ICCS240: Assignment 2

Kriangsak Thuiprakhon
kriangsak.thi@student.mahidol.edu
Collaborator: Prof and 1409 ppl

2: Relational Database Design Theory

---

(1) Alice *vs* Bob Given

```
Contracts( c_no, supp_no, proj_no, dept_no, part_no, qty, val)
```

Let $\mathcal{F}_A$ and $\mathcal{F}_B$ denote the set of functional dependencies of Alice and Bob respectively. Here, for simplicity, we rewrite the relation to be

```
Contracts( C, S Pr, D, Pa, Q, V)
```

then.

$$
\begin{aligned}
\mathcal{F}_A = \{ \quad & C \to SPrDPaQV, & PrPa \to C, \quad & SD \to Pa\} \\
\mathcal{F}_B = \{ \quad & C \to SPrD, & PrPa \to C, \quad & SDPr \to CPaQV\}
\end{aligned}
$$

As per Bob's schema, it was designed such that for Pa to be determined, it needs $SDPr$ altogether to do so, meanwhile, in Alice's, only SD can determine Pa. Therefore, the two designs are different.

(2) Given three disjoint ordered sets of attributes α, β, γ, consider relations $R = (\alpha, \beta, \gamma)$ and its decomposition: $R_1 = (\alpha, \beta)$and $R_2 = (\alpha, \gamma)$.That is, $R_1 = \Pi_{R_1}(R)$ and $R_2 = \Pi_{R_2}(R)$. Assume $R_1 \cap R_2 \to R_1$, that is ($\alpha \to \alpha\beta$). Prove/disprove that

Claim 0.1.

$$R = R_1 \bowtie R_2$$

Proof. Let's take a look at RHS, as discussed in class that Natural join is a derived operation that can be expressed as:

Let

- Each of the attribute sets have m elements
- A denote the set of attributes in R_1
- B denote the set of attributes in R_2

$$
\begin{aligned}
R_1 \bowtie R_2 &= \Pi_{A \cup B}(\sigma_{\alpha=\varepsilon}(\rho_{\alpha \to \varepsilon}(R_1) \times R_2))) && (1) \\
&= \Pi_{A \cup B}(\sigma_{\alpha=\varepsilon}(\{(\alpha_i, \beta_i, \varepsilon_j, \gamma_j) \quad \forall i, j = \{1, 2, 3, \cdots m\}\})) && (2) \\
&= \Pi_{A \cup B}(\{(\alpha_i, \beta_i, \varepsilon_j, \gamma_j) \quad | \quad \alpha_i = \varepsilon_j)\}) && (3) \\
&= \Pi_{A \cup B}(F(\alpha^*, \beta^*, \varepsilon^*, \gamma^*)) \qquad \text{For some relation } F && (4) \\
&= \Pi_{A \cup B}(F(\alpha^*, \beta^*, \varepsilon^*, \gamma^*)) && (5) \\
&= F(\alpha^*, \beta^*, \gamma^*) && (6) \\
& && (7)
\end{aligned}
$$

From here on, we now need to show that the relations R and F are equal.

Claim 0.2. *Relation R is equal to resulting relation F in (7). That is,*

$$R(\alpha, \beta, \gamma) = F(\alpha^*, \beta^*, \gamma^*) \Leftrightarrow (\alpha, \beta, \gamma) = (\alpha^*, \beta^*, \gamma^*)$$

Proof. From above, line (3), it is clear that each remaining pair of tuples was compared only by the value of α, such that $\alpha_i = \varepsilon_j$. This will happen if and only if $i = j$. This implies that $(\alpha, \beta, \gamma) = (\alpha^*, \beta^*, \gamma^*)$ □

From *claim 0.2*, it is shown that relations$R_1 \bowtie R_2 = R$. Hence proved. □

(3) Consider relation $R(A, B, C, D)$,with $\mathcal{F} = \{AB \to C, \quad BC \to D, \quad CD \to A\}$

- Is R in BCNF?

  To answer this question, we first need to identify the candidate keys. The candidate keys in this relation are AB, BC since $AB^+, BC^+ = \{A, B, C, D\}$ Now we can tell that this relation is not in $BCNF$ since it is *not* the case that *every* determinant is a superkey.

- Decompose the $R(A, B, C, D)$ into $R_1(B, C, D)$ and $R_2(C, D, A)$. Let us consider each decomposed relations at a time.

  (i) For $R_1(B, C, D)$ We can now determine the new set of functional dependency. That is, we will only take the subset of $\mathcal{F}$ that R_1 is relevant to. Then,

  $$\mathcal{F}_{R_1} = \{BC \to D\}$$

  (ii) For $R_2(C, D, A)$

  $$\mathcal{F}_{R_2} = \{CD \to A\}$$

  Now, the two relations are of BCNF since each of them satisfies the condition: *every* determinant is a superkey.

- The decomposition is lossless if at least one of the following functional dependencies are in $\mathcal{F}^+$ (closure of every attribute or attribute sets in $\mathcal{F}$).

