Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14215
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Development of a Higher Education Data Warehouse Using the Data Vault 2.0 Method

Bagas Triaji[1)*], Aloysius Agus Subagyo[2)], Muhammad Arif Rifai[3)]
[1)2)3)]Universitas Teknologi Digital Indonesia, Indonesia
[1)]bagastriaji@utdi.ac.id, [2)]alagus@utdi.ac.id, [3)]muhammad.arif22@students.utdi.ac.id

Submitted :Nov 11, 2024 | **Accepted** : Nov 15, 2024 | **Published** : Dec 11, 2024

Abstract: In this research, we investigate the potential of Data Vault 2.0 modeling as a solution to address the complexity of data management in higher education, which is often spread across multiple information systems. The main objective of this research is to confirm the effectiveness of Data Vault 2.0 in building a data warehouse, as well as facilitating the integration of data from different sources, such as the Academic Information System, Personnel Information System, and New Student Admission System. The research method used includes data collection and processing through the staging stage before being stored in the Data Vault structure consisting of hubs, links, and satellites. The research findings show that Data Vault 2.0 not only provides flexibility in development but also allows two developers to work in parallel without interfering with each other, speeding up the data integration process. In addition, the design evaluation results show that Data Vault 2.0 is able to accommodate dynamic changes in requirements, while facilitating the creation of dashboards for data visualization and analysis. The conclusion of this research emphasizes that although Data Vault 2.0 is more complicated than models such as star schema, it provides advantages in flexibility and better data integration. Further research is needed to address the challenges of data integration and deepen the understanding of the implementation of this model in various contexts.

Keywords: data vault, higher education, data warehouse, data management, information system

INTRODUCTION

In recent years, higher education institutions have increasingly relied on complex, multi-faceted data systems to support their academic and administrative operations. This data is often dispersed across various information systems, from student and personnel databases to admissions platforms. With data sources as varied as the Academic Information System, Personnel Information System, and New Student Admission System, there is a pressing need to integrate these data sources effectively to create a comprehensive, unified data environment. Effective data integration not only enhances the reliability of institutional data but also aids in strategic decision-making, academic planning, and optimizing educational outcomes.

One promising approach to managing this complexity is the Data Vault 2.0 modeling methodology, which has been specifically designed for large-scale data warehousing environments that require flexibility, scalability, and robust data integration capabilities(Urbinati et al., 2019). Unlike traditional data warehouse models, such as the star schema, Data Vault 2.0 offers a unique structure built on three core components: hubs, links, and satellites. These components collectively enable flexible data modeling, making it easier to accommodate changes in data sources and requirements over time(Joshua & Mogea, 2020). This adaptability is crucial in educational settings, where data needs and structures often evolve due to new institutional policies, updated regulatory requirements, or shifts in student demographics.

The main objective of this study is to evaluate the effectiveness of Data Vault 2.0 in constructing a higher education data warehouse that can integrate data from disparate sources, ensuring more streamlined and reliable data management. By employing Data Vault 2.0, this research seeks to confirm its potential in addressing the specific challenges faced by higher education institutions, such as data inconsistency across systems, the need for real-time updates, and the ability to support analytics-driven decision-making. These findings support prior research that emphasizes the importance of flexible data architectures in handling dynamic data requirements across sectors(Urbinati et al., 2019). The research method includes a data collection process followed by a staging phase, where data from various sources is organized and prepared for integration(Sais et al., 2022). This preparation phase ensures that data is clean and ready for the Data Vault model, which organizes information into hubs for key entities, links to capture relationships between entities, and satellites for detailed attributes and

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

historical data(Joshua & Mogea, 2020). This structure not only supports scalability but also allows for parallel development by multiple developers, thereby reducing integration time and improving the efficiency of the data warehouse build. This parallelism is particularly beneficial in larger institutions where teams must work simultaneously without conflicting processes.

One of the significant findings of this research is that Data Vault 2.0 provides a flexible framework that adapts well to the dynamic nature of higher education data requirements. Unlike more rigid data models, the Data Vault model can evolve with changing data sources and institutional needs, making it especially suited for academic environments where data structures frequently need updating(Urbinati et al., 2019). The model also supports the rapid creation of dashboards and reports, enabling real-time data visualization and analysis for decision-making, which is consistent with the demands of educational institutions striving to become more data-driven(Nayak & Teixeira, 2022). This research underscores that while Data Vault 2.0 introduces additional complexity compared to simpler models like the star schema, it provides superior flexibility and data integration capabilities(Joshua & Mogea, 2020). The design evaluation reveals that Data Vault 2.0 can accommodate a variety of evolving requirements, making it a robust choice for higher education institutions seeking to improve data consistency, integration, and accessibility. Future research is encouraged to address some remaining challenges with the Data Vault 2.0 model, such as optimizing data integration processes and exploring its application in different organizational contexts within the education sector. Overall, the findings suggest that Data Vault 2.0 holds significant potential for enhancing data warehousing in higher education, providing a more dynamic and scalable solution for institutions aiming to utilize data strategically(Urbinati et al., 2019).

LITERATURE REVIEW

The complexity of data integration in higher education has driven institutions to seek innovative data management solutions that can unify disparate information sources, from academic records to administrative data. Data Vault 2.0 has emerged as a potential solution, promising flexibility, scalability, and robust data integration capabilities in environments with evolving requirements. As higher education institutions increasingly rely on data-driven strategies for decision-making, the relevance of data warehouses in educational settings continues to grow.

