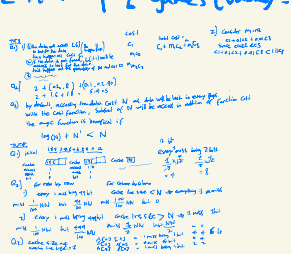
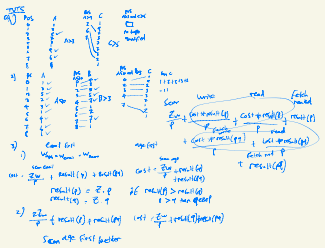
TUT 1

Q1) (2), Walmart generates a large amount of data (Volume), the data is generated frequently with day to day transaction of customer transaction (velocity), the data generated by walmart is trustworthy (Veracity) due to walmart reputation and can come in any different format (variety) for us to use the data to analyse Walmart customer purchasing behaviour (values).

(4), AI requires alot of training data (Volume), the data is generated fast for training the AI model (velocity) the data should be trustworthy to train a good model (veracity) the data come in different format (variety) . we train the AI model so that we can use it to play games (value).

Q2) 1) Velocity 3) Volume 5) Values
2) Veracity 4) Variety





TUT2

Q1) 1) key A1 value A2,A3
2) step1 Left Join Table 1 and Table 2 on T1.A3 = T2.A3
Step2 take A1 as primary key (Uniquely represent the dataset)
Step3 Set key A1 and value A2;A3;B1;B2

Q2) Step1 Left Join Table 3 and Table 1 on T3.A1 = T1.A1 and then Left Join Table 3 and Table 2 on T3.B1 = T2.B1
Step2 take A1;B1 as primary key
Step3 Set key as A1;B1 and value A2;B2

Q3) key-value model can be partially convert into relation with key can be convert into primary key and the value can be Partially convert as other attribute

Q4) Student (SID, Name) Register (SID, CID)
Course (CID, Name)

TUT 3

1) Step 1 Left Join Engineer-manager with Employer with EM.EID = E.EID and Left Join the result with Manager with EM.MID = M.MID to make T(EID, MID, E_Salary, M_Salary)
Step 2 make EID as primary key
Step4 key EID value MID, ESalary, M_Salary

2) 2 + (0.6 10) + (0.4 0.2 100)
2 + 4 + 8 = 14 ns

3) 10 pages → 10240 int
Q1 1, x
Q2 x+1, y
Q3 y+1, z
Q4 z+1, 10240
cost magic func = 0
adding Q1 to Q4 → all the int and 10 pages → Cost = 10 I/O

best case = 10 I/O s
but (x x+1)
(y y+1) → could be in same page
(z z+1)
worst case = 13 I/O s

So worst case the 4 scans might scans 3 pages they all scan

possible case 10, 11, 12, 13

4) for first 32 int
1 miss 3 hit miss $\frac{1}{4} \times \frac{8}{32} = 8$
hit $\frac{3}{4} \times \frac{8}{32} = 24$

hit = 24 + 98(16 + 12) = 2768
miss = 8 + 98(4) = 400

assume we use f-trip() first
we need to use b-trip() next so we can utilize the existing 16 int in cache without miss
16 hit out of 32
remaining 16 1 miss 3 hit miss $\frac{1}{4} \times 16$ miss = 4
hit $\frac{3}{4} \times 16$ hit 12
hit = 16

TUT8

Q① 1) data is partitioned into k/M parts
2) master will allocate tasks to each slave machines
3) each slave perform computation to the task and send back to master if necessary
4) master will perform aggregation to the result. and repeat step 2 if necessary

② S data send across M machines
M machines each return Top K to master

$O(S + Mk)$

TUT 10

70000 is distinct keys

1) L_0 (memory buffer) 5000 5 000
L_1 5000 * 4 20 000
L_2 5000 * 4 * 4 80 000 0 20000 40000 60000

2) b < c

3) 1) assume first K is in level j
case 1 i > j 0, the search terminated before going to level i
2)
case 2 i ≤ j 1, incur 1/0 from fence pointer

TUT 8

1] Map (String key, String value) {
if (key = "Student-Table") {
StudentID = Split(value).first;
CourseID = split(value).second;
Score = Split(value).third);
Emit-Intermediate (StudentID, CourseID);
}

Map output samples
(S001, C001) (S001, C002) (S002, C001)

Reduce (String key, Iterator Values) {
List distinct_course;
int n = 0;
for (V in Values) {
if (V not in distinct_course) {
distinct_course.add (V);
n++;
}
}
Emit (key, n);
}

Reduce input samples
(S001, {C001, C002}) (S002, C001)

Reduce output samples
(S001, 2) (S002, 1)

b) map (String key, String value){
if (key = "Student-Table"){
SID = split(value).first
Emit = IntermDist(courseID
}
}

map output sample
(Cool; Sem1, S001), (Cool; Sem2, S002), (Cool; Sem1, S001)
↳ this will be diffrict as the Student(S001) taking this course is 1 individual so cannot regist

Reduce input sample
(Cool; Sem1, {S001, S002})

Reduce (~~String key, Iterator Values~~){
Sem_list;
for(v in values){
if(v not in sem_list)
sem_list.insert(v, 1)
else
semlist[v]++;
}
}

Reduce output if both semester > 50
(Cool; Sem1, 51), (Cool; Sem2, 60)

input of map 1
map < String, int > X

Cnt
for(s in Sem_list)
if
Cnt++

c) Map(String key String value);
the Sem
Emit-Intermediate
(course; sem, ID);
}

map output
(Cool; 202151, S001), (Cool; 202151, S002)
(Cool; 202151, P001) ....
(all these are distinct!)

Reduce(String key, Iterator Value
String P;
List students;
for(V in values)
if(V start with P)
P = V;
else
students.add(V);
for S in students
Emit(S, P);

Reduce input
(Cool; 202151, {S001, S002, P001}), (Cool; 202152, {S1, P1}
↑ will only have 1 prof cause of my key student will be unique also

Reduce output
(P001, S001), (P001, S002) ....

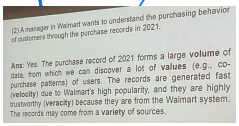
TVT 1

Question 1

Which of the following are related to big data applications, why?

(1) A computer scientist at MIT is trying to prove NP!=P (note: a famous computer science conjecture). (no, require mathematical skills)

(2) A manager in Walmart wants to understand the purchasing behavior of customers through the purchase records in 2021. Record of huge data possibility (relate it to Big data 5Vs)

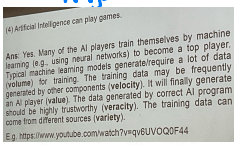

(2) A manager in Walmart wants to understand the purchasing behavior of customers through the purchase records in 2021.

Ans: Yes. The purchase record of 2021 forms a large volume of data, from which we can discover a lot of values (e.g., co-purchase patterns) of users. The records are generated fast (velocity) due to Walmart's high popularity, and they are highly trustworthy (veracity) because they are from the Walmart system. The records may come from a variety of sources.

(3) A programmer spends 10 hours in debugging his code. (no, basically use his own knowledge)

(4) Artificial Intelligence can play games. huge Value dataset

(Can be yes/no
yes → 5Vs
NO → some AI are rule based)

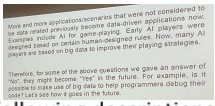

(4) Artificial Intelligence can play games.

Ans: Yes. Many of the AI players train themselves by machine learning (e.g., using neural networks) to become a top player. Typical machine learning models generate/require a lot of data (volume) for training. The training data may be frequently generated by other components (velocity). It will finally generate an AI player (value). The data generated by correct AI program should be highly trustworthy (veracity). The training data can come from different sources (variety).

E.g., https://www.youtube.com/watch?v=qv6UVOQ0F44

Question 2 [Open Question]

Is it possible that an application is not considered a big data application now, but can possibly become a big data application in the future? yes, as improvement is possible additional dataset might be required when 4 Volume, 4 Variety


More and more applications/scenarios that were not considered to be data-related previously become data-driven applications now. Examples include AI for game-playing. Early AI players were designed based on certain human-designed rules. Now, many AI players are based on big data to improve their playing strategies.

Therefore, for some of the above questions we gave an answer of "No", they might become "Yes" in the future. For example, is it possible to make use of big data to help programmers debug their code? Let's see how it goes in the future.

Question 3

For each of the following descriptions, which of the 5V's it is the most related?

(1) There are more than 60000 searches per second in the Google search engine. Velocity

(2) Suppose there is a database storing all kinds of enterprise related data, including the ratios of male/female in different companies. Ben wrote the following C program to collect the ratio of male/female for a company.

```
int cal_ratio(){
        int num_Male=getMales();
        int num_Female=getFemales();
        return num_Male/num_Female;
}
```

~~Variety~~ ratio is not fractional number long in algorithm, output biased data (Veracity)

(3) ImageNet is a 150GB dataset that holds 1,281,167 images for training and 50,000 images for validation, organized in 1,000 categories. Volume

(4) A website allows users to upload different forms of documents such as Excel, JPG, PDF, video. in different format Variety Value

(5) A researcher performs data mining algorithms on Amazon's purchase records, and he successfully predicts the best seller in the next month. ~~Veracity~~ Value

CE/CZ 4123 Tutorial 2—Data Models

Question 1

Given the following relational schemas containing two tables (a.k.a., relations), and their primary keys are underlined:

Table1(A1, A2, A3)
Table2(A3, B1, B2)

Attribute A3 is the primary key of Table2 and is also a foreign key in Table1.

(1) how do you convert Table 1 **alone** into key-value model? Give one possible solution.

key A_1 Value A_2, A_3

(2) how to convert them into key-value data model? Give one possible solution.

Step 1 Join table1 and Table2 (left) using attr A3 we have bigger table T_3 (A_1, A_2, A_3, B_1, B_2)
Step 2 make A_1 primary key (key)
A_2, A_3, B_1, B_2 (value)

Question 2

Given the following relational schema containing three tables (primary keys are underlined):

Table1(A1, A2)
Table2(B1, B2)
Table3(A1, B1)

Please give one possible way to convert the above relational model into key-value model. (you may assume that query is always issued with respect of A1, B1)

Step1 Left Join T3 with T_1 on $T_1.A_1 = T_3.A_1$ and with T_2 on $T_2.B_1 = T_3.B_1$
Step 2 make A1 and B1 as key
A2 and B2 as Value

Question 3

Can key-value model be converted into a relation? Why?

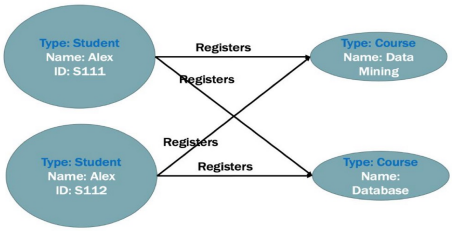
Ans: Key-value model can be trivially converted into a relation as follows

Table(key, value)

Note: This solution, however, is not able to uncover any schema that exists in **value**.

Question 4

Given the following instance based on the graph mode. Please convert it into the instance based on the relational model-based instance. (Hint: converting different types of nodes/edges to tables.)



Step 1 : Consider number of tables

Student (Name, ID)
Course (Name)
Register (studentID, CourseName) (many to many)
! Can try construct dummy table

Question 5 (open discussion)

Why do we need the graph model? Discuss it from the physical storage's perspective.

Ans: Graphs can be stored in the order based on adjacency list. In fundamental graph related operations such as graph traversal, the data access order aligns with the adjacency list. In contrast, relational tables often require costly "join" for these operations.

Example: Reconsider the example for Q4, we have two students S1, S2, and two courses C1, C2.
Nodes: S1, S2, C1, C2; **Edges:** (S1, C1), (S1, C2), (S2, C1), (S2, C2)

Adjacency list:

S1's neighbor list: C1, C2
S2's neighbor list: C1, C2
C1's neighbor list: S1, S2 (if we do not consider the edge direction)
C2's neighbor list: S1, S2 (if we do not consider the edge direction)

graph can be faster when accessing

CE/CZ 4123 Tutorial 3 – Memory Hierarchy

Question 1

We consider three-layer memory hierarchy, L(i), L(i+1), L(i+2). Their access costs are c1, c2, c3, respectively, with c1<c2<c3. Assume that the data access always checks the existence of data in the order of L(i), L(i+1), and L(i+2). The miss rates of L(i) and L(i+1) are m1 and m2, respectively.

(1) Show that we can estimate the data access cost by c1+m1c2+m1m2c3.

(2) Suppose m1=m2=0.1, show that the over cost is at most 1.11c3.

Question 2

Consider reading data from memory hierarchy consisting of L1-cache, L2-cache, and main memory. Their read access times and hit ratios are given below:

L1-cache:
read access time: 2 nanoseconds; hit ratio: 0.8
L2-cache:
read access time: 8 nanoseconds: hit ratio: 0.9
Main memory:
read access time: 90 nanoseconds.

Please estimate the average data read cost (considering L1, L2 caches and main memory only).

Question 3

Consider the 2nd magic function we mentioned in the lecture, i.e., the magic function that can tell us which pages contain the qualified data. In practice, such magic function is implemented by a certain data structure and hence it incurs some cost when call the function.

Suppose the cost of calling the function is equal to accessing log(N) pages, where N is the number of pages used to store the data. Please give a condition about when is beneficial to use the function.

Question 4

Consider the array-scanning scenario introduced in the lecture. In the lecture, we consider a single query for x>4. In big data systems, many queries are issued together.

Suppose our system needs to handle the following two queries:

1) Select x>4
2) Select x<2

Please explain an efficient way of finishing these two queries together and analyze the number of page accesses needed.

CE/CZ 4123 Big Data Management Tutorial 4
Cache Conscious Designs

Question 1

We have a 12-integer array in main memory as follows
1, 3, 5, 2, 4, 6, 4, 6, 8, 9, 11, 12
Let cache size be the size of 6 integers, and cache line size (transfer unit) be the size of 3 integers. Suppose initially the cache is empty, and there is a program sequentially accessing the whole array. The cache replace mechanism is the same as in the lecture notes: i.e., first cached first evicted.

(1) After the execution of the program, what are the final values stored in the cache? Please give the cache state after every access of the array element.
(2) How many cache hits and cache misses during the program execution? Please give the hit/miss state after every access of the array element.

Question 2

Suppose we have a 2-dimentional integer array A[N][N]. We consider the two ways of array scanning: row-by-row and column-by-column. Let cache size = 5000 integers.

(1) If N>5000 and cache line size (transfer unit)=100 integers, please give a formal analysis of the cache hits/misses of the two ways of writing codes.
(2) If N=250 and cache line size (transfer unit)=500 integers, please give a formal analysis of the cache hits/misses of the two ways of writing codes.

Question 3

We have an 8-integer array A in the main memory. Let cache size be 4 (integers), and cache line size be 2 (integers). Suppose that initially the cache is empty, and the cache replacement policy is the same as the one introduced in the lecture notes, i.e., first cached first evicted.

Let the round-trip scanning be scanning an array in the order from the beginning to the end (i.e., A[0] to A[7]), and then from the end to the beginning (i.e., A[7] to A[0]). How many cache hits and cache misses during the round-trip scanning? Please explain your answer.

Q2) 1) for row by row, cache line = 100
every cache miss brings 100 ints
(1 miss and 99 cache hit)

overall cache miss: row column N.N / 100
overall cache hit: row column N.N / 100 . 99

for column by column with cache line = 100
every access = cache miss
overall cache miss: N.N
overall cache hit: 0

2) for row by row, cache line = 500
every 1 miss 499 hit

overall cache miss: N.N / 500
overall cache hit: N.N / 500 . 499

for column by column
every 1 miss = 1 hit (hit next comm.)

overall cache miss: N.N / 2
overall cache hit: N.N / 2

Q1) 1 3 5 2 4 6 4 6 8 9 11 12
cache size = 6
cache line size = 3
cache miss = 0
cache hit = 0

initial
CM=0 CH=0
1 3 5 | 2 4 6 | 4 6 8 | 9 11 12
1 2 3 4
Cache 6/3=2

access pattern 1 3 5 4 6 7 8 9 10 11 12

CM=1 CH=0 | CM=1 CH=1 | CM=1 CH=2 | CM=2 CH=2 | CM=2 CH=3 | CM=2 CH=4 | CM=3 CH=4
468246 | 468246 | 468 9 11 12 | 468 9 11 12 | 468 9 11 12 ①
CM=3 CH=5 | CM=3 CH=6 | CM=4 CH=6 | CM=4 CH=7 | CM=4 CH=8 ②

Q3) access pattern
0,1,2,3,4,5,6,7 < 7,6,5,4,5,2,1,0
forward | backward
cache size 4
cache line 2

A0 A1 A2 A3 A4 A5 A6 A7
0 1 2 3 4 5 6 7

forward
0 | code: A0 A1 H:0 M:1
01 | A0 A1 H:1 M:1
012 | A0A1A2A3 H:1 M:2
0123 | A0A1A2A3 H:2 M:2
01234 | A2A3A4A5 H:2 M:3
012345 | A2A3A4A5 H:3 M:3
0124456 | A4A5A6A7 H:3 M:4
01234567 | A4A5A6A7 H:4 M:4

backward
7 | A6A7 H:5 M:4
76 | A6A7A4A5 H:6 M:4
765 | A4A5A6A7 H:7 M:4
7654 | A4A5A6A7 H:8 M:4
76543 | A2A3A4A5 H:8 M:5
765432 | A2A3A4A5 H:9 M:5
7654321 | A2A3A4A5 H:9 M:6
76543210 | A2A3A4A5 H:10 M:6

CE/CZ 4123 Big Data Management Tutorial 5
Column Stores

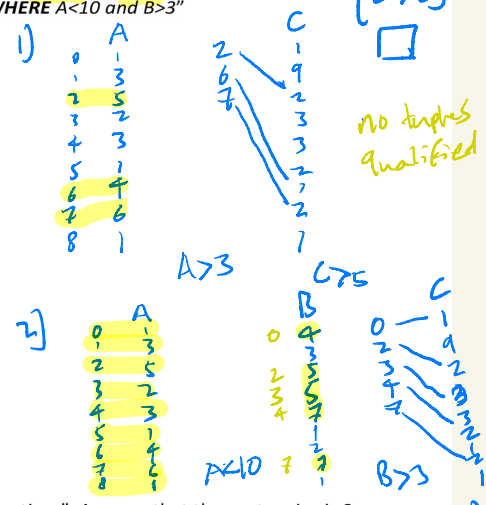
Question 1

Given column store table T as follow.

(1) Give the flow chart using "column at a time" for the query "***SELECT*** *min(B)* ***FROM*** *T* ***WHERE*** *A>3 and C>5*"

(2) Give the procedure using "column at a time" for the query "***SELECT*** *sum(C)* ***FROM*** *T* ***WHERE*** *A<10 and B>3*"

A	B	C
1	4	1
3	3	9
5	5	2
2	5	3
3	7	3
1	1	2
4	2	1
6	7	2
1	1	1



Question 2

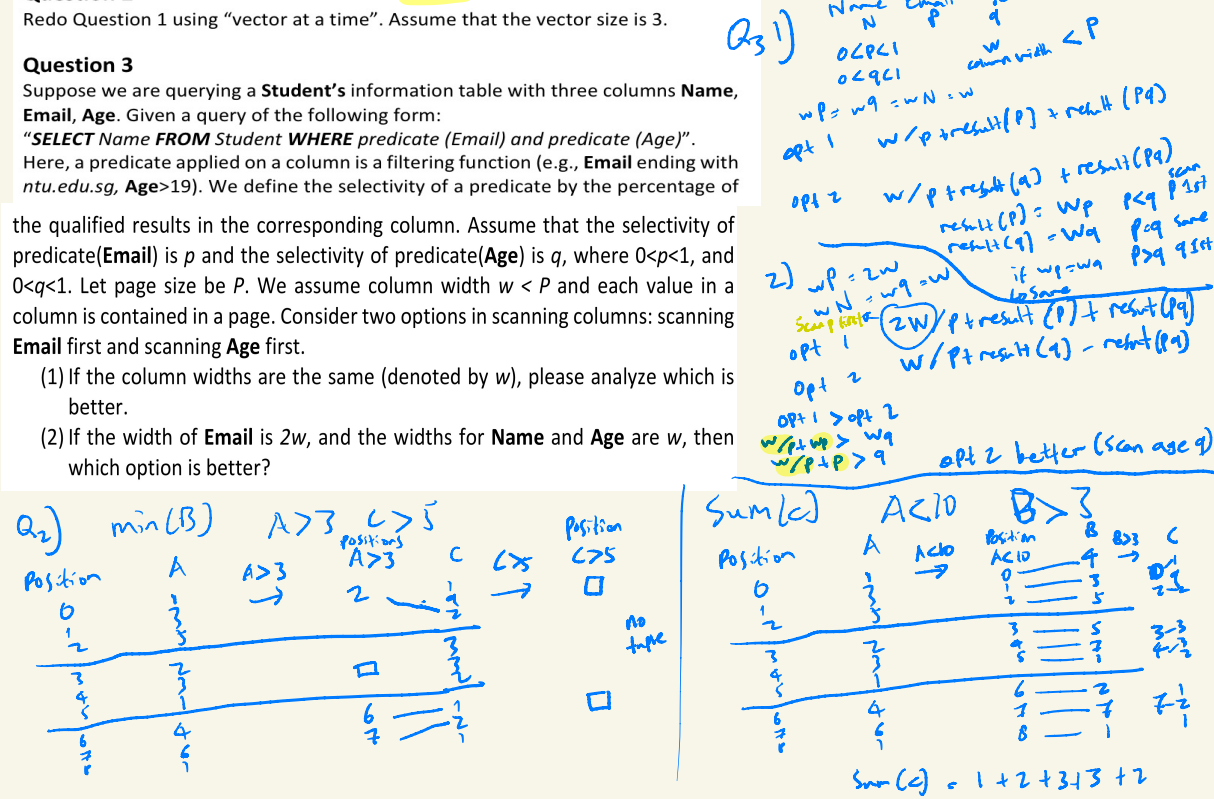
Redo Question 1 using "vector at a time". Assume that the vector size is 3.