  (i) $R_1 \cap R_2 \to R_1$. Apply our relations,

  $$R_1 \cap R_2 = AC \to ABC$$

  (ii) $R_1 \cap R_2 \to R_2$Apply our relations,

  $$R_1 \cap R_2 = AC \to ACD$$

  From the given $\mathcal{F}$, the closure of attributes are as follows:

  $$AB^+ = \{A, B, C, D\}$$
  $$BC^+ = \{A, B, C, D\}$$
  $$CD^+ = \{C, D, A\}$$

  We can now see that there is neither $AC \to ABC$ nor $AC \to ACD$ belongs to $\mathcal{F}$, hence, the decomposition is not lossless.

- Relation R is of 3NF if at least one of the following requirements is met:
  For any $X \to Z$ in $\mathcal{F}$

  (i) X is *not a superkey*

  (ii) A is not *prime*

  Since $CD \to A$ only violates the *(ii)*, the relation R is of **3NF**

3: Storage and Indexing

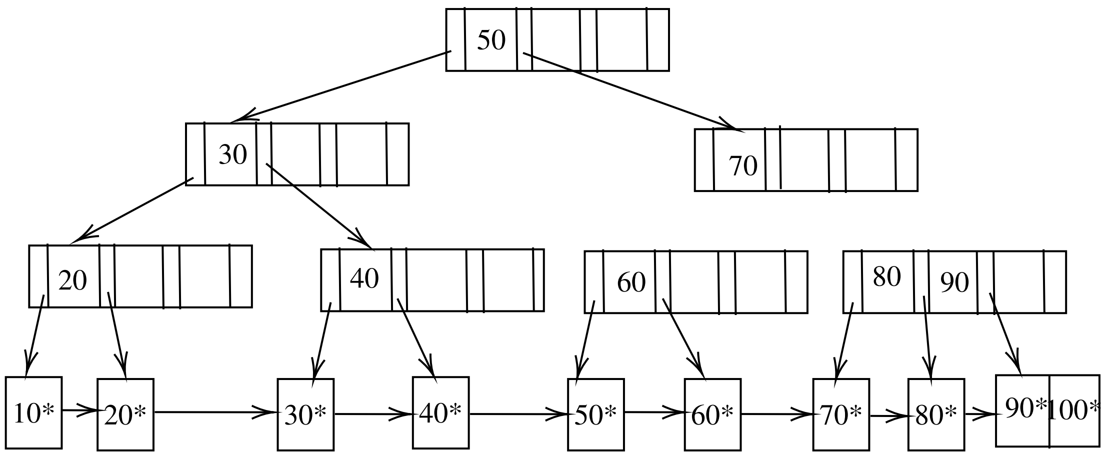
---

(1) which would most likely require the fewest I/Os for processing the query.

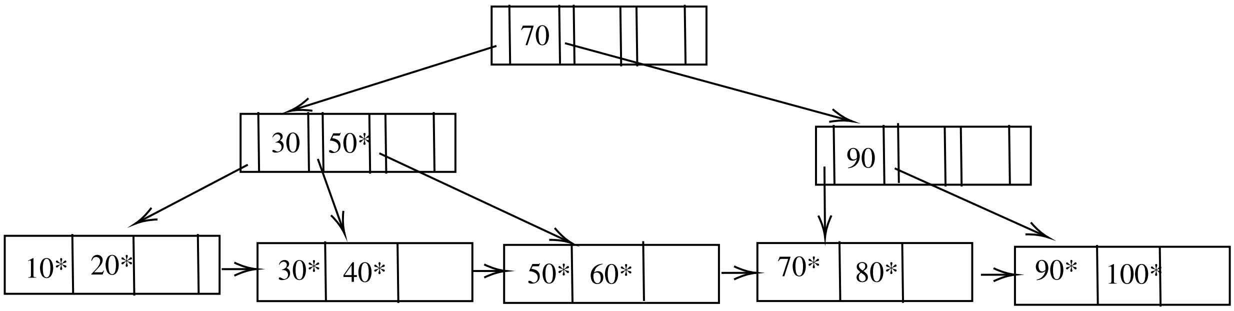
- scanning through the whole heap file for R since what is needed is the tuple of the relation R, doing so will take $O(n)$.
- Use B+ tree, this is done in the similar fashion of Binary Search. From Data Structure, BST, takes at most $O(\log_{\frac{k}{2}} n)$ *w.h.p.*
- Use Hash Index, this takes amortized $O(1)$.

