RAID Performance Analysis

We have six 1 TB disks with 6 ms average seek time. They rotate at 5400 RPM and have a transfer rate of 40 MB/sec. The minimum unit of transfer to each disk is a 512 byte sector. Assume the controller overhead is negligible. We want to compare five arrangements of these disks:

6 independent disks
6 disks arranged in RAID-1 (mirroring)
6 disks arranged in a RAID-5 (block-interleaved, rotated parity) with a block size of 0.5 MB
6 disks arranged in a RAID-6 (block-interleaved, rotated parity, 2 parity blocks per stripe) with a block size of 0.5 MB
6 disks arranged in a RAID-10 (striping over mirroring) – block size of 0.5 MB over 3 logical drives each logical drive made up of a mirrored pair

We will use the parameter N to refer to the number of drives and in this case N = 6.
We will use the parameter DiskTransferRate = 40 MB/sec.

We will compare these arrangements for reads (large and small) and writes (large and small). Assume a small I/O is 512 bytes (S_s) and a large I/O is 30MB (S_L) (chosen so that it is evenly divisible by all 3 stripe sizes). For all answers, show your work.

1) **Capacity. (3 points)** How much real data (not redundant copies) can we store in each of these arrangements?

Table 1. Capacity

	Capacity (GB)
Independent Disks	
RAID-1	
RAID-5	
RAID-6	
RAID-10	

2) **Latency Of Individual Operations. (15 points)** Assume I/Os are distributed randomly, every I/O takes an average seek, and the disk does not reorder requests. You do not need to consider any queuing delays.
   1. How much time does it take to do an average seek and half of a rotation? Call the answer to this question AvgSeekRotate.

   2.
      - What is the total time to complete a small I/O operation on a single disk? (Note: It does not matter if it is a read or a write.) Call the answer to this question L_s.

      - What percentage of the latency of a small I/O operation is seek and rotational delay?

   3.
      - What is the total time to complete a large I/O operation on a single disk? (Note: It does not matter if it is a read or a write.) Call the answer to this question L_L.

- What percentage of the latency of a large I/O operation is seek and rotational delay?

4. Fill in the table with latencies for the specified configurations. The following questions will help you think about issues you need to consider when answering these questions.
   - How many of the boxes in the table below are the same as Ls above? Which ones? Why?
   - How many of the boxes in the table below are the same as L_L above? Which ones? Why?
   - For RAID-5, is the time for a small write the same as the time for a small read? Specifically is it L_s, L_s *2, L_s *4, L_s *6?
   - For RAID-6, is the time for a small write the same as the time for a small read? Specifically is it L_s, L_s *2, L_s *4, L_s *6?
   - For RAID-10, is the time for a small write the same as the time for a small read? Specifically is it L_s, L_s *2, L_s *4, L_s *6?
   - For RAID-5, how much data (not parity) is in each stripe?
   - For RAID-6, how much data (not parity) is in each stripe?
   - For RAID-10, how much data (not parity) is in each stripe?
   - For RAID-5, is the time for a large write the same as the time for a large read? Why or why not?
   - For RAID-6, is the time for a large write the same as the time for a large read? Why or why not?
   - For RAID-10, is the time for a large write the same as the time for a large read? Why or why not?

Here ask which of these is the same answer as L_s? 8 of 20 boxes below is just Ls above, and which of these is the same as L_L above; 4 of boxes below is just LL

Then for RAID 5 small write ask is it L_s, L_s *2, L_s *4, L_s *6,
Same for RAID 6 small write

Then for the remaining 6 ask as first what is stripe size and number of stripes and then
Ask of writes are same as reads for each of RAID-5, RAID-6 and RAID-10

Table 2. Latency

	Small read (ms)	Small write (ms)	Large read (ms)	Large write (ms)
Independent Disks				
RAID-1				
RAID-5				
RAID-6				

RAID-10				

3) **Throughput of Many Operations. (15 points)** Compute the maximum number of I/Os per second for each request type. Assume individual I/Os on the same disk cannot be overlapped. In reality, writes would include time to compute new parity and would sometimes be destined for the same disk.
   1. What is the maximum number of small I/Os per second for a single disk? Call the answer to this question T_s.

