

Especificación formal de cubos de datos aplicados a la administración de las actividades docentes

Especificação formal de cubos de dados aplicada à administração de atividades de ensino

Pedro Fletes Gudiño

Instituto Tecnológico de Colima

pfletes@itcolima.edu.mx

Nicandro Farías Mendoza

Instituto Tecnológico de Colima

nfarias@itcolima.edu.mx

Resumen

Durante los últimos años las bases de datos multidimensionales se han convertido en algo común en el mundo de los negocios y científico. En este trabajo se propone un modelo para la especificación formal de base de datos multidimensionales (BDMD) o cubos de datos, que nos permita a través de un proceso clasificar, filtrar información del área de docencia, seccionando los registros en ejes y capas. La meta que se pretende alcanzar a través de la propuesta de este modelo es la especificación formal del cubo de datos.

Palabras clave: base de datos multidimensionales, cubos de datos, docencia y modelo.

Resumo

Nos últimos anos, os bancos de dados multidimensionais tornaram-se comuns no mundo dos negócios e da ciência. Este artigo apresenta um modelo para a especificação formal de uma base de dados multidimensional (BDMD) ou cubos de

dados é proposto, que nos permite através de um processo de triagem, filtragem da informação área de ensino, o corte de toras e machados camadas. A meta a ser alcançada através da proposta deste modelo é a especificação formal do cubo de dados.

Palavras-chave: base de dados multidimensional, cubos de dados e modelo de ensino..

Fecha recepción: Febrero 2013

Fecha aceptación: Mayo 2013

Introdução

Razões para escolher o tópico

Durante os últimos vinte anos, minha principal função era como um professor do Instituto Tecnológico de Colima, sempre no Departamento de Sistemas e Computação. Uma área que sempre tem uma licenciatura em ciência da computação estão entusiasmados com bases de dados e por esta razão este projeto de tese foi selecionado para continuar com o tema atual dos bancos de dados multidimensionais.

Contexto do problema

Actualmente, a nível nacional, não a exceção sendo o nosso instituto, estamos enfrentando um problema por elevadas taxas de insucesso dos alunos. Desde sempre que eles têm procurado várias maneiras de reduzir essas taxas, mas a nível institucional, há uma ferramenta computacional para manter o controle sobre eles, que também fornece informações estatísticas de diferentes anos e de diferentes parâmetros. Isso nos levou a desenvolver a nossa proposta vai gerar, através da especificação formal, um esquema computacional (cubo de dados), capaz de armazenar informações pertinentes sobre as taxas de insucesso acima e produzir informações de forma rápida e eficaz sobre as diferentes causas desaprovação originam na escola.

Alvo

Desenvolver um modelo formal para a especificação de bancos de dados multidimensionais que podem ser representados com SQL para uma orientada para a gestão de casos Professores atividades de estudo.

Objetivos específicos

- Definir o modelo formal.
- Especifique o estudo de caso.
- Detalhe o fluxo de dados para o estudo de caso.

Estado de Conhecimento

Contexto Histórico

Este projecto teve o seu início no Departamento de Sistemas e Informática Ensino Projetos Sede, onde cada controlar metade dos planejamentos de os oferecidos leva. Os professores no início do planejamento semestre entregues cada uma das matérias ensinadas e os instrumentos de ensino de cada uma das unidades. Durante o curso do semestre, a classificação do professor e do progresso de entrega no final dos registros de entrega de curso de qualificação, onde a taxa de aprovação e desaprovação do grupo aparece.

Framework 1.2 Contextual

Neste capítulo, uma breve descrição de outros trabalhos apresentados na modelagem conceitual e lógica dos bancos de dados multidimensionais. Para facilitar a compreensão deste trabalho e unificar a terminologia, previamente uma introdução às estruturas e operações de modelos multidimensionais serão apresentados.

Base de dados multidimensionais

bancos de dados são projetados para desenvolver aplicações muito específicas, como a criação Cubos OLAP (Online Analytical Processing). Basicamente não diferem muito do banco de dados relacional (uma tabela em um banco de dados relacional também poderia estar em um banco de dados multidimensional), a diferença é sim um nível conceitual; nos campos de bancos de dados multidimensionais ou atributos de uma tabela pode ser de dois tipos, ou representam dimensões da tabela, ou representar métrica que você deseja estudar.

Versus bancos de dados multidimensionais Cubos OLAP

Cada uma destas tabelas pode ser comparada a um hipercúbica ou mais especificamente, se as ferramentas OLAP é um cubo OLAP, onde as dimensões das áreas correspondem ao tamanho do quadro (campos 'di ...'), e valor armazenado em cada célula do cubo é igual a métrica ou métricas (fi ... " campos) armazenado na tabela.

Este tipo de base de dados é aplicado ao sistema também chamado cubo OLAP multidimensional ou hiperligação. Ele consiste de fatos numéricos chamados medidas

que são classificadas por tamanho. Os metadados cubo é geralmente criado a partir de um esquema em estrela (ver Fig. 1) ou do floco de neve (ver Fig. 2) e usando tabelas em um banco de dados relacional.

