6 Solutions

Solution 6.1

6.1.1

a.	Auto Pilot	Keypad – Input, Human Display – Output, Human Alarms – Output, Human Control Surfaces – I/O, Machine
b.	Automated Thermostat	Keypad – Input, Human Control Signals – Output, Machine

6.1.2

a.	Auto Pilot	Keypad – 0.0001 Mbit/sec Display – 800 Mbit/sec Alarms – 0.00001 (highly variable) Mbit/sec Control Surfaces – 0.1 (highly variable) Mbit/sec
b.	Automated Thermostat	Keypad – 0.0001 Mbit/sec Control Signals – 0.00001 Mbit/sec

6.1.3

a.	Auto Pilot	Keypad – Operation Rate Display – Data Rate Alarms – Operation Rate Control Surfaces – Operation Rate for most applications
b.	Automated Thermostat	Keypad – Operation Rate Control Signals – Operation Rate

Solution 6.2

6.2.1

a.	1096 days	26,304 hours
b.	2558 days	61,392 hours

6.2.2

a.	0.9990875912%
b.	0.9988272088%

6.2.3 Availability approaches 1.0. With the emergence of inexpensive drives, having a nearly 0 replacement time *for hardware* is quite feasible. However, replacing file systems and other data can take significant time. Although a drive manufacturer will not include this time in their statistics, it is certainly a part of replacing a disk.

6.2.4 MTTR becomes the dominant factor in determining availability. However, availability would be quite high if MTTF also grew measurably. If MTTF is 1000 times MTTR, the specific value of MTTR is not significant.

Solution 6.3

6.3.1

a.	14.011 ms
b.	10.025 ms

6.3.2

a.	14.022
b.	10.05

6.3.3 The dominant factor for all disks seems to be the average seek time, although RPM would make a significant contribution as well. Interestingly, by doubling the block size, the RW time changes very little. Thus, block size does not seem to be critical.

Solution 6.4

6.4.1

a.	No	An aircraft control system will process frequent requests for small amounts of information. Increasing the sector size will decrease the rate at which requests can be processed.
b.	No	A phone switch processes frequent requests for small data elements. Increasing sector size will potentially reduce performance.

6.4.2

a.	No	An aircraft control system is not typically I/O limited. Faster access to disk may be useful in some situations, but not normal operation.
b.	No	A phone switch should not be I/O limited. Faster access to disk may be useful, but may improve performance in limited scenarios.

6.4.3

a.	No	Failure in an aircraft control system is not tolerable. Increasing disk failure rate for faster data access is not acceptable.
b.	No	Failure in a phone switch is not tolerable. Increasing disk failure rate for faster data access is not acceptable.

Solution 6.5

6.5.1 There is no penalty for either seek time or for the disk rotating into position to access memory. In effect, if data transfer time remains constant, performance should increase. What is interesting is that disk data transfer rates have always outpaced improvements with disk alternatives. FLASH is the first technology with potential to catch hard disk.

6.5.2

a.	No	Increased drive performance is not an issue in an aircraft controller.
b.	No	Increased drive performance is not an issue in a phone switch.

6.5.3

a.	No	
b.	No	

Solution 6.6

6.6.1 Note that some of the specified FLASH memories are controller limited. This is to convince you to think about the system rather than simply the FLASH memory.

a.	9.77 ms
b.	10.85 ms

6.6.2 Note that some of the specified FLASH memories are controller limited. This is to convince you to think about the system rather than simply the FLASH memory.

a.	4.89 ms
b.	5.43 ms

6.6.3 On initial thought, this may seem unexpected. However, as the FLASH memory array grows, delays in propagation through the decode logic and delays propagating decoded addresses to the FLASH array account for longer access times.

Solution 6.7

6.7.1

a.	Asynchronous. The printer is electrically distant from the CPU.
b.	Asynchronous. Scanner inputs are relatively infrequent in comparison to other inputs. The scanner itself is electrically distant from the CPU.

6.7.2 For all devices in the table, problems with long, synchronous busses are the same. Specifically, long synchronous busses typically use parallel cables that are subject to noise and clock skew. The longer a parallel bus is, the more susceptible it is to environmental noise. Balanced cables can prevent some of these issues, but not without significant expense. Clock skew is also a problem with the clock at the end of a long bus being delayed due to transmission distance or distorted due to noise and transmission issues. If a bus is electrically long, then an asynchronous bus is usually best.

