Database Design Proposal - Deciphering Big Data

Total word count = 1092

Team: *Chantelle Fields (12698973)*
Kalifa Burton (12699422)
Daniella Heaton (12698971)
Anna Kakovka (12696017)

Client Organisation

Organisation Name: MedicaidUK

Sector: UK Private Healthcare Services

Services: Consultations with GP, Specialist clinics (e.g. Dermatology, Rheumatology), diagnostic tests, patient management and billing services.

Objectives: MedicaidUK seeks to improve the efficiency and safety of medical record management through a logical database system that ensures secure patient data storage, manages appointments, treatments, and billing, complies with UK healthcare regulations (GDPR and NHS), and scales to meet growing demand and integrate with digital health platforms.

Our team are developers in this project to build a robust, scalable database solution designed to grow and adapt with our clients' evolving needs.

Logical Design

1) Data items/entities

Patient	Stores personal and medical information of patients
Healthcare Professionals	Stores professional details and specialisations of healthcare professionals (Nurses, doctors, consultants, therapists etc.)
Appointment Details	Details of patient appointments (times/dates, appointment reason etc.)
Treatment	Information about treatments, diagnostics, medications and procedures
Billing	Records of payments, invoices and insurance claims
Department	Information about departments, specialities and treatments offered within the clinic
Insurance Provider	Records/ information for insurance companies that are outside of the

	system and this is tied to the claims for billing purposes.

2) Data types/formats

INT for unique IDs	For unique IDs (both patients' and healthcare professionals')
VARCHAR	For text fields (e.g. names, addresses, e-mails etc.)
DATE / DATETIME	For dates and timestamps (appointment dates/times, medical record entry dates/times etc.)
DECIMAL(8,2)	For financial figures
CHAR(1)	For single character data, such as gender.
TEXT	For long descriptive fields (e.g. treatment descriptions etc.)
ENUM	For gender, status, consultation type etc.

3) Attributes of data items

Patient	PatientID (Primary Key, INT, auto-increment) FirstName (VARCHAR, 50) LastName (VARCHAR, 50) DateOfBirth (DATE) MaritalStatus (VARCHAR, 20) MedicalHistorySummary (TEXT) Gender (ENUM ('M', 'F','O') PhoneNumber (VARCHAR, 20) Email (VARCHAR, 100) EmergencyContactName and Phone (VARCHAR, 100) and (VARCHAR, 15) Address (VARCHAR, 255) NHSNumber (VARCHAR, 15) BloodType(VARCHAR, 3)
Healthcare Professionals	HealthcareProfessionalID (Primary Key, INT, auto-increment)

	FirstName (VARCHAR, 50) LastName (VARCHAR, 50) Specialisation (VARCHAR, 50) Qualifications (TEXT) PhoneNumber (VARCHAR, 15) Email (VARCHAR, 100) DepartmentID (INT) LicenceNumber (VARCHAR,30) YearsOfExperience (INT)
Appointment Details	AppointmentID (Primary Key, INT, auto-increment) PatientID (Foreign Key → Patient.PatientID) DoctorID (Foreign Key → Doctor.DoctorID) AppointmentDate (DATETIME) DurationMins (INT) Reason (VARCHAR, 255) Status (ENUM) ConsultationType(ENUM) Notes (TEXT)
Treatment	TreatmentID (Primary Key, INT, auto-increment) AppointmentID (Foreign Key → Appointment.AppointmentID) TreatmentType (VARCHAR, 50) Description (TEXT) Cost (DECIMAL(10,2) MedicalPrescribed (TEXT) FollowUpRequired (TINYINT) Outcome (TEXT)
Billing	BillID (Primary Key, INT, auto-increment) InvoiceNumber (VARCHAR, 30) PatientID (Foreign Key → Patient.PatientID) TreatmentID (Foreign Key → Treatment.TreatmentID) AmountPaid (DECIMAL 10,2) OutstandingBalance (DECIMAL 10,2) PaymentDate (DATE) PaymentMethod ENUM(...) DueDate (DATE) ClaimStatus ENUM (...)
Department	DepartmentID (Primary Key, INT, auto-increment)

	DepartmentName (VARCHAR, 50) Location (VARCHAR, 100)
Insurance Provider	InsuranceID (Primary Key, INT, auto-increment) ProviderName (VARCHAR, 100) ContactNumber (VARCHAR, 20) PolicyDetails (TEXT)