Question 3

Suppose we are querying a **Student's** information table with three columns **Name**, **Email**, **Age**. Given a query of the following form:
"***SELECT*** *Name* ***FROM*** *Student* ***WHERE*** *predicate (Email) and predicate (Age)*".
Here, a predicate applied on a column is a filtering function (e.g., **Email** ending with *ntu.edu.sg,* **Age**>19). We define the selectivity of a predicate by the percentage of the qualified results in the corresponding column. Assume that the selectivity of predicate(**Email**) is *p* and the selectivity of predicate(**Age**) is *q*, where 0<*p*<1, and 0<*q*<1. Let page size be *P*. We assume column width *w* < *P* and each value in a column is contained in a page. Consider two options in scanning columns: scanning **Email** first and scanning **Age** first.

(1) If the column widths are the same (denoted by *w*), please analyze which is better.

(2) If the width of **Email** is *2w*, and the widths for **Name** and **Age** are *w*, then which option is better?



CE/CZ 4123 Big Data Management Tutorial 7

More Practice for Preparing for Quizzes

1) Converting to key-value model :

Step 1 : Left Join Employee-Manager table(EM) with Employee(E) table based on EM.EID = E.EID and Left Join with Manager table(M) based on EM.MID = M.MID to output table T1 where T1(EID, MID, ESalary, MSalary)

Step 2 : Convert to key-value

key = EID Value = MID; ESalary; MSalary

1. Given the following three tables (primary keys are underlined):

   Employee(EID, Salary)

   Manager(MID, Salary)

   Employee-Manager(EID, MID)

   Each manager supervises at least one employees. Employee-Manager is a table that contains the manager ID (i.e., MID) for each employee (i.e., EID). How to convert the relational data model to a key-value data model? Consider that the main purpose of the conversion is for the query "Given an employee ID, find the salary of the employee's manager". The conversion should retain the information as much as possible.

2) • L1 hit = 2ns
• L1 miss L2 hit = 0.4 . 10ns
• L1 miss L2 miss = 0.4 . 0.2 . 100 ns
2 + 4 + 8 = 14ns

L1 hit cost 2ns, if L1 miss and L2 hit, it happens at L1 miss ratio and cost of L2, if L2 missed it will go to mainmemory which happens at L1 miss ratio, L2 miss ratio with main memory access price

2. Consider reading data from memory hierarchy consisting of L1 Cache, L2 Cache, and main memory with the following parameters.

   - L1 Cache:

     Read access time: 2 nanoseconds

     Miss ratio: 0.4

   - L2 Cache:

     Read access time: 10 nanoseconds

     Miss ratio: 0.2

   - Main memory:

     Read access time: 100 nanoseconds.

   Estimate the average data read cost and explain your answer. (Note: consider L1, L2 caches and main memory only).

3) Page Size = 1024
Page line size = 10
best Q1 + Q2 + Q3 + Q4 = exactly 10 I/Os

x n x+1
y n y+1
z n z+1
} Share 1 page I/Os = 3 So need to access twice

worst 10 + 3 I/Os

3. In the lecture, we introduced a cost-free magic function telling which pages locate the qualified data for a query. Consider a disk page size is 1024 integers. There are 10240 integers, which are 1, 2, 3, …, 10240, sequentially stored at 10 consecutive disk pages.

   Consider the following 4 queries using 4 scans over the data, where each query range is decided by three integers x, y, z, and 1<x<y<z<10240.

   Query 1: searching values in the range [1, *x*] (i.e., values at least 1 and at most *x*)
   Query 2: searching values in the range [*x*+1, *y*]
   Query 3: searching values in the range [*y*+1, *z*]
   Query 4: searching values in the range [*z*+1, 10240]

   List **all possible** total number of read I/Os needed for the 4 scans, *with* the magic function. Please explain your answer.

4) Cache size 16
f-trip()
cache line size 4
0 31
each miss bring 3 hit
b-trip()
31 0
each miss bring 3 hit
1 array = 32 int
cache size
1 trip
miss = 32/4 = 8
hit = 32 . 3 / 4 = 24
miss = 8
hit = 24
1 scan of array = 32 int
16 int

	hit	miss
1 f-trip	24	8
99 trip	28	4 (after 16)
b-trip		
f-trip		

4 miss
32 − 4 = 28 hit

miss 4.98 + 8 = 400
hit 28.98 + 24 = 2760

4. We have a 32-integer array *A* in the main memory. Let cache size be 16 (integers), and cache line size be 4 (integers). Suppose that initially the cache is empty, and the cache replacement policy is the same as the one introduced in the lectures, i.e., first cached first evicted. Let the *f-trip* and *b-trip* scanning be the following.

```
f-trip(){
        for (int j=0; j<32; j++){
                Access A[j];// Access does not change the data
        }
}
b-trip(){
        for (int j=0; j<32; j++){
                Access A[31-j];
        }
}
```

   If we need to do 99 scans of the array, and we can select each scan to be either f-trip or b-trip. Please give one best selection strategy that gives the minimum number of misses and explain your answer. Please also compute the number of cache hits and cache misses in the best strategy.

CE/CZ 4123 Big Data Management Tutorial 7
Distributed Systems and MapReduce

Question 1

Amazon wants to estimate the Top-K best sold products from S purchase records of L products in the form of a list of (User id, Product id) pairs. Assume that L is a multiple of M. Suppose there is a distributed system with 1 master machine and M slave machines. Design a distributed computation procedure to finish the task. Please describe

(1) how the data is distributed, computed and aggregated?

(2) how much data is sent across machines?

Question 2:

Consider the MapReduce paradigm and answer the following questions.

(1) In a MapReduce job, the output of Map phase is a list of key-value pairs: (A, 1) (C, 2), (A, 5), (C, 6), (B, 3), (E, 3), (C, 8). Please list the possible input to the Reduce function.

(2) Based on the answer to Q2(a), write a Reduce function (in pseudocode) so that the MapReduce output is: (2, A), (3, C), (6, A), (7, C), (4, B), (4, E), (9, C).

(3) Consider an employee table containing three columns (EmployeeID, age, monthly-salary) where age and monthly-salary are integers. Use MapReduce to collect the number of employees falling into each of the following two categories:

- Category 1: The age of the employee is between 30 and 40 (including 30 and 40). His/her monthly salary is at most 7000.
- Category 2: The age of the employee is between 40 and 50 (including 40 and 50). His/her monthly salary is more than 7000.

Please use only one MapReduce Job to achieve this task and write down the pseudocode of the Map function and Reduce function. The input key and value for Map function are an employee's age and monthly-salary respectively.

(Example: if there are 100 employees in Category 1 and 50 employees in Category 2, then the MapReduce output will contain two key-value pairs: (1, 100), (2, 50))

Question 3: (important)

Design MapReduce algorithms for the multiplication of two matrices A, B of n by n. Elements in the matrices are integers. The input key for *Map* function is *MatrixName*; the input value for *Map* function is in the form of *i;j;v*, indicating that the value of the *i*-th row and *j*-th column is *v*.

(Matrix Multiplication: Given matrix A[nxn] and matrix B[nxn], compute matrix C such that $C[i][z] = \sum_{j=0}^{n-1} A[i][j] \times B[j][z]$, for i,z in [0, n-1]).

(1) Use at most two MapReduce jobs to finish the computation.
(2) Furthermore, can the multiplication be finished using one MapReduce job?

Big Data Management Tutorial - MapReduce

Consider a *Student* table containing three columns (*studentID*, *courseID*, *semester*). Each tuple in the *Student* table records that a student registered for the course in the corresponding semester. Also consider a *Professor* table containing three columns (*professorID*, *courseID*, *semester*). Each tuple in the *Professor* table records that a professor teaches a course in the corresponding semester. A course may open in multiple semesters. If a student fails a course, he may retake the course in later semesters. **Example tuples** for the two tables are as follows.

studentID	courseID	semester
S001	C001	2021S1
S002	C001	2021S1
S002	C002	2021S2
S003	C001	2021S2

Table Q4.1: Example Student Table

professorID	courseID	semester
P001	C001	2021S1
P002	C002	2021S1
P001	C001	2021S2

Table Q4.2: Example Professor Table

The *Professor* table and the *Student* table are stored together in a file named *input_file*, with each tuple per line. There is an additional attribute to indicate whether this tuple is from the *Student* Table or the *Professor* Table. Based on the above example tuples, the file content is as follows.

Student-Table S001;C001;2021S1
Student-Table S002;C001;2021S1
Student-Table S002;C002;2021S2
Student-Table S003;C001;2021S2
Professor-Table P001;C001;2021S1
Professor-Table P002;C002;2021S1
Professor-Table P001;C001;2021S2

Please use MapReduce for the following scenarios and write down the pseudocode. Your pseudocode should start with a *Map* function that takes each line of the *input_file* as the input. The key in the *Map* function is the additional attribute (e.g., "*Student-Table*") in the line, and the value in the *Map* function is the remaining of the line (e.g., *S001;C001;2021S1*). You also need to design *Reduce* function if necessary. You can use multiple MapReduce jobs.

(a) Use MapReduce to collect for each student (represented by *studentID*) the number of distinct courses (represented by *courseID*s) he has registered.

(b) Use MapReduce to collect the courses (represented by *courseID*s) that have more than 50 registered students for at least two semesters. For example, a course will be output if there are 55 students for 2021S1 and 60 students for 2021S2.

(c) Use MapReduce to output every pair of (student, professor) (represented by *studentID* and *professorID*) that the student has attended at least one courses taught by the professor.

CE/CZ 4123 Big Data Management Tutorial 10
Key-Value Stores

Question 1:
Consider a leveling LSM-tree with a size ratio 4. The memory buffer (Level 0) can store 5000 key-value pairs. Initially the LSM-tree is empty. After inserting 70000 key-value pairs with distinct keys continuously, how many levels are formed?

Question 2:
Consider a leveling LSM-tree with a size ratio 4. The LSM-tree has 5 levels (excluding the memory buffer level), and it is incorporated with **both** fence pointers and Bloom Filters. Assume that a key-value pair is always entirely stored within a disk page. Consider the procedure of Get(K) for a key K, which of the followings sequence are possible to be the I/O costs from Level-0 to Level-5? (can select multiple answers)

(a) 1, 1, 1, 1, 1, 1
(b) 0, 1, 1, 1, 1, 1
(c) 0, 0, 1, 0, 0, 1
(d) 0, 0, 0, 0, 0, 1
(e) 1, 0, 0, 0, 0, 0

Question 3:
Consider a leveling LSM-tree with a size ratio 4. The LSM-tree has L levels (excluding the memory buffer level), and it is **only** incorporated with fence pointers (without Bloom Filters). Assume that a key-value pair is always entirely stored within a disk page. Consider the procedure of Get(K) for a key K,

(1) What is the possible I/O cost of accessing Level-i (i is in [1, L])?
(2) If K exists in the LSM-tree, what is the expected I/O cost at Level-i (i is in [1, L])? (hint: divide the cases based on the first-appearing location of the key K)

Question 4:
Consider a leveling LSM-tree with a size ratio 4. The LSM-tree has L levels (excluding the memory buffer level), and it is incorporated with **both** fence pointers and Bloom Filters. Assume that a key-value pair is always entirely stored within a disk page. Consider the procedure of Get(K) for a key K,

(1) What is the possible I/O cost of accessing Level-i (i is in [1, L])?
(2) If K exists in the LSM-tree and the FPR of the Bloom filter at Level-i (i is in [1, L]) is P (P is in [0, 1]), what is the expected I/O cost at Level-i? (hint: divide the cases based on the first-appearing location of the key K)

CZ/CE 4123 Tutorial 1 : Big Data 5V's

Question 1

Which of the following are related to big data applications, why?

(1) A computer scientist at MIT is trying to prove NP!=P (note: a famous computer science conjecture).

(2) A manager in Walmart wants to understand the purchasing behavior of customers through the purchase records in 2021.

(3) A programmer spends 10 hours in debugging his code.

(4) Artificial Intelligence can play games.

Question 2 [Open Question]

Is it possible that an application is not considered a big data application now, but can possibly become a big data application in the future?

Question 3

For each of the following descriptions, which of the 5V's it is the most related?

(1) There are more than 60000 searches per second in the Google search engine.

(2) Suppose there is a database storing all kinds of enterprise related data, including the ratios of male/female in different companies. Ben wrote the following C program to collect the ratio of male/female for a company.

```
int cal_ratio(){
        int num_Male=getMales();
        int num_Female=getFemales();
        return num_Male/num_Female;
}
```

(3) ImageNet is a 150GB dataset that holds 1,281,167 images for training and 50,000 images for validation, organized in 1,000 categories.

Volume

(4) A website allows users to upload different forms of documents such as Excel, JPG, PDF, video.

Variety

(5) A researcher performs data mining algorithms on Amazon's purchase records, and he successfully predicts the best seller in the next month.

Value

CE/CZ 4123 Tutorial 2—Data Models

Question 1

Given the following relational schemas containing two tables (a.k.a., relations), and their primary keys are underlined:

Table1(A1, A2, A3)
Table2(A3, B1, B2)

Attribute A3 is the primary key of Table2 and is also a foreign key in Table1.

(1) how do you convert Table 1 **alone** into key-value model? Give one possible solution.

key A1 value A2; A3

(2) how to convert them into key-value data model? Give one possible solution.

Left join T1 and T2 by A3
key A1 value A2; A3; B1; B2

Question 2

Given the following relational schema containing three tables (primary keys are underlined):

Table1(A1, A2)
Table2(B1, B2)
Table3(A1, B1)

Please give one possible way to convert the above relational model into key-value model.

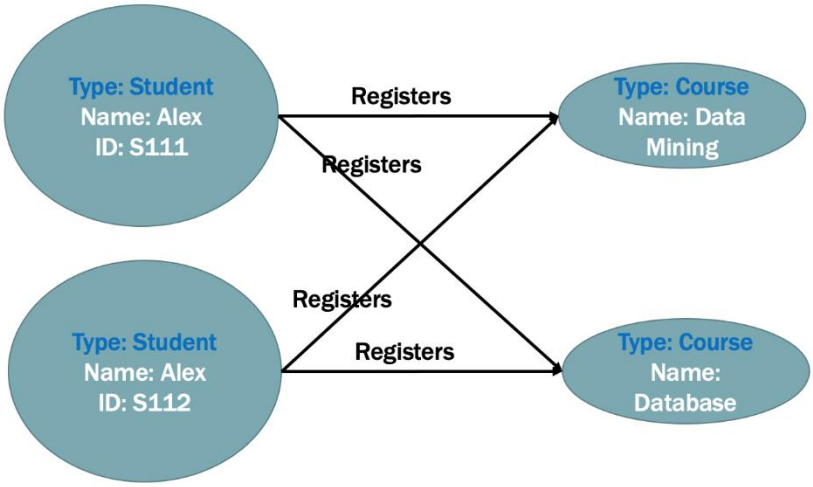
T3 Left Join T1 by A1 and
Left Join T2 by B1 to get
T4 (A1, B1, A2 B2)
key (A1; B1) value A2, B2
ASSUMING (A1, B1) is interested of user

Question 3

Can key-value model be converted into a relation? Why?

Question 4

Given the following instance based on the graph mode. Please convert it into the instance based on the relational model-based instance. (Hint: converting different types of nodes/edges to tables.)



Student
Name ID

Register
Student Course

Course
Name Database

Question 5 (open discussion)

Why do we need the graph model? Discuss it from the physical storage's perspective.

CE/CZ 4123 Tutorial 3 – Memory Hierarchy

Question 1

We consider three-layer memory hierarchy, L(i), L(i+1), L(i+2). Their access costs are c1, c2, c3, respectively, with c1<c2<c3. Assume that the data access always checks the existence of data in the order of L(i), L(i+1), and L(i+2). The miss rates of L(i) and L(i+1) are m1 and m2, respectively.

(1) Show that we can estimate the data access cost by c1+m1c2+m1m2c3.

(2) Suppose m1=m2=0.1, show that the over cost is at most 1.11c3.

Question 2

Consider reading data from memory hierarchy consisting of L1-cache, L2-cache, and main memory. Their read access times and hit ratios are given below:

L1-cache:
read access time: 2 nanoseconds; hit ratio: 0.8
L2-cache:
read access time: 8 nanoseconds: hit ratio: 0.9
Main memory:
read access time: 90 nanoseconds.

Please estimate the average data read cost (considering L1, L2 caches and main memory only).

Question 3

Consider the 2nd magic function we mentioned in the lecture, i.e., the magic function that can tell us which pages contain the qualified data. In practice, such magic function is implemented by a certain data structure and hence it incurs some cost when call the function.

Suppose the cost of calling the function is equal to accessing log(N) pages, where N is the number of pages used to store the data. Please give a condition about when is beneficial to use the function.

Question 4

Consider the array-scanning scenario introduced in the lecture. In the lecture, we consider a single query for x>4. In big data systems, many queries are issued together.

Suppose our system needs to handle the following two queries:

1) Select x>4
2) Select x<2

Please explain an efficient way of finishing these two queries together and analyze the number of page accesses needed.

CE/CZ 4123 Big Data Management Tutorial 4
Cache Conscious Designs

Question 1

We have a 12-integer array in main memory as follows
1, 3, 5, 2, 4, 6, 4, 6, 8, 9, 11, 12
Let cache size be the size of 6 integers, and cache line size (transfer unit) be the size of 3 integers. Suppose initially the cache is empty, and there is a program sequentially accessing the whole array. The cache replace mechanism is the same as in the lecture notes: i.e., first cached first evicted.

(1) After the execution of the program, what are the final values stored in the cache? Please give the cache state after every access of the array element.
(2) How many cache hits and cache misses during the program execution? Please give the hit/miss state after every access of the array element.

Question 2

Suppose we have a 2-dimentional integer array A[N][N]. We consider the two ways of array scanning: row-by-row and column-by-column. Let cache size = 5000 integers.

(1) If N>5000 and cache line size (transfer unit)=100 integers, please give a formal analysis of the cache hits/misses of the two ways of writing codes.
(2) If N=250 and cache line size (transfer unit)=500 integers, please give a formal analysis of the cache hits/misses of the two ways of writing codes.

Question 3

We have an 8-integer array A in the main memory. Let cache size be 4 (integers), and cache line size be 2 (integers). Suppose that initially the cache is empty, and the cache replacement policy is the same as the one introduced in the lecture notes, i.e., first cached first evicted.

Let the round-trip scanning be scanning an array in the order from the beginning to the end (i.e., A[0] to A[7]), and then from the end to the beginning (i.e., A[7] to A[0]). How many cache hits and cache misses during the round-trip scanning? Please explain your answer.

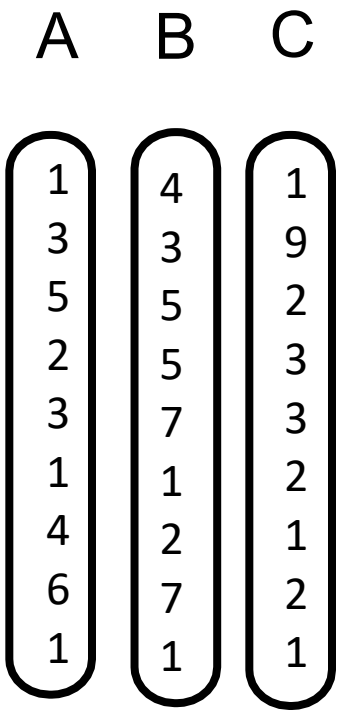
CE/CZ 4123 Big Data Management Tutorial 5
Column Stores

Question 1

Given column store table T as follow.

(1) Give the flow chart using "column at a time" for the query "***SELECT*** *min(B)* ***FROM*** *T* ***WHERE*** *A>3 and C>5*"

(2) Give the procedure using "column at a time" for the query "***SELECT*** *sum(C)* ***FROM*** *T* ***WHERE*** *A<10 and B>3*"



Question 2

Redo Question 1 using "vector at a time". Assume that the vector size is 3.

Question 3

Suppose we are querying a **Student's** information table with three columns **Name**, **Email**, **Age**. Given a query of the following form:
"***SELECT*** *Name* ***FROM*** *Student* ***WHERE*** *predicate (Email) and predicate (Age)*".
Here, a predicate applied on a column is a filtering function (e.g., **Email** ending with *ntu.edu.sg,* **Age**>19). We define the selectivity of a predicate by the percentage of

the qualified results in the corresponding column. Assume that the selectivity of predicate(**Email**) is p and the selectivity of predicate(**Age**) is q, where $0<p<1$, and $0<q<1$. Let page size be P. We assume column width $w < P$ and each value in a column is contained in a page. Consider two options in scanning columns: scanning **Email** first and scanning **Age** first.

(1) If the column widths are the same (denoted by w), please analyze which is better.

(2) If the width of **Email** is $2w$, and the widths for **Name** and **Age** are w, then which option is better?

CE/CZ 4123 Big Data Management Tutorial 7

More Practice for Preparing for Quizzes

1. Given the following three tables (primary keys are underlined):

   Employee(EID, Salary)

   Manager(MID, Salary)

   Employee-Manager(EID, MID)

   Each manager supervises at least one employees. Employee-Manager is a table that contains the manager ID (i.e., MID) for each employee (i.e., EID). How to convert the relational data model to a key-value data model? Consider that the main purpose of the conversion is for the query "Given an employee ID, find the salary of the employee's manager". The conversion should retain the information as much as possible.

