香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen
SCHOOL OF DATA SCIENCE
數據科學學院

CSC3170
Final Review

Chenhao Ma
School of Data Science
The Chinese University of Hong Kong, Shenzhen

SCHOOL OF DATA SCIENCE

Final Exam

- **December 12, 2024 (Thursday)**
- Time: **8:20 AM – 10:20 AM**
- Location: **SHAW F302 & F301 (逸夫F302、F301)**
- **Seat Numbers:** Please check BB

Final Exam

- The exam is **closed-book**.

- **One A4-sized cheat sheet** is allowed:
  - You may write on both sides of the sheet.
  - Handwritten notes only; printed materials are not permitted.

- **No electronic devices** (e.g., calculators, phones, or smartwatches) are allowed.

General advice

- Arrive early to find your assigned seat.
- Bring all necessaries, including pens and student ID.
- Manage your time wisely during the exam.

Key points

The key points may not include all the exam content, and not all of them will necessarily be tested.

- Relational Model
  - Relational Algebra
- SQL (basic usage + advanced usage)
- Storage
  - File Storage
  - Page Layout
  - Tuple Layout
  - OLAP + OLTP
  - Row store + column store
  - Compression

Key points

- Buffer pool
  - Buffer pool manager
  - Optimizations
    - Multiple buffer pools, prefetching, scan sharing, buffer pool bypass
  - Replacement policies
- Hash Tables
  - Static + Dynamic Hashing Schemes
- B+Tree
  - Insert + Delete
  - Clustered / Unclustered

Key points

- Index Concurrency Control
  - Hash table latching
  - B+Tree Latching (basic + better algorithm)
- Sorting & Aggregations
  - External merge sort
  - Aggregations
- Join Algorithms
  - Nested Loop Join (naïve, block, index)
  - Sort-merge Join
  - Hash Join (GRACE)
  - Cost analysis

Key points

- Query execution
  - Processing models
  - Access methods
  - Expression evaluation
  - Parallel (inter- / intra- query)
- Query optimization
  - Logical vs Physical Plans
  - Search Space of Plans
  - Cost Estimation of Plans
- DB Design
  - E-R Diagram
  - Functional Dependency

Key points

- Transactions
  - ACID
  - Dependency graph

Sample questions

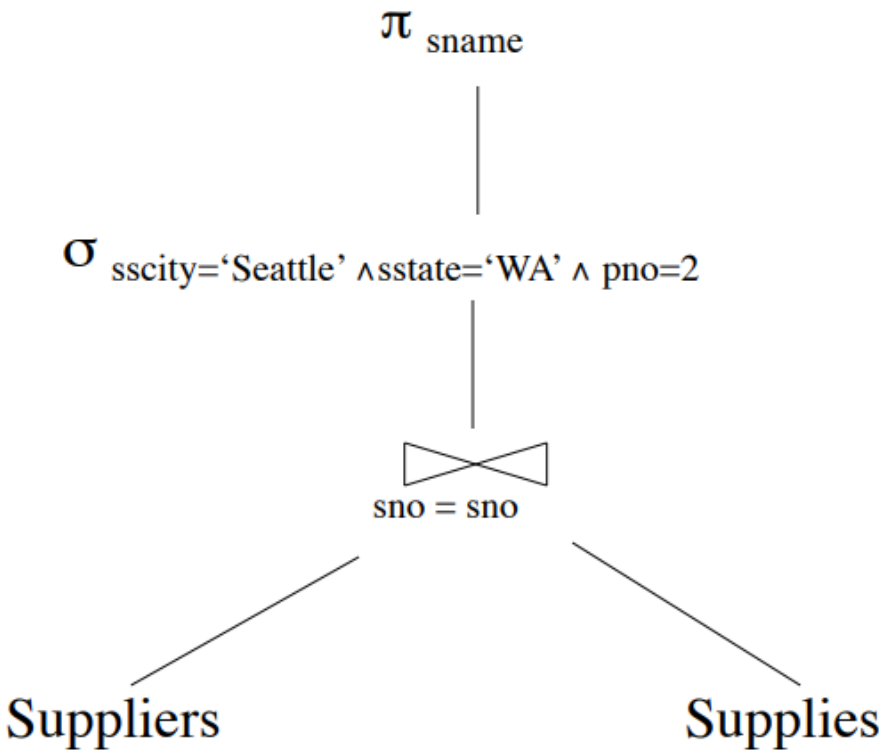
- Show the result of inserting 12, 10, 15, 4, 1, 17, 3, 13, and 8 into an initially empty 5-way search B+tree one-by-on. To maintain consistency in answers, please follow the following rules: You should split nodes whenever there is an overflow due to insertion. When splitting a leaf node due to insertion overflow, keep half (rounded up) in the left node and half (rounded down) in the right.

Sample questions

- For the hash table using linear probing, insert the following values in order: 4371, 1323, 6173, 4199, 4344, 9679 and 1989. Assume the table size is 10 and that the primary hash function is h(k) = k % 10. You do not need to resize the tables. If an element cannot be successfully inserted, state why. You need only show the final table.

Sample questions

- Assume sno in supplies is a foreign key.
- suppliers have 2,000 tuples and 200 pages.
- supplies have 10,000 tuples and 1,000 pages.
- What is the cost of different join algorithms?
  - Partitioned hash join
  - Sort-Merge
  - Block-nested
- The size of the join output?



Sample questions

- **What is the average time complexity of a hash table lookup operation?**
  - A) $O(n)$
  - B) $O(\log n)$
  - C) $O(1)$
  - D) $O(n^2)$

Sample questions

- What does the following SQL statement do?
- `SELECT * FROM Employees WHERE Age BETWEEN 30 AND 40;`
  - A) Retrieves employees exactly 30 years old
  - B) Retrieves employees under 30 and over 40 years old
  - C) Retrieves employees aged 30 to 40, inclusive
  - D) Throws an error because of incorrect syntax