



# TEAM PROJECT

Project Report

Ruth Allison, Nicholas Bandy, Mika Fowles

## Overview

Uni Prep Ltd is a private education company that enables access to UK universities by providing courses for international students who are not eligible for direct entry. The company has 10 colleges partnered with UK universities plus a head office with central teams such as admissions and compliance. Students apply from their home country and, on completion of their course at the required level, can then progress onto their chosen undergraduate degree. Uni Prep Ltd requires a student management system for use by staff at different points in the students' journeys – admissions, finance, academic staff, support staff and the data team.

## User requirements

To determine the system requirements, staff in a range of college-based roles were surveyed about their use of the current student management system, KSS, and what improvements would be beneficial (Appendix 1).

## Requirements of the database model design

### Data storage

Information held at Uni Prep Ltd. contains a range of staff and student data as shown in Figure 1.

Relevant student information:

- Approx 4500 students. Average 350, max 2000
- 2 programmes, 20 modules
- 2 start dates per year (Sept, Jan)
- 2 programme lengths (2 term, 3 term)
- 2 finish dates (May, July)

### User access

The proposed DBMS must support many concurrent users including approximately 1,200 staff (e.g., teachers, managers and administration team members in 10 different colleges in the network and the head office staff members). The average number of concurrent users is 50 with the maximum of 200.

### Manipulation and retrieval of data

The database model design must allow efficient data entry, deletion and updating. Data retrieval via search and query functionality is a key transaction.

Another important function based on the user feedback is the reporting and analytic element which allows users to generate predefined or custom reports.

## Data control issues

Some crucial data control issues are security, backup and recovery, and concurrency control (Gillenson, 2011). Data security requirements include: user authentication, authorisation and encryption. The student data contains personal details of vulnerable individuals (under 18 and/or with disabilities). Given the sensitive nature of data, compliance requirements, such as GDPR, must be adhered to. Reliable data backup and recovery must also be part of the data management system. In the multi-user environment at Uni Prep Ltd. where users attempt to access the same data simultaneously, the database model must be designed in such a way that data updates will not interfere with multi-user transactions.

## Proposed database model design

The proposal is a SQL database designed to easily handle data entry, updates, or deletion, while minimizing the amount of storage space required to store all the information.

The following role-level security should be applied to protect personal information:

Students – View and update their own personal information, view their classes and grades.

Teachers – Add notes and grades for classes they're teaching, update their own personal information, View details relevant to their classes.

Administrative staff – View everything, add students, teachers, modules, classes.

Database administrator – Full access

Apart from the database administrator, all groups should be viewing this information through a user interface. This prevents any user from writing and attempting to run personalized scripts. This measure improves data quality as varchar fields such as address or assessment type can be selected from a list in a form, making the data consistent.

Depending on the role of some administrative staff, they may be given access to view data with a customized script, but not to insert or delete.

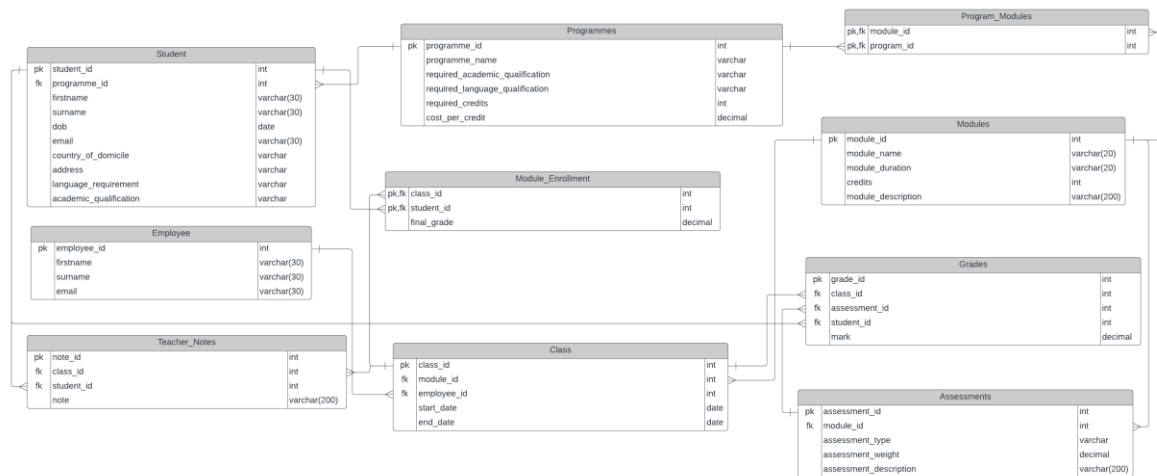


Figure 1: Proposed database model design

## Data management pipeline

Data is collected in three stages – first the admissions team input applicant data to an online form, creating multiple individual student records. Then college staff sort the student records and aggregate them by programme to populate the module instances. Finally, teaching staff enter student grades for each module assessment. All these steps have challenges in ensuring the integrity of the data.

