Designing a Relational Database for Classroom Observations

Contents

Abstract

A relational database is proposed in this report to replace the Excel spreadsheets used for tracking and reporting on classroom observations in [OMMITED FOR ANONYMITY]. The concept of relational databases is explored, alongside the pros and cons of utilising them in [OMMITED FOR ANONYMITY] as well as the legislative, ethical and security implications thereof. A design of a relational database for observations is introduced and discussed in terms of its purpose and implementation. The overall suitability of using the designed relational database instead of Excel spreadsheets is then summarised.

Introduction

The data model proposed in this report, to replace Excel spreadsheets used for tracking classroom observations within further education provider [OMMITED FOR ANONYMITY] ([OMMITED FOR ANONYMITY]), is a relational database. A relational database can be described as multiple tables which are related to one another to form a database. The way in which the relationships are established, is by creating an attribute in one table, called a primary key (PK), that can be identified and referenced within another table, where it would be a foreign key (FK), as defined by Google (2023) Cloud, a provider of database services.

There are several potential gains that could be taken advantage of by using relational databases in [OMMITED FOR ANONYMITY]. One advantage is that relational databases can eliminate duplication of data (Jatana et al, 2012, p.5), thus contributing to database normalisation. This could be beneficial within [OMMITED FOR ANONYMITY] where information is often repeated across different systems, such as student information.

It is also generally easy to report on data that is structured as a relational database as well. Since the schema is defined, the desired information from it can be queried and collated through programming such as with Structured Query Language (SQL) (Ordonez, Song and Garcia-Alvarado, 2010, p.1). The Reporting team within [OMMITED FOR ANONYMITY] could make use of this functionality.

Another advantage of relational databases is that it is usually easy to adjust the schema, once it has been created (Jatana et al, 2012, p.2). This may prove useful to [OMMITED FOR ANONYMITY] since, in education, changes can occur, such as with legislation, which may in turn require the way data is utilised to be altered. Hence, employing a relational database may simplify the process of adjusting the data model to meet these requirements.

Despite the benefits, one of the more prominent drawbacks with relational databases is that they are often not effective with handling unstructured data (Ordonez, Song and Garcia-Alvarado, 2010, p.1). For this reason, it may be prudent to analyse the structure of the data in question from [OMMITED FOR ANONYMITY] beforehand when considering if it should be made into a relational database.

Relational databases also do not offer efficient scalability (Jatana et al, 2012, p.3). This may be problematic for [OMMITED FOR ANONYMITY] if, for example, a relational database was created and over time the data it stores grows too large and measures such as hardware upgrades or distribution across servers need to be implemented. This outcome could be costly, both financially and timely for [OMMITED FOR ANONYMITY].

Additionally, relational databases can become complex in structure since it can only store data in a tabular format (Jatana et al, 2012, p.3). This may be challenging for [OMMITED FOR ANONYMITY] if

the relationships between tables are so complex that it is difficult to discern for the Reporting team and too complex to explain to other stakeholders, such as the executives.

Legislation must also be considered with relational databases. Schools in the UK must abide by the General Data Protection Regulation (GDPR) alongside the Data Protection Act (DPA), as specified by the Department for Education (2023). [OMMITED FOR ANONYMITY] acknowledges this within the Data Protection Policy. Since [OMMITED FOR ANONYMITY] hosts data that is relevant to these laws, appropriate measures must be created and enforced with regards to relational databases.

Security is another concern for relational databases. Since they can contain sensitive data, steps should be taken to protect it. Creating backups to prevent data loss or only allowing access to data for persons with the company network to avoid breaches (National Cyber Security Centre, 2023) are some methods that should be utilised within [OMMITED FOR ANONYMITY] to ensure sufficient security for its databases.

Workplace Problem

Currently in [OMMITED FOR ANONYMITY], classroom observations are recorded and tracked using Microsoft Excel spreadsheets. While spreadsheets have been satisfactory, the downsides of this format are significant enough that a relational database is proposed in this report to replace this method.

Table 1 – Sample row of main table from [OMMITED FOR ANONYMITY] Observations Excel spreadsheet

Date	Site	Department	Team	Offering	Observer	Observee	Type	Grade	Attendance%
01/01/2023	1	Sciences & Development Studies	English	GCSE English	Matt Jones	John Smith	Standard	4 – Good	90%

