



قواعد معطيات متقدمة

Query processing معالجة الاستعلام



مخطط الفصل

- الخطوات الأساسية في معالجة الاستعلام
- قياس كلفة الاستعلام
- تعلية الاختيار
- الفرز
- الدمج
- عمليات أخرى
- تقييم التعابير

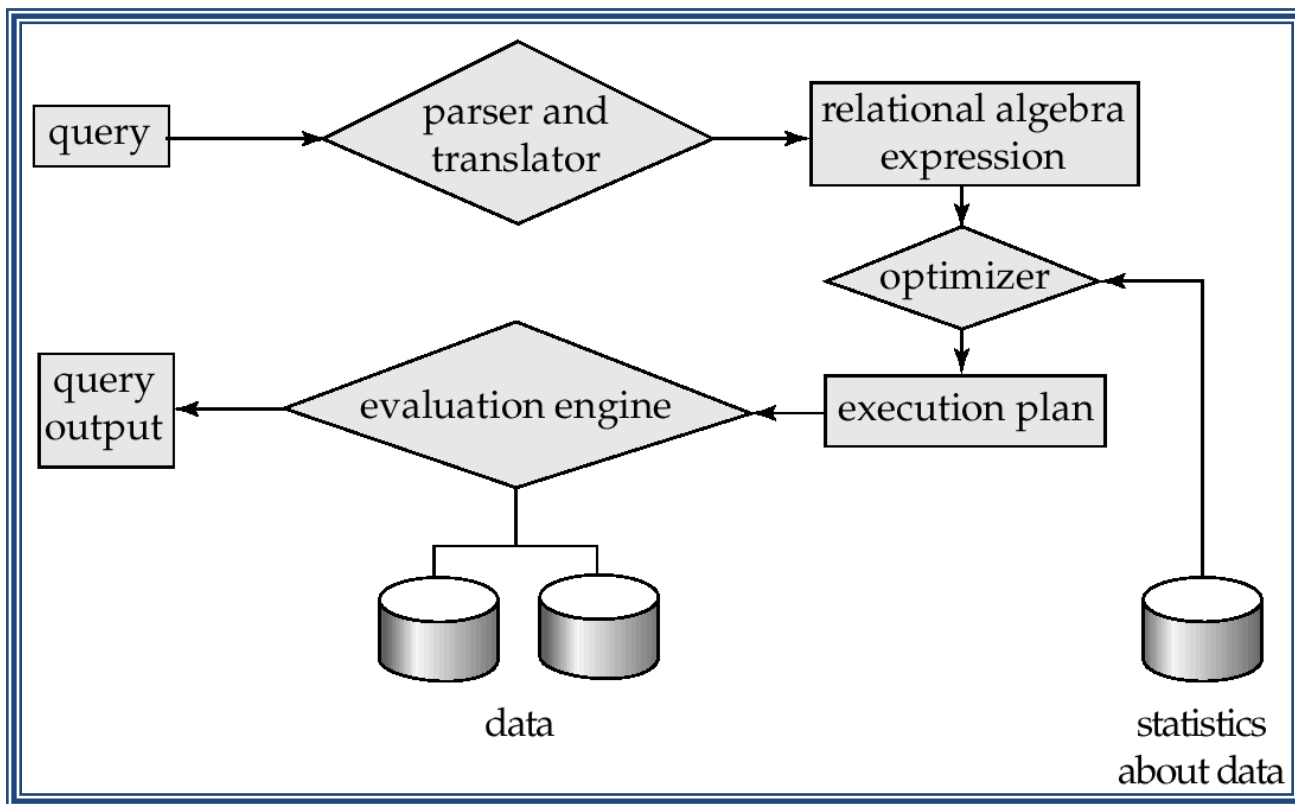


الخطوات الأساسية في معالجة الاستعلام

التفسير والترجمة Parsing and translation

الاختزال Optimization

التقييم Evaluation





الخطوات الأساسية في معالجة الاستعلام

□ التفسير والترجمة

- ترجمة الاستعلام إلى شكل داخلي، الذي بدوره يتم تحويله إلى استعلام مكتوب بلغة الجبر العلاقتي.
- يقوم المفسر parser بالتحقق من وجود العلاقات والتأكد من صحة الاستعلام قواعدياً

□ التقييم Evaluation

- يقوم محرك تنفيذ الاستعلام بتنفيذ مخطط تقييم الاستعلام query-evaluation plan، ومن ثم يقوم بإعادة الأجوبة لهذا الاستعلام



الخطوات الأساسية في معالجة الاستعلام

- يمكن إيجاد أكثر من تعبير جبري مكافئة للاستعلام المطروح

– مثال : لإيجاد مجموعة الحسابات التي رصيدها أقل من ٢٥٠٠

- Select account-number, balance from account where balance < 2500
- التعابير المكافئة بلغة الجبر العلاقتي هي :

- $\sigma_{balance < 2500}(\Pi_{balance}(account))$
- $\Pi_{balance}(\sigma_{balance < 2500}(account))$

- يجري تنفيذ الاستعلام (تقييمه) بعدة طرق، ندعو الطريقة المحددة (المختارة) لتنفيذ الاستعلام بمخطط التنفيذ evaluation plan

مثال : ١- استخدام فهرس على balance لإيجاد الحسابات التي رصيدها أقل من ٢٥٠٠ أو

٢- مسح العلاقة بشكل كامل وحذف الحسابات التي رصيدها أكبر أو يساوي ٢٥٠٠

- الاختزال Optimization : من بين التعابير المتكافئة، يجري اختيار التعبير الموافق لمخطط تنفيذ أقل كلفة

– حساب الكلفة المتوقعة باستخدام معلومات إحصائية من كاتالوج قاعدة المعطيات

- مثال : عدد الأسطر في كل علاقة، حجم السطر أو التسجيل، وغيرها ...