2. Consider reading data from memory hierarchy consisting of L1 Cache, L2 Cache, and main memory with the following parameters.

   - L1 Cache:

     Read access time: 2 nanoseconds

     Miss ratio: 0.4

   - L2 Cache:

     Read access time: 10 nanoseconds

     Miss ratio: 0.2

   - Main memory:

     Read access time: 100 nanoseconds.

   Estimate the average data read cost and explain your answer. (Note: consider L1, L2 caches and main memory only).

3. In the lecture, we introduced a cost-free magic function telling which pages locate the qualified data for a query. Consider a disk page size is 1024 integers. There are 10240 integers, which are 1, 2, 3, …, 10240, sequentially stored at 10 consecutive disk pages.

Consider the following 4 queries using 4 scans over the data, where each query range is decided by three integers x, y, z, and 1<x<y<z<10240.

Query 1: searching values in the range [1, *x*] (i.e., values at least 1 and at most *x*)
Query 2: searching values in the range [*x*+1, *y*]
Query 3: searching values in the range [*y*+1, *z*]
Query 4: searching values in the range [*z*+1, 10240]

List **all possible** total number of read I/Os needed for the 4 scans, *with* the magic function. Please explain your answer.

4. We have a 32-integer array *A* in the main memory. Let cache size be 16 (integers), and cache line size be 4 (integers). Suppose that initially the cache is empty, and the cache replacement policy is the same as the one introduced in the lectures, i.e., first cached first evicted. Let the *f-trip* and *b-trip* scanning be the following.

```
f-trip(){
        for (int j=0; j<32; j++){
                Access A[j];// Access does not change the data
        }
}
b-trip(){
        for (int j=0; j<32; j++){
                Access A[31-j];
        }
}
```

If we need to do 99 scans of the array, and we can select each scan to be either f-trip or b-trip. Please give one best selection strategy that gives the minimum number of misses and explain your answer. Please also compute the number of cache hits and cache misses in the best strategy.

CE/CZ 4123 Big Data Management Tutorial 7
Distributed Systems and MapReduce

Question 1

Amazon wants to estimate the Top-*K* best sold products from *S* purchase records of *L* products in the form of a list of (User id, Product id) pairs. Assume that *L* is a multiple of *M*. Suppose there is a distributed system with 1 master machine and *M* slave machines. Design a distributed computation procedure to finish the task. Please describe

(1) how the data is distributed, computed and aggregated?

(2) how much data is sent across machines?

Question 2:

Consider the MapReduce paradigm and answer the following questions.

(1) In a MapReduce job, the output of Map phase is a list of key-value pairs: (A, 1) (C, 2), (A, 5), (C, 6), (B, 3), (E, 3), (C, 8). Please list the possible input to the Reduce function.

(2) Based on the answer to Q3(a), write a Reduce function (in pseudocode) so that the MapReduce output is: (2, A), (3, C), (6, A), (7, C), (4, B), (4, E), (9, C).

(3) Consider an employee table containing three columns (EmployeeID, age, monthly-salary) where age and monthly-salary are integers. Use MapReduce to collect the number of employees falling into each of the following two categories:

- Category 1: The age of the employee is between 30 and 40 (including 30 and 40). His/her monthly salary is at most 7000.

- Category 2: The age of the employee is between 40 and 50 (including 40 and 50). His/her monthly salary is more than 7000.

Please use only one MapReduce Job to achieve this task and write down the pseudocode of the Map function and Reduce function. The input key and value for Map function are an employee's age and monthly-salary respectively.

(Example: if there are 100 employees in Category 1 and 50 employees in Category 2, then the MapReduce output will contain two key-value pairs:
(1, 100), (2, 50))

Question 3:

Design MapReduce algorithms for the multiplication of two matrices A, B of n by n. Elements in the matrices are integers. The input key for *Map* function is *MatrixName*; the input value for *Map* function is in the form of *i;j;v*, indicating that the value of the *i*-th row and *j*-th column is *v*.

(Matrix Multiplication: Given matrix A[nxn] and matrix B[nxn], compute matrix C such that $C[i][z] = \sum_{j=0}^{n-1} A[i][j]xB[j][z]$, for i,z in [0, n-1]).

(1) Use at most two MapReduce jobs to finish the computation.
(2) Furthermore, can the multiplication be finished using one MapReduce job?

Big Data Management Tutorial - MapReduce

Consider a *Student* table containing three columns (*studentID*, *courseID*, *semester*). Each tuple in the *Student* table records that a student registered for the course in the corresponding semester. Also consider a *Professor* table containing three columns (*professorID*, *courseID*, *semester*). Each tuple in the *Professor* table records that a professor teaches a course in the corresponding semester. A course may open in multiple semesters. If a student fails a course, he may retake the course in later semesters. **Example tuples** for the two tables are as follows.

studentID	courseID	semester
S001	C001	2021S1
S002	C001	2021S1
S002	C002	2021S2
S003	C001	2021S2

Table Q4.1: Example Student Table

professorID	courseID	semester
P001	C001	2021S1
P002	C002	2021S1
P001	C001	2021S2

Table Q4.2: Example Professor Table

The *Professor* table and the *Student* table are stored together in a file named *input_file*, with each tuple per line. There is an additional attribute to indicate whether this tuple is from the *Student* Table or the *Professor* Table. Based on the above example tuples, the file content is as follows.

Student-Table S001;C001;2021S1
Student-Table S002;C001;2021S1
Student-Table S002;C002;2021S2
Student-Table S003;C001;2021S2
Professor-Table P001;C001;2021S1
Professor-Table P002;C002;2021S1
Professor-Table P001;C001;2021S2

Please use MapReduce for the following scenarios and write down the pseudocode. Your pseudocode should start with a *Map* function that takes each line of the *input_file* as the input. The key in the *Map* function is the additional attribute (e.g., "*Student-Table*") in the line, and the value in the *Map* function is the remaining of the line (e.g.,

S001;C001;2021S1). You also need to design *Reduce* function if necessary. You can use multiple MapReduce jobs.

(a) Use MapReduce to collect for each student (represented by *studentID*) the number of distinct courses (represented by *courseID*s) he has registered.

(b) Use MapReduce to collect the courses (represented by *courseID*s) that have more than 50 registered students for at least two semesters. For example, a course will be output if there are 55 students for 2021S1 and 60 students for 2021S2.

(c) Use MapReduce to output every pair of (student, professor) (represented by *studentID* and *professorID*) that the student has attended at least one courses taught by the professor.

CE/CZ 4123 Big Data Management Tutorial 10
Key-Value Stores

Question 1:

Consider a leveling LSM-tree with a size ratio 4. The memory buffer (Level 0) can store 5000 key-value pairs. Initially the LSM-tree is empty. After inserting 70000 key-value pairs with distinct keys continuously, how many levels are formed?

Question 2:

Consider a leveling LSM-tree with a size ratio 4. The LSM-tree has 5 levels (excluding the memory buffer level), and it is incorporated with **both** fence pointers and Bloom Filters. Assume that a key-value pair is always entirely stored within a disk page. Consider the procedure of Get(K) for a key K, which of the followings sequence are possible to be the I/O costs from Level-0 to Level-5? (can select multiple answers)

(a) 1, 1, 1, 1, 1, 1
(b) 0, 1, 1, 1, 1, 1
(c) 0, 0, 1, 0, 0, 1
(d) 0, 0, 0, 0, 0, 1
(e) 1, 0, 0, 0, 0, 0

Question 3:

Consider a leveling LSM-tree with a size ratio 4. The LSM-tree has L levels (excluding the memory buffer level), and it is **only** incorporated with fence pointers (without Bloom Filters). Assume that a key-value pair is always entirely stored within a disk page. Consider the procedure of Get(K) for a key K,

(1) What is the possible I/O cost of accessing Level-i (i is in [1, L])?
(2) If K exists in the LSM-tree, what is the expected I/O cost at Level-i (i is in [1, L])? (hint: divide the cases based on the first-appearing location of the key K)

Question 4:

Consider a leveling LSM-tree with a size ratio 4. The LSM-tree has L levels (excluding the memory buffer level), and it is incorporated with **both** fence pointers and Bloom Filters. Assume that a key-value pair is always entirely stored within a disk page. Consider the procedure of Get(K) for a key K,

(1) What is the possible I/O cost of accessing Level-i (i is in [1, L])?

(2) If K exists in the LSM-tree and the FPR of the Bloom filter at Level-i (i is in [1, L]) is P (P is in [0, 1]), what is the expected I/O cost at Level-i? (hint: divide the cases based on the first-appearing location of the key K)

test 1

Question 1

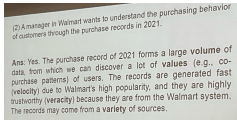
Which of the following are related to big data applications, why?

(1) A computer scientist at MIT is trying to prove NP!=P (note: a famous computer science conjecture).

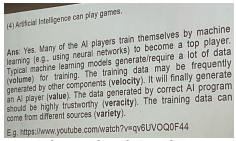
(2) A manager in Walmart wants to understand the purchasing behavior of customers through the purchase records in 2021.

(3) A programmer spends 10 hours in debugging his code.

(4) Artificial Intelligence can play games.


(2) A manager in Walmart wants to understand the purchasing behavior of customers through the purchase records in 2021.

Ans: Yes. The purchase record of 2021 forms a large volume of data, from which we can discover a lot of values (e.g., co-purchase patterns) of users. The records are generated fast (velocity) due to Walmart's high popularity, and they are highly trustworthy (veracity) because they are from the Walmart system. The records may come from a variety of sources.

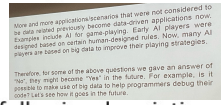

(4) Artificial Intelligence can play games.

Ans: Yes. Many of the AI players train themselves by machine learning (e.g., using neural networks) to become a top player. Typical machine learning models generate/require a lot of data (volume) for training. The training data may be frequently generated by other components (velocity). It will finally generate an AI player (value). The data generated by correct AI program should be highly trustworthy (veracity). The training data can come from different sources (variety).

E.g. https://www.youtube.com/watch?v=qv6UVOQ0F44

Question 2 [Open Question]

Is it possible that an application is not considered a big data application now, but can possibly become a big data application in the future?


More and more applications/scenarios that were not considered to be data related previously become data-driven applications now. Examples include AI for game-playing. Early AI players were designed based on certain human-designed rules. Now, many AI players are based on big data to improve their playing strategies.

Therefore, for some of the above questions we gave an answer of "No", they might become "Yes" in the future. For example, is it possible to make use of big data to help programmers debug their code? Let's see how it goes in the future.

Question 3

For each of the following descriptions, which of the 5V's it is the most related?

(1) There are more than 60000 searches per second in the Google search engine.

(2) Suppose there is a database storing all kinds of enterprise related data, including the ratios of male/female in different companies. Ben wrote the following C program to collect the ratio of male/female for a company.

```
int cal_ratio(){
        int num_Male=getMales();
        int num_Female=getFemales();
        return num_Male/num_Female;
}
```

(3) ImageNet is a 150GB dataset that holds 1,281,167 images for training and 50,000 images for validation, organized in 1,000 categories.

(4) A website allows users to upload different forms of documents such as Excel, JPG, PDF, video.

(5) A researcher performs data mining algorithms on Amazon's purchase records, and he successfully predicts the best seller in the next month.

CE/CZ 4123 Tutorial 2—Data Models

Question 1

Given the following relational schemas containing two tables (a.k.a., relations), and their primary keys are underlined:

Table1(A1, A2, A3)
Table2(A3, B1, B2)

Attribute A3 is the primary key of Table2 and is also a foreign key in Table1.

(1) how do you convert Table 1 **alone** into key-value model? Give one possible solution.

(2) how to convert them into key-value data model? Give one possible solution.

Question 2

Given the following relational schema containing three tables (primary keys are underlined):

Table1(A1, A2)
Table2(B1, B2)
Table3(A1, B1)

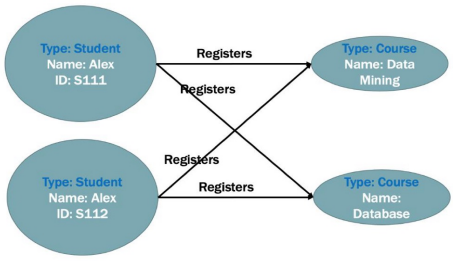
Please give one possible way to convert the above relational model into key-value model.

Question 3

Can key-value model be converted into a relation? Why?

Question 4

Given the following instance based on the graph mode. Please convert it into the instance based on the relational model-based instance. (Hint: converting different types of nodes/edges to tables.)



Question 5 (open discussion)

Why do we need the graph model? Discuss it from the physical storage's perspective.

Q1) 1) key A1 Value A2; A3
2) Step 1 Left Join Table 1 and Table 2 on T1.A3 = T2.A3 to create T3 table

Step 2 determine primary key assuming A1 a
T3 (A1, A2, A3, B1, B2)

Step 3 determine key A1
value A2; A3; B1; B2

Q2 assuming user want to keep all information in Table 3
Step 1 left Join Table 3 with Table 1 based on T3.A1 = T1.A1 then Left Join with Table 2 based on T3.B1 = T2.B1 to create table T4
T4 (A1, B1, A2, B2)
Step 2 determine primary key table A1 and B1
Step 3 determine key as A1; B1 and value as A2; B2

Q3 key value can convert primary key however it cannot cover the other attribute in Value 2

Q4) Student (Name, SID)
Register (SID, CID)
Course (CID, name)

Q5) None of scenario cost lesser to access information through graph stored in the order based on adjacency list

example 2 Student S1, S2
2 Course C1, C2

S1 — C1
S2 — C2

we can check the Course S1 takes, without joining table

CE/CZ 4123 Tutorial 3 – Memory Hierarchy

Question 1

We consider three-layer memory hierarchy, L(i), L(i+1), L(i+2). Their access costs are c1, c2, c3, respectively, with c1<c2<c3. Assume that the data access always checks the existence of data in the order of L(i), L(i+1), and L(i+2). The miss rates of L(i) and L(i+1) are m1 and m2, respectively.

(1) Show that we can estimate the data access cost by c1+m1c2+m1m2c3.

(2) Suppose m1=m2=0.1, show that the over cost is at most 1.11c3.

Question 2

Consider reading data from memory hierarchy consisting of L1-cache, L2-cache, and main memory. Their read access times and hit ratios are given below:

L1-cache:
read access time: 2 nanoseconds; hit ratio: 0.8
L2-cache:
read access time: 8 nanoseconds: hit ratio: 0.9
Main memory:
read access time: 90 nanoseconds.

Please estimate the average data read cost (considering L1, L2 caches and main memory only).

Question 3

Consider the 2nd magic function we mentioned in the lecture, i.e., the magic function that can tell us which pages contain the qualified data. In practice, such magic function is implemented by a certain data structure and hence it incurs some cost when call the function.

Suppose the cost of calling the function is equal to accessing log(N) pages, where N is the number of pages used to store the data. Please give a condition about when is beneficial to use the function.

Question 4

Consider the array-scanning scenario introduced in the lecture. In the lecture, we consider a single query for x>4. In big data systems, many queries are issued together.

Suppose our system needs to handle the following two queries:

1) Select x>4
2) Select x<2

Please explain an efficient way of finishing these two queries together and analyze the number of page accesses needed.

CE/CZ 4123 Big Data Management Tutorial 4
Cache Conscious Designs

Question 1

We have a 12-integer array in main memory as follows

1, 3, 5, 2, 4, 6, 4, 6, 8, 9, 11, 12

Let cache size be the size of 6 integers, and cache line size (transfer unit) be the size of 3 integers. Suppose initially the cache is empty, and there is a program sequentially accessing the whole array. The cache replace mechanism is the same as in the lecture notes: i.e., first cached first evicted.

(1) After the execution of the program, what are the final values stored in the cache? Please give the cache state after every access of the array element.

(2) How many cache hits and cache misses during the program execution? Please give the hit/miss state after every access of the array element.

Question 2

Suppose we have a 2-dimentional integer array A[N][N]. We consider the two ways of array scanning: row-by-row and column-by-column. Let cache size = 5000 integers.

(1) If N>5000 and cache line size (transfer unit)=100 integers, please give a formal analysis of the cache hits/misses of the two ways of writing codes.

(2) If N=250 and cache line size (transfer unit)=500 integers, please give a formal analysis of the cache hits/misses of the two ways of writing codes.

Question 3

We have an 8-integer array A in the main memory. Let cache size be 4 (integers), and cache line size be 2 (integers). Suppose that initially the cache is empty, and the cache replacement policy is the same as the one introduced in the lecture notes, i.e., first cached first evicted.

Let the round-trip scanning be scanning an array in the order from the beginning to the end (i.e., A[0] to A[7]), and then from the end to the beginning (i.e., A[7] to A[0]). How many cache hits and cache misses during the round-trip scanning? Please explain your answer.

Column Stores

Question 1

Given column store table T as follow.

(1) Give the flow chart using "column at a time" for the query "***SELECT*** *min(B)* ***FROM*** *T* ***WHERE*** *A>3 and C>5*"

(2) Give the procedure using "column at a time" for the query "***SELECT*** *sum(C)* ***FROM*** *T* ***WHERE*** *A<10 and B>3*"

A	B	C
1	4	1
3	3	9
5	5	2
2	5	3
3	7	3
1	1	2
4	2	1
6	7	2
1	1	1

Question 2

Redo Question 1 using "vector at a time". Assume that the vector size is 3.

Question 3

Suppose we are querying a **Student's** information table with three columns **Name**, **Email**, **Age**. Given a query of the following form:
"***SELECT*** *Name* ***FROM*** *Student* ***WHERE*** *predicate (Email) and predicate (Age)*".
Here, a predicate applied on a column is a filtering function (e.g., **Email** ending with *ntu.edu.sg,* **Age**>19). We define the selectivity of a predicate by the percentage of the qualified results in the corresponding column. Assume that the selectivity of predicate(**Email**) is *p* and the selectivity of predicate(**Age**) is *q*, where 0<*p*<1, and 0<*q*<1. Let page size be *P*. We assume column width *w* < *P* and each value in a column is contained in a page. Consider two options in scanning columns: scanning **Email** first and scanning **Age** first.

(1) If the column widths are the same (denoted by *w*), please analyze which is better.

(2) If the width of **Email** is *2w*, and the widths for **Name** and **Age** are *w*, then which option is better?

CE/CZ 4123 Big Data Management Tutorial 7

More Practice for Preparing for Quizzes

1. Given the following three tables (primary keys are underlined):

   Employee(EID, Salary)

   Manager(MID, Salary)

   Employee-Manager(EID, MID)

   Each manager supervises at least one employees. Employee-Manager is a table that contains the manager ID (i.e., MID) for each employee (i.e., EID). How to convert the relational data model to a key-value data model? Consider that the main purpose of the conversion is for the query "Given an employee ID, find the salary of the employee's manager". The conversion should retain the information as much as possible.

Q1) Step 1: Left Join Employee-manager table with Employee table based on EM.EID = E.EID and then left Join the table with Manager table based on EM.MID = M.MID (this ensure we retain all EID)

Step 2: make EID the primary key of Confine result and the rest as Value

T(EID, MID, E_Salary, M_Salary)

Step 3: Convert to key-Value data model with key = EID, Value = MID; E.Salary; M.Salary

2. Consider reading data from memory hierarchy consisting of L1 Cache, L2 Cache, and main memory with the following parameters.

   - L1 Cache:

     Read access time: 2 nanoseconds

     Miss ratio: 0.4

   - L2 Cache:

     Read access time: 10 nanoseconds

     Miss ratio: 0.2

   - Main memory:

     Read access time: 100 nanoseconds.

   Estimate the average data read cost and explain your answer. (Note: consider L1, L2 caches and main memory only).

Q2) 2ns + 0.4 · 10ns + 0.4 · 0.2 100ns

2 + 4 + 8 = 14ns

3. In the lecture, we introduced a cost-free magic function telling which pages locate the qualified data for a query. Consider a disk page size is 1024 integers. There are 10240 integers, which are 1, 2, 3, …, 10240, sequentially stored at 10 consecutive disk pages.

   Consider the following 4 queries using 4 scans over the data, where each query range is decided by three integers x, y, z, and 1<x<y<z<10240.

   Query 1: searching values in the range [1, *x*] (i.e., values at least 1 and at most *x*)
   Query 2: searching values in the range [*x*+1, *y*]
   Query 3: searching values in the range [*y*+1, *z*]
   Query 4: searching values in the range [*z*+1, 10240]

   List **all possible** total number of read I/Os needed for the 4 scans, *with* the magic function. Please explain your answer.

Q3) 10240 int, page size 1024

Query 1 to 4, best case needed 10 scans (access 10 pages on disk, all 10240 integers)

as all the range in query 1 to 4 add up to be

1 to 10240 (x, y, z) however values could be in the middle of a page

0 —— 1024

1 2

same page need to be access again in different time

the worse scenario is when x, y, z are all in middle of the pages

0 —— 1024

page page ... page

which requires 3 extra I/Os

which add up to 13 I/Os

4. We have a 32-integer array *A* in the main memory. Let cache size be 16 (integers), and cache line size be 4 (integers). Suppose that initially the cache is empty, and the cache replacement policy is the same as the one introduced in the lectures, i.e., first cached first evicted. Let the *f-trip* and *b-trip* scanning be the following.

```
f-trip(){
    for (int j=0; j<32; j++){
        Access A[j]://Access does not change the data
    }
}
b-trip(){
    for (int j=0; j<32; j++){
        Access A[31-j];
    }
}
```

   If we need to do 99 scans of the array, and we can select each scan to be either f-trip or b-trip. Please give one best selection strategy that gives the minimum number of misses and explain your answer. Please also compute the number of cache hits and cache misses in the best strategy.