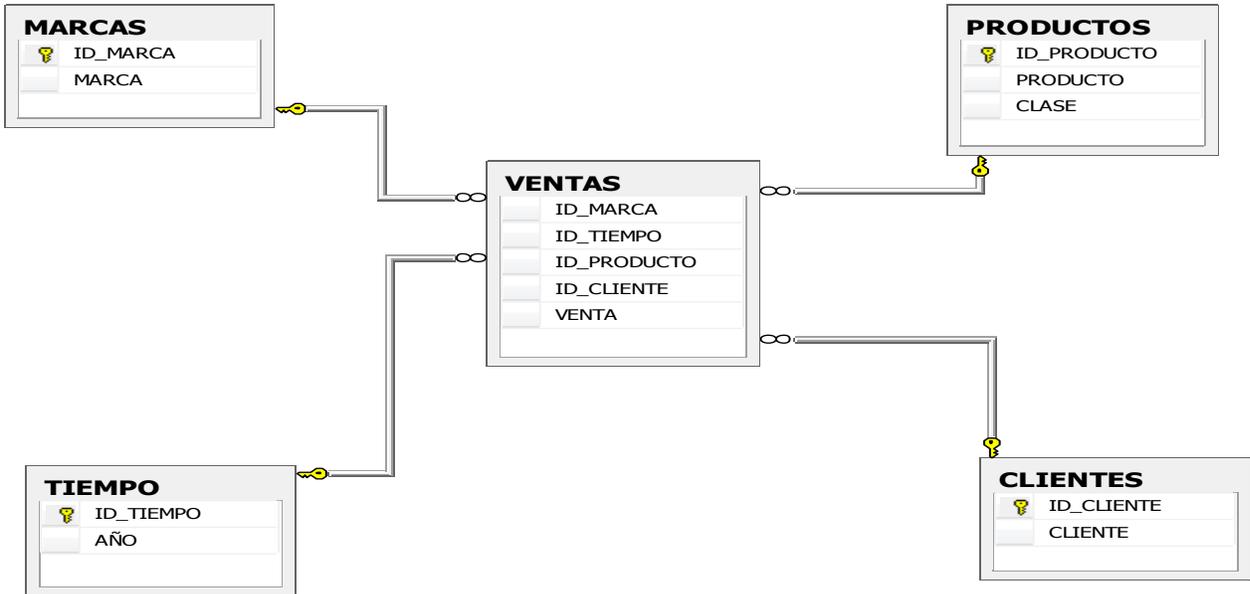


Fig.1 Esquema de estrella

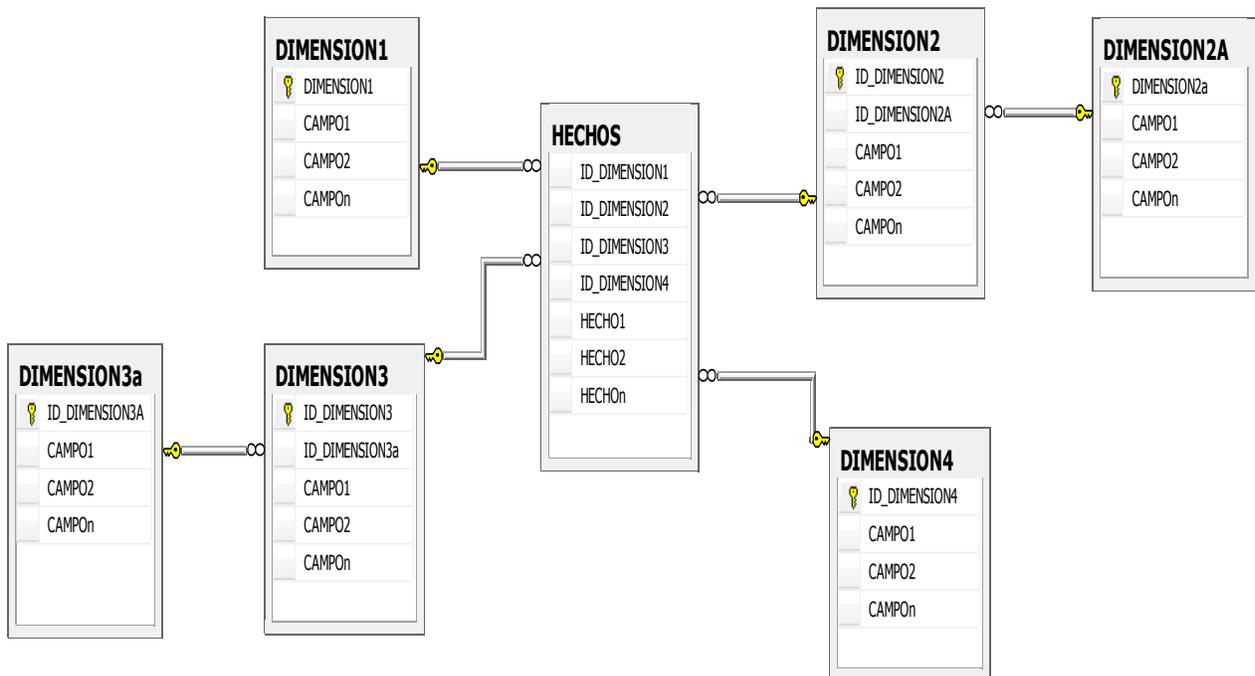


Fig.2 Esquema de copo de nieve

Trabalho e as contribuições relacionadas

(Seng & Habibollah, 2013) KSOM pretende usar para organizar os dados não estruturados para superfícies fechadas. KSOM utilizada neste trabalho, testando a sua capacidade de organizar os dados KSOM imagiologia médica que é usado principalmente para a construção de campo de engenharia de dados. (Hsiao, Wo-Shun, e Petchulat, 2011) mostram um protótipo funcional de um sistema OLAP cliente-centrado, com um middleware personalizado no lado do servidor e um cliente web que incorpora um motor de luz para os dados OLAP consultas na memória. Eles desenvolveram três ferramentas para visualização de dados interativos em execução no mecanismo de dados no lado do cliente. (Mansmann & Scholl, 2007) apresentam uma abordagem para explorar cubos de dados multidimensionais com técnicas de visualização hierárquicos. Um trabalho cuja principal contribuição é uma abordagem chamada hierárquica diferença de dispersão (HDS). Esta proposta permite vários níveis de hierarquia e explicitamente expostas nas diferenças entre eles no contexto da posição absoluta dos valores de articulação; É apresentado por (Piringer, Buchetics, Hauser, & Gröller, 2009). (Lafon, Bouali, Guinot & Venturini, 2013), temos várias formas de reorganizar um cubo OLAP em termos do qual todos os membros da reorganização for selecionado: todos os membros, os membros só irá mostrar certo nível (nível abordagem nível). O trabalho visa combinar a rota de células do tamanho e testes estatísticos paramétricos para identificar diferentes métricas significativas entre as células do cubo, é apresentado por (Ordonez, Chen, & Garcia-Garcia, 2011). (Takama

& Yamada, 2009) nos oferece uma exibição balde para interação modelagem na análise das tendências temporais e espaciais exploratória de dados. (Pitarch, Laurent & Poncelet, 2009) propõem um modelo conceitual para modelar hierarquias personalizadas bancos de dados multidimensionais. A proposta de construção de um algoritmo baseado em elipsóides (ER-árvore) apresentados por (Dankoand & Skopal, 2009) mostra que elas afetam significativamente a velocidade, indexação e desempenho de consulta em grandes bancos de dados, tais como multimídia , medicina e geografia. (Stolte, Tang, e Hanrahan, 2006) propõem uma linguagem de consulta visual que permite a visualização gráfica de dados com base em tabela. Este comandos de linguagem compila ambas as consultas que forem necessárias para gerar a exibição de gráficos, permitindo-nos para projetar sistemas que integram fortemente a análise e visualização. (Yaghmaie, Bertossi, & Ariyan, 2012) propõem a formalização do caminho do esquema relacional, que se torna a base para a obtenção de reparações dimensionais, aqui demonstrado que a estrela ordinária relacional e multidimensional esquemas de neve bancos de dados não são a melhor escolha para este processo. TEXplorer um novo sistema permite aos usuários procurar por palavras-chave e agregação OLAP e elegante, explorando os objetos de texto em um edifício em um banco de dados multidimensional texto que nos é apresentado (Zhao, hub Lin Ding, e Han, 2011). (Esch-Bussenmarkers & Cremers, 2004) propõem um tutorial de um sistema experimental em que o objetivo do projeto foi avaliar os diferentes tipos de acesso (voz, gráficos e modo de toque) para acessar e apresentar certos tipos de informação, e certas estratégias Pesquisar para pesquisar e navegar um banco de dados de dados multidimensionais musicais usando um celular simulado.