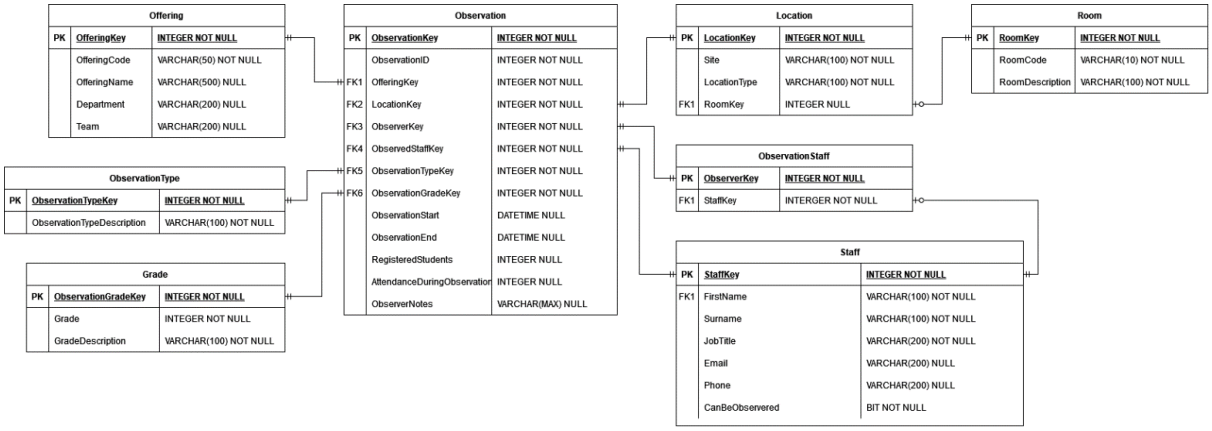
One of the key problems with this format is that these spreadsheets have no relations to [OMMITED FOR ANONYMITY]'s data warehouse. This means that collating data from observations with the data warehouse is either not possible or challenging to achieve. If a relational database were to be implemented instead, then existing tables from the warehouse could be used to make this possible and improve overall reporting.

The spreadsheets also use many columns, since all the data is stored in a single table, while also not always containing all relevant information. A relational database could improve performance and increase the number of overall attributes. This would be achieved by having several related tables with more attributes to retrieve the desired information through queries.

Design

The below entity relationship diagram (ERD) is designed as a relational database that could replace the [OMMITED FOR ANONYMITY] observations spreadsheets.

Figure 1 – Entity relationship diagram of proposed observations relational database



‘Observation’ is the main table as it is intended for recording the details of classroom observations such as: observers/observed staff, the start/finish date & time, students on register versus attended and observer notes. The ‘ObservationKey’ attribute uniquely identifies each row as the PK. This differs to the ‘ObservationID’ which is not unique as it identifies each observation case, since an observation could have more than one row where, for example, there may be more than one observer for an observation. ‘Observation’ also allows for additional details to be retrieved, when needed, through its relations to other tables in the database. The tables which ‘Observation’ has FKs to are:

- ‘Offering’ which contains details of [OMMITED FOR ANONYMITY]s courses.
- ‘ObservationType’ that contains values such as “standard” for defining the type of observation.
- ‘Grade’ used for the observer score of an observation.
- ‘Location’ which tracks the whereabouts of all classes in [OMMITED FOR ANONYMITY].
- ‘ObservationStaff’ is all the Observations Team staff.
- ‘Staff’ that stores all employee details.

Furthermore, ‘Staff’ has a one-to-zero-or-one relationship with ‘ObservationStaff’ as all employees are staff members but not all belong to the Observations Team. This is so that ‘ObservationStaff’ exists as its own entity with a separate configuration to ‘Staff’. It may be advantageous, for instance, to allow the Observations Team to keep ‘ObservationStaff’ up to date but not be able to change ‘Staff’. The employee details of the Observations Team can be looked up through the relationship between ‘ObservationStaff’ and ‘Staff’ to avoid data redundancy by not needing the information in both entities. ‘Staff’ also has column ‘CanBeObservered’ to ensure that staff who would never need classroom observations are not accidentally entered into ‘Observation’; this could be implemented as a constraint.

‘Room’ also exists with a one-to-zero-or-one relationship to ‘Location’. The reason for this is to enable a specific classroom to be findable if one exists, although it may not exist in cases such as virtual classrooms.

The main benefit to this model is that it could fit into [OMMITED FOR ANONYMITY]’s data warehouse which may improve the overall reporting ability of the college. To illustrate, ‘Offerings’

already exists as a table within the data warehouse and classroom observations could be added as one of the indicators for measuring the success of a particular course.

There are also more details in this model regarding classroom observations compared to the spreadsheet version. However, it should still be straightforward to report on since the requested data can be retrieved using queries on the relevant entities instead of it existing in one large table.

Despite this, there are drawbacks to this model. There are a lot of text attributes in the form of VARCHARs, and this could potentially result in a large storage cost; particularly over time as the database grows. The 'ObserverNotes' especially may be costly as it is a VARCHAR(MAX), but this cannot realistically be reduced as the Observations Team require the ability to record as few or many notes as they need.

