile component that supports the management decision-making process. DWs integrate analytical data from various sources in a multi-dimensional form, allowing analyses from different perspectives at varying levels of granularity (3). However, when it comes to integrating process data and operational data, DWs are often less active (4). To optimize care processes, hospital management can follow procedures suggested by business analysts, such as eliminating unnecessary medical imaging activities, thereby reducing workload and treatment costs. Addressing this issue requires an integrated global analysis tool that ensures both process data and additional patient information are considered.

A practical DW applied in the business processes analysis is known as the Process Warehouse (PW), defined as "a DW which stores histories of engineering processes and products for experience reuse and provides situated process support" (4). PW is an appropriate tool for identifying optimization potential and analyzing business process performance by applying DW technology and online analytical processing (OLAP) tools (5). These tools facilitate the definition, computation, and monitoring of key performance indicators across various aspects. According to (6), the typical dimensions in PWs consist of process, time, resource, and location.

Digital twin data analysis enables the creation of complex, dynamic, and real-time virtual representations of physical entities. Data warehousing technologies like Amazon Redshift and Google BigQuery are specifically designed to handle the storage and analysis of large volumes of structured data, improving the efficiency of query execution for complex analytical tasks. A process digital twin generates a virtual business process model, which, using a causal process model including agents, capacity, randomness, and context, is both real-time and forward-looking.

In most existing process mining techniques, the entire event log is typically considered, which makes the process models complex, hard to understand, and overloaded, especially among the unstructured data (e.g., health field) due to the individuality of patients (7). Grouping patients of similar features and analyzing the process per group is more appropriate (8); otherwise, the heterogeneity of patients could result in a spaghetti model that obscures the particular features' influence. To address this, event logs can be filtered to focus on specific subsets of cases or activities, but analyzing and comparing multiple groups of patients requires additional effort. This fact necessitates the introduction of an approach that enables analyst(s) to separate event logs into groups of common cases concerning the flexible nature of homogeneous features. By doing so, a customized process model can be mined and compared for each group.

This approach can be materialized through multidimensional process mining (MPM), where the data cube concept is applied (7). In MPM, the attributes of the described event log and the formation of a multidimensional data space are the primary concerns (9). Each combination of dimension values generates a cell in the cube that embodies a sub-set of the event log (sub-log) concerning these dimensions' values. Analysts can manipulate the data cube and define specific views of the data to formulate new hypotheses using OLAP operators.

Since the disease registry is saturated with valuable, critical data on human life and supports clinicians and professionals in their decision-making and patient monitoring (10), the aim was to develop a Care Process Digital Twin (CPDT) as our overall work. Since the process warehouse is the first step in constructing the CPDT, the focus here was on the central component of the CPDT Platform. The primary contribution of this work was to demonstrate how to gather heterogeneous data from different sources, transform it into process-oriented data within a single repository, and express the OLAP queries through SQL to promote data filtering and aggregation.

[Downloaded from jsurgery.bums.ac.ir on 2026-11-25]

[DOI: 10.61186/jsurgtrauma.13.2.63]

Methods

Data Introduction

Operated by the Sina Trauma and Surgery Research Center at Sina Hospital in Tehran, the National Trauma Registry of Iran (NTRI) functions as a multi-hospital registry that systematically collects comprehensive injury-related data for individual patients. This registry compiles de-identified clinical abstracts from diverse sources, including patient records, medical examinations, standardized questionnaires, and HIS (Hospital information system). Three qualified nurses, trained as specialized registrars, input trauma-related details into standardized forms and manually enter this information via the NTRI's web-based platform. The NTRI database includes records for over 26,000 patients and 500,000 medical interactions spanning seven Iranian hospitals between 2017 and 2021. The data consist of (1) Demographics details (e.g., age, gender, underlying diseases), (2) injury description and body region (e.g., injury mechanism and nature, Injury Severity Score (ISS) (ISS describe severity of injury in a trauma patient and is validated as follows: <9 : Mild, 9 – 15 :Moderate, 16–24 : Severe, and >=25 : Profound), Abbreviated Injury Scale (AIS) (AIS is an anatomically-based injury severity scoring system which classifies each injury by body region as follows: Minor, Moderate, Serious, Severe, Critical, and Maximal injury.)), (3) Prehospital care (e.g. Timestamps for ambulance dispatch/arrival, transport details, and initial vital signs), (4) emergency department information (e.g. Time of admission, vital signs at arrival, and respiratory interventions), (5) Hospital care trajectory (e.g. types of diagnostic, therapeutic procedures, medical procedures, and specialist consultations), (6) outcome information (e.g., LOS, mortality rates, and discharge status), and (7) financial data (e.g., costs).

Cases were excluded if (1) patients were admitted, hospitalized, or treated outside Sina Hospital, or (2) documentation was incomplete (e.g., missing timestamps for critical events). After applying these criteria, the refined event log comprised 4,498 cases, 44,344 events, and 104 distinct clinical activities between 2017 and 2021. Key characteristics of the analyzed dataset are summarized in Table 1.

A short description of the dataset used in this applied study is represented in Table 1. The primary causes of traumatic injuries included road traffic accidents, falls, and injuries involving sharp objects. Males predominated in road accident cases, whereas females were more frequently represented in fall-related incidents. Despite a 5.5-fold higher incidence of trauma among males compared to females, the average age of female patients was significantly greater (53 years vs. 37 years for males). Summer months accounted for the highest proportion of referrals (30%), with 51% of events recorded during evening hours. Hospitalization durations were brief for most patients (75% ≤7 days), and the majority were discharged (97%), with a mortality rate of 3%. This study adhered to the ethical principles outlined in the Declaration of Helsinki. Approval for the research protocol was granted by the Research Ethics Board of Tehran University of Medical Sciences (Approval ID: IR.TUMS.MEDICINE.REC.1403.509). Prior to granting researchers access, all identifying details (e.g., patient names, birthdates) within the dataset were rigorously de-identified and anonymized by the Information Technology division of the Sina Trauma and Surgery Research Center, ensuring compliance with confidentiality protocols.