4) Relationships and associations of data

- **Patient ↔ Appointment:** One-to-Many (One patient can have multiple appointments)
- **Patient ↔ Treatment:** One-to-Many (One patient can receive multiple treatments)
- **HealthcareProfessional ↔ Appointment:** One-to-Many (One professional can have multiple appointments)
- **HealthcareProfessional ↔ Department:** One-to-One or One-to-Many (One professional generally belongs to one department, but may belong to many)
- **Appointment ↔ Treatment:** One-to-Many (Each appointment can have multiple treatments)
- **Treatment ↔ BloodTransfusion:** One-to-One (only some treatments are transfusions)
- **Patient ↔ Billing:** One-to-Many (Each patient can have multiple bills)
- **Treatment ↔ Billing:** One-to-One or One-to-Many (Depending on multiple bill entries)
- **Billing ↔ Insurance Provider** → Many-to-One (numorous claims controlled by one provider).

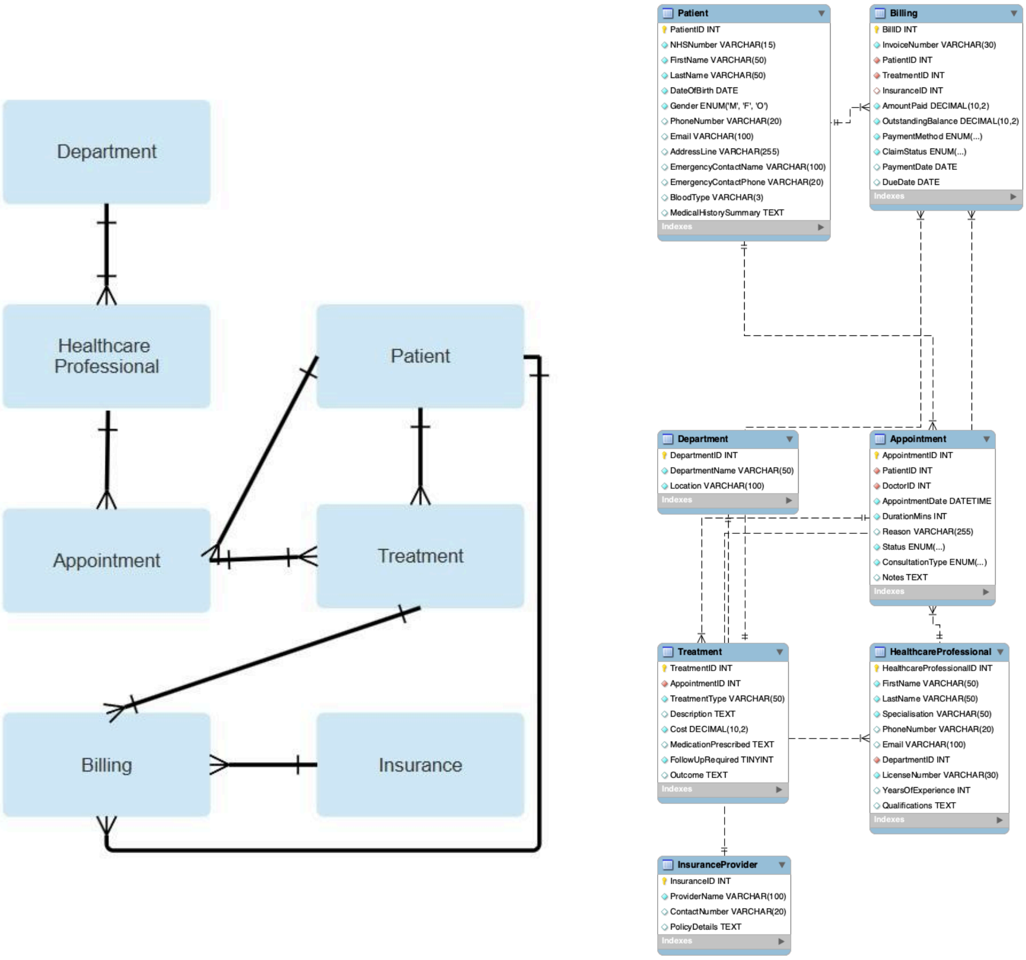

Figure 1. Entity-Relationship diagram, depicting the associations within the database, Oracle (2025).

Proposed Database Build

a) **Proposed Database Management System (DBMS):** MySQL

Reason: Offers robust relational data support, ACID compliance for secure healthcare data, strong community backing, extensibility, cost efficiency, and compatibility with web apps and analytics tools.

b) User Access and Security: Role-based access; Admin (full access), Healthcare Professionals (access patient records and treatments only), Receptionists (appointments and billing), Auditors/Managers (read-only to monitor compliance), and Data encryption for sensitive fields (e.g. personal contact details).

c) Data Manipulation and Retrieval: CRUD operations for all entities

Queries to: Retrieve patient history, track appointments and treatments, generate billing reports, monitor departmental workload.

