【OS】Day42

Class	Operating System: Three Easy Pieces
Date	@February 23, 2022

【Ch38】Redundant Arrays of Inexpensive Disks(2)

38.4 RAID Level 0: Striping(2)

Back To RAID-0 Analysis

Let us now evaluate the capacity, reliability, and performance of striping:

- Given N disks each of size B blocks, striping delivers $N \times B$ blocks of useful capacity
- From the standpoint of reliability, striping is perfect in a bad way. ☹️ Any disk failure will lead to data loss.
- Finally, performance is excellent: all disks are utilized, often in parallel, to service user I/O requests.

Evaluating RAID performance

We can consider two different performance metrics in analysing RAID performance.

The first is single-request latency. Understanding the latency of a single I/O request to a RAID is useful as it reveals how much parallelism can exist during a single logical I/O operation.

The second is steady-state throughput of the RAID(i.e. the total bandwidth of may concurrent requests).

Let’s assume, for this discussion, that there are two types of workloads: sequential and random.

- With a sequential workload, we assume that requests to the array come in large contiguous chunks.
- For random workloads, we assume that each request is rather small, and that each request is to a different random location on disk.

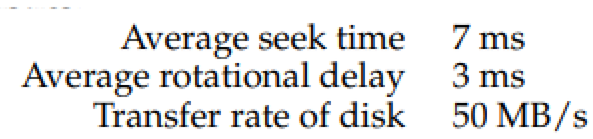
Sequential and random workloads will result in widely different performance characteristics from a disk.

- With sequential access, a disk operates in its most efficient mode, spending little time seeking and waiting for rotation and most of its time transferring data.
- With random access, just the opposite is true: most time is spent seeking and waiting for rotation and relatively little time is spent transferring data.

To capture this difference in our analysis, we will assume that a disk can transfer data at S MB/s under a sequential workload, and R MB/s when under a random workload.

Let's calculate S and R given the following disk characteristics.

Assume a sequential transfer of size 10 MB on average, and a random transfer of 10 KB on average. Also, assume the following disk characteristics:



Average seek time	7 ms
Average rotational delay	3 ms
Transfer rate of disk	50 MB/s

To compute S, we need to first figure out how time is spent in a typical 10 MB transfer. First, we spend 7 ms seeking, and then 3ms rotating.

Finally, transfer begins; 10 MB @ 50 MB/s leads to 1/5th of a second, or 200ms. Thus, for each 10 MB request, we spend 210ms completing the request. To compute S, we just need to divide:

$$S = \frac{Amount\ of\ Data}{Time\ to\ access} = \frac{10\ MB}{210\ ms} = 47.62\ MB/s$$

Because of the large time spent transferring data, S is very near the peak bandwidth of the disk.

We can compute R similarly. Seek and rotation are the same; we then compute the time spent in transfer, which is 10KB @ 50MB/s, or 0.195ms.

$$R = \frac{Amount\ of\ Data}{Time\ to\ access} = \frac{10\ KB}{10.195\ ms} = 0.981\ MB/s$$

Back To RAID-0 Analysis

Let’s now evaluate the performance of striping. It is generally good.

From a latency perspective, the latency of a single-block requests should be just about identical to that of a single disk; RAID-0 will simply redirect that request to one of its disks.

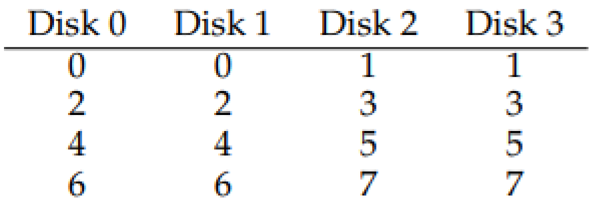
From the perspective of steady-state sequential throughput, we’d expect to get the full bandwidth of the system. Thus, throughput equals N(the number of disks) multiplied by S(the sequential bandwidth of a single disk).

For a large number of random I/Os, we can again use all of the disks, and thus obtain $N \times R$ MB/s.

38.5 RAID Level 1: Mirroring

Our first RAID level beyond striping is known as RAID level 1, or mirroring. With a mirrored system, we simply make more than one copy of each block in the system. Each copy should be placed on a separate disk. By doing so, we can tolerate disk failures.

In a typical mirrored system, we will assume that for each logical block, the RAID keeps two physical copies of it. Here is an example:



Disk 0	Disk 1	Disk 2	Disk 3
0	0	1	1
2	2	3	3
4	4	5	5
6	6	7	7

Figure 38.3: **Simple RAID-1: Mirroring**

In the example, disk 0 and disk 1 have identical contents, and disk 2 and disk 3 do as well; the data is striped across these mirror pairs.

There are a number of different ways to place block copies across the disks. The arrangement above is a common one and is sometimes called RAID-10(or RAID 1+0, stripe of mirrors) because it uses mirrored pairs(RAID-1) and then stripes(RAID-0) on top of them.

When reading a block from a mirrored array, the RAID has a choice: it can read either copy. For example, if a read to logical block 5 is issued to the RIAD, it is free to read it from either disk 2 or disk 3.

When writing a block, the RAID must update both copies of the data, in order to preserve reliability. Do note, though, that these writes can take place in parallel.

RAID-1 Analysis

- From a capacity standpoint, RAID-1 is expensive; with the mirroring level = 2, we only obtain half of our peak useful capacity. With N disks of B blocks, RAID-1 useful capacity is $(N \times B)/2$.
- From a reliability standpoint, RAID-1 does well. It can tolerate the failure of any one disk. A mirrored system(with mirroring level of 2) can tolerate 1 disk failure for certain, and up to N/2 failures depending on which disks fail. Thus, most people consider mirroring to be good for handling a single failure.
- Finally, we analyse performance. From the perspective of the latency of a single read request, we can see it is the same as the latency on a single disk; all the RAID-1 does is direct the read to one of its copies.

A write is a little different: it requires two physical writes to complete before it is done. These two writes happen in parallel, and thus the time will be roughly equivalent to the time of a single write.

To analyse stead-state throughput, let us start with the sequential workload.

- When writing out to disk sequentially, each logical write must result in two physical writes. Thus, we can conclude that the maximum bandwidth obtained during sequential writing to a mirrored array is $(N/2 \times S)$, or half the peak bandwidth.
- Unfortunately, we obtain the exact same performance during a sequential read. One might think that a sequential read could do better, because it only needs to read one copy of the data, not both.

  However, let's use an example to illustrate why this doesn't help much.

  Imagine we need to read blocks 0,1,2,3,4,5,6, and 7. Let's say we issue the read of 0 to disk 0, the read of 1 to disk 2, the read of 2 to disk 1, and the read of 3 to disk 3. We continue by issuing read to 4, 5, 6, and 7 to disks 0, 2,1, and 3.

  Consider the requests a single disk receives(say disk 0). First, it gets a request for block 0; then, it gets a request for block 4(skipping block 2). In fact, each disk receives a request for every other block. And thus, the sequential read will only obtain a bandwidth of $(N/2 \times S)$ MB/s.
- For random reads, RAID-1 delivers $N \times R$ MB/s.
- For random writes, RIAD-1 delivers $N/2 \times R$ MB/s. Each logical write must turn into two physical writes, and thus while all the disks will be in use, the client will only perceive this as half the available bandwidth.