The two types of technical architecture prevail in this context: the *Inmon architecture*, where the term DW is introduced (2), and the *Kimball architecture*, where the DW Toolkit first edition is published (3). The top-down approach begins with the corporate data model in Inmon's architecture, while, in Kimball's architecture, the opposite holds due to dimensional modeling, where the back-room and front-room, the fundamental concept of the star schema, are proposed. Since the bottom-up DW approaches (star schemes) usually perform better by creating more constrained construction between descriptive (dimensions) and subjected tables (fact tables), it can be a structured schema for Registry Event Log Warehouse (REWH). Like the traditional schema, the fact table is positioned in the DW center containing foreign keys for all dimension tables to store the cells of the data cube—a specific combination of foreign keys references to the cells' dimension values. Instead of the cell's value, the fact table stores a unique ID (2). This scheme is named the star schema because its entity-relation diagram resembles a star, with rays radiating from the table center. Next to the simplicity of this schema, it is comprehendible by the users and flexible for future changes when the designer can add dimension(s). The system's efficiency, due to the few connections between facts and dimensions, which reduce the scans in retrieving the information, is affected by this scheme.

Definition 1: *(Eventlog warehouse)* is a fact-dimension relation consisting of five elements (F, D_S, R, MD, BD), where F is the fact table, D_S is the set of dimensions, R is the set of fact-dimension or dimension-dimension relations between entities (each relation r is a pair of entities $r = (e_1, e_2)$), MD is metadata of the event warehouse, and BD is other optional data.

Definition 2: *(Star Schema)* based on this definition, an empty star schema is created for each data object associated with a special scenario of business process P (e.g., the star schema of fall

[Downloaded from jsurgery.bums.ac.ir on 2026-11-25]

[DOI: 10.61186/jsurgtrauma.13.2.63]

trauma treatment). For each data object o, $\forall$ o $\in$ P a star schema S and a fact table $f \in S$ are created such that o. name = f. name.

The fact table consists of a five-element $(F, key, a_F, r, measurements)$ in which F represents facts. Most facts are numerical and additive and can be aggregated across all dimensions (e.g., the number of MRI scans). Some are semi-additive (e.g., account balance), and some are non-additive (e.g., unit price), which contain discrete descriptive data applied in constraining a query instead of forming aggregation in a computation. The primary key of the fact table includes a foreign keys subset from all dimension tables represented by the tuple $(id, name)$. $measurements$ is a set of the numerical attributes that quantify the amount of an event (e.g., in-hospital LOS). r is a set of dimensional attributes defining the fact-dimensions relations. a_F is a set of quantitative attributes, in fact, a table such that $\forall$ o. name = f. name , $f \in S, a = quantitative \rightarrow a \in a_F$.

Definition 3: *(Dimension creation)* Dimension is a four-tuple $(D, key, a_D, [r])$ in which D refers to a dimension, a_D is dimensional features and [] is the optional feature objects. The first-level dimension is created for each data object o associated with fact table f, $\forall\, o \in f$ a dimension $d \in S$ is created such that $o.name = d.name$. for each dimension, the second-level dimensions are created and associated with other classes, $\forall\, o \in d$ a dimension $\acute{d} \in S$ is created such that o. name = $\acute{d}$. name on the condition that $\exists\, r \in R\ r = (d, \acute{d})$.

Definition 4: *(Dimensional Attributes)* Based on this definition, the characteristics of the dimensional objects are added to the schema. $\forall\, d \in D, \exists\, r \in R$ such that $\forall\, a \neq a_k \rightarrow a \in a_D$, in which a_k is a key feature and a_D is a dimensional feature.

Definition 5: *(Basic dimension creation)* The basic dimension includes Time, Event ID, and Process instance ID,... which are added based on the features types of the fact table. $\forall$ o. name = f. name, $f \in S, (a) \in \{eventID, InstanceID, timestamp, \dots\}$ create a dimension $d \in S \wedge d.name = a.name$

Definition 6: *(Key Attributes)* The dimensional attributes associated with other data are considered key attributes. A hierarchical structure is applied to generate and manage numerous categories within a dimension for drill-down traces.

$$\forall\, d \in D, \exists\, o \in O, \text{o. name} = \text{d. name}, \quad \forall\, a = a_k \rightarrow a \in a_k$$

$\forall\, \acute{d} \in D,$
$that\ \acute{d}\ is\ a\ second\ level\ dimension, \forall\, a = a_{\acute{d}} \rightarrow a \in a_D$
$\forall\, d \in D,\quad that\ d\ is\ a\ first\ level\ dimension, \forall\, a = a_D \rightarrow a \in a_{FactTable}$

This schema of the constructed PW, as an example in the trauma registry, is illustrated in Figure (1), which includes the following six dimensions.

