Homework 3

Simon Zheng
260744353

April 5th, 2018

1

1.1 a)

1.1.1 i)

- In general: 3000 (page) accesses (have to search everywhere)
- Concrete values: 3000 (page) accesses (same reason)

1.1.2 ii)

- In general: 3000 (page) accesses (unclustered index for range search not very useful)
- Concrete values: 3000 (similar reason)

1.1.3 iii)

- In general:
- Concrete values: 10 entries + 1 leaf page

1.1.4 iv)

- In general: 10+1 (using X index as it is more useful)
- Concrete values: Same thing

1.2 b)

Yes, especially on price, as it is very useful for range look-ups. It allows us to make less I/O since we can use the assumption that the range is in order on the pages.

2

2.1 a)

$$\begin{aligned} OuterPages \ + \ & |Outer| \times (\text{cost of finding matching tuples in inner relation}) \\ & = 3,000 + 200,000 \times (1 \ data \ page + 1 \ leaf \ page)) = 403,000 \end{aligned}$$

where $Inner$ is $Parts$ since it has the primary key index.

2.2 b)

$$\begin{aligned} OuterPages \ + \ & \frac{OuterPages}{|bp_{Outer}|} \times (InnerPages) \\ & = 1,500 + \frac{1,500}{51-1} \times 3,000 = 91,500 \end{aligned}$$

where $Outer$ is $Parts$ and $Inner$ is $Catalog$ (as specified).

2.3 c)

We first calculate the number of passes for each and then add up the costs of sorting each and merging:

$$1 + \lceil \log_{50-1}(\frac{1,500}{50}) \rceil = 2$$
$$1 + \lceil \log_{50-1}(\frac{4,000}{50}) \rceil = 3$$
$$3,000 \times 3 + 1,500 \times 2 + 3,000 + 1,500 = 16,500$$

2.4

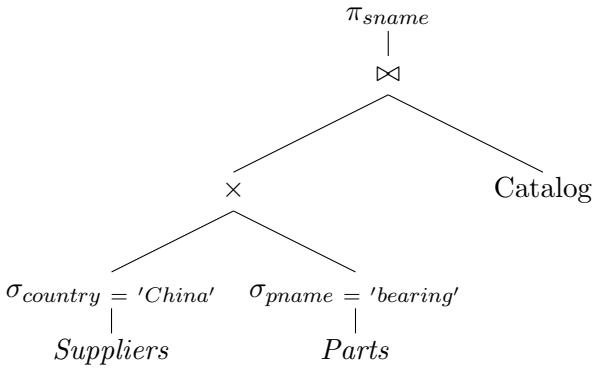
Number of output tuples: $|Catalog|$ for all of them.

3

3.1 a)

$\pi_{sname} \left(\left(\sigma_{country \ = \ 'China'}(Suppliers) \times \sigma_{pname \ = \ 'bearing'}(Parts) \right) \bowtie Catalog \right)$
We could also use projection after each selection to only project the column needed to join on and the column used in the final projection (pid and $sname$).

3.2 b)


π_{sname}
$\bowtie$
$\times$
Catalog
$\sigma_{country\ =\ 'China'}$
$\sigma_{pname\ =\ 'bearing'}$
Suppliers
Parts