University of Essex

Department of Computing

Deciphering Big Data 2023

Executive Summary

Date: 27/05/2024

Word Count: 1972

Database Proposal for HR Hiring Process

Executive Summary

1. Introduction

The purpose of this executive summary is to communicate the completed design and build of a single logical database that streamlines the hiring process for the Human Resources (HR) department at the request of the organization. This database ensures that the organization is compliant with legal requirements such as the GDPR and makes it easy for HR employees to maintain data integrity. A robust data system specially created for the hiring process such as this is essential for optimizing management of employee and candidate data and will help the organization make strategic decisions to achieve their goals.

The project involved various phases of research, design, and development. Research was done to understand current HR data management practices and the structure of HR databases. After visualizing the design of the logical flow of the hiring system and understanding what dimensions may be necessary for the database, the design and structure of the database was created. This was done while ensuring that the database would remain accurate and compliant with the organization's standards and protocols by following a data management pipeline.

2. Overview of Project Activities

As the organization requested a single logical database to streamline their HR hiring process, the first step in designing the database was deciding what model of database to create. As the data from the hiring process is structured and there is a need for a high degree of data integrity, consistency, and security, a relational database model best suited the needs of the organization (Pattinson, 2022). Harnessing the power of structured query language (SQL), relational databases allow for complex data retrieval and manipulation, while keeping data organized and accurate (Bryan, 2023).

When it came to what tools to utilize to create this model, Oracle – a database solution with seamless integration with extract, transform, and load tools – was chosen as the database solution due to its effectiveness in handling structured data (Nguyen, 2023). Various diagrams were created to map out the relationships between the data, allowing for a better understanding of the data flow within the HR hiring system. A data management pipeline was also put into place to effectively normalize and transform the data, ensuring that the database could operate successfully, efficiently, and that all data was in the correct format (Antkowiak & Nowaczyk, 2021). Once the necessary dimensions for the database became apparent, a database design and build were created that properly accommodated the tracking of candidates throughout the hiring process, as well as existing employees and their data.

3. Key Project Outcomes

3.1 Overview of Database Design

The final database design features a single relational model with six core dimensions: "CANDIDATE", "EMPLOYEE", JOB_OPENING", "DEPARTMENT", "EE_LEVEL" and "EE_TYPE". These dimensions represent various HR entities, for example the "EMPLOYEE" dimension stores the data and status of an employee in the organization. Further explanation on what each dimension corresponds to can be found in Table 1.

Dimension	Corresponding Data
CANDIDATES	The data and status of a candidate for employment
EMPLOYEES	The data and status of an employee
JOB_OPENING	An available job opening within the organization
DEPARTMENT	All departments within the organization
EE_LEVEL	The job level of an employee, e.g. Entry, Senior, Executive, etc.
EE_TYPE	The type of employment an employee is under, e.g. Full-Time, Part-Time, Contractor, etc.

Table 1. Dimensions and the data they correspond to.

When new information relevant to the database needs to be added, employees will manually input the necessary data. The logical data flow shown in Figure 1 displays the flow of the data within the HR hiring process, and displays when data might be updated, added, or removed from the database. HR employees will have full access to the "CANDIDATE", "DEPARTMENT", "EE_LEVEL", and "EE_TYPE" dimensions, with

read access to the other two dimensions. HR staff will be expected to be able to use SQL to query, manipulate, and retrieve data. Management-level staff will be able to manipulate the "EMPLOYEE" and "JOB_OPENING" dimensions, however all other dimensions will be read access only. No other permissions will be given to other staff. This design ensures that all HR data is easily accessible and well organized.
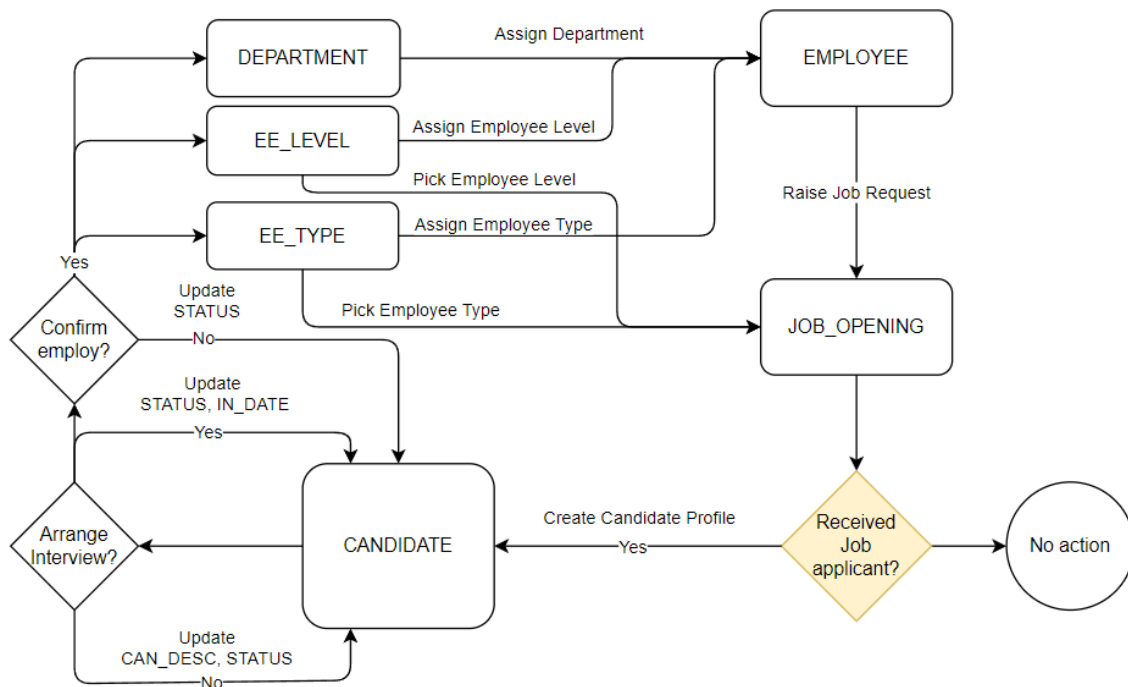

