

**Universidade do Minho**

Mestrado Integrado em Engenharia Informática

Licenciatura em Ciências da Computação

**Unidade Curricular de**

**Bases de Dados**

Ano Letivo de 2019/2020

**Agendamento e realização de testes clínicos de atletas**

**Grupo 42**

**Hugo Cunha A84656**

**Maria Pires A86268**

**Susana Marques a84167**

Janeiro, 2020

**BD**

|  |  |
| --- | --- |
| Data de Recepção |  |
| Responsável |  |
| Avaliação |  |
| Observações |  |

**Agendamento e realização de testes clínicos de atletas**

**Grupo 42**

**Hugo Cunha A84656**

**Maria Pires A86268**

**Susana Marques a84167**

Janeiro, 2020

# Resumo

Este relatório foi realizado no âmbito do desenvolvimento de uma Base de Dados para a *Clínica Alta Performance*, que foca a sua atividade na área da saúde principalmente na gestão da realização de testes clínicos a atletas de alta competição. Desde o agendamento destes testes até ao resultado dos mesmos, passando pelas provas realizadas pelos atletas que se submetem aos testes, este software facilita e permite de forma eficiente organizar a realização destes exames.

Neste relatório são apresentadas todas as etapas realizadas para o desenvolvimento desta base de dados até à implementação física da mesma.

Inicialmente, estudou-se o meio onde o software implementado será inserido. Analisou-se as motivações e os objetivos da clínica mencionada, como esta mesma surgiu e de que modo a implementação de uma base de dados será economicamente rentável para a mesma.

A seguir, por métodos de análise, levantou-se e aprovou-se os requisitos essenciais para o sistema que guiaram todo o processo que foi desenvolvido. Com estes requisitos, iniciou-se a modulação conceptual que depois, em conformidade com as regras do mapeamento ER, serviu para desenvolver o modelo lógico na ferramenta *MySQL Workbench*.

Concluído o modelo lógico, implementou-se fisicamente a Base de Dados através também do *MySQL Workbench*, garantindo a validação de todo o trabalho realizado anteriormente e permitindo implementar uma solução correta para o problema apresentado.

Recorreu-se também a um sistema NoSQL quando houve necessidade de uma Base de Dados complementar com maior escalabilidade e “performance” para alcançar a solução para as dificuldades sentidas com o aumento de agendamentos de atletas.

Com a resolução deste problema, depois de satisfeitas as necessidades encontradas, demos por concluído este projeto.

**Área de Aplicação:** Desenho, arquitetura, desenvolvimento e implementação de Sistemas de Bases de Dados.

**Palavras-Chave:**  Bases de Dados, Bases de Dados Relacionais, Análise de Requisitos, Entidades, Atributos, Relacionamentos, Modelo Conceptual, Modelo Lógico, Modelo Físico, Normalização, Interrogações, Transações, Triggers, Índices, Vistas de Utilização, Backup, MySQL WorkBench, MySQL.

# Índice

[1. Introdução 1](#_Toc22767017)

[1.1. Contextualização 1](#_Toc22767018)

[1.2. Apresentação do Caso de Estudo 1](#_Toc22767019)

[1.3. Motivação e Objectivos 1](#_Toc22767020)

[1.4. Estrutura do Relatório 1](#_Toc22767021)

[2. Sugestões para Escrita do Relatório 2](#_Toc22767022)

[2.1. Sugestões Gerais 2](#_Toc22767023)

[2.2. Termos Estrangeiros 2](#_Toc22767024)

[2.3. Tabelas e Figuras 2](#_Toc22767025)

[2.4. Siglas e Acrónimos 3](#_Toc22767026)

[2.5. Referências Bibliográficas 3](#_Toc22767027)

[2.6. Tipo de Ficheiro 3](#_Toc22767028)

[3. Levantamento e Análise de requisitos 4](#_Toc22767029)

[3.1. Método de Levantamento e de Análise de Requisitos Adotados 4](#_Toc22767030)

[3.2. Requisitos Levantados 4](#_Toc22767031)

[3.3. Análise de Requisitos 4](#_Toc22767032)

[4. Modelação Conceptual 5](#_Toc22767033)

[4.1. Apresentação da Abordagem de Modelação Realizada 5](#_Toc22767034)

[4.2. Identificação e Caracterização das Entidades 5](#_Toc22767035)

[4.3. Identificação e Caracterização dos Relacionamentos 5](#_Toc22767036)

[4.4. Identificação e Caracterização da Associação dos Atributos com as Entidades e Relacionamentos 5](#_Toc22767037)

[4.5. Domínio dos Atributos 5](#_Toc22767038)

[4.6. Chaves Candidatas, Primárias e Alternativas 5](#_Toc22767039)

[4.7. Detalhe ou Generalização de Entidades 5](#_Toc22767040)

[4.8. Apresentação e Explicação do Diagrama ER 5](#_Toc22767041)

[4.9. Validação do Modelo de Dados com o Utilizador 5](#_Toc22767042)

[5. Modelação Lógica 6](#_Toc22767043)

[5.1. Construção e Validação do Modelo de Dados Lógico 6](#_Toc22767044)

[5.2. Desenho do Modelo Lógico 6](#_Toc22767045)

[5.3. Validação do Modelo através da Normalização 6](#_Toc22767046)

[5.4. Validação do Modelo com as Interrogações do Utilizador 6](#_Toc22767047)

[5.5. Validação do Modelo com as Transações Estabelecidas 6](#_Toc22767048)

[5.6. Revisão do Modelo Lógico com o Utilizador 6](#_Toc22767049)

[6. Implementação Física 7](#_Toc22767050)

[6.1. Seleção do Sistema de Gestão de Base de Dados 7](#_Toc22767051)

[6.2. Tradução do Esquema Lógico para o Sistema de Gestão de Bases de Dados escolhido em SQL 7](#_Toc22767052)

[6.3. Tradução das Interrogações do Utilizador para SQL 7](#_Toc22767053)

[6.4. Tradução das transações estabelecidas para SQL 7](#_Toc22767054)

[6.5. *Triggers* SQL 7](#_Toc22767055)

[6.6. Escolha, Definição e Caracterização de Índices em SQL 7](#_Toc22767056)

[6.7. Estimativa do Espaço em Disco da Base de Dados e Taxa de Crescimento Anual 7](#_Toc22767057)

[6.8. Definição e Caracterização das vistas de utilização em SQL 8](#_Toc22767058)

[6.9. Definição e Caracterização dos Mecanismos de Segurança em SQL 8](#_Toc22767059)

[6.10. Revisão do Sistema Implementado com o Utilizador 8](#_Toc22767060)

[7. Sistema NoSQL: ?? 9](#_Toc22767061)

[7.1. Justificação da utilização de um sistema NoSQL 9](#_Toc22767062)

[7.2. Identificação e descrição dos objetivos da base de dados, em termos de aplicações e de utilizadores 9](#_Toc22767063)

[7.3. Identificação e explicação do tipo de questões (necessidades) que serão realizadas sobre o sistema de dados NoSQL 9](#_Toc22767064)

[7.4. Definição da estrutura base para o sistema de dados NoSQL que satisfaça os requisitos e as questões apresentadas anteriormente 9](#_Toc22767065)

[7.5. Identificação dos objetos de dados no sistema SQL que serão utilizados para alimentar o novo sistema 10](#_Toc22767066)

[7.6. Mapeamento do processo de migração de dados, descrevendo o processo de conversão dos vários objetos de d 10](#_Toc22767067)

[7.7. Explicação do processo de migração de dados, explicando de forma detalhada as suas principais etapas - extração, transformação e carregamento 10](#_Toc22767068)

[7.8. Apresentação e descrição da implementação do processo de migração de dados 10](#_Toc22767069)

[7.9. Apresentação da forma como as questões identificadas anteriormente podem ser satisfeitas com o novo sistema, utilizando a linguagem de interrogação do sistema NoSQL 10](#_Toc22767070)

[8. Conclusões e Trabalho Futuro 11](#_Toc22767071)

[Referências 12](#_Toc22767072)

[Lista de Siglas e Acrónimos 13](#_Toc22767073)

[Anexos 14](#_Toc22767074)

[I. Anexo 1 15](#_Toc22767075)

# Índice de Figuras

[Figura 1 - Relacionamento Atleta - Teste Clínico 7](#_Toc28267460)

[Figura 2 - Relacionamento Teste Clínico – Médico 7](#_Toc28267461)

[Figura 3 - Relacionamento Médico - Clínica 8](#_Toc28267462)

[Figura 4 - Relacionamento Atleta - Prova 9](#_Toc28267463)

[Figura 5 - Modelo Conceptual 12](file:////Users/mariapires/Desktop/BD/relatorio.docx#_Toc28267464)

[Figura 6 - Modelo Lógico 18](#_Toc28267465)

[Figura 7 - Criação da tabela Atleta 23](#_Toc28267466)

[Figura 8 - Criação da tabela TesteClinico 23](#_Toc28267467)

[Figura 9 - Criação da Tabela Médico 24](#_Toc28267468)

[Figura 10 - Criação da tabela Clínica 24](#_Toc28267469)

[Figura 11 - Criação da tabela Contacto 24](#_Toc28267470)

[Figura 12 - Criação da Tabela Prova\_TipoTeste 25](#_Toc28267471)

[13 - Criação da tabela Prova 25](#_Toc28267472)

[Figura 14 - Criação da tabela Atleta\_Prova 25](#_Toc28267473)

[Figura 15 - Criação da tabela Morada 25](#_Toc28267474)

# Índice de Tabelas

[Tabela 1 – Dicionário de dados das entidades 6](#_Toc28267475)

[Tabela 2 - Dicionário de relacionamentos 9](#_Toc28267476)

[Tabela 3 - Caraterização de todos os atributos existentes 10](#_Toc28267477)

[Tabela 4 - Representação da entidade Atleta 14](#_Toc28267478)

[Tabela 5 - Representação da entidade TesteClinico 15](#_Toc28267479)

[Tabela 6 - Representação da entidade Medico 15](#_Toc28267480)

[Tabela 7- Representação da entidade Clínica 15](#_Toc28267481)

[Tabela 8 - Representação da entidade Prova 15](#_Toc28267482)

[Tabela 9 - Espaço ocupado em disco por cada tipo de dados 26](#_Toc28267483)

[Tabela 10 - Espaço ocupado pela tabela Atleta 27](#_Toc28267484)

[Tabela 11 - Espaço ocupado pela tabela Clínica 27](#_Toc28267485)

[Tabela 12 - Espaço ocupado pela tabela Médico 27](#_Toc28267486)

[Tabela 13 - Espaço ocupado pela tabela Teste Clínico 28](#_Toc28267487)

[Tabela 14 - Espaço ocupado pela tabela Prova 28](#_Toc28267488)

[Tabela 15 - Espaço ocupado pela tabela Morada 28](#_Toc28267489)

[Tabela 16 - Espaço ocupado pela tabela Contacto 28](#_Toc28267490)

[Tabela 17 - Espaço ocupado pela tabela Atleta\_Prova 28](#_Toc28267491)

[Tabela 18 - Espaço ocupado pela tabela Prova\_TipoTeste 29](#_Toc28267492)

[Tabela 19 - Espaço ocupado em disco pela população atual 29](#_Toc28267493)

[Tabela 20 - Espaço ocupado em disco pela população no futuro 29](#_Toc28267494)

1. Introdução
   1. Contextualização

Antes de iniciar a prática de qualquer atividade física é importante realizar uma avaliação médica detalhada, especialmente quando um individuo deseja praticar desporto a nível competitivo, no qual terá que realizar um esforço físico intenso ou contínuo.

