AP Exam Review 4: Binary Numbers, Compression, Problems, Parallelism Simulations & Data

Bits & Binary Numbers

AP EXAM VOCABULARY

Binary digits

- A **bit** (stands for **binary digit**) is a single unit of data that represents one of two values.
- **One single bit can represent: true** or **false, on** or **off, yes** or **no, 1** or **0**
- **All** digital data that you have on your computers (numbers, text, images, sound, video, etc.) is stored as **binary data**, specifically as a **binary sequence** of **bits** which are all either **1** or **0**
- For example, the *text string* "`hello`" is represented on your computer by this **binary sequence** of **40** individual bits:

`0110100001100101011011000110110001101111` ← **40** bits to represent "hello" in binary

- The *number* `5723` is represented by this **binary sequence** of **13** bits:

`1011001011011` ← **13** bits to represent 5723 in binary

5

- A **byte** is **8 bits**. The term "byte" is used more commonly when talking about computer memory because the physical hardware of memory stores binary data in groups of 8 bits (bytes); you're probably familiar with the terms gigabyte, megabyte, and kilobyte.
- This **binary sequence** of **40 bits** for "hello" is **5 bytes** (5 groups of 8 bits):

`01101000` `01100101` `01101100` `01101100` `01101111`

- A **bit** can represent **two** possible values: 1 (on) or 0 (off)

0	**off**
1	**on**

1 bit gives us 2 different values:
`0, 1`

- If we need to represent more than two possible values, we can do this using *more than one bit!*
- For example, we can represent 4 different values using 2-bits

bit 2	bit 1
0	**0**
0	**1**
1	**0**
1	**1**

2 bits gives us 4 different values:
`00, 01, 10, 11`

- Possible values for a 3-bit sequence:

bit 3	bit 2	bit 1
0	**0**	**0**
0	**0**	**1**
0	**1**	**0**
0	**1**	**1**
1	**0**	**0**
1	**0**	**1**
1	**1**	**0**
1	**1**	**1**

3 bits gives **8** possible values:
`000`
`001`
`010`
`011`
`100`
`101`
`110`
`111`

What's the pattern?

0
1

1 bit gives us 2 different values:
0, 1

$2^1 = 2$ values

0	0
0	1
1	0
1	1

2 bits gives us 4 different values:
00, 01, 10, 11

$2^2 = 4$ values

0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

3 bits gives **8** possible values:
000
001
010
011
100
101
110
111

$2^3 = 8$ values

A. What does "adding one more bit" do to the total number of possible values that can be represented? **doubles it**

B. How many values could we represent with 4 bits? **16** (= 8 doubled!) $2^4 = 16$ values

C. **Generalize:** What's the relationship between number of bits and number of values that can be represented with those bits? 2^n **where n = number of bits**

20

4 bits = 2^4 = 16 possible values

bit 4	bit 3	bit 2	bit 1
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

0000
0001
0010
0011
0100
0101
0110
0111
1000
1001
1011
1100
1101
1110
1111

What could we represent with 16 values?

Would 4 bits, or 16 values, be enough to represent the 26 the letters of the alphabet?

Why or why not?

No! Because $2^4 = 16$ and $16 < 26$ needed to represent the letters of the alphabet.

0	0	0	0	A
0	0	0	1	B
0	0	1	0	C
0	0	1	1	D
0	1	0	0	E
0	1	0	1	F
0	1	1	0	G
0	1	1	1	H
1	0	0	0	I
1	0	0	1	J
1	0	1	0	K
1	0	1	1	L
1	1	0	0	M
1	1	0	1	N
1	1	1	0	O
1	1	1	1	P

What is the minimum number of bits we would need to represent 26 different values?

5 bits! Since each additional bit DOUBLES the amount of possible values, 5 bits would give us double the amount of possible values that 4 bits does (and 4 bits gives us 16), so it would give us 32 possible values. 32 values is enough for 26 values (and we would have 6 unused).