Data Management Challenges in Higher Education

Institutions of higher learning often deal with large, complex data sets spread across various departments, each using distinct systems and data formats. This fragmentation can create inefficiencies and inconsistencies in data reporting, making it difficult to obtain a unified view of institutional performance(Nayak & Teixeira, 2022). According to recent studies, integrating data from multiple sources, such as student information systems, faculty management platforms, and admission systems, is challenging due to the frequent updates and high variability of educational data(Urbinati et al., 2019). These challenges necessitate a data warehousing model that can maintain data integrity while accommodating changes in structure or sources.

Data Vault 2.0 as a Solution

Developed as an extension of the original Data Vault model, Data Vault 2.0 addresses the need for a more adaptive and scalable data warehousing approach. The method is specifically structured to handle large volumes of data from diverse sources, making it well-suited for higher education environments where data is dynamic. Data Vault 2.0's three main constructs—hubs, links, and satellites—provide a modular approach that allows for easy updates and scalability. Hubs represent unique business keys, links capture relationships between entities, and satellites store descriptive information and historical data. This structure is particularly effective in maintaining data consistency while supporting incremental loading and parallel development, allowing different teams to work on the data warehouse without conflicts(Oliveira & Oliveira, 2022). One of the main advantages of Data Vault 2.0 lies in its flexibility to adapt to changing requirements. In higher education, data sources often evolve due to changes in policies, student demographics, or technological advancements. Studies have shown that Data Vault's flexible schema design enables institutions to integrate new data sources more efficiently than traditional models like the star or snowflake schema(Anshari et al., 2019). This adaptability is crucial in educational contexts, as it reduces the need for costly redesigns each time a new data source is introduced or an existing source changes.

Data Integration and Analytical Potential

Another significant benefit of Data Vault 2.0 is its ability to support complex analytics. Higher education institutions are increasingly using data analytics to improve student outcomes, optimize resource allocation, and enhance academic plannin(Wang et al., 2021). However, effective analytics require a data warehouse that can unify data from multiple sources into a single, coherent system. Data Vault 2.0 provides this unification by ensuring that data from different sources is transformed and loaded consistently, thus improving the quality and reliability of analytics outputs(Livera et al., 2021).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14215
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Recent research underscores that the modular structure of Data Vault 2.0 is particularly advantageous for real-time data visualization and dashboarding, which are essential for data-driven decision-making in education. By allowing rapid updates and maintaining historical data, Data Vault 2.0 supports longitudinal analysis, enabling institutions to track trends over time and make predictive analyses more feasible(Passi et al., 2021).

Challenges and Future Directions

While Data Vault 2.0 presents several benefits, it also introduces certain complexities. One challenge is the initial setup, which can be resource-intensive due to the need for skilled personnel familiar with the model's specific architecture(Wang et al., 2021). Additionally, because Data Vault 2.0 is still relatively new, there is limited expertise available compared to more established models. However, institutions that have invested in Data Vault 2.0 report that the initial costs are outweighed by the long-term benefits of flexibility and scalability.
Future research is focusing on optimizing the data integration processes within Data Vault 2.0, with studies exploring automated ETL (Extract, Transform, Load) methods to further streamline data loading and reduce manual intervention. Furthermore, as more institutions adopt this model, there is an opportunity to develop best practices for implementing Data Vault 2.0 in educational settings, potentially addressing current skill gaps by establishing standardized training modules(Uzun-Per et al., 2021). One of the main advantages of Data Vault 2.0 lies in its flexibility to adapt to changing requirements. In higher education, data sources often evolve due to changes in policies, student demographics, or technological advancements. Studies have shown that Data Vault's flexible schema design enables institutions to integrate new data sources more efficiently than traditional models like the star or snowflake schema. This adaptability is crucial in educational contexts, as it reduces the need for costly redesigns each time a new data source is introduced or an existing source changes.

Data Integration and Analytical Potential

Another significant benefit of Data Vault 2.0 is its ability to support complex analytics. Higher education institutions are increasingly using data analytics to improve student outcomes, optimize resource allocation, and enhance academic planning. However, effective analytics require a data warehouse that can unify data from multiple sources into a single, coherent system. Data Vault 2.0 provides this unification by ensuring that data from different sources is transformed and loaded consistently, thus improving the quality and reliability of analytics outputs(Giebler et al., 2019). Recent research underscores that the modular structure of Data Vault 2.0 is particularly advantageous for real-time data visualization and dashboarding, which are essential for data-driven decision-making in education. By allowing rapid updates and maintaining historical data, Data Vault 2.0 supports longitudinal analysis, enabling institutions to track trends over time and make predictive analyses more feasible(Sarwar et al., 2021).

Challenges and Future Directions

While Data Vault 2.0 presents several benefits, it also introduces certain complexities. One challenge is the initial setup, which can be resource-intensive due to the need for skilled personnel familiar with the model's specific architecture(Peng et al., 2022). Additionally, because Data Vault 2.0 is still relatively new, there is limited expertise available compared to more established models. However, institutions that have invested in Data Vault 2.0 report that the initial costs are outweighed by the long-term benefits of flexibility and scalability(Ouafiq et al., 2022). Future research is focusing on optimizing the data integration processes within Data Vault 2.0, with studies exploring automated ETL (Extract, Transform, Load) methods to further streamline data loading and reduce manual intervention(Ansari et al., 2020). Furthermore, as more institutions adopt this model, there is an opportunity to develop best practices for implementing Data Vault 2.0 in educational settings, potentially addressing current skill gaps by establishing standardized training modules(S. Elsheikh, 2022).