(2)

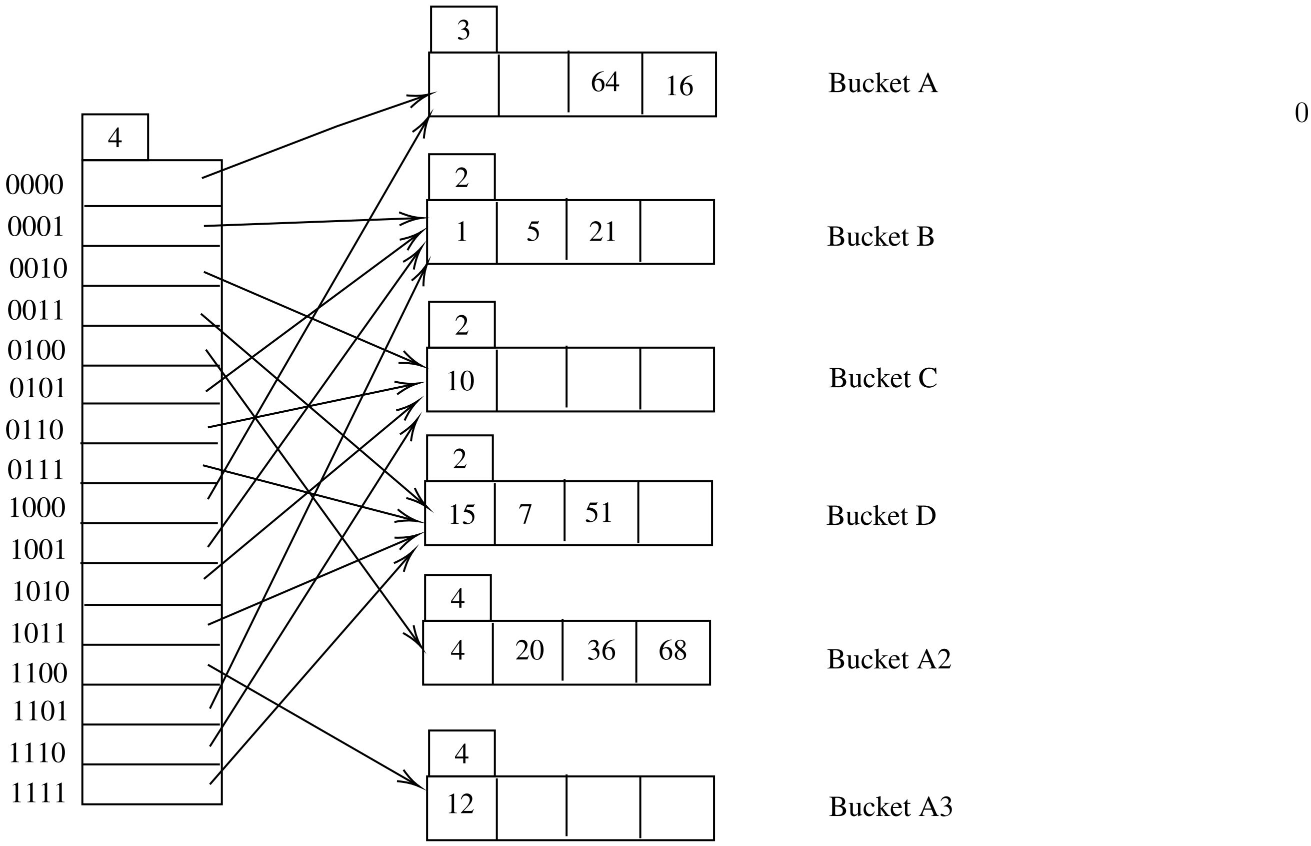
- B+ tree with this exact order of insertion: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100. The tree will be built based on the following property:
  - oder(k) = 3
  - pointer(k+1) = 4
  - minKey (k/2) = 3/2 = 1:
  - maxKey (k-1) = 2



- reconstruct a B+ tree in descending order of insertion. Notice that the height decreases down to 3: it was 4 in the previous tree.



(3)



(a)

(b) The minimal set of deletion to trigger a merge is $\{64, 16\}$. That is, when this set is deleted, there can be a merge between bucket A and A2. One might think that the answer should be $\{10\}$, but, in this case, the bucket C is already a primary page and it has to be left as a place holder.

Exam Revisit:

---

(1) Set-model will not give the right answer in querying for the average score of a class. This is because set model will contain no duplicates and the exam scores are not guaranteed to be unique. Therefore, Bag Model it is!

(2)

```
SELECT DISTINCT maker FROM Computer WHERE model in(
SELECT model FROM PC WHERE speed in (
SELECT speed FROM PC ORDER BY  speed  DESC LIMIT 1);
```

(3) given $R(A, B, C, D)$ and $\mathcal{F} = \{AC \to B, B \to A, BD \to C, D \to A\}$ Find the candidate keys.

- $BD \quad \because BD^+ = \{A, B, C, D\}$
- $CD \quad \because CD^+ = \{A, B, C, D\}$

(4) Basis and Minimal basis

- The right hand side of the dependency cannot contain more than one attribute to be a ***minimal basis*** .
- $\mathcal{G}$ is a ***basis*** of $\mathcal{F}$ and vice versa. This is because the two sets of dependencies are equivalent.
- $\mathcal{M} = \{A \to B\ A \to C\}$ is a minimal basis of $\mathcal{F}$ since it meets the following three criteria:
  - (i) All FDs in $\mathcal{M}$ have singleton right sides.
  - (ii) Any FD in $\mathcal{M}$ is removed, then the result is no longer a basis
  - (iii) For any FD in $\mathcal{M}$ we remove one or more attributes from the left side of F, the result is no longer a basis.