



데이터베이스

질의 최적화





학습목표

- ➔ 데이터베이스 관리 시스템에서 질의 처리 단계를 설명할 수 있다.
- ➔ 질의 최적화의 원리 및 방법을 설명할 수 있다.



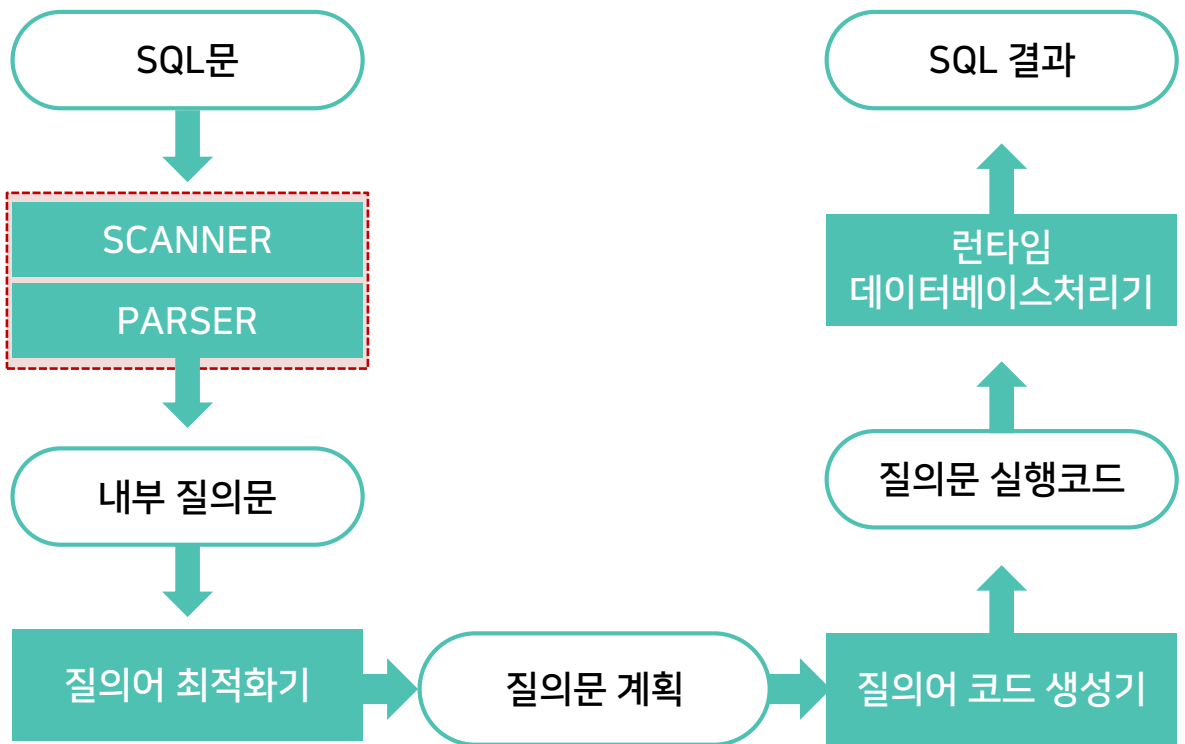
학습내용

- ➔ 질의 처리
- ➔ 질의 최적화기



