Ari Fauzi Mukti Nugroho[1]
Master of Information Technology,
Universitas Teknologi Digital Indonesia,
Yogyakarta, Indonesia
email: arikrw@gmail.com

Rikie Kartadie
Department of Computer Engineering,
Universitas Teknologi Digital Indonesia,
Yogyakarta, Indonesia
email: rikie@utdi.ac.id

Latifah Nurrohmah Handayani
PT Dua Empat Tujuh,
Yogyakarta, Indonesia,
email: student.latifahnh24@mti.utdi.ac.id

Nenen Isnaeni
Department of Informatics,
Telkom University,
Purwokerto, Jawa Tengah, Indonesia
email: nenisna@telkomuniversity.ac.id

Moh. Abdul Kholik
Department of Computer System,
Faculty of Electrical Engineering and Informatics,
Universitas Surakarta
Surakarta, Jawa Tengah, Indonesia
email: abdul_kholik@unsa.ac.id

Building Data Warehouse For Employee Training Ministry of Law And Human Rights

In order to improve and expand each employee's competency and knowledge, education and training are crucial for the Ministry of Law and Human Rights. They may also be utilized for employee mapping. In the past, it was necessary to gather data on things like the number of employees who attended a particular training, which training had reached its participant quota, and the number of graduates in each training. This required extensive processing and repeated cross-checking of data sources to make sure the data was accurate and legitimate before it could be compiled into a table for analysis. Information technology may be used to immediately process employee competence data and education and training results into information. Therefore, it is expected that the Nine Step method, which is part of the Kimball and Ross (2010), methodology will simplify and accelerate the process of processing training data into information presented for analysis and reporting purposes at the leadership level in each work unit.

***KeyWords**: Data Warehouse, OLAP, ETL, Pentaho, Kemenkumham, Training*

This Article was:
submited: 28-11-24
accepted: 03-12-24
publish on: 10-11-24

How to Cite:
A. F. M. Nugroho, et al, "Building Data Warehouse For Employee Training Ministry of Law And Human Rights", Journal of Intelligent Software Systems, Vol.3, No.2, 2024, pp.22–26, 10.26798/jiss.v3i2.1500

1 Introduction

The digital era, an era where everything is required to be digital and also the level of data traffic is required to be very fast as if there is no pause in the speed of transformation, both data retrieval and data processing into information. So that people are helped in getting and delivering information anywhere, anytime in just a split second[1]. In its development, to accommodate these needs, there needs to be data processing that can facilitate requests for information presentation applications by creating a data warehouse. Where all data is collected in a data warehouse, which contains data that has been grouped according to its needs. From here the data will be accessed by the information presentation application[2].
The use of this data warehouse is widely used in marketplaces, social media, companies, education and so on. Therefore, the author intends to apply it in the Ministry of Law and Human Rights in the field of education and training. This research was taken because of the large amount of education and training data and the results of employee competency tests of the Ministry of Law and Human Rights which are still scattered and not integrated with each other because they come from more than one application, so that when processed to become reports and analysis results are still manual by comparing with other data sources and then arranged in one table format so that it will take a long time to process. Therefore, the author took the initiative to create a data warehouse from these data, which is useful for simplifying and accelerating the presentation of information needed for analysis and various information and report needs.
The implementation of a data warehouse within the Ministry of Law and Human Rights' education and training sector is expected to address these inefficiencies by centralizing and integrating data from various sources. With a data warehouse, disparate datasets can be transformed into a unified structure, enabling faster and more accurate data retrieval. This centralized repository will allow stakeholders to access reliable and up-to-date information, which can be used for competency assessments, training evaluations, and strategic decision-making. Moreover, the integration of historical and real-time data in a structured format will facilitate more comprehensive analyses, empowering the Ministry to identify trends, monitor progress, and optimize training programs effectively.
The data warehouse will serve as a foundation for implementing advanced analytics tools such as business intelligence dashboards and predictive analytics. These tools can provide insights into employee performance, highlight areas requiring improvement, and forecast future training needs. By reducing manual processing and minimizing the risk of human error, the system will save time and resources, enabling the Ministry to focus on enhancing the quality of its training programs. This initiative not only aligns with the demands of the digital era but also supports the Ministry's mission to foster a skilled and competent workforce capable of addressing

[1]Corresponding Author.

evolving challenges in the legal and human rights sectors.

