

INSTITUTO FEDERAL

FELIPE PEREIRA DE ARAUJO

Cálculo Relacional: Fundamentos e Aplicações na Ciência de Dados

CAMPOS DO JORDÃO

2024

INSTITUTO FEDERAL

FELIPE PEREIRA DE ARAUJO

Cálculo Relacional: Fundamentos e Aplicações na Ciência de Dados

Trabalho de conclusão de curso apresentado à INSTITUTO FEDERAL
como parte dos requisitos para a obtenção do grau de BANCO DE DADOS 1

Orientador: Paulo Giovani

]

Campos do Jordão

2024

INTRODUÇÃO

O **Cálculo Relacional** é uma linguagem declarativa usada em consultas de bancos de dados, onde é especificado o que deve ser recuperado, mas não a maneira pela qual o resultado será obtido. Baseando-se na **lógica de primeira ordem** e no **cálculo de predicados**, essa abordagem permite expressar consultas sem demonstrar os passos exatos para a sua execução. Utilizando de algumas fórmulas lógicas compostas por predicados (proposições que podem ou não, ser verdadeiras), operadores lógicos como “e”, “ou” e “negação”, além de variáveis e quantificadores. Essa é uma área fundamental na área de bancos de dados, usada para especificar consultas em bancos de dados relacionais. Ele se baseia em lógica e é a base teórica para linguagens como o SQL.

Lógica de primeira ordem

Lógica de primeira ordem (também chamada de **lógica de predicados**) é um sistema formal usado em matemática, filosofia e ciência da computação para representar e analisar a estrutura lógica de afirmações. Ela se expande em relação à lógica proposicional, permitindo o uso de **quantificadores** e **predicados**.

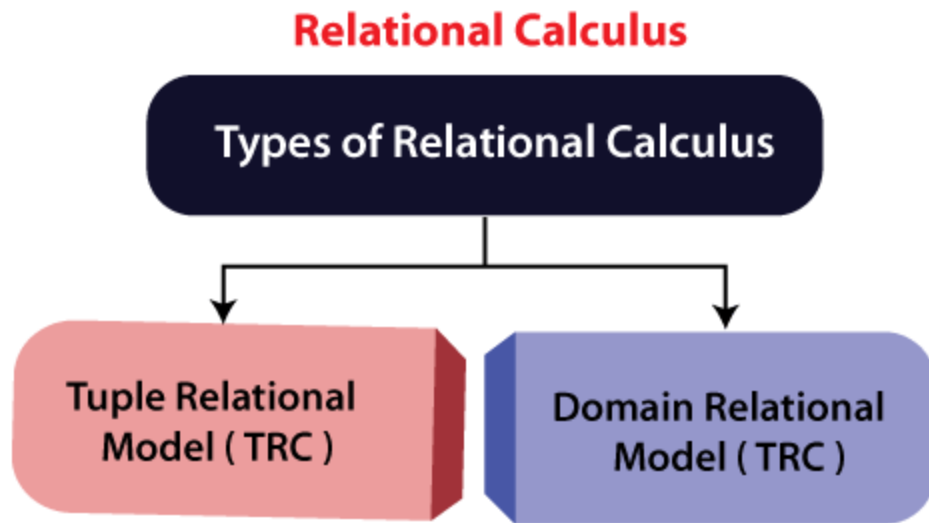
Na lógica de primeira ordem, expressões lógicas são construídas com:

- **Variáveis**: símbolos que podem representar elementos de um domínio.
- **Predicados**: expressões que afirmam algo sobre as variáveis (exemplo: $P(x)$ pode significar "x é um número primo").

Essa lógica é mais expressiva, pois permite falar sobre propriedades de objetos e relações entre eles, não apenas sobre a verdade ou falsidade de proposições.

Cálculo de predicados

Cálculo de predicados (ou lógica de **predicados**) é o sistema formal que usa a **lógica de primeira ordem** para descrever e manipular afirmações com variáveis, predicados e quantificadores. Ele permite fazer **inferências e provas formais com base nessas expressões lógicas**.



Existem duas principais variantes do cálculo relacional:

1. **Cálculo com variáveis de domínio**, em que as variáveis substituem os atributos das tuplas.
2. **Cálculo com variáveis n-tuplas**, onde as variáveis representam as tuplas inteiras.