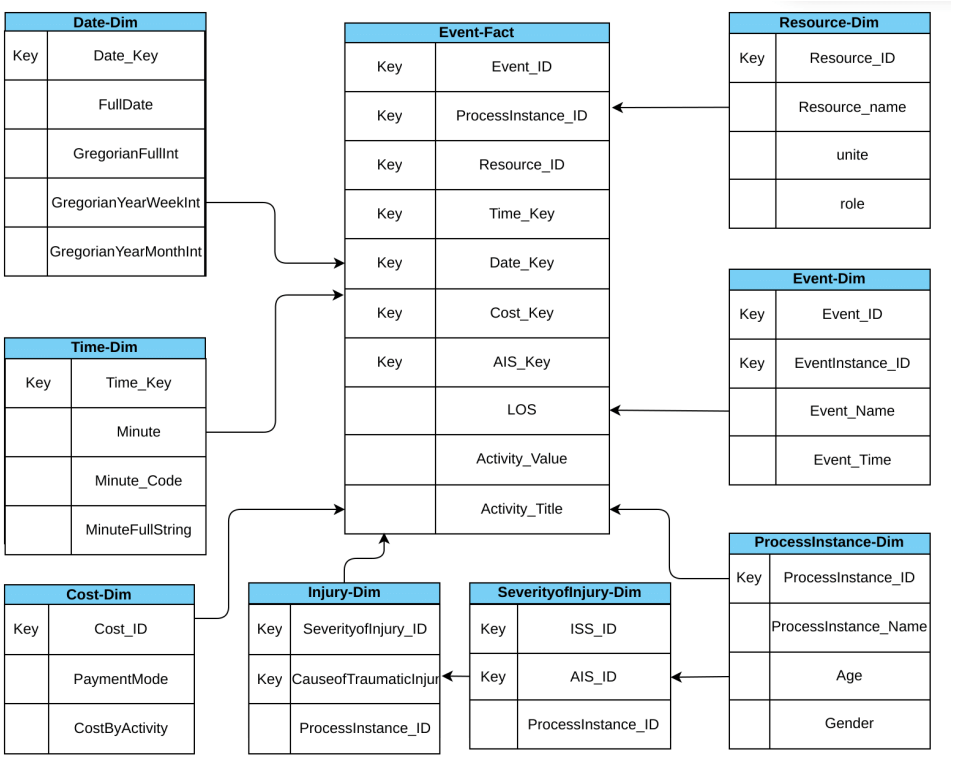

Figure 1. Star schema of the Registry Process Warehouse

[Downloaded from jsurgery.bums.ac.ir on 2026-11-25]

[DOI: 10.61186/jsurgtrauma.13.2.63]

- **Time Dimension** is the main dimension in the REWH. It stores temporal information about activity execution. Here, Time and Date calendars are modeled as separate dimensions. In the activity level, the information is given in the minute unit, whereas analysts may interested in viewing this information on a weekly or monthly level.
- **Event Dimension** consists of activities identified by a unique ID. A combination of the process instance ID is used to distinguish between activities with a similar name in different traces, which determines which instance they belong to. By doing so, different criteria can be applied at the process instances level or events level
- **Process instance Dimension** stores information about process participants, e.g., age, gender, and ID.
- **Resource dimension** stores information about activity performers, e.g., staff's name or ID.
- **Cost Dimension** stores information about the final cost paid by patients, payment method, insurance information, and the cost per activity.
- **Injury Dimension** stores information about the external cause of injury, severity of injury, and final diagnostics of injury.

No single value exists in the cells of the process cube, while the opposite holds for complex data. Since the available DW tools cannot handle such data, the REWH requires specific solutions for storing the event log data. The structure of event logs to a process cube is mapped through this approach, where cases and events at different levels become organized. Here, the cells contain sets of cases and a collection of events. The attributes of the cases (A_1 to A_p), composed of the patient's properties' metadata: 1) the traumatic injury causation, 2) the traumatized person's position, and 3) AIS, are considered the dimensions that construct the multidimensional data cube. Each case is mapped to a cell according to its attributes' values. A cell of the cube is identified by any combination of its dimension values. According to their respective case, the events and their attributes (B_1 to B_p) are stored inside the REWH. Similar to case attributes, the event attributes are interpreted as dimensions, Figure (2). At the event level, each cell in the process cubes consists of a set of events identified by a combination of event dimension values. All dimensions may have an arbitrary count of hierarchically structured dimension levels at both levels.

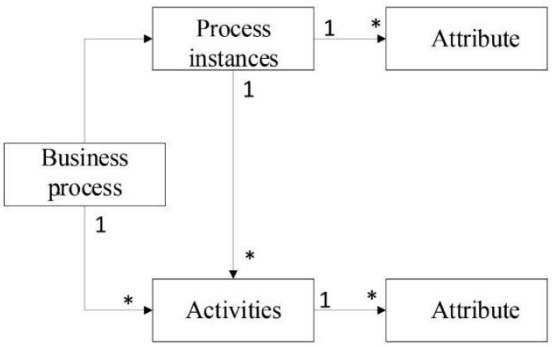

Figure 2. Typical structure of a process as UML class diagram. The activities are associated with a respective process instance (case), and the ordered sequence of activities for an instance forms a path. Both the instances and activities include different types of attributes.

Applying the OLAP operators on a set of base operators like the selection and aggregation, filtering (slice or dice) the event data, or changing its aggregation level (roll-up or drill-down) on both the case and the event levels becomes possible. The union of all the cells' cases is formed by aggregating cells on the case level. Aggregating the cube along the Gender and Age dimensions generates a single cell containing all cases for genders of all ages for a specific year; aggregating cells on the event level will merge all events into a single abstract event (event class). The event class forms a granularity layer with more details for each type of activity, thus avoiding the aggregation of different event types. This layer is beneficial when the analyst is interested in whether an MRI test is run, regardless of the MRI type or how many times. Regarding the case level, the REWH is filtered by both case and event attributes. In an aggregation function, the min, max, avg, sum, or count is specified to select cases exceeding a maximum medical activity count.