2 Research Methodology

This study uses a research approach that has been carried out by [3] namely the Design of a Data Warehouse for Mapping Student Data in the Information Systems Department of STMIK Sumedang with the method used is the Nine Step Methodology. This data warehouse is used to facilitate the mapping of student data.
The research conducted by the author is to build a data warehouse for education and training carried out within the Ministry of Law and Human Rights which aims to facilitate obtaining information from training data that is being and has been carried out by its employees easily and quickly, so that it can be used for various reporting and decision-making purposes. This research was conducted using the same method as the research conducted by [3] namely using the Nine Step Methodology and the application used is Pentaho Data Integration. For data sources, the author uses training data from the Ministry of Law and Human Rights institutions originating from the e-learning application with the url [4] and the CBHRIS (Competency Base Human Resources Information System) application with the http://cbhris.kemenkumham.go.id and data stored in excel files.

2.1 Data Collection. Data collection was carried out by collecting data sources (operational data), namely the database from the training application managed by the Human Resources Development Agency for Law and Human Rights, namely the e-learning application with the https://e-learning.kemenkumham.go.id and the CBHRIS (Competency Base Human Resources Information System) application with the https://cbhris.kemenkumham.go.id The application only displays general training data, while the data used for research can only be accessed by admin accounts. By using the Pentaho Data Integration 9.3 application, the data will be processed ETL (Extract, Transform and Loading) into a data warehouse by storing the data into several dimension tables and fact tables. Furthermore, the data can be used for analysis purposes using the OLAP (OnLine Analytical Processing) application. The data collection process was carried out in December 2022.

2.2 System Design. In this study, the architecture used for the data warehouse is as follows:

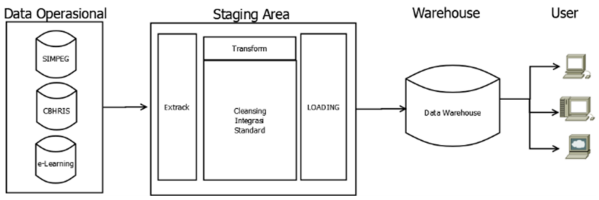

Fig. 1 Data Warehouse Architecture

From this architecture, data processing is divided into 4 areas;

(1) Operational data; The data included in Operational Data is the main data or data source from the CBHRIS application, e-Learning and data stored in excel files.
(2) Staging Area; This is an area for the ETL (Extract, Transform and Loading) data process, where the data will be first aligned in structure or format before the data is stored in the warehouse area. The tables in this area consist of employee tables, rank tables, participant tables, training tables, organizer tables and work unit tables.
(3) Warehouse Area. The data from the staging area will then be transformed into this warehouse area, this is where the process of compiling data is grouped according to the dimension table and fact table. The tables include the d_employee table, d_rank table, d_participant table, d_diklat table, d_organizer table, d_work unit table, d_time and f_organization