A principal diferença entre cálculo relacional de tupla e cálculo de domínio

cálculo relacional está em tipos de variáveis em consultas. No cálculo **relacional de tuplas**, as variáveis representam tuplas (geralmente de alguma relação, mas também podem representar todas as tuplas no banco de dados), enquanto no **cálculo relacional de domínio** as variáveis representam **valores únicos** de domínios de **atributos de relações**.

Comparação entre Álgebra Relacional e Cálculo Relacional

Embora sejam abordagens relativamente distintas, a álgebra relacional e o cálculo relacional possuem a mesma ordem de expressão. Em 1972, **Edgar Frank Codd** demonstrou que qualquer consulta feita em uma dessas linguagens pode ser escrita como uma fórmula, estabelecendo assim a equivalência entre ambas. Porém, a álgebra relacional é mais sobre procedimento, uma vez que especifica a ordem das operações a serem realizadas, enquanto o cálculo relacional é mais focado na declaração, permitindo que o sistema determine a forma mais eficiente de avaliação.

Uma analogia interessante é compará-lo a um **planejamento de viagem**: o cálculo relacional permite especificar que se deseja partir de uma cidade e chegar a outra, sem fornecer os detalhes de como isso será feito.

Manipulação de Dados no Modelo Relacional

O **modelo relacional**, que sustenta a maioria dos sistemas de banco de dados, define dois mecanismos principais para manipulação de dados:

1. **Álgebra relacional** que se baseia em operações algébricas.
2. **Cálculo relacional**, que usa a lógica matemática.

Ambos os mecanismos operam sobre as **tabelas** do banco de dados, e os resultados de cada operação são **relações**. A principal diferença entre eles está no nível de procedimentalidade. Enquanto a álgebra relacional é mais detalhada na execução de operações, o cálculo relacional apenas define as condições que as tuplas resultantes devem satisfazer, sendo, assim, uma linguagem mais não-procedimental e declarativa.

Como dissera o matemático **John Wilder Tukey**: "*Uma resposta aproximada para a pergunta certa vale muito mais do que uma resposta exata para a pergunta errada*". No contexto de bancos de dados, o cálculo relacional ajuda a formular as perguntas certas sem a necessidade de detalhar como o sistema deve calcular as respostas, deixando a otimização para o sistema de gerenciamento de banco de dados (DBMS).

<i>Símbolo</i>	<i>Operação</i>	<i>Sintaxe</i>	<i>Tipo</i>
σ	Seleção / Restrição	$\sigma_{\text{condição}} (\text{Relação})$	Primitiva
π	Projeção	$\pi_{\text{expressões}} (\text{Relação})$	Primitiva
\cup	União	$\text{Relação1} \cup \text{Relação2}$	Primitiva
\cap	Intersecção	$\text{Relação1} \cap \text{Relação2}$	Adicional
-	Diferença de conjuntos	$\text{Relação1} - \text{Relação2}$	Primitiva
\times	Produto cartesiano	$\text{Relação1} \times \text{Relação2}$	Primitiva
$ x $	Junção	$\text{Relação1} x \text{Relação2}$	Adicional
\div	Divisão	$\text{Relação1} \div \text{Relação2}$	Adicional
ρ	Renomeação	$\rho_{\text{nome}} (\text{Relação})$	Primitiva
\leftarrow	Atribuição	$\text{variável} \leftarrow \text{Relação}$	Adicional

Um dos seus focos é o desenvolvimento de **teorias formais de bancos de dados** e a aplicação de métodos lógicos em grandes sistemas de informação.

Otimização de consultas: Melhora a eficiência em sistemas de bancos de dados.

Lógica temporal e espacial: Integra dimensões temporais no cálculo relacional para assim modelar eventos que mudam ao longo do tempo.

Autômatos e álgebra relacional: Estudos que exploram como métodos formais de autômatos podem ser aplicados a sistemas relacionais.

Livros e artigos acadêmicos em russo muitas vezes abordam esses tópicos com um foco prático, dado o papel fundamental que bancos de dados e sistemas de informação desempenham no desenvolvimento de infraestrutura tecnológica e industrial no país.

Na Índia por exemplo, a pesquisa sobre cálculo relacional tem uma abordagem bastante prática, com foco na aplicação em grandes bancos de dados para sistemas financeiros, governamentais e empresariais. Algumas áreas de destaque incluem:

Consultas complexas e big data: O uso do cálculo relacional para otimizar a execução de consultas em grandes volumes de dados.

Interoperabilidade de sistemas de bancos de dados: Com foco na integração de dados de diferentes fontes (heterogêneas), um campo que usa intensivamente o cálculo relacional para garantir que diferentes sistemas possam se comunicar de maneira eficaz.

