Leo Forster
Maximilian Prinz
Bastian Simon

1. Cost-based Optimisation

1.1 (a)

1. Dataset:

Costs $B, B.s = 12$:

- scan: $300,000$
- index: $252,018.19$

=> use index

Costs $C.z \neq 17$:

- scan: $50,000$
- index: $21,015.61$

=> use index

Costs $C.c < 123$:

- scan: $50,000$
- index: $105.015,61$

=> use scan

2. Dataset:

Costs $B, B.s = 12$:

- scan: $70,000$
- index: $117,616.10$

=> use scan

Costs $C.z \neq 17$:

- scan: $150,000$
- index: $94,517.19$

=> use index

Costs $C.c < 123$:

- scan: $150,000$
- index: $472.517,19$

=> use scan

Thus, for dataset 1:

- $\sigma_{B.s=12}$ should be replaced with $\sigma\_IndexBased_{B.s=12}$
- $\sigma_{C.c<123}$ should be replaced with $\sigma\_ScanBased_{C.c<123}$
- $\sigma_{C.z\neq 17}$ should be replaced with $\sigma\_IndexBased_{C.z\neq 17}$

For dataset 2:

- $\sigma_{B.s=12}$ should be replaced with $\sigma\_ScanBased_{B.s=12}$
- $\sigma_{C.c<123}$ should be replaced with $\sigma\_ScanBased_{C.c<123}$
- $\sigma_{C.z\neq 17}$ should be replaced with $\sigma\_IndexBased_{C.z\neq 17}$

1.2 (b)

Given the rule, to use the index if: $scan(T) > index(T, p)$:

$$|T| > \log_2(|T|) + 42 \cdot \text{sel}(p) \cdot |T|$$

Solving this for $\text{sel}(p)$ we get:

$$\text{sel}(p) < \frac{|T| - \log_2(|T|)}{|T| \cdot 42}$$

Leo Forster
Maximilian Prinz
Bastian Simon

With large table sizes this approximates:

$$\mathrm{sel}(p) \approx \frac{1}{42} \approx 0.024$$