Figure 1. Logical flow of data within the hiring process.

3.2 Data Security and Compliance

As the organization works within the EU/EEA, it is of utmost importance that the database is compliant with the relevant legal requirements, in particular the General Data Protection Regulation (GDPR). The GDPR is a legal framework created by the European Union that sets requirements for the protection of personal data that must be

met by member states of the European Union (Calder, 2020). Personal data is defined as "…any information relating to an identified or identifiable natural person…" by the GDPR (Regulation (EU) 2016/679, 2016). The database enforces individual privileges and permissions, ensuring that only select staff will have access to sensitive HR data. Staff should undergo GDPR training, change passwords regularly, never leave the database open unattended, and be fully informed of the consequences of data breaches.

While restricting permissions to certain staff and only allowing read access to others helps protect data, more needs to be done to properly protect sensitive HR information. Oracle is an amazing tool when it comes to ensuring data security. Encryption is a necessity for proper data protection, and Oracle offers transparent data encryption. The encrypted columns in the database are all columns with personal data, such as names, telephone numbers, addresses, etc. When users input data into encrypted columns, Oracle automatically encrypts the data. When authorized users – users who have a password associated with the master key – select the column, Oracle automatically decrypts the data (Oracle, 2017). This ensures that only select people who know the password will have knowledge of the personal data, and if a third party gets access to the database, they will be unable to decrypt the columns with sensitive data.

Oracle also allows for database auditing, which can track who made changes within the database, when they made changes, and what those changes were. This is compliant with Article 24 of the GDPR, which requires organizations to demonstrate the security measures implemented (Yehuda, 2018). Documenting all database activities

can help the organization if EU officials question the security of the database. Finally, Oracle ensures that if any security incidents take place, the organization will be alerted to the attempt. Once this occurs, the organization can take steps to mitigate the risk of a data breach by changing passwords, restricting access, and looking through the database audits to see whether the database was affected by the breach.

4. Evaluation and Analysis

4.1 Strengths and Weaknesses

The single relational database model chosen offers several strengths. First is the database solution that it is a part of. As Oracle is the database solution, tools such as the Oracle Data integrator, Oracle Business Intelligence, and more can be used to help streamline and visualize the HR hiring process (Oracle, 2021). Oracle also helps ensure that the database is compliant with the GDPR, as seen in section 3.2. Relational databases also have specific benefits over other database models. SQL is a common and relatively simple programming language that allows HR employees to easily add, update, and delete data from the database. The tabular form of the database also makes any changes that are required, such as new columns or rows, simple to add.

While the database model has many strengths, there are a few limitations to consider. First is the number of resources that are required to install this new database. Oracle has many benefits that can be utilized; however, it is a very high-cost system. The organization would also have to potentially hire new employees that understand and can use SQL to query the new database or spend resources training already

available employees in this skill. Ongoing maintenance of the database, as well as regular updates of employee and candidate information, will also require significant amounts of time and expertise. Relational databases also struggle with very large datasets and complex queries, which will be expanded upon in the next section.

4.2 Comparative Analysis with Alternative Data Models

While relational databases are a great way to deal with structured data, and a common solution to HR data management, it may not be the most efficient method to store large quantities of data. Moura (2021) argues that while relational databases are common to use in Human Resources, they are highly inefficient when it comes to querying graph data, the most common way data is stored in HR. Instead of a relational database, a graph database performs much better when executing queries, with Moura stating that graph databases with Cypher queries having "…execution time values 200 times smaller than those obtained in SQL.". While the relational database built for the organization will perform tasks adequately, the performance of the database might suffer if queries become larger and as the amount of data to traverse increases.

Bara et al. (2015) argue that relational databases are a thing of the past, and that the future of human resources is in NoSQL databases due to the faster data transactions. They state "…relational databases lack the necessary features to meet faster data transactions in the big data era". While the database created will be adequate for the foreseeable future for the organization, it would be unwise not to

consider alternatives if the amount of data the organization needs to query starts to impact database performance.