Alternatively, you could do $2^5 = 32$ and see that 5 is the minimum power of 2 that is greater than 26

Representing 26 values with 5 bits (32 values)

0	0	0	0	0	A
0	0	0	0	1	B
0	0	0	1	0	C
0	0	0	1	1	D
0	0	1	0	0	E
0	0	1	0	1	F
0	0	1	1	0	G
0	0	1	1	1	H
0	1	0	0	0	I
0	1	0	0	1	J
0	1	0	1	0	K
0	1	0	1	1	L
0	1	1	0	0	M
0	1	1	0	1	N
0	1	1	1	0	O
0	1	1	1	1	P

1	0	0	0	0	Q
1	0	0	0	1	R
1	0	0	1	0	S
1	0	0	1	1	T
1	0	1	0	0	U
1	0	1	0	1	V
1	0	1	1	0	W
1	0	1	1	1	X
1	1	0	0	0	Y
1	1	0	0	1	Z
1	1	0	1	0	
1	1	0	1	1	
1	1	1	0	0	
1	1	1	0	1	
1	1	1	1	0	
1	1	1	1	1	

True or false: 3 bits would be enough to store 9 unique values in memory. *If false, correct it to make it true!*	
What is the fewest number of bits you would need to represent 80 different values?	
How many *different* values can be represented with: A. 3 bits B. 4 bits C. 8 bits D. 9 bits E. 16 bits	A. B. C. D. E.
How does the number of possible values change with the addition of *one* more bit?	
How many times *more* values can be represented with: A. 8 bits than with 4 bits? B. 16 bits than with 15 bits? C. 32 bits than with 31 bits	**A**. **B**. **C**.

a. How many **bits** does this binary sequence of bits represent? `100101100011010101100101` **b.** How many **bytes** does it represent?	

Check answers

AP Exam Practice Question

Which of the following can be represented by a single binary digit?
Select two answers.

(A) The position of the minute hand of a clock

(B) The remainder when dividing a whole number by 2

(C) The value of a Boolean variable

(D) The volume of a car radio

Your answers (select two):

Check your answer!

AP Exam Practice Question

A store uses binary numbers to assign a unique binary sequence to each item in its inventory. What is the minimum number of bits required for each binary sequence if the store has 600 items in its inventory?

A. 9 bits

B. 10 bits

C. 600 bits

D. 1024 bits

Your answer:

Check your answer!

AP Exam Practice Question

A video game character can face toward one of four directions: north, south, east, and west. Each direction is stored in memory as a sequence of four bits. A new version of the game is created in which the character can face toward one of eight directions, adding northwest, northeast, southwest, and southeast to the original four possibilities. Which of the following statements is true about how the eight directions must be stored in memory?

(A) Four bits are not enough to store the eight directions. Five bits are needed for the new version of the game.

(B) Four bits are not enough to store the eight directions. Eight bits are needed for the new version of the game.

(C) Four bits are not enough to store the eight directions. Sixteen bits are needed for the new version of the game.

(D) Four bits are enough to store the eight directions.

My answer:

Check!

AP Exam Practice Question

Which of the following can be represented by a single binary digit?

Select two answers.

[A] The position of the minute hand of a clock

[B] The remainder when dividing a whole number by 2

[C] The value of a Boolean variable

[D] The volume of a car radio

My answer:

Check!

AP Exam Practice Question

An online store uses 6-bit binary sequences to identify each unique item for sale. The store plans to increase the number of items it sells and is considering using 7-bit binary sequences. Which of the following best describes the result of using 7-bit sequences instead of 6-bit sequences?

A. 2 more items can be uniquely identified.

B. 10 more items can be uniquely identified.

C. 2 times as many items can be uniquely identified.

D. 10 times as many items can be uniquely identified.

My answer:

Check!