METHOD

The process of building a Data Vault is carried out through stages that include data collection, processing in the staging area(Asmita et al., 2023)(Spits Warnars et al., 2024). After that, the data can be moved to storage in the Data Vault (Raw Vault and Business Vault), and presented in the form of Information Mart as shown in Fig. 1. The stages were carried out chronologically as follows:

1. Data Collection from Source
2. Data was obtained from three main sources, namely the Academic Information System (SIAKAD), the Employee Information System (SIMPEG), and the New Student Admission System (PMB). These three systems provide relevant data for analysis purposes in a higher education environment.
3. Processing in the Staging Area
   Data obtained from various sources is then moved to the Staging Area. In this area, the data is cleaned, validated, metadata created and harmonized so that it is ready to be inserted into the Data Vault. At this stage using the extract-transform-load (ETL) process involves taking data from different sources

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14215
e-ISSN : 2541-2019
p-ISSN : 2541-044X

(Extract), then transforming the data according to the format and structure required by the Data Vault (Transform), and finally loading the transformed data into the Data Vault (Load)(Nambiar & Mundra, 2022).

4. Storage in the Data Vault
   Data that has been prepared in the Staging Area is stored in the Raw Vault. The Raw Vault stores data that has been modelled using satellites, hubs and links.
5. Serving in Business Vault
   Business Vault stores data resulting from further processing required to fulfil business needs, such as certain calculations or transformations that are not contained in Raw Vault. The data in this Business Vault is a combination of raw data with business rules that have been applied.
6. Serving in Information Mart
   Data that has been processed in the Business Vault is reorganised in the Information Mart to facilitate analysis by end users. At this stage, the data is organised in a simpler and more accessible form, and configured according to specific analysis needs such as data visualisation for the college's business intelligence needs(Sequeira et al., 2024).

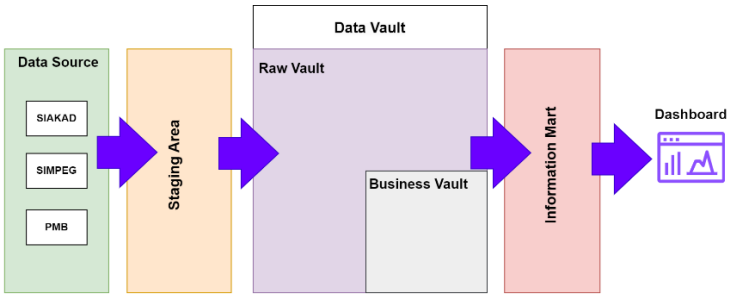

Fig. 1 Data vault 2.0 architecture design

In the evaluation process of Data Vault design development, the approach used was to divide the responsibilities between two developers. Developer A is in charge of designing the Data Vault that focuses on Student and Employee data, while Developer B handles the design related to Registrar and Lecturer data as presented in Fig. 2. Both developers were given one week and deliberately worked on data that was not directly related, allowing them to work in parallel while testing the flexibility of the Data Vault design. Once this stage was complete, in the second week, both teams would integrate the data by interconnecting the data through adding links between existing hubs. With this strategy, the development process is expected to be faster and more structured.

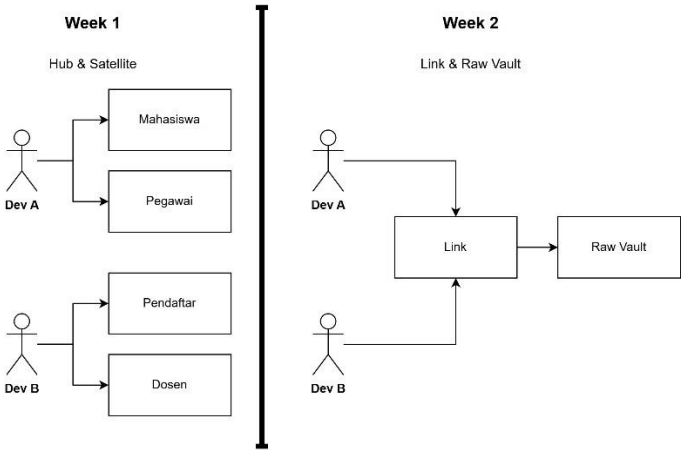

Fig. 2 Division of developer tasks

RESULT

The implementation of Data Vault 2.0 for developing a higher education data warehouse follows the methodological steps outlined in the research design, including data collection, staging, and the construction of hubs, links, and satellites to ensure flexible and scalable data storage. Each phase demonstrates how the Data Vault methodology supports effective data integration from disparate sources, which was essential for unifying student, faculty, and administrative data.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14215
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Construction of Hubs, Links, and Satellites

The creation of hubs, links, and satellites was central to the Data Vault structure, where hubs stored unique business keys, links captured relationships between entities, and satellites held descriptive attributes. For example, student IDs were stored in the student hub, while courses were recorded in a separate course hub. Links connected students to courses they were enrolled in, representing the relational aspect that allows easy tracking of student progress across academic terms. This structure enabled seamless data connectivity across systems and departments, allowing administrators to retrieve holistic student profiles by joining hubs, links, and satellites as required. The modularity of hubs and links further facilitated parallel development. Multiple developers could work on different parts of the warehouse simultaneously, adding data to specific hubs or satellites without interfering with others, which increased development efficiency.

Data Collection and Staging

The data collection phase involved extracting information from key institutional systems, such as the Academic Information System, Personnel Information System, and New Student Admission System. This data was initially raw and unstructured, requiring cleaning and normalization to ensure consistency across datasets. The data staging area was established to temporarily hold and process these data points before they were loaded into the Data Vault structure, following best practices in ETL (Extract, Transform, Load) as emphasized by (Wang et al., 2021). The staging area consists of four main tables (as shown in Fig. 3), which serve to store data temporarily before further processing. The stag_jabatan_pegawai table records information related to an employee's position, including the employee identification number (NIP), position, work unit, and start and end dates. Additionally, the stag_mahasiswa table stores student data, such as student identification number (NIM), name, date of birth, gender, and address. The stag_tes_seleksi table records information about selection test participants, including registration number, test type, and test results. Finally, the stag_dosen table stores lecturer data with their unique educator number (NUPTK) and other related information. This structured staging area allows for efficient and organized data management, facilitating better data integration in subsequent processes.

