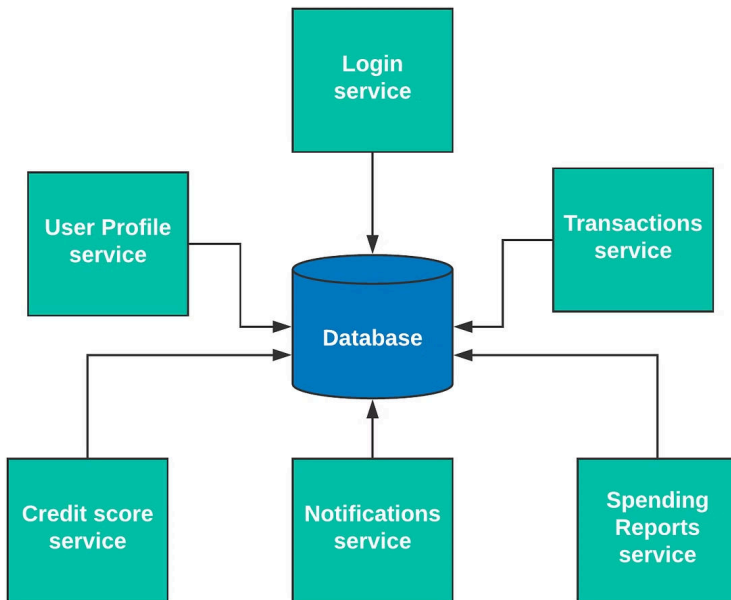


Banco de Dados Integrado das Olimpíadas

O projeto de banco de dados integrado das Olimpíadas tem como objetivo centralizar e organizar as informações sobre a participação de atletas nas Olimpíadas, com foco inicial nos dados sobre atletas australianos. Os dados foram retirados do site Kaggle e incluem informações detalhadas sobre atletas, modalidades, medalhas e eventos olímpicos.



Alimentação do Banco (Fase 1)



1

Separação dos Dados

Na fase inicial, o dataset completo em formato CSV foi filtrado para conter apenas informações sobre atletas da Austrália. Esse processo envolveu a utilização de um script em Python (`separate-tables.py`) para separar os dados condensados em um único CSV em cinco arquivos CSV distintos, cada um representando uma entidade específica.

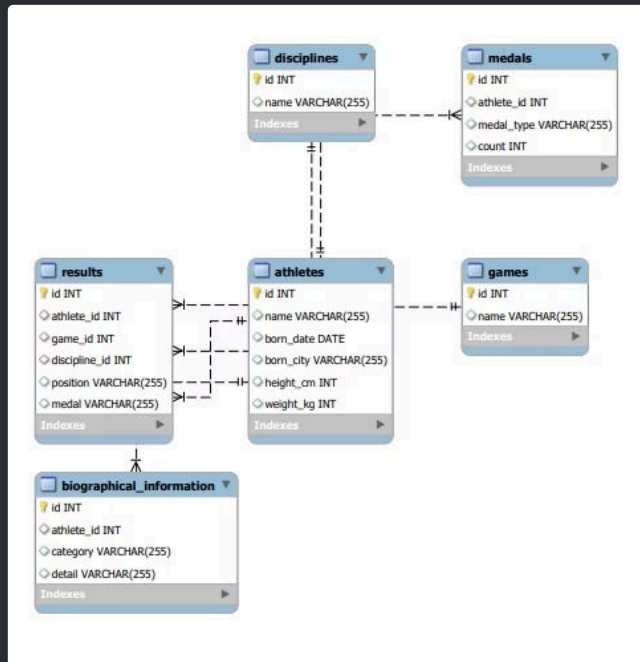
2

Geração de Querys

Após a separação, outro script (`create-queries.py`) foi utilizado para gerar as queries de inserção no banco de dados. Os arquivos resultantes e suas respectivas funções são: `/queries:` Contém os arquivos com as queries de inserção para cada tabela. `comandos.sql`: Arquivo consolidado com todos os comandos SQL para inserção de dados.

