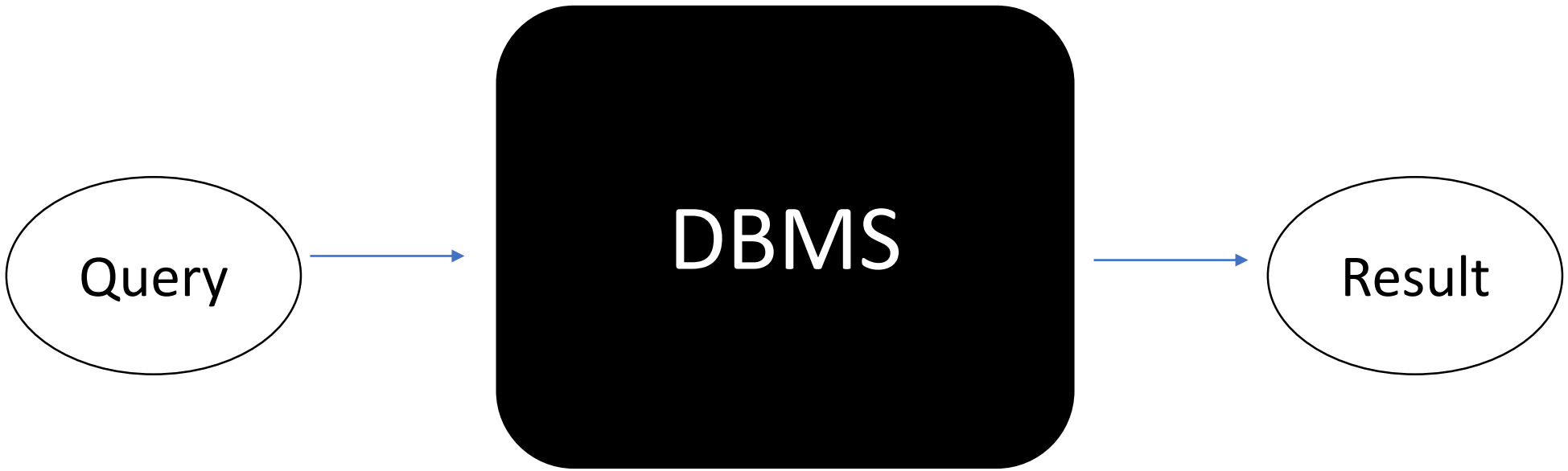
Database Internals

Leanne Wu

lewu@ucalgary.ca

Department of Computer Science

The DBMS as a black box



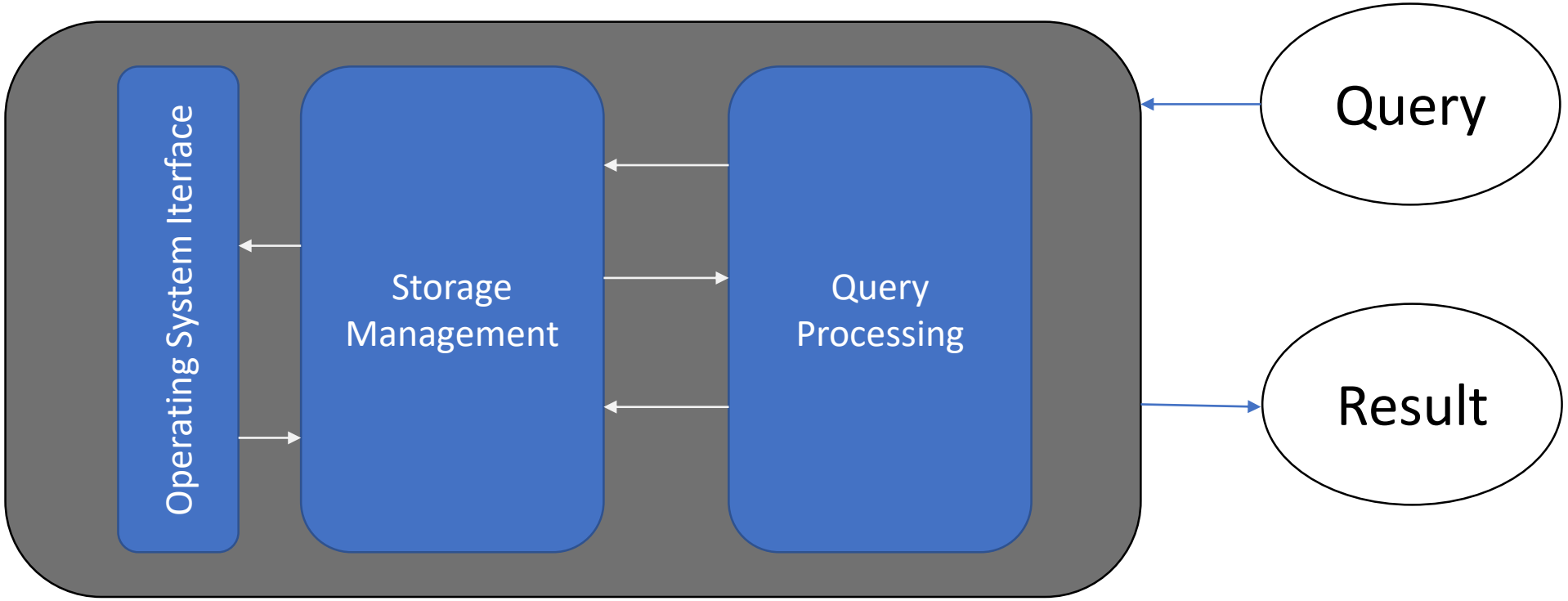
Tracing a query through the system