Atletas de alta competição devem realizar regularmente exames médicos como uma medida preventiva essencial à proteção da sua saúde, através da avaliação da sua pré-disposição a lesões e deste modo contribuir para reduzir o risco de ocorrência. Estes testes são também fundamentais para avaliar a evolução das suas diferentes capacidades de adaptação e resposta do aparelho cardiorrespiratório a um trabalho muscular do tipo aeróbio ou de resistência, necessário na maioria dos desportos de alta competição.

Existe uma grande variedade de exames de alto rendimento para avaliar as capacidades e saúde do atleta que dependem da modalidade da competição.

É também necessário realizar, por vezes, testes para a deteção do uso de substâncias de melhoria de performance, cujo uso por desportistas de alta competição é proibido por lei.

Assim sendo, a *Clínica Alta Performance* nasceu com o objetivo de permitir a qualquer atleta de alta competição realizar todos os seus exames médicos periódicos com acompanhamento especializado face às suas necessidades desportivas. É obviamente uma prioridade empresarial automatizar todos os processos, manter um histórico da condição física e saúde dos atletas promovendo uma prática desportiva segura oferecendo orientação e preparação física de modo a que os pacientes possam obter o máximo rendimento das suas qualidades.

* 1. Apresentação do Caso de Estudo

A *Clínica Alta Performance* é uma empresa composta por um conjunto de clínicas que se encontram em 3 pontos do país.

A clínica disponibiliza todo o tipo de exames médicos necessários para atestar a saúde do atleta...

É ainda de destacar o fator notável que distingue esta empresa da sua concorrência, o seu acompanhamento ......

* 1. Fundamentação da Implementação da Base de Dados

<<Esta secção acolherá os diversos motivos, acompanhados por uma breve descrição, que conduziram à proposta e ao desenvolvimento do trabalho, assim como a apresentação detalhada dos diversos objectivos a alcançar com a sua realização.>>

* 1. Análise da viabilidade do projeto

Este projeto consiste na implementação de uma Base de Dados Relacional que possibilita estudar e relacionar os dados que a empresa possui de modo a melhorar o processo de gestão de agendamentos e de dados dos pacientes. Possuir uma Base de Dados Relacional é essencial pois permite a manipulação e a extração de conhecimento de uma forma metodológica, podendo assim assegurar a consistência dos dados uma vez que, em larga medida, a necessidade de redundância desaparece. A nova base de dados, como será implementada num sistema de gestão de base de dados relacional baseado no uso de transações, permitirá efetuar inúmeras sequências de operações em simultâneo sem que a consistência e integridade dos dados seja sacrificada.

O investimento na criação de uma Base de Dados Relacional permitirá à empresa aumentar a qualidade do seu serviço, uma vez que permite uma maior facilidade no acesso à informação de um cliente de modo a conhecer as suas limitações físicas e todo o historial de doenças e lesões e sobretudo ser capaz de avaliar o estado de saúde do atleta antes de competir.

* 1. Estrutura do Relatório

O relatório do projeto é constituído por 5 capítulos.

O primeiro e atual capítulo é tratada a **Definição do sistema.** É apresentada a contextualização do problema, os objetivos e a motivação para a realização do projeto, incluindo também a análise da sua viabilidade.

No segundo capítulo, é feito o **Levantamento e Análise de Requisitos,** é apresentado o método utilizado e posteriormente descritos os vários tipos de requisitos e a sua análise.

No terceiro capítulo, é demonstrado todo o processo a realizar para a implementação da nossa **Base de Dados Relacional**. É apresentada a **Modelação Conceptual**, onde são identificadas as entidades, os seus relacionamentos e atributos. Seguidamente, é feita a **Modelação** **Lógica**, nomeadamente, a derivação de relações do modelo lógico e validação destas através da normalização e verificação das restrições de integridade. Por ultimo é exposto o processo de **Implementação física.**

No quarto capítulo, **Base de Dados Não Relacional** que teve como base a implementada no capítulo anterior. Para tal, demonstramos a realização da migração dos dados para o Neo4j, assim como a realização das queries definidas anteriormente.

Finalmente, é apresentada a **Conclusão** do projeto.

1. Levantamento e Análise de requisitos
   1. Método de Levantamento e de Análise de Requisitos Adotados

Quanto à abordagem no levantamento dos requisitos, decidimos que a mais adequada a seguir seria a de agrupar todos os que dizem respeito a cada perfil de utilização de maneira a formar um único conjunto correspondente à totalidade da base de dados. Para isso recorremos a várias técnicas de *fact-finding* para tentar perceber o comportamento do sistema atual com o intuito de garantir que a resultado final satisfaz as necessidades dos utilizadores.

* 1. Requisitos Levantados
  2. Análise de Requisitos

1. Modelação Conceptual
   1. Apresentação da Abordagem de Modelação Realizada

Após o levantamento e análise de requisitos do sistema começa a fase do planeamento do design da Base de Dados. Para que a base de dados que será construída esteja de acordo com as necessidades dos seus utilizadores finais recorremos ao **diagrama ER** para efetuar a modelação conceptual. O objetivo do modelo conceptual é representar os dados presentes nos requisitos do utilizador.

O diagrama ER utiliza uma abordagem *top-down*, isto é, começamos por identificar os dados relevantes ao problema, as entidades, os relacionamentos entre esses dados, seguidamente completamos a informação do modelo com os seus atributos, quer nas entidades, que nos relacionamentos entre as mesmas e respetivos domínios. Finalmente, adicionamos a cardinalidade dos relacionamentos. Ao longo da sua construção é fundamental que o modelo seja testado e validado com o utilizador e com os seus requisitos, assegurando assim, a coerência e capacidade de resolução de problemas da futura base de dados.

Um aspeto muito importante desta metodologia é também o facto de ser completamente independente dos futuros detalhes de implementação.

* 1. Identificação e Caracterização das Entidades

Atendendo aos requisitos levantados foram identificadas as entidades apresentadas em seguida.

**Atleta**

O atleta é paciente da clínica que será submetido a diversos testes. Cada atleta possui atributos próprios, tem uma existência autónoma e pode ser identificado univocamente, sendo assim, uma entidade. É representado pelo seu **id**, **nome**, **sexo**, **data de nascimento**, **morada** e **contacto**.

**Teste Clínico**

O teste clínico refere-se a qualquer tipo de teste ou exame realizado pelo atleta. É representado pelo seu **id**, **designação** (nome do exame), **resultado, data e hora** em que foi ou será realizado, **preço**, **id do médico** que realizou ou realizará o exame e o **id do paciente** que se refere ao atleta sujeito ao teste.

**Médico**

A entidadeMédico representa o funcionário responsável pela realização dos testes clínicos. É caraterizado pelo seu **id**, **nome**, **data de nascimento**, a sua **morada**, **data de início de serviço**, **especialidade, id da Clínica** em que trabalha, e ainda o **id do chefe.**

**Clínica**

A entidadeClínica representa cada uma das clínicas da empresa. É caraterizada pelo seu **id**, **nome** e **morada**.

**Prova**

A entidade prova refere-se às provas em que os atletas competem. É caraterizada pelo **id**, **designação** (nome da prova), **modalidade** (tipo de prova), **categoria,** que se refere ao intervalo de idades dos atletas para essa prova e ainda **tipoTesteClinico** que representa os vários tipos de testes clínicos que são obrigatórios para um atleta poder competir nesta.

**3.2.1 Dicionário de dados das entidades**

|  |  |  |  |
| --- | --- | --- | --- |
| Entidade | Descrição | Alcunha | Ocorrência |
| Atleta | Entidade que recorre aos serviços da clínica | Paciente | O atleta é uma entidade fundamental uma vez que é este que recorre à clínica. |
| Teste Clínico | Entidade que contém toda a informação sobre um exame médico realizado pelo atleta | Exame | O teste clínico é o motivo pelo qual os pacientes recorrem aos serviços da clínica. |
| Médico | Entidade que representa o responsável pela realização de testes | - | O médico é a base do funcionamento dos exames realizados na clínica. |
| Clínica | Entidade que representa o local em que um atleta é submetido a um determinado exame | - | A clínica é o local onde os pacientes são submetidos aos testes clínicos. |
| Prova | Entidade que contém a informação das provas em que o atleta compete | - | A prova é a atividade que conduz à necessidade de efetuar testes clínicos. |

Tabela 1 – Dicionário de dados das entidades

* 1. Identificação e Caracterização dos Relacionamentos

**Relacionamento: Atleta – Teste Clínico**

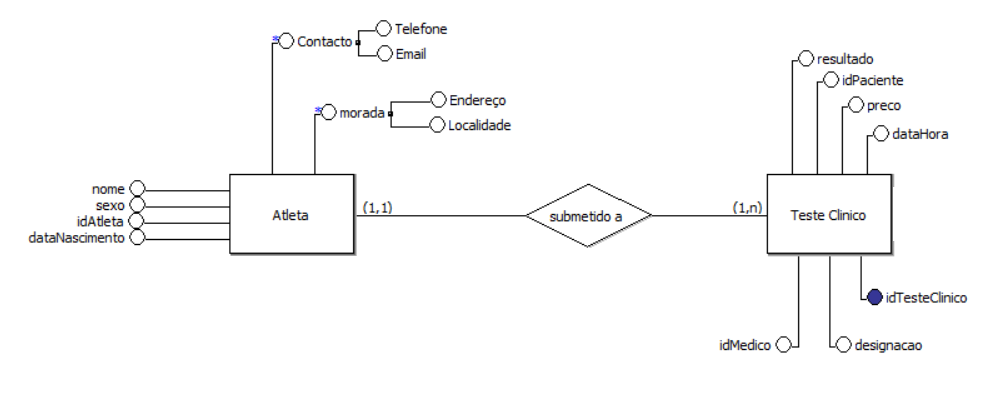


Figura 1 - Relacionamento Atleta - Teste Clínico

**Relacionamento:** Atleta é submetido a um Teste Clínico

**Descrição:** Os atletas recorrem à Clínica Alta Performance para realizar exames médicos. Assim, é importante preservar a informação dos testes médicos anteriormente realizados, para além de salvaguardar a informação de agendamentos futuros**.**

**Cardinalidade:** N para N

Um atleta tem que realizar pelo menos um teste clínico ou ter algum teste agendado para ser uma paciente da clínica. Entretanto podem ser realizados diferentes testes clínicos ao mesmo tempo.