5. Next Steps

For the business's commercial needs, the following needs to take place as soon as possible:

- The finalization of the database implementation.
- Thorough testing to ensure the database system is reliable and up to standard.
- Training HR staff on the new system, potentially hiring or contracting staff that can act as database coordinators.

After these immediate actions are taken, the organization can focus on long-term strategies for continuous improvement and compliance with the GDPR. Training HR staff on the importance of data protection and security, ensuring that passwords are regularly changed every few months, and updating the database to ensure it meets the needs of the organization are all important actions that should be taken and upkept for as long as the database is being used within the organization.

If the organization notices that the relational database is performing slowly, then it may be time to consider using a different form of database that is more suited to the Big Data era, as recommended in section 4.2 (Bara et al., 2015). However, this is unlikely to happen soon, only when the organization has expanded and has a large amount of data that the SQL queries need to traverse, should this switch be considered. While relational databases offer a wide variety of features, many organizations do not make use of them (Jatana et al., 2012). If the organization notices that it is not utilizing

most of the features provided by Oracle, then it is recommended to consider moving the database to an open-source software such as MySQL to save on the high costs Oracle imposes. New actions would need to be taken to ensure the database is still compliant with the GDPR, however.

6. Final Thoughts

The single relational database suits the organization's current needs when it comes to organizing and optimizing the hiring process for the human resources department. The database solution Oracle simplifies the route to compliance with the GDPR, with only a few additional steps needing to be taken by the organization to ensure full compliance with the legal requirements of the EU. While installing the database will require many resources from the organization, a functioning database that is GDPR compliant will be necessary to ensure the organization does not suffer from heavy fines for breaches of data privacy. Although the database might fit with the current needs of the organization, research suggests that relational databases might not be as relevant in the future, and the organization may wish to make a switch to a non-relational database or a graph database for faster querying and better performance. The implementation of a database is essential for optimizing the management of human resource data and will ensure the organization makes strategic decisions based on accurate and reliable data.

References:

Antkowiak M. & Nowaczyk M. (2021) Data cleansing examples. Available from: https://medium.com/transparent-data-eng/data-cleansing-examples-24581c3d14f1 [Accessed 15 April 2024]

Bara, A., Botha, J., Belciu, A. &Nedelcu, B. (2015) Exploring Data in Human Resources Big Data. *Database Systems Journal* VI(3/2015): 3-10. Available from: https://www.dbjournal.ro/archive/21/21_1.pdf [Accessed 20 May 2024]

Bryan, D. (2023) Relational Databases: Unleashing the Power of Structured Data. Available from: https://medium.com/@dezsays/relational-databases-unleashing-the-power-of-structured-data-fcd657740660 [Accessed 15 May 2024]

Calder, A. (2020) *EU GDPR – An international guide to compliance*. Ely: IT Governance Publishing. Available from: https://learning.oreilly.com/library/view/eu-gdpr/9781787782549/ [Accessed 23 May 2024]

GDPR Regulation at: Regulation (EU) 2016/679. European Union. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1532348683434&uri=CELEX:02016R0679-20160504 [Accessed 20 May 2024]

Jatana, N., Puri, S., Ahuja, M., Kathuria, I. & Gosain, D. (2012) A Survey and Comparison of Relational and Non-Relational Database. *International Journal of Engineering Research & Technology* 1(6): 1-5. Available from: https://www.ijert.org/research/a-survey-and-comparison-of-relational-and-non-relational-database-IJERTV1IS6024.pdf [Accessed 11 May 2024]

Moura, R. (2021) Graph Databases for HR Relationships. Available from:

https://repositorio-aberto.up.pt/handle/10216/137426 [Accessed May 24 2024]

Nguyen S. (2023) The Benefits of Oracle DBMS for Your Organization. Available from:

https://blog.dreamfactory.com/the-benefits-of-oracle-dbms-for-your-

organization/#:~:text=This%20popular%20and%20powerful%20relational,a%20s

ecure%20hybrid%20cloud%20environment. [Accessed 11 Apr 2024]

Oracle (2017) 2 Day + Security Guide Release 12.2. Available from:

https://docs.oracle.com/en/database/oracle/oracle-

database/12.2/tdpsg/encrypting-data-with-oracle-transparent-data-

encryption.html#GUID-3D055E66-51CB-4C5F-880D-66482FF25DF6 [Accessed

10 May 2024]

Oracle (2021) 10 Benefits of Oracle's Data Management Platform. Available from:

https://www.oracle.com/a/ocom/docs/10-benefits-of-oracle-data-management-

platform.pdf [Accessed 15 April 2024]

Pattinson, T. (2022) Relational vs. Non-Relational Databases. Available from:

https://www.pluralsight.com/blog/software-development/relational-vs-non-

relational-databases [Accessed 21 May 2024]

Yehuda, Y. (2018) 5 Ways to Make Your Database GDPR-Compliant. Available from:

https://dzone.com/articles/5-ways-to-make-your-database-gdpr-compliant

[Accessed 22 May 2024]