3 Results and Discussions

The testing was conducted using the Pentaho Data Integration application, the process begins with preparing the main application database (operational data) which is then extracted to the staging area. In this area, data is selected by cleansing according to needs (not all data is taken, but only what is needed). After that, the data is prepared to be transformed into dimension tables and fact tables in the data warehouse area[5][6][7]. The fact table is a table that stores measurement data, metrics, or facts related to business operations. So in this study, a fact table was created, namely f_organization (in Figure 2 it is colored orange).
Where this table will later contain the primary key derived from the following dimension tables (in Figure 2 it is colored light blue): d_employee table, d_rank table, d_participant table, d_diklat table, d_organizer table, d_satker table, and d_waktu table. This fact table has foreign keys including: id_diklat, id_peserta, and id_waktu, and in several other dimension tables also have foreign keys from other dimension tables including in the participant table which has a foreign key from the d_pegawai table and the d_pegawai table has a foreign key from the d_pangkat and d_satker tables. So if a snowflake-shaped scheme is made. Because this snowflake scheme is more complex, in the process it is necessary to normalize the dimension table.
By using the JOIN query to compare several dimension tables. The more data there is in a table, the merging process also takes longer. But because the time count is still in seconds, it can be said that it is still relatively fast and does not cause significant problems. After the dimension table and fact table are formed, the transformation process can be carried out as before. By using this scheme, the storage size of each table becomes more efficient because the data is divided into separate tables and is not made into 1 table[8][9]

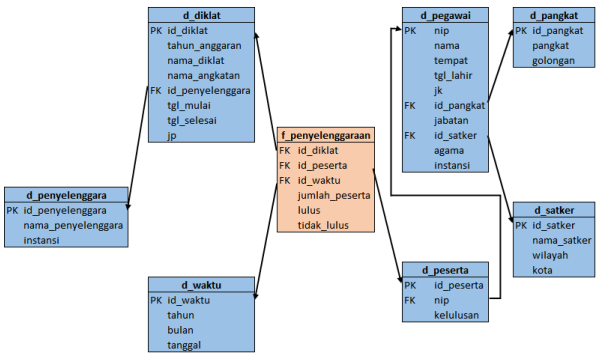

Fig. 2 Snowflake Scheme

From this data warehouse area, the data will be processed using certain queries to be used by analysis applications or reporting applications. The following is the process flow for ETL (Extract, Transform, Loading):

3.1 Extract. Data originating from data sources (operational data) applications (Competency Base Human Resources Information System) with the url https://e-learning.kemenkumham.go.id, as well as data in the excel file are selected. Because the data in the table still contains incomplete data, data filtering needs to be done to match the data needed in this ETL process[10][11].
In this case, the researcher filtered using a query, so that the data taken was completely correct and there was no missing / corrupt data. This is done to reduce the risk of data failure to be stored in the staging area. After the data is correct, the next process is that

the data will be entered into the staging area. The process steps are as shown in the image below:

Fig. 3 Employee Data Extraction Process From Source to Staging

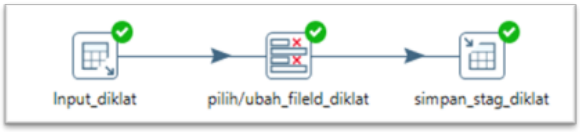

Fig. 4 Training Data Extraction Process From Source to Staging

Fig. 5 Participant Data Extraction Process From Source to Staging

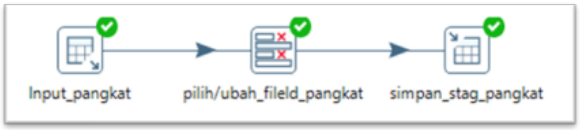

Fig. 6 Rank Data Extraction Process From Source to Staging

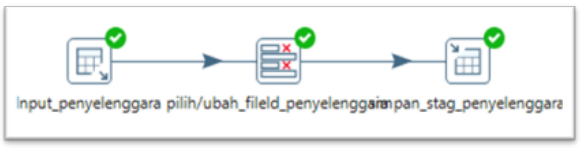

Fig. 7 Organizer Data Extraction Process From Source to Staging

Fig. 8 Work Unit Data Extraction Process From Source to Staging

The images above show the flow of the data process that is extracted and then stored in the staging area. The successful process is marked with a green check mark at each step. The extraction process uses the following SQL query:

```
SELECT nip,nama_pegawai,tempat_lahir,(SELECT
CONCAT((SUBSTRING(nip,1,4)),"-
",(SUBSTRING(nip,5,2)),"-
",(SUBSTRING(nip,7,2)))) AS
"tgl_lahir",jenis_kelamin,idpangkat,jabatan,ids
atker
FROM cbhris_backup.pegawai
WHERE idpangkat IS NOT NULL AND LENGTH(nip)>10
AND (tgl_lahir IS NOT NULL AND tempat_lahir IS
NOT NULL)
```