Q4) 32 int array, cache size = 16 int, cache line = 4 int, need to do 99 scans

each scans

Each 1 miss, brings 3 hits

first scan total miss = 1/4 · 32 = 8 miss

using f-trip total hit = 3/4 · 32 = 24 hits

Second scan use B-trip() so the first 16 int will be a hit

the subsequent 16 int follows 1 miss, 7 hit

total miss = 1/4 · 16 = 4 miss

total hits = 3/4 · 16 = 12 hits

after that switch back to f-trip() then B-trip() to alternate and get the first 16 hits at start of every scans

in total

total hits = first scan + subsequent scan = 24 + 98(16+12) = 2768 hits

total miss = 8 + 98(4) = 400 misses

CE/CZ 4123 Big Data Management Tutorial 7
Distributed Systems and MapReduce

Question 1

Amazon wants to estimate the Top-*K* best sold products from *S* purchase records of *L* products in the form of a list of (User id, Product id) pairs. Assume that *L* is a multiple of *M*. Suppose there is a distributed system with 1 master machine and *M* slave machines. Design a distributed computation procedure to finish the task. Please describe

(1) how the data is distributed, computed and aggregated?

(2) how much data is sent across machines?

Question 2:

Consider the MapReduce paradigm and answer the following questions.

(1) In a MapReduce job, the output of Map phase is a list of key-value pairs: (A, 1) (C, 2), (A, 5), (C, 6), (B, 3), (E, 3), (C, 8). Please list the possible input to the Reduce function.

(2) Based on the answer to Q2(a), write a Reduce function (in pseudocode) so that the MapReduce output is: (2, A), (3, C), (6, A), (7, C), (4, B), (4, E), (9, C).

(3) Consider an employee table containing three columns (EmployeeID, age, monthly-salary) where age and monthly-salary are integers. Use MapReduce to collect the number of employees falling into each of the following two categories:

- Category 1: The age of the employee is between 30 and 40 (including 30 and 40). His/her monthly salary is at most 7000.
- Category 2: The age of the employee is between 40 and 50 (including 40 and 50). His/her monthly salary is more than 7000.

Please use only one MapReduce Job to achieve this task and write down the pseudocode of the Map function and Reduce function. The input key and value for Map function are an employee's age and monthly-salary respectively.

(Example: if there are 100 employees in Category 1 and 50 employees in Category 2, then the MapReduce output will contain two key-value pairs: (1, 100), (2, 50))

Question 3:

Design MapReduce algorithms for the multiplication of two matrices A, B of n by n. Elements in the matrices are integers. The input key for *Map* function is *MatrixName*; the input value for *Map* function is in the form of *i;j;v*, indicating that the value of the *i*-th row and *j*-th column is *v*.

(Matrix Multiplication: Given matrix A[nxn] and matrix B[nxn], compute matrix C such that $C[i][z] = \sum_{j=0}^{n-1} A[i][j] \times B[j][z]$, for i,z in [0, n-1]).

(1) Use at most two MapReduce jobs to finish the computation.
(2) Furthermore, can the multiplication be finished using one MapReduce job?

Big Data Management Tutorial - MapReduce

Consider a *Student* table containing three columns (*studentID*, *courseID*, *semester*). Each tuple in the *Student* table records that a student registered for the course in the corresponding semester. Also consider a *Professor* table containing three columns (*professorID*, *courseID*, *semester*). Each tuple in the *Professor* table records that a professor teaches a course in the corresponding semester. A course may open in multiple semesters. If a student fails a course, he may retake the course in later semesters. **Example tuples** for the two tables are as follows.

studentID	courseID	semester
S001	C001	2021S1
S002	C001	2021S1
S002	C002	2021S2
S003	C001	2021S2

Table Q4.1: Example Student Table

professorID	courseID	semester
P001	C001	2021S1
P002	C002	2021S1
P001	C001	2021S2

Table Q4.2: Example Professor Table

The *Professor* table and the *Student* table are stored together in a file named *input_file*, with each tuple per line. There is an additional attribute to indicate whether this tuple is from the *Student* Table or the *Professor* Table. Based on the above example tuples, the file content is as follows.

Student-Table S001;C001;2021S1
Student-Table S002;C001;2021S1
Student-Table S002;C002;2021S2
Student-Table S003;C001;2021S2
Professor-Table P001;C001;2021S1
Professor-Table P002;C002;2021S1
Professor-Table P001;C001;2021S2

Please use MapReduce for the following scenarios and write down the pseudocode. Your pseudocode should start with a *Map* function that takes each line of the *input_file* as the input. The key in the *Map* function is the additional attribute (e.g., "*Student-Table*") in the line, and the value in the *Map* function is the remaining of the line (e.g., *S001;C001;2021S1*). You also need to design *Reduce* function if necessary. You can use multiple MapReduce jobs.

(a) Use MapReduce to collect for each student (represented by *studentID*) the number of distinct courses (represented by *courseID*s) he has registered.

(b) Use MapReduce to collect the courses (represented by *courseID*s) that have more than 50 registered students for at least two semesters. For example, a course will be output if there are 55 students for 2021S1 and 60 students for 2021S2.

(c) Use MapReduce to output every pair of (student, professor) (represented by *studentID* and *professorID*) that the student has attended at least one courses taught by the professor.

CE/CZ 4123 Big Data Management Tutorial 10
Key-Value Stores

Question 1:

Consider a leveling LSM-tree with a size ratio 4. The memory buffer (Level 0) can store 5000 key-value pairs. Initially the LSM-tree is empty. After inserting 70000 key-value pairs with distinct keys continuously, how many levels are formed?

Question 2:

Consider a leveling LSM-tree with a size ratio 4. The LSM-tree has 5 levels (excluding the memory buffer level), and it is incorporated with **both** fence pointers and Bloom Filters. Assume that a key-value pair is always entirely stored within a disk page. Consider the procedure of Get(K) for a key K, which of the followings sequence are possible to be the I/O costs from Level-0 to Level-5? (can select multiple answers)

(a) 1, 1, 1, 1, 1, 1
(b) 0, 1, 1, 1, 1, 1
(c) 0, 0, 1, 0, 0, 1
(d) 0, 0, 0, 0, 0, 1
(e) 1, 0, 0, 0, 0, 0

Question 3:

Consider a leveling LSM-tree with a size ratio 4. The LSM-tree has L levels (excluding the memory buffer level), and it is **only** incorporated with fence pointers (without Bloom Filters). Assume that a key-value pair is always entirely stored within a disk page. Consider the procedure of Get(K) for a key K,

(1) What is the possible I/O cost of accessing Level-i (i is in [1, L])?
(2) If K exists in the LSM-tree, what is the expected I/O cost at Level-i (i is in [1, L])? (hint: divide the cases based on the first-appearing location of the key K)

Question 4:

Consider a leveling LSM-tree with a size ratio 4. The LSM-tree has L levels (excluding the memory buffer level), and it is incorporated with **both** fence pointers and Bloom Filters. Assume that a key-value pair is always entirely stored within a disk page. Consider the procedure of Get(K) for a key K,

(1) What is the possible I/O cost of accessing Level-i (i is in [1, L])?
(2) If K exists in the LSM-tree and the FPR of the Bloom filter at Level-i (i is in [1, L]) is P (P is in [0, 1]), what is the expected I/O cost at Level-i? (hint: divide the cases based on the first-appearing location of the key K)

```
0.
sudo apt install openjdk-8-jdk-headless
---
1.1.
sudo apt install ssh pdsh
cd ~/Downloads
wget https://downloads.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz
tar xzf hadoop-3.4.0.tar.gz
sudo mv hadoop-3.4.0 /usr/share/hadoop
---
1.2.
gedit ~/.bashrc

export PDSH_RCMD_TYPE=ssh
export HADOOP_HOME=/usr/share/hadoop
export PATH=$PATH:$HADOOP_HOME:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

gedit $HADOOP_HOME/etc/hadoop/hadoop-env.sh
export JAVA_HOME="/usr/lib/jvm/java-8-openjdk-amd64/"

hadoop jar /usr/share/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.4.0.jar pi 16 1000
---
2.
cd ~/.ssh
rm ./id_rsa*
ssh-keygen -t rsa
ssh-copy-id -i ~/.ssh/id_rsa.pub localhost
ssh localhost
exit
---
3.
gedit $HADOOP_HOME/etc/hadoop/core-site.xml

<configuration>
    <property>
        <name>fs.defaultFS</name>
        <value>hdfs://localhost:9000</value>
    </property>
    <property>
        <name>hadoop.tmp.dir</name>
        <value>file:/usr/share/hadoop/tmp</value>
    </property>
</configuration>

gedit $HADOOP_HOME/etc/hadoop/hdfs-site.xml

<configuration>
    <property>
        <name>dfs.replication</name>
        <value>1</value>
    </property>
    <property>
        <name>dfs.namenode.name.dir</name>
        <value>file:/usr/share/hadoop/tmp/dfs/name</value>
    </property>
    <property>
        <name>dfs.datanode.data.dir</name>
        <value>file:/usr/share/hadoop/tmp/dfs/data</value>
    </property>
    <property>
        <name>dfs.namenode.datanode.registration.ip-hostname-check</name>
        <value>false</value>
    </property>
</configuration>

gedit $HADOOP_HOME/etc/hadoop/mapred-site.xml

<configuration>
    <property>
        <name>mapreduce.framework.name</name>
```

```
        <value>yarn</value>
    </property>
    <property>
        <name>mapreduce.application.classpath</name>
        <value>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/lib/*</value>
    </property>
</configuration>

gedit $HADOOP_HOME/etc/hadoop/yarn-site.xml

<configuration>
    <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
    </property>
    <property>
        <name>yarn.nodemanager.env-whitelist</name>
        <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_HOME,PATH,LANG,TZ,HADOOP_MAPRED_HOME</value>
    </property>
</configuration>
---
4.
hdfs namenode -format -force
start-dfs.sh
start-yarn.sh

jps
http://localhost:9870/
http://localhost:8088/

hdfs dfs -mkdir -p /user/hadoop
hdfs dfs -mkdir -p /user/input
hdfs dfs -put /usr/share/hadoop/etc/hadoop/*.xml /user/input

hadoop jar /usr/share/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.4.0.jar grep /user/input output 'dfs[a-z.]+'
hdfs dfs -cat output/*
```

BIG DATA MANAGEMENT

CZ/CE4123

Tutorial – Big data 5V's

QUESTION1

Which of the following are related to big data applications, why?

(1) A computer scientist at MIT is trying to prove NP!=P (note: a famous computer science conjecture).

(2) A manager in Walmart wants to understand the purchasing behavior of customers through the purchase records in 2021.

(3) A programmer spends 10 hours in debugging his code.

(4) Artificial Intelligence can play games.

(1) A computer scientist at MIT is trying to prove NP!=P (note: a famous computer science conjecture).

(1) A computer scientist at MIT is trying to prove NP!=P (note: a famous computer science conjecture).

Ans: No. This is not that related to big data applications. It requires mathematical skills.

(2) A manager in Walmart wants to understand the purchasing behavior of customers through the purchase records in 2021.

(2) A manager in Walmart wants to understand the purchasing behavior of customers through the purchase records in 2021.

Ans: Yes. The purchase record of 2021 forms a large **volume** of data, from which we can discover a lot of **values** (e.g., co-purchase patterns) of users. The records are generated fast (**velocity**) due to Walmart's high popularity, and they are highly trustworthy (**veracity**) because they are from the Walmart system. The records may come from a **variety** of sources.

(3) A programmer spends 10 hours in debugging his code.

(3) A programmer spends 10 hours in debugging his code.

Ans: No. This is about a programmer using his own knowledge in debugging. Not related to (big) data.

(4) Artificial Intelligence can play games.

(4) Artificial Intelligence can play games.

Ans: Yes. Many of the AI players train themselves by machine learning (e.g., using neural networks) to become a top player. Typical machine learning models generate/require a lot of data (**volume**) for training. The training data may be frequently generated by other components (**velocity**). It will finally generate an AI player (**value**). The data generated by correct AI program should be highly trustworthy (**veracity**). The training data can come from different sources (**variety**).

E.g. https://www.youtube.com/watch?v=qv6UVOQ0F44

(4) Artificial Intelligence can play games.

Note: It is also acceptable to say that some AI is not related to big data. For example, some earlier AI uses rule-based algorithms.

QUESTION2

[Open Question]:
Is it possible that an application is not considered a big data application now, but can possibly become a big data application in the future?

Please also give some concrete examples during the discussion.

More and more applications/scenarios that were not considered to be data related previously become data-driven applications now. Examples include AI for game-playing. Early AI players were designed based on certain human-designed rules. Now, many AI players are based on big data to improve their playing strategies.

Therefore, for some of the above questions we gave an answer of “No”, they might become “Yes” in the future. For example, is it possible to make use of big data to help programmers debug their code? Let’s see how it goes in the future.

QUESTION3

Question 3:

For each of the following descriptions, which of the 5V’s is the most related?

QUESTION3

(1) There are more than 60000 searches per second in the Google search engine.

QUESTION3

(1) There are more than **60000 searches per second** in the Google search engine.

Ans: **60000 searches per second → velocity**

QUESTION3

(2) Suppose there is a database storing all kinds of enterprise related data, including the ratios of male/female in different companies. Ben wrote the following C program to collect the ratio of male/female for a company.

```
int cal_ratio(){
  int num_Male=getMales();
  int num_Female=getFemales();
  return num_Male/num_Female;
}
```

QUESTION3

(2) Suppose there is a database storing all kinds of enterprise related data, including the ratios of male/female in different companies. Ben wrote the following C program to collect the ratio of male/female for a company.

```
int cal_ratio(){
  int num_Male=getMales();
  int num_Female=getFemales();
  return num_Male/num_Female;
}
```

Suppose num_Male=1500, num_Female=1000, then the program returns “1” instead of “1.5”.

Ans: The code in red indicates that the ratio is not a fractional number. It is a bug in the algorithm, which outputs biased data. → Veracity

QUESTION3

(3) ImageNet is a 150GB dataset that holds 1,281,167 images for training and 50,000 images for validation, organized in 1,000 categories.

QUESTION3

(3) ImageNet is a 150GB dataset that holds 1,281,167 images for training and 50,000 images for validation, organized in 1,000 categories.

Ans: 150GB dataset that holds 1,281,167 images → Volume

(4) A website allows users to upload different forms of documents such as Excel, JPG, PDF, video.

(4) A website allows users to upload different forms of documents such as Excel, JPG, PDF, video.

Ans: Structured data (Excel) and Unstructured data(video, PDF, JPG)→Variety

(5) A researcher performs data mining algorithms on Amazon’s purchase records, and he successfully predicts the best seller in the next month.

(5) A researcher performs data mining algorithms on Amazon's purchase records, and he successfully predicts the best seller in the next month.

Ans: Data mining → Value

BIG DATA MANAGEMENT

CZ/CE4123

Tutorial 2
Data Models

QUESTION1

Given the following relational schemas containing two tables (a.k.a., relations), and their primary keys are underlined:

Table1(<u>A1</u>, A2, A3)
Table2(<u>A3</u>, B1, B2)

Attribute A3 is the primary key of Table2 and is also the foreign key in Table1.

1) how do you convert Table 1 **alone** into key-value model? Give one possible solution.

2) how to convert them into key-value data model? Give one possible solution.

QUESTION1

Given the following relational schemas containing two tables (a.k.a., relations), and their primary keys are underlined:

Table1(A1, A2, A3)
Table2(A3, B1, B2)

Attribute A3 is the primary key of Table2 and is also the foreign key of Table2.

1) how do you convert Table 1 **alone** into key-value model? Give one possible solution.

Ans:
Key: A1
Value: A2;A3

QUESTION1

Given the following relational schemas containing two tables (a.k.a., relations), and their primary keys are underlined:

Table1(A1, A2, A3)
Table2(A3, B1, B2)

Attribute A3 is the primary key of Table2 and is also the foreign key of Table1. (You may assume that the query is always issued with respect to A1)

2) how to convert them into key-value data model? Give one possible solution.

Ans:
Step 1: (Left) join Table 1 and Table 2 using attribute A3, so that we have a bigger table T3(A1, A2, A3, B1, B2).
Step 2: Key: A1 Value: A2;A3;B1;B2

QUESTION2

Given the following relational schema containing three tables (primary keys are underlined):

Table1(A1, A2)

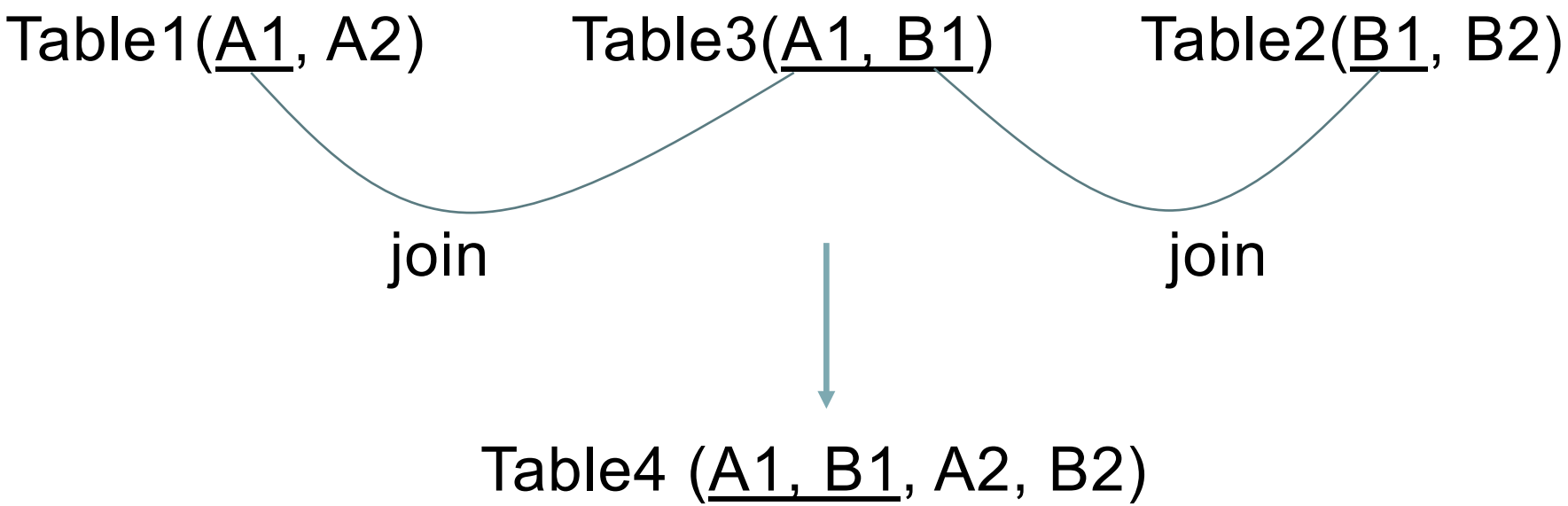
Table2(B1, B2)

Table3(A1, B1)

Please give one possible way to convert the above relational model into key-value model.
(You may assume that the query is always issued with respect to A1,B1)

QUESTION2

Given the following relational schema containing three tables (primary keys are underlined):

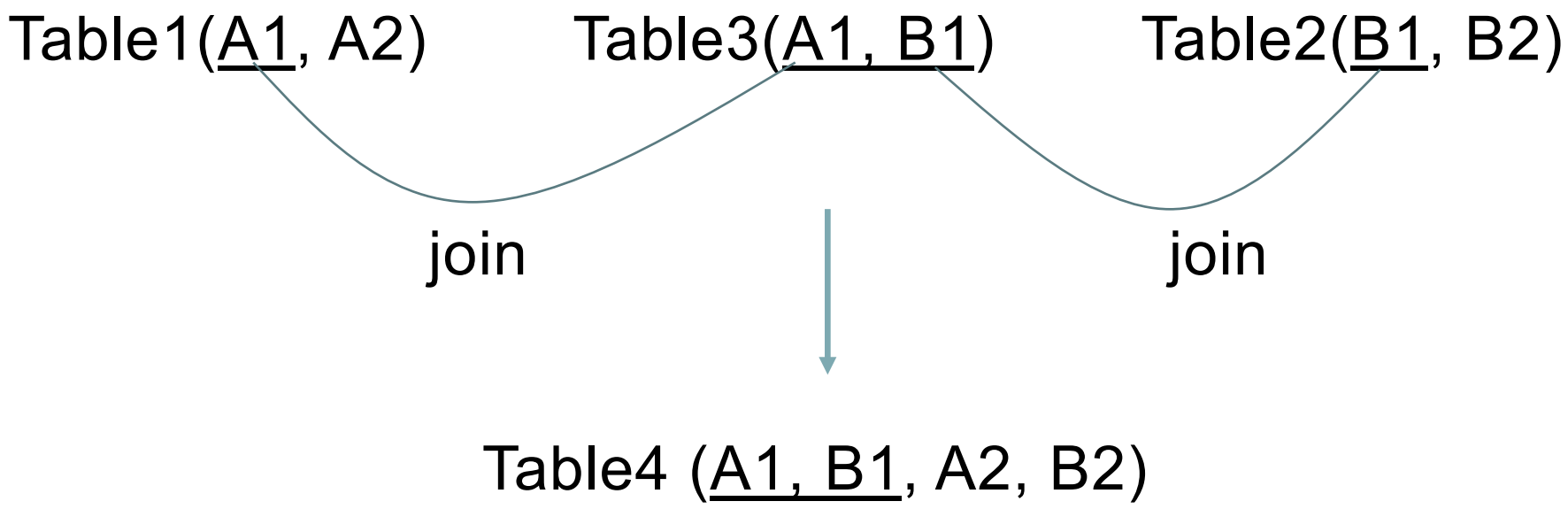


QUESTION2

Assume that the user is interested in querying the information for the combination of (A1,B1), then

Key: A1;B1

Value: A2;B2.