질의어 처리과정

01 처리과정

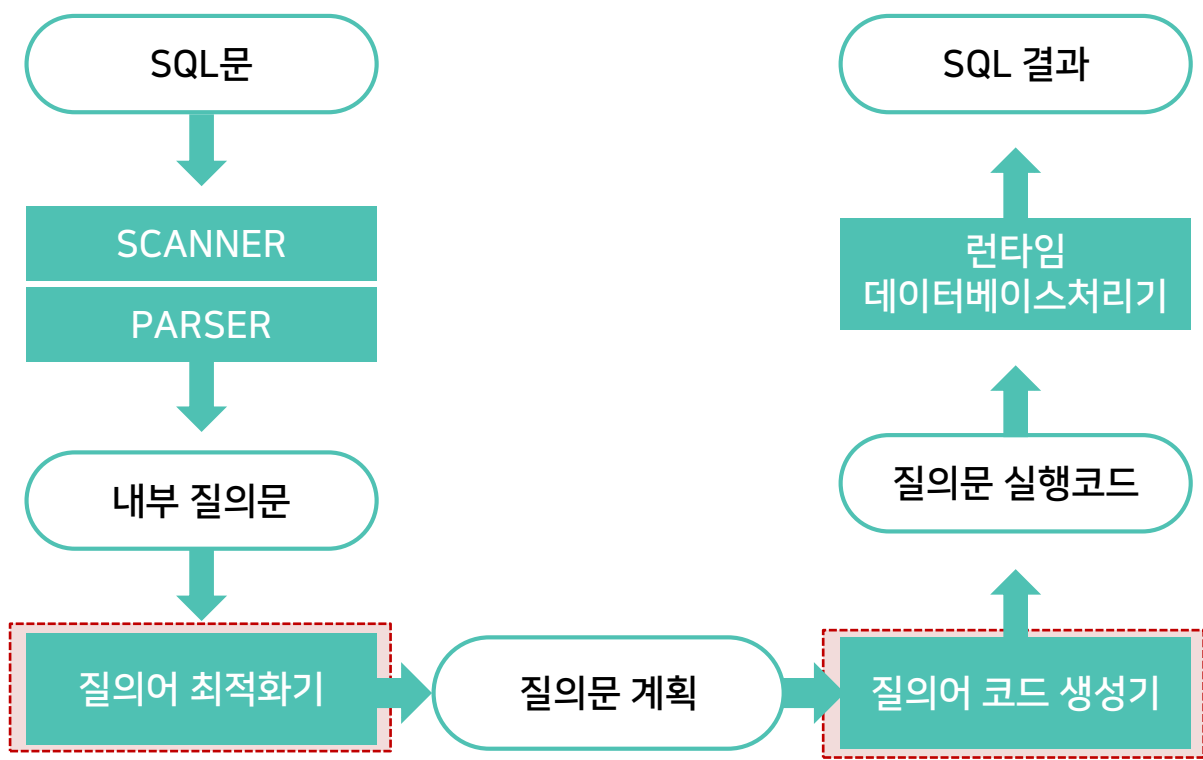


SCANNER (또는 Tokenizer)	• SQL 질의문을 문법적 단위(Token)들로 쪼개는 모듈
PARSER(구문분석기)	• 질의문이 문법에 맞는지 검사하고 내부 질의문으로 변환하는 모듈



