database tables

| **employees** (*E* rows) | |
|---|---|
| name | empID |
| Goldberg | 352 |
| Le | 90 |
| Abdulhamid | 229 |
| Zawad | 61 |
| Trinh | 251 |
| Wong | 230 |
| Elguera | 302 |
| Sorkin | 273 |
| Cheng | 312 |
| Chu | 494 |
| Martin | 71 |
| Lee | 284 |
| Nguyen | 191 |
| Ortiz | 360 |
| Hoang | 263 |
| Dao | 337 |

| **projects** (*P* rows) | |
|---|---|
| projID | empID |
| 1 | 61 |
| 2 | 251 |
| 3 | 352 |
| 4 | 61 |
| 5 | 229 |
| 6 | 230 |
| 7 | 61 |
| 8 | 230 |
| 9 | 61 |
| 10 | 71 |
| 11 | 251 |
| 12 | 191 |
| 13 | 71 |
| 14 | 352 |
| 15 | 284 |
| 16 | 284 |
| 17 | 312 |
| 18 | 229 |
| 19 | 494 |
| 20 | 337 |
| 21 | 494 |
| 22 | 273 |
| 23 | 191 |
| 24 | 337 |
| 25 | 302 |

desired result

| **projects NATURAL JOIN employees** | | |
|---|---|---|
| projID | empID | name |
| 1 | 61 | Zawad |
| 2 | 251 | Trinh |
| 3 | 352 | Goldberg |
| 4 | 61 | Zawad |
| 5 | 229 | Abdulhamid |
| 6 | 230 | Wong |
| 7 | 61 | Zawad |
| 8 | 230 | Wong |
| 9 | 61 | Zawad |
| 10 | 71 | Martin |
| 11 | 251 | Trinh |
| 12 | 191 | Nguyen |
| 13 | 71 | Martin |
| 14 | 352 | Goldberg |
| 15 | 284 | Lee |
| 16 | 284 | Lee |
| 17 | 312 | Cheng |
| 18 | 229 | Abdulhamid |
| 19 | 494 | Chu |
| 20 | 337 | Dao |
| 21 | 494 | Chu |
| 22 | 273 | Sorkin |
| 23 | 191 | Nguyen |
| 24 | 337 | Dao |
| 25 | 302 | Elguera |

three algorithms for performing the join:

## 1. nested loop join with 'projects' as outer table

loop over rows of 'projects', look up corresponding row of 'employees' for each row of 'projects':

| projID | name |
|---|---|
| 1 | Zawad |
| 2 | Trinh |
| 3 | Goldberg |
| 4 | Zawad |
| 5 | Abdulhamid |
| 6 | Wong |
| 7 | Zawad |
| 8 | Wong |
| 9 | Zawad |
| 10 | Martin |
| 11 | Trinh |
| 12 | Nguyen |
| 13 | Martin |
| 14 | Goldberg |
| 15 | Lee |
| 16 | Lee |
| 17 | Cheng |
| 18 | Abdulhamid |
| 19 | Chu |
| 20 | Dao |
| 21 | Chu |
| 22 | Sorkin |
| 23 | Nguyen |
| 24 | Dao |

| | |
|---|---|
| 25 | Elguera |

exercise 1: what is asymptotic complexity in terms of *P* and *E* ?

four scenarios, based on whether empID is indexed in…

| 'projects' | 'employees' | asymptotic complexity |
|---|---|---|
| Y | Y | O(?) |
| N | Y | O(?) |
| Y | N | O(?) |
| N | N | O(?) |