The admissions teams work remotely, dealing with specific geographical markets. Regional differences in data such as date format are common. Hellenstein (2008) suggests that this can be mitigated by a robust data input interface with constraints on the data types and values accepted, however a study by Barchard and Pace (2011) found that only 0.06% of data entry errors were blank or outside the acceptable input range. A solution is to have drop-down menus, minimising the input of free text. Blank fields are dealt with by blocking the data submission until an option has been chosen. While this will help with the consistent input of data it won't necessarily address the most common input issues. Huxley (2020) identifies six possible data entry errors – transposition, loss, substitution, insertion, omission, and duplication.

College staff work directly with the student records, having a series of drop-down options when they are creating class lists for each module. This is integrated with specialist timetabling software to ensure that students cannot be put in two concurrent classes.

The data management step most vulnerable to error is the addition of student grades. The accuracy depends on the person inputting the grades. If the grade added is in an acceptable range (0-100%) any errors are unlikely to be spotted. Good practice is to have an independent check the grades awarded tally with those input.

A final step in data cleaning is to identify duplication of student records. This occurs when a student has changed programme or repeated a programme. Each instance must be examined to determine if the duplicate record can be deleted or not.

## Proposal of the Database management system (DBMS)

Our proposed database model design involves the relational database management system (RDBMS). Compared to other types of DBMS such as the hierarchical and network approaches which require navigation through the hierarchy or network of stored data to be able to find specific data in question (Gillenson, 2011), the RDBMS presents several advantages including simplicity and flexibility. The RDBMS has been in increasing use in the past four decades (Hassan, 2021) and its data structure allows reliable and efficient transactions, which would be ideal for the requirements of Uni Prep Ltd.

Based on this, our proposed database management system is Microsoft Azure.

Azure is a cloud computing platform provided by Microsoft. This online portal offers reliable and scalable Platform-as-a-Service (PaaS) on a pay-as-you-go basis, which makes it a flexible and cost-effective solution. Furthermore, several aspects in data security are covered, e.g., network security, access management, threat protection and information protection (Microsoft, 2023a). Azure also complies with various privacy regulations including the GDPR and ISO/IEC 27001 (Microsoft, 2023b). Azure SQL database allows users to extract specific information using an XML-based query (Chao, 2013).

## References

Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behaviour* (27): 1834–1839.

<https://www.sciencedirect.com/science/article/pii/S0747563211000707> [Accessed 1 June 2023]

Chao, L. (2013) *Cloud database development and management*. CRC Press.

Gillenson, M. L. (2011) *Fundamentals of database management systems* (2<sup>nd</sup> ed.). Wiley.

Hassan, M. A. (2021) Relational and NoSQL databases: The appropriate database model choice. In *2021 22nd International Arab Conference on Information Technology, ACIT 2021*. DOI:

[10.1109/ACIT53391.2021.9677042](https://doi.org/10.1109/ACIT53391.2021.9677042) [Accessed 3 June 2023].

Hellerstein, J. M. (2008). Quantitative data cleaning for large databases (White Paper). United Nations Economic Commission for Europe. Available from

<http://db.cs.berkeley.edu/jmh/papers/cleaning-unece.pdf> [Accessed 1 June 2023]

Huxley, K. (2020) 'Data Cleaning', in Atkinson, P., Delamont, S., Cernat, A., Sakshaug, J.W., & Williams, R.A. (eds) *Quantitative Data Preparation & Secondary Data Analysis*. SAGE Research Methods Foundations. DOI: <https://doi.org/10.4135/9781526421036842861>

Microsoft (2023a) *What is database security?: Learn how to secure your database and protect it from threats*. Available from: <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-database-security/#what-is-database-security> [Accessed 10 June 2023].

Microsoft (2023b) *Privacy in Azure*. Available from: <https://azure.microsoft.com/en-gb/explore/trusted-cloud/privacy> [Accessed 10 June 2023].

## Appendix 1. User survey regarding the current system



KSS user survey  
Results.docx