



Universidade do Minho
Departamento de Informática

Janeiro 2025

Laboratórios de Informática III

Relatório Fase II

Grupo 77

Hélder Tiago Peixoto da Cruz (A104174)

André Miguel Rego Trindade Pinto (A104267)

Bruno Miguel Vieira Ribeiro (A94662)

Conteúdo

Introdução e Objetivos.....	3
Batch Mode e Queries	4
Sistema.....	5
Estrutura Modular	5
Fluxo Geral de Operação	6
Esboço da Estrutura do Sistema.....	7
Programa Interativo	8
Conclusão	10

Introdução e Objetivos

O desenvolvimento da segunda fase do projeto focou-se na expansão e refinação de funcionalidades existentes, bem como na integração de novos componentes essenciais ao sistema de streaming de música. Esta etapa introduziu desafios significativos, como a adaptação ao aumento de dados e à inclusão de novas entidades, como álbuns e histórico de utilização. Além disso, a necessidade de incorporar um modo de execução interativo e novas queries exigiu um esforço considerável para manter a modularidade e encapsulamento do sistema.

Dificuldades iniciais emergiram devido ao aumento da complexidade no manuseio e validação de dados, sobretudo com o volume expandido de entradas CSV e a implementação de novas estruturas de dados. Para contornar esses desafios, optou-se por uma abordagem iterativa, reforçando a integração entre módulos já existentes e desenvolvendo soluções específicas para problemas identificados. Por exemplo, foi necessária a criação de uma estratégia robusta de gestão de memória, especialmente para evitar fugas de memória reportadas em testes com ferramentas como o Valgrind.

Adicionalmente, a implementação de novas queries trouxe à tona a importância de uma análise detalhada de desempenho e consumo de recursos, obrigando-nos a justificar a escolha de estruturas como tabelas hash para melhorar a eficiência. Assim, esta fase não apenas consolidou o progresso alcançado, mas também proporcionou uma oportunidade para ampliar as capacidades técnicas do grupo, garantindo a conformidade com os rigorosos critérios de avaliação definidos.

Batch Mode e Queries

Este modo utiliza o módulo principal para carregar os dados, interpretar os comandos e executar as queries, armazenando os resultados em ficheiros na pasta resultados. A integração com o sistema de armazenamento permitiu uma gestão eficiente das entidades carregadas, garantindo que os dados estavam prontos para serem processados antes da execução das queries.

Desenvolvimento e Estrutura

O processo no *Batch Mode* começa com o carregamento dos dados das entidades a partir de ficheiros CSV, validando e populando as tabelas hash. Posteriormente, o sistema lê o ficheiro de comandos, interpreta os parâmetros de cada query e invoca a função correspondente no módulo dedicado.

Cada query foi estruturada de forma modular, facilitando a sua integração e manutenção:

- A **Query 1** e a **Query 2** foram concluídas com sucesso, permitindo o acesso a resumos e rankings conforme os critérios definidos.
- Para a **Query 3**, conseguimos identificar géneros mais populares com base nos parâmetros fornecidos, mesmo com desafios relacionados à agregação de dados.
- As **Queries 4, 5 e 6** foram parcialmente implementadas, com progresso significativo na estruturação dos cálculos necessários, mas limitadas pela complexidade do processamento de dados históricos e pela dificuldade de implementar recomendações avançadas.

Desafios e Soluções

O maior desafio no Batch Mode foi garantir que todas as queries funcionassem corretamente sob condições de carga elevada. A utilização de tabelas hash revelou-se eficaz para minimizar o tempo de busca e permitir um acesso rápido às entidades. No entanto, o tratamento de cenários como validações ausentes ou dados incompletos exigiu um esforço adicional, resultando em melhorias contínuas durante o desenvolvimento.

