

# **BD Avanzada**

## **Tema 3: Optimización de Consultas**

# Optimización de Consultas

## Objetivo General:

- Al completar este tema el estudiante estará en la capacidad de manejar los conceptos sobre Optimización de Consultas.

**Objetivos Específicos:** Al completar este tema el estudiante deberá estar en la capacidad de discutir sobre los siguientes conceptos:

- Aspectos de optimización de consultas.
- Usar consultas correlacionales usando SQL.

# Optimización de Consultas

- **Optimización de consultas**
  - Es el proceso de selección del plan de evaluación de las consultas más eficiente de entre las muchas estrategias generalmente disponibles para el procesamiento de una consulta dada, especialmente si la consulta es compleja.
- **Optimizadores**
  - Hacen uso de la información estadística sobre las relaciones, como los tamaños de las relaciones y las profundidades de los índices
- **Etapas generación de planes de evaluación de consultas:**
  1. La generación de expresiones que sean equivalentes lógicamente a la expresión dada y
  2. La anotación de las expresiones resultantes en maneras alternativas de generar planes de evaluación de consultas alternativos.

# Optimización de Consultas

- **Reglas de equivalencia**

- Especifican el modo de transformar una expresión en otra equivalente lógicamente.

- **Información del catálogo**

- $n_r$ , el número de tuplas de la relación  $r$ .
- $b_r$ , el número de bloques que contienen tuplas de la relación  $r$ .
- $t_r$ , el tamaño de cada tupla de la relación  $r$  en bytes.
- $f_r$ , el factor de bloqueo de la relación  $r$ , es decir, el número de tuplas de la relación  $r$  que caben en un bloque.
- $V(A, r)$ , el número de valores distintos que aparecen en la relación  $r$  para el atributo  $A$ .

# Optimización de Consultas

- **Histograma:**
  - En los histogramas los valores del atributo se dividen en una serie de rangos, y con cada rango el histograma asocia el número de tuplas cuyo valor del atributo se halla en ese rango.
- **Estimación del tamaño de la selección**
  - Se considerará un solo predicado de igualdad,
  - Luego un solo predicado de comparación y
  - Finalmente, combinaciones de predicados.

# Optimización de Consultas

- **Selecciones complejas:**
  - **Conjunción:** Una *selección conjuntiva* es una selección de la forma
$$\sigma_{\theta_1 \wedge \theta_2 \wedge \dots \wedge \theta_n}(r)$$
  - **Disyunción:** Una *selección disyuntiva* es una selección de la forma
$$\sigma_{\theta_1 \vee \theta_2 \vee \dots \vee \theta_n}(r)$$
  - **Negación:** En ausencia de valores nulos el resultado de una selección  $\sigma_{\theta}(r)$  es simplemente las tuplas de  $r$  que no están en  $\sigma_{\theta}(r)$ .

# Optimización de Consultas

- **Estimación del tamaño de otras Operaciones**

- **Proyección:** El tamaño estimado (número de registros de las tuplas) de una proyección de la forma  $\Pi A(r)$  es  $V(A, r)$ , ya que la proyección elimina los duplicados.
- **Agregación:** El tamaño de  $A \text{ GROUP BY } F(r)$  es simplemente  $V(A, r)$ , ya que hay una tupla de  $A \text{ GROUP BY } F(r)$  por cada valor distinto de  $A$ .
- **Operaciones de conjuntos:** Si las dos entradas de una operación de conjuntos son selecciones de la misma relación se puede reescribir la operación de conjuntos como disyunciones, conjunciones o negaciones.
- **Reunión externa:** El tamaño estimado de  $r \bowtie s$  es el tamaño de  $r \bowtie s$  más el tamaño de  $r$ ; el de  $r \ltimes s$  es simétrico, mientras que el de  $r \ltimes s$  es el tamaño de  $r \bowtie s$  más los tamaños de  $r$  y de  $s$ .

# Optimización de Consultas

- **Expresiones Equivalentes**
  - Si en cada ejemplar legal de la base de datos, las dos expresiones generan el mismo conjunto de tuplas.
  -
- **Ejemplar legal de la base de datos**
  - Satisface todas las restricciones de integridad especificadas en el esquema de la base de datos.
- **Reglas de equivalencia**
  - Las expresiones de dos formas son equivalentes.
- **Conjunto de reglas de equivalencia mínimo**
  - No se puede obtener ninguna regla a partir de una reunión de las demás.



# Optimización de Consultas

- **Optimizadores basados en el costo**
  - Generan una gama de planes de evaluación a partir de la consulta dada empleando las reglas de equivalencia y escogen el de costo mínimo.
- **Algoritmos de programación dinámica**
  - Almacenan los resultados de los cálculos y los reutilizan, un procedimiento que puede reducir enormemente el tiempo de ejecución.
- **Optimización heurística**
  - Reducir el número de elecciones que hay que hacer de una manera basada en los costos.
- **Regla heurística**
  - Llevar a cabo las operaciones de selección tan pronto como sea posible.
  - Llevar a cabo las proyecciones tan pronto como sea posible.
  - La fase de **selección del plan de acceso** del optimizador heurístico selecciona la estrategia más eficiente para cada operación.

# Optimización de Consultas

- **Optimización de las subconsultas anidadas**
  - **Variables de correlación**, parámetros o las variables de la consulta del nivel externo que se utilizan en la subconsulta anidada
  - **Evaluación correlacionada**, técnica para evaluar una consulta con una subconsulta anidada.
  - **Descorrelación**, proceso de sustituir una consulta anidada por una consulta con una reunión.

# Optimización de Consultas

- **Vista materializada**
  - Es una vista cuyo contenido se calcula y se almacena.
- **Mantenimiento de la vista**
  - Tarea de mantener actualizada una vista materializada con los datos subyacentes.
- **Mantenimiento incremental de la vista**
  - Modificar sólo las partes afectadas de la vista materializada.