QUESTION 3

Can key-value model be converted into a relation? Why?

QUESTION 3

Can key-value model be converted into a relation? Why?

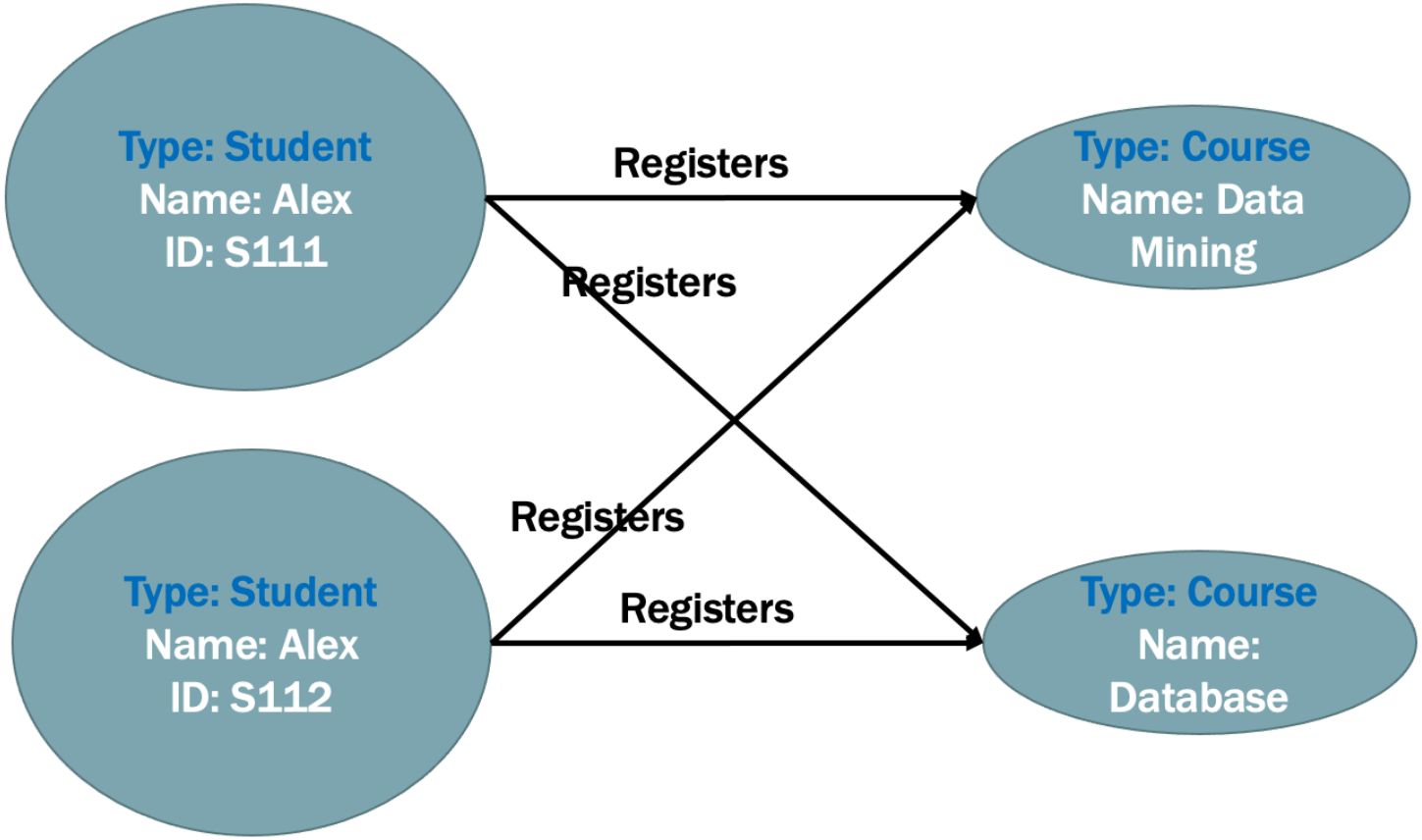
Ans: Key-value model can be trivially converted into a relation as follows

Table(<u>key</u>, value)

Note: This solution, however, is not able to uncover any schema that exists in **value**.

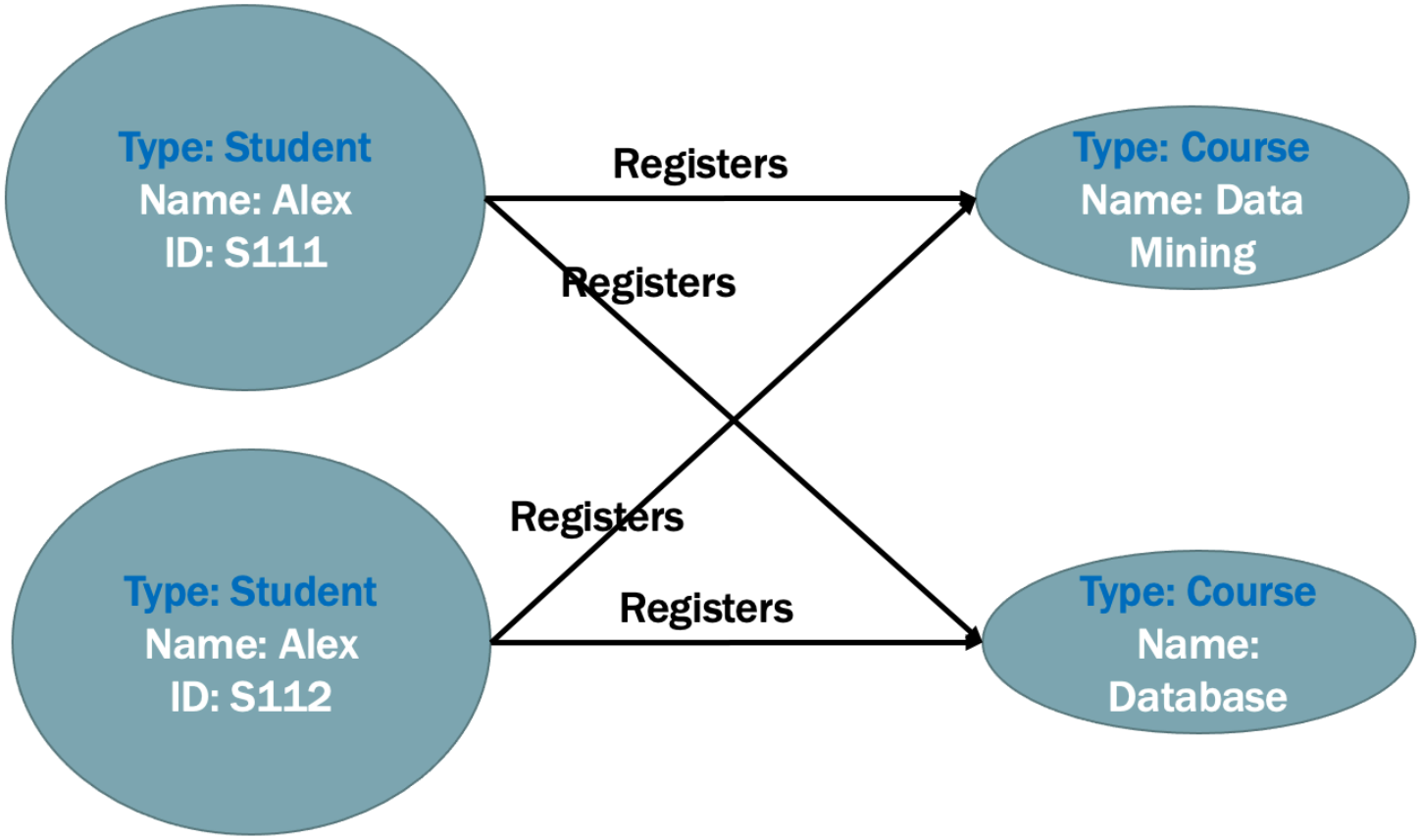
QUESTION 4

Given the following graph model, please convert it into relational data model.



QUESTION 4

Given the following graph model, please convert it into relational data model.

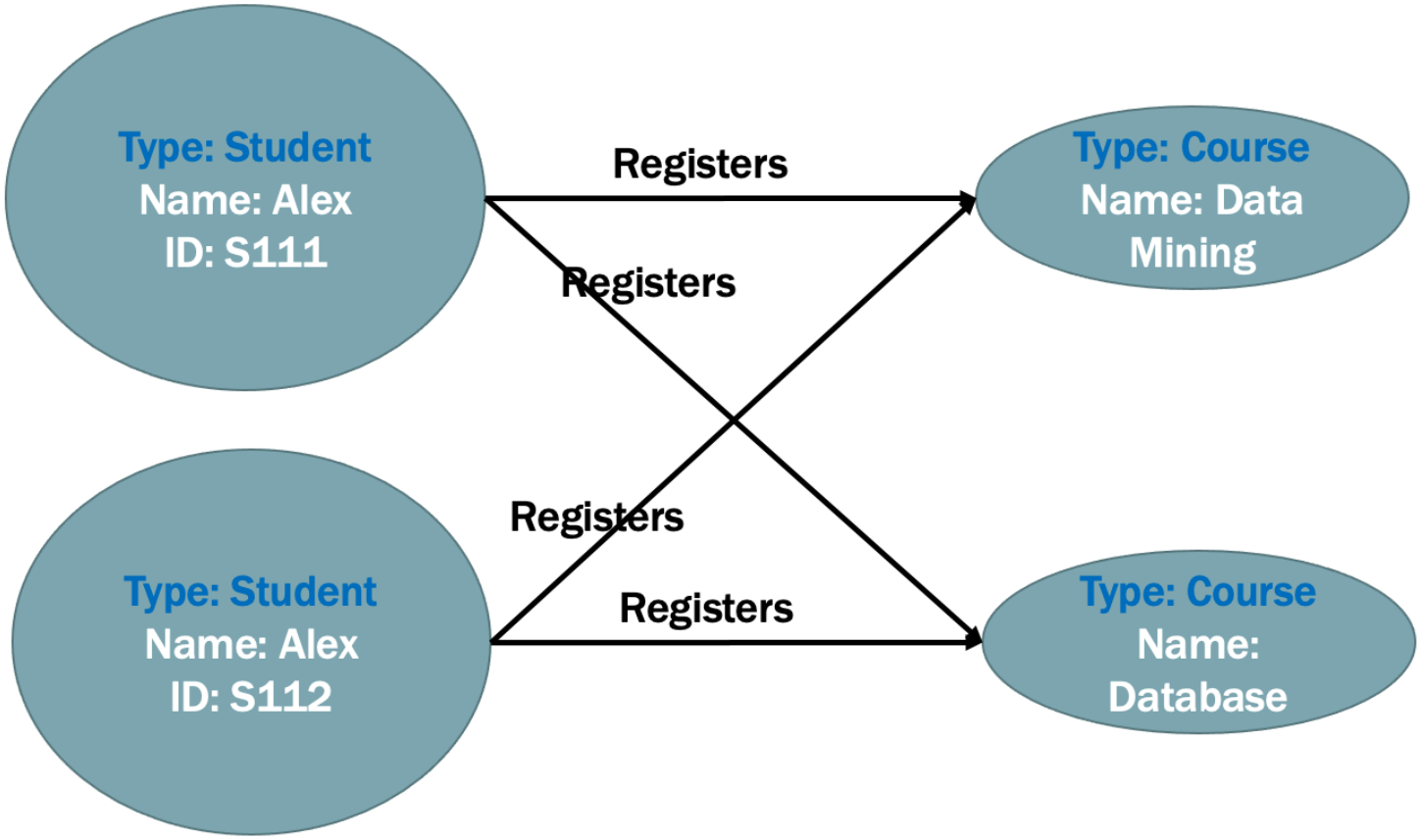


Step 1: consider how many tables are needed.

- ❑ *Type Student*→ **Student** Table
- ❑ *Type Course*→ **Course** Table

QUESTION 4

Given the following graph model, please convert it into relational data model.

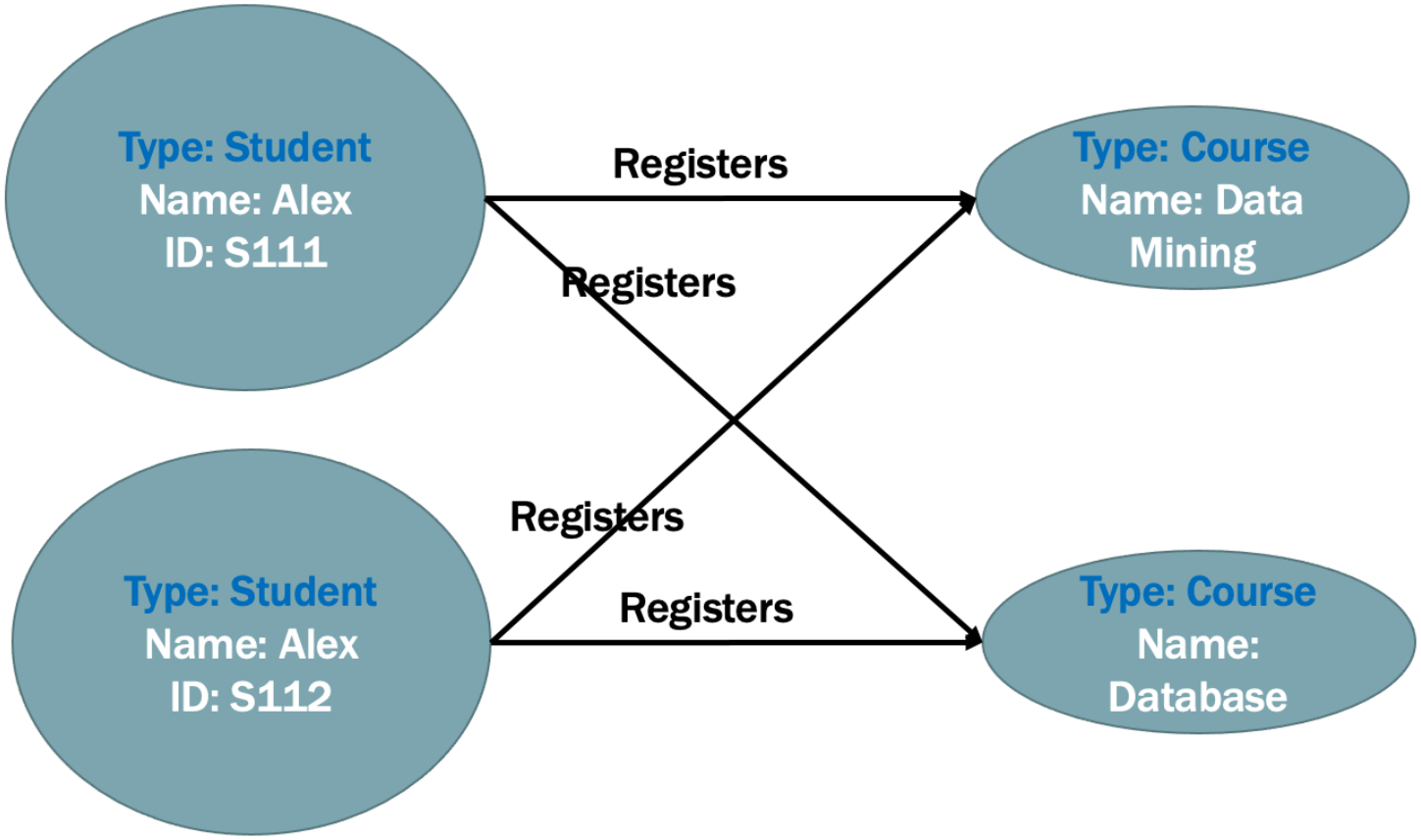


Step 1: consider how many tables are needed.

- ❑ *Type "Student"*→ **Student** Table
- ❑ *Type "Course"*→ **Course** Table
- ❑ *Relationship "Registers"*→ **Register** Table

QUESTION 4

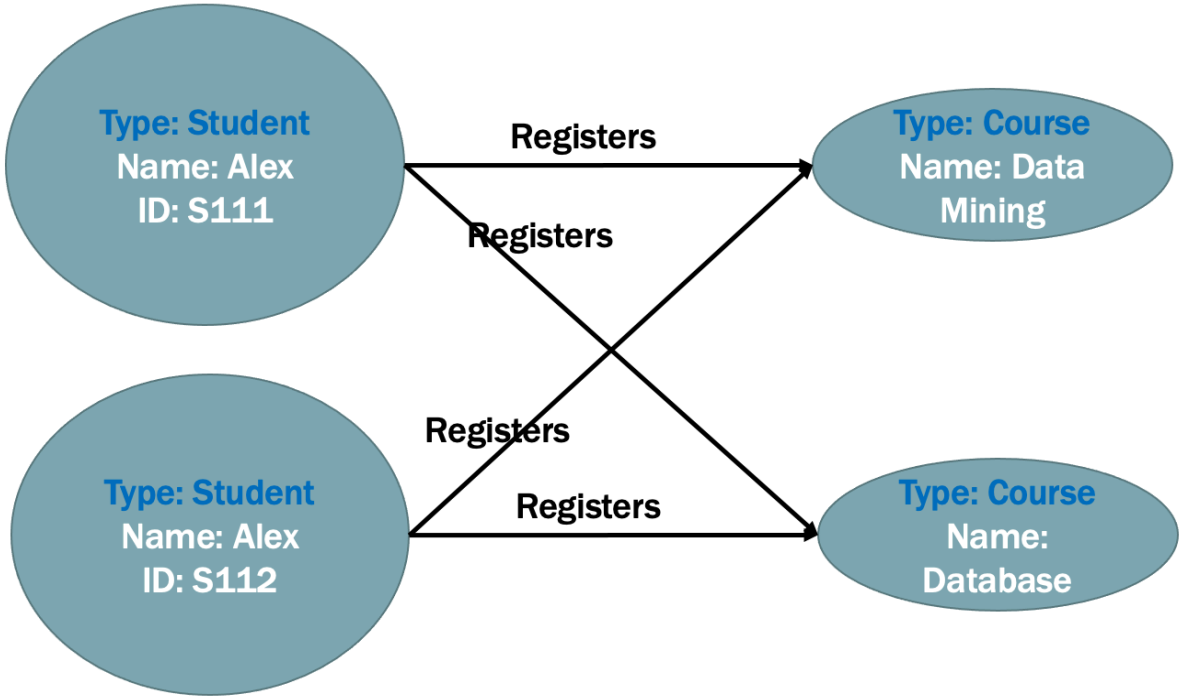
Given the following graph model, please convert it into relational data model.



Step 2: find attributes for each table.

Student (Name, ID)
Course (**Name**)
Register (**StudentID, CourseName**)

QUESTION 4



Student

ID	Name
S111	Alex
S112	Alex

Course

Name
Data Mining
Database

Register

StudentID	Name
S111	Data Mining
S111	Database
S112	Data Mining
S112	Database

QUESTION 5

Why we need the graph model? Discuss it from the physical storage's perspective.

Ans: Graphs can be stored in the order based on adjacency list. In fundamental graph related operations such as graph traversal, the data access order aligns with the adjacency list. In contrast, relational tables often require costly “join” for these operations.

Example: Reconsider the example for Q4, we have two students S1, S2, and two courses C1, C2.
Nodes: S1, S2, C1, C2; **Edges**: (S1, C1), (S1, C2), (S2, C1), (S2, C2)

Adjacency list:

S1’s neighbor list: C1, C2
S2’s neighbor list: C1, C2
C1’s neighbor list: S1, S2 (if we do not consider the edge direction)
C2’s neighbor list: S1, S2 (if we do not consider the edge direction)

BIG DATA MANAGEMENT

CZ/CE4123

Tutorial 3
Memory Hierarchy

QUESTION1

We consider three-layer memory hierarchy, L(i), L(i+1), L(i+2). Their access costs are c1, c2, c3, respectively, with c1<c2<c3. Assume that the data access always checks the existence of data in the order of L(i), L(i+1), and L(i+2). The miss rates of L(i) and L(i+1) are m1 and m2, respectively.

(1) Show that we can estimate the data access cost by c1+m1c2+m1m2c3.