Fig. 9 The Employees Filtering Query

```
SELECT id, nama_diklat, idpenyelenggara,
tgl_mulai, tgl_akhir
FROM diklat
WHERE nama_diklat IS NOT NULL AND
idpenyelenggara IS NOT NULL
AND YEAR(tgl_akhir)>=2022
```

Fig. 10 The Training Data Filtering Query

```
SELECT id, nama_pangkat, gol_ruang
FROM pangkat
WHERE gol_ruang!=""
```

Fig. 11 Rank Data Filtering Query

```
SELECT id, nama_penyelenggara
FROM penyelenggara
```

Fig. 12 Organizer Data Filtering Query

```
SELECT idsatker, nama_satker, kode_wilayah,
nama_kota
FROM satker
WHERE idsatker IS NOT NULL
```

Fig. 13 Employees Data Filtering Query

The queries above are used to select and filter data that is truly intact (no NULL or less) and limit the data taken with the criteria of the year of implementation in 2020 and above, so that data before 2020 will not be taken and entered into the staging area. The data selection process is also carried out by adding a step select value, which is useful for matching the names of the data fields in the source data with the data stored in the staging area.
For participant data originating from an excel file, the process is almost the same as extraction from a database, namely by directing the data source in Pentaho Data Integration to the excel file, then

Table 1 Log In The Data Transformation Process

2023/03/20 11:45:32	-	pilih/ubah_field_penyelnggara.0 - Finished processing (i=0, O=0, R=11, W=11, U=0, E=0)
2023/03/20 11:45:33	-	Input_pangkat.0 - Finished Processing (I=13, O=0, R=0, W=13, U=0, E=0)
2023/03/20 11:45:33	-	pilih/ubah_field_pangkat.0 - Finished Processing (I=0, O=0, R=13, W=13, U=0, E=0)
2023/03/20 11:45:33	-	Input_satker.0 - Finished reading query, closing connection
2023/03/20 11:45:33	-	Input_pegawai.0 - Finished reading query, closing connection
2023/03/20 11:45:34	-	pilih/ubah_field_pegawai.0 - Finished Processing (I=0, O=0, R=1045, W=1045, U=0, E=0)
2023/03/20 11:45:34	-	pilih/ubah_field_satker.0 - Finished Processing (I=0, O=0, R=984, W=984, U=0, E=0)
2023/03/20 11:45:34	-	input_peserta_excel.0 - Finished Processing (I=24, O=0, R=0, W=24, U=0, E=0)
2023/03/20 11:45:34	-	input_pegawai.0 - Finished Processing (I=1045, O=0, R=0, W=1045, U=0, E=0)
2023/03/20 11:45:34	-	input_satker.0 - Finished Processing (I=984, O=0, R=0, W=984, U=0, E=0)
2023/03/20 11:45:34	-	Select_values.0 - Finished Processing (I=0, O=0, R=24, W=24, U=0, E=0)
2023/03/20 11:45:34	-	sort_peserta.0 - Finished Processing (I=0, O=0, R=24, W=24, U=0, E=0)
2023/03/20 11:45:34	-	simpan_stag_penyelenggara.0 - Finished Processing (I=0, O=11, R=11, W=11, U=0, E=0)
2023/03/20 11:45:34	-	simpan_stag_pangkat.0 - Finished Processing (I=0, O=13, R=13, W=13, U=0, E=0)
2023/03/20 11:45:34	-	Input_diklat.0 - Finished reading query, closing connection
2023/03/20 11:45:34	-	Input_diklat.0 - Finished Processing (I=2592, O=0, R=0, W=2592, U=0, E=0)
2023/03/20 11:45:34	-	pilih/ubah_field_diklat.0 - Finished Processing (I=2592, O=0, R=0, W=2592, U=0, E=0)
2023/03/20 11:45:34	-	sort_diklat.0 - Finished Processing (I=0, O=0, R=1296, W=1296, U=0, E=0)
2023/03/20 11:45:34	-	simpan_stag_satker.0 - Finished Processing (I=0, O=984, R=984, W=984, U=0, E=0)
2023/03/20 11:45:34	-	simpan_stag_pegawai.0 - Finished Processing (I=0, O=1045, R=1045, W=1045, U=0, E=0)
2023/03/20 11:45:34	-	simpan_stag_diklat.0 - Finished Processing (I=0, O=1296, R=1296, W=1296, U=0, E=0)
2023/03/20 11:45:35	-	Merge join.0 - Finished Processing (I=0, O=0, R=1320, W=24, U=0, E=0)
2023/03/20 11:45:35	-	Merge join.0 - Finished Processing (I=0, O=24, R=24, W=24, U=0, E=0)
2023/03/20 11:45:35	-	simpan_stag_diklat.0 - Finished Processing (I=0, O=1296, R=1296, W=1296, U=0, E=0)
2023/03/20 11:45:35	-	input_peserta.0 - linenr 50000