**Atributos**: Este relacionamento não possui atributos.

**Relacionamento: Teste Clínico – Médico**

Uma imagem com mapa, texto

Descrição gerada automaticamente

Figura 2 - Relacionamento Teste Clínico – Médico

**Relacionamento:** Teste Clínico é efetuado pelo Médico

**Descrição:** O medico é a entidade responsável pela supervisão e realização do exame do atleta. É importante ser capaz de identificar a identidade do medico que realizou determinado exame não só para o acompanhamento ser consistente e permitir que o paciente seja examinado pelo mesmo médico mas também para em caso de negligencia poder apurar os responsáveis.

**Cardinalidade:** 1 para N

Um teste clínico é realizado somente por um médico. Um médico, por sua vez, pode realizar vários exames.

**Atributos:** Este relacionamento não possui atributos.

**Relacionamento: Médico – Clínica**

Uma imagem com texto

Descrição gerada automaticamente

Figura 3 - Relacionamento Médico - Clínica

**Relacionamento:** Médico trabalha na Clínica

**Descrição:**  O médico é a entidade que trabalha na clínica. Uma vez que a Clínica Alta Performance é uma empresa constituída por 3 clínicas diferentes é importante identificar o local de trabalho de cada médico.

**Cardinalidade:** N para 1

Vários médicos trabalham numa clínica, contudo cada médico apenas pode trabalhar numa clínica.

**Atributos:** Este relacionamento não possui atributos.

**Relacionamento: Atleta – Prova**

Uma imagem com texto, mapa

Descrição gerada automaticamente

Figura 4 - Relacionamento Atleta - Prova

**Relacionamento:** Atleta compete numa Prova

**Descrição:** Um Atleta compete em provas distintas, que por sua vez, necessitam da realização de testes distintos para apurar a condição física do atleta que vai competir, assim através deste relacionamento conseguimos identificar os testes a que determinado atleta se submeteu e verificar o seu estado de saúde.

**Cardinalidade:** N para N

Várias atletas podem competir em várias provas.

**Atributos:** Este relacionamento não possui atributos.

* + 1. Dicionário de relacionamentos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Entidade | Multiplicidade | Relacionamento | Multiplicidade | Entidade |
| Atleta | 1..N | Compete | 1..N | Prova |
| Atleta | 1..1 | Submetido a | 1..N | Teste |
| Teste | 1..1 | Realizado por | 1..1 | Médico |
| Médico | 1..N | Trabalha | 1..1 | Clínica |

Tabela 2 - Dicionário de relacionamentos

* 1. Identificação e Caracterização da Associação dos Atributos com as Entidades e Relacionamentos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Entidade** | **Atributos** | **dESCRIÇÃO** | **tipo do atributo** | **tipo de dados e tamanho** |
| **Atleta** | Id | Identificador do atleta | Chave primária | INT |
| Nome | Nome do atleta | Simples | VARCHAR(100) |
| Sexo | Género do atleta | Simples | VARCHAR(5) |
| Data de Nascimento | Data de Nascimento do atleta | Simples | DATE |
| Contacto:  Email  Telefone |  | Composto |  |
| Email do atleta | VARCHAR(45) |
| Número de telefone do atleta | INT |
| Morada:  Endereço  Localidade |  | Composto |  |
| Endereço postal do atleta | VARCHAR(100) |
| Localidade do atleta | VARCHAR(100) |
| **Teste Clínico** | IdTesteClínico | Identificador do teste clínico | Chave primária | INT |
| DataHora | Data e hora da realização do teste clínico | Simples | DATE |
| Preço | Preço do exame | Simples | DECIMAL(6,2) |
| Designação | Nome do exame | Simples | VARCHAR(100) |
| IdAtleta | Identificador do atleta submetido ao exame | Simples | INT |
| IdMedico | Identificador do médico que realiza o exame | Simples | INT |
| **Médico** | IdMédico | Identificador do médico | Chave primária | INT |
| Nome | Nome do médico | Simples | VARCHAR(100) |
| Data de Nascimento | Data de Nascimento do médico | Simples | DATE |
| Morada  Endereço  Localidade |  | Composto |  |
| Endereço postal do médico | VARCHAR(100) |
| Localidade do médico | VARCHAR(100) |
| Especialidade | Área de especialização do médico | Simples |  |
| Data de Inicio de Serviço | Data de inicio de serviço do médico na clínica | Simples | DATE |
| IdHospital | Identificador do hospital | Simples | INT |
| **Clínica** | IdClinica | Identificador da clínica | Chave primária | INT |
| Nome | Nome da Clínica | Simples |  |
| Morada  Endereço  Localidade |  | Composto |  |
| Endereço postal da clínica | VARCHAR(100) |
| Localidade da clínica | VARCHAR(100) |
| **Prova** | IdProva | Identificador da prova | Chave primária | INT |
| tipoTesteClínico | Tipos de testes clínicos necessários para poder realizar a prova | Multivalorado | VARCHAR(100) |
| Categoria | Intervalo de idades dos atletas que podem competir numa prova | Simples | VARCHAR(100) |
| Designação | Nome da prova | Simples | VARCHAR(100) |
| Modalidade | Tipo de prova | Simples | VARCHAR(100) |

Tabela 3 - Caraterização de todos os atributos existentes

* 1. Chaves Candidatas, Primárias e Alternativas

Uma vez identificadas as entidades, os seus atributos e relacionamentos, podemos agora verificar quais desses atributos poderão ser chaves primárias. É importante referir que uma chave primária corresponde a um atributo, que pode identificar a identidade da entidade inequivocamente. As chaves candidatas são todos os atributos da entidade que a identificam univocamente e os que não forem escolhidos para chave primária denominam-se chaves alternativas.

**Atleta**

Chaves candidatas: IdAtleta, Contacto

Chave Primária: IdTesteClínico

Chave Alternativa: Contacto

**Teste**

Chaves candidatas: IdTesteClínico

Chave Primária: IdTeste

Chave Alternativa: Nenhuma

**Médico**

Chaves candidatas: IdMédico

Chave Primária: IdMédico

Chave Alternativa: Nenhuma

**Clínica**

Chaves candidatas: IdClínica, Morada

Chave Primária: IdClínica

Chave Alternativa: Morada

**Prova**

Chaves candidatas: IdProva

Chave Primária: IdProva

Chave Alternativa: Nenhuma

* 1. Detalhe ou Generalização de Entidades

No nosso projeto não utilizamos generalização nem especialização de entidades no Modelo Conceptual.

* 1. Apresentação e Explicação do Diagrama ER

Depois de explicada a importância e o significado de cada uma das entidades existentes no nosso sistema, cada um dos relacionamentos entre elas e os respetivos atributos, tanto das entidades, como dos relacionamentos, apresentamos o modelo conceptual:

Uma imagem com texto, mapa

Descrição gerada automaticamente

Figura 5 - Modelo Conceptual

O funcionamento do nosso sistema começa a ganhar textura. Traduzindo o Modelo Conceptual ilustrado anteriormente para texto temos que: O Atleta que compete em provas é submetido a Testes Clínicos. Os Testes Clínicos são efetuados por um Médico que trabalha numa Clínica.

* 1. Validação do Modelo de Dados com o Utilizador

Uma vez terminado o Modelo Conceptual, é necessário efetuar a sua validação. Para tal, é fundamental ser possível responder a todas as perguntas que o utilizador possa elaborar. Assim, escolhemos as queries posteriormente expostas, de modo a verificar a viabilidade com o Modelo Conceptual.

1. **Consultar os testes realizados antes de determinada prova para um dado atleta ordenados do mais recente para o mais antigo.**

Para obter a informação necessária para a respostas recorremos às entidades *Prova, Atleta* e *Teste Clínico*. Tendo em conta o relacionamento entre o *Atleta* e a *Prova*, temos acesso a todas as provas em que o atleta competiu e, por conseguinte, à sua data. Através do relacionamento entre o *Atleta* e o *Teste Clínico* obtemos todos os exames a que o atleta foi submetido e as suas datas. Assim é possível determinar quais os exames realizados antes da data da prova.

1. **Consultar os nomes dos atletas de determinada modalidade que acusaram positivo nos testes de dopping.**

Através das entidades *Atleta, Prova* e *Teste Clínico* conseguimos dar resposta a esta interrogação. Do relacionamento *Atleta* e *Prova* obtemos o nome e modalidade em que os atletas concorrem e da relação entre *Atleta* e *Teste Clínico* conseguimos descobrir se o resultado dos exames de dopping realizados alguma vez deu positivo.

1. **Consultar os nomes dos pacientes/atletas que foram submetidos a determinado teste**

Através das entidades *Teste Clínico* e *Atleta* conseguimos identificar todos os atletas submetidos a determinado teste.

1. **Consultar o número de testes clínicos realizados por um dado médico num determinado ano e o montante faturado.**

Através das entidades *Teste Clínico* e *Medico* é possível determinar o médico que realizou cada exame, a sua data e preço.

1. **Consultar agendamentos para um determinado tipo de exame numa determinada clínica.**

Através da entidade *Médico* conseguimos determinar a clínica em que determinado médico trabalha, podendo assim determinar através da junção deste com a entidade *TesteClinico* os tipos de exames efetuados.

1. Modelação Lógica

Terminada a conceptualização do problema apresentado através do modelo anteriormente elaborado, torna-se agora necessária a construção do mesmo num Modelo Lógico. Esta será uma etapa fundamental para o desenvolvimento do nosso sistema de gestão da base de dados uma vez que nos permitirá derivar os relacionamentos. Assim, quando o modelo lógico estiver validado, tendo em conta as formas normais, este irá suportar o nosso problema.

* 1. Construção e Validação do Modelo de Dados Lógico

Para a elaboração do Modelo Lógico, começamos por derivar/criar todas as relações que retratam as entidades, os relacionamentos e os atributos previamente identificados no Modelo Conceptual.

* + 1. Entidades fortes

Uma entidade forte identifica-se pela posse de uma de uma chave primária que a identifica e pela independência de outras entidades. No Modelo Logico, cada entidade forte será́ representada numa tabela, em que cada um dos seus atributos constitui uma coluna. Para os atributos compostos, na tabela apenas serão apresentados os atributos simples que o constituem.