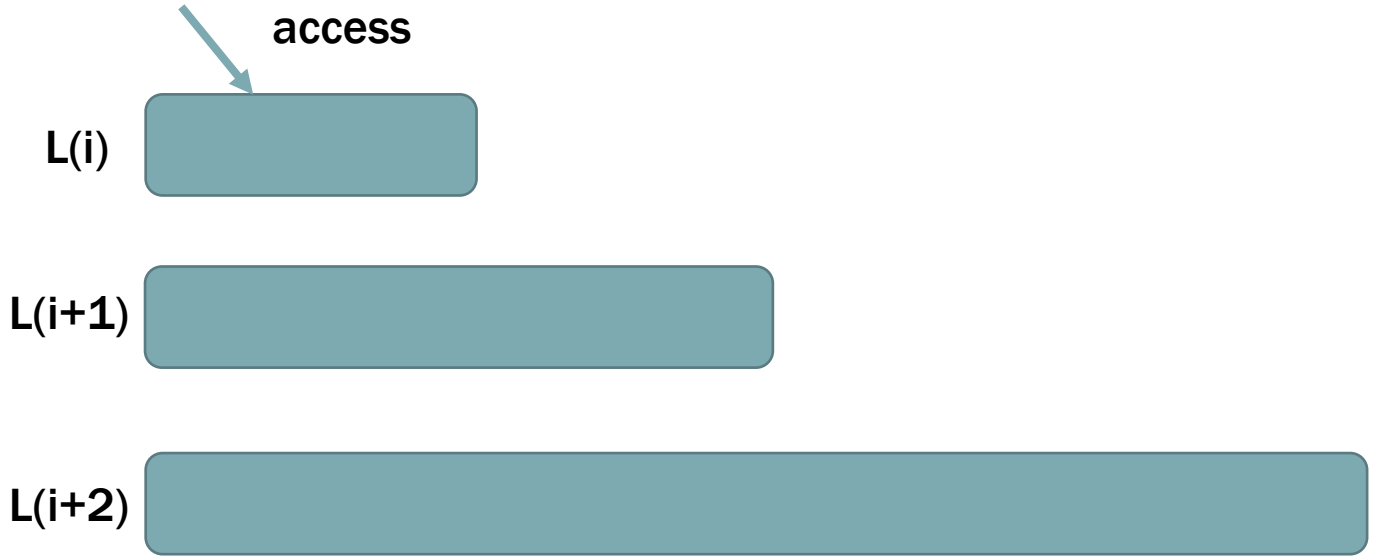
(2) Suppose m1=m2=0.1, show that the overall cost is at most 1.11c3.

QUESTION1

We consider three-layer memory hierarchy, L(i), L(i+1), L(i+2). Their access costs are c1, c2, c3, respectively, with c1<c2<c3. Assume that the data access always checks the existence of data in the order of L(i), L(i+1), and L(i+2). The miss rates of L(i) and L(i+1) are m1 and m2, respectively.

(1) Show that we can estimate the data access cost by c1+m1c2+m1m2c3.

Ans:
First access L(i), having a cost c1;



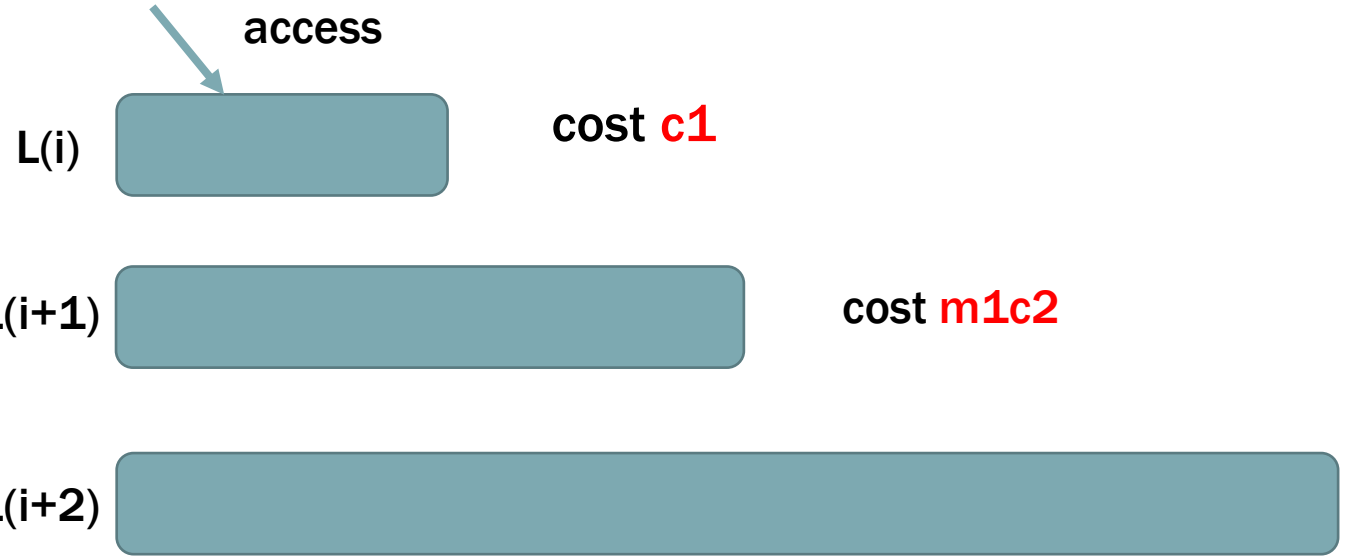
QUESTION1

We consider three-layer memory hierarchy, L(i), L(i+1), L(i+2). Their access costs are c1, c2, c3, respectively, with c1<c2<c3. Assume that the data access always checks the existence of data in the order of L(i), L(i+1), and L(i+2). The miss rates of L(i) and L(i+1) are m1 and m2, respectively.

(1) Show that we can estimate the data access cost by c1+m1c2+m1m2c3.

Ans:
First access L(i), having a cost c1;

Then, if there is an L(i) miss, we need to access L(i+1). This happens with probability m1 (miss rate) and a cost of c2.



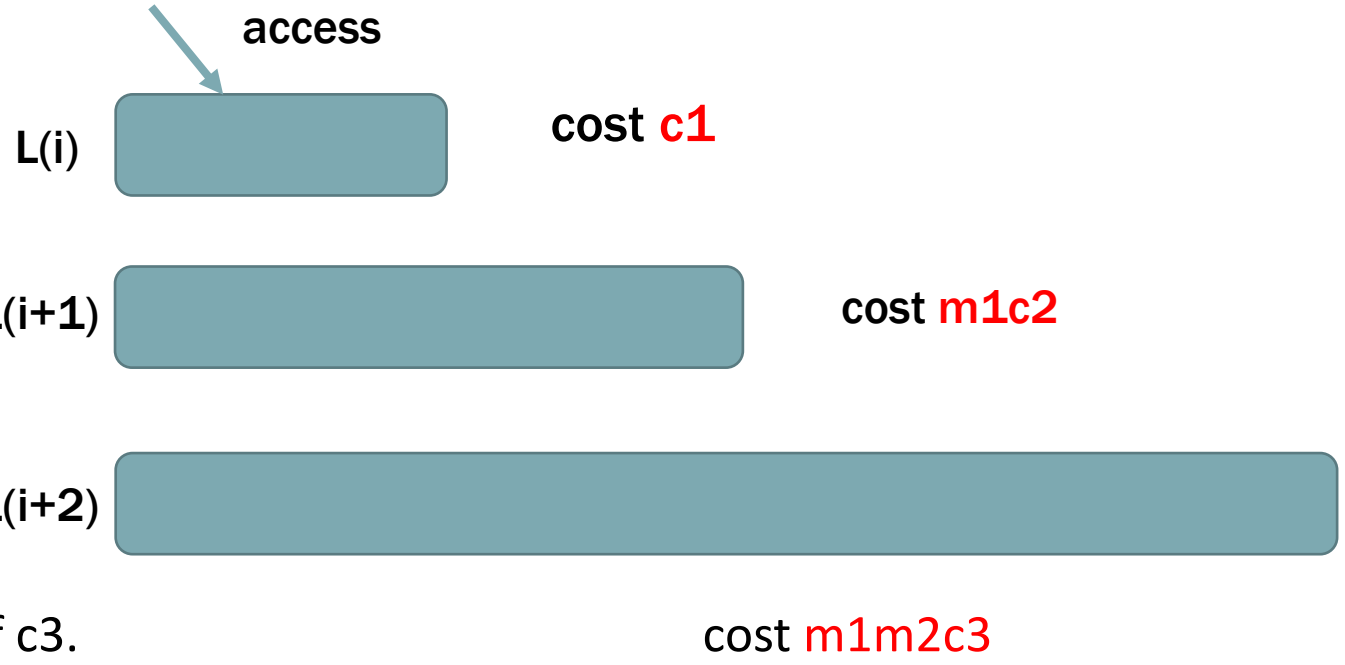
QUESTION1

We consider three-layer memory hierarchy, L(i), L(i+1), L(i+2). Their access costs are c1, c2, c3, respectively, with c1<c2<c3. Assume that the data access always checks the existence of data in the order of L(i), L(i+1), and L(i+2). The miss rates of L(i) and L(i+1) are m1 and m2, respectively.

(1) Show that we can estimate the data access cost by c1+m1c2+m1m2c3.

Ans:

First access L(i), having a cost c1;



Then, if there is an L(i) miss, we need to access L(i+1). This happens with probability m1 (miss rate) and a cost of c2.

Then, if there is an L(i+1) miss, we need to access L(i+2). This happens with probability m1m2 (both L(i) and L(i+1) miss) and a cost of c3.

QUESTION1

We consider three-layer memory hierarchy, L(i), L(i+1), L(i+2). Their access costs are c1, c2, c3, respectively, with c1<c2<c3. Assume that the data access always checks the existence of data in the order of L(i), L(i+1), and L(i+2). The miss rates of L(i) and L(i+1) are m1 and m2, respectively.

(1) Show that we can estimate the data access cost by c1+m1c2+m1m2c3.

The other solution:

The case for L(i) hit: (1-m1)c1

The case for L(i) miss and L(i+1) hit: m1(1-m2)(c1+c2)

The case for L(i) miss and L(i+1) miss: m1m2(c1+c2+c3)

Hence, the total cost is (1-m1)c1+m1(1-m2)(c1+c2)+m1m2(c1+c2+c3)=c1+m1c2+m1m2c3

DETAILS

$$
\begin{aligned}
&(1-m_1)c_1 + m_1(1-m_2)(c_1+c_2) + m_1m_2(c_1+c_2+c_3) \\
=&(1-m_1)c_1 + m_1(1-m_2)c_1 + m_1(1-m_2)c_2 + m_1m_2c_1 + m_1m_2c_2 + m_1m_2c_3 \\
=&(1-m_1+m_1-m_1m_2+m_1m_2)c1 + (m_1-m_1m_2+m_1m_2)c_2 + m_1m_2c_3 \\
=&c_1 + m_1c_2 + m_1m_2c_3
\end{aligned}
$$

QUESTION1

We consider three-layer memory hierarchy, L(i), L(i+1), L(i+2). Their access costs are c1, c2, c3, respectively, with c1<c2<c3. Assume that the data access always checks the existence of data in the order of L(i), L(i+1), and L(i+2). The miss rates of L(i) and L(i+1) are m1 and m2, respectively.

(2) Suppose m1=m2=0.1, show that the overall cost is at most 1.11c3.

Ans:

$$
\begin{aligned}
& c1+m1c2+m1m2c3 \\
= \quad & c1+ 0.1c2+ 0.01c3 \\
< \quad & c3+0.1c3+0.01c3 \\
= \quad & 1.11c3
\end{aligned}
$$

ADDITIONAL DISCUSSION

What about having more than 3 layers?

QUESTION2

Consider reading data from memory hierarchy consisting of L1-cache, L2-cache and main memory. Their read access times and hit rates are given below:

L1-cache:
read access time: 2 nanoseconds; hit rate: 0.8
L2-cache:
read access time: 8 nanoseconds: hit rate: 0.9
Main memory:
read access time: 90 nanoseconds.

Please estimate the average data read cost (considering L1, L2 caches and main memory only).

QUESTION2

Consider reading data from memory hierarchy consisting of L1-cache, L2-cache and main memory. Their read access times and hit ratios are given below:

L1-cache:
read access time: 2 nanoseconds; hit rate: 0.8
L2-cache:
read access time: 8 nanoseconds: hit rate: 0.9
Main memory:
read access time: 90 nanoseconds.
Please estimate the average data read cost (considering L1, L2 caches and main memory only).

Ans:
L1 access time + L1 miss rate * L2 access time + L1 miss rate * L2 miss rate * Memory access time
=2 + (1-0.8)*8+(1-0.8)*(1-0.9)*90
=2+1.6+1.8
=5.4 nanoseconds

QUESTION3

Consider the 2nd magic function we mentioned in the lecture, i.e., the magic function that can tell us which pages contain the qualified data. In practice, such magic function is implemented by a certain data structure and hence it incurs some cost when calling the function.

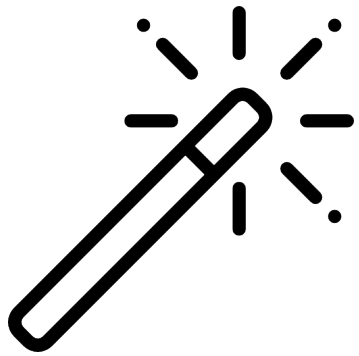
Suppose the cost of calling the function is equal to accessing log(N) pages, where N is the number of pages for the data. Please give a condition about when it is beneficial to use the function.

Let us recap what has been discussed in the lectures

- Consider another cost-free magic function:

The function tells you **which pages** will contain the qualified keys (i.e., x<4).
Can we do better with such a function?

In some situations, yes!



Cost: 40 Bytes

Query x < 4 from the following data

Qualified results

(size=120 bytes)
Memory Layer i

Memory Layer i+1

Magic function read Page 2

Page 1	Page 2	Page 3
5, 10, 7, 4, 12	2, 8, 9, 11, 7	7, 11, 3, 9, 8

Each integer: 8 bytes

Each page: 8 bytes x 5 = 40 bytes

Cost: 40 Bytes

Query x < 4 from the following data

Qualified results

(size=120 bytes)
Memory Layer i

2, 8, 9, 11, 7

2

bring

Magic function read Page 2

Memory Layer i+1

5, 10, 7, 4, 12

2, 8, 9, 11, 7

7, 11, 3, 9, 8

Page 1

Page 2

Page 3

Each integer: 8 bytes

Each page: 8 bytes x 5 = 40 bytes

Cost: 80 Bytes

Query x < 4 from the following data

Qualified results

(size=120 bytes)
Memory Layer i

2, 8, 9, 11, 7

7, 11, 3, 9, 8

2, 3

bring

Memory Layer i+1

Magic function read Page 3

5, 10, 7, 4, 12

2, 8, 9, 11, 7

7, 11, 3, 9, 8

Page 1

Page 2

Page 3

Each integer: 8 bytes

Each page: 8 bytes x 5 = 40 bytes

QUESTION3

In the setting, it is not cost-free… We have two options:

1) We can use the function, so that we have an **additional cost** of log(N) page accesses. The benefit is, we may only need to access a **subset** of data pages.
2) We can also prefer not to use the function, so that we have the total cost of scanning **all** the pages storing the data, i.e., incurring N pages.

QUESTION3

Ans:

1) We can use the function, so that we have **additional cost** of log(N) page accesses. The benefit is, we may only need to access a **subset** of data pages. Denote the number of pages in the subset is N'.

2) We can also prefer not to use the function, so that we have the total cost of scanning **all** the pages storing the data, i.e., incurring N pages.

Then case 1) is beneficial only when log(N)+N'<N.

Just this is enough

QUESTION3

Ans:

Then case 1) is beneficial only when log(N)+N’<N.

More formally, we define the page selectivity of the query to be c. Here, the selectivity is the ratio between the qualified pages and all pages. Then, we have

log(N)+N’<N

→ log(N)+N*c<N

→ log(N)<(1-c)N

DISCUSSION

When the selectivity for a query is low, the benefit of using the additional data structure is high.

Note: the benefit is

$N-(\log(N)+N') = (1-c)N-\log(N)$

QUESTION4

Consider the array-scanning scenario introduced in the lecture. In the lecture, we consider a single query for x>4. In big data systems, many queries are issued together.

Suppose our system needs to handle the following two queries:

1) Select x>4
2) Select x<2

Please explain an efficient way of finishing these two queries together, and analyze the number of page accesses.

QUESTION4

Handle the following two queries:

1) Select x>4
2) Select x<2

Think:

The data are stored in pages.

If we separately process the two queries:

(i) We load the data pages to memory, and check the data in each page for value>4.

(ii) We load the data pages again to memory, and check the data in each page for value<2.

Overall, there are two scans of the array.

Can we reduce the number of scans to one?

QUESTION4

Handle the following two queries:

1) Select x>4
2) Select x<2

Ans:

The data are stored in pages.

We can process the two queries in one single scan:

We load the data pages to memory, and for each page, we check the data in each page

(i) for value>4.

(ii) for value<2.

Overall, there is one scan of the array.

BIG DATA MANAGEMENT

CZ/CE4123

Tutorial 4
Cache Conscious Designs

QUESTION1

We have a 12-integer array in main memory as follows
1, 3, 5, 2, 4, 6, 4, 6, 8, 9, 11, 12
Let cache size be the size of 6 integers, and cache line size (transfer unit) be the size of 3 integers. Suppose initially the cache is empty, and there is a program sequentially accessing the whole array. The cache replacement mechanism is the same as that in the lecture notes: i.e., first cached first evicted.

(1) After the execution of the program, what are the final values stored in the cache? Please give the cache state after every access of the array element.
(2) How many cache hits and cache misses during the program execution? Please give the hit/miss state after every access of the array element.

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

Memory

1	3	5	2	4	6	4	6	8	9	11	12
1	2	3	4	5	6	7	8	9	10	11	12

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

Cache Miss: 1
Cache Hit: 0

Memory

1 2 3 4 5 6 7 8 9 10 11 12

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

Cache Miss: 1
Cache Hit: 1

Memory

1 2 3 4 5 6 7 8 9 10 11 12

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

Cache Miss: 1
Cache Hit: 2

Memory

1 2 3 4 5 6 7 8 9 10 11 12

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

1	3	5	2	4	6

Cache Miss: 2
Cache Hit: 2

Memory

1	3	5	2	4	6	4	6	8	9	11	12
1	2	3	4	5	6	7	8	9	10	11	12

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

Cache Miss: 2
Cache Hit: 3

Memory

1 2 3 4 5 6 7 8 9 10 11 12

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

1	3	5	2	4	6

Cache Miss: 2
Cache Hit: 4

Memory

1	3	5	2	4	6	4	6	8	9	11	12
1	2	3	4	5	6	7	8	9	10	11	12

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

4	6	8	2	4	6

Cache Miss: 3
Cache Hit: 4

Memory

1	3	5	2	4	6	4	6	8	9	11	12
1	2	3	4	5	6	7	8	9	10	11	12

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

4	6	8	2	4	6

Cache Miss: 3
Cache Hit: 5

Memory

1	3	5	2	4	6	4	6	8	9	11	12
1	2	3	4	5	6	7	8	9	10	11	12

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

4	6	8	2	4	6

Cache Miss: 3
Cache Hit: 6

Memory

1	3	5	2	4	6	4	6	8	9	11	12
1	2	3	4	5	6	7	8	9	10	11	12

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

4	6	8	9	11	12

Cache Miss: 4
Cache Hit: 6

Memory

1	3	5	2	4	6	4	6	8	9	11	12
1	2	3	4	5	6	7	8	9	10	11	12

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

4	6	8	9	11	12

Cache Miss: 4
Cache Hit: 7

Memory

1	3	5	2	4	6	4	6	8	9	11	12
1	2	3	4	5	6	7	8	9	10	11	12

QUESTION1 SOLUTION

Access pattern: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Cache

4	6	8	9	11	12

Cache Miss: 4
Cache Hit: 8

Memory

1	3	5	2	4	6	4	6	8	9	11	12
1	2	3	4	5	6	7	8	9	10	11	12

QUESTION2

Suppose we have a 2-dimentional integer array A[N][N]. We consider the two ways of array scanning: row-by-row and column-by-column. Let cache size = 5000 integers.

(1) If N>5000 and cache line size (transfer unit)=100 integers, please give a formal analysis of the cache hits/misses of the two ways of writing codes.

(2) If N=250 and cache line size=500 integers, please give a formal analysis of the cache hits/misses of the two ways of writing codes.

(note: for easy analysis, you can assume the 1st cache line starts from A[0][0])

SOLUTION FOR (1)

Solution X

```
for (int i=0;i<N;i++)
    for(int j=0;j<N;j++)
        A[i][j]=1;
```

Cache friendly

1 cache miss will bring 99 cache hits

SOLUTION FOR (1)

Ans:

For solution X, whenever there is a cache miss, it brings consecutive 100 integers (along the row) into the main memory, getting 99 cache hits.

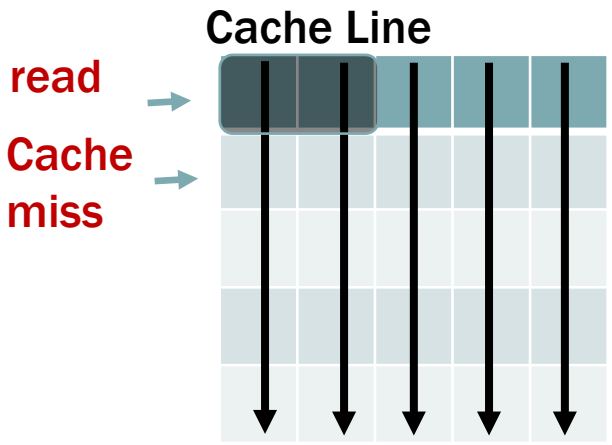
So, the overall cache misses: NxN/100

the overall cache hits: NxNx99/100

SOLUTION FOR (1)

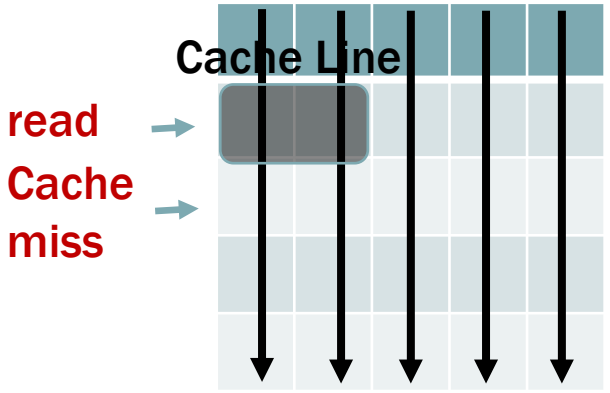
Solution Y

```
for (int j=0;j<N;j++)
    for(int i=0;i<N;i++)
        A[i][j]=1;
```



Cache unfriendly

All are cache misses



Cache unfriendly

All are cache misses

SOLUTION FOR (1)

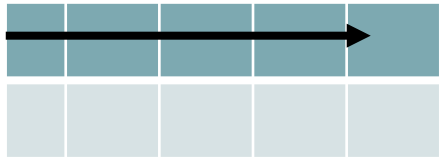
Ans:

For solution Y, since N>5000/100 (cache size/cache-line size), any cache miss at A[i][j] cannot guarantee a cache hit at A[i][j+1]. So every access incurs one cache miss.