the file data is first formatted so that it is appropriate and can be moved to the database in the staging area. In this case, the data merging process requires a merger process with the training data in the database. This aims to ensure that the data in the empty excel table can be filled in according to the reference data from the database, this is useful for avoiding data failure due to empty or NULL data.

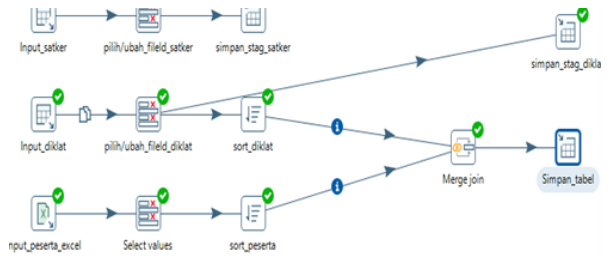

Fig. 14 Data Merge Process From Excel File with MySQL Database

3.2 Transform. This transform is a process where data that has been extracted is then sent/saved to staging or to the warehouse. In the transformation process, the data to be sent should not have a NULL value, because it will result in the data not being able to be processed by the OLAP (OnLine Analytical Processing) application, especially data used for calculations.
So that data with a NULL value needs to be deleted or the NULL value changed first with a default value, for example NULL is replaced with a value of 0, so that no errors or problems occur during the transform process. The following is a log in the transformation process that displays the amount of data that has been successfully transformed.

3.3 Loading. This process is carried out to store data into the data warehouse area with the condition that the data has been filtered and is clean from incomplete / lacking data, NULL values and so on. While to see the progress of the ETL process, the Pentaho Data Integration application also provides a monitoring feature called Step Matrix. So that it can always be monitored every process such as the amount of data sent and successfully sent, time, data delivery speed.