Consider the following query

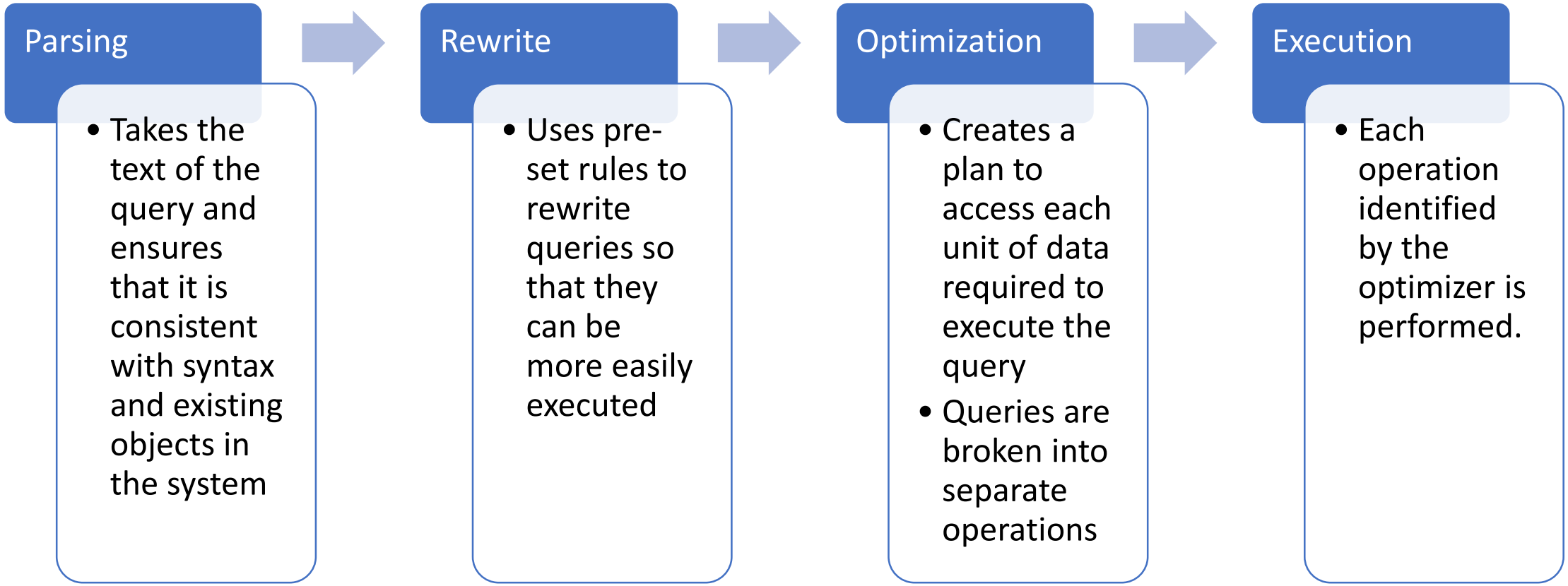
```
SELECT library, postal code
FROM library_locations
WHERE Monday_Close > ‘6:00’;
```

What happens to this query once the database management system receives it?