**Entidades fortes resultantes:**

**Atleta** (idAtleta, nome, dataNascimento, sexo, morada)

**Chave primária:** idAtleta

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| idAtleta | Nome | Data de Nascimento | Sexo | Morada |
| 1 | Maria José Borges Pires | 27-10-1998 | F | Rua de Barros nº32, Braga |
| (...) | (...) | (...) | (...) | (...) |

Tabela 4 - Representação da entidade Atleta

**TesteClínico** (idTesteClinico, designação, dataHora, preço, resultado, idMedico, idAtleta)

**Chave primária:** idTesteClinico

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| idTesteClinico | Designação | Data e Hora | Preço | Resultado | idMedico | idAtleta |
| 1123 | Eletrocardiograma | 10-12-19 10h30 | 50,00 | Ok | 250 | 12 |
| (...) | (...) | (...) | (...) | (...) | (...) | (...) |

Tabela 5 - Representação da entidade TesteClinico

**Médico** (idMedico, nome, dataNascimento, morada, especialidade, dataInicioServico, idClinica, idChefe)

**Chave Primária:** idMedico

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| idMedico | Nome | Data de Nascimento | Morada | Especialidade | Data de Inicio de Serviço | idClinica | idChefe |
| 201 | Susana Marques | 01-04-1986 | Rua 25 de Abril nº12, Braga | Cardiologia | 03-10-2002 | 2 | 100 |
| (...) | (...) | (...) | (...) | (...) | (...) | (...) |  |

Tabela 6 - Representação da entidade Medico

**Clínica** (idClinica, nome, morada)

**Chave primária:** idClinica

|  |  |  |
| --- | --- | --- |
| idClinica | Nome | morada |
| 1 | Clínica Alta Performance - Braga | Sete Fontes - São Victor, Braga |
| (...) | (...) | (...) |

Tabela 7- Representação da entidade Clínica

**Prova** (idProva, designação, modalidade, categoria)

**Chave primária:** idProva

|  |  |  |  |
| --- | --- | --- | --- |
| idProva | Designação | Modalidade | Categoria |
| 1 | Corrida de Obstáculos: 100m | Corrida de Obstáculos | 18-21 |
| (...) | (...) | (...) | (...) |

Tabela 8 - Representação da entidade Prova

* + 1. Relacionamentos 1:N

Neste tipo de relacionamento, a entidade que apresenta multiplicidade N possui como atributo a chave primária da entidade com multiplicidade 1. Este atributo é considerado uma chave estrangeira.

**TesteClinico** (idTesteClinico, designação, data, hora, preço, resultado, idMedico, idAtleta)

**Chave primária:** idTesteClinico

**Chaves estrangeiras:** idMedico, idAtleta

**Médico** (idMedico, nome, dataNascimento, morada, especialidade, dataInicioServico, idClinica, idChefe)

**Chave primária:** idMedico

**Chaves estrangeiras:** idClinica, morada

Como os atletas, médicos e cada clínica possuem uma morada, o atributo morada e este irá repetir-se algumas vezes, sentimos a necessidade de criar uma tabela *Morada*. A nova tabela terá́ como atributos o endereço, a localidade e o idMorada, este último corresponde à chave primária e que será́ uma chave estrangeira em Médico, como visto anteriormente e nas tabelas *Clinica* e *Atleta*, como se pode verificar de seguida.

**Clínica** (idClinica, nome, morada)

**Chave primária:** idClinica

**Chaves estrangeiras:** morada

**Atleta** (idAtleta, nome, dataNascimento, sexo, morada)

**Chave primária:** idAtleta

**Chaves estrangeiras:** morada

* + 1. Relacionamentos N:M

Este tipo de multiplicidade gera um novo relacionamento, onde a chave primária é composta pelas chaves primárias de cada uma das entidades envolvidas.

**Atleta\_Prova** (id\_Atleta, id\_Prova, data)

**Chave primária:** id\_Atleta, id\_Prova

* + 1. Atributos Multivalorados

Este tipo de atributo é utilizado, quando para a mesma entidade, o atributo pode assumir múltiplos valores. Um atributo multivalorado cria um novo relacionamento, em que a chave primária é composta pela chave primária da entidade e o próprio atributo.

**Contacto** (idMorada, telefone, email, idAlteta)

**Chave primária:** idMorada

**Chaves estrangeiras:** idAtleta

**Prova\_TipoTeste** (idProva, designcação)

**Chave primária:** Não possui uma chave primária, a sua existência depende da entidade Prova

**Chaves estrangeiras:** idProva

* 1. Desenho do Modelo Lógico

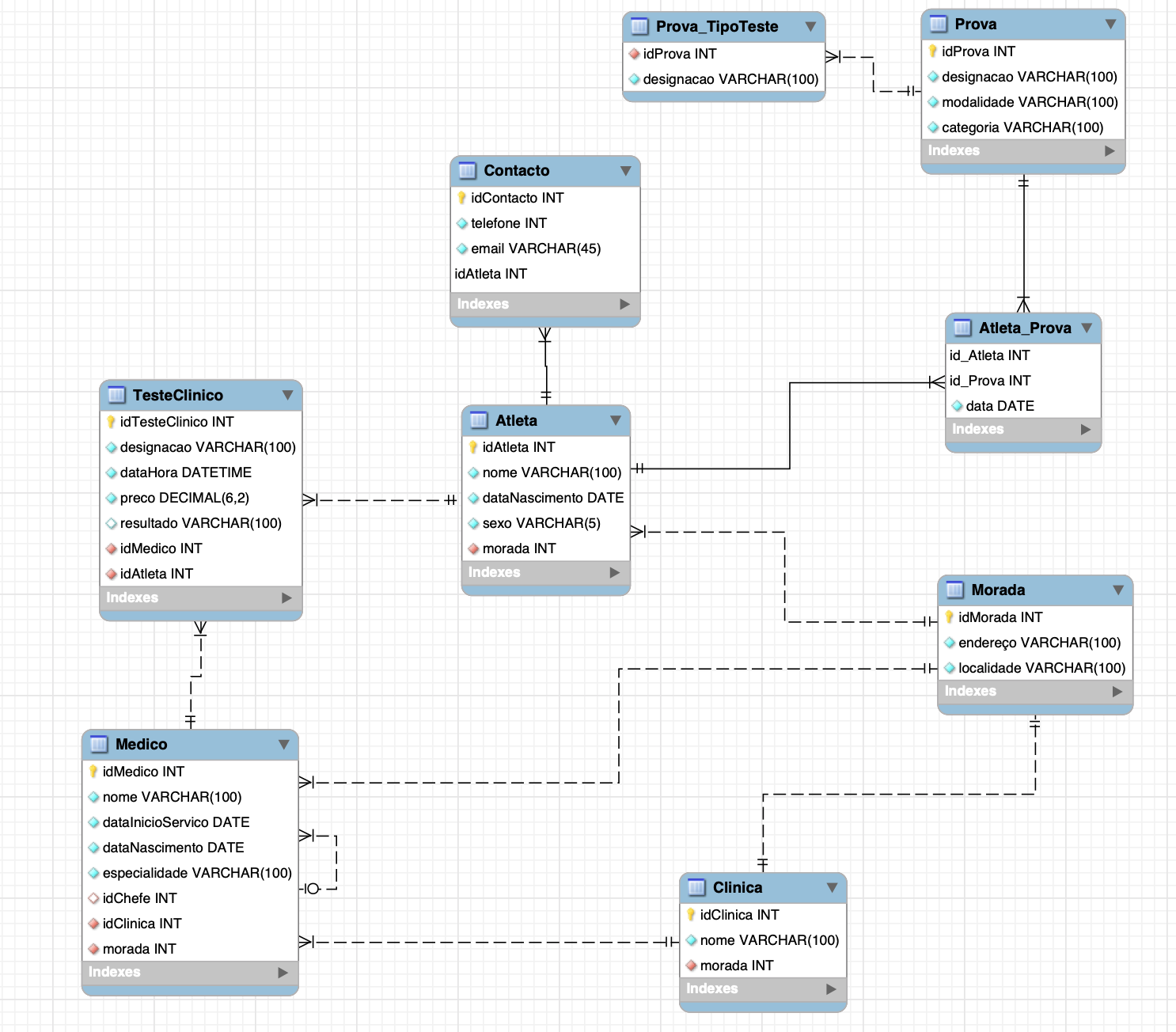


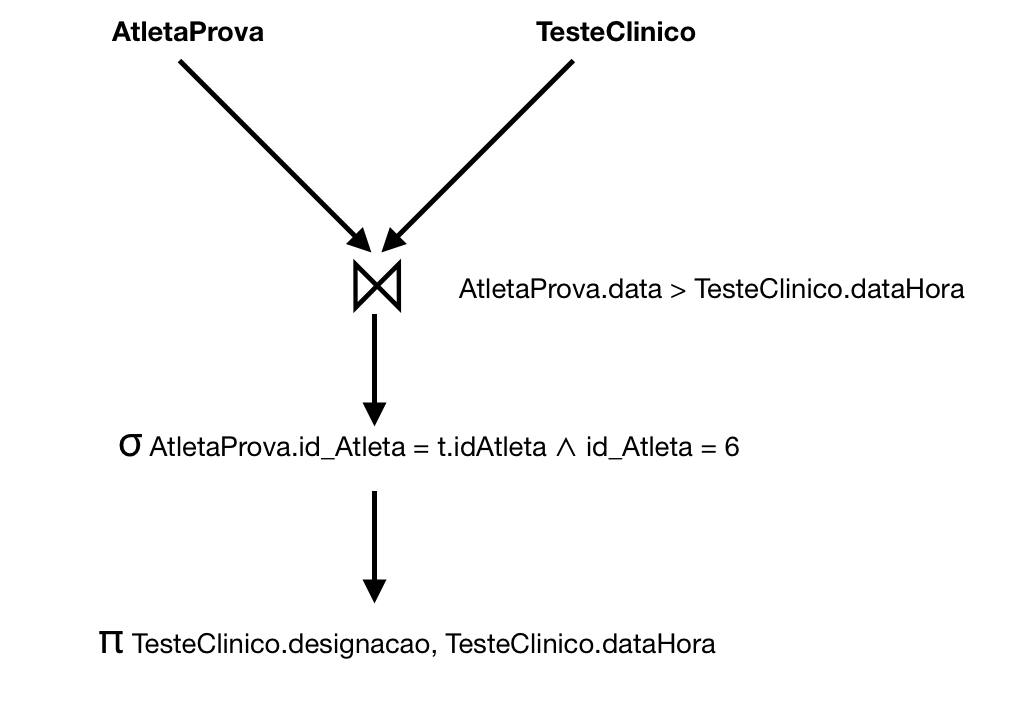
Figura 6 - Modelo Lógico

* 1. Validação do Modelo através da Normalização

No modelo relacional de uma base de dados o objetivo é agrupar os atributos em relações, uma vez que existem dependências entre estes. Assim, é possível construir um modelo lógico consistente em termos estruturais e com a menor redundância possível. Uma vez que a normalização tem como objetivo garantir que as relações têm um número mínimo, mas suficiente de atributos necessários para suportar as exigências relativas aos dados a guardar, podemos perceber melhor cada atributo e o que este representa na base de dados. Para proceder à validação do modelo através de normalização, começamos por identificar as dependências funcionais entre os atributos.