Data Management Pipeline

a) Data capturing: Sources; manual entry at clinic reception, patient self-service portals, integration with diagnostic equipment and imported historical records from external healthcare databases, integration with internal laboratory information management (LIMS) systems.

b) Data cleansing techniques:
Standardisation - Consistent formatting of phone numbers, addresses and dates.

Validation - Mandatory fields for critical data (e.g. PatientID, DateOfBirth etc.)

De-duplication - Remove duplicate patient entries, merge duplicate patient records.

Error checking - Ensure numeric fields contain valid numbers and values (e.g. billing amounts).

c) Stage of data cleansing:
Data collection - Gathered from all sources, checked for completeness

Validation and formatting - Applied rules for consistent and accurate formats

Error correction - Fixed inconsistencies in names, dates, and codes

De-duplication - Merged duplicate records

Final verification - Cross checked with source records for accuracy

Summary/Evaluation

Advantages of proposed design

The proposed design provides a secure, regulation-compliant framework for managing healthcare data, with role-based access that reduces insider breaches (Marquis, 2024) while allowing flexible analysis to enhance patient care (Batko & Ślęzak, 2022). It also includes audit and compliance features essential for regulated services and supports clinical reporting.

Challenges/limitations:

a) **Data privacy and compliance**

Challenge: Healthcare data is highly sensitive and must comply with GDPR and the UK Data Protection Act. Any breach could result in severe legal and financial penalties.
Mitigation: Strong encryption, role-based access control, regular audits, and secure backups.

b) **User training and implementation**

Challenge: Staff may resist adopting a new system due to unfamiliarity or fear of change. Incorrect use of the database could introduce errors or inefficiencies.
Mitigation: Provide training sessions, user-friendly interfaces, and ongoing support. ‘Gamified’ training proved to be more impactful (Bitrián et al., 2024).

c) **Integration with existing systems**

Challenge: The clinic probably already uses a system for collating all this information (e.g. patient notes, billing, or test results). Data migration and system interoperability can be complex and error-prone.
Mitigation: Use ETL (Extract, Transform, Load) pipelines and API-based integration strategies.

d) **Scalability of system**

Challenge: The system may suit small clinics, but performance could decline with growth, as relational databases can slow under heavy transaction loads without optimisation.
Mitigation: Use indexing, partitioning, and plan for cloud scalability (AWS RDS, Azure SQL).

Proposed Future Development

The current proposed database focuses on private healthcare, a centralised healthcare system reduces testing redundancy, decreasing costs, and increasing efficiency (Tariq, Tariq and Ahmad Adnan Shoukat, 2023).

Integrating Machine Learning into the data pipeline using predictive modelling can provide more accurate care (Na et al., 2023).

Reference list

Batko, K. and Ślęzak, A. (2022). The Use of Big Data Analytics in Healthcare. *Journal of Big Data*, [online] 9(3), pp.1–24. doi:https://doi.org/10.1186/s40537-021-00553-4.

Bitrián, P., Buil, I., Catalán, S. and Merli, D. (2024). Gamification in Workforce training: Improving Employees' self-efficacy and Information Security and Data Protection Behaviours. *Journal of Business Research*, [online] 179, pp.114685–114685. doi:https://doi.org/10.1016/j.jbusres.2024.114685.

Marquis, Y.A. (2024). From Theory to Practice: Implementing Effective Role-Based Access Control Strategies to Mitigate Insider Risks in Diverse Organizational Contexts. *Journal of Engineering Research and Reports*, [online] 26(5), pp.138–154. doi:https://doi.org/10.9734/jerr/2024/v26i51141.

Na, L., Carballo, K.V., Pauphilet, J., Haddad-Sisakht, A., Kombert, D., Boisjoli-Langlois, M., Castiglione, A., Khalifa, M., Hebbal, P., Stein, B. and Bertsimas, D. (2023). *Patient Outcome Predictions Improve Operations at a Large Hospital Network*. [online] arXiv.org. Available at: https://arxiv.org/abs/2305.15629.

Oracle (2025) *My SQL Workbench* [Computer Program]. Available at: https://www.mysql.com/products/workbench/ (Accessed: 06 Sept 2025).

Tariq, S., Tariq, S. and Ahmad Adnan Shoukat (2023). Centralized healthcare database for ensuring better healthcare: Are we lagging behind? *Pakistan Journal of Medical Sciences*, [online] 40(3). doi:https://doi.org/10.12669/pjms.40.3.9084.