질의어 처리과정

01 처리과정



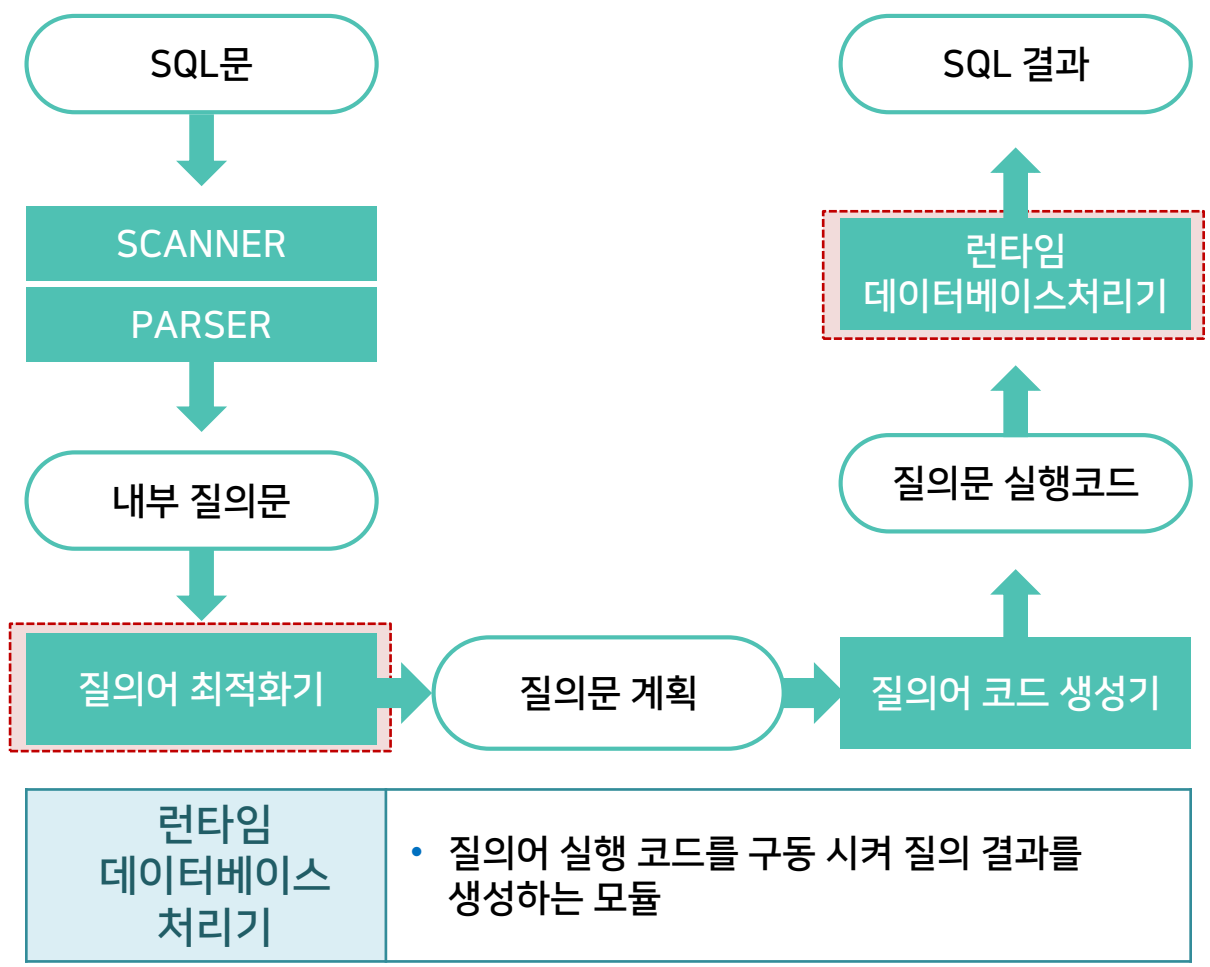
질의어 최적화기	<ul style="list-style-type: none">내부 질의문에 대한 가장 효과적인 실행 계획을 찾는 모듈
질의어 코드 생성기	<ul style="list-style-type: none">생성된 실행 계획에 따른 질의어 실행 코드 생성 모듈



질의어 처리과정

01

처리과정





질의어 처리과정

02

내부표현과 형식론

내부표현 ·>

사용자의 질의문을 컴퓨터가 처리하기 적절한 형태로 표현되어야 함

형식론 ·>

- 질의문 문법으로 표현할 수 있는 모든 것이 내부 표현으로 표현 가능해야 함
- 최적화 단계에 영향을 미치지 않는 독립적인 표현 방법이어야 함 → **관계 대수(Relational Algebra)**