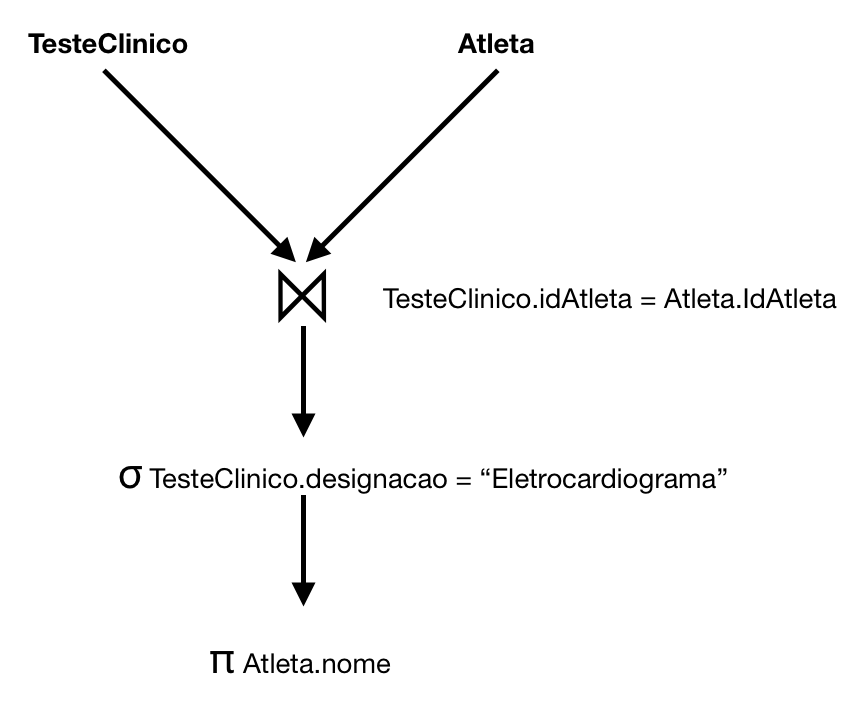
* + 1. Primeira Forma Normal (1FN)
    2. Segunda Forma Normal (2FN)
    3. Terceira Forma Normal (3FN)
  1. Validação do Modelo com as Interrogações do Utilizador

1. Consultar os testes realizados antes de determinada prova para um dado atleta, ordenados do mais recente para o mais antigo.

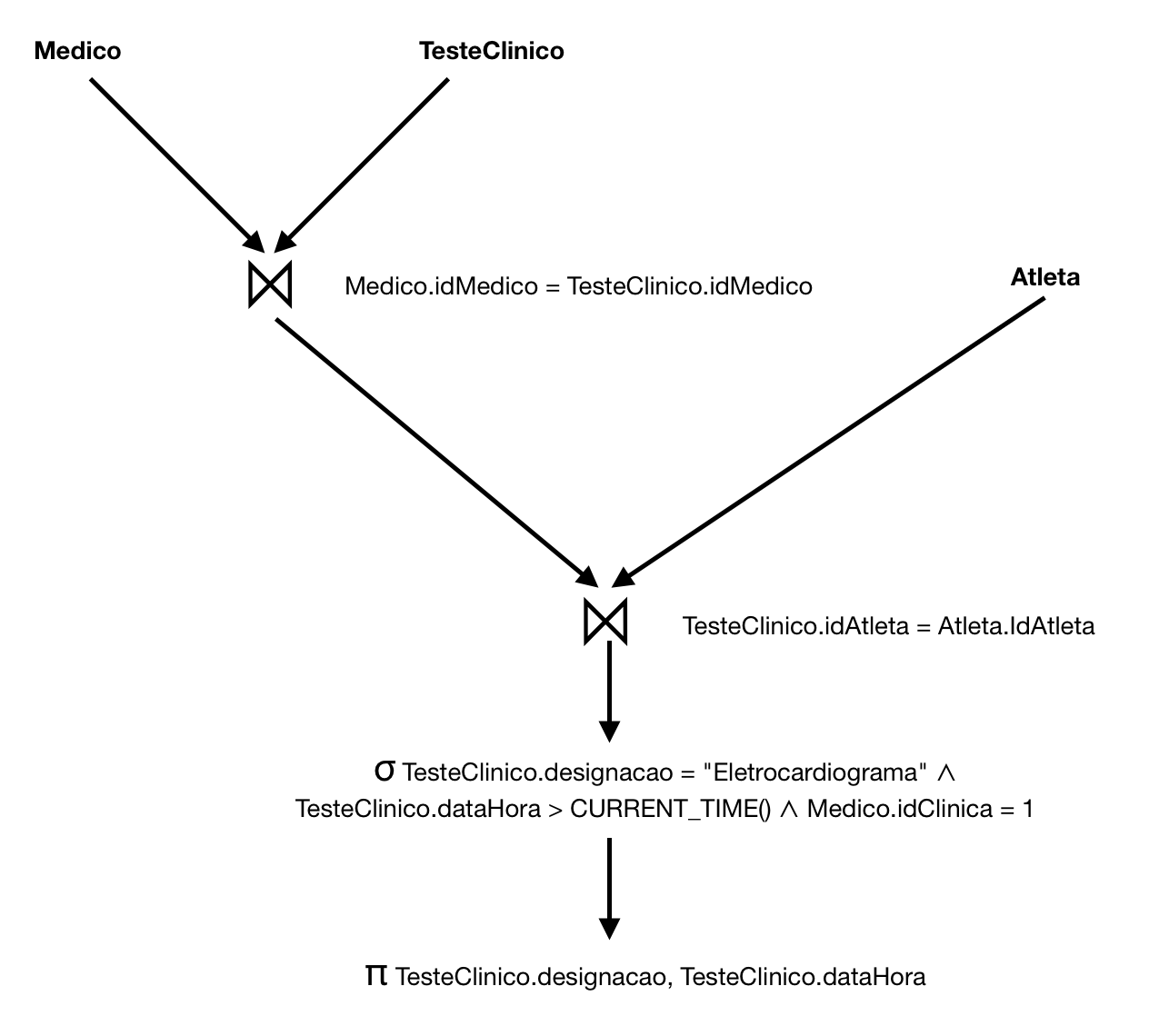


1. Consultar os nomes dos atletas de determinada modalidade que acusaram positivo nos testes de dopping. ∧
2. Consultar os nomes dos pacientes/atletas que foram submetidos a determinado teste.

πnome(Atleta ⨝ ( πidAtleta(σdesignacao=’Eletrocardiograma’(TesteClinico))))



1. Consultar o número de testes clínicos realizados por um dado médico num determinado ano e o montante faturado.
2. Consultar agendamentos para um determinado tipo de exame numa determinada clínica.



* 1. Validação do Modelo com as Transações Estabelecidas

Foram estabelecidas algumas transações que serão posteriormente utilizadas pelos utilizadores (com as devidas permissões) da base de dados. Por se tratarem de transações, fica assegurado que cada uma destas é interpretada pelo sistema de base de dados como um bloco linear e atómico, pois as operações dentro de cada transação serão efetuadas sequencialmente nos dados. Deste modo, fica então garantido que as tabelas envolvidas na transação não serão editadas entre cada operação no bloco da transação por uma outra operação que interaja com esses mesmos dados.

* **Registar um atleta:** Para registar um atleta começa-se por elaborar um novo registo na tabela *Atleta,* adicionando a morada na tabela *Morada* caso essa ainda não exista. De seguida, é necessário inserir os contactos telefónicos e email do mesmo, efetuando um novo registo na tabela *Contactos*. Finalmente, é preenchida a tabela *Atleta\_Prova*, com as provas futuras que levaram o atleta a recorrer aos serviços da clínica.

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 7 - Mapa de Transações para registar um atleta

* **Efetuar um agendamento:** Para efetuar um agendamento é necessário adicionar uma nova entrada na tabela *TesteClinico.* Esta entrada possui uma data futura e o resultado do teste é *null.*

**Uma imagem com captura de ecrã

Descrição gerada automaticamente**

Figura 8 - Tabela TesteClinico

* **Adicionar uma prova:** Para adicionar um novo tipo de prova ao sistema, adiciona-se primeiro uma entrada na tabela *Prova* ede seguida são registados os exames obrigatórios para competir nessa prova na tabela *Prova\_TipoTeste.*

**Uma imagem com captura de ecrã

Descrição gerada automaticamente**

Figura 9 - Mapa de Transações para adicionar uma prova

* 1. Verificação das Restrições de Integralidade

Para garantir que o modelo lógico representa os dados de forma fiável, é necessário a aplicar algumas regras, nomeadamente, as restrições de integridade, sendo possível assim garantir, que a base de dados é consistente e completa. As restrições aplicadas são as seguintes:

**Dados Necessários:** Numa Base de Dados existem campos de preenchimento obrigatório e que devem possuir um valor válido, como é o caso das chaves primarias. No SGBD que estamos a implementar, todos os campos são de preenchimento obrigatório.

**Integridade de Entidade:** Para que todas as entidades sejam identificadas e que não exista repetição das mesmas, é necessário garantir a existência de uma chave primária, tendo esta de ser única e não nula. Como as chaves primárias definidas no Modelo Conceptual transitam para o modelo lógico, é possível verificar as que identificam cada entidade na seção *3.5.*

**Integridade de Domínio:** Esta restrição refere-se aos atributos, dado que todos possuem um domínio de valores admissíveis. É possível observar o tipo de dados de cada atributo no dicionário de dados dos atributos das entidades e relacionamentos.

**Integridade Referencial:** Esta restrição tem por base a relação entre chaves estrangeiras e primárias, de modo a salvaguardar os relacionamentos entre as tabelas quando é efetuado uma inserção, atualização ou remoção de um registo. Esta garante que os dados permanecem consistentes em todas as tabelas.

**Restrições de Multiplicidade:** As multiplicidades são definidas na criação do modelo concetual, e como tal devem transitar para o modelo lógico. Contudo, existem casos particulares como os atributos multivalorados que devem ser analisados, como é retratado na seção *4.1*, onde é possível encontrar os relacionamentos estabelecidos.

* 1. Revisão do Modelo Lógico com o Utilizador

Após concluído o desenvolvimento do modelo lógico foi agendada nova reunião com a *Clínica Alta Performance* de modo a apresentar todo o trabalho desenvolvido e obter autorização para avançar para a próxima fase do projeto. Uma vez que este modelo é significativamente mais complexo que o modelo conceptual, o primeiro passo foi apresentar em traços gerais como analisar e interpretar este tipo de modelos, para que o cliente possa, de maneira fiável, confirmar que o modelo atende às necessidades da empresa. Cumpridas todas as exigências do cliente, ele deu autorização para avançar para a próxima fase do projeto.

1. Implementação Física
   1. Seleção do Sistema de Gestão de Base de Dados

O sistema de gestão de bases de dados utilizado para o nosso projeto foi o *MySQL.* A nossa escolha baseou-se essencialmente no facto de este ser de fácil utilização, possuir um bom desempenho e ser *open-source*. Adicionalmente, o facto de o sistema de armazenamento e gestão de transações do *MySQL* seguir os princípiosACID (Atomicidade, Consistência, Isolamento e Durabilidade) também contribuiu positivamente para a nossa escolha.

* 1. Tradução do Esquema Lógico para o Sistema de Gestão de Bases de Dados escolhido em SQL

Durante todo o processo decidimos usar a ferramenta de desenho, desenvolvimento e administração de base de dados, *MySQL Workbench*. Para gerar o modelo físico através do modelo logico recorremos à ferramenta *Forward Engineer*. Este mecanismo, tal como o seu nome elucida permite automatizar o processo *top-down* de construção de componentes de baixo nível, neste caso a implementação no sistema de gestão de base de dados, através de uma abstração de alto nível, o modelo lógico. Tendo obtido o código resultante deste processo apenas tivemos que o executar para que a base de dados fosse criada.