Converting between Binary & Decimal

Decimal Number System (dec = 10)

- Humans generally use **base 10** ("decimal") digits to write numbers.
- In **base 10**, there are **ten digits** (0-9), and *each place is worth **ten times*** as much as the place to its right (powers of 10):

Example:

		6	7	5	0	2
1000000's (10^6)	**100000's** (10^5)	**10000's** (10^4)	**1000's** (10^3)	**100's** (10^2)	**10's** (10^1)	**1's** (10^0)

$$67502 = (6 \times 10000) + (7 \times 1000) + (5 \times 100) + (0 \times 10) + (2 \times 1)$$
$$= (6 \times 10^4) + (7 \times 10^3) + (5 \times 10^2) + (0 \times 10^1) + (2 \times 1^0)$$
$$= 60000 + 7000 + 500 + 0 + 2$$
$$= 67502$$

Binary Number System (bi = 2)

- ***Computers*** use **base 2** ("binary") digits to write numbers.
- In **base 2**, there are **two digits** (0-1), and *each place is worth **two times*** as much as the place to its right (powers of 2):

Example:

1	0	1	1	0	1	0
64's (2^6)	**32's** (2^5)	**16's** (2^4)	**8's** (2^3)	**4's** (2^2)	**2's** (2^1)	**1's** (2^0)

$$1011001 = (1 \times 64) + (0 \times 32) + (1 \times 16) + (1 \times 8) + (0 \times 4) + (1 \times 2) + (0 \times 1)$$
$$= (1 \times 2^6) + (0 \times 2^5) + (1 \times 2^4) + (1 \times 2^3) + (0 \times 2^2) + (1 \times 2^1) + (0 \times 2^0)$$
$$= 64 + 0 + 16 + 8 + 0 + 2 + 0$$
$$= 90$$

$1011001_2 = 90_{10}$ ➡ "The binary representation of the decimal number 90 is 1011001"

Converting from Binary to Decimal

Example: Determine what decimal (base 10) number is represented by the binary (base 2) number `110111`

(in other words, convert `110111` to decimal)

Step 1: Write out the binary number:

1	1	0	1	1	1
32	16	8	4	2	1

Step 2: Write the powers of 2 directly underneath, right to left, starting with 1 (2^0), followed by 2 (2^1), 8 (2^3), etc.

Step 3: Add up all the powers of 2 that have a 1 in the place value

$32 + 16 + 4 + 2 + 1 = 55$

➡ `110111` is 55 in base 10 (decimal)

Incrementing in Binary

- In the binary number system, when we increment (add 1), the carries happen every 2 -- not 10! -- so things move more quickly

						Value in base 10: 52
1	1	0	1	0	0	
						Value in base 10: 53
1	1	0	1	0	1	
				+1		Value in base 10: 54
1	1	0	1	1	0	
						Value in base 10: 55
1	1	0	1	1	1	
		+1		+1	+1	Value in base 10: 56
1	1	1	0	0	0	
						Value in base 10: 57
1	1	1	0	0	1	

				+1			Value in base 10:
1	1	1	0	1	0		58
							Value in base 10:
1	1	1	0	1	1		59
			+1	+1			Value in base 10:
1	1	1	1	0	0		60
							Value in base 10:
1	1	1	1	0	1		61
				+1			Value in base 10:
1	1	1	1	1	0		62
							Value in base 10:
1	1	1	1	1	1		63
+1	+1	+1	+1	+1	+1		Value in base 10:
1	0	0	0	0	0	0	64

We need another 2's place (bit)!

							Value in base 10:
1	0	0	0	0	0	1	65
					+1		Value in base 10:
1	0	0	0	0	1	0	66
							Value in base 10:
1	0	0	0	0	1	1	67
				+1	+1		Value in base 10:
1	0	0	0	1	0	0	68
							Value in base 10:
1	0	0	0	1	0	1	69

From Decimal to Binary

- So we converted *from* 110111 (base 2) *to* 55 (base 10)
- But how do we go the *other* way? *from* 55 (base 10) to 110111 (base 2)?