Referencial Teórico

Estes são os itens incluídos neste projeto são detalhados:

Bases de dados

Um sistema de gerenciamento de banco de dados (Sistema de Gestão de Base de Dados DBMS) é uma coleção de dados inter-relacionados e um conjunto de programas para acessar os dados (Silberschatz, Kotrh, e Sudarshan, 2002, p. 1). A coleção compartilhada de dados relacionados logicamente, juntamente com uma descrição destes dados, que são projetados para atender as necessidades de informação de uma organização (Connolly & Begg, 2005, pág. 15).

Base de dados multidimensionais

Banco de dados orientado com base em consultas complexas com estrutura de alta performance. Você pode usar uma estrela de RDBMS (multidimensional nível lógico banco de dados) ou SGBDM (banco de dados multidimensional para níveis lógicos e físicos ou Pura Multidimensional Banco de Dados) (De la Herran Gascon, 2004).

Cubos de dados

O OLAP Cube, que cunhou o nome de sua característica multidimensional é um banco de dados que tem várias dimensões.

Cubos de dados são usados em sistemas de processamento analítico online (OLAP) para apoiar a tomada de decisão. Construído a partir de bases de dados de uma empresa, este sistema de visualização interativa apresenta uma condicional 1D árvore estrutura cúbica hierárquica para representar cubos de dados e usando ícones gráficos

2D para ilustrar os elementos de dados. Os usuários podem, então, explorar de forma interativa de dados multi-dimensionais em níveis hierárquicos.

Desenvolvimento de cubo de dados aplicada à administração de atividades de ensino

Análise de Sistemas

Em todos os institutos tecnológicos em todo o país, existem departamentos acadêmicos que gerenciam as carreiras oferecidas e, em cada sede lá ensinando. Actualmente, existe um sistema automatizado que nos permite fazer um registro de dados, em seguida, uma informação como essa demanda hoje.

Portanto, não é possível obter informações de controle (aprovação estatísticas, as taxas de insucesso e de abandono) de cada um dos professores (por assunto, o horário de aula, a metade) das raças, o departamento eo Instituto. Além disso, não é possível obter facilmente a partir de informações semestres anteriores.

Modelo Conceptual

O diagrama multidimensional proposto (ver Figura 1) mostra como a informação da SIITEC¹ e diferentes áreas de ensino de departamentos acadêmicos, a informação está a sofrer transformações e segmentar (passos 1 e 2) para atingir o cubo de dados (passo 3) e já não pode ser processado para consultas, conforme necessário (passo 4).

¹ Sistema de Información Integral del Instituto Tecnológico de Colima.

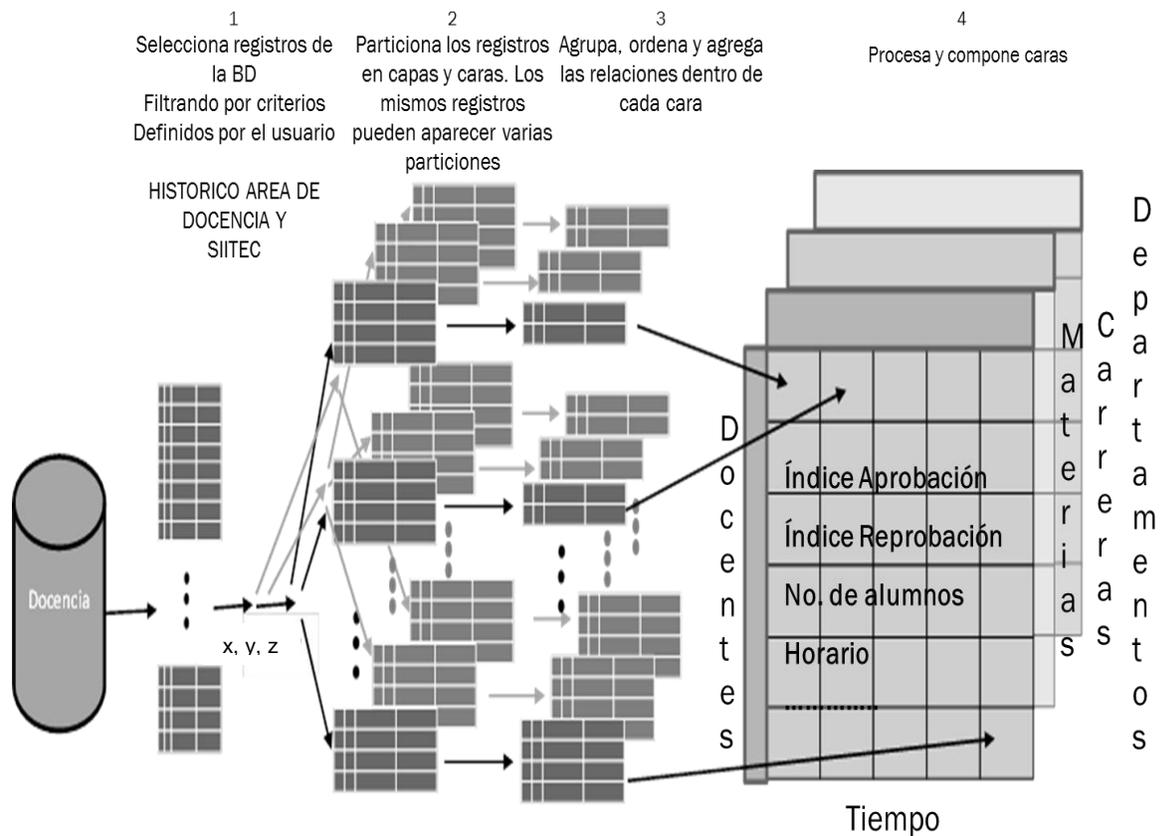


Figura 1 Big Picture del esquema multidimensional propuesto.

Requisitos do sistema

Neste ponto, considerando todas as informações você pode obter o cubo de dados para avaliar os aspectos para os quais a reprovação de alunos em cada um dos cursos de assuntos durante a sua estada no Instituto e vai acontecer:

- Indivíduos com maior taxa de insucesso
- Programação com o maior índice de reprovação
- Os professores com maior taxa de insucesso
- Aspectos que influenciam a desaprovação
- estado civil dos estudantes.

- Regime de trabalho dos estudantes (se tiverem um emprego).
- Sexo dos estudantes.
- Localização
- Estudantes (locais ou estrangeiras).
- Grant (se o aluno tem algum apoio).
- Se você tiver qualquer corrida anterior.
- O cronograma da matéria.

Toda esta informação pode ser obtida a partir do ponto de vista de cada uma das dimensões do cubo: o ensino, o tempo, departamento, carreira e matéria.

Modelo de especificação formal para Multidimensional Data Base (BDMD) ou cubos de dados

O modelo para a especificação formal de cubos de dados consiste nos seguintes cinco vezes:

$$\mathbf{MCD}^2 = (C, OAT, OPAT, PR, OBD)$$

Onde:

$C = \{ \text{campos ordinais, campos quantitativos, dimensões, medidas} \}$ são os princípios básicos.