* + 1. Desenho das relações base

**Relação Atleta**

Domínio IdAtleta: Inteiro

Domínio nome: String de tamanho variável

Domínio dataNascimento: Data

Domínio sexo: String de tamanho fixo

Domínio morada: Inteiro

**Atleta**(

idAtleta INT NOT NULL,

nome VARCHAR(100) NOT NULL,

dataNascimento DATE NOT NULL,

sexo VARCHAR(5) NOT NULL,

morada INT NOT NULL,

**PRIMARY KEY** (idAtleta),

**FOREIGN KEY** (morada)

**REFERENCES** Morada(idMorada));

**Relação Médico**

Domínio idMedico: Inteiro

Domínio nome: String de tamanho variável

Domínio dataInicioServico: data

Domínio dataNascimento: data

Domínio especialidade: String de tamanho variável

Domínio idChefe: Inteiro

Domínio idClinica: Inteiro

Domínio morada: Inteiro

**Medico**(

idMedico INT NOT NULL,

nome VARCHAR(100) NOT NULL,

dataInicioServico DATE NOT NULL,

dataNascimento DATE NOT NULL,

especialidade VARCHAR(100) NOT NULL,

idChefe INT NULL,

idClinica INT NOT NULL,

morada INT NOT NULL,

**PRIMARY KEY**(idMedico),

**FOREIGN KEY**(idChefe) **REFERENCES**  Medico (idMedico)

**FOREIGN KEY**(idClinica) **REFERENCES** Clinica(idClinica)

**FOREIGN KEY**(morada) **REFERENCES** Morada(idMorada));

**Relação TesteClinico**

Domínio idTesteClinico: Inteiro

Domínio designação: String de tamanho variável

Domínio dataHora: Data e hota

Domínio preco: Número décimal

Domínio resultado: String de tamanho variável

Domínio idMedico: Inteiro

Domínio idAtleta: Inteiro

**TesteClinico**(

idTesteClinico INT NOT NULL,

designacao VARCHAR(100) NOT NULL,

dataHora DATETIME NOT NULL,

preco DECIMAL(6,2) NOT NULL,

resultado VARCHAR(100) NULL,

idMedico INT NOT NULL,

idAtleta INT NOT NULL,

**PRIMARY KEY**(idTesteClinico),

**FOREIGN KEY**(idMedico) **REFERENCES** Medico(idMedico)

**FOREIGN KEY**(idAtleta) **REFERENCES** Atleta(idAtleta));

**Relação Clínica**

Domínio idClinica: Inteiro

Domínio nome: String de tamanho variável

Domínio morada: Inteiro

**Clinica**(

idClinica INT NOT NULL,

nome VARCHAR(100) NOT NULL,

morada INT NOT NULL,

**PRIMARY KEY** (idClinica),

**FOREIGN KEY** (morada) **REFERENCES** Morada(idMorada));

**Relação Morada**

Domínio idMorada: Inteiro

Domínio endereço: String de tamanho variável

Domínio localidade: String de tamanho variável

**Morada(**

idMorada INT NOT NULL,

endereço VARCHAR(100) NOT NULL,

localidade VARCHAR(100) NOT NULL,

**PRIMARY KEY** (idMorada));

**Relação Prova**

Domínio idProva: Inteiro

Domínio designação: String de tamanho variável

Domínio modalidade: String de tamanho variável

Domínio categoria: String de tamanho variável

**Prova**(

idProva INT NOT NULL,

designacao VARCHAR(100) NOT NULL,

modalidade VARCHAR(100) NOT NULL,

categoria VARCHAR(100) NOT NULL,

**PRIMARY KEY** (idProva));

**Relação Atleta\_Prova**

Domínio idAtleta: Inteiro

Domínio idProva: Inteiro

Domínio data: Data

**Atleta\_Prova**(

id\_Atleta INT NOT NULL,

id\_Prova INT NOT NULL,

data DATE NOT NULL,

PRIMARY KEY (id\_Prova, id\_Atleta),

**FOREIGN KEY** (id\_Atleta) **REFERENCES** Atleta (idAtleta)

**FOREIGN KEY** (id\_Prova) **REFERENCES** Prova(idProva));

**Relação Prova\_TipoTeste**

Domínio idProva: inteiro

Domínio designação: String de tamanho variável

**Prova\_TipoTeste**(

idProva INT NOT NULL,

designacao VARCHAR(100) NOT NULL,

**FOREIGN KEY** (idProva) **REFERENCES** Prova(idProva));

**Relação Contacto**

Domínio idContacto: Inteiro

Domínio telefone: Inteiro

Domínio email: String de tamanho variável

Domínio idAtleta: Inteiro

**Contacto**(

idContacto INT NOT NULL,

telefone INT NOT NULL,

email VARCHAR(45) NOT NULL,

idAtleta INT NOT NULL,

**PRIMARY KEY** (idContacto, idAtleta),

**FOREIGN KEY** (idAtleta) **REFERENCES** Atleta(idAtleta));

* + 1. Desenho das restrições

Nesta parte serão expostas as restrições gerais do problema agregadas por relação.

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 10 - Criação da tabela Atleta

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 11 - Criação da tabela TesteClinico

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 12 - Criação da Tabela Médico

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 13 - Criação da tabela Clínica

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 14 - Criação da tabela Contacto

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 15 - Criação da Tabela Prova\_TipoTeste

Uma imagem com captura de ecrã

Descrição gerada automaticamente

16 - Criação da tabela Prova

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 17 - Criação da tabela Atleta\_Prova

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 18 - Criação da tabela Morada

* 1. Tradução das Interrogações do Utilizador para SQL

De seguida serão apresentadas as respostas a algumas das interrogações realizadas pelo utilizador.

1. **Consultar os testes realizados antes de determinada prova para um dado atleta ordenados do mais recente para o mais antigo.**

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 19 - Resolução da interrogação do utilizador nº1

1. **Consultar os nomes dos atletas de determinada modalidade que acusaram positivo nos testes de dopping.**

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 19 - Resolução da interrogação do utilizador nº2

1. **Consultar os nomes dos pacientes/atletas que foram submetidos a determinado teste.**

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 20 - Resolução da interrogação do utilizador nº3

1. **Consultar o número de testes clínicos realizados por um dado médico num determinado ano e o montante faturado.**

**Uma imagem com captura de ecrã

Descrição gerada automaticamente**

Figura 21 - Resolução da interrogação do utilizador nº4

1. **Consultar agendamentos para um determinado tipo de exame numa determinada clínica.**

**Uma imagem com captura de ecrã

Descrição gerada automaticamente**

Figura 22 - Resolução da interrogação do utilizador nº5

* 1. Tradução das transações estabelecidas para SQL
  2. Escolha, Definição e Caracterização de Índices em SQL

Apesar de a opção de *Foward Engeneering* criar índices por predefinição, para as chaves primárias por exemplo, no nosso sistema de bases de dados optamos por não implementar índices, pois consideramos que este não dispõe de uma dimensão e complexidade significativa.

* 1. Estimativa do Espaço em Disco da Base de Dados e Taxa de Crescimento Anual

Ao elaborar a base de dados é necessário ter em atenção o espaço que ocupará em disco. Para efetuar uma previsão do espaço ocupado em disco, começamos por identificar o tamanho de cada tipo de dados que utilizamos. De seguida, para cada tabela do modelo lógico verificamos o espaço ocupado por esta.

|  |  |
| --- | --- |
| Tipo de Dados | Tamanho (bytes) |
| INT | 4 |
| DATE | 4 |
| DATETIME | 8 |
| VARCHAR(N) | N+1 |
| DECIMAL(I,D) | 4+4\* |

Tabela 9 - Espaço ocupado em disco por cada tipo de dados

\*4 para o pior caso pois varia consoante a parte decimal

|  |  |  |  |
| --- | --- | --- | --- |
| ATLETA | Atributos | Tipo de Dados | Espaço Ocupado |
| idAtleta | INT | 4 |
| nome | VARCHAR(100) | 101 |
| dataNascimento | DATE | 4 |
| sexo | VARCHAR(5) | 6 |
| morada | INT | 4 |
| TOTAL | 119 | | |

Tabela 10 - Espaço ocupado pela tabela Atleta

|  |  |  |  |
| --- | --- | --- | --- |
| Clínica | Atributos | Tipo de Dados | Espaço Ocupado |
| idClinica | INT | 4 |
| nome | VARCHAR(100) | 101 |
| morada | INT | 4 |
| TOTAL | 109 | | |

Tabela 11 - Espaço ocupado pela tabela Clínica

|  |  |  |  |
| --- | --- | --- | --- |
| Médico | Atributos | Tipo de Dados | Espaço Ocupado |
| idMedico | INT | 4 |
| nome | VARCHAR(100) | 101 |
| dataNascimento | DATE | 4 |
| dataInicioServico | DATE | 4 |
| especialidade | VARCHAR(100) | 101 |
| idChefe | INT | 4 |
| idClinica | INT | 4 |
| morada | INT | 4 |
| TOTAL | 226 | | |

Tabela 12 - Espaço ocupado pela tabela Médico

|  |  |  |  |
| --- | --- | --- | --- |
| TesteClínico | Atributos | Tipo de Dados | Espaço Ocupado |
| idTesteClinico | INT | 4 |
| designacao | VARCHAR(100) | 101 |
| dataHora | DATE | 4 |
| preco | DECIMAL(6,2) | 4+1 |
| resultado | VARCHAR(100) | 101 |
| idMedico | INT | 4 |
| idAtleta | INT | 4 |
| TOTAL | 223 | | |

Tabela 13 - Espaço ocupado pela tabela Teste Clínico

|  |  |  |  |
| --- | --- | --- | --- |
| Prova | Atributos | Tipo de Dados | Espaço Ocupado |
| idProva | INT | 4 |
| designacao | VARCHAR(100) | 101 |
| modalidade | VARCHAR(100) | 101 |
| categoria | VARCHAR(100) | 101 |
| TOTAL | 307 | | |

Tabela 14 - Espaço ocupado pela tabela Prova

|  |  |  |  |
| --- | --- | --- | --- |
| Morada | Atributos | Tipo de Dados | Espaço Ocupado |
| idMorada | INT | 4 |
| endereço | VARCHAR(100) | 101 |
| localidade | VARCHAR(100) | 101 |
| TOTAL | 206 | | |

Tabela 15 - Espaço ocupado pela tabela Morada

|  |  |  |  |
| --- | --- | --- | --- |
| Contacto | Atributos | Tipo de Dados | Espaço Ocupado |
| idContacto | INT | 4 |
| telefone | INT | 4 |
| email | VARCHAR(100) | 101 |
| idAtleta | INT | 4 |
| TOTAL | 113 | | |

Tabela 16 - Espaço ocupado pela tabela Contacto

|  |  |  |  |
| --- | --- | --- | --- |
| Atleta\_Prova | Atributos | Tipo de Dados | Espaço Ocupado |
| idAtleta | INT | 4 |
| idProva | INT | 4 |
| data | DATE | 4 |
| TOTAL | 12 | | |