Example: Convert the decimal value 55 (base 10) to binary (base 2)

Step 1: Write the powers of 2, right to left, starting with 1 (2^0), followed by 2 (2^1), 4 (2^2), 8 (2^3), etc. until you get to one that is BIGGER than the number

64 32 16 8 4 2 1

64 > 55, so we leave it out!

32 16 8 4 2 1

32 16 8 4 2 1

Step 2: Subtract the largest power of 2 from your number and place a 1 in that position

55
− 32
23

32 16 8 4 2 1

1

Step 2: Subtract the largest power of 2 from your number and place a 1 in that position

55
− 32
23

Step 3: *Try* to subtract the *next* largest power of 2 from the *remainder*; if it's possible, put a 1 in that position, if not, put a 0

32	16	8	4	2	1
1					

$$\begin{array}{rr} 55 & 23 \\ -\ 32 & -\ 16 \\ \hline 23 & 7 \end{array}$$

Step 3: *Try* to subtract the *next* largest power of 2 from the *remainder*; if it's possible, put a 1 in that position, if not, put a 0

32	16	8	4	2	1
1	1				

$$\begin{array}{rr} 55 & 23 \\ -\ 32 & -\ 16 \\ \hline 23 & 7 \end{array}$$

Step 4: Repeat step 3 for the next power of two and the most recent remainder.

32	16	8	4	2	1
1	1				

$$\begin{array}{rrr} 55 & 23 & 7 \\ -\ 32 & -\ 16 & -\ 8 \\ \hline 23 & 7 & \end{array}$$

Step 4: Repeat step 3 for the next power of two and the most recent remainder.

32	16	8	4	2	1
1	1				

$$\begin{array}{rrr} 55 & 23 & 7 \\ -\ 32 & -\ 16 & -\ 8 \\ \hline 23 & 7 & \end{array}$$

Doesn't work! You can't subtract 8 from 7

Step 4: Repeat step 3 for the next power of two and the most recent remainder.

32	16	8	4	2	1
1	1	0			

55	23	7	
− 32	− 16	− 8	
23	7		PUT A 0 since you can't subtract 8 from 7

Step 5: Repeat step 3 for the next power of two and the most recent remainder.

32	16	8	4	2	1
1	1	0			

55	23	7	7
− 32	− 16	− 8	− 4
23	7		3

32	16	8	4	2	1
1	1	0	1		

55	23	7	7
− 32	− 16	− 8	− 4
23	7		3

Step 6: Continue in this fashion until you have done all powers of 2 and you have a 1 or 0 in each place.

32	16	8	4	2	1
1	1	0	1		

55	23	7	7	3
− 32	− 16	− 8	− 4	− 2
23	7		3	1

32	16	8	4	2	1
1	1	0	1	1	

Step 6: Continue in this fashion until you have done all powers of 2 and you have a 1 or 0 in each place.

55	23	7	7	3	
− 32	− 16	− 8	− 4	− 2	
23	7		3	1	

32	16	8	4	2	1
1	1	0	1	1	

55	23	7	7	3	1
− 32	− 16	− 8	− 4	− 2	− 1
23	7		3	1	0

32	16	8	4	2	1
1	1	0	1	1	1

55	23	7	7	3	1
− 32	− 16	− 8	− 4	− 2	− 1
23	7		3	1	0

32	16	8	4	2	1
1	1	0	1	1	1

55	23	7	7	3	1
− 32	− 16	− 8	− 4	− 2	− 1
23	7		3	1	0

Done! We have 1's and 0's for all powers of 2

The decimal value 55 in binary is 110111

You can always check your work by converting *back* to decimal:

1	1	0	1	1	1
32	16	8	4	2	1

32 + 16 + 4 + 2 + 1 = 55 ✅

Translate/convert each **binary** (base 2) representation below *into* **decimal** (base 10)

Binary	Decimal
A. 10	
B. 101	
C. 111	
D. 1000	
E. 1011	
F. 11000	

[Check answers]

Complete the table below with binary representations for all decimal numbers 0 - 15. Use the "incrementing" strategy we discussed in class to "carry" the 1's.
Note that a "left zero padded" binary number like 0011 is the SAME as 11.