In the process cube Figure (3), events are grouped in cells according to case properties, event class, and time dimensions. The event class consists of MRI, ultrasonography, radiography, echocardiography, CT scan, and angiography activity classes in the process cube, and the activity name depends on the activity. The time dimension may refer to a specific day, week, month, or year. A process cell in Fig. (3) contains the aggregation of cases, including specific events, concerning the *traffic road trauma care* process diagnosed by *FAST* in *ultrasonography* activity class during *2018* and *2019*, like {1007770: <Getting injured, Calling ambulance, Admission by ED, FAST, Chest radiography, Brain CT scan, Face and Sinus CT scan, Transfer from ED to Neurosurgery, Brain CT scan^2, Discharge>}.

[Downloaded from jsurgery.bums.ac.ir on 2026-11-25]

[DOI: 10.61186/jsurgtrauma.13.2.63]

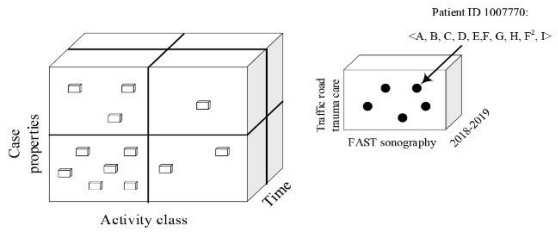

Figure 3. Aggregation of events and cases

Results

All experiments are run on a laptop with Intel Core i5-4200M 2.5 GHz CPU, 6 GB DDR3 RAM, and WDC HDD running on 64-bit Windows 10. To establish the practical feasibility of this REWH approach, a solution was implemented for the analysis services of Microsoft SQL Server 2019.

The beginning point defined by a user is the OLAP queries, through which the user describes a logical view of the data cube. This leads to a set of cells containing a sub-log at all case and event attributes and expressing the appropriate filtering and aggregation operations. Each sub-log is mined separately through an arbitrary process discovery algorithm to develop a process model to reflect the sub-log behavior. Reducing the event count may simplify the mined process model because its node count is reduced. This reduction initiates flexible views on the event data and convenient analysis of processes for different patient cohorts. It is possible to compare the clinical pathways of different clusters of patients.

The structure of a query pattern that detects sequences based on the *ActivityName* and *ActivityValue*, (i.e., select all patient IDs where a specific sub-process 'P_X = [1], [2], [3], [12]' is executed), is shown in Fig. (4) Sub-query1. The activities are labeled by numbers (e.g., the *Calling ambulance* activity is referred to [1]). The patients are ordered based on Event ID in Tables T1 and T2 in Figure (4) to restore the sequence of events for each case. This query returns three patients.

The Sub-query2 selects the activities of all case IDs that satisfy a specific condition. Each SQL query represents a specific cell; consequently, the cube-forming dimensions must correspond with the respective dimension values. Provided that a cell represents all patients of the year 2020, the filter conditions preceding WHERE must contain the D.MonthInt>=202005 AND like D.MonthInt<=202006 expression, where D. MonthInt is the table representing the time dimension at the month level. The *LEFT JOIN* in Figure (5) returns all records from Table 1 and then matches records with Table 2, which results in 169 cases.

```
With T1 AS
(
    Select ROW_NUMBER() Over (partition By PatientID order by EventID) AS R,
    patientID,
    ActivityName
    --ActivityValue,
    --ActivityName
    From FactProcess
) ,
T2 As
(
Select *
From T1
pivot(min(ActivityName) for R in ([1],[2],[3],[4],[5],[6],[7],[8],[9],[10],[11],[12])) AS P
)
Select * From T2
Where
([1] Like '%TimeStamp of Calling ambullance%')
And ([2] Like N'%TimeStamp of Admission by ED %' )
And ([3] Like N'%Transfer from ED to ICU%' )
And ([4] Like N'%TimeStamp of Getting injured %' )
And ([5] like N'%Radiography%')
And ([7] like N'%FAST%')
```

Figure 4. Sub-query1 for selecting patients with specifically executed sub-processes

[Downloaded from jsurgery.bums.ac.ir on 2026-11-25]

[DOI: 10.61186/jsurgtrauma.13.2.63]

```
Select *
 FROM [HospitalDW].[dbo].[FactProcess] F
 Left Join DimDate D ON F.DateKey=D.DateKey
 Where
    ActivityTitle like N'%Ultrasonography%'
    And F.ActivityValue like N'%FAST%'
    And CarryOutSurgery like N'%NO%'
    --And FinalDiagnosisOfInjury1 like N'%T07 -- Unspecified multiple injuries%'
    And D.PersianYearMonthInt>=139505
    And D.PersianYearMonthInt<=139906
```

Figure 5. Sub-query2 for selecting specific patients with a specific executed activity at a specific time

Table 1. Summary of NTRI dataset (4498 patients during 2017-2021)

Attribute	Value
Number of patients	4498
Gender (male/female)	3842/656
Age (mean/st. dev)	41.7/18.7
Cause of traumatic injury	
• Road traffic injuries	2313
• Fall	1020
• Sharp force injuries	874
• Blunt	211
• Gun violence trauma	55
• Other transport accident	11
• Poisoning	4
• Electrical injuries	4
• Animal attacks	2
• Other	4
Arrival mode	
• Ambulance (ground/helicopter)	3038
• Self-presentation	1367
• Public services	84
• Air ambulances	9
Length of stay (LOS in day) (mean/st Dev)	
• ICU	2/ 5.14
• In-hospital	5.69/ 10.44
Discharge status	
• Discharged to home	4255
• Inter-hospital patient transfer	108
• Passed away	101
• Escaped	34

The different kinds of case selection over event attributes ($\exists, \forall, aggregation$) highlights variations in the patterns for sub-queries. The sub-query for the cases with a specific occurrence frequency selection per case that matches a specific activity class is shown in Figure (6), where *Select* is replaced by a list of database attributes that are to be loaded. By inserting a GROUP BY patient ID statement, the cases are to be merged into a single high-level attribute form group. All patient IDs of an event are selected through this statement to meet the count condition for a specific Activityclass (i.e., all patients with at least one MRI during their treatment). The Dim ActivityTitle aggregates all events of a case referring to the same activity to a new high-level activity named Activityclass, regardless of the Activityvalue.