Apesar de esforços intensos, não conseguimos atingir a implementação completa de todas as queries devido à sua complexidade intrínseca e às limitações temporais. Ainda assim, os módulos e algoritmos foram concebidos de forma escalável, permitindo continuidade no desenvolvimento futuro. Embora o Batch Mode esteja funcional para a maioria das queries, reconhecemos que há espaço para otimizações e refinamentos.

Sistema

O sistema desenvolvido para o projeto é estruturado com base nos princípios de modularidade e encapsulamento, assegurando a separação clara entre as diferentes funcionalidades e promovendo a manutenibilidade e escalabilidade do código. A organização segue uma arquitetura modular composta por múltiplos módulos principais, cada um dedicado a um aspeto específico do sistema de streaming de música.

Estrutura Modular

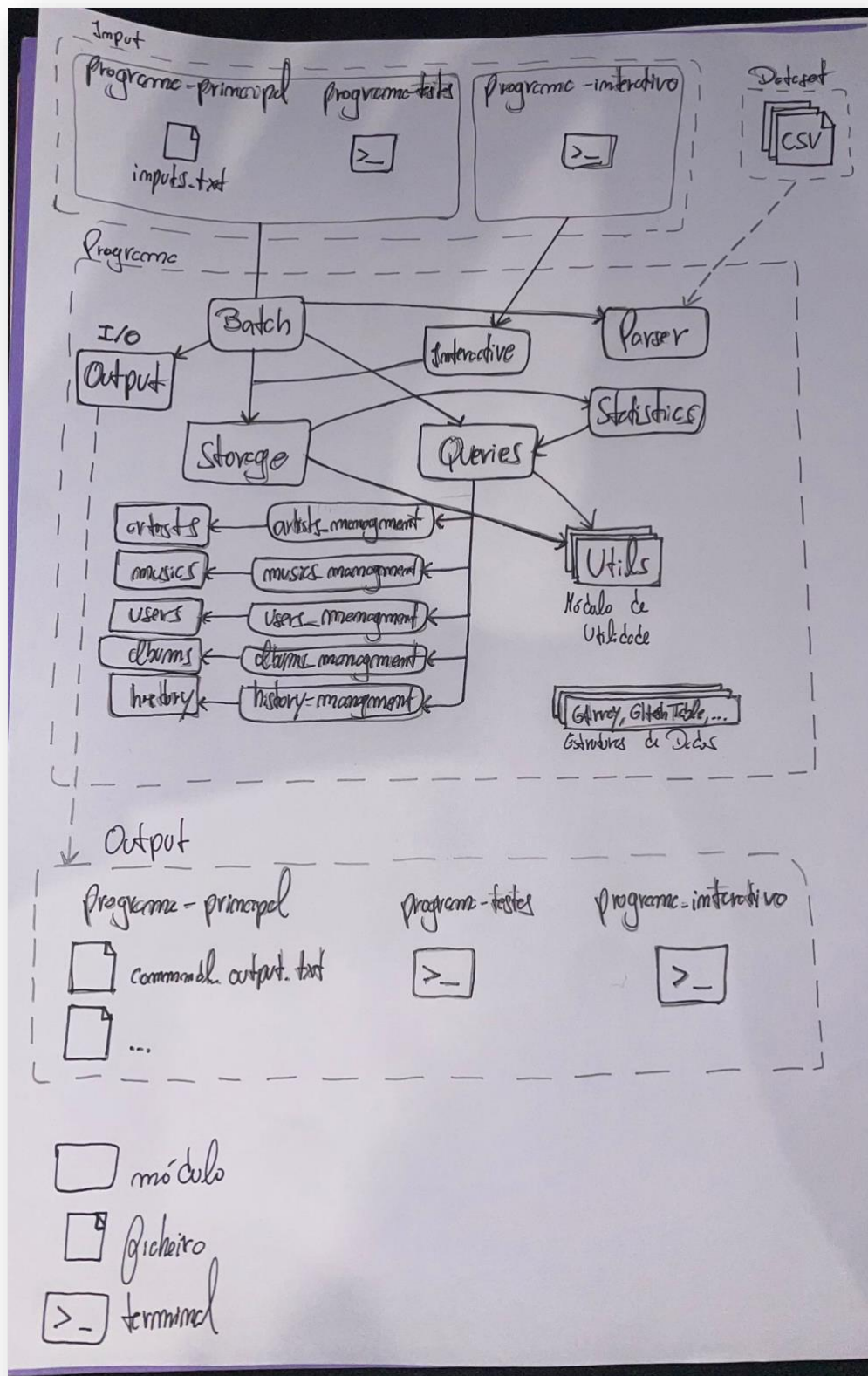
1. **Gestores de Entidades:** Cada entidade principal do sistema, como **utilizadores**, **artistas**, **músicas**, **álbuns** e **histórico**, possui um módulo dedicado para gestão. Estes módulos são responsáveis por operações como criação, inserção e validação de dados, utilizando tabelas hash da biblioteca GLib para armazenamento eficiente. Exemplos incluem:
 - `users_management.c` para gerir utilizadores;
 - `artists_management.c` para gerir artistas;
 - `musics_management.c` para gerir músicas.
2. **Parser:** O módulo `parser.c` é responsável pela leitura e validação dos ficheiros CSV de entrada, garantindo que os dados são carregados corretamente para as estruturas internas. Além disso, regista erros encontrados em ficheiros específicos na pasta resultados, seguindo as normas do projeto.
3. **Execução de Queries:** As queries estão implementadas como funções dedicadas nos seus respetivos ficheiros, como *querie1.c*, *querie2.c*, entre outros. Estas funções extraem e processam dados das estruturas do sistema para produzir os resultados solicitados.
4. **Armazenamento Geral:** O módulo *storage.c* centraliza a gestão de todas as tabelas hash, encapsulando as operações de inicialização e destruição de memória, bem como a disponibilização de acesso às entidades para outros módulos.