Decimal	Binary		Decimal	Binary
0	0000		**8**	
1	0001		**9**	
2	0010		**10**	
3			**11**	
4			**12**	
5			**13**	
6			**14**	
7			**15**	

True or false: The decimal value 16 can be represented in binary using 4 bits.	

Check answers

Translate/convert each **decimal** (base 10) representation below *into* **binary** (base 2)	
Decimal	**Binary**
A. 27	
B. 28	
C. 63	
D. 64	

Check answers

AP Exam Practice Question

Consider the 4-bit binary numbers 0011, 0110, and 1111. Which of the following decimal values is NOT equal to one of these binary numbers?

- (A) 3
- (B) 6
- (C) 9
- (D) 15

My answer:	

Check!

AP Exam Practice Question

Each student that enrolls at a school is assigned a unique ID number, which is stored as a binary number. The ID numbers increase sequentially by 1 with each newly enrolled student. If the ID number assigned to the last student who enrolled was the binary number 1001 0011, what binary number will be assigned to the next student who enrolls?

A 1001 0100

B 1001 0111

C 1101 0100

D 1101 0111

My answer:

Check!

AP Exam Practice Question

ASCII is a character-encoding scheme that uses 7 bits to represent each character. The decimal (base 10) values 65 through 90 represent the capital letters A through Z, as shown in the table below.

Decimal	ASCII Character
65	A
66	B
67	C
68	D
69	E
70	F
71	G
72	H
73	I
74	J
75	K
76	L
77	M

Decimal	ASCII Character
78	N
79	O
80	P
81	Q
82	R
83	S
84	T
85	U
86	V
87	W
88	X
89	Y
90	Z

What ASCII character is represented by the binary (base 2) number 1001010 ?

A H

B I

C J

D K

My answer:

Check!

AP Exam Practice Question

Consider the following numeric values.

- Binary 1011
- Binary 1101
- Decimal 5
- Decimal 12

Which of the following lists the values in order from least to greatest?

A Decimal 5, binary 1011, decimal 12, binary 1101

B Decimal 5, decimal 12, binary 1011, binary 1101

C Decimal 5, binary 1011, binary 1101, decimal 12

D Binary 1011, binary 1101, decimal 5, decimal 12

My answer:	

Check!

Overflow Errors

AP EXAM VOCABULARY

- **Overflow error:** An error that occurs when a computer attempts to handle a number outside the possible range.

Overflow

- Let's say a simple calculator computer program uses **4 bits** to represent a number:

- What would happen if the user tried to calculate 12 + 6?

  12 + 6 = 18 = `10010` in binary, which requires 5 bits!

  Since our computer program uses only 4 bits for numbers, the decimal value 18 could not be stored; the maximum possible number it can hold is `1111` (15) with 4 bits.

  So if the user tried to enter 12 + 6 into the calculator, what would happen is called an **overflow error, which is an error that occurs when a computer attempts to handle a number outside the possible range.**

Overflow

- When an **overflow error** occurs, the computer *might* "reset" and "rollover" the extra value:

  `18 - 15 = 3` → `3` in binary is `11`, which the computer could "zero pad" and store as `0011`:

0	0	1	1

This would lead to 3 being output as the result to 12 + 6, instead of 18! 😱

- **Or**, the computer program might just crash and quit on you!
- Either way, overflow errors are not good and lead to unexpected results!

Practice AP Exam Question

A particular program uses 3 bits to represent whole numbers. When that program adds the numbers 3 and 6, the result is given as 0. Identify the best explanation of the result.

A. The result was incorrect due to data being corrupted during the operation due to a technical glitch.

B. The result was incorrect due to a round-off error.

C. The result was incorrect due to an overflow error.

D. The result is correct when all values are converted to binary.

Your answer:

[ ]

Check your answer!