Tabela 17 - Espaço ocupado pela tabela Atleta\_Prova

|  |  |  |  |
| --- | --- | --- | --- |
| Prova\_TipoTeste | Atributos | Tipo de Dados | Espaço Ocupado |
| idProva | INT | 4 |
| designacao | VARCHAR(100) | 101 |
| TOTAL | 105 | | |

Tabela 18 - Espaço ocupado pela tabela Prova\_TipoTeste

Considerando a dimensão da população presente na nossa base de dados, em seguida será ilustrado o espaço ocupado no disco por esta.

|  |  |
| --- | --- |
| Tabela | Espaço Ocupado em Disco(bytes) |
| Atleta | 35\*119 = 4165 |
| Clínica | 3\*109 = 327 |
| Médico | 15\*226 = 3390 |
| TesteClinico | 99\*223 = 22077 |
| Prova | 39\*307 = 11973 |
| Morada | 49\*206 = 10094 |
| Contacto | 43\*113 = 4859 |
| Atleta\_Prova | 45\*12 = 540 |
| Prova\_TipoTeste | 47\*105 = 4935 |
| Total | 62360 |

Tabela 19 - Espaço ocupado em disco pela população atual

Como podemos observar na tabela acima, a estimativa para o tamanho da nossa base de dados com o povoamento final é de 62360 bytes. No entanto, é fundamental verificar como se comportará com uma quantidade de dados significante, ou seja, com uma escala real. Atendendo ao crescimento da procura dos serviços da clínica estima-se que o número de atletas quadruplique, o que implicará contratar mais médicos. O número de testes clínicos e provas vai aumentar numa escala aproximadamente proporcional ao aumento do número de atletas. Na tabela seguinte podemos observar a estimativa para o crescimento da população da base de dados no próximo ano.

|  |  |
| --- | --- |
| Tabela | Espaço Ocupado em Disco(bytes) |
| Atleta | 140\*119 = 16660 |
| Clínica | 3\*109 = 327 |
| Médico | 35\*226 = 7910 |
| TesteClinico | 400\*223 = 89200 |
| Prova | 138\*307 = 42366 |
| Morada | 190\*206 = 39140 |
| Contacto | 215\*113 = 24295 |
| Atleta\_Prova | 135\*12 = 1620 |
| Prova\_TipoTeste | 141\*105 = 14805 |
| Total | 236323 |

Tabela 20 - Espaço ocupado em disco pela população no futuro

Assim, se a estimativa se mantiver, podemos concluir que anualmente o crescimento esperado para a base de dados é de 174 *KBytes* por ano.

* 1. Definição e Caracterização das vistas de utilização em SQL
  2. Definição e Caracterização dos Mecanismos de Segurança em SQL

Em qualquer projeto é necessário garantir que sejam cumpridos todos os requisitos de segurança necessários para assegurar a proteção dos dados. Assim é essencial que a base de dados seja ela própria o primeiro mecanismo de defesa contra qualquer tipo de ações mal-intencionadas. Assim a primeira medida foi, a criação de vários utilizadores que irão interagir de forma regular com a base de dados com permissões segundo a seguinte tabela. Existirá ainda um administrador que terá́ todos os privilégios.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Médico** | | | | **Rececionista** | | | |
| **Tabela** | **Select** | **Insert** | **Update** | **Delete** | **Select** | **Insert** | **Update** | **Delete** |
| **Atleta** | x |  |  |  | x | x | x | x |
| **Medico** | x |  |  |  | x |  |  |  |
| **TesteClinico** | x | x | x |  | x | x | x | x |
| **Clínica** |  |  |  |  | x |  |  |  |
| **Contacto** | x |  |  |  | x | x | x | x |
| **Morada** | x |  |  |  | x | x | x | x |
| **Prova** | x |  |  |  | x | x | x | x |
| **Atleta\_Prova** | x |  |  |  | x | x | x | x |
| **Prova\_TipoTeste** | x | x | x | x | x |  |  |  |

Tabela 21 - Permissões dos utilizadores

* 1. Revisão do Sistema Implementado com o Utilizador

Após a conclusão da implementação da base de dados foi agendada uma reunião final com a empresa de forma a validar a base de dados e possivelmente dar como concluído o desenvolvimento do sistema. Aqui, foram analisados todos os requisitos estabelecidos e foi revisto novamente todas as fases do trabalho já realizado, desde a elaboração do Modelo Conceptual até ao Modelo Físico. Esclarecidas todas as dúvidas que poderiam existir sobre a implementação o cliente deu autorização para avançar para a instalação do sistema.

1. Sistema *NoSQL*: *Neo4j*
   1. Justificação da utilização de um sistema *NoSQL*

Com a criação de uma base de dados relacional, a valorização da clínica no mercado aumentou drasticamente (devido à sua gestão de qualidade) e mais atletas e responsáveis pelas competições têm contactado a clínica para a realização de testes.

Deste modo, alguns problemas têm ocorrido, com cada vez mais agendamentos, a informação tem crescido intensamente e cada vez mais, é necessário usar um sistema de base de dados complementar, com maior escalabilidade, e melhor capacitado à obtenção de respostas rápidas a “queries” (interrogações).

Surge assim a necessidade de utilização de um sistema NoSQL, no nosso caso o motor Neo4j que tem em vista bases de dados com uma enorme quantidade de informação(“Performance”), tal como a Clínica Alta Performance, que permite a sua alteração face às necessidades dos médicos e dos atletas e aos requisitos de negócio(Flexibilidade e Agilidade), permitindo assim corresponder de forma positiva e eficiente às dificuldades encontradas.

* 1. Identificação dos objetivos da base de dados

O objetivo da implementação deste sistema *NoSQL* é possibilitar a realização de interrogações que envolvem a junção de várias relações de uma forma muito mais rápida. Estas interrogações correspondem às interrogações anteriormente implementadas em *SQL,* que devido a necessidade de melhoria na qualidade do serviço serão agora realizadas em *Neo4j,* tirando proveito das suas funcionalidades de travessia em grafos.

* 1. Identificação e explicação do tipo de questões (necessidades) que serão realizadas sobre o sistema de dados NoSQL

De acordo com os objetivos antes mencionados, irá surgir um subconjunto dos dados da base de dado relacional apresentada que serão utilizados para efetuar as questões (“queries”) que mais se destacam para a resolução do problema atual.

Decidiu-se demonstrar as queries que mais sobrecarregam a base de dados, ou seja, as que recorrem a *inner joins*, visto que estas são as que justificam a utilização do sistema NoSQL:

- Consultar a lista de todos os atletas de uma determinada modalidade.

- Consultar a lista de agendamentos efetuados para um determinado médico.

- Obtenção do gasto de um atleta com todos os testes realizados até ao momento.

- Consultar o número total de atletas que obteve resultado positivo em testes de dopping.

- Consultar todos os testes clínicos realizados por um atleta antes de uma determinada data.

* 1. Definição da estrutura base para o sistema de dados NoSQL que satisfaça os requisitos e as questões apresentadas anteriormente

De modo a satisfazer todos os requisitos apresentados na questão anterior surgiu a seguinte estrutura:

Uma imagem com captura de ecrã

Descrição gerada automaticamente

Figura 16 – Estrutura Neo4j

A cada nodo “Prova” estão ligados pelo relacionamento “realiza” desde 0 até N nodos “Atleta”. 0 até N nodos “Atleta” estão ligados ao nodo “TesteClinico” pelo relacionamento “faz”. A cada nodo “TesteClinico” está ligado 1 nodo “Medico” pela relação “executa” que por sua vez está ligado ao nodo “Clinica” pela relação “trabalha”.