6.7.3 The only real drawback to an asynchronous bus is the time required to transmit bulk data. Usually, asynchronous busses are serial. Thus, for large data sets, transmission can be quite high. If a device is time sensitive, then an asynchronous bus may not be the right choice. There are certainly exceptions to this rule of thumb such as FireWire, an asynchronous bus that has excellent timing properties.

Solution 6.8

6.8.1

a.	USB due to distance from the CPU and low bandwidth requirements. FireWire would not be as appropriate due to its daisy chaining implementation.
b.	PCI due to higher throughput. No need for hot swap capabilities and the device will be close to the CPU.

6.8.2

Bus Type	Protocol
PCI	Uses a single, parallel data bus with control lines for each device. Individual devices do not have controllers, but send requests and receive commands from the bus controller through their control lines. Although the data bus is shared among all devices, control lines belong to a single device on the bus.
USB	Similar to the PCI bus except that data and control information is communicated serially from the bus controller.
FireWire	Uses a daisy chain approach. A controller exists in each device that generates requests for the device and processes requests from devices after it on the bus. Devices relay requests from other devices along the daisy chain until they reach the main bus controller.
SATA	As the name implies, Serial ATA uses a serial, point-to-point connection between a controller and device. Although both SATA and USB are serial connections, point-to-point implies that unlike USB, data lines are not shared by multiple connections. Like USB and FireWare, SATA devices are hot swappable.

6.8.3

Bus Type	Drawbacks
PCI	The parallel bus used to transmit data limits the length of the bus. Having a fixed number of control lines limits the number of devices on the bus. The trade-off is speed. PCI busses are not useful for peripherals that are physically distant from the computer.
USB	Serial communication implies longer communication distances, but the serial nature of the communication limits communication speed. USB busses are useful for peripherals with relatively low data rates that must be physically distant from the computer.
FireWire	Daisy chaining allows adding theoretically unlimited numbers of devices. However, when one device in the daisy chain dies, all devices further along the chain cannot communicate with the controller. The multiplexed nature of communication on FireWire makes it faster than USB.
SATA	The high-speed nature of SATA connections limits the length of the connection between the controller and devices. The distance is longer than PCI, but shorter than FireWire or USB. Because SATA connections are point-to-point, SATA is not as extensible as either USB or FireWire.

Solution 6.9

6.9.1 A polled device is checked by devices that communicate with it. When the devices requires attention or is available, the polling process communicates with it.

a.	No. Interface may be handled by polling, but not control or sensor inputs.
b.	Yes

6.9.2 Interrupt driven communication involves devices raising interrupts when they require attention and the CPU processing those interrupts as appropriate. While polling requires a process to periodically examine the state of a device, interrupts are raised by the device and occur when the device is ready to communicate. When the CPU is ready to communicate with the device, the handler associated with the interrupt runs and then returns control to the main process.

a.	Aircraft surfaces generate interrupts caused by movements. Controller generates signals back to control surfaces. User displays can be managed by either polling or interrupts.
b.	Polling is okay.

6.9.3 Basically, each interface is designed in a similar way with memory locations identified for inputs and outputs associated with devices.

a.	The autopilot is an input/output device. It inputs 32 single word values from various sensors on control surfaces and generates 32 single word values as control signals to actuators. Status for 32 potential alarm values is stored in one word while four words store navigational information.
b.	An automated thermostat is a simple device, but it has both input and output functions. It uses a keypad for communication to the user and on/off outputs to communicate with a furnace and air conditioner. The keypad memory should hold values input by toggle switches and numeric entries. The on/off outputs can be mapped to single bits in memory.

6.9.4

a.	The autopilot is an input/output device that requires significant I/O with a user and control surfaces. User I/O can be handled by commands that fetch input information. Similarly, control surfaces can be controlled by issuing individual commands or issuing commands with state for several sensors.
b.	An automated thermostat is a simple device, but it has both input and output functions. It uses a keypad for communication to the user and on/off outputs to communicate with a furnace and air conditioner. The keypad memory should hold values input by toggle switches and numeric entries. The on/off outputs can be mapped to single bits in memory.

6.9.5 Absolutely. A graphics card is an excellent example. A memory map can be used to store information that is to be displayed. Then, a command can be used to actually display the information. Similar techniques would work for other devices from the table.