The model also must allow for additional rows per classroom observation if there is more than one: observer, observed staff, start/end date & time or observer notes. Though a particular classroom observation can still be identified via the 'ObservationID', this could add to both the complexity and storage costs of the database.

Implementation

Regarding implementing this data model, this would be added to [OMMITED FOR ANONYMITY]'s existing data warehouse, which is on its own server and hosted using SQL Server. This method is chosen as it would be consistent with how most other reporting databases are implemented in [OMMITED FOR ANONYMITY]. The ERD in figure 1 was also designed with SQL Server in mind, such as with 'CanBeObservered' since the data type for this is assigned as a BIT. This column is intended to be a true/false column though BOOLEAN does not exist in SQL Server, hence BIT was chosen as this only allows for 0, 1 or NULL values and works in SQL Server. This implementation would also allow the observations database to be utilised with the Reporting platforms used by [OMMITED FOR ANONYMITY] such as Power BI or SQL Server Reporting Services.

To ensure that this data is stored securely, existing procedures, as well as ones specific to the observations database, will be employed. For starters, the data warehouse server is held in a secure location in [OMMITED FOR ANONYMITY] to prevent physical risk, including theft or damage. Further protection will also be applied such as encryption and strong passwords to the server and database.

Permissions to the data warehouse server and observations database for read, write, and edit access, will be restricted to the Reporting Team and Information Technology Team only. Separate security could also be implemented for allowing the Observations Team read-only access to the observations database to enable them to generate their own reporting, without the ability to manipulate the data. These permissions would be enforced via active directory groups and roles within Windows and SQL Server. This should block unwanted persons from accessing the data and prevent unethical use of it. For instance, if Observations or Teaching Staff had edit access to the database, someone may want to manipulate the recorded data to appear more favourable towards them when reported on.

Several Backups will be taken of the database in separate locations as this should mitigate data loss for threats such as flooding (National Cyber Security Centre, 2023). These measures are to contribute to the efforts of safely storing and safeguarding data as per the UK GDPR and Data Protection Act that schools in the UK must follow (Department for Education, 2023).

Actions for this data model should also be applied with respect to the legislation of UK classroom observations. One such action that could be put in place is around the permitted duration of

observations as they are not to exceed more than three hours (The National Archives, 2006). In this case, a trigger could be added to the schema to comply with this by creating an alert for when the difference between 'ObservationStart' and 'ObservationEnd' for the same 'ObservationID' exceeds three hours; or the trigger could disallow this from occurring altogether. Protocols such as this can be utilised to ensure the observations database complies with the relevant laws.

Conclusion

In the matter of whether this proposed relational database trumps the currently used spreadsheets, the advantages potentially gained with using this model should improve upon the current format for tracking and reporting on classroom observations in [OMMITED FOR ANONYMITY]. In other words, the improved functionality and reporting ability of this relational database should benefit the organisation more compared to the simplicity of using Excel spreadsheets.

The proposed relational database is not perfect however as there are further improvements that could be made. To name one, the way the model handles having more than one observer, staff observed, or observer notes is by having more rows for these in 'Observation' and linking these to a single 'ObservationID'. This may cause unnecessary duplication of data for other columns in 'Observation', such as 'RegisteredStudents', which would not change for an observation, regardless of how many observers, staff observed, or notes there are. A bridging table could be added to the model instead to link a classroom observation to one-or-many observers, staff observed, or observer notes so that the 'Observation' table always has a single row per observation and thus eliminating duplication.

References

Department for Education (2023) *Data protection in schools* [Online]. Available at: https://www.gov.uk/guidance/data-protection-in-schools/data-protection-policies-and-procedures (Accessed 27th May 2023)

Google (2023) *What is a relational database?* [Online]. Available at: https://cloud.google.com/learn/what-is-a-relational-database (Accessed 25th May 2023)

Jatana, N. et al. (2012) 'A survey and comparison of relational and non-relational database', *International Journal of Engineering Research & Technology*, 1(6) [online]. Available at: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=14897e1d7a48840ac77809966376a29fe9a09d2b (Accessed 25th May 2023)

The National Archives (2006) *The Education (School Teacher Performance Management) (England) Regulations 2006* [Online]. Available at: https://www.legislation.gov.uk/uksi/2006/2661/regulation/17/made (Accessed 8th June 2023)

National Cyber Security Centre (2023) *10 Steps to Cyber Security* [Online]. Available at: https://www.ncsc.gov.uk/collection/10-steps/data-security (Accessed 27th 2023)

Ordonez, C., Song, I.Y. and Garcia-Alvarado, C. (2010) 'Relational versus non-relational database systems for data warehousing'. *Proceedings of the ACM 13th international workshop on Data warehousing and OLAP,* p67-68 [online]. Available at: https://dl-acm-org.ezproxy.neu.edu/doi/abs/10.1145/1871940.1871955 (Accessed 27th May 2023)