$OAT = \{A, B, C, P, Q, R\}$ são os operadores da álgebra de tabelas.

$OPAT = \{+, X, / \}$ são os operadores da álgebra de tabelas.

$PR = \{ \text{Filtro, Fila } (i), \text{ Columna } (j), \text{ Capa } (k) \}$ são os predicados.

$OBD = \{ \text{Agregados, dim, SUM, AVG} \}$ Operadores de Banco de Dados.

² Modelo de cubo de datos.

Este modelo proposto é baseado em uma álgebra de tabelas que captura a estrutura dos quadros do modelo relacional e executa a conversão destes para um formato de cubos de dados formalismo. Conceitos utilizados:

Campos ordinais: A escala ordinal é discretamente apresentado como cabeçalhos ou de tipos diferentes.

Campos quantitativos: quantitativa contínua e os campos são mostrados como eixos ou como valores continuamente variáveis.

Mesas de álgebra

Nós definimos uma álgebra como o mecanismo formal para especificar tabelas de configuração. A configuração completa consiste em três tabelas distintas expressões de álgebra. Duas expressões definir a configuração de eixos X e mesa eixo Y seccionamento os eixos e as colunas da tabela. O terceiro termo define o eixo Z da mesa, que parte nas camadas desdobradas. As expressões X, Y e Z são cláusulas idioma.

Os operandos na álgebra de nomes de tabela são campos ordinais ou quantitativas entre o banco de dados. Ele está usando A, B e C para representar campos ordinais e P, Q e R representar campos quantitativos. Sequências de atribuir valores para cada símbolo de campo da seguinte forma: Campos ordinais para atribuir membros de um domínio ordenado de campo e define campos quantitativos dos itens individuais contendo o nome do campo é atribuído.

Campos ordinais e quantitativos gerar tabelas com diferentes estruturas (1) - (2). Campos da tabela segmento ordinais em linhas e colunas usando cabeçalhos, enquanto campos quantitativos geram eixos.

$$A = \text{domain}(A) = \{a_1, \dots, a_n\} \quad (1)$$

$$P = \{P\} \quad (2)$$

Uma expressão válida em álgebra consiste em um ou mais símbolos com operadores entre cada par de operandos adjacentes e parênteses utilizados para alterar a precedência de operadores.

Operadores de mesas de álgebra

Concatenación (+) El operador + (3)-(5) concatena dos secuencias de la siguiente forma:

$$\begin{aligned} A + B &= \{a_1, \dots, a_n\} + \{b_1, \dots, b_m\} \\ &= \{a_1, \dots, a_n, b_1, \dots, b_m\} \end{aligned} \quad (3)$$

$$\begin{aligned} A + P &= \{a_1, \dots, a_n\} + \{P\} \\ &= \{a_1, \dots, a_n, P\} \end{aligned} \quad (4)$$

$$\begin{aligned} P + Q &= \{P\} + \{Q\} \\ &= \{P, Q\} \end{aligned} \quad (5)$$

Producto: (X) O operador do produto (6) - (7) leva o produto cartesiano de duas seqüências.

$$\begin{aligned} A \times B &= \{a_1, \dots, a_n\} \times \{b_1, \dots, b_m\} \\ &= \{a_1b_1, \dots, a_1b_m, \\ &\quad a_2b_1, \dots, a_2b_m, \dots, \\ &\quad a_nb_1, \dots, a_nb_m\} \end{aligned} \quad (6)$$

$$\begin{aligned} A \times P &= \{a_1, \dots, a_n\} \times P \\ &= \{a_1P, \dots, a_nP\} \end{aligned} \quad (7)$$

Projeção (/), a saliência (8) é semelhante ao operador o operador do produto, mas apenas cria sequências para as quais existem registros.

$$A/B = \{a_i b_j \mid \exists r \in R \text{ st } A(r) = a_i \ \& \ B(r) = b_j\} \quad (8)$$

A interpretação intuitiva do operador de projeção B em A.

Por exemplo, dados os campos trimestre e mês, a expressão trimestre / mês poderia ser interpretada como aqueles meses dentro de cada trimestre, resultando em três bilhetes cada trimestre. Em contraste, X mês trimestre poderia ser de 12 entradas para cada trimestre. Cubos de dados representam hierarquias e não precisa explicitamente calcular a relação de mapeamento.

Consultas de processamento de dados e geração de no banco de dados

Um aspecto importante do formalismo proposto é a transformação de dados. Assim, o formalismo deve suportar toda a gama de transformações de dados em uma consulta em uma linguagem de consulta SQL como, inclusive operadores comuns relacionais: seleção, filtragem, agrupamento e sistema de agregação.

Fluxo total de dados no modelo MCD

O formalismo deve ser compatível com toda a gama de possíveis transformações de dados em uma linguagem de consulta SQL como, incluindo operadores relacionais comuns: seleção, filtragem, agrupamento, agregação e classificação. Pode ser

demonstrado que qualquer consulta SQL definida pode ser expressa como um MCD especificação formalismo.

Nesta proposta, como campos são adicionados enquanto os campos de dimensão são inseridos em um GROUP BY estatuto com dimensões adicionais campos especificados em detalhe. Cada dimensão pode também ser encomendados, e as diferentes funções de agregação pode ser associada a cada medida e estas opções pode ser escolhido. Há também um filtro que representa os elementos na cláusula ONDE. Finalmente, também são apresentados os cálculos gerais e junta.

É referência à Figura 1, é explicado em pormenor cada um dos passos mostrados na figura descrita pelo modelo MCD.

Passo 1. Seleção de registros. A primeira fase do fluxo de dados recupera registros do banco de dados, usando filtros definidos pelo usuário para selecionar subconjuntos de dados.

Para um campo ordinal U_m , o utilizador pode especificar um subconjunto do domínio de campo como o filtro (A) é um subconjunto seleccionado pelo utilizador. Em seguida, um predicado relacional expressando o filtro para A é:

$A \text{ in } filter(A)$

Para um campo quantitativa, o utilizador pode definir um subconjunto do domínio do campo como válido se $min(P)$ y $max(P)$ extensões são usuário subconjunto definido. O predicado relacional expressar o filtro para P é:

$(P \geq \min(P) \text{ and } P \leq \max(P))$

Podemos definir o filtro de predicado relacional como uma combinação de filtros de todos os campos individuais. Em seguida, a primeira fase do processamento de dados da rede é equivalente para a seguinte declaração SQL:

```
SELECT *  
WHERE {filtros}
```

É possível dentro de todo o formalismo para definir um filtro mais sofisticada, como unidades de campo múltipla cruzada produto filtro ou sistema de filtros (filtro A é calculada em relação ao filtro B).

Passo 2. Painéis de partição de registos: A segunda fase do fluxo de dados partições recuperados registos em grupos correspondentes a cada painel na tabela. A tabela é dividida em linhas, colunas e camadas correspondentes às entradas nestes conjuntos.