Linguagens de consulta baseadas em lógica: O desenvolvimento de novas linguagens ou extensões de SQL que incorporem diretamente o cálculo relacional, a fim de melhorar a expressividade e a eficiência das consultas.

Institutos indianos como o *Indian Institute of Technology (IIT)* e o *National Informatics Centre (NIC)* têm projetos que aplicam a teoria do cálculo relacional a problemas de infraestrutura nacional, como o gerenciamento de dados em larga escala para serviços públicos.

Comparações com Outras Abordagens

No contexto internacional, o cálculo relacional oferece uma forma declarativa de descrever consultas, em contraste com a álgebra relacional que é procedural. Muitos estudiosos de países de língua russa e hindi veem o cálculo relacional como um caminho promissor para melhorar a eficiência e a escalabilidade dos sistemas de bancos de dados modernos.

Manipulação de Dados no Modelo Relacional

O **modelo relacional**, que forma a base da maioria dos sistemas modernos de banco de dados, utiliza dois mecanismos principais para a manipulação de dados:

1. **Álgebra relacional:** Um conjunto de operações algébricas que manipula e retorna relações (tabelas).
2. **Cálculo relacional:** Um sistema baseado em lógica que descreve as condições que os dados retornados devem satisfazer.

Ambos os mecanismos operam sobre relações, e os resultados de suas operações também são relações. A álgebra relacional é mais detalhada, exigindo que o programador especifique o procedimento exato, enquanto o cálculo relacional define apenas as condições que os dados devem satisfazer, permitindo que o DBMS faça a otimização automaticamente.

Otimização de Consultas

Uma das vantagens do cálculo relacional é sua aplicabilidade em técnicas de **otimização de consultas**. Como o cálculo relacional é declarativo, ele oferece mais flexibilidade para o DBMS reorganizar e otimizar a execução das consultas. Isso é particularmente útil em grandes bancos de dados, onde a eficiência da execução pode impactar significativamente o desempenho do sistema.

O **otimizador de consultas** de um DBMS pode usar heurísticas e algoritmos de otimização para transformar consultas baseadas em cálculo relacional em planos de execução eficientes, minimizando o uso de recursos como tempo de CPU e memória.

Cálculo Relacional em Lógica Temporal e Espacial

Outro campo de aplicação do cálculo relacional é a **lógica temporal e espacial**, onde se introduzem dimensões temporais para modelar eventos que ocorrem ao longo do tempo. Essa abordagem é especialmente importante em sistemas de controle, como monitoramento de redes de tráfego ou sistemas de controle industrial. O cálculo relacional pode ser expandido para lidar com aspectos temporais e espaciais, permitindo consultas que consideram a evolução dos dados ao longo do tempo.

Cálculo Relacional e Big Data

No contexto de **Big Data**, o cálculo relacional tem mostrado ser uma ferramenta poderosa para a otimização de consultas em grandes volumes de dados. Em sistemas que processam petabytes de informações diariamente, como aqueles utilizados por empresas de tecnologia, bancos e governos, o cálculo relacional permite formular consultas complexas que podem ser automaticamente otimizadas pelo sistema. Isso é vital para garantir que o tempo de resposta e o uso de recursos sejam eficientes, mesmo com grandes volumes de dados.

Comparações Internacionais e Pesquisas Avançadas

Em diferentes partes do mundo, o cálculo relacional é estudado e aplicado de maneiras variadas. Nos países de língua russa, por exemplo, há um grande foco em **lógica temporal e**

espacial, aplicada a sistemas industriais de controle e infraestrutura. Esses estudos são essenciais para a otimização de sistemas que envolvem grandes volumes de dados distribuídos ao longo de redes de controle e sistemas de automação.

Na Índia, centros de pesquisa como o **Indian Institute of Technology (IIT)** e o **National Informatics Centre (NIC)** têm explorado o uso do cálculo relacional em **grandes bancos de dados** voltados para sistemas financeiros e governamentais. O cálculo relacional é visto como uma ferramenta vital para a integração de sistemas de dados heterogêneos, melhorando a interoperabilidade entre diferentes plataformas de gerenciamento de dados.

Cálculo Relacional e Machine Learning

Com a crescente integração entre bancos de dados e **machine learning (ML)**, há um interesse crescente em como o cálculo relacional pode ser combinado com algoritmos de aprendizado de máquina para melhorar a eficiência de sistemas de inteligência artificial. Pesquisadores estão explorando maneiras de usar o cálculo relacional para gerar consultas mais eficientes que possam ser usadas em pipelines de aprendizado de máquina, particularmente na extração e pré-processamento de dados antes da aplicação de algoritmos de ML.