Selection of all cases through the aggregated event attributes becomes possible when each case with an attribute is assumed to represent the individual resource for the care procedure implementation (e.g., having at least an average execution duration of 20 minutes per Resource1 when Activity title= MRI and Activity value= hand). This procedure enables the analyst(s) to define a specific filtering of cases. The sub-query of this filtering is shown in Figure (7), with the results tabulated in Table 1. Applying the SUM, AVG, MIN, or MAX arbitrary SQL aggregation functions for any event/case attribute becomes possible depending on the analysis question.

[Downloaded from jsurgery.bums.ac.ir on 2026-11-25]

[DOI: 10.61186/jsurgtrauma.13.2.63]

```
Select patientID,Count(*)
FROM [HospitalDW].[dbo].[FactProcess]
Where
 ActivityTitle like N'%CT Scan%'
 and Activityvalue like N'%Lung%'
   Group By patientID
Having Count(*)>=3
```

Figure 6. Sub-query3 for selecting patient IDs with at least one event matching a condition

```
SELECT AVG(F.HospitalLengthOfStay_Day)
FROM [HospitalDW].[dbo].[FactProcess] F
WHERE F.Gender=N'Male' AND F.CauseOfTraumaticInjuryKey=5
```

Figure 7. Sub-query4 for selecting patient IDs by aggregated event attributes

Table 1. Results of sub-query4

Cause of traumatic injuries	Average hospital length of stay
Falls	101
Traffic road injuries	46
Gun violence	13
Assaults	6
Cut wound	4

To demonstrate the applications of REWH in process process-oriented analysis, a sub-process model of Queries 1-3, where the alpha miner in PM4PY is applied, is discovered, Fig. (8), where the data are extracted from the corresponding PW and the process cube. The process mining preparation is through the PW and the process cube, where the required conditions are selected from the PW, and the outcome is allocated to the process cube. At this stage, the process mining, like process discovery and conformance checking, is supported by the process cube.

As described by [25], the severity of injuries should be evaluated through ISS within 5 minutes of admission. Also, experts must perform the necessary imaging in the first 40 minutes of admission. By defining a performance evaluation measure with if-conditional rules, it is possible to check the treatment performance. The rule in this case is that *IF* the *Time duration from admission to FAST < 30 min. Then the performance is good*, and *IF* the *Time duration from admission to FAST<120 min, then the performance is average*, and *otherwise, the performance is bad*. In addition, one can filter the results according to the *cause of traumatic injury* or the *final diagnostics of injury*. The same criterion can be used for imaging activities such as MRI, CT scan, etc. Figure 9 indicates the histogram of sub-query 5. The treatment performance of 169 patients is ranked as good level, i.e., the time interval of *admission* to *FAST* lasted less than 30 minutes. While 200 patients report average values, i.e., time interval less than 2 hours, FAST did not perform well for the rest of the patients (969 out of 1328).

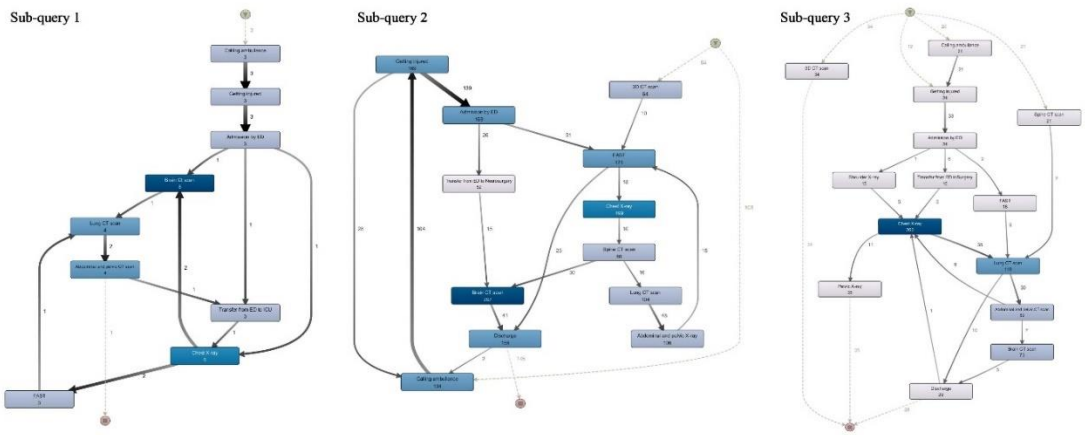

Figure 8. Discovered process model of a) sub-query1, b) sub-query2, and c) sub-query3

[Downloaded from jsurgery.bums.ac.ir on 2026-11-25]

[DOI: 10.61186/jsurgtrauma.13.2.63]