Table 2 Monitoring The Data Extraction Process To The Staging Area

Stepname	Copy	Read	Write	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	Input/ Output
Input_pangkat	0	0	13	13	0	0	0	0	Finished	0,3s	52	-
Input_peserta	0	0	40823	40823	0	0	0	0	Finished	2,8s	14.554	-
Input_diklat	0	0	472	472	0	0	0	0	Finished	0,4s	1.276	-
Input_penyelenggara	0	0	11	11	0	0	0	0	Finished	0,3s	44	-
Input_pegawai	0	0	513	513	0	0	0	0	Finished	0,3s	1.819	-
pilih/ubah_field_peserta	0	40823	40823	0	0	0	0	0	Finished	3,7s	11.024	-
Input_satker	0	0	984	984	0	0	0	0	Finished	0,3s	3.144	-
pilih/ubah_field_penyelnggara	0	11	11	0	0	0	0	0	Finished	0,2s	5	-
pilih/ubah_field_pegawai	0	513	513	0	0	0	0	0	Finished	0,3s	1.639	-
pilih/ubah_field_diklat	0	472	472	0	0	0	0	0	Finished	0,4s	1.157	-
simpan_stag_diklat	0	472	472	0	472	0	0	0	Finished	0,5s	908	-
simpan_stag_peserta	0	40823	40823	0	40823	0	0	0	Finished	4,7s	8.704	-
pilih/ubah_field_satker	0	984	984	0	0	0	0	0	Finished	0,4s	2.811	-
simpan_stag_satker	0	984	984	0	984	0	0	0	Finished	0,5s	1.836	-
simpan_stag_penyelenggara	0	11	11	0	11	0	0	0	Finished	0,4s	31	-
pilih/ubah_field_pangkat	0	13	13	0	0	0	0	0	Finished	0,3s	52	-
simpan_stag_pangkat	0	13	13	0	13	0	0	0	Finished	0,3s	42	-
simpan_stag_pegawai	0	513	513	0	513	0	0	0	Finished	0,5s	1.051	-

For the data transformation process carried out from the data source to the staging area, it is by first deleting the existing data in the destination table (staging area) after that the data from the source is copied to be entered into the staging table, with the condition that the source data has been filtered beforehand to avoid corrupt data and inappropriate data formats that cause the data transformation process to fail. So that the data will always be updated. In the data transformation process, there are also different times and speeds, this is influenced by the amount of data and the size of each table which is different, so that the processing can take longer and shorter. However, in terms of completion time, it is still relatively fast because it is still in seconds.

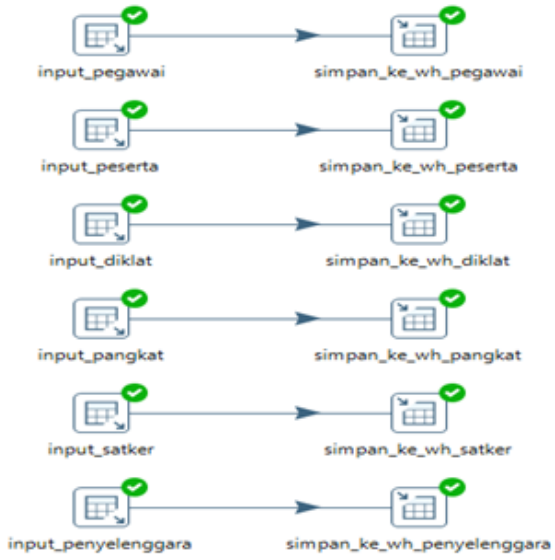

Fig. 15 Data Extraction Process From Staging To Warehouse

Data that has been stored in the data warehouse can then be presented using an OLAP application as shown in the image below:

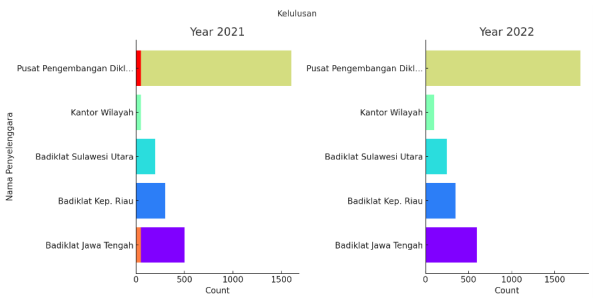

Fig. 16 Results of Processing Using OLAP Application

Table 3 Comparison of Training Participants Based on SK and Training Status

Organizer Name		2022	Comparison
Badiklat Jawa Tengah	Participants who participated	216.818	80.794
	Total Participants (According to Decree)	297.612	
Badiklat Kep. Riau	Participants who participated	139.969	53.985
	Total Participants (According to Decree)	193.954	
Badiklat Sulawesi Utara	Participants who participated	50.920	21.232
	Total Participants (According to Decree)	72.152	
Pusat Pengembangan Diklat Fungsional dan HAM	Participants who participated	115.974	47.290
	Total Participants (According to Decree)	163.264	
Pusat Pengembangan Diklat Teknis dan Kepemimpinan	Participants who participated	1.075.794	385.547
	Total Participants (According to Decree)	1.461.341	

The image above is a visualization of the comparison of participant data in the Participant Summons Decree and participant data whose status is willing and has participated in the training which is then presented using a bar graph and in the form of a regular comparison table to make it easier to read and analyze the data.

4 Conclusion

From the research that has been done, it can be concluded that the application of data warehouse for training data processing in the Ministry of Law and Human Rights is very important and necessary, because it can help leaders in analyzing, concluding and deciding the results and information presented from the data warehouse and can also be used for reporting and evaluation easily, quickly, precisely, and accurately. So that the presentation of the data can be maximized.

References

[1] Haleem, A., Javaid, M., Qadri, M. A., and Suman, R., 2022, "Understanding the role of digital technologies in education: A review," Sustainable Operations and Computers, **3**(May), pp. 275–285.

[2] Pelser, T. G. and Gaffley, G., 2020, "Implications of digital transformation on the strategy development process for business leaders," *Promoting Inclusive Growth in the Fourth Industrial Revolution*, IGI Global, Chap. 1.

[3] Udayana, G. A., Mahendra, I. M. Y., Sukawirasa, I. K. A., Dimastawan Saputra, G. D., and Mahendra, I. B. M., 2021, "Implementasi Data Warehouse Dan Penerapannya Pada PHI-Minimart Dengan Menggunakan Tools Pentaho dan Power BI," JELIKU (Jurnal Elektronik Ilmu Komputer Udayana), **10**(1), p. 163.

[4] Bpsdm, 2023, "E-learning Kemenkumham," https://e-learning.kemenkumham.go.id/, Accessed: 2024-12-05.

[5] Lokaadinugroho, I., Girsang, A. S., and Burhanudin, B., 2021, "Tableau Business Intelligence Using the 9 Steps of Kimball's Data Warehouse & Extract Transform Loading of the Pentaho Data Integration Process Approach in Higher Education," Engineering, Mathematics and Computer Science (EMACS) Journal, **3**(1), pp. 1–11.

[6] Raminhos, R. and Moura-Pires, 2007, "Extraction and transformation of data from semi-structured text files using a declarative approach," *ICEIS 2007 - 9th International Conference on Enterprise Information Systems*, Vol. DISI, pp. 199–205, doi: 10.5220/0002364201990205.

[7] Ummah, M. S., 2019, "Building Open Source ETL Solutions with Pentaho Data Integration," http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y, Accessed: 2024-12-05.

[8] Celko, J., 2014, *Complete Guide to NoSQL*, Elsevier.

[9] Kwon, I. H., 2014, "Book Review: Process Mining: Discovery, Conformance and Enhancement of Business Processes," Healthcare Informatics Research, **20**(2), p. 157.

[10] Ben Moussa, N. and El Arbi, R., 2020, "The impact of Human Resources Information Systems on individual innovation capability in Tunisian companies: The moderating role of affective commitment," European Research on Management and Business Economics, **26**(1), pp. 18–25.

[11] Università, L., Sociali, S., Guido, C., Roma, and Chiara Benfatto, D. M., 2022, "Dottorato di Sistemi Informativi Aziendali XXII Ciclo Human Resource Information Systems and the performance of the Human Resource Function Tesi di dottorato," Unpublished.