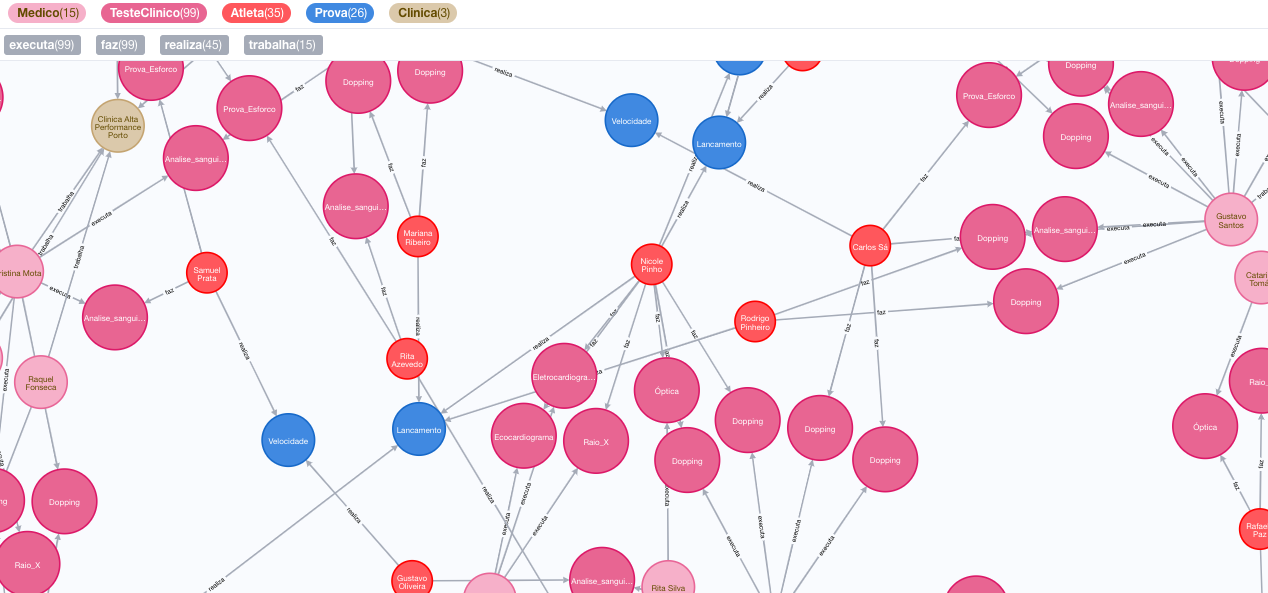


Figura 17– Ilustração da base de dados NoSQL criada em Neo4j

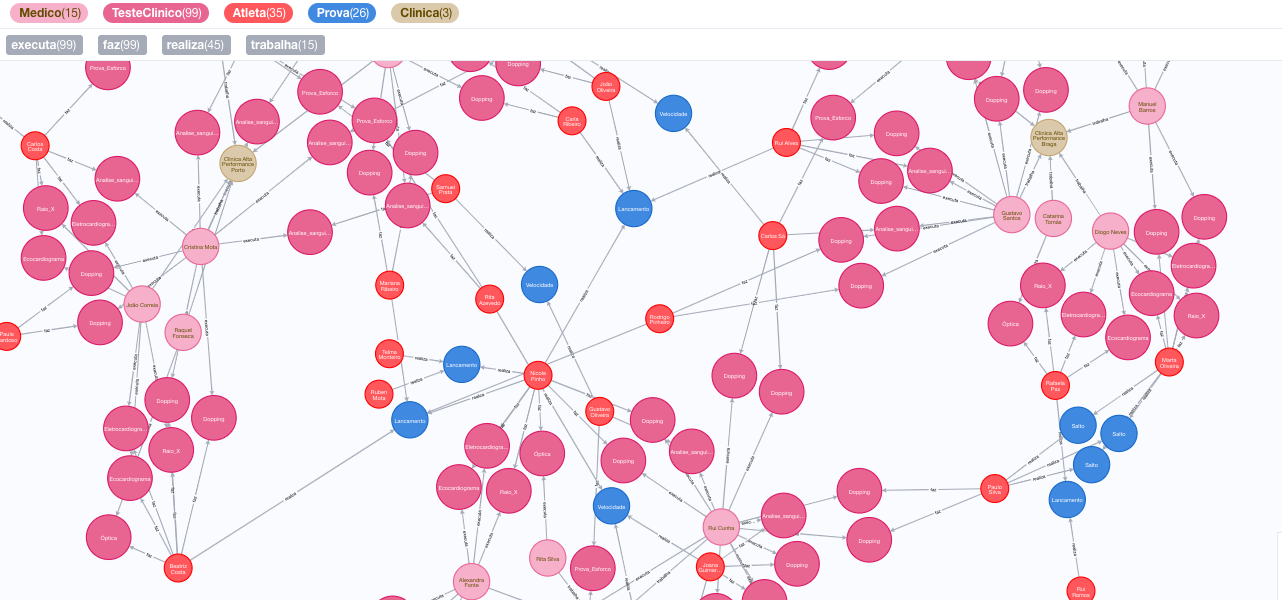


Figura 18 – Ilustração da base de dados NoSQL criada em Neo4j

Uma imagem com texto, mapa

Descrição gerada automaticamente

Figura 19 – Ilustração da base de dados NoSQL criada em Neo4j

* 1. Identificação dos objetos de dados no sistema SQL que serão utilizados para alimentar o novo sistema

Para a base de dados NoSQL os objetos utilizados serão provenientes de 6 tabelas da base de dados do sistema relacional: Clinica, Atleta, TesteClinico, Atleta\_Prova, Prova e Medico.

Obtemos cada nodo através de uma tabela no modelo relacional. O nodo Atleta obtém-se através da tabela Atleta, o nodo Clinica obtém-se através da tabela Clinica, o nodo Prova obtém-se através da tabela Prova, o nodo TesteClinico através da tabela e o nodo Medico através da tabela Medico.

Os relacionamentos “realiza” são gerados a partir da tabela Atleta\_Prova, os relacionamentos “faz” e “executa” são gerados a partir da tabela TesteClinico e o “trabalha” a partir da tabela Medico.

* 1. Mapeamento do processo de migração de dados, descrevendo o processo de conversão dos vários objetos de dados

O processo de conversão dos vários objetos de dados deu-se em duas fases: o carregamento dos nodos e o carregamento dos relacionamentos entre esses nodos.

**Carregamento dos nodos:**

-As entradas da tabela Clinica foram convertidas para nodos do tipo Clinica com os atributos idClinica e nome;

-As entradas da tabela Atleta foram convertidas para nodos do tipo Atleta com os atributos idAtleta, nome, dataNascimento e sexo;

-As entradas da tabela Prova foram convertidas para nodos do tipo Prova com os atributos idProva, designacao, modalidade e categoria;

-As entradas da tabela TesteClinico foram convertidas para nodos do tipo TesteClinico com os atributos idTesteClinico, designacao, dataHora, preco e resultado;

-As entradas da tabela Medico foram convertidas para nodos do tipo Medico com os atributos idMedico, nome, dataInicioServico, dataNascimento e especialidade;

**Carregamento dos relacionamentos:**

-A tabela Atleta\_Prova será utilizada para criar o relacionamento “realiza” com direção do Atleta para Prova e que irá guardar a data da prova;

A tabela TesteClinico será utilizada para dois tipos de relacionamento:

-O “executa” se será criado com o idMedico e o idTesteClinico (Medico executa TesteClinico);

-O relacionamento “faz” com direção Atleta para TesteClinico;

-A tabela Medico será utilizada para criar o relacionamento “trabalha” com diração do Medico para a Clinica;

* 1. Explicação do processo de migração de dados, explicando de forma detalhada as suas principais etapas - extração, transformação e carregamento

Para a conversão dos vários objetos de dados usamos as tabelas da base de dados relacional mencionadas acima (SELECT \* FROM TABELADESEJADA) exportando-as e convertendo-as em ficheiros CSV que podem ser carregados na base de dados não relacional.

Para estes ficheiros serem carregados pelo método usado no neo4j com a linguagem cypher :“LOAD CSV from ‘file:///ficheiro.csv’ AS line” teremos que coloca-los na diretoria correta para poderem ser invocados com o “file:///”. Esta diretoria está predefinida como a diretoria import do neo4j, conseguindo-se chegar lá pelo path: usr/local/Cellar/neo4j/\*3.5.12/libexec/import.

Todas as relações e nodos base que foram migradas para a nova base de dados foram alvo de projeções uma vez que nem todos os atributos constantes nestes eram necessárias para as interrogações criadas.

O processo é realizado sistematicamente e parcialmente. É sempre feita a extração parcial das entradas, a sua transformação, e, por fim,, o seu carregamento na base de dados. A extração é parcial para solucionar os problemas de memória que o respetivo programa pode ter, desta forma este funciona para bases de dados de “qualquer” dimensão.

\*A versão pode ser diferente dependendo do driver instalado

* 1. Apresentação e descrição da implementação do processo de migração de dados

Para a migração da base de dados foi criado o seguinte script em linguagem cypher que pode ser utilizado para o neo4j;

Como referido anteriormente, em primeiro lugar devemos iniciar a criação dos nodos principais:

**Criação do nodo Clinica:**

USING PERIODIC COMMIT

LOAD CSV FROM 'file:///Clinica.csv' AS line

CREATE (:Clinica { idClinica: line[0], nome: line[1]})

**Criação do nodo Atleta:**

USING PERIODIC COMMIT

LOAD CSV FROM 'file:///Atleta.csv' AS line

CREATE (:Atleta { idAtleta: line[0], nome: line[1], dataNascimento: line[2], sexo: line[3]})

**Criação do nodo Prova:**

USING PERIODIC COMMIT

LOAD CSV FROM 'file:///Prova.csv' AS line

CREATE (:Prova { idProva: line[0], designacao: line[1], modalidade: line[2],categoria:line[3]})

**Criação do nodo TesteClinico:**

USING PERIODIC COMMIT

LOAD CSV FROM 'file:///Teste\_Clinico.csv' AS line

CREATE (:TesteClinico { idTesteClinico: line[0], designacao: line[1], dataHora: line[2], preco: line[3], resultado: line[4]})

**Criação do nodo Medico:**

USING PERIODIC COMMIT

LOAD CSV FROM 'file:///Medico.csv' AS line

CREATE (:Medico { idMedico: line[0], nome: line[1], dataInicioServico: line[2],dataNascimento: line[3], especialidade: line[4]})

Foram criados índices para assegurar uma procura mais quando for feita a criação de relacionamentos num passo seguinte:

CREATE INDEX ON : Clinica(idClinica);

CREATE INDEX ON : Atleta(idAtleta);

CREATE INDEX ON : Prova(idProva);

CREATE INDEX ON : TesteClinico(idTesteClinico);

CREATE INDEX ON : Medico(idMedico);

Após a criação dos índices e dos nodos, é feita a criação dos relacionamentos:

**Criação do relacionamento entre o nodo Atleta e o nodo Prova:**

USING PERIODIC COMMIT

LOAD CSV FROM 'file:///Atleta\_Prova.csv' AS line

MATCH (a:Atleta {idAtleta: line[0]})

MATCH (p:Prova {idProva: line[1]})

MERGE (a)-[r:realiza {data: line[2]}]->(p);

**Criação do relacionamento entre o nodo Atleta e o nodo TesteClinico:**

USING PERIODIC COMMIT

LOAD CSV FROM 'file:///Teste\_Clinico.csv' AS line

MATCH (a:Atleta {idAtleta: line[6] })

MATCH (t:TesteClinico {idTesteClinico: line[0] })

MERGE (a)-[r:faz]->(t);

**Criação do relacionamento entre o nodo Medico e o nodo Clinica:**

USING PERIODIC COMMIT

LOAD CSV FROM 'file:///Medico.csv' AS line

MATCH (m:Medico {idMedico: line[0]})

MATCH (c:Clinica {idClinica: line[6] })

MERGE (m)-[r:trabalha]->(c);

**Criação do relacionamento entre o nodo Medico e o nodoTesteClinico:**

USING PERIODIC COMMIT

LOAD CSV FROM 'file:///Teste\_Clinico.csv' AS line

MATCH (m:Medico {idMedico: line[5]})

MATCH (t:TesteClinico {idTesteClinico: line[0] })

MERGE (m)-[r:executa]->(t);

É boa prática o uso de “USING PERIODIC COMMIT” quando criamos nodos ou relacionamentos dado a estarmos a trabalhar com bases de dados de grandes dimensões que devem ser atualizadas (commit) frequentemente ao carregarem os dados.

* 1. Apresentação da forma como as questões identificadas anteriormente podem ser satisfeitas com o novo sistema, utilizando a linguagem de interrogação do sistema NoSQL

A resolução para qualquer uma das questões levantadas anteriormente é obtida de frma eficiente fácil ao recorrermos ao Neo4j.

Decidiu-se demonstrar as queries que mais sobrecarregam a base de dados, ou seja, as que recorrem a *inner joins*, visto que estas são as que justificam a utilização do sistema NoSQL.

As queries que fazem apenas a leitura de atributos não sobrecarregam a base de dados MySQL pelo que a performance destas já é eficiente, mesmo perante uma enorme quantidade de dados.

8. Conclusões e Trabalho Futuro

<<Elaborar uma apreciação crítica sobre o trabalho realizado, apontando os seus pontos fortes e fracos. Adicionalmente, caso se aplique, enunciar eventuais tarefas a realizar futuramente ou novas opções para estender o trabalho realizado.>>

Referências

Connolly, T. and Begg, C. (2005). Database Systems, A Practical Approach to Design, Implementation, and Management. 4th ed. Addison-Wesley.

Lista de Siglas e Acrónimos

**BD** Base de Dados

**SGBD** Sistema de Gestão de Base de Dados

**SQL** *Structured Query Language*

**ER** *Entity–Relationship*

**ACID** Atomicidade, Consistência, Isolamento e Durabilidade

**NoSQL** Not only SQL

# Anexos

1. **Script Completo de Criação**
2. **Script Parcial do Povoamento**
3. **Queries em SQL**
4. **Exportação de Dados**
5. Anexo 1 – Script Completo de Criação
6. Anexo 2 – Script Parcial de Povoamento
7. Anexo 3 – Queries em SQL
8. Anexo 4 – Exportação de dados