03

질의문 트리

- **구문분석기의 결과로 생성되는 자료 구조**
- **관계대수식을 트리 형태로 사용**

트리의
단말 노드

피연산자인 릴레이션

트리의
비단말 노드

관계 대수 연산자



질의어 처리과정



04 질의문 트리의 실행

- 필요한 피연산자가 모두 사용 가능한 서브 트리에 대하여 먼저 **비단말 노드 연산자를 실행**



그 결과 릴레이션을 실행된 서브 트리로 대체

루트 노드의 실행 → **질의문의 실행 결과**



질의어 처리과정

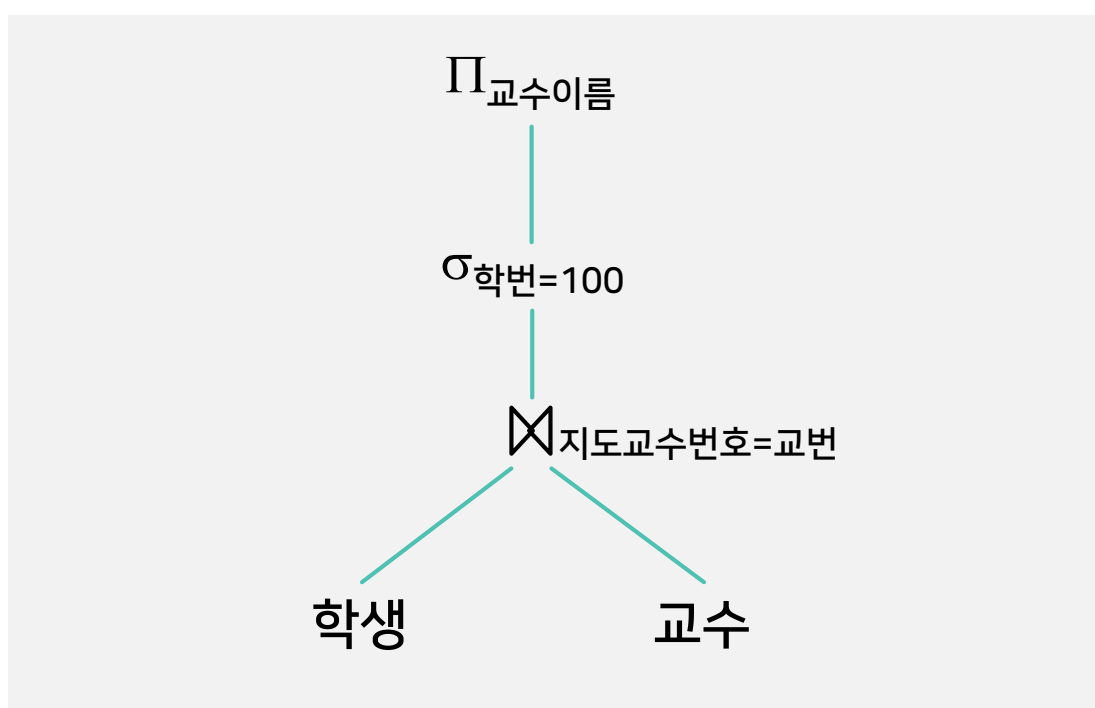
05

질의문 트리의 예

예 | 학번이 100번인 학생의 지도 교수명을 검색하시오.

- 관계대수식

$$\Pi_{\text{교수이름}} (\sigma_{\text{학번}=100} (\text{학생} \bowtie_{\text{지도교수번호}=\text{교번}} \text{교수}))$$





단순 질의문의 변환

01

단일 테이블 검색에 대한 질의문 변환 방법

SELECT절	FROM절	WHERE절
프로젝션 연산자(Π)	대상 테이블	선택 연산자(σ)
<p>SELECT절</p> <p>▶ 학생 테이블에서 학번이 100번인 학생의 이름을 구하시오.</p>	<p>SQL문</p> <p>▶ SELECT SNAME</p> <p>▶ FROM절 STUDENT</p> <p>▶ WHERE SNUM = 100</p>	<p>관계대수식</p> <p>$\Pi_{SNAME}(\sigma_{SNAME=100}(STUDENT))$</p>



조인 질의문의 변환

01 FROM절에 대상 테이블이 2개 이상 존재하는 경우

기본적으로는 단순 질의문 변환과 동일

- 차이점은 FROM절에 나타나는 2개 이상의 테이블을 카티션 프로덕트로 연결

SELECT절

프로젝션 연산자(Π)

FROM절

대상 테이블들의
카티션 프로덕트(\times)

WHERE절

셀렉션 연산자(σ)

02 조인 질의문 변환의 예

질의문

- 학번이 100번인 학생의 지도교수 명을 검색하시오.

SQL문

```
SELECT PNAME
FROM절 STUDENT, PROFESSOR
WHERE STUDENT.SNAM = 100 AND STUDENT.PNUM
= PROFESSOR.PNUM
```



조인 질의문의 변환

02

조인 질의문 변환의 예

관계대수식

$$\Pi_{PNAME}(\sigma_{STUDENT.SNAM=100 \wedge STUDENT.PNUM=PROFESSOR.PNUM}(STUDENT \bowtie PROFESSOR))$$

- ▶ 선택 연산자에 AND(\wedge) 연산자가 있으면 2개의 선택 연산자로 분리할 수 있음

$$\Pi_{PNAME}(\sigma_{STUDENT.SNAM=100}(\sigma_{STUDENT.PNUM=PROFESSOR.PNUM}(STUDENT \bowtie PROFESSOR)))$$

- ▶ 선택 연산자의 조건이 두 개의 테이블간의 조건을 기술하는 것이면 이는 조인 조건임
- ▶ 조인조건을 포함하는 선택 연산자와 카티션 프로덕트는 조인 연산자 하나로 표현할 수 있음

$$\Pi_{PNAME}(\sigma_{STUDENT.SNAM=100}(STUDENT \bowtie_{STUDENT.PNUM=PROFESSOR.PNUM} PROFESSOR)))$$



질의 최적화의 개념

01 개념

01 질의어의 효율적인 실행 전략 계획

- ▶ 질의문을 실행할 수 있는 후보 질의문 계획을 평가하고 그 중에서 **최소 비용의 계획을 결정**하는 것

* **최적화란?** : 더 효율적인 실행 방안을 생성하는 것
→ 어떤 것을 더 효율적인 것이라고 하는가?

