



EXAM TOPICS AND DESCRIPTION

GOAL

- Clarify exam Process
- Clarify exam Topics

FORMALIA

TIME AND PLACE

Exam take place:

- Monday June 18th U66
- Tuesday June 19th U66


Starting 8.30

Each student 20-25 min. in total (including deciding grade and receiving it)

WHEN TO BE THERE

 In principle, you should be at the venue for the whole day.

- For your convenience you will be assigned a timeslot. But note that the schedule is not exact, since some students may not show up
- If a student doesn't show up, the next will be taken
- The external examiner may leave if there is no more student to be taken

 Please show up plenty early (2 hours before your timeslot unless you are first)

SPORTSMANSHIP

- 💡 Be fair to your peers and announce if you are sick or cannot show up (or intend to skip)

PROCESS

Draw a topic (no preparation time)

10 minutes for presentation on the topic - might be interrupted with questions or cut short for another topic

5-10 minutes of questions about the other topics


In English - but you can explain terms or concepts in Danish if you are stuck with the language

❗ You can bring brief notes (no reading from the notes!) **No slides are allowed**

TOPICS

1. Storage and Indexing
2. Relational Operator Evaluation
3. Query Optimization
4. External Sorting
5. Concurrency Control with Locking
6. Concurrency Control without Locking
7. Crash Recovery
8. Database Tuning
9. Distributed Query Processing
10. Distributed Transactions

1 - STORAGE AND INDEXING

- Describe the cost model, i.e. what are we optimizing for, and why (Explain the elements of access time)
 - Describe file organization and indexing into a file
 - Describe access methods and what the difference between a clustered and an unclustered index have on it
-  include the task of a buffer manager and how does it process a page request.

2 - RELATIONAL OPERATOR EVALUATION

- Explain the cost model and compare them
- Iteration
 - Selection and projection: on the fly
 - Join: nested-loop join (page-based, block-based)
- Index-based:
 - explain how to match an index with a predicate
 - index-based join
- Explain sort-merge join and hash join
 - Hybrid hash join



Pay attention to the buffer usage!

3 - QUERY OPTIMIZATION

Relational algebra equivalences

How to estimate the cost of a plan

- how to estimate the selectivity (reduction factor)
- Histograms

Dynamic programming to choose the "best" plan

- interesting order
- be careful its difference with greedy algorithm
- describe limitations

4 - EXTERNAL SORTING

Explain multi-pass merge sort and its cost model

Explain how you can make longer initial runs and how it reduces the cost

If you have more time: Explain blocked I/O and double buffering

 Pay attention to the buffer usage!

5 - CONCURRENCY CONTROL WITH LOCKING

Basic concepts of concurrency control,

- such as ACID properties, serializability, etc.

Explain 2PL and strict 2PL

- Why they work

Multi-granularity Lock

6 - CONCURRENCY CONTROL WITHOUT LOCKING

Basic concepts of concurrency control,

- such as ACID properties, serializability, etc.

Timestamp-based CC

Multiversion CC

7 - CRASH RECOVERY

Explain the roles of Buffer Management

- Steal/No-Steal,
- Force/No-Force

Explain write-ahead logging and why it works.

Explain the ideas and mechanisms of checkpointing

ARIES Algorithm

8 - DATABASE TUNING

Index Selection

- The overall procedure of index selection
 - What index can be used for various predicates, including equality selection, range selection and joins
 - Clustered vs Unclustered Index; how to decide which index should be clustered
 - Index-only plans
- Horizontal Decomposition
- Query Tuning

9 - DISTRIBUTED QUERY PROCESSING

Parallelism vs. Distributed

Parallel Sorting and Parallel Join

Distributed Join

- Semi-join
- Bloom-join

10 - DISTRIBUTED TRANSACTIONS

2PC & 3PC

- What is 2PC
- Why 2PC?
- Why 2PC is blocking and how 3PC solve the problem

WHAT ELSE?

Expect also questions on:

- Clarification of what you just said
- Something that you haven't talked about
 - Your projects and NoSQL
 - Data replication

ADVICE

- Present the contents with a good logical structure
- Prepare in groups
- Don't just present “how” but also explain “why”
- Don't give the same lecture as I gave at the class



Use the blackboard and **don't** wipe out what you write

- Use diagrams as much as possible to illustrate the algorithms and your points
 - Probably a simplified version of the diagrams in the slides

QUESTIONS