CS186 Week 2 Discussion Worksheet

External Sorting

1. List the differences between 2-way external merge sort and general external merge sort:

2-way utilizes 2 input buffers.

general utilizes B-1 input buffers.

2. Your system has 640 KB of memory allocated for the buffer for external sorting and you have infinite space for scratch disks. Each page holds 64 KB of data.

*Note: 1024 KB = 1 MB.*

How many pages can your buffer hold?

640 KB / 64 KB = 10 pages

How many pages are in a 4 MB file?

4 \* 1024 KB / 64 KB = 64 pages

How many passes would it take to externally merge sort a 4 MB file?

1 + ceil(log10 - 1ceil(64/10)) = 1 + ceil(log9ceil(7)) = 1 + 1 = 2 passes

How many I/O’s are needed to to externally merge sort a 4 MB file?

2 \* 2 \* 64 = 256 I/Os

What is the maximum file size that can be sorted with just 2 passes in this system?

10 \* (10 - 1) = 90 pages = 5760 KB

External Hashing

3. Why can we process B \* (B - 1) pages of data with external hashing in just two passes (divide and conquer phases)?

B-1 partitions in Pass 1 and each of them should be no more than B pages in size.

**4. If you’re processing exactly B \* (B - 1) pages of data, is it likely that you’ll have to perform recursive external hashing? Why?**

The hash function should evenly distributes the data into B - 1 partitions to process the data. This is hard to achieve. So I may assume some partitions larger than B.

**5. While you recursively perform external hashing, you reuse the same hash functions for partitioning. What’s the problem with this?**

If one partition is too big for the RAM. Reusing the same functions will not change this problem.

Single Table SQL

Songs (song\_id, song\_name, album\_num, weeks\_in\_top\_40)

Artists(artist\_id, artist\_name, first\_year\_active)

Album (album\_id, album\_name, artist\_num, year\_released, genre)

Write SQL expressions for the following queries:

6. Find the name of every song that spent more than 10 weeks in the top 40.

select song\_name from Songs

where weeks\_in\_top\_40 > 10;

7. Find the name and first year active of every artist whose name starts with the letter ‘A’.

select artist\_name, first\_year\_active from Artists

where artists\_name like “A%”;

8. Find the number of “Rap” albums released each year.

select year\_released, count(1) from Album

where genre = “Rap”

group by year\_released;

9. Find the number of albums released in each genre; don’t include genres that have a count of less than 10.

select genre, count(1) from Album

group by genre

having count(1) >= 10;