Solution 6.10

6.10.1 Low-priority interrupts are disabled to prevent them from interrupting the handling of the current interrupt that is higher priority. The status register is saved to assure that any lower priority interrupts that have been detected are handled when the status register is restored following handling of the current interrupt.

6.10.2 Lower numbers have higher interrupt priorities.

a.	Ethernet Controller Data: 2	Mouse Controller: 3	Reboot: 1
b.	Mouse Controller: 3	Power Down: 2	Overheat: 1

6.10.3

Power Down Interrupt	Jump to an emergency power down sequence and begin execution.
Ethernet Controller Data Interrupt	Save the current program state. Jump to the Ethernet controller code and handle data input. Restore the program state and continue execution.
Overheat Interrupt	Jump to an emergency power down sequence and begin execution.
Mouse Controller Interrupt	Save the current program state. Jump to the mouse controller code and handle input. Restore the program state and continue execution.
Reboot Interrupt	Jump to address 0 and reinitialize the system.

6.10.4 If the enable bit of the Cause register is not set then interrupts are all disabled and no interrupts will be handled. Zeroing all bits in the mask would have the same affect.

6.10.5 Hardware support for saving and restoring program state prior to interrupt handling would help substantially. Specifically, when an interrupt is handled that does not terminate execution, the running program must return to the point where the interrupt occurred. Handling this in the operating system is certainly feasible, but this solution requires storing information on the stack, in registers, in a dedicated memory area, or some combination of the three. Providing hardware support removes the burden of storing program state from the operating system. Specifically, program state information need not be pulled from the CPU and stored in memory.

This is essentially the same as handling a function call, except that some interrupts do not allow the interrupted program to resume execution. Like an interrupt, a function must store program state information before jumping to its code. There are sophisticated activation record management protocols and frequently supporting hardware for many CPUs.

6.10.6 Priority interrupts can still be implemented by the interrupt handler in roughly the same manner. Higher priority interrupts are handled first and lower priority interrupts are disabled when a higher priority interrupt is being handled. Even though each interrupt causes a jump to its own vector, the interrupt system implementation must still handle interrupt signals.

Both approaches have roughly the same capabilities.

Solution 6.11

6.11.1 Yes. The CPU initiates the data transfer, but once the data transfer starts, the device and memory communicate directly with no intervention from the CPU.

6.11.2

a.	No. The dataflow back and forth from a mouse is insignificant.
b.	Possibly. One thought is the Ethernet controller handles significant amounts of data. However, that data is typically in relatively small packets. Depending on the functionality performed by the controller, it may or may not make sense to have it use DMA.

DMA is useful when individual transactions with the CPU may involve large amounts of data. A frame handled by a graphics card may be huge but is treated as one display action. Conversely, input from a mouse is tiny.

6.11.3

a.	No. The mouse controller will not use DMA.
b.	No. The Ethernet controller will not use DMA.

Basically, any device that writes to memory directly can cause the data in memory to differ from what is stored in cache.

6.11.4 Virtual memory swaps memory pages in and out of physical memory based on locations being addressed. If a page is not in memory when an address associated with it is accessed, the page must be loaded, potentially displacing another page. Virtual memory works because of the principle of locality. Specifically, when memory is accessed, the likelihood of the next access being nearby is high. Thus, pulling a page from disk to memory due to a memory access not only retrieves the memory to be accessed, but likely the next memory element being accessed.

Any of the devices listed in the table could cause potential problems if it causes virtual memory to thrash, continuously swapping in and out pages from physical memory. This would happen if the locality principle is violated by the device. Careful design and sufficient physical memory will almost always solve this problem.

Solution 6.12

6.12.1

a.	Not typically, although it is possible.
b.	Yes.

6.12.2

a.	N/A
b.	No. Online chat is dominated by transactions, not the size of those transactions.

6.12.3 See the previous problem for explanations.

a.	N/A
b.	Yes.

6.12.4 Polling would be more inappropriate for applications where numbers of transactions handled is a good performance metric. When data throughput dominates numbers of transactions, then polling could potentially be a reasonable approach.

The selection of command driven or memory mapped I/O is more difficult. In most situations, a mixture of the two approaches is the most pragmatic approach. Specifically, use commands to handle interactions and memory to exchange data. For transaction dominated I/O, command driven I/O will likely be sufficient.

Solution 6.13

6.13.1

a.	Large, concurrent data reads and writes.
b.	Large numbers of small, concurrent transactions.