5. **Modo Interativo:** Para suportar o modo interativo, foi desenvolvido um conjunto de utilitários (*interactive.c* e *interactive_utils.c*) que fornecem menus dinâmicos, validação de entradas e processamento de comandos diretamente no terminal.
6. **Utilitários Gerais:** O módulo *utils.c* contém funções auxiliares amplamente utilizadas no sistema, como manipulação de datas, cálculos de métricas e registo de resultados.

Fluxo Geral de Operação

- **Inicialização:** O programa começa por carregar os dados dos ficheiros CSV, validando-os e armazenando-os em tabelas hash encapsuladas.
- **Execução:** Consoante o modo (principal, interativo ou testes), processa comandos, queries ou entradas do utilizador, produzindo os resultados esperados.
- **Validação e Saída:** Dados inválidos são registados em ficheiros específicos, enquanto os resultados das queries são armazenados na pasta resultados, garantindo conformidade com o formato esperado.

Esboço da Estrutura do Sistema



Programa Interativo

Execução exemplo em modo interativo da query 2 com os argumentos “10” e “Portugal” com o dataset regular de exemplo proveniente da equipa docente.

```
C:\Program Files\WSL\wsl.exe X + v
=====
|| BEM-VINDO AO PROGRAMA INTERATIVO ||
=====
Introduza o caminho do dataset (pressione Enter para usar o default): |
```

```
C:\Program Files\WSL\wsl.exe X + v
=====
|| A CARREGAR O DATASET ||
=====
Aguarde por favor...
Faltam 1 segundo(s)... |
```

```
C:\Program Files\WSL\wsl.exe X + v
=====
|| MENU INTERATIVO ||
=====
1. Query 1 - Resumo de Utilizador/Artista
2. Query 2 - Top N Artistas por Discografia
3. Query 3 - Géneros mais Populares por Faixa Etária
4. Query 4 - Artista no Top 10 mais vezes
5. Query 5 - Recomendação de Utilizadores
6. Query 6 - Resumo Anual de Utilizador
7. Ajuda
0. Sair
-----
Escolha uma opção: |
```

```
C:\Program Files\WSL\wsl.exe X + v
=====
|| A PROCESSAR A QUERY 2 ||
=====
Introduza o número de artistas a listar (N): 10
Introduza o filtro de país (pressione Enter para ignorar): Portugal

Resultados da Query 2:
SeaShell Should;individual;02:47:18;Portugal
Source Home;individual;01:52:42;Portugal
Deborah;individual;01:29:35;Portugal
Coach Population;group;01:13:43;Portugal
Lil Stephen;individual;01:07:07;Portugal
Megan;individual;00:59:29;Portugal
Back Less;group;00:58:28;Portugal
Lil Sandy;individual;00:57:59;Portugal
The though;group;00:51:59;Portugal
James George;individual;00:51:11;Portugal

Pressione ENTER para voltar ao menu...
|
```