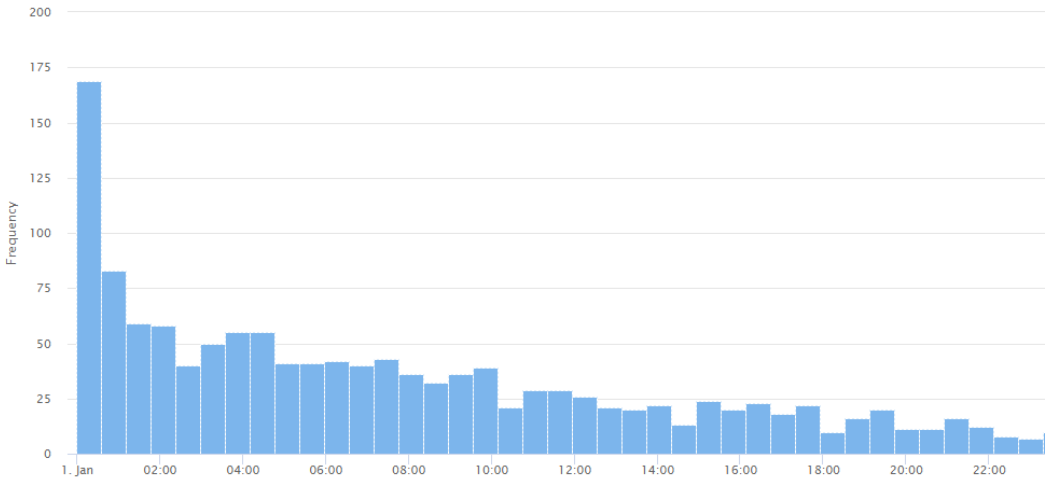

Figure 9. Histogram of sub-query5

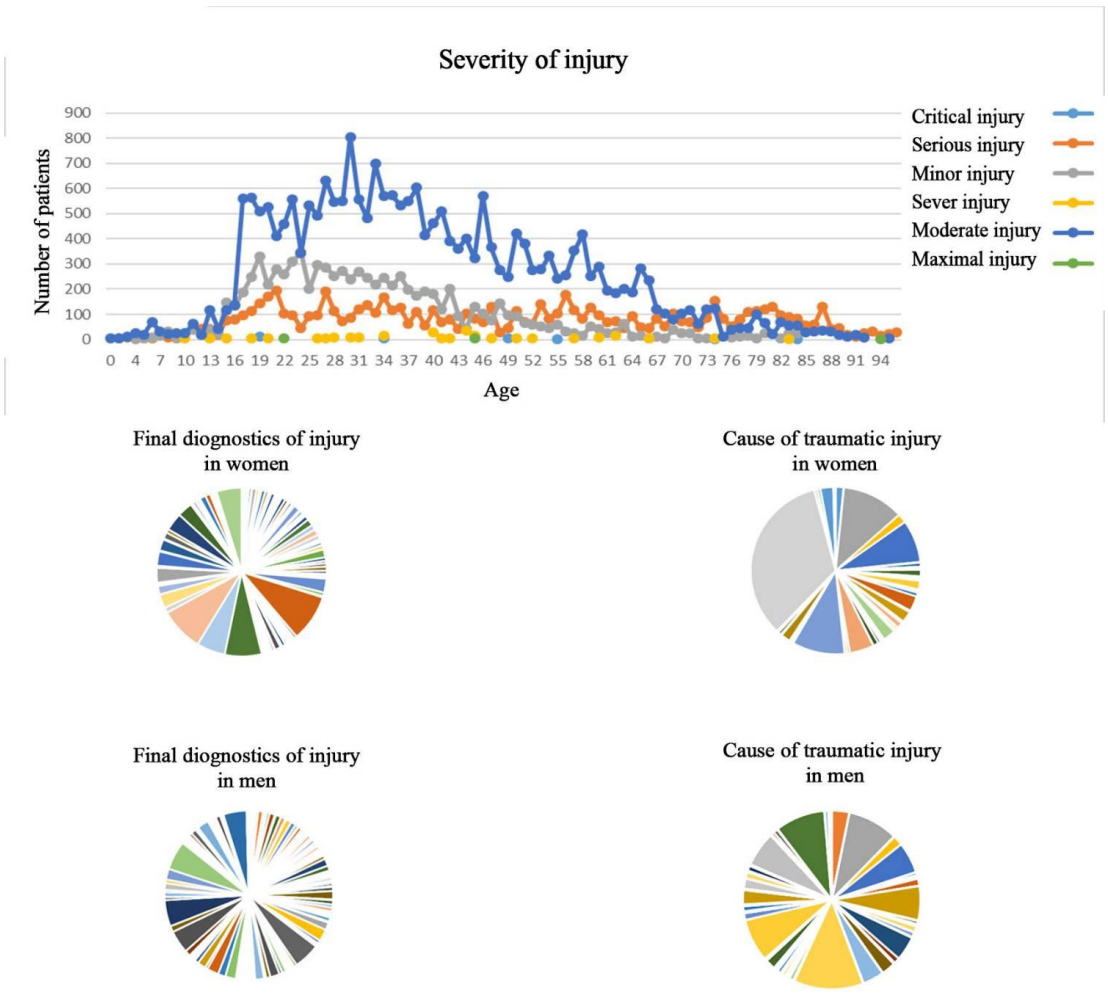

Figure 10. Designed dashboard based on gender, age, and injury characteristics

Next to process-oriented analyses, REWH can provide data-oriented. The dashboard in Figure 10 reports the severity of injuries across different age categories. As it is shown, in the age range of 70 and over, serious injuries are recorded more than moderate injuries (unlike other age categories), and

[Downloaded from jsurgery.bums.ac.ir on 2026-11-25]

[DOI: 10.61186/jsurgtrauma.13.2.63]

the frequency of occurrence of severe injuries is higher in the age range of 20-50 years. The bottom results in the dashboard indicate the cause of traumatic injury and the diagnosis of injury by gender. The most *common injury* mechanisms in the *male* group are road accidents by V42.4 (frequency=4144), V02.1 (frequency=3027), X99 (frequency=2959), W17 (frequency=2539), V22.4 and W77 (frequency=2100), respectively. Whereas females are mostly victims of falls and sharp objects by W77 (with a frequency of 2012), V02.1 and W17 and V03.1 (with a frequency of 700), and V43.5 and W25 (with a frequency of 150). Women were pedestrians in road accidents. The anatomic region most often injured in men is the leg and hand, with common diagnoses as S82.9 (with frequency 1926), S52.5 and S82.1 (with frequency 1700), S82 and T07 (with frequency 1500), S52 and S92.3 (with frequency 800). The anatomic region most often injured in women is the hip joint, femur, and pelvis, with common diagnoses as S72.1 and S52.5 (frequency=500), S72.0 (frequency=700), as well as S82.8 and S82.1 (frequency=300).