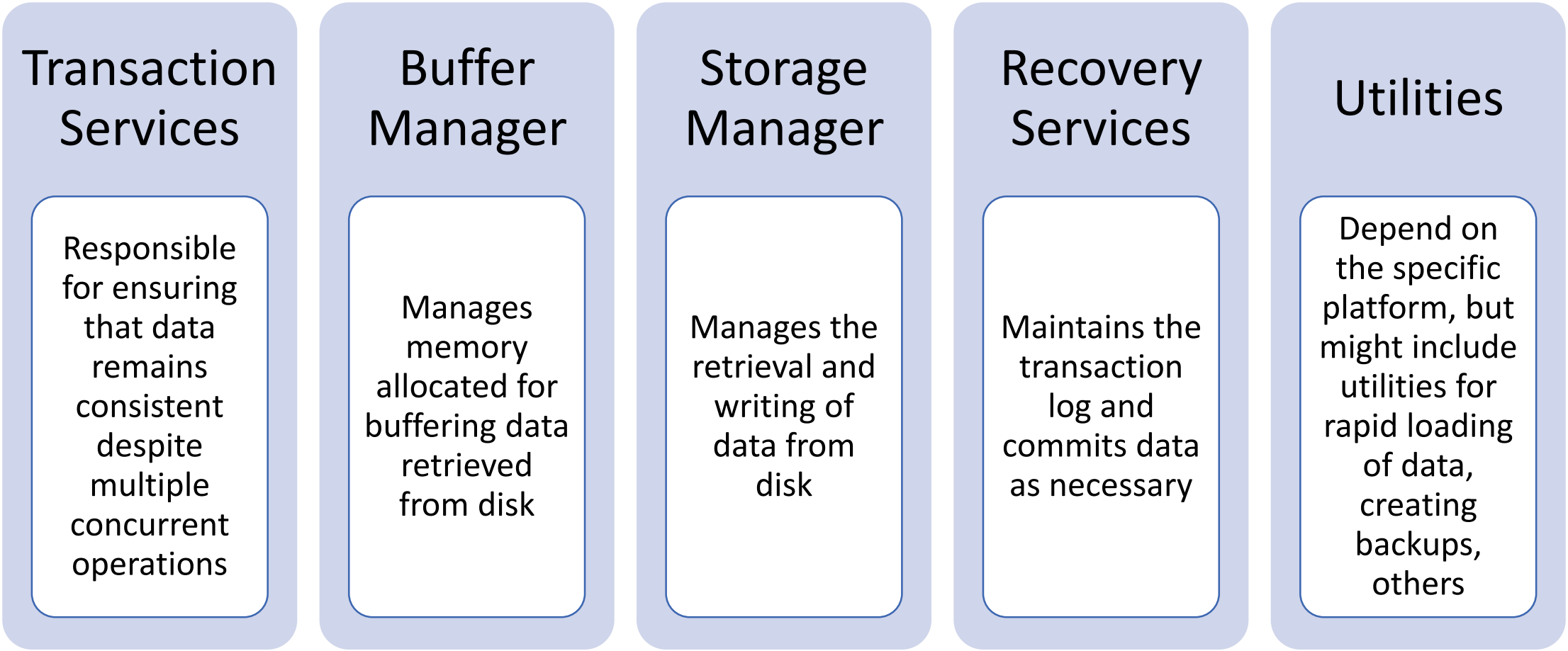
Inside the black box



Query Processing



Storage Management



Where is the data?

- The simplest format for storing data is to keep each table in its own file

- Tables are logically organized into *tablespaces*, which organize disk storage for all space used by a database
  - determines the physical location of the data
  - may determine the physical format of the data
  - may determine the size of each individual block of data
  - may determine how the operating system participates in managing physical storage

The physical organization of individual tables

- Each table is split into small blocks or pages
  - common default size of these blocks is 8 kB (but this can be changed)
    - transactional systems prefer smaller blocks
    - analytics-heavy systems may prefer larger blocks
- Data is usually not stored in any specified order
  - *Indexes* are a physical structure which may be built on specified keys to allow traversal of data in a particular order
    - Indexes are typically based on B+-trees but this varies by platform
    - *Clustered* indexes can be built to guarantee physical ordering of data
    - The creation of indexes takes processing time when records are inserted or modified and additional physical storage

Buffering

- More important on platforms other than MySQL
- Large areas of memory (commonly called: buffer, buffer cache, bufferpool) caches data which is retrieved from disk to minimize I/O
  - Cache hit ratio: ratio of read operations (i.e. how many pages or blocks) satisfied by an element in the buffer vs those which must be read by disk
  - Most platforms allow multiple buffer areas per database and can be assigned to a set of tables

Storage and Queries

- The optimizer must decide where to pull data from, how much, and at what point
  - There may be more than one possible combination of operations that can be used to retrieve data
  - Typical operations:
    - table scan: scan through entire table to retrieve values
    - index scan: use index to traverse entire table to retrieve values
  - Goal is to minimize I/O, keep cardinality (the number of rows) small, do as much work in parallel as possible

A note on NoSQL databases

- Architecture varies very much by platform
- For example: MySQL document store
  - Each collection is a table
  - Each document in a collection is its own row in the collection's table
    - Two attributes: JSON and a unique ID
- Some design rules may help to improve query performance
  - keep JSON objects relatively flat
  - put similar documents together (the semantics of collections might help to improve usability & query performance)