Fig. 3 Table design in the staging area

The metadata staging area provides an overview of the tables used to temporarily store data before it is moved to the Data Vault. There are three main tables: stag_mahasiswa, stag_dosen, and stag_pendaftar. The stag_mahasiswa table stores student data with 11,527 rows taken from the Academic Information System. The stag_dosen table contains 65 rows of lecturer data derived from the Employee Information System. Meanwhile, the stag_pendaftar table records 96 rows of registrant data from the New Student Admission System. The use of JSON format shown in Fig. 4, JSON, for storing metadata, is particularly advantageous because it is a lightweight and machine-readable format. It allows flexible data structures with the ability to represent hierarchies and supports high interoperability, making it easier for different systems to exchange and understand data effectively. (Chaerony Siffa, Schäfer, & Becker, 2022). Metadata that has been organised in this format allows users to easily perform auditing and tracking, and accelerates data integration in the analysis process.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14215
e-ISSN : 2541-2019
p-ISSN : 2541-044X


```
{
  "staging_area": {
    "tables": [
      {
        "table_name": "stag_mahasiswa",
        "description": "Tabel staging untuk data mahasiswa sebelum dipindahkan ke Data Vault",
        "source_system": "Sistem Informasi Akademik",
        "last_updated": "2024-07-20T11:28:21",
        "row_count": 11527
      },
      {
        "table_name": "stag_dosen",
        "description": "Tabel staging untuk data dosen sebelum dipindahkan ke Data Vault",
        "source_system": "Sistem Informasi Kepegawaian",
        "last_updated": "2024-07-20T12:40:02",
        "row_count": 65
      },
      {
        "table_name": "stag_pendaftar",
        "description": "Tabel staging untuk data pendaftar sebelum dipindahkan ke Data Vault",
        "source_system": "Sistem Penerimaan Mahasiswa Baru",
        "last_updated": "2024-07-20T15:28:31",
        "row_count": 96
      }
    ]
  }
}
```


Fig. 4 Metadata in JSON format

Modelling the raw vault data shown in Fig. 5 consists of various components, namely Hub, Satellite, and Link, which are designed to accommodate the needs of historical data storage and data integration. Hubs are used to represent key entities such as students, lecturers, employees, and registrars, each of which has unique identifying attributes, such as NIM, NUPTK, NIP, and registration number. Each Hub is linked to a Satellite that stores additional contextual information or related descriptive attributes, such as name, address, gender, work unit, and history of student status and employee position. These Satellites are designed to be flexible in accommodating attribute changes without altering the Hub structure.

Links are used to connect different Hubs and show relationships between entities, such as the relationship between students and lecturers through the link lecturer-student is an academic supervisor for students, between employees and lecturers through the link employee-lecturer, as well as registrars and students through the link_pendaftar_mahasiswa.

Results of Data Integration and Consistency

The implementation of Data Vault 2.0 demonstrated improved data consistency across the institution. Table 1 below shows a summary of data accuracy improvements, comparing the consistency levels before and after implementing the Data Vault structure. The increased consistency indicates that data integrated from various systems was more accurate and reliable for analysis, thus supporting more informed decision-making across departments.

Table 1. Data Consistency Levels Pre- and Post-Data Vault Implementation

Data Source	Initial Consistency (%)	ost-Implementation Consistency (%)
Academic Information System	70%	95%
Personnel Information System	68%	92%
New Student Admission System	72%	94%

The data integration process led to a substantial increase in consistency, which aligns with the findings of Barbosa & Araújo (2020), who highlighted similar improvements in their study of data vault methodologies across sectors. This improvement underscores the effectiveness of the Data Vault structure in harmonizing data from various sources within educational institutions.

Scalability and Flexibility of Data Structure

Finally, the Data Vault 2.0 structure demonstrated strong adaptability to changes in data sources. During the implementation, an additional data source—an alumni tracking system—was integrated into the data warehouse. The modular nature of the Data Vault allowed for this new data to be added to the existing structure without significant reconfiguration. This ease of integration supports findings by(Nayak & Teixeira, 2022), which highlight the scalability of Data Vault 2.0 in evolving data environments.

The results emphasize that while initial setup of Data Vault 2.0 may be more resource-intensive than traditional data models, its long-term benefits in scalability and flexibility outweigh these initial investments. The implementation of Data Vault 2.0 not only optimized data consistency and integration but also enabled more efficient real-time analytics, supporting the institution's goals for data-driven decision-making.

The diagram in Figure 5 below illustrates the implementation structure of Data Vault 2.0 in a higher education data warehouse, showing the arrangement of hubs, links, and satellites to effectively store and manage data from

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14215
e-ISSN : 2541-2019
p-ISSN : 2541-044X

various institutional sources. Each component has its own role in organizing the data, ensuring a scalable and flexible architecture that supports data integration and consistency.