According to the REWH structure, resource performance queries can also be run, like "the average duration time of activity execution by resource R1 when Activitytitle=MRI and Activityvalue=Spine". However, resource-related queries will have no output because of the lack of resource information due to data privacy issues.

Discussion

In OLAP approaches, the schema usually consists of a fact table where the data values of the dimensions and their hierarchies are stored and linked to other tables. Each dimension is stored in a single table representing its levels in the star schema, while each dimension level is stored in its table in the snowflake schema (1).

Enhancing analysis of business processes and storing process data by applying DW and OLAP technologies has gained momentum over the last 15 years (11). Nishiyama introduces the PW concept, consisting of many aspects of target technology compiled into a practical information matrix for software process improvement (12). Casati et al. (13) proposed a generic solution for warehousing business processes over HP and its customer processes, where the abstracted process models map the process progression (i.e., associating the beginning and the completion of each step) to events occurring in the source systems. 14-Koncilia et al. (14) focused on analyzing the complex workflow logs, like splits/joins and loops, through the OLAP tools and defined sequences of events captured from the trace log after being initiated into a DW. A framework consisting of 1) a generic data model capable of capturing and consolidating the process variants into a reference process model and 2) a PW capable of running OLAP operations on different variations is presented in (15).

To analyze the underlying processes for identifying the in-situ problems, researchers (11, 12, 16) focused on the traditional PW concepts, which do not store complete event logs but measure the events' process performance and cases where cube dimensions are formed. An analysis is run in (7) where the measures along the dimensions are aggregated; nevertheless, these approaches generally do not support process mining.

The next-generation PW designs applied to process mining analysis have been developed. According to (17), the idea of partitioning the event data into 3-D sub-logs as class type, event class, and time window is proposed. Each cell in the process cube corresponds to a set of events and can apply different algorithms to discover a process model and expose the multiple process models in a grid. An event cube model is introduced by (18) for efficient multi-dimensional event pattern analysis at different abstraction levels, where an event pattern hierarchy that integrates specified complex event patterns is formed by applying the sequence, negation, and concept abstractions. The information retrieval technique is adopted to devise an index over an outdated structured event log from which a data cube is extracted. Instead of containing raw event data, each cell of an event cube contains the precomputed dependency measure, which generates a single process model through which the dimensions of each value are mapped into a different path. A framework is developed in (19) to construct a process cube for event data comparison, where a hierarchy level is defined only in the time dimension. The Event Cubes and Process Mining Cube (PMC) are subject to a MOLAP approach, where the associated operations like slice, dice, roll-up, and drill-down prevail. According to (20), the loading time concerning the data volume in PMC is somewhat lengthy. In PMC, the filtering is restricted for particular cases and events without allowing event aggregation into high-level events.

Clinical data is distributed among many related entities and is collected in heterogeneous formats that make different choices without common management or interoperability. Some technical and organizational solutions are developed in (21, 22), where the consistency of clinical data is improved (e.g., the clinical data warehouse (CDW)). In a CDW, the data, throughout different electronic systems, is collected into a consolidated database for analysis, improvement in care, and research.

[Downloaded from jsurgery.bums.ac.ir on 2026-11-25]

[DOI: 10.61186/jsurgtrauma.13.2.63]

Concerning CDW implementations, many approaches reveal that the practice and acceptance of evidence-based medicine next to strategic decision-making are due to CDW implementation. The possibilities of applying DW and OLAP technologies in public health care and the gained experience thereof on the outpatient data at the national level are described in (23). According to (24), the DW integration, OLAP, and data mining techniques in healthcare provide an easy and applicable decision support platform for caretakers and clinical managers.

In this context, a PW for the clinical registry is absent with all said and performed. Healthcare processes, where different heterogeneous resources and activities are of concern, are more complex than business-based processes. Multidimensional modeling is assumed to be a promising solution for applying contemporary analysis in the healthcare process, especially in trauma care domains comprising multidisciplinary medical teams. This modeling allows one to observe the data from different perspectives and granularities. A registry PW requires an extended ETL approach by enriching the contextual information and processing data. To the best of the authors' knowledge, no study has implemented a DW/PW to support the analytical needs of MPM in the healthcare field.

According to the REWH structure, resource performance queries can also be run, like "the average duration time of activity execution by resource R1 when Activitytitle=MRI and Activityvalue=Spine". However, resource-related queries will have no output because of the lack of resource information due to data privacy issues.

Conclusions

OLAP allows the analyst to obtain information rapidly and consider it a combination of many queries in one query by applying the rolling up and drilling down hierarchy concept. The multidimensional process of mining is featured by its explorative approach. The OLAP queries gradually become modified to allow the analysis of processes in different contexts.

The Registry Process Warehouse, a concept for warehousing disease registry processes, which is the central component of the process-aware recommender system, is presented in this article. As to future work, the cube will be applied for operational support through recommendations like which resources are to be assigned to the next activity in terms of time or cost and next to proper activity over a running process instance.