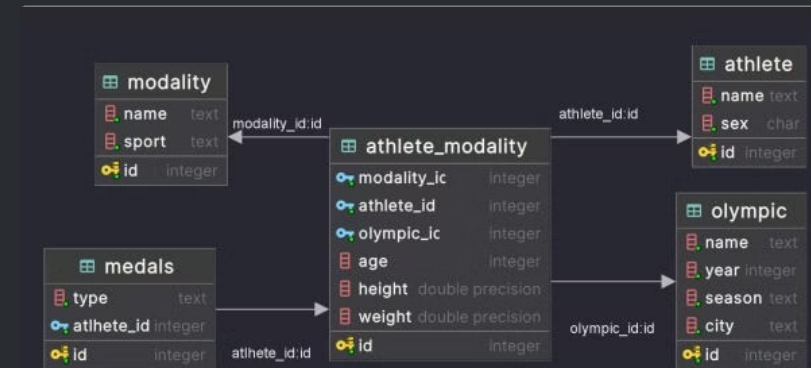
Esquemas Relacionais

Banco Japão



Colunas: id (inteiro, chave primária), name (texto), sex (booleano), nationality (texto), age (inteiro), height (inteiro), weight (inteiro)

Banco Austrália



Colunas: id (inteiro, chave primária), name (texto), sport (texto)

olympic

Colunas: id (inteiro, chave primária), year (inteiro), season (texto), city (texto), type (tex

Esquema Relacional Integrado

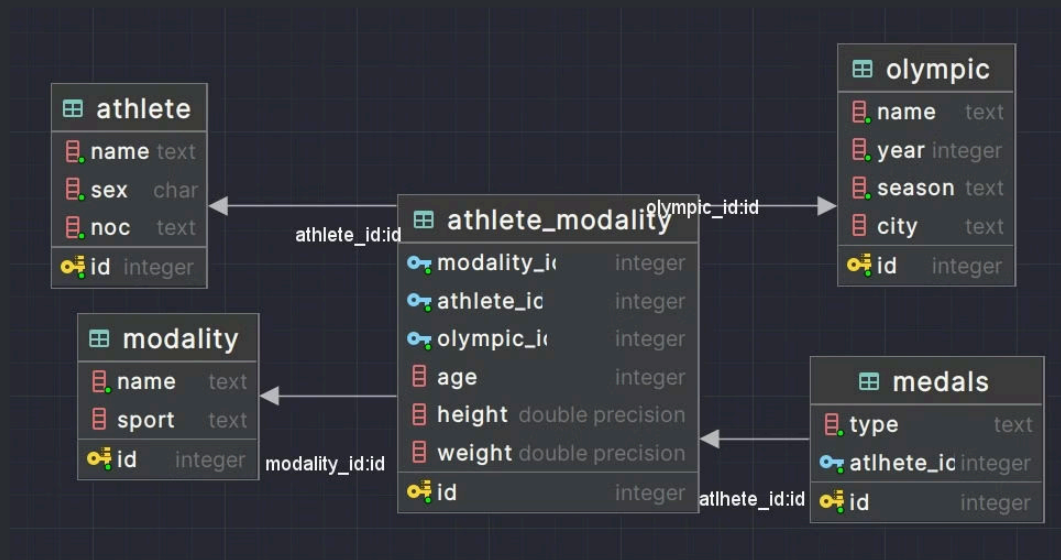


Tabela de Junção "Atleta Modalidade" (Athlete Modality)

Tabela: athlete_modality

Colunas: athlete_id (inteiro, chave estrangeira referenciando athlete(id)), modality_id (inteiro, chave estrangeira referenciando sport(id))

Chave Primária Composta: A combinação de athlete_id e modality_id garante que cada atleta esteja associado a cada modalidade uma única vez.

Tabela de Junção "Jogos Olímpicos" (Olympic Games)

tabela: olympic_games

Colunas: olympic_id (inteiro, chave estrangeira referenciando olympic(id)), athlete_id (inteiro, chave estrangeira referenciando athlete(id)), modality_id (inteiro, chave estrangeira referenciando sport(id))

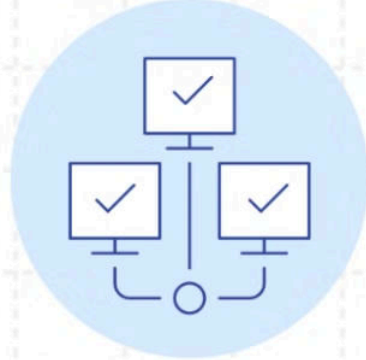
Chave Primária Composta: A combinação de olympic_id, athlete_id, e modality_id garante a unicidade dos registros.