SOLUTION FOR (2)

Solution X

```
for (int i=0;i<N;i++)
    for(int j=0;j<N;j++)
        A[i][j]=1;
```

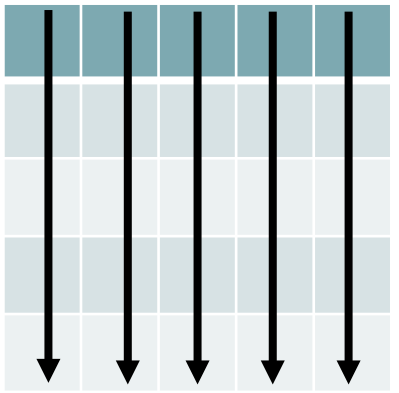


N=250

1 cache line can contain two rows.
1 miss will still bring 499 cache hits

Solution Y

```
for (int j=0;j<N;j++)
    for(int i=0;i<N;i++)
        A[i][j]=1;
```



N=250

1 cache line can contain two rows.
1 cache miss will bring 1 cache hit.

SOLUTION FOR (2)

Ans:

For solution X, the overall cache misses: NxN/500

the overall cache hits: NxNx499/500

For solution Y, since N=250>5000/250(cache size/row size), 1 cache miss will bring 1 cache hit.

So the overall cache misses: NxN/2

the overall cache hits: NxN/2

QUESTION 3

We have an 8-integer array A in the main memory. Let cache size be 4 (integers), and cache line size be 2 (integers). Suppose that initially the cache is empty, and the cache replacement policy is the same as the one introduced in the lecture notes, i.e., first cached first evicted.

Let the round-trip scanning be scanning an array in the order from the beginning to the end (i.e., A[0] to A[7]), and then from the end to the beginning (i.e., A[7] to A[0]). How many cache hits and cache misses during the round-trip scanning? Please explain your answer.

QUESTION 3


Forward Trip
A[0]
A[1]
A[2]
A[3]
A[4]
A[5]
A[6]
A[7]
Cache
Backward Trip
A[0]
A[1]
A[2]
A[3]
A[4]
A[5]
A[6]
A[7]

QUESTION 3

We have an 8-integer array A in the main memory. Let cache size be 4 (integers), and cache line size be 2 (integers). Suppose that initially the cache is empty, and the cache replacement policy is the same as the one introduced in the lecture notes, i.e., first cached first evicted.

Let the round-trip scanning be scanning an array in the order from the beginning to the end (i.e., A[0] to A[7]), and then from the end to the beginning (i.e., A[7] to A[0]). How many cache hits and cache misses during the round-trip scanning? Please explain your answer.

Ans: 10 hits, 6 misses.

1. The 1st trip (forward), one miss always brings one hit. We call it “1 miss 1 hit” procedure. So there are 4 hits and 4 misses.
2. The 2nd trip (backward), The first 4 integers of backward trip give 4 hits; the last 4 integers resume the cases of 1 miss and 1 hit.
3. In total, 10 hits and 6 misses.

BIG DATA MANAGEMENT

CZ/CE4123

Tutorial 5:
Column Stores

QUESTION1

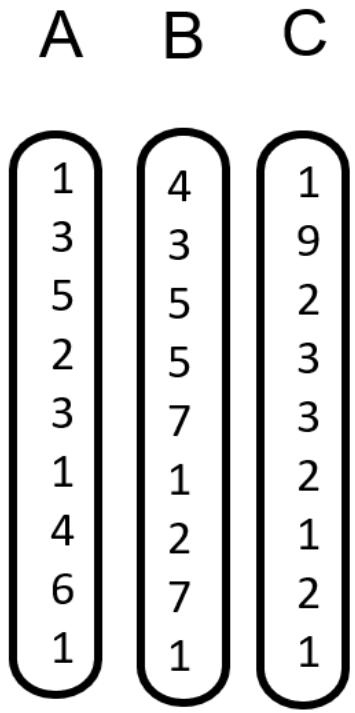
Given column store table T as follow.

(1)Give the flow chart (the flow graph presented in the lecture slides) using "column at a time" for the query

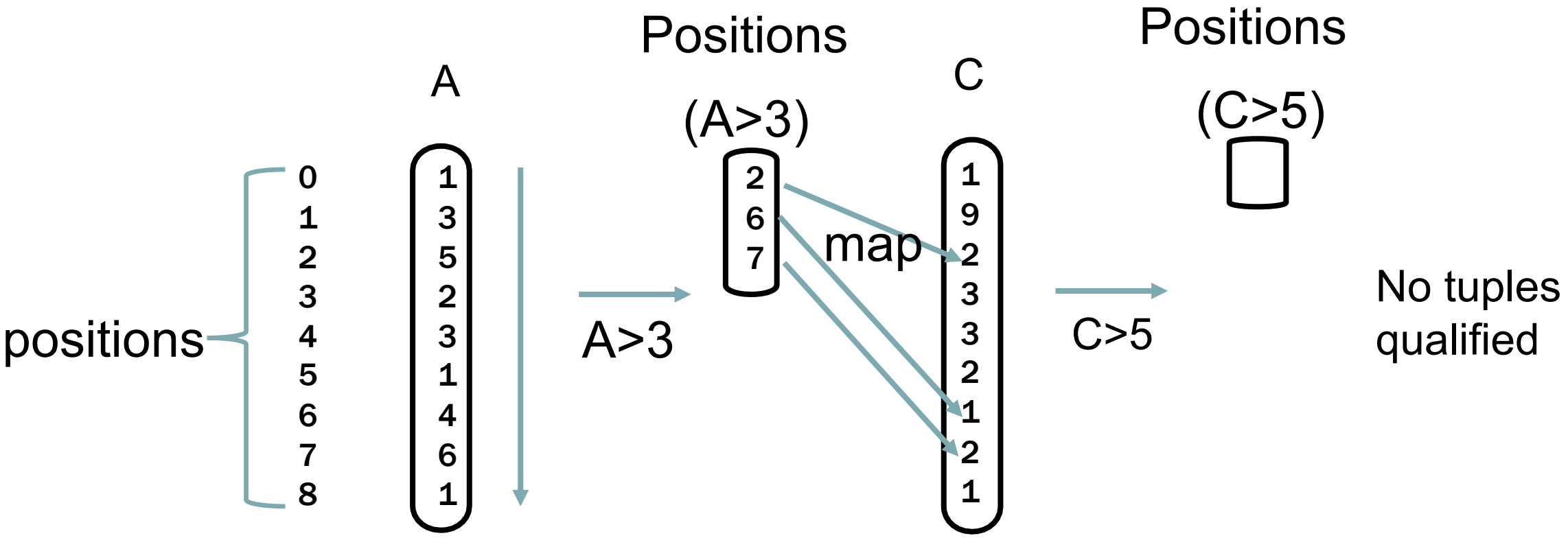
"***SELECT*** *min(B)* ***FROM*** *T* ***WHERE*** *A>3 and C>5*"

(2)Give the flow chart using "column at a time" for the query

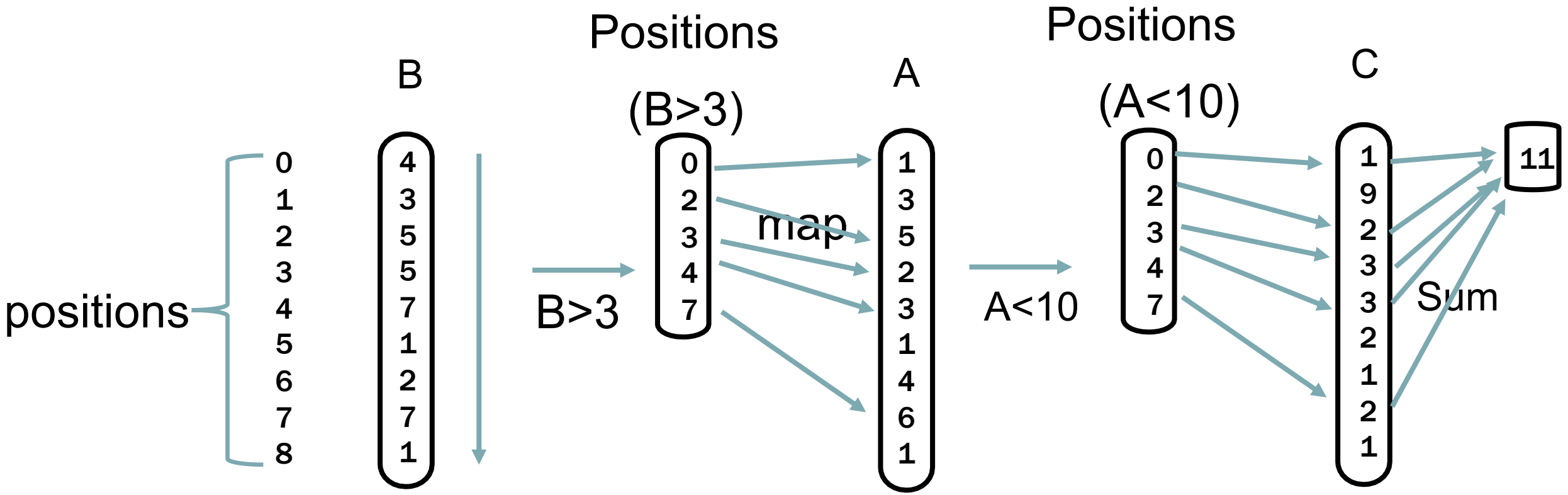
"***SELECT*** *sum(C)* ***FROM*** *T* ***WHERE*** *A<10 and B>3*"



SOLUTION FOR (1)



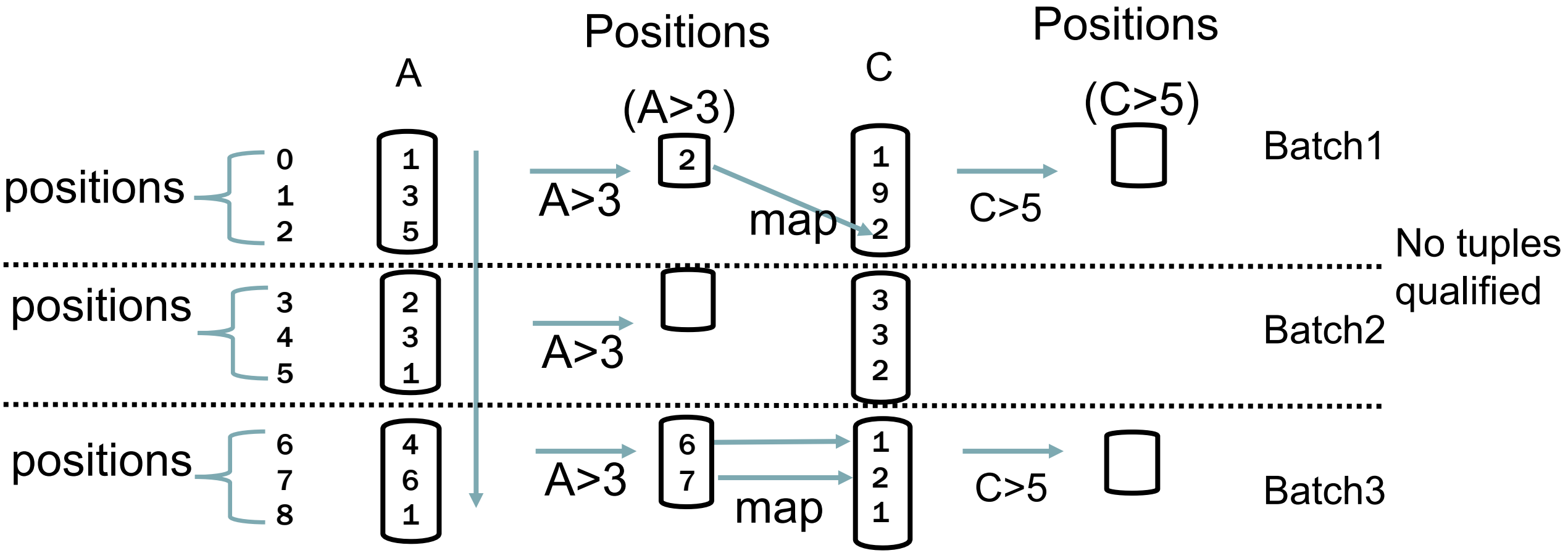
SOLUTION FOR (2)



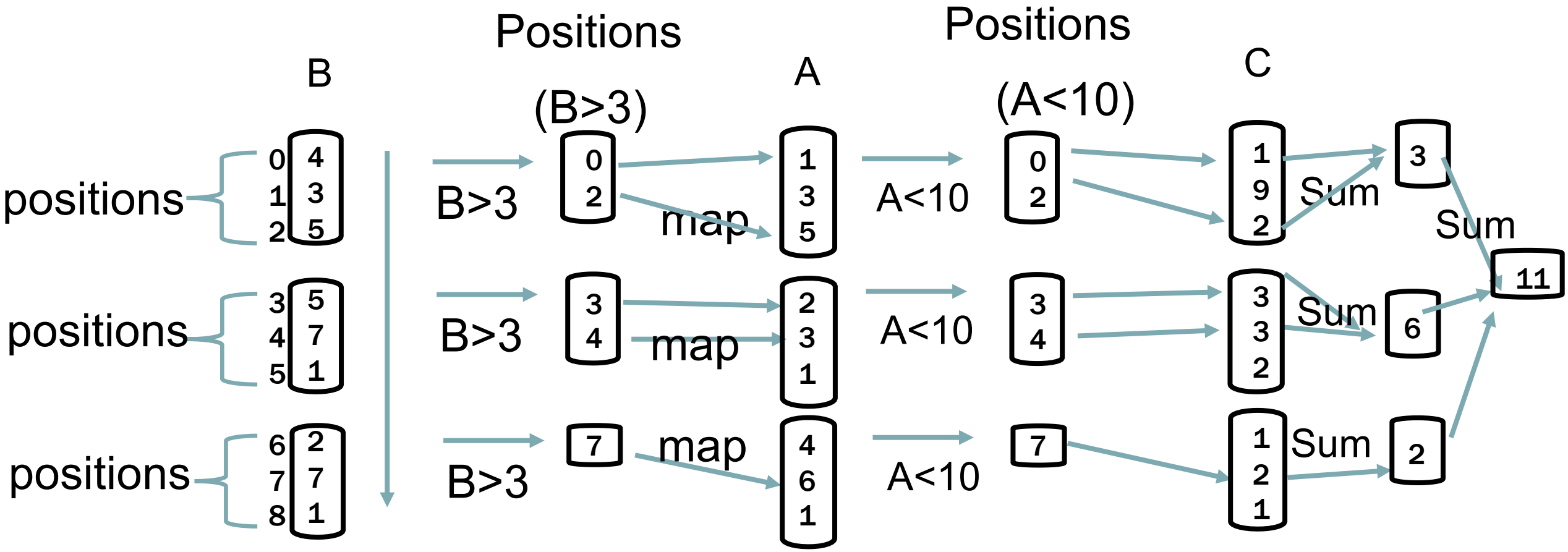
QUESTION 2

Redo Question 1 using “vector at a time”. Assume that the vector size is 3.

SOLUTION FOR (2)



SOLUTION FOR (2)



QUESTION 3

Suppose we are querying a **Student** information table with three columns **Name**, **Email**, **Age**. Given a query of the following form:

"***SELECT*** *Name* ***FROM*** *Student* ***WHERE*** *predicate (Email) and predicate (Age)*".

Here, a predicate applied on a column is a filtering function (e.g., **Email** ending with *ntu.edu.sg,* **Age**>19). We define the selectivity of a predicate by the percentage of the qualified results in the corresponding column. Assume that the selectivity of predicate(**Email**) is p and the selectivity of predicate(**Age**) is q, where $0<p<1$, and $0<q<1$. Let page size be P. We assume each column width is less than P and each value in a column is contained in a page. Consider two options in scanning columns: scanning **Email** first and scanning **Age** first.

(1) If the column widths are the same (denoted by w), please analyze which is better.

(2) If the width of **Email** is $2w$, and the widths for **Name** and **Age** are w, then which option is better?

RECAP

Cost for column store (number of page access):

$Zw/P+2result(A)*4/P+result(A)+2result(AB)*4/P+result(AB)$

$=Zw/P+result(A)*(8/P+1)+result(AB)*(8/P+1)$

$\approx Zw/P+result(A)+result(AB)$

SOLUTION TO (1)

Answer:
Use the formula we learnt in the lectures.
Let Email be *A*, and Age be *B*. Let *Z* be the length of a column
The cost of option 1 (Scanning Email first) is approximately
Zw/P+result(A)+result(AB)

The cost of option 2 (Scanning Age first) is approximately
Zw/P+result(B)+result(AB)

SOLUTION TO (1)

Answer:
Use the formula we learnt in the lectures.
Let Email be *A*, and Age be *B*. Let *Z* be the length of a column
The cost of option 1 (Scanning Email first) is approximately
Zw/P+result(A)+result(AB)

The cost of option 2 (Scanning Age first) is approximately
Zw/P+result(B)+result(AB)

Note that **result(A)=Zp, result(B)=Zq**
Then, if p<q, then scanning Email first is better; if p=q, equally good; if p>q, then scanning Age first is better.

SOLUTION TO (2)

Recap the formula we learnt in the lectures.

Zw/P+result(A)+result(AB)

Width of 1st accessed column

SOLUTION TO (2)

Answer:
Let Email be Column *A*, and Age be Column *B*. Let width of Email be 2*w*. Then the width of Age is *w*.
Revise the formula we learnt in the lectures.

The cost of option 1 (Scanning Email first) is approximately

2Zw/P+result(A)+result(AB)

The cost of option 2 (Scanning Age first) is approximately

Zw/P+result(B)+result(AB)

Note that **result(A)=Zp, result(B)=Zq**

SOLUTION TO (2)

Answer:

Then, option 1 is worse (or option 2 is better) when

2Zw/P+Zp+result(AB) > Zw/P+Zq+result(AB)

→ Zw/P+Zp > Zq

→ w/P+p>q

(Note: it is also okay to use more refined formulas, i.e., considering the cost of positions.)

FURTHER DISCUSSION

In practice, how do we know p and q?

BIG DATA MANAGEMENT

CZ/CE4123

Tutorial 6
More Practice for quiz

MAIN TAKE-AWAY MESSAGE

1. Procedures/reasons are important.
2. Consider all cases.
3. Just memorizing the knowledge is not perfect.

QUESTION1

Given the following three tables (primary keys are underlined):

Employee(EID, Salary)
Manager(MID, Salary)
Employee-Manager(EID, MID)

Each manager supervises at least one employees. Employee-Manager is a table that contains the manager ID (i.e., MID) for each employee (i.e., EID). How to convert the relational data model to a key-value data model? Consider that the main purpose of the conversion is for the query “Given an employee ID, find the salary of the employee’s manager”. The conversion should retain the information as much as possible.

Step 1: Join Employee and Employee-Manager on attribute EID

Step 2: further (Left) join Manager on attribute MID to get Table

(EID, MID, Manager-Salary, Employee-Salary)

Given the main purpose, we set

key as EID;

As the conversion needs to keep the information as much as possible, we set

value as MID;Manager-Salary;Employee-Salary;

COMMON MISTAKES

1. Only give the answer without procedures;
2. Incorrectly set the key to be EID;MID;
3. Only retain part of the information;

QUESTION 2

Consider reading data from memory hierarchy consisting of L1 Cache, L2 Cache, and main memory with the following parameters.

L1 Cache:

Read access time: 2 nanoseconds

Miss ratio: 0.4

L2 Cache:

Read access time: 10 nanoseconds

Miss ratio: 0.2

Main memory:

Read access time: 100 nanoseconds.

Estimate the average data read cost and explain your answer. (Note: consider L1, L2 caches and main memory only).

SOLUTION

First access L(i), having a cost c1;

Then, if there is an L(i) miss, we need to access L(i+1). This happens with Probability m1 (miss rate) and a cost of c2.

Then, if there is an L(i+1) miss, we need to access L(i+2). This happens with probability m1m2 (both L(i) and L(i+1) miss) and a cost of c3.

Average cost = **c1+m1c2+m1m2c3 = 2+0.4*10+0.4*0.2*100=14 nanoseconds.**

ALTERNATIVE SOLUTION

If L1 cache hit,
probability = 1-0.4=0.6
If L1 cache miss and L2 cache hit,
probability = 0.4*(1-0.2)=0.32
If L1,L2 caches miss (data in memory),
probability = 0.4*0.2=0.08

Average data read cost = 2*0.6 + (2+10)*0.32+(2+10+100)*0.08=14ns

QUESTION 3

In the lecture, we introduced a cost-free magic function telling which pages locate the qualified data for a query. Consider a disk page size is 1024 integers. There are 10240 integers, which are 1, 2, 3, …, 10240, sequentially stored at 10 consecutive disk pages. Consider the following 4 queries using 4 scans over the data, where each query range is decided by three integers x, y, z, and $1<x<y<z<10240$.