Os valores ordinais em cada entrada de definir os critérios estabelecidos pela qual os registos são dispostos em cada linha, coluna e camada. Dada a linha (i) ser o predicado que representa os critérios de seleção para o i-th linha, coluna (j) é o predicado da coluna j-th, e Layer (k) o predicado para a camada k-th . Por exemplo, se a tabela de eixo Y é definido pelo conjunto padronizado:

$\{a_1b_1P, a_1b_2P, a_2b_1P, a_2b_2P\}$

Abaixo estão quatro linhas da tabela, cada uma definida por uma entrada nesse grupo, ea linha é definida como:

Fila (1) = (A = a₁ and B = b₁)

Fila (2) = (A = a₁ and B = b₂)

Fila (3) = (A = a₂ and B = b₁)

Fila (4) = (A = a₂ and B = b₂)

Dadas essas definições, os registros em que o painel é dividido é a interseção da linha i-th e j-ésima coluna e a camada k-th podem ser recuperados com a seguinte consulta:

```
SELECT *  
WHERE {Fila (i) and Columna (j) and Capa (k)}
```

Passo 3. Transformação de registros dentro dos painéis: A última fase do fluxo de dados é a transformação dos registros em cada painel. Se a especificação visual inclui agregação, em seguida, todas as medidas no esquema de banco de dados deve ser atribuído a um operador de agregação.

Por exemplo, se o banco de dados contém os lucros, vendas e folha de pagamento campos quantitativos, e que o usuário tenha especificado explicitamente que a média de vendas deve ser calculado, em seguida, os agregados são definidas como:

Aggregate = SUM (Profit), AVG (vendas), SUM (Payroll)

Filtros agregar campos (por exemplo, SUM (Profit > 500), não pôde ser avaliado na Etapa 1, com todos os outros filtros, porque os agregados ainda não tinha sido calculada. Por isso, os filtros devem ser aplicadas em esta fase. Nós definimos os filtros de predicados relacionais como no passo 1 para campos não agregadas.

Além disso, definir as seguintes listas:

G: Os nomes dos campos no grupo de plataformas,

S: Os nomes dos campos na plataforma de triagem, e

Dim: Dimensões em banco de dados.

A transformação necessária pode ser expressa pelo acórdão SQL:

```
SELECT {dim}, {agregados}
GROUP BY {G}
HAVING {filters}
ORDER BY {S}
```

Se os campos adicionados não estão incluídos na especificação, o processamento restante classifica apenas os registos em ordem:

```
SELECT *
ORDER BY {S}
```

Projeto

Porque o esquema proposto é a geração de um cubo de dados para a exploração da informação na área de ensino através de uma especificação formal, toma como entrada uma base de dados relacional esquema. Na Figura 2 o esquema multidimensional que especifica cada uma das dimensões e medidas da tabela de fato (GROUP) propostos é mostrado.

Projeto arquitetônico

Propõe-se que o grupo tabela de fatos também contém os atributos já mencionados (Índice de aprovação, taxa de reprovação, número de alunos e horas matéria), dados adicionais para complementar as informações e vamos oferecer a possibilidade de extrair informações a partir de diferentes pontos de vista. A Figura 2 mostra o modelo

conceitual do esquema proposto é mostrado. A seguir todos os atributos de cada uma das dimensões e a tabela de verdade são especificados.

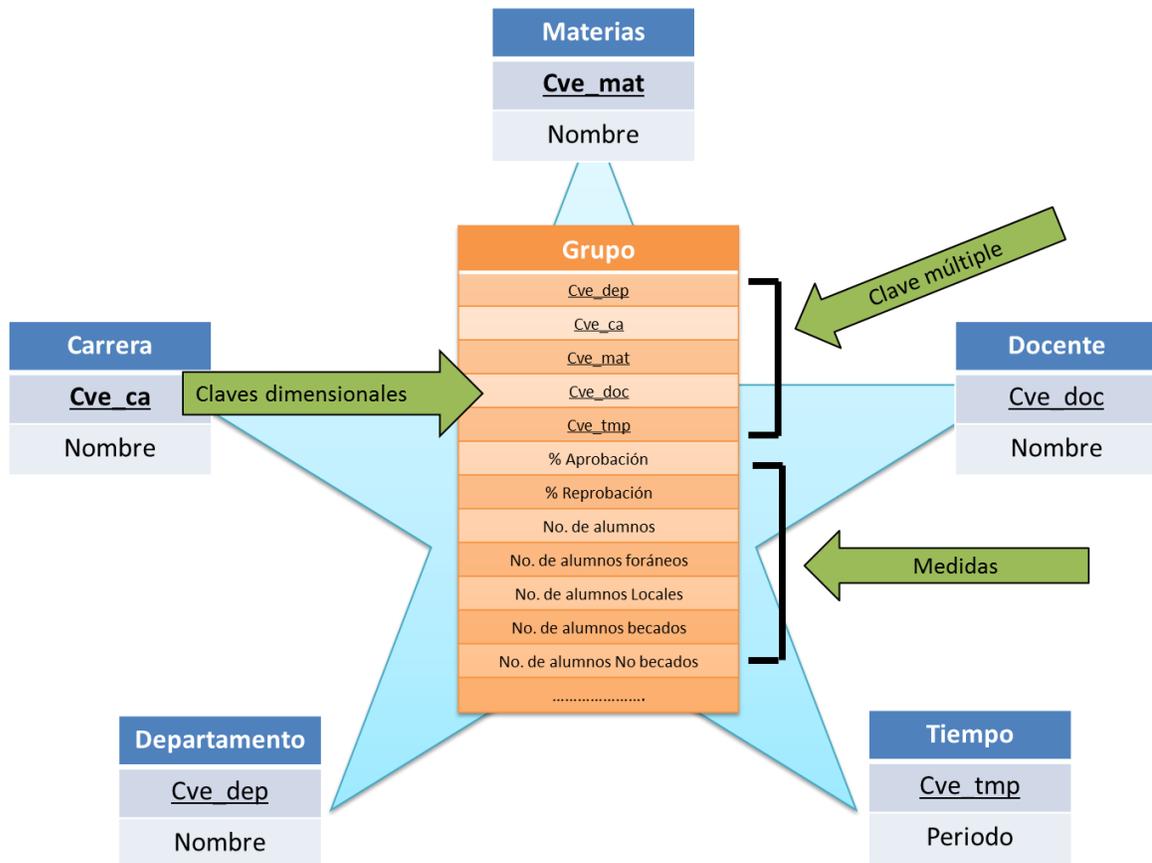


Figura 2 Modelo Conceptual.

Projeto cubo de dados

Com referência ao modelo conceitual da Figura 2, o design proposto é mostrado no seguinte esquema E-R (ver Figura 3).

