

Resumen: Materialized Views and Data Warehouses

Los + Aplicados

1 Introducción

El artículo comienza planteando las diferentes maneras en las que podemos interpretar a las vistas: Un programa, Datos, Índices, Agregación OLAP. Según el autor: una vista puede ser alguna o todas las anteriores, dependiendo de la implementación que se utilice.

2 Maneras de intepretar las vistas

Las vistas no materializadas son la especificación de un programa, este artículo menciona que las primeras aplicaciones de este tipo de vista fueron: modificaciones de búsquedas y búsquedas compiladas. Esta aplicación es un programa puro. Además menciona que el costo de generar estas vistas es aproximadamente el mismo cada vez que se ejecutan.

Las vistas materializadas son una extensión de aplicación anterior, ya que podemos realizar búsquedas empleando las vistas materializadas para generar nuevas vistas; algo que el artículo llama "datos derivados". Mientras que realizar un *snapshot* convierte a la vista materializada en datos puros, esta información ya no se puede modificar.

Los índices de una vista generan *pointers* a los datos sobre los cuales se generó una vista, la importancia que se le dé a esta aplicación depende de la organización y los datos que maneje, el artículo llama a esto "índices puros". Se pueden crear vistas parcialmente materializadas, las cuales contienen datos; y además referencia a datos que no son almacenados directamente como **pointers**.

Un conjunto de vistas o índices materializados se denomina un "cubo de datos", estos conjuntos son proyecciones de datos almacenados con perspectiva multidimensional a dimensiones menores.

3 Descubrimiento y Reutilización de Vistas

De acuerdo al texto, los sistemas manejadores de bases de datos generan vistas materializadas continuamente; sin embargo, los sistemas más comerciales descartan dichas vistas al ser entregadas al usuario o al pasar a la siguiente fase en la ejecución de una rutina; detectar y rescatar dichas vistas es útil para ahorrar costos. Sin embargo, esto solo es posible si el optimizador de búsquedas logra descubrirlas.

La técnica más utilizada para realizar este proceso se denomina "subsunción", el artículo no ahonda en dicho proceso; pero menciona que en general es un problema indecidible, sin embargo: en las búsquedas más comunes puede reducirse a un problema NP-completo, y para búsquedas simples existen algoritmos en tiempo polinomial bastante optimizados que resuelven el problema.

En un almacén de datos es importante reutilizar vistas ya que reduce costos de operación y hace labores subsiguientes más eficientes.

4 Procesamiento de Vistas

El procesamiento de vistas es una labor que incluye las acciones de actualizar una vista y el escaneo de la base de datos. El escaneo de una vista no materializada cuesta lo mismo que volver a ejecutar la búsqueda que la generó en primer lugar; por lo que no existe beneficio al emplearlas.

Escanear una vista materializada tiene un costo que depende de la proporción de tuplas útiles para resolver una búsqueda, esto se denomina "densidad de la vista". En la forma indexada de una vista, es importante que los pointers estén bien organizados para reducir el costo computacional.

El costo de actualizar una vista materializada depende de si se realiza al mismo tiempo que durante el escaneo, o si se realiza después; mientras que en una vista no materializada el costo es el mismo que el de volver a crearla.

En un almacén de datos se suelen actualizar las vistas en momentos diferentes al escaneo.

5 Minería y Aprovechamiento de las Vistas

Las vistas coontienen información útil como distribución de datos numéricos y otros valores; esto mejora las estimaciones sobre el comportamiento estadístico de los datos

Emplear vistas materializadas de forma frecuente pueden aprovecharse para obtener patrones de acceso, los cuales mejoran el manejo de la memoria y reduce el sobre coste sobre la base de datos. Por lo que es útil incorporarlas a nuestra lista de herramientas para manejar una base de datos.