Query 1: searching values in the range [1, x] (i.e., values at least 1 and at most x)
Query 2: searching values in the range [x+1, y]
Query 3: searching values in the range [y+1, z]
Query 4: searching values in the range [z+1, 10240]

List **all possible** total number of read I/Os needed for the 4 scans, *with* the magic function. Please explain your answer.

QUESTION 5

Based on the magic function, the total #pages of 4 searches is roughly 10 (but with possible additional cost) because the 4 ranges are disjoint and covering all the range [1, 10240].

The best case happens when x, y, z are in the exact page boundaries (right boundary), i.e., x and x+1 are stored at different pages; y and y+1 are stored at different pages; z and z+1 are stored at different pages. In this case, total I/Os=10.

Also note that, if x is in the middle of a page, i.e., x and x+1 are in the same page, then this page is accessed by 2 searches.

Hence, the worst case is 10+3 = 13pages, where x, y, z are in the middle.

Finally, easy to see 10, 11, 12, 13 are all possible, because

(1)Incurring 11 page reads happens when only one of (x,y,z) is in the middle of a page.

(2)Incurring 12 page reads happens when only two of (x,y,z) are in the middle of a page.

QUESTION 4

We have a 32-integer array *A* in the main memory. Let cache size be 16 (integers), and cache line size be 4 (integers). Suppose that initially the cache is empty, and the cache replacement policy is the same as the one introduced in the lectures, i.e., first cached first evicted. Let the *f-trip* and *b-trip* scanning be the following.

```
f-trip(){
    for (int j=0; j<32; j++){
        Access A[j];// Access does not change the data
    }
}
b-trip(){
    for (int j=0; j<32; j++){
        Access A[31-j];
    }
}
```

If we need to do 99 scans of the array, and we can select each scan to be either f-trip or b-trip. Please give one best selection strategy that gives the minimum number of misses and explain your answer. Please also compute the number of cache hits and cache misses in the best strategy.

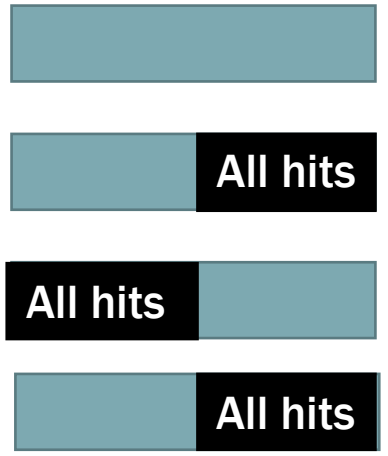
QUESTION 2

The best strategy:

Take turns to use f-trip and b-trip, namely,

f-trip, b-trip, f-trip, …

The idea is that, starting from the 2nd trip, we can always make use of the cache from the previous trip.



Number of hits & misses

1st f-trip: one miss brings three hits. So there are 8 misses and 24 hits.

The remaining 98 trips: first 16 accesses are hits; the last 16 accesses also have the pattern – one miss brings three hits. In total: 28 hits and 4 misses.

Total hits: 24+28*98=2768

Total misses: 8+4*98=400

BIG DATA MANAGEMENT

CZ/CE4123

Tutorial 7
Distributed Systems and MapReduce

QUESTION 1

Amazon wants to estimate the Top-K best sold products from S purchase records of L products in the form of a list of (User id, Product id) pairs. Suppose there is a distributed system with 1 master machine and M slave machines. Assume that L is a multiple of M. Design a distributed computation procedure to finish the task. Please describe

(1) how the data is distributed, computed and aggregated?

(2) how much data is sent across machines?

SOLUTION – MAIN IDEA

Let us roughly think…

First, we let each machine handle L/M products. Distribute the records corresponding to Product 1, 2, …, L/M to Machine 1, the records corresponding to Product L/M+1, …, $2L/M$ to Machine 2, …

Second, in each machine, we can simply compute the top-K sold products locally.

Third, when there are local Top-K best sold products computed from each machine, aggregate the results to finally form the global top-K results.

SOLUTION – DETAILS

Firstly, the master machine distributes the S purchase records to the slave machines, so that the records of Product 1, 2, …, L/M are sent to Machine 1, the records of Product L/M+1, …, $2L/M$ are sent to Machine 2, …

Secondly, each machine computes the top-K sold products locally. In particular, for each product, we collect the frequencies of it being sold, and sort the frequencies and find the K products *with* top-K frequencies. Only send the top-K pairs of (Product id, Frequency) to master.

SOLUTION – DETAILS

Finally, the master machine aggregates all the MK pairs of (Product id, Frequency). Output the products with the Top-K frequencies.

SOLUTION – DATA SENT

Data Distribution Phase: *S* records have been sent out.
Aggregation Phase: *MK* pairs of (Product id, Frequencies) have been sent out.

In total, the data sent out is in the scale of $O(S+MK)$.

QUESTION 2

(1) In a MapReduce job, the output of Map phase is a list of key-value pairs:

(A, 1) (C, 2), (A, 5), (C, 6), (B, 3), (E, 3), (C, 8). Please list the possible input to the Reduce function.

The input to the *Reducer* aggregates the *Map* output keys.

Hence, the possible input to the reducer is (A, {1, 5}); (B, {3}); (C, {2, 6, 8}); (E, {3}).

(2) Based on the answer to Q2(1), write a Reduce function (in pseudocode) so that the MapReduce output is: (2, A), (3, C), (6, A), (7, C), (4, B), (4, E), (9, C).

```
Reduce(int key, iterator values){
    for(each v in values){
        Emit(v+1, key);
    }
}
```

(3) Consider an employee table containing three columns (EmployeeID, age, monthly-salary) where age and monthly-salary are integers. Use MapReduce to collect the number of employees falling into each of the following two categories:

- Category 1: The age of the employee is between 30 and 40 (including 30 and 40). His/her monthly salary is at most 7000.

- Category 2: The age of the employee is between 40 and 50 (including 40 and 50). His/her monthly salary is more than 7000.

Please use **only one MapReduce Job** to achieve this task and write down the pseudocode of the Map function and Reduce function. The input key and value for Map function are an employee's age and monthly-salary respectively.

```
Map(int age, int salary){
    if(age>=30 and age<=40 and salary<=7000){
        Emit-Intermediate(1, 1);
    }
    if(age>=40 and age<=50 and salary>7000){
        Emit-Intermediate(2, 1);
    }
}

Reduce(int category, iterator values){
     int total_frequency = 0;
     for(int frequency: values)
    {
         total_frequency+=frequency;
     }
     Emit(category, total_frequency);
}
```

QUESTION 3

Design MapReduce algorithms for the multiplication of two matrices *A*, *B* of *n* by *n*. Elements in the matrices are integers. The input key for *Map* function is *MatrixName*; the input value for *Map* function is in the form of *i;j;v*, indicating that the value of the *i*-th row and *j*-th column is *v*.

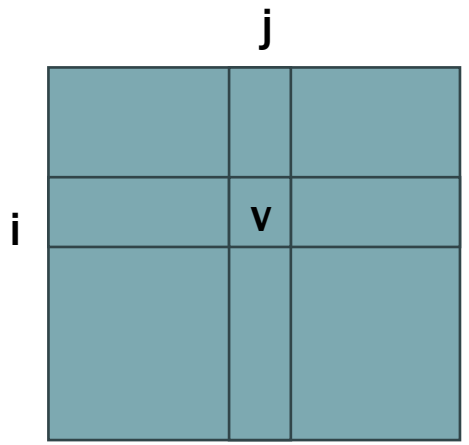
(Matrix Multiplication: Given matrix *A*[*n*x*n*] and matrix *B*[*n*x*n*], compute matrix *C* such that $\mathrm{C[i][z]} = \sum_{\mathrm{j}=0}^{\mathrm{n}-1} \mathrm{A[i][j]xB[j][z]}$, for i,z in [0, n-1]).

(1) Use at most two MapReduce jobs to finish the computation.

(2) Furthermore, can the multiplication be finished using one MapReduce job?

SOLUTION – 1ST JOB

```
Map(String MatrixName,  String value){
    int i = get_i(value);// get row
    int j = get_j(value);// get column
    int v = get_v(value);// get value
    if(MatrixName==“A”)
        Emit-Intermediate(j, ToString(MatrixName, i, v));// combine
    else
        Emit-Intermediate(i, ToString(MatrixName, j, v));
}
```



Column as key

Row as key

For each intermediate key j, the values can contain

(1) A[u][j] for any u;

(2) B[j][v] for any v;

What needs to be done in reduce()?

For any (u, v), send out A[u][j]*B[j][v] for aggregation of C[u][v]

SOLUTION – 1ST JOB

```
Reduce(String key,  Iterator<String> values){
    int A_start[n];  int B_end[n];
    for(String value : values){
       String MatrixName=get_first_element(value);
       if(MatrixName==“A”){
            int i = get_second_element(value);// get row index in matrix A
            A_start[i]=get_third_element(value);// get the matrix element
       }
       else{
            int j = get_second_element(value);// get column index in matrix B
            B_end[j] = get_third_element(value);// get the matrix element
       }
    }
}
```

```
for (int u=0;u<n;u++)
  for(int v=0;v<n;v++)
  {
     Emit(ToString(u,v), A_start[u]*B_end[v]);
  }
```

A[u][j] B[j][v]

THE 2ND JOB

Need to aggregate all the values for each pair (u,v).

So,

Map() does nothing;

Reduce() conducts the aggregation.

SOLUTION - 2ND JOB

```
Map(String key, String value){
    Emit-Intermediate(key, value);
}
```

SOLUTION - 2ND JOB

```
Reduce(String key, Iterator<String> values){
    int sum=0;
    for (String value in values){
        sum+=ToInteger(value);
    }
    Emit(key, sum);
}
```

Can we use a single job only?

Think…

For any A[i][j], which elements of C would it contribute to?

Can we use a single job only?

Think…

For any A[i][j], which elements of C would it contribute to?

C[i][k] for any k

For any B[i][j], which elements of C would it contribute to?

C[k][j] for any k

CAN WE FINISH THESE TASKS WITH ONE JOB ONLY?

```
Map(String MatrixName,  String value){
    int i = get_i(value);
    int j = get_j(value);
    int v = get_v(value);
    if(MatrixName==“A”)
        for(int k=0;k<n;k++)
            Emit-Intermediate(ToString(i, k), ToString(MatrixName, j, v));// each A[i][j] may contribute
                                                                          //to C[i][k] for any k
    else
        for(int k=0;k<n;k++)
                Emit-Intermediate(ToString(k, j), ToString(MatrixName, i, v)); // each B[i][j] may
                                                                          //contribute to C[k][j] for any k
}
```

For each intermediate key (pair (i,j)), it aggregates

(1) The values of A[i][u], for any u

(2) The values of B[u][j], for any u

CAN WE FINISH THESE TASKS WITH ONE JOB ONLY?

```
Reduce(String index_pair_for_C,  Iterator<String> values){
    int A_middle[n]; int B_middle[n];
    for(String value in values){
        String MatrixName=get_first_element(value);
        if(MatrixName==“A”){
            int j = get_second_element(value);
            A_middle[j]=get_third_element(value);
        }
        else{
            int i = get_second_element(value);
            B_middle[i] = get_third_element(value);
        }
    }
}
```

```
int sum=0;
for (int u=0;u<n;u++)
  sum+=A_middle[u]*B_middle[u];

Emit(index_pair_for_C, sum);
```

EXAMPLE

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

EXAMPLE-MAP

A[0][0] → ((0;0), (A; 0; A[0][0]))

→ ((0;1), (A; 0; A[0][0]))

A[0][1] → ((0;0), (A; 1; A[0][1]))

→ ((0;1), (A; 1; A[0][1]))

A[1][0] → ((1;0), (A; 0; A[1][0]))

→ ((1;1), (A; 0; A[1][0]))

A[1][1] → ((1;0), (A; 1; A[1][1]))

→ ((1;1), (A; 1; A[1][1]))

B[0][0] → ((0;0), (B; 0; B[0][0]))

→ ((1;0), (B; 0; B[0][0]))

B[0][1] → ((0;1), (B; 0; B[0][1]))

→ ((1;1), (B; 0; B[0][1]))

B[1][0] → ((0;0), (B; 1; B[1][0]))

→ ((1;0), (B; 1; B[1][0]))

B[1][1] → ((0;1), (B; 1; B[1][1]))

→ ((1;1), (B; 1; B[1][1]))

EXAMPLE-REDUCE

Take (1,0) for example intermediate key

It aggregates {(A; 0; A[1][0]), (B; 0; B[0][0]), (A; 1; A[1][1]), (B; 1; B[1][0])}

Hence, we can compute C[1][0] by A[1][0]*B[0][0]+A[1][1]*B[1][0]

EXAMPLE-REDUCE

Take (1,0) for example intermediate key

It aggregates {(A; 0; A[1][0]), (B; 0; B[0][0]), (A; 1; A[1][1]), (B; 1; B[1][0])}

Hence, we can compute C[1][0] by A[1][0]*B[0][0]+A[1][1]*B[1][0]

BIG DATA MANAGEMENT

CZ/CE4123

Tutorial 9
MapReduce Designs

QUESTION1

```
Map(string key, string value){
    if (key.equals("Student-Table"){
        studentID=split(value).first;
        courseID=split(value).second;
        Emit(studentID, courseID);
    }
}
```

QUESTION1

```
Reduce(string key, iterator values){
    int s=0;
    Map<string, string> distinct_course;
    for(each v in values){
        if(distinct_course does not contain v)
         {
            distinct_course.insert<v,1>;
            s++;
         }
    }
    Emit(key, s);
}
```

QUESTION 2

```
Map(string key, string value){
    if(key.equals("Student-Table"){
        studentID=split(value).first;
        courseID=split(value).second;
        semester=split(value).third;
        Emit(courseID, semester);
    }
}
```

QUESTION 2

```
Reduce(string key, iterator semesters){
    Map<string, string> semester_freq;
    for(each sem in semesters){
      if(semester_freq does not contain sem)
      {
            semester_freq.insert<sem, 1>;
       }
      else
            semester_freq[sem]++;
    }
    int cnt=0;
    for(each <semester, freq> in semester_freq){
        if(freq>50){
            cnt++;
        }
    }
    if (cnt>=2)
        Emit(courseID, NULL);
}
```

QUESTION 3

```
Map1(string key, string value){
    if(key.equals("Student-Table")){
        studentID=split(value).first;
        courseID=split(value).second;
        semester=split(value).third;
        Emit(toString(courseID, ";", semester), toString("s;", studentID));
    }
    if(key.equals("Professor-Table")){
        professorID=split(value).first;
        courseID=split(value).second;
        semester=split(value).third;
        Emit(toString(courseID, ";", semester), toString("p;", professorID));
    }
}
```

QUESTION 3

```
Reduce1(string key, iterator values){
    List professors;
    List students;
    for(each value in values){
        if(value starts with “s”){
            students.add(getStudentID(value));
        }
        if(value starts with “p”){
            professors.add(getProfessorID(value));
        }
    }
    for(each student in students){
        for(each professor in professors){
            Emit (student, professor);
        }
}
```

QUESTION 3

//Map2 takes the output of *Reduce1*, with the purpose to remove duplicates.

```
Map2(string student, string professor){
    Emit(toString(student,";", professor), "1");
}
Reduce2(string key, iterator values){
    student=split(key).first;
    professor=split(key).second;
    Emit(student, professor);
}
```

BIG DATA MANAGEMENT

CZ/CE4123

Tutorial 10
Key-Value Stores

QUESTION 1

Consider a leveling LSM-tree with a size ratio 4. The memory buffer (Level 0) can store 5000 key-value pairs. Initially the LSM-tree is empty. After inserting 70000 key-value pairs with distinct keys continuously, how many levels are formed?

QUESTION 1-SOLUTION

This question is related to understanding the capacity.
The capacity of a level is defined as “the maximum number of key-value pairs that can be stored in the level”.

Size Ratio = 4
Level 0 capacity = 5000
Level 1 capacity = 5000x4=20000
Level 2 capacity = 5000x4x4=80000

70000 > Level 0 capacity + Level 1 capacity (=25000)
70000 < Level 0 capacity + Level 1 capacity + Level 2 capacity (= 105000)
So there are at least 3 levels (Level 0, Level 1, Level 2)

QUESTION 1-SOLUTION

We further verify that Level 3 is not created. (why?)

If level 2's actual size is larger than Level 2 capacity – Level 1 capacity, it can already trigger the merge

Before creating Level 3, it must trigger the sort-merge of Level 2. Since the 70000 keys are distinct, the actual size of Level 2 must be a multiple of the capacity of Level 1 (i.e., 20000). So only when the actual size of Level 2 reaches 80000 it can trigger a merge, (note that 0, 20000, 40000, 60000 are not larger than {Level 2 capacity – Level 1 capacity}, and hence will not trigger a merge). However, since there are only 70000 keys, it is impossible for Level 2 to reach a size of 80000. Hence, Level 3 will not be created. So finally there are 3 levels (Levels 0, 1, 2).

QUESTION 2

Consider a leveling LSM-tree with a size ratio 4. The LSM-tree has 5 levels (excluding the memory buffer level), and it is incorporated with **both** fence pointers and Bloom Filters. Assume that a key-value pair is always entirely stored within a disk page. Consider the procedure of Get(*K*) for a key *K*, which of the followings sequence are possible to be the I/O costs from Level-0 to Level-5? (can select multiple answers)

(a) 1, 1, 1, 1, 1, 1

(b) 0, 1, 1, 1, 1, 1

(c) 0, 0, 1, 0, 0, 1

(d) 0, 0, 0, 0, 0, 1

(e) 1, 0, 0, 0, 0, 0

SOLUTION

(a) and (e) are not possible because Level 0 is memory buffer, and hence no I/O costs.

(b) (c) (d) are possible because if *K* is not in the LSM-tree, then each level's I/O cost fully depends on the accuracy of the Bloom filter (We consider Fence Pointer will incur 1 I/O when Bloom filter returns true).

- (b) is possible: BFs in Levels 1-5 all generate false-positives
- (c) is possible: BFs in Levels 2 and 5 generate false-positives
- (d) is possible: BF in Level 5 generates a false-positive

QUESTION 3

Consider a leveling LSM-tree with a size ratio 4. The LSM-tree has *L* levels (excluding the memory buffer level), and it is **only** incorporated with fence pointers (without Bloom Filters). Assume that a key-value pair is always entirely stored within a disk page. Consider the procedure of Get(*K*) for a key *K*,

(1) What is the possible I/O cost of accessing Level-*i* (*i* is in [1, *L*])?

(2) If *K* exists in the LSM-tree, what is the expected I/O cost at Level-*i* (*i* is in [1, *L*])? (hint: divide the cases based on the first-appearing location of the key *K*)

SOLUTION FOR Q3(1)

The possible I/O cost at Level-*i* can be 0 or 1.

0 is possible:

if *K* first appears in a level smaller than Level-*i*, then the search is terminated before Level-*i*. Hence, no I/O cost at Level-*i*.

1 is possible:

if *K* first appears in Level-*i* or larger levels, then fence pointer is used and can incur 1 page read, or 1 I/O. (Note: We assume that using fence pointers always incurs 1 I/O.)

SOLUTION FOR Q3(2)

Let *j* be the first level that contains key *K*.

Divided into two cases:

- If *i*>*j*, then the (expected) I/O cost is 0, because the search ends at Level *j*.

- If *i*<=*j*, then the (expected) I/O cost is 1, because at level *i* fence pointer is used and incurs 1 I/O.

QUESTION 4

Consider a leveling LSM-tree with a size ratio 4. The LSM-tree has *L* levels (excluding the memory buffer level), and it is incorporated with **both** fence pointers and Bloom Filters. Assume that a key-value pair is always entirely stored within a disk page. Consider the procedure of Get(*K*) for a key *K*,

(1) What is the possible I/O cost of accessing Level-*i* (*i* is in [1, *L*])?

(2) If *K* exists in the LSM-tree and the FPR of the Bloom filter at Level-*i* is *P* (*P* is in [0, 1]), what is the expected I/O cost at Level-*i* (*i* is in [1, *L*])? (hint: divide the cases based on the first-appearing location of the key *K*)

SOLUTION FOR Q4(1)

The possible I/O cost at Level-*i* can be 0 or 1.

0 is possible: if *K* is not in the LSM-tree, and the BF in level-*i* returns FALSE. Then, the search within the disk for this level is skipped.

1 is possible: if *K* is not in the LSM-tree, and the BF in level-*i* returns TRUE. Then, fence pointer is used and incurs 1 page read, or 1 I/O. (Note: We assume that when using fence pointers always incurs 1 I/O.)

SOLUTION FOR Q4(2)

Since *K* exists in the LSM-tree. The cases is divided based on the first-appearing level of key *K*.

Let Level-*j* be the level that first contains *K*.

Case 1: If $i>j$, the search is up to Level-*j*, and terminates before reaching Level-*i*, and hence the I/O cost at Level-*i* is 0;

Case 2: If $j=i$, the I/O cost at level-*i* must be 1 because the key first appears at Level-*i*.

Case 3: If $i<j$, the I/O cost depends on the FPR of the Bloom filter:

- Since the key *K* does not exist in Level-*i*, then with probability *P* the Bloom filter returns TRUE, and later the fence pointer incurs 1 I/O.
- Since the key *K* does not exist in Level-*i*, then with probability 1-*P* the Bloom filter returns FALSE, and no I/O cost is incurred.
- To summarize, the expected I/O cost is : $P*1+(1-P)*0=P$.