## 2. sort-merge join

(a) sort each table by joined column:

| **employees** | |
|---|---|
| name | empID |
| Zawad | 61 |
| Martin | 71 |
| Le | 90 |
| Nguyen | 191 |
| Abdulhamid | 229 |
| Wong | 230 |
| Trinh | 251 |
| Hoang | 263 |
| Sorkin | 273 |
| Lee | 284 |
| Elguera | 302 |
| Cheng | 312 |
| Dao | 337 |
| Goldberg | 352 |
| Ortiz | 360 |
| Chu | 494 |

| **projects** | |
|---|---|
| projID | empID |
| 9 | 61 |
| 4 | 61 |
| 7 | 61 |
| 1 | 61 |
| 10 | 71 |
| 13 | 71 |
| 23 | 191 |
| 12 | 191 |
| 18 | 229 |
| 5 | 229 |
| 6 | 230 |
| 8 | 230 |
| 11 | 251 |
| 2 | 251 |
| 22 | 273 |
| 15 | 284 |
| 16 | 284 |
| 25 | 302 |
| 17 | 312 |
| 20 | 337 |
| 24 | 337 |
| 14 | 352 |
| 3 | 352 |
| 19 | 494 |
| 21 | 494 |

(b) step through both tables, matching joined tuples:

| projID | empID |
|---|---|
| 9 | Zawad |
| 4 | Zawad |
| 7 | Zawad |
| 1 | Zawad |
| 10 | Martin |
| 13 | Martin |
| 23 | Nguyen |
| 12 | Nguyen |
| 18 | Abdulhamid |
| 5 | Abdulhamid |
| 6 | Wong |
| 8 | Wong |
| 11 | Trinh |
| 2 | Trinh |
| 22 | Sorkin |
| 15 | Lee |
| 16 | Lee |
| 25 | Elguera |
| 17 | Cheng |
| 20 | Dao |
| 24 | Dao |
| 14 | Goldberg |
| 3 | Goldberg |
| 19 | Chu |
| 21 | Chu |

exercise 2: what is asymptotic complexity in terms of *P* and *E* ?

part (a): O(?)

part (b): O(?)

total: O(?)

## 3. hash join

(a) build hash table of first table, with joined values as keys:

'projects' represented as a hash table

| empID | maps to | projID |
|---|---|---|
| 302 | → | 25 |
| 191 | → | 12, 23 |
| 337 | → | 20, 24 |
| 251 | → | 2, 11 |
| 284 | → | 15, 16 |
| 352 | → | 3, 14 |
| 312 | → | 17 |
| 229 | → | 5, 18 |
| 71 | → | 10, 13 |
| 273 | → | 22 |
| 61 | → | 1, 4, 7, 9 |
| 494 | → | 19, 21 |
| 230 | → | 6, 8 |

(b) (i) compute hashes of second table, (ii) creating matching tuples:

(i) look up hash value:

| name | empID | maps to | |
|---|---|---|---|
| Goldberg | 352 | → | 3, 14 |
| Le | 90 | → | null |
| Abdulhami | 229 | → | 5, 18 |
| Zawad | 61 | → | 1, 4, 7, 9 |
| Trinh | 251 | → | 2, 11 |
| Wong | 230 | → | 6, 8 |
| Elguera | 302 | → | 25 |
| Sorkin | 273 | → | 22 |
| Cheng | 312 | → | 17 |
| Chu | 494 | → | 19, 21 |
| Martin | 71 | → | 10, 13 |
| Lee | 284 | → | 15, 16 |
| Nguyen | 191 | → | 12, 23 |
| Ortiz | 360 | → | null |
| Hoang | 263 | → | null |
| Dao | 337 | → | 20, 24 |

(ii) output:

| name | projID |
|---|---|
| Goldberg | 3 |
| Goldberg | 14 |
| Abdulhamid | 5 |
| Abdulhamid | 18 |
| Zawad | 1 |
| Zawad | 4 |
| Zawad | 7 |
| Zawad | 9 |
| Trinh | 2 |
| Trinh | 11 |
| Wong | 6 |
| Wong | 8 |
| Elguera | 25 |
| Sorkin | 22 |
| Cheng | 17 |
| Chu | 19 |
| Chu | 21 |
| Martin | 10 |
| Martin | 13 |
| Lee | 15 |
| Lee | 16 |
| Nguyen | 12 |
| Nguyen | 23 |
| Dao | 20 |
| Dao | 24 |

exercise 3: what is asymptotic complexity in terms of $P$ and $E$ ?

part (a): O(?)

part (b): O(?)

total: O(?)