تقدير الكلفة - Cost Estimation

- تقاس عادة الكلفة بالزمن الكلي اللازم للإجابة عن الاستعلام
 - تدخل عدة عوامل في تقدير كلفة الزمن
 - الوصول إلى القرص، سرعة وحدة المعالجة المركزية CPU ، أو حدث الاتصال بالشبكة
- الوصول إلى القرص نسبياً سهل التقدير، ويقاس أخذين بالاعتبار عدة عوامل :
 - عدد مرات الانتقال في القرص * متوسط كلفة الانتقال في القرص
 - عدد الكتل المقروءة * متوسط كلفة قراءة كتلة
 - عدد الكتل المكتوبة * متوسط كلفة كتابة كتلة
- عادة كلفة كتابة كتلة أكبر من كلفة قراءة كتلة (نحتاج إلى إعادة قراءة الكتلة بعد كتابتها للتأكد من أن عملية الكتابة ناجحة).
- لتبسيط حساب الكلفة نستخدم فقط عدد الكتل المنقولة من وإلى القرص كقياس للكلفة
 - نهمل بذلك الفرق في الكلفة بين الدخل/الخرج التسلسلي والعشوائي للتبسيط
 - نهمل أيضاً كلفة المعالجة في CPU للتبسيط، في أنظمة الزمن الحقيقي يؤخذ أيضاً كلفة CPU بالحسبان

قياس الكلفة

- تعتمد الكلفة أيضاً على حجم الـ buffer في الذاكرة الرئيسية
 - وجود ذاكرة كبيرة يقلص الحاجة إلى الوصول إلى القرص
 - كمية الذاكرة المتاحة للـ buffer تعتمد على معالجات نظام التشغيل الحالية، ومن الصعب تحديدها قبل التنفيذ الفعلي.
 - نستخدم غالباً تقديرات أسوأ حالة مفترضين بأن كمية الذاكرة المتاحة لدينا هي فقط أقل كمية من الذاكرة التي نحتاجها لإنجاز العملية.
- لا نُضمن كلفة كتابة النتيجة على القرص

تعليمية الاختيار

- مسح الملف File scan - نستخدم خوارزميات البحث التي تقوم على تحديد وإيجاد مجموعة التسجيلات المحققة لشرط الاختيار.

- خوارزمية A1 (البحث الخطي linear search).

– الكلفة المتوقعة (عدد كتل القرص الممسوحة) $b_r =$

- b_r يرمز لعدد الكتل التي تحوي تسجيلات العلاقة r (حجم العلاقة)

– إذا كان البحث باستخدام المفتاح

- $cost = (b_r / 2)$

- التوقف عند إيجاد التسجيلة

– يمكن تطبيق البحث الخطي بغض النظر عن شرط الاختيار أو ترتيب التسجيلات في العلاقة، أو وجود الفهارس.

تعليمية الاختيار

• خوارزمية A2 (البحث الثنائي Binary Search). مطبق في حال كان الاختيار مقارنة على الواصفة المفروز بها الملف.

– بفرض أن كتل الملف المخزن فيه العلاقة موضوعة بشكل متتالي

– الكلفة المتوقعة (عدد الكتل التي يجري مسحها قبل الوصول إلى الإجابة)

• كلفة الوصول إلى أول سطر من العلاقة باستخدام البحث الثنائي على الكتل $\lceil \log_2(b_r) \rceil$

• بالإضافة إلى عدد الكتل الحاوية على تسجيلات محققة للشرط

– بفرض $SC(A, r)$ عدد التسجيلات التي تحقق شرط الاختيار، و f_r عدد التسجيلات في الكتلة

» عدد الكتل التي نحتاج إلى قراءتها هو $\lceil sc(A, r) / f_r \rceil$

$$= \lceil \log_2(b_r) \rceil + \lceil sc(A, r) / f_r \rceil - 1$$

✓ (-1) لأن الوصول لأول تسجيلية حسبناه باللوغار يتم

الاختيار بوجود الفهارس

- مسح الفهرس **Index scan** : خوارزميات البحث المستخدمة للفهارس
 - يجب أن يكون شرط البحث على مفتاح البحث للفهرس
- A3 (فهرس رئيسي على مفتاح مرشح، مساواة). إيجاد تسجيلة وحيدة تحقق شرط المساواة المطلوب
 - $Cost = HT_i + 1$ حيث HT_i هي كلفة الوصول إلى التسجيلة المعنية في الفهرس
 - ال 1 هو وصول واحد على **main table** لأن الفهرس رئيسي والمفتاح مرشح ..المفتاح المرشح: كل قيمة تقابلها تسجيلة واحدة فقط
- A4 (فهرس رئيسي علي واصفة ليست مفتاح، مساواة). إيجاد مجموعة من التسجيلات متواجدة في كتل متعاقبة
 - $Cost = HT_i + \text{number of blocks containing retrieved records}$
 - $Cost = HT_i + \lceil sc(A, r) / f_r \rceil$
- A5 (مساواة على مفتاح بحث لفهرس ثانوي)
 - تسجيلة واحدة إذا كان مفتاح البحث مفتاحاً مرشحاً
 - $Cost = HT_i + 1$
 - عدة تسجيلات إذا لم يكن مفتاح البحث مفتاحاً مرشحاً
 - $Cost = HT_i + \text{number of records retrieved} = HT_i + sc(A, r)$
 - يمكن أن تكون كل تسجيلة في كتلة مختلفة.