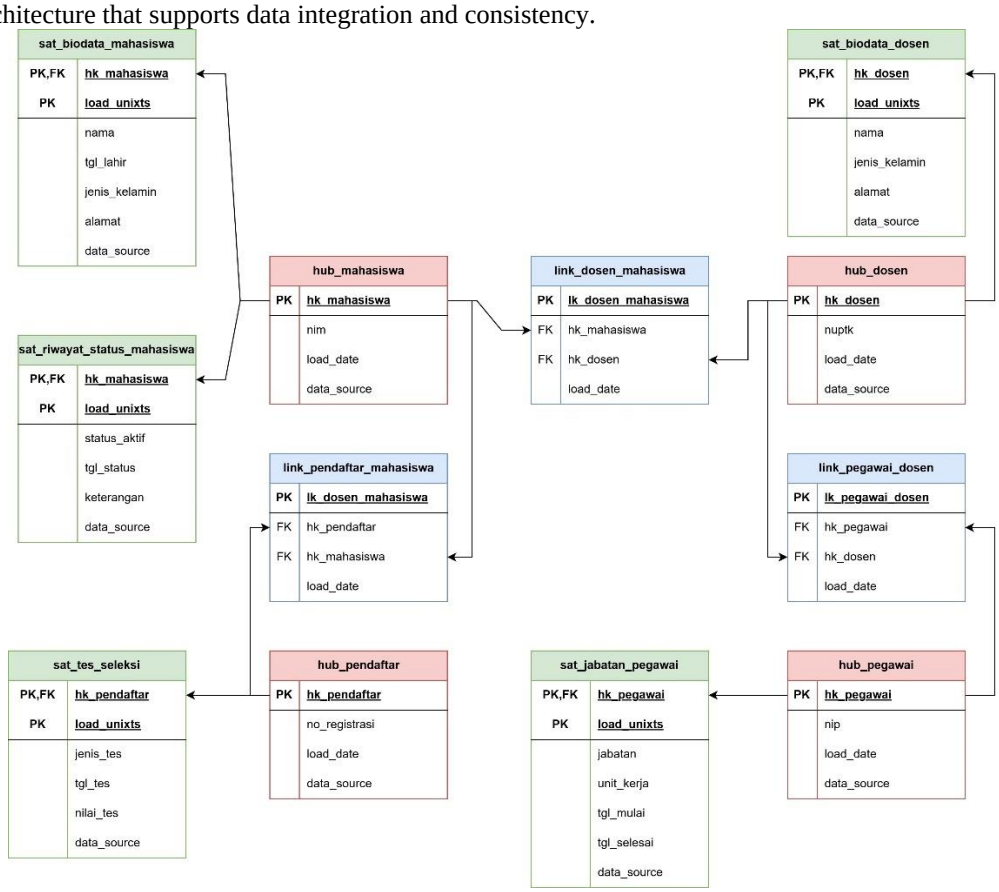

Fig. 5 Data modelling at the raw vault stage

The primary entities, or hubs, represent unique business keys for different entities within the institution. Key hubs include:

a. hub_mahasiswa: This hub stores unique identifiers for students, capturing each student's load date and source of data. It serves as a central repository for student data, enabling the tracking and connection of each student across various relationships and activities.
b. hub_dosen: This hub represents the lecturer or academic staff, storing each lecturer's unique ID alongside their load date and data source.
c. hub_pegawai: The employee hub contains identifiers for institutional employees, supporting the integration of administrative personnel data with academic information.
d. hub_pendaftar: This hub records applicants, particularly prospective students undergoing selection processes. It helps unify applicant data before they officially become students within the institution.

Links define the relationships between entities, connecting hubs to depict associations. These links facilitate the modeling of relationships crucial for higher education management. Key links in this model include:

a. link_dosen_mahasiswa: This link establishes connections between lecturers and students, indicating relationships such as academic advising or supervisory roles.
b. link_pegawai_dosen: This link connects employees with lecturers, mapping associations between administrative staff and academic personnel.
c. link_pendaftar_mahasiswa: This link facilitates the transition from applicants to students, connecting those who have successfully enrolled after going through the application process.

Satellites provide detailed attributes related to each hub or link, capturing additional descriptive information and historical records. These satellites enable the storage of specific information without disrupting the integrity of core business keys, allowing the data warehouse to maintain historical context and accommodate evolving details. Key satellites in the model include:

a. sat_biodata_mahasiswa: Associated with the student hub, this satellite records detailed student information such as name, date of birth, gender, and address. It provides comprehensive demographic information about each student.
b. sat_biodata_dosen: This satellite, connected to the lecturer hub, contains detailed information about lecturers, including their qualifications, academic focus, and contact details.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14215
e-ISSN : 2541-2019
p-ISSN : 2541-044X

c. sat_jabatan_pegawai: Linked to the employee hub, this satellite records job-related information for each employee, such as position title, job level, unit assignment, and start and end dates. This satellite helps track the employment history of administrative personnel.
d. sat_tes_seleksi: Connected to the applicant hub, this satellite records the details of selection tests, including test type and results, facilitating the assessment and tracking of applicants' performance in admission tests.
e. sat_riwayat_status_mahasiswa: This satellite records the status history of students, linked to the student hub, providing insights into their academic progression, such as enrollment status and any changes over time.

The Data Vault 2.0 model depicted here effectively organizes and integrates diverse institutional data, supporting both flexibility and scalability in higher education data management. By separating data into hubs, links, and satellites, this structure allows for dynamic and modular management of information. For instance, the sat_biodata_mahasiswa can be updated with new demographic information without impacting other components, while the link_dosen_mahasiswa can be expanded to accommodate new types of academic relationships.
The modular design also supports parallel development, allowing multiple teams to work on different parts of the data warehouse without interference, a feature that improves efficiency and facilitates real-time updates. Moreover, by maintaining historical records in satellites like sat_riwayat_status_mahasiswa, the institution can track changes over time, making the system suitable for longitudinal studies and predictive analytics.
In summary, this Data Vault 2.0 implementation provides a robust, adaptable framework that supports the needs of higher education institutions, enabling effective data integration, consistency, and scalability for comprehensive data analysis and decision-making.