6.13.2 Standard benchmarks help when trying to compare and contrast different systems. Ranking systems with benchmarks is generally not useful. However, understanding trade-offs certainly is.

6.13.3 It does not make much sense to evaluate an I/O system outside the system where it will be used. Although benchmarks help simulate the environment of a system, nothing replaces live data in a live system.

CPUs are particularly difficult to evaluate outside of the system where they are used. Again, benchmarks can help with this, but frequently Amdahl's Law makes spending resources on improving CPU speed have diminishing returns.

Solution 6.14

6.14.1 Striping forces I/O to occur on multiple disks concurrently rather than on a single disk.

a.	No, unless computations force the system to access disk frequently.
b.	No. The bottleneck in such systems is network throughput not disk I/O.

6.14.2 The MTBF is calculated as MTTF+MTTR, with MTTF as the dominating factor. For the RAID 1 system with redundancy to fail, both disks must fail. The probability of both disks failing is the product of a single disk failing. The result is a substantially increased MTBF.

In all applications, decreasing the likelihood of data loss is good. However, online database and video services are particularly sensitive to resource availability. When such systems are offline, revenue loss is immediate and customers lose confidence in the service.

6.14.3 RAID 1 maintains two complete copies of a dataset while RAID 3 maintains error correction data only. The trade-off is storage cost. RAID 1 requires two times the actual storage capacity while RAID 3 requires substantially less. This must be viewed both in terms of the cost of disks, but also power and other resources required to keep the disk array running.

In the previous applications, large online services like database and video services would definitely benefit from RAID 3. Video and sound editing may also benefit from RAID 3, but these applications are not as sensitive to availability issues as online services.

Solution 6.15

6.15.1

a.	DEE8
b.	7B25

6.15.2

a.	F030
b.	78E9

6.15.3 RAID 4 is more efficient because it requires fewer reads to generate the next parity word value. Specifically, RAID 3 accesses every disk for every data write no matter which disk is being written to. For smaller writes where data is located on a single disk, RAID 4 will be more efficient.

RAID 3 has no inherent advantages to RAID 4.

6.15.4 RAID 5 distributes parity blocks throughout the disk array rather than on a single disk. This eliminates the parity disk as a bottleneck during disk access. For applications with high numbers of concurrent reads and writes, RAID 5 will be more efficient. For lower volume, RAID 5 will not significantly outperform RAID 4.

6.15.5 As the number of disks grows by 1, the number of accesses required to calculate a parity word in RAID 3 also grows by 1. In contrast, RAID 4 and 5 continue to access only existing values of data being stored. Thus, as the number of disks grows, RAID 3 performance will continue to degrade while RAID 4 and 5 will remain constant.

There is no performance advantage for RAID 4 or 5 over RAID 3 for small numbers of disks. For two disks, there is no difference.

Solution 6.16

6.16.1

a.	8000
b.	7500

6.16.2

	16 Disks		8 Disks		4 Disks		2 Disks	
	IOPS	Bottleneck?	IOPS	Bottleneck?	IOPS	Bottleneck?	IOPS	Bottleneck?
a.	28000	No	14000	No	7000	Yes	3500	Yes
b.	14000	No	7000	Yes	3500	Yes	1750	Yes

6.16.3

	PCI Bus		DIMM		Front Side Bus	
	IOPS	Bottleneck?	IOPS	Bottleneck?	IOPS	Bottleneck?
a.	31250	No	83375	No	165625	No
b.	15625	No	41687.5	No	82812.5	No

6.16.4 The assumptions made in approximating I/O performance are extensive. From the approximation of I/O commands generated by the executing system through sequential and random I/O events handled by disks, the approximations are extensive. By benchmarking in a full system, or executing an actual application, an engineer can see actual numbers that are far more accurate than approximate calculations.

Solution 6.17

6.17.1 Runtime characteristics vary substantially from application to application. All three applications perform some kind of transaction processing, but those

transactions may be different in nature. A web server processes numerous transactions typically involving small amounts of data. Thus, transaction throughput is critical. A database server is similar, but the data transferred may be much larger. A bioinformatics data server will deal with huge data sets where transactions processed is not nearly as critical as data throughput.

When identifying the runtime characteristics of the application, you are implicitly identifying characteristics for evaluation. For a web server, transactions per second is a critical metric. For the bioinformatics data server, data throughput is critical. For a database server, you will want to balance both criteria.