The REWH integrates heterogeneous data (beginning at text attributes and ending at process-oriented data) extracted from different sources into a single repository and expresses the OLAP queries through the SQL to attempt data filtering and aggregation. To demonstrate the practicality and feasibility of REWH, process models of each sub-query based on PM4PY and the top of the warehouse are discovered.

Acknowledgments

We want to acknowledge the National Trauma Registry of Iran (NTRI) for providing the essential data that made this research possible.

Funding

No funding information.

Conflict of Interest

The authors declare that they have no competing interests.

References

1. Kimball R Margy R. The data warehouse toolkit: The definitive guide to dimensional modeling, 3rd ed. Wiley. 2019.
2. Inmon WH. Building the data warehouse, 3rd ed. Wiley, New York. 2002.
3. Kimball R. The data warehouse toolkit: Practical techniques for building dimensional data Warehouses, John and Wiley & Sons, Inc.1996.
4. Jarke M, List T, Koller J. The challenge of process data warehousing. In: Proceedings of the 26th International Conference on Very Large Data Bases (VLDB'00), Cairo. 2000: 473- 483.
5. Benker T. A generic process data warehouse schema for BPMN workflows. In Business Information Systems: 19th International Conference, BIS 2016. Lecture Notes in Business Information Processing; Springer International Publishing. 2016;255:222-234
6. Shahzad K. Extending the REA ontology for the evaluation of process wrehousing approaches. In Lecture Notes in Computer Science. Springer Berlin Heidelberg. 2009; 5736:196-206
7. Vogelgesang T, Appelrath HJ. A relational data warehouse for multidimensional process mining. In Data-Driven Process Discovery and Analysis: 5th IFIP WG 2.6 International Symposium, SIMPDA 2015.Springer International Publishing. 2017;244:155-184.
8. Eili MY, Rezaeenour J, Bidgoly AJ. Mining trauma care flows of patient cohorts. Intelligence-Based Med. 2024;10:100150.
9. Vogelgesang T, Appelrath HJ. PMCube: a data-warehouse-based approach for multidimensional process mining. In Business Process Management Workshops: BPM 2015, 13th International Workshops, Innsbruck, Austria, August 31–

[Downloaded from jsurgery.bums.ac.ir on 2026-11-25]

[DOI: 10.61186/jsurgtrauma.13.2.63]

September 3, 2015. Springer International Publishing. 2016;256:167-178.
10. Kohzadi Z, Nickfarjam AM, Shokrizadeh Arani L, Kohzadi Z, Mahdian M. A comprehensive evaluation of ensemble learning methods and decision trees for predicting trauma patient discharge status using real-world data. Arch Trauma Res. 2023;12(3):137-149.
11. Kiss C, List B. Analysing collaborative workflows with a data warehouse-A case study in the insurance sector. In Vom Data Warehouse zum Corporate Knowledge Center. Physica, Heidelberg. 2002:491-505.
12. Nishiyama T. Using a 'process warehouse' concept: A practical method for successful technology transfer. *ISORC'99 Proceedings*.1999: 117-120.
13. Casati F, Castellanos M, Dayal U, Salazar N. A generic solution for warehousing business process data. In Proceedings of the 33rd international conference on Very large data bases. 2007:1128-1137.
14. Koncilia C, Pichler H, Wrembel R. A generic data warehouse architecture for analyzing workflow logs. In Advances in Databases and Information Systems; Lecture Notes in Computer Science. Springer. 2015;9282:106-119.
15. Berberi L. Analysis of process variants with a process warehouse approach [Doctoral dissertation]. Alpen-Adria-Universität Klagenfurt ;2021.
16. Schiefer, J., List, B., & Bruckner, R.M., 2003. Process data store: A real-time data store for monitoring business processes. Lecture Notes in Computer Science; Database and Expert Systems Applications (DEXA). 2003;2736:760-770.
17. Van Der Aalst WM. Process cubes: Slicing, dicing, rolling up and drilling down event data for process mining. In Asia Pacific Business Process Management: First Asia Pacific Conference, AP-BPM 2013, Beijing, China, August 29-30, 2013. Springer International Publishing. 2013;159:1-22.
18. Liu M, Rundensteiner E, Greenfield K, Gupta C, Wang S, Ari I, et al. E-cube: multi-dimensional event sequence analysis using hierarchical pattern query sharing. In Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. 2011: 889-900.
19. Vogelgesang T, Appelrath HJ. Multidimensional process mining with PMCube explorer. BPM (Demos).2015: 90-94.
20. Bolt A, van der Aalst WM. Multidimensional process mining using process cubes. In: Enterprise, Business-Process and Information Systems Modeling. Lecture Notes in Business Information Processing; Springer International Publishing. 2015; 214;:102-116.
21. Doutreligne M, Degremont A, Jachiet PA, Lamer A, Tannier X. Good practices for clinical data warehouse implementation: A case study in France. PLOS Digit Health. 2023;2(7):e0000298.
22. Schaaf J, Chalmers J, Omran H, Pennekamp P, Sitbon O, Wagner TOF, et al. The registry data warehouse in the European reference network for rare respiratory diseases - background, conception and implementation. Stud Health Technol Inform. 2021;278:41-48.
23. Hristovski D, Rogac M, Markota M. Using data warehousing and OLAP in public health care. Proc AMIA Symp. 2000:369-73.
24. Stolba N, Tjoa AM. The relevance of data warehousing and data mining in the field of evidence-based medicine to support healthcare decision making. Int J Computer Syst Sci Eng. 2006; 3: 12-17.
25. Bloom BA, Gibbons RC. Focused Assessment With Sonography for Trauma. [Updated 2023 Jul 24]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK470479/

[Downloaded from jsurgery.bums.ac.ir on 2026-11-25]

[DOI: 10.61186/jsurgtrauma.13.2.63]