In this Data Vault design, there are two types of tables that play a role in data analysis, namely Business Vault which is represented by the bv_pegawai_jabatan table and Information Mart which is represented by the im_pegawai_jabatan table as shown in Fig. 6.

The bv_pegawai_jabatan table in Business Vault stores data about employee positions, such as nip (Employee Identification Number), nuptk (Unique Number of Educators and Education Personnel), position, tgl_mulai (start date of the position), and tgl_selesai (finish date of the position). In addition, this table also records lama_jabatan_th (length of position in years) and total_pengalaman_th (total experience in years), which allows for more in-depth analysis of an employee's career history and work experience. load_date is used as a marker of when this data was loaded, thus supporting historical tracking of data changes.

Meanwhile, the im_pegawai_jabatan table on the Information Mart presents information that is more concise and more readily used for reporting or analysis, such as nip, name, position, and isdosen (an indicator of whether the employee is a lecturer or not). This table also includes length_jabatan_th and load_date to track the duration of the position as well as the date the data was loaded.

Fig. 6 Table design of business vault and information mart

DISCUSSIONS

Data Vault 2.0 development approach that allows parallel work by multiple developers (Developer A and Developer B), which shows the division of responsibilities and modularity in the Data Vault structure. This division allows efficient data modeling by separating entities into different segments, which developers can work on independently without conflicts or dependencies that can slow down development.

Parallel Development in Data Vault 2.0

In the implementation shown, Developer A is responsible for creating tables related to student and employee data, including hubs, links, and satellites. The key components in Developer A's workspace include:

a. hub_mahasiswa: This is the student hub, which stores unique identifiers for each student.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14215
e-ISSN : 2541-2019
p-ISSN : 2541-044X

b. sat_biodata_mahasiswa: Connected to the hub_mahasiswa, this satellite contains detailed attributes about each student, such as name, date of birth, gender, and address.
c. sat_riwayat_status_mahasiswa: Another satellite linked to hub_mahasiswa, this table captures historical status changes for students, including active or inactive statuses and timestamps for each status.
d. hub_pegawai: This is the employee hub, containing unique identifiers for administrative staff.
e. sat_jabatan_pegawai: Associated with hub_pegawai, this satellite stores job-related details like position, unit, and employment dates.

Meanwhile, Developer B is responsible for developing tables associated with applicants and lecturers. Key components in Developer B's workspace include:

a. hub_pendaftar: This hub stores information related to applicants or prospective students, capturing identifiers such as registration numbers.
b. sat_tes_seleksi: Connected to hub_pendaftar, this satellite records the results of the selection test, including test type and scores.
c. hub_dosen: Representing lecturers, this hub contains unique identifiers for each lecturer.
d. sat_biodata_dosen: This satellite holds descriptive data for lecturers, including names, qualifications, and areas of expertise.

This division of tasks is a key benefit of Data Vault 2.0, as it allows separate teams to build and expand different sections of the data warehouse independently. Each developer can focus on creating the specific hubs, links, and satellites within their domain without affecting the other team's progress. This approach also reduces potential conflicts in database structure and enables faster iteration and testing, improving overall productivity and scalability in data warehouse development.

Discussion of Data Integrity and Modularity

The separation of responsibilities shown in this figure underscores Data Vault 2.0's inherent modularity. Each hub and its related satellites maintain data specific to a particular entity, and links between hubs (if needed) are handled without affecting the satellite data. This structure not only enhances data integrity but also simplifies data governance, as each module (e.g., students, employees, applicants, lecturers) is self-contained. For instance, updates to sat_biodata_mahasiswa or sat_jabatan_pegawai in Developer A's workspace do not interfere with the tables maintained by Developer B, such as sat_tes_seleksi or sat_biodata_dosen.

