ages_sum = {}
ages_count = {}

with open_csv('data.csv', 'r') as f:
    for tuple in f:
        rank = fields[2]
        if rank not in ages_sum:
            ages_sum[rank] = 0
        if rank not in ages_count:
            ages_count[rank] = 0
        ages_sum[rank] += fields[3]
        ages_count[rank] += 1

For l
Student (sid, name) Course (cid, title) Signed Up For (sid, cid)

for student in 'students.csv':
    if student.name == "James T. Kirk":
        for taken in 'taken.csv':
            if taken.sid == student.sid:
                for course in 'courses.csv':
                    if taken.cid == course.cid
                        print course.title

SELECT title
FROM student, taken,
course
WHERE
    student.name = "James T. Kirk"
    
    cid = cid
    sid = sid
SELECT name, cid
FROM student, taken
taken.sid = student.sid
WHERE

Block Nested Loop
$h(x) = x \mod 4$

2-pass Hash Index Join
Step 1: Build dictionary

Step 2: Scan through T, find pairings

Run Hash Index Join

Optional: 1-pass Tree Index Join

S1, S3, S4, S5, S6, S10, S11, S12, S13
for block1 in students.csv
for block2 in taken.csv
for student in block1
for taken in block2
SORT-MERGE JOIN
Have index on T

for student in 'student.csv'
for T.index lookup(student)
taken in
print student.name, taken.cid

Index Nested Loop Join