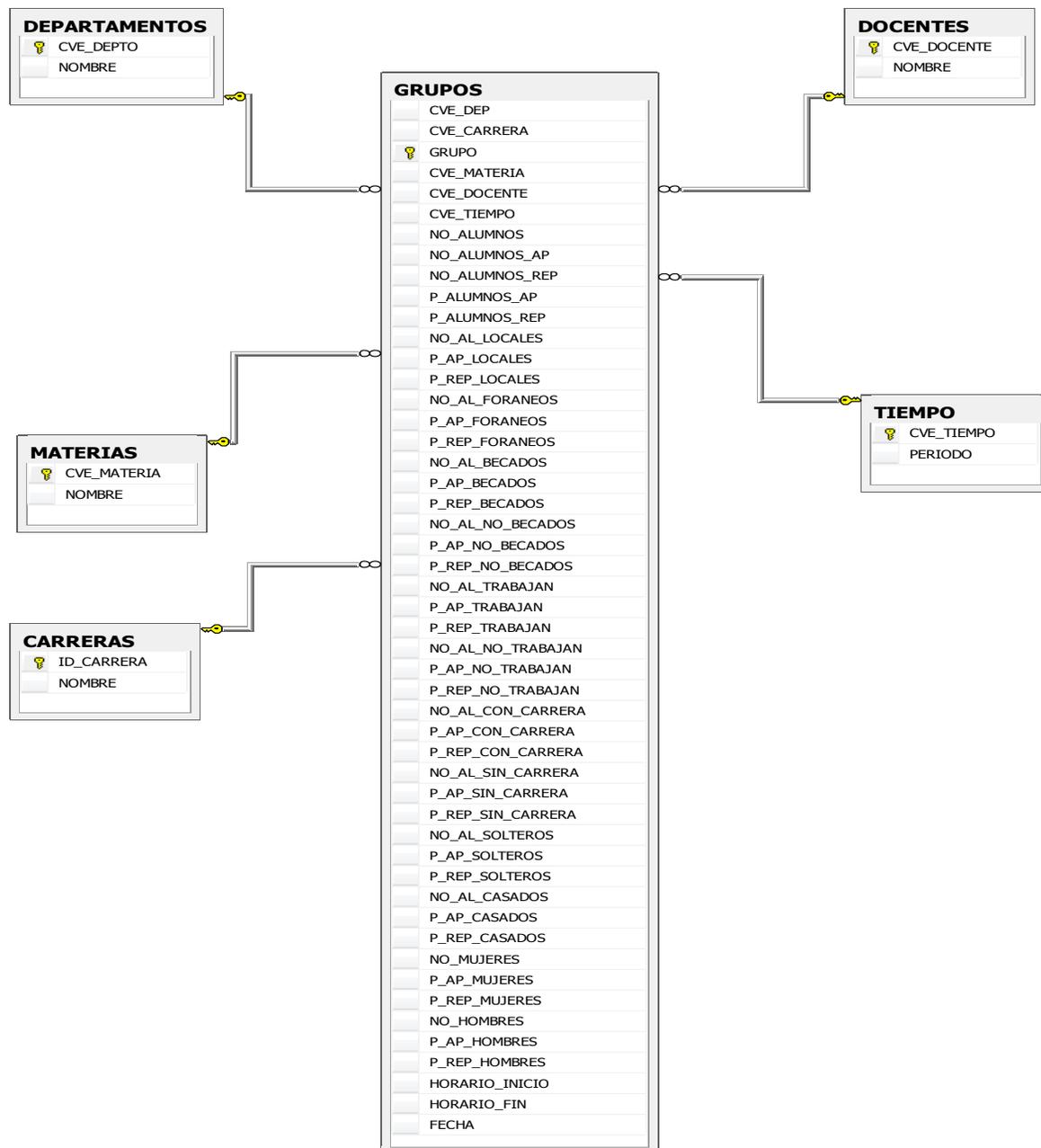


Figura 3 Diagrama de E-R del cubo de datos.

No = Número, P = Porcentaje, AP = Aprobación, REP = Reprobación.

Estudo de caso

O seguinte caso mostra como a partir do esquema relacional na Figura 4 e usando o modelo MCD atinge o cubo de dados Figura 3.

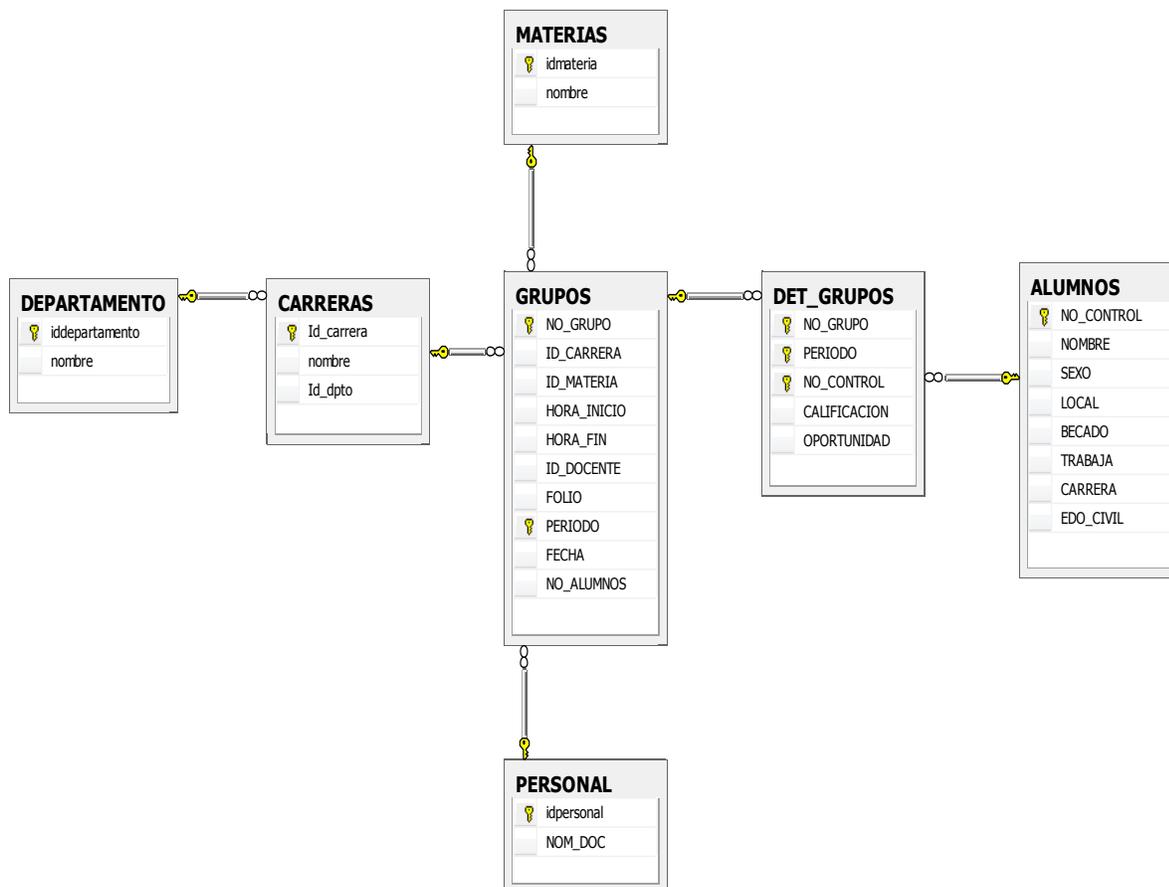


Figura 4. Modelo relacional de docencia.