   2. What is the maximum number of large I/Os per second for a single disk. Call the answer to this question T_L.

3. Fill in the table with maximum number of I/Os per second for the specified configurations. Whenever possible, assume that the independent requests do not conflict with one another (i.e. assume that they access different disks if possible.). Note: When there is more than one disk, throughput is rarely the inverse of latency because multiple operations can occur in parallel.
    - Are any of the boxes in the table below the same as Ts above? Why or why not?
    - Are any of the boxes in the table below the same as T_L above? Why or why not?
    - Some of the boxes in the table below are the same as Ts*N above. Which ones?
    - Some of the boxes in the table below are the same as T_L *N above. Which ones?
    - For RAID-1, is the throughput for small writes the same as for small reads? Why or why not? If different how?
    - For RAID-5, is the throughput for small writes the same as for small reads? Why or why not? If different how?
    - For RAID-6, is the throughput for small writes the same as for small reads? Why or why not? If different how?
    - For RAID-10, is the throughput for small writes the same as for small reads? Why or why not? If different how?
    - For RAID-1, is the throughput for large writes the same as for large reads? Why or why not? If different how?
    - For RAID-5, is the throughput for large writes the same as for large reads? Why or why not? If different how?
    - For RAID-6, is the throughput for large writes the same as for large reads? Why or why not? If different how?
    - For RAID-10, is the throughput for large writes the same as for large reads? Why or why not? If different how?
    - For which operations are all the drives used? Why or why not?

Table 3. Throughput

	Small read (IOs/sec)	Small write (IOs/sec)	Large read (IOs/sec)	Large write (IOs/sec)
Independent Disks				
RAID-1				

RAID-5				
RAID-6				
RAID-10				

4) Qualitatively how would answers in throughput table change if we did not make the assumption that independent requests do not conflict with each other?

5) For which configurations are we less likely to get hot spots?

6) **Availability. (3 points)** How many drives can fail in each configuration without data loss? Explain your answer.

Table 4. Availability

	Best Case: Maximum Number of Disks That Can Fail Without Data Loss	Worst Case: Minimum Number of Failed Drives Needed To Cause Data Loss
Independent Disks		
RAID-1		
RAID-5		
RAID-6		
RAID-10		

7) **RAID5 Failure Case (5 points)** For the following questions, describe in detail what happens in a RAID-5 system when one disk fails. You may assume a system with 6 disks.
   1. (1 point) What happens when reading a piece of data from which the parity is stored on the failed disk?

   2. (1 point) What happens when reading a piece of data for which the actual data is stored on the failed disk?

3. (1 point) If we replace the failed disk with a new disk, how could the system be returned to its original state?

4. (2 points) How much data would need to be read and written during recovery? Estimate how long recovery would take if there were no competing requests in the system.
    a) How much total data must be read and how much written?
    b) For reconstruction of one single stripe, what of this could be done in parallel? Can the write for one stripe be done in parallel with the reads for that stripe?
    c) Could the write for one stripe be done in parallel with the reads of the next stripe?
    d) How much time would it take to read one entire drive if you assumed no seek and rotation delay?
    e) In the very best case scenario, you could imagine data being read from many drives in parallel and then the write being in done in parallel with the next batch of reads? Describe some ways in which this best case scenario is unrealistic?

5. (2 points) What might happen if we want to write a new piece of data while there is still a failed disk in the system. List at least two options. For each option, say how this would complicate your answer to the previous question.

8) **Applications (3 points)** Use the information in the preceding tables to recommend a configuration for the following workloads. Justify your answers.

1. Transaction Processing. This workload does a large number of small I/Os both reads and writes. Data integrity and availability is very important.

2. Personal home directories. These are typically stored on a file server. For file servers, capacity is important. Backups are done nightly.

3. Storage for video on demand. This workload is dominated by large reads. Availability and capacity are important.