6.17.2 It is relatively easy to use online resources to identify potential servers. You may also find advertisements in periodicals from your professional societies or trade journals. You should be able to identify one or more candidates using the criteria identified in 6.17.1. If your reasons for selecting the server don't follow from the criteria in 6.17.1, something is not right.

6.17.3 In Problem 6.16, we used characteristics of a Sun Fire x4150 to attempt to predict its performance. You can use the same data and characteristics here. Remember that the Sun Fire x4150 has multiple configurations. You should consider this when you perform your evaluation.

Find similar measurements for the server that you have selected. Most of this data should be available online. If not, contact the company providing the server and see if such data is available.

It's a reasonably simple task to use a spreadsheet to evaluate numerous configurations and systems simultaneously. If you design your spreadsheet carefully, you can simply enter a table of data and make comparisons quickly. This is exactly what you will do in industry when evaluating systems.

6.17.4 Although analytic analysis is useful when comparing systems, nothing beats hands-on evaluation. There are a number of test suites available that will serve your needs here. Virtually all of them will be available online. Look for benchmarks that generate transactions for the web server, those that generate large data transfers for the bioinformatics server, and a combination of the two for the database server.

Solution 6.18

6.18.1

a.	8.76
b.	9.125

6.18.2

	7 Years	10 Years
a.	26.28	189.8
b.	32.85	237.25

6.18.3 Average failure rates of the drives with longer longevity for 7 and 10 years are:

	7 Years	10 Years
a.	16.06	36.5
b.	12.775	38.325

It is not surprising that with failure rates starting to double 3 years later, we have to replace far fewer disks in the second situation than the first. The ratio of the number of drives replaced in the first scenario to the number replaced in the second should give us the multiple that we want:

	7 Years	10 Years
a.	1.64	5.2
b.	2.57	6.19

Solution 6.19

6.19.1 In all cases, no. The objective of the customer is not known. Thus, improving any performance metric by nearly doubling the cost may or may not have a price impact on the company.

6.19.2 As a search engine provider paid by ad hits, throughput is critical. Most HTTP traffic is small, so the network is not as great a bottleneck as it would be for large data transfers. RAID 0 may be an effective solution. However, RAID 1 will almost certainly not be an effective solution. Increased availability makes our product more attractive, but a 1.6 cost multiple is most likely too high.

RAID 0 is going to increase throughput by 70%, meaning the potential exists to serve 1.7 times as many ads. The cost of this gain is 0.6 of the original price. 1.7 times as many ads for 1.6 times the original cost may justify the upgrade cost.

6.19.3 This problem is not as simple as it would seem at first glance. As an online backup provider, availability is critical. Thus, using RAID 1 where failure

rate decreases for a 1.6 times cost increase might be worthwhile. However, online backup is more appealing when services are provided quickly making RAID 0 appealing. Remember Amdahl's law. Will increasing throughput in the disk array for long data reads and writes result in performance improvements for the system? The network will be our throughput bottleneck, not disk access. RAID 0 will not help much.

RAID 1 has more potential for increased revenue by making the disk array available more. For our original configuration, we are losing between 12 and 19 disks per 1000 to 1500 every 7 years. If the system lifetime is 7 years, the RAID 1 upgrade will almost certainly not pay for itself even though it addresses the most critical property of our system. Over 10 years, we lose between 30 and 50 drives. If repair times are small, then even over a 10-year span the RAID 1 solution will not be cost effective.

Solution 6.20

6.20.1 The approach to solving this problem is relatively simple once parameters of a bioinformatics simulation are understood. Simulations tend to run days or months. Thus, losing simulation data or having a system failure during simulation are catastrophic events. Availability is therefore a critical evaluation parameter. Additionally, the disk array will be accessed by 1000 parallel processors. Throughput will be a major concern.

The primary role of the power constraint in this problem is to prevent simply maximizing all parameters in the disk array. Adding additional disks and controllers without justification will increase power consumption unnecessarily.

6.20.2 Remember that your system must provide both backup and archiving. Thus, you will need multiple copies of your data and may be required to move those copies offsite. This makes none of the solutions optimal.

RAID or a second backup array provides high-speed backup, but does not provide archival capabilities. Magnetic tape allows archiving, but can be exceptionally slow when comparing to disk backups. Online backup automatically achieves archiving, but can be even slower than disks.

6.20.3 Your benchmarks must evaluate backup throughput. Most other parameters that govern selection of a system are relatively well understood—portability and cost being the primary issues to be evaluated.