AP Exam Practice Question

A certain programming language uses 4-bit binary sequences to represent nonnegative integers. For example, the binary sequence `0101` represents the corresponding decimal value `5`. Using this programming language, a programmer attempts to add the decimal values `14` and `15` and assign the sum to the variable `total`. Which of the following best describes the result of this operation?

(A) The correct sum of `29` will be assigned to the variable `total`.

(B) An overflow error will occur because 4 bits is not large enough to represent either of the values `14` or `15`.

(C) An overflow error will occur because 4 bits is not large enough to represent `29`, the sum of `14` and `15`.

(D) A round-off error will occur because the decimal values `14` and `15` are represented as approximations due to the fixed number of bits used to represent numbers.

My answer: [ ]

Check!

Storing Decimals, Text, Images & Sound as Binary

AP EXAM VOCABULARY

Decimals as Binary

- Computers use **floating-point representation** to represent decimals, fractions, and all *non-integer numbers* (like 2/5, 1.2641, 9.9999999, or pi) in binary.

In binary, *only fractions whose denominator is a power of 2 can be represented exactly because computers have a fixed number of bits.* So 1/16 can be represented exactly in binary, but 1/10 can't. When 1/10 is converted to binary, there is an infinitely repeating sequence of 1's and 0's that occurs, which needs to be *rounded off* because a computer does *not* have an unlimited number of bits. So the binary representation of 0.1 is just slightly too big. **Doing math with values that can't be exactly represented in binary using a fixed number of bits can lead to round-off errors.**

- **Round-off errors** occur when the computer *tries* to represent a some non-integer values exactly using a fixed number of bits; this leads to weird results like:

```
print(0.2 + 0.4)
```
prints
```
0.6000000000000001
```
instead of 0.6

*Note that **round-off** errors are **different** from **overflow** errors.*

Text as Binary

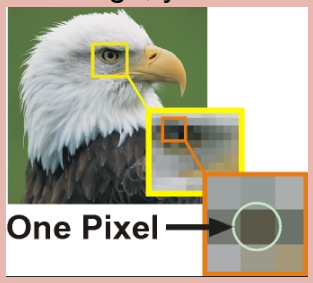
How do **text strings** (letters and characters) get stored in binary? For example:
"**hello**" in binary is a 40-bit binary sequence: `0110100001100101011011000110110001101111`

Each character in "**hello**" is a **byte** (8 bits):
The 5 bytes of "hello" are `01101000 01100101 01101100 01101100 01101111`

Characters are translated to binary using the **Unicode** table, which has binary representations for over 65000 characters. You can see the Unicode table here, which uses "UTF-8," an international standard encoding scheme used to convert characters to binary using Unicode. It's an agreed upon table (kind of like a protocol!). **You do not need to know the details of Unicode or UTF-8, but you *should* know that text is converted to binary using "translation" tables that are internationally accepted standards that the world uses.**

Images as Binary

Images are just a bunch of **pixels**, and each pixel is just a "single square of color." Below is an image of an eagle, and if you zoom in far enough, you can see each individual pixel:



Each pixel represents a color, and each color in the spectrum can be converted to three numbers using the **RGB color model**, which is an *additive* color model in which red (R), green (G), and blue (B) light are added together in various ways to reproduce a broad array of colors. Every color can be represented using three numbers, one for R ("how much red"), one for G ("how much green"), and one for B ("how much blue"). Each value is on the scale 0-255, where 0 is none of that color, and 255 is the maximum amount of that color.

Here are some examples.

Each pixel is a set of three numbers, for example **(183, 72, 201)**.