O modo interativo do sistema foi desenvolvido para oferecer uma experiência acessível e funcional, permitindo a execução de queries diretamente no terminal de forma intuitiva. Este modo é iniciado com a apresentação de um menu interativo que exhibe as opções disponíveis, incluindo as seis queries implementadas, *ajuda* e a possibilidade de saída do programa.

Ao iniciar o programa, o utilizador é solicitado a fornecer o caminho para o dataset de entrada. Caso o caminho não seja especificado, é utilizado um caminho predefinido. Uma vez carregados os dados, o sistema realiza uma validação para garantir que as tabelas de armazenamento estão corretamente populadas. Este processo é acompanhado por mensagens de feedback e uma contagem regressiva que simula o carregamento dos dados, melhorando a interação visual com o utilizador.

Cada query pode ser executada ao selecionar a opção correspondente no menu. Dependendo da query, o programa solicita ao utilizador a introdução de parâmetros adicionais, como IDs, intervalos de idade ou filtros opcionais. Por exemplo, a Query 1 necessita que o utilizador forneça o id de um utilizador ou artista, enquanto a Query 2 solicita o valor N (número de artistas a listar) e o filtro de país (opcional).

As entradas do utilizador são validadas rigorosamente para evitar comportamentos inesperados, como crashes. Em caso de erro, mensagens informativas são exibidas e o sistema retorna ao menu principal. Os resultados das queries são apresentados no terminal ou em ficheiros temporários que são processados e removidos após a exibição.

A arquitetura modular do modo interativo garante que cada funcionalidade está isolada e é facilmente extensível. Funções auxiliares foram desenvolvidas para tarefas específicas, como validação de IDs, captura de números inteiros e gestão de filtros opcionais. Este design assegura que o programa interativo não apenas cumpre os requisitos do projeto, mas também proporciona uma interface eficiente e robusta para a exploração dos dados do nosso projeto.

Conclusão

Ao longo deste projeto, enfrentámos vários desafios, desde a adaptação inicial à escala e complexidade do sistema na Fase 1 até à integração de novas funcionalidades e à expansão significativa de dados na Fase 2. Estes desafios testaram não apenas as nossas competências técnicas, mas também a nossa capacidade de nos organizarmos enquanto equipa e de encontrarmos soluções criativas para problemas complexos.

Superar as dificuldades da Fase 1, como garantir a validação eficaz dos dados e a implementação modular do sistema, deu-nos a confiança necessária para enfrentar os obstáculos ainda maiores da Fase 2. Embora não tenhamos conseguido implementar todas as funcionalidades a 100%, fizemos progressos notáveis e conseguimos construir um sistema robusto e funcional, capaz de responder à maioria das queries e de proporcionar uma experiência interativa envolvente.

Olhando para o trabalho desenvolvido, sentimos orgulho no que conseguimos alcançar. Apesar das limitações impostas pelo tempo e pela complexidade, conseguimos criar um projeto que reflete o nosso esforço, dedicação e capacidade de superação. Este processo foi uma experiência de aprendizagem valiosa, que certamente nos preparará melhor para enfrentar futuros desafios com ainda mais determinação.