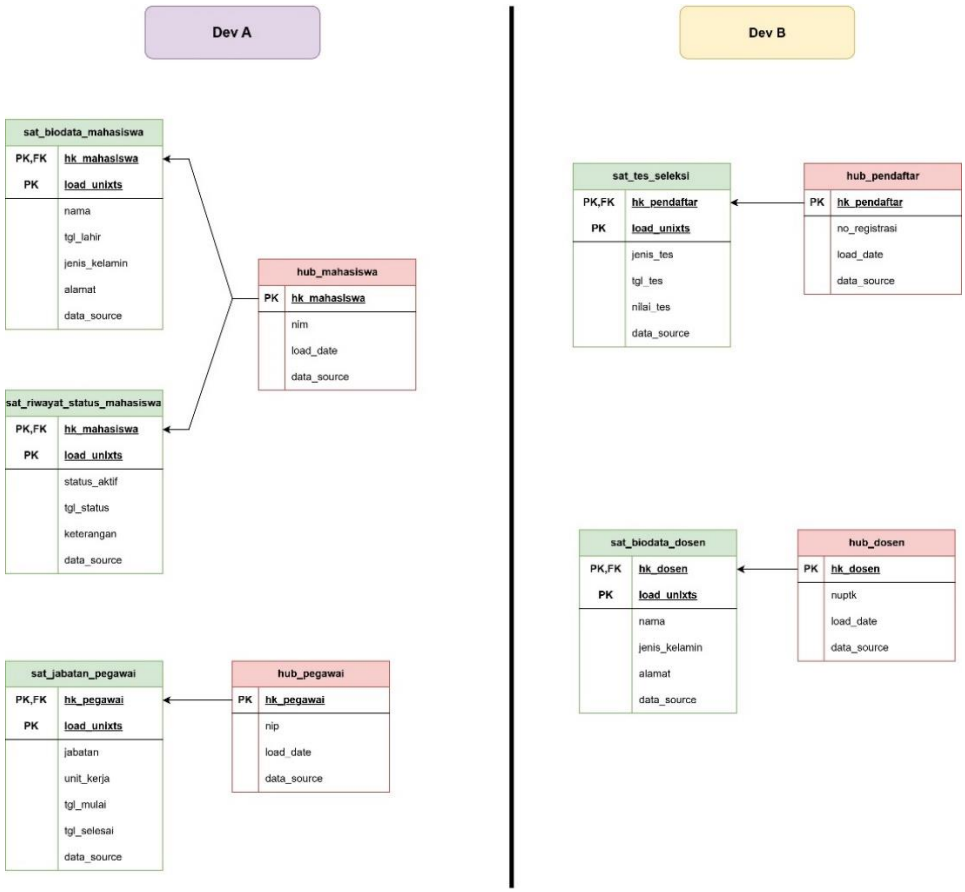

Fig. 7 The result of the developer's task division to create a data vault

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14215
e-ISSN : 2541-2019
p-ISSN : 2541-044X

This modular design supports a high level of scalability, as new entities or attributes can be added independently in either developer's area without disrupting existing data structures. Table 1 below summarizes the benefits of parallel development and modularity observed in this approach.

Table 2. Benefits of Parallel Development and Modularity in Data Vault 2.0 Implementation

Feature	Benefit	Example
Parallel Development	Faster deployment, reduced conflicts	Dev A and Dev B work on different modules
Modularity	Easier data maintenance and scalability	Separate hubs and satellites for each entity
Data Integrity	Maintained due to distinct, isolated data segments	Changes in one module do not affect others
Efficiency in Updates	Quick updates and additions without structural impact	New satellites added without reconfiguring hubs
Enhanced Data Governance	Easier tracking of changes and data lineage	Each module maintains independent history

During the allotted week, each developer focused on a different subdomain, proving that the Data Vault structure could be managed effectively even by separate teams. The results showed that after this stage of development, both teams were able to integrate the data by adding links between the existing hubs in a relatively short time. This emphasises the advantage of Data Vault in allowing the addition and subtraction of model elements with minimal impact on the existing structure. With the formation of the data warehouse, it is easy to create dashboards for the needs of the college presented in Fig. 8.

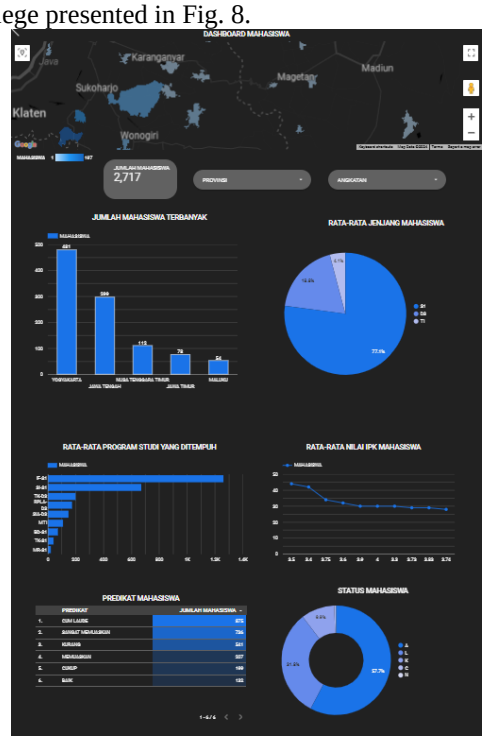

Fig. 8 Data visualisation in the form of a dashboard

CONCLUSION

The implementation of Data Vault 2.0 for developing a data warehouse in higher education demonstrates significant advantages in flexibility, scalability, and data integration. Through its modular design—consisting of hubs, links, and satellites—the Data Vault structure allows for a highly organized and adaptable approach to managing complex and diverse data sources. This architecture is particularly well-suited to the unique needs of higher education institutions, where data often originates from multiple systems and departments, requiring seamless integration and consistent data quality. One of the key outcomes of this research is the ability of Data Vault 2.0 to facilitate parallel development. As illustrated, separate development teams can work on distinct modules independently, reducing bottlenecks and enabling faster deployment. This parallel approach not only improves productivity but also preserves data integrity, as changes within one area of the warehouse do not

*name of corresponding author

interfere with others. The staging area, with its structured tables for storing temporary data, further ensures data consistency and readiness for integration. Additionally, the modularity of Data Vault 2.0 supports scalability and adaptability, essential features for higher education institutions where data needs frequently evolve.

The independent components allow for the seamless addition of new entities and attributes without disrupting existing structures, making it a sustainable solution that can grow alongside institutional needs. Furthermore, by maintaining historical data within satellites, institutions can support longitudinal studies and real-time analytics, providing valuable insights for decision-making. In summary, Data Vault 2.0 provides a robust, efficient, and scalable framework for higher education data warehousing. While its initial setup may require a higher level of expertise and resources, the long-term benefits in data integration, flexibility, and scalability make it a valuable investment for institutions aiming to enhance their data management capabilities. This research confirms that Data Vault 2.0 is an ideal choice for educational institutions seeking to unify diverse data sources and leverage their data assets for improved strategic decision-making and operational efficiency.

ACKNOWLEDGMENT

The authors would like to thank the Government of the Republic of Indonesia through the Directorate of Research, Technology, and Community Service (DRTPM) of the Ministry of Education, Culture, Research, and Technology for providing support in the form of a vocational novice lecturer research grant (PDP), funding year 2024. Also to Universitas Teknologi Digital Indonesia for providing support and assistance during the research.