No caso mostrado abaixo da geração das dimensões do cubo ea tabela de fatos é exibido (aqui você tem apenas o grupo-chave de dimensões -carrera chave, materiais, professor e em tempo departamento, quando Clique em Iniciar, fim da classe tempo, fólho, número e percentagem de estudantes que passam).

Instruções SQL que usam o código ficaria assim:

```

Select L.No_Grupo, L.Id_Carrera, L.Id_Materia, L.Hora_inicio, L.Hora_fin,
      L.Id_Docente, L.Folio, L.Fecha, L.No_Alumnos, P.Iddepartamento,
      Count (*), Count (*)/L.No_Alumnos
From Grupos L join Det_Grupos M   on L.No_Grupo = M.No_Grupo
              join Alumnos N     on M.No_control = N.No_Control
              join Carreras O     on L.Id_Carrera = O.Id_Carrera
              join Departamento P on O.Id_Dpto   = P.Iddepartamento
Group By L.No_Grupo
Having M.Calificacion >= 70

```

MCD utilizando o método proposto, a especificação seria como se segue:

```

Select A1, A2, A3, B1, B2, A4, B3, B4, P1, A5, Q1, Q2
From Grupos L / Det_Grupos M / Alumnos N / Carreras O / Departamento P
Group By A1
Having Filtro1

```

Onde:

A ₁ = L.No_Grupo	A ₄ = L.Id_Docente	Q ₁ = COUNT (*)
A ₂ = L.Id_Carrera	B ₃ = L.Folio	Q ₂ = COUNT (*) / P ₁
A ₃ = L.Id_Materia	B ₄ = L.Fecha	Filtro ₁ = {M.Calificacion >= 70}
B ₁ = L.Hora_inicio	P ₁ = L.No_Alumnos	
B ₂ = L.Hora_fin	A ₅ = P.Iddepartamento	

y L, M, N, O y P são variáveis de tupla ou Grupos tabelas de alias, grupos, estudantes,

Racing e Departamento respectivamente. Exibido nesta consulta é gerado:

- Os departamentos com atributo de dimensão P.Iddepartamento,
- A dimensão da carreira com atributo L.Id_Carrera,
- Os materiais de tamanho, com atributo L.Id_Materia,
- O atributo dimensão Professores com L.Id_Docente
- A dimensão de tempo com o atributo L.Fecha.

Para os atributos No_alumnos_Rep y P_Alumnos_Rep, o filtro é alterado para:

Filtro₁ = {M.Calificacion < 70}

Para os atributos No_AI_Locales, P_Ap_Locales y P_Rep_Locales, o filtro será:

Filtro₁ = {M.Calificación >= 70 and N.Local = 'L'} y

Q₂ = COUNT(*)/P₁

Q₃ = 100 – Q₂ y se añadiría al Select

Para os atributos No_AI_Foráneos, P_Ap_Foráneos y P_Rep_Foráneos, el filtro sería:

Filtro₁ = {M.Calificación >= 70 and N.Local = 'F'} y

Q₂ = COUNT(*)/P₁

Q₃ = 100 – Q₂ y se añadiría al Select

Para os atributos No_AI_Becados, P_Ap_Becados y P_Rep_Becados, el filtro sería:

Filtro₁ = {M.Calificación >= 70 and N.Becado = 'S'} y

Q₂ = COUNT(*)/P₁

Q₃ = 100 – Q₂ y se añadiría al Select

Para os atributos No_AI_No_Becados, P_Ap_No_Becados y P_Rep_No_Becados, el

filtro sería:

Filtro₁ = {M.Calificación >= 70 and N.Becado = 'N'} y

Q₂ = COUNT(*)/P₁

Q₃ = 100 – Q₂ y se añadiría al Select

Para os atributos No_AI_Trabajan, P_Ap_Trabajan y P_Rep_Trabajan, el filtro sería:

Filtro₁ = {M.Calificación >= 70 and N.Trabaja = 'S'} y

Q₂ = COUNT(*)/P₁

Q₃ = 100 – Q₂ y se añadiría al Select

Para os atributos No_AI_No_Trabajan, P_Ap_No_Trabajan y P_Rep_No_Trabajan, el filtro sería:

Filtro₁ = {M.Calificación >= 70 and N.Trabaja = 'N'} y
Q₂ = COUNT(*)/P₁
Q₃ = 100 – Q₂ y se añadiría al Select

Para os atributos No_AI_Con_Carrera, P_Ap_Con_Carrera y P_Rep_Con_Carrera, el filtro sería:

Filtro₁ = {M.Calificación >= 70 and N.Carrera = 'S'} y
Q₂ = COUNT(*)/P₁
Q₃ = 100 – Q₂ y se añadiría al Select

Para os atributos No_AI_Sin_Carrera, P_Ap_Sin_Carrera y P_Rep_Sin_Carrera, el filtro sería:

Filtro₁ = {M.Calificación >= 70 and N.Carrera = 'N'} y
Q₂ = COUNT(*)/P₁
Q₃ = 100 – Q₂ y se añadiría al Select

Para obtener los atributos No_AI_Solteros, P_Ap_Solteros y P_Rep_Solteros, el filtro sería:

Filtro₁ = {M.Calificación >= 70 and N.Edo_Civil = 'S'} y
Q₂ = COUNT(*)/P₁
Q₃ = 100 – Q₂ y se añadiría al Select

Para os atributos No_AI_Casados, P_Ap_Casados y P_Rep_Casados, el filtro sería:

Filtro₁ = {M.Calificación >= 70 and N.Edo_Civil = 'C'} y
Q₂ = COUNT(*)/P₁

$Q_3 = 100 - Q_2$ y se añadiría al Select

Para os atributos No_Mujeres, P_Ap_Mujeres y P_Rep_Mujeres, el filtro sería:

Filtro₁ = {M.Calificación >= 70 and N.Sexo = 'F'} y

$Q_2 = \text{COUNT}(\ast)/P_1$

$Q_3 = 100 - Q_2$ y se añadiría al Select

Para os atributos No_Hombres, P_Ap_Hombres y P_Rep_Hombres, el filtro sería:

Filtro₁ = {M.Calificación >= 70 and N.Sexo = 'M'} y

$Q_2 = \text{COUNT}(\ast)/P_1$

$Q_3 = 100 - Q_2$ y se añadiría al Select.

Com o acima exposto é formalmente especificado o cubo de dados, utilizando o modelo de MCD.