Cálculo Relacional na Prática

O cálculo relacional tem uma aplicação prática significativa em sistemas de bancos de dados, especialmente em grandes **sistemas financeiros** que processam volumes **massivos de dados em tempo real**. Sua abordagem declarativa simplifica a expressão de consultas complexas, tornando-o uma ferramenta crucial para a otimização de consultas em ambientes onde a eficiência é fundamental. Um exemplo típico de uso do cálculo relacional em bancos é na detecção de fraudes. As instituições financeiras precisam verificar milhões de transações diariamente para identificar padrões suspeitos. O cálculo relacional permite a formulação de consultas que cruzam diferentes atributos, como montante da transação, localização e frequência de operações, para identificar comportamentos atípicos.

Essa abordagem permite que, ao invés de especificar como a verificação das transações deve ser feita, os bancos simplesmente descrevam as condições necessárias, e o sistema se encarrega de otimizar o processo, tornando possível processar grandes quantidades de dados rapidamente. Além disso, o cálculo relacional facilita a interoperabilidade entre diferentes sistemas de banco de dados, permitindo consultas padronizadas que podem ser aplicadas a diversas fontes de dados de maneira eficiente.

A precisão dos resultados em bancos de dados é aprimorada com o uso de quantificadores, garantindo que as consultas retornem os resultados corretos, sem omitir ou incluir dados irrelevantes. Em cenários onde o tempo de resposta é crítico, como em sistemas financeiros ou de e-commerce, o uso de \exists e \forall em conjunto com o cálculo relacional torna o processo mais eficiente e confiável.

Conclusão

O cálculo relacional é uma ferramenta essencial para a teoria e prática dos bancos de dados modernos. Sua natureza declarativa oferece flexibilidade e permite a otimização automática de consultas, o que é crucial em sistemas grandes e complexos. Seja em sistemas de controle industriais, grandes bancos de dados empresariais ou no desenvolvimento de novas linguagens de consulta, o cálculo relacional continua a ser um campo de pesquisa e aplicação de grande relevância.

Referências:

이상철. (2019). 데이터베이스 관리 시스템: 이론과 실무 (*Database Management Systems: Theory and Practice*)

고영수. (2011). 관계형 데이터베이스 이론과 응용 (*Relational Database Theory and Applications*)

गर्ग, सुनील कुमार – डेटाबेस प्रबंधन प्रणाली (*Database Management System*). वाणी प्रकाशन, 2012.

शर्मा, विनोद कुमार – रिलेशनल डेटाबेस थ्योरी और एप्लीकेशन (*Relational Database Theory and Application*). प्रभात प्रकाशन, 2018.

Иванов, В.А. – *Реляционные базы данных: Теория и практика*. СПб: Питер, 2015.

Сергей Новиков – *Модели и методы реляционных баз данных*. Издательство "Физматлит", 2008.

Гржибовский, И.В. – *Теория баз данных: Модели и методы*. Москва: Издательство МГУ, 2004.

Codd, Edgar F. – "A Relational Model of Data for Large Shared Data Banks." *Communications of the ACM*, vol. 13, no. 6, 1970, pp. 377-387.

JAVA TPOINT. DBMS Relational Calculus. Disponível em:
<https://www.javatpoint.com/dbms-relational-calculus>. Acesso em: 1 out. 2024.

KENT, Andrew. Relational Algebra, Relational Calculus, and SQL. Disponível em:
https://cs.nyu.edu/~jcf/classes/CSCI-GA.2433-001_sp15/slides/session5/RelationalAlgebra-RelationalCalculus-SQL.pdf. Acesso em: 1 out. 2024.

TUNG, Anthony K. Relational Calculus. Disponível em:
<https://cseweb.ucsd.edu/classes/fa19/cse132A-a/slides/relational-calculus.pdf>. Acesso em: 3 out. 2024.

ESPINASSE, Bernard. Algèbre et Calcul Relationnels. Disponível em:
<https://pageperso.lis-lab.fr/bernard.espinasse/wp-content/uploads/2021/12/AlgebreCalculRelationnelle-4p.pdf>. Acesso em: 2 out. 2024.

ANONYMOUS. [relational calculus in DBMS]. YouTube. Disponível em:
<https://youtu.be/InZ23qi-588>. Acesso em: 4 out. 2024.