02 질의문 하나에는 보통 여러 개의 후보 질의문 계획이 존재

- ▶ 제한된 수의 질의문 계획을 생성
- ▶ 탐색 공간의 축소
- ▶ 가장 비용이 적게 드는 계획을 선택

02 성능

DISK I/O의 횟수	일반적인 기준으로 • 메모리: ns 단위 • DISK: ms 단위 } 최소 1000배 느림 → CPU를 더 쓰더라도 DISK I/O를 줄이는 것이 DB 분야에서는 중요한 이슈임
중간 결과의 크기	• 분산 DB와 같이 여러 개의 DB들로 구성되어 있는 상황 • 중간 결과의 크기가 크면 네트워크의 사용량이 많아지고 그만큼 느려 짐
처리시간	• 결과물 모두를 사용자에게 알려주는데 걸리는 시간
응답시간	• 사용자에게 결과물(일부분이라도)을 보여주는데 걸리는 시간

대화형 시스템의 경우 사용자에게 빨리 결과를 보여주는 것이 중요함



질의 최적화기



질의 최적화의 개념

03 최적화 단계

01 질의문의 내부적 표현을 생성

02 동등한 내부적 표현들을 생성

03 후보 프로시저를 선정

04 질의문 계획의 평가 및 결정

04 최적화의 예

» 예 | 셀렉션 연산은 가능한 빨리하자.

» 예 | 프로젝션 연산은 가능한 빨리하자.

$$\sigma_{Sc}(R \bowtie S) \rightarrow R \bowtie (\sigma_{Sc}(S))$$

05 최적화의 구분



경험론적 최적화

비용 기반 최적화



질의 최적화의 개념

06

비용 기반 최적화의 예

```

SELECT SNAME
FROM S, E
WHERE S.SNO=E.SNO AND E.CNO='C413'

```

통계 정보

$$|S|=100, |E|=10000, |E.CNO='C413'|=50$$

- ▶ E.SNO는 S.SNO의 외래키이고, Null값일 수 없음
→ 실행전략

- A: $\Pi_{SNAME}(\sigma_{E.CNO='c413'}(E \bowtie_{S.SNO=E.SNO} S))$
- B: $\Pi_{SNAME}((\sigma_{E.CNO='c413'}(E)) \bowtie_{S.SNO=E.SNO} S)$



질의 최적화의 개념

06

비용 기반 최적화의 예

통계 정보

$|S|=100, |E|=10000, |E.CNO='C413'|=50$

비용 A

- $\Pi_{SNAME}(\sigma_{E.CNO='C413'}(\bowtie_{S.SNO=E.SNO} S))$
- $R1=(\bowtie_{S.SNO=E.SNO} S): 100*10000+10000$
- $R2=\sigma_{E.CNO='C413'}(R1): 10000+50$
- $R3=\Pi_{SNAME}(R2): 50$
- 총: 1020100번 I/O

비용 B

- $\Pi_{SNAME}(\sigma_{E.CNO='C413'}(\bowtie_{S.SNO=E.SNO} S))$
- $R1=\sigma_{E.CNO='C413'}(E): 10000+50$
- $R2=R1 \bowtie_{S.SNO=E.SNO} S: 50*100+50$
- $R3=\Pi_{SNAME}(R2): 50$
- 총: 15150번 I/O



변환규칙

01

변환규칙

- 질의 내부 표현을 동등하면서도 효율적인 형태로 변환
- 문법 또는 의미를 가지고 생성

R1 논리곱으로 연결된 선택 조건

- 일련의 개별적인 선택 조건

$$\sigma_{c1 \text{ AND } c2 \text{ AND } \dots \text{ AND } c_n}(R) \equiv \sigma_{c1}(\sigma_{c2}(\dots(\sigma_{c_n}(R))\dots))$$