Flujo de datos total da Big Picture usando o estudo de caso

Para explicar cada una das etapas do Big Picture, Figura 1, vamos usar o exemplo para a obtenção do modelo de dados relacional com o cubo de dados na seção anterior.

Se nós temos o seguinte formalização:

```
Select A1, A2, A3, B1, B2, A4, B3, B4, P1, A5, Q1, Q2
From Grupos L / Det_Grupos M / Alumnos N / Carreras O / Departamento P
Group By A1
Having Filtro1
```

Onde:

$A_1 = L.No_Grupo$

$A_4 = L.Id_Docente$

$Q_1 = \text{COUNT}(\ast)$

$A_2 = L.Id_Carrera$ $B_3 = L.Folio$ $Q_2 = COUNT (*) / P_1$
 $A_3 = L.Id_Materia$ $B_4 = L.Fecha$ $Filtro_1 = \{M.Calificacion \geq 70\}$
 $B_1 = L.Hora_inicio$ $P_1 = L.No_Alumnos$
 $B_2 = L.Hora_fin$ $A_5 = P.Idepartamento$

y L, M, N,O y P são variáveis de tupla ou Grupos tabelas de alias, grupos, estudantes, Racing e Departamento respectivamente.

Passo 1. Seleccione os registros do banco de dados de filtragem por critérios definidos pelo usuário.

Esta etapa é definido através $Filter1 = \{M.Calificación \geq 70\}$, neste caso com o filtro o número ea percentagem de estudantes aprovados são obtidos.

Passo 2. Segmento registros em camadas e caro. Os mesmos registros podem aparecer em várias partições.

Deste modo, os dados são agrupados em faces através da cláusula:

```

Select A1, A2, A3, B1, B2, A4, B3, B4, P1, A5, Q1, Q2
From Grupos G / Det_Grupos D / Alumnos A / Carreras C / Departamento P
    
```

En el Select las variables A₅, A₃, A₂, A₄ y B₄, determinar as dimensões do Departamento, Matéria, Carreira, Professor e hora, respectivamente. Ligação de dados por meio da operação de projeção na cláusula FROM.

Passo 3. Group, tipos e adiciona as relações dentro de cada face.

Nesta etapa agrupados, condenada a ser adicionado em cada lado através da instrução:

```
Select A1, A2, A3, B1, B2, A4, B3, B4, P1, A5, Q1, Q2
```

Onde cada atributo é definido acima.

Passo 4. Processos e fazia caretas.

E com as informações armazenadas no cubo de dados através de instruções SQL pode ser obtido indicadores, conforme necessário.

Análise dos resultados

O escopo deste projeto é limitado à especificação formal do cubo de dados através do modelo MCD, que através do estudo de caso demonstraram.

Outro resultado é que qualquer desenho de um cubo de dados pode ser representado e visualizados através desta especificação.

Além disso, sem qualquer designer implementar o hub pode capturar e exibir o cubo de dados esboço final.

Conclusões e recomendações

Observando os resultados, podemos dizer que qualquer expressão representado pela linguagem SQL pode ser formalmente especificados usando o modelo de cubo de dados desenvolvido nesta pesquisa. Pelo acima, observa-se que os objetivos sejam

atingidos no início e, portanto, é integralmente adoptada a hipótese de trabalho para esta investigação.

A importância deste projecto consiste na concepção mostram um cubo de dados através de uma especificação formal usando o modelo MCD, neste caso, como uma solução proposta para ser indicadores de taxas de falhas que revelam as causas destes tarifas em diferentes carreiras oferecido pelo Instituto Tecnológico de Colima e dobre.

Também é importante notar que o esquema multidimensional obtida é analítico e capaz de armazenar tanto informações atuais e históricas, dando a oportunidade de gerir de forma eficiente as informações derivadas das atividades de ensino do Instituto.

Para continuar no futuro com esta pesquisa sugere:

- Use o outro esquema de representação como um cubo de dados é o floco de neve para apresentar as vantagens ou desvantagens desta abordagem.
- Aumente o tamanho dos cubos, que nos dá a capacidade de gerenciar outro tecnológica agrupados em regiões geográficas.

Referencias

Connolly, T. & Begg, C. (2005). *Sistemas de Bases de datos*. Madrid: Addison Wesley.

Dankoand, T. & Skopal, T. (2009). *Elliptic Indexing of Multidimensional Databases*. Wellington, New Zealand: Australasian Databases Conference.

De la Herrán Gascón, M. (2004). *Red Científica*. Recuperado el 28 de Noviembre de 2013, de <http://www.redcientifica.com/oracle/c0001p0005.html>

Esch-Bussenmarkers, V. & Cremers, A. (2004). *User Walkthrough of Multimodal Access to Multidimensional Databases*. Pennsylvania, USA: ICMI.

Hsiao, T.; Wo-Shun, L.; & Petchulat, S. (2011). *Data Visualization on Web-based OLAP*. Glasgow, Scotland, UK: DOLAP'2011.

Lafon, S.; Bouali, F.; Guinot, C. & Venturini, G. (2013). *Hierarchical Reorganization of Dimensions in OLAP Visualizations*. Francia: IEEECS.

Mansmann, S. & Scholl, M. H. (2007). *Exploring OLAP Aggregates with Hierarchical Visualization Techniques*. Seoul, Korea: SAC'07.

Ordonez, C.; Chen, Z. & García-García, J. (2011). *Interactive Exploration and Visualization of OLAP*. Glasgow, Scotland, UK: DOLAP'11.

Piringer, H.; Buchetics, M.; Hauser, H. & Gröller, E. (2009). *Hierarchical Difference Scatterplots*. París, Francia: VAKD'09.

Pitarch, Y; Laurent, A. & Poncelet, P. (2009). *A Conceptual Model For Handling Personalized Hierarchies in Multidimensional Databases*. Lyon, France: MEDES.

Seng, P. & Habibollah, H. (2013). *Cube Kohonen Self-Organizing*. Malaysia: IEEE Transactions on Neural Networks and Learning Systems Vol.24.

Silberschatz, A.; Kotrh, H. & Sudarshan , S. (2002). *Fundamentos de Bases de datos*. España: McGraw-Hill.

Stolte, C.; Tang, D. & Hanrahan, P. (2006). *Polaris: A System for Query, Analysis and Visualization of Multidimensional Databases*. ACM.

Takama, Y. & Yamada, T. (2009). *Visualization Cube: Modeling Interaction for Exploratory Data Analysis of Spatiotemporal Trend Information*. Tokyo, Japan: IEEE/WIC/ACM.

Yaghmaie, M.; Bertossi, L.; & Ariyan, S. (2012). *Repair-Oriented Relational Schemas for Multidimensional Databases*. Berlín, Alemania: EDBT.

Zhao, B.; Lin, x.; Ding, B. & Han, J. (2011). *TEXplorer: Keyword-based Object Search and Exploration in Multidimensional Text Databases*. Glasgow, Scotland, UK: CIKM.