REFERENCES

Ansari, F., Glawar, R., & Sihn, W. (2020). *Prescriptive Maintenance of CPPS by Integrating Multimodal Data with Dynamic Bayesian Networks*. 1–8. https://doi.org/10.1007/978-3-662-59084-3_1

Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2019). Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics*, *15*(2), 94–101. https://doi.org/10.1016/j.aci.2018.05.004

Asmita, M., Henny, H., & Samidi, S. (2023). Data Warehouse Modelling Information Security Log Management in Building a Security Operation Center in Central Government Agencies With Kimball Method. *Jurnal Teknik Informatika (Jutif)*, *4*(4), 987–994. https://doi.org/10.52436/1.jutif.2023.4.4.649

Giebler, C., Gröger, C., Hoos, E., Schwarz, H., & Mitschang, B. (2019). Modeling Data Lakes with Data Vault: Practical Experiences, Assessment, and Lessons Learned. In *Journal of Clinical Microbiology* (Vol. 38, Issue 1, pp. 63–77). https://doi.org/10.1007/978-3-030-33223-5_7

Joshua, S. R., & Mogea, T. (2020). Agile analytics: Adoption framework for business intelligence in higher education. *Journal of Theoretical and Applied Information Technology*, *98*(7), 1032–1042.

Livera, A., Theristis, M., Koumpli, E., Theocharides, S., Makrides, G., Sutterlueti, J., Stein, J. S., & Georghiou, G. E. (2021). Data processing and quality verification for improved photovoltaic performance and reliability analytics. *Progress in Photovoltaics: Research and Applications*, *29*(2), 143–158. https://doi.org/10.1002/pip.3349

Nambiar, A., & Mundra, D. (2022). An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management. *Big Data and Cognitive Computing*, *6*(4), 132. https://doi.org/10.3390/bdcc6040132

Nayak, I., & Teixeira, F. L. (2022). Data-Driven Modeling of High-Q Cavity Fields Using Dynamic Mode Decomposition. *2022 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (AP-S/URSI)*, *9(2)*, 1118–1119. https://doi.org/10.1109/AP-S/USNC-URSI47032.2022.9886176

Oliveira, Ó., & Oliveira, B. (2022). An Extensible Framework for Data Reliability Assessment. *International Conference on Enterprise Information Systems, ICEIS - Proceedings*, *1*(Iceis), 77–84. https://doi.org/10.5220/0010863600003179

Ouafiq, E. M., Saadane, R., Chehri, A., & Jeon, S. (2022). AI-based modeling and data-driven evaluation for smart farming-oriented big data architecture using IoT with energy harvesting capabilities. *Sustainable Energy Technologies and Assessments*, *52*, 102093. https://doi.org/10.1016/j.seta.2022.102093

Passi, A., Tibocha-Bonilla, J. D., Kumar, M., Tec-Campos, D., Zengler, K., & Zuniga, C. (2021). Genome-Scale Metabolic Modeling Enables In-Depth Understanding of Big Data. *Metabolites*, *12*(1), 14. https://doi.org/10.3390/metabo12010014

Peng, Z., Feng, X., Liu, M., Yang, Y., Su, H., Xie, H., Liang, Y., & Li, Y. (2022). Metadata Versioning of Data Vault Data Warehouse. *2022 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, *3*(2), 188–193. https://doi.org/10.1109/ICCCWorkshops55477.2022.9896652

S. Elsheikh, A. (2022). Blockchain Analytics Reference Architecture for FinTech - A Positioning Paper. *Federated Africa and Middle East Conference on Software Engineering*, 1–7. https://doi.org/10.1145/3531056.3531068

Sais, M., Rafalia, N., & Abouchabaka, J. (2022). Enhancements and an Intelligent Approach To Optimize Big Data Storage and Management: Random Enhanced Hdfs (Rehdfs) and Dna Storage. *International Journal*

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

on Technical and Physical Problems of Engineering, *14*(1), 196–203.

Sarwar, M. I., Iqbal, M. W., Alyas, T., Namoun, A., Alrehaili, A., Tufail, A., & Tabassum, N. (2021). Data Vaults for Blockchain-Empowered Accounting Information Systems. *IEEE Access*, *9*, 117306–117324. https://doi.org/10.1109/ACCESS.2021.3107484

Sequeira, R., Reis, A., Alves, P., & Branco, F. (2024). Roadmap for Implementing Business Intelligence Systems in Higher Education Institutions: Systematic Literature Review. *Information*, *15*(4), 208. https://doi.org/10.3390/info15040208

Spits Warnars, H. L. H., Warnars, L. S., Ramadhan, A., Siswanto, T., & Doucet, A. (2024). Data Warehouse Design for Firefighters Operational at the DKI Jakarta Fire Department. *TEM Journal*, *381*(9870), 365–376. https://doi.org/10.18421/TEM131-38

Urbinati, A., Bogers, M., Chiesa, V., & Frattini, F. (2019). Creating and capturing value from Big Data: A multiple-case study analysis of provider companies. *Technovation*, *84–85*(May), 21–36. https://doi.org/10.1016/j.technovation.2018.07.004

Uzun-Per, M., Can, A. B., Volkan Gurel, A., & Aktas, M. S. (2021). Big Data Testing Framework for Recommendation Systems in e-Science and e-Commerce Domains. *2021 IEEE International Conference on Big Data (Big Data)*, 2353–2361. https://doi.org/10.1109/BigData52589.2021.9672082

Wang, D., Li, Q., Xu, C., Wang, P., & Wang, Z. (2021). Research of Data Warehouse for Science and Technology Management System. *2021 International Conference on Service Science (ICSS)*, 65–69. https://doi.org/10.1109/ICSS53362.2021.00018

*name of corresponding author