R2 교환적인 선택 연산

$$\sigma_{c1}(\sigma_{c2}(R)) \equiv \sigma_{c2}(\sigma_{c1}(R))$$

R3 연속적인 프로젝트 연산(Π)

- 마지막 것만 실행

$$\Pi_1(\Pi_2(\dots(\Pi_n(R))\dots)) \equiv \Pi_1(R)$$

R4 셀렉트의 조건 c가 프로젝트 애트리뷰트만 포함하고 있다면 교환적임

$$\sigma_c(\Pi(R)) \equiv (\Pi(\sigma_c(R)))$$



변환규칙

01

변환규칙

R5

선택의 조건이 두 개의 릴레이션이 연결되어 있다면
조인조건

$$\sigma_c(R \times S) \equiv R \bowtie_c S$$

$$\sigma_{c_1}(R \bowtie_{c_2} S) \equiv R \bowtie_{c_1 \wedge c_2} S$$

R6

선택의 조건이 조인 또는 카티션 프로덕트에 관련된
릴레이션 하나와만 관련이 되어있을 때

$$\sigma_c(R \bowtie S) \equiv \sigma_c(R) \bowtie S$$

$$\sigma_c(R \times S) \equiv \sigma_c(R) \times S$$

R7

c1은 릴레이션 R과 관련되어 있고, c2는 릴레이션 S와
관련이 되어 있을 때 c=(c1 AND c2)

$$\sigma_c(R \bowtie S) \equiv (\sigma_{c_1}(R)) \bowtie (\sigma_{c_2}(S))$$

$$\sigma_c(R \times S) \equiv (\sigma_{c_1}(R)) \times (\sigma_{c_2}(S))$$



질의 최적화기



변환규칙



01 변환규칙

R8 $\times, \cup, \cap, \bowtie$ 는 교환적임

$$R \times S \equiv S \times R$$

$$R \cup S \equiv S \cup R$$

$$R \cap S \equiv S \cap R$$

$$R \bowtie S \equiv S \bowtie R$$

R9

L1은 릴레이션 R에 관련되어있고,
L2는 릴레이션 S에 관련되어 있을 때 $L = (L1, L2)$

$$\Pi_L(R \bowtie S) \equiv (\Pi_{L1}(R)) \bowtie (\Pi_{L2}(S))$$

$$\Pi_L(R \times S) \equiv (\Pi_{L1}(R)) \times (\Pi_{L2}(S))$$



변환규칙



01 변환규칙

R10 집합연산과 관련된 선택의 변환

$$\sigma_c(R \cup S) \equiv \sigma_c(R) \cup \sigma_c(S)$$

$$\sigma_c(R \cap S) \equiv \sigma_c(R) \cap \sigma_c(S)$$

$$\sigma_c(R - S) \equiv \sigma_c(R) - \sigma_c(S)$$

R11 합집합과 관련된 프로젝트의 변환

$$\Pi(R \cup S) \equiv (\Pi(R)) \cup (\Pi(S))$$



질의 최적화기



의미론적 최적화

01

의미론적 최적화

- 질의문의 의미와 데이터의 범위, 제약조건 등을 이용하여 최적화 진행
- 비용기반 최적화보다 더 효율적임
- 실제 발생될 확률은 매우 적음

02

예시

질의

5학년 이상 학생들의 이름을 찾아보시오.



도메인 무결성 제약조건

학년은 1~4까지의 값만 가짐



결과

결과가 없음을 쉽게 알 수 있음



질의 최적화기



의미론적 최적화

01

의미론적 최적화

I 질의

$$\Pi_{CNO}(S \bowtie_{S.SNO=E.SNO} E)$$

- 제약조건 SNO
 - ▶ S의 기본 인덱스
 - ▶ E의 외래키
 - ▶ NULL이 아님CNO(E)



$$\Pi_{CNO}(E)$$



1 질의 처리

- ✓ 질의어를 내부적 표현으로 변환하고 이를 수행시켜 질의 결과를 얻는 일련의 절차
- ✓ SCANNER → PASER → 질의 최적화기 → 질의어 코드 생성기 → 런타임 데이터베이스 처리기
- ✓ 내부적 표현으로 관계대수식을 트리 형태로 표현

2 질의 최적화기

- ✓ 질의문을 실행할 수 있는 후보 질의문 계획을 평가하고 그 중에서 최상의 비용 계획을 결정하는 것