Storage



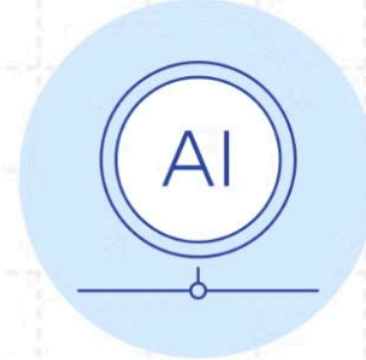
Cloud-based



ETL vs ELT



Real-time data



AI-powered

Integração dos Bancos

1

Importação Manual de Dados Adicionais

Dados adicionais de um segundo banco de dados do Japão foram importados manualmente, verificando e evitando duplicidades.

2

Verificação de Duplicidades

Para cada novo registro, verificamos se já existia um registro correspondente no banco de dados principal. Caso contrário, criávamos um novo registro.

3

Ajustes Manuais

Realizamos ajustes manuais nos registros, como a substituição de textos e a criação de novos IDs, para assegurar a integridade dos dados.

Consultas Propostas

Medalhas por Atleta

```
WITH MedalCounts AS (  
  
    SELECT  
  
        id_atlhete_modality,  
  
        SUM(CASE WHEN Medal = 'Bronze' THEN 1 ELSE 0 END) AS Bronze_Medals,  
  
        SUM(CASE WHEN Medal = 'Silver' THEN 1 ELSE 0 END) AS Silver_Medals,  
  
        SUM(CASE WHEN Medal = 'Gold' THEN 1 ELSE 0 END) AS Gold_Medals,  
  
        SUM(CASE WHEN Medal IN ('Bronze', 'Silver', 'Gold') THEN 1 ELSE 0 END) AS Total_Medals  
  
    FROM medals  
  
    GROUP BY id_atlhete_modality  
  
)  
  
SELECT  
  
    id_atlhete_modality,  
  
    Bronze_Medals,  
  
    Silver_Medals,  
  
    Gold_Medals,  
  
    Total_Medals  
  
FROM MedalCounts  
  
ORDER BY Total_Medals DESC  
  
LIMIT 10;
```

#	id_atlhete_modality	#	Bronze_Medals	#	Silver_Medals	#	Gold_Medals	#	Total_Medals
	387		1		0		0		1
	495		1		0		0		1
	725		1		0		0		1
	786		0		0		1		1
	828		0		1		0		1
	1194		1		0		0		1
	1216		0		0		1		1
	1246		0		0		1		1
	8		1		0		0		1
	143		0		0		1		1

Eventos Olímpicos por Ano

SELECT


Season,

COUNT(*) AS Number_of_Games

FROM olympic

GROUP BY Season

ORDER BY Season

 Season	# Number_of_Games
Summer	28
Winter	19

Modalidades Populares

SELECT sport, COUNT(*) AS Total

FROM modality

GROUP BY sport

ORDER BY Total DESC;

T	sport	#	Total
	Athletics		56
	Swimming		40
	Cycling		29
	Shooting		26
	Canoeing		24
	Sailing		19
	Wrestling		19
	Rowing		18
	Judo		15
	Weightlifting		15
	Gymnastics		14
	Boxing		14

Conclusão

Banco de Dados Integrado

O projeto de banco de dados integrado das Olimpíadas permitiu centralizar e organizar as informações sobre a participação de atletas nas Olimpíadas, com foco inicial nos dados sobre atletas australianos. O processo envolveu a separação dos dados, a geração de queries, a criação de esquemas relacionais e a integração dos bancos de dados.

Consultas Relevantes

As consultas SQL elaboradas permitem obter estatísticas e informações relevantes, como medalhas por atleta, modalidades populares e eventos olímpicos por ano. Essas informações são valiosas para análises e tomada de decisões relacionadas ao desempenho olímpico.

