

Quiz 5, Fall 2018
CSCI 4380 Database Systems
Time: 20 minutes

Name 1: _____
Name 2: _____
Name 3: _____

Rules.

- Open book and notes. Do not use any electronic tools including your computer.
- You can talk to anyone in class. If you know the answer, help someone else in class.
- Each quiz must be by at least two and at most three people. Most quizzes should be by three people, so find someone to talk to. Put your name on one quiz only.

Question (Points: a:3 b-d:7). You are given the following table that stores candidate information in no particular order (TUPLES(Candidates) = 8000, PAGES(Candidates) = 500):

Candidates(personid, eid, year, partyname, ballotposition, numvotes)

Index name	Indexed Attributes	Structure	# pages at leaf level
CanIdx1	Candidates(eid, year, personid)	3 levels (root, internal, leaf)	80 pages
CanIdx2	Candidates(partyname, year)	3 levels (root, internal, leaf)	50 pages

You can use this information to find approximately how many tuples are stored per leaf node in each index and per data page in the table. In addition, you are given the number of tuples for each of the following conditions:

Condition	Number of tuples
eid = 10	200
year = 2018	160
partyname = 'Morty Party'	2000
personid = 101	8
eid = 10 and year = 2018	4
partyname = 'Morty Party' and year = 2018	40
eid = 10 and personid = 101	8

The following are the queries over this table.

Q1: select partyname, numvotes from candidates where eid = 10 and year = 2018 ;
Q2: select * from candidates where partyname = 'Morty Party' and year = 2018 ;
Q3: select year from candidates where eid = 10 and personid = 101 ;

Answer the following with explanations of your computation.

- (a) What is the cost of Q1 using sequential scan? Explain with one sentence.

Answer here.

We must scan all pages in the relation: Cost= 500.

- (b) How can you answer Q1 using index CanIdx1? What is the potential cost? Explain.

Answer here.

Note: Index CanIdx1 has 80 pages in the leaf node that address 8000 tuples, which means $8000/80=100$ tuples per leaf node.

Scan index for eid = 10 and year = 2018, which finds 4 tuples which will fit in 1-2 leaf nodes of the index. So, 3-4 index pages followed by reading the matching tuples from the relation: 4 pages max. Cost = 7-8 pages.

- (c) How can you answer Q2 using index `CanIdx2`? What is the potential cost? Explain.

Answer here.

Note: Index `CanIdx2` has 50 pages in the leaf node that address 8000 tuples, which means $8000/50=160$ tuples per node.

Scan index for `partyname = 'Morty Party'` and `year = 2018` which is 40 tuples, which will fit in 1/2 leaf nodes. We then have to read the matching 40 tuples which can be in 40 different data pages. Cost: index 3-4 pages + data 40 = 43-44 total.

- (d) How can you answer Q3 using index `CanIdx1`? What is the potential cost? Explain.

Answer here.

`personid` is not useful for scanning, so we must scan the index for `eid=10`, which 200 tuples which will fit in 2-3 leaf nodes. We will find all attributes needed for this query in the index, so the total cost is 4-5 disk pages.