These three *decimal* numbers can be converted to *binary*:

183 in binary = 10110111
72 in binary = 1001000 → 01001000 with zero padding to make 8 bits (1 byte)
201 in binary = 11001001

So a *single purple pixel* could be represented with this 3-byte (24-bit) sequence:

`10110111 010010001 1001001`

But to convert an entire *image* to binary, you need to repeat this process for *every pixel.*

It turns out, the *same* sequence of bits can represent *different* types of data in different contexts. For example, the same binary sequence shown below can either be a *number* or *text*, depending on how the data is being used, for example:

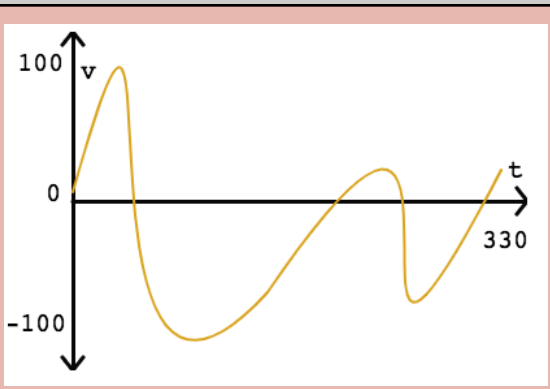
The binary sequence **011100110110101101111001**
converted to a *decimal number* is **2939600**

The *same* binary sequence **011100110110101101111001**
converted to *text* is "**sky**"

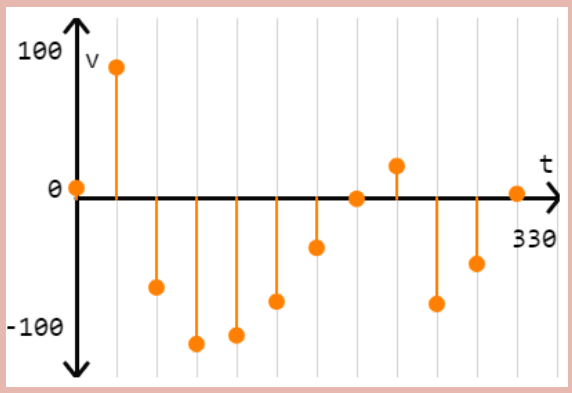
Sound as Binary

Not all data are naturally digital (that is, they may not be individual values that can be represented in the form of binary sequences). Some real-world values, such as the pitch and volume of music, the colors of a painting, or the position of a sprinter during a race, ***change smoothly over time*** or position; they are **analog**. When analog data is encoded digitally (as bits on a computer), their values are *approximated* by **sampling** the data. **Sampling** means measuring values of the analog signal at regular intervals called **samples**. The values of the samples are converted to bits to be stored digitally.

Here is an **analog** waveform representing part of an audio song (measured in volts):



If we want to store this song on our computer or stream it through Spotify, it needs to get converted to binary data (0's and 1's). This is done by **sampling** it, or measuring the voltage at a specific interval, say every 30 ms:



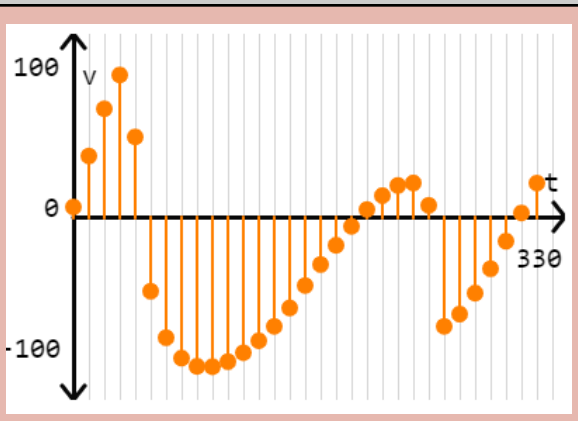
A **sampling** interval of 30 ms gives us 12 **samples** of the signal between 0 and 330 milliseconds. We can capture these 12 samples as a series of points (time, voltage):

(0, 7)
(30, 96)
(60, -71)
(90, -106)
(120, -97)
(150, -70)
(180, -29)
(210, 6)
(240, 24)
(270, -74)
(300, -31)
(330, 24)

This sample data is now just ***numbers***. **And we know how to convert numbers to binary.** You might note that there are negative numbers above; for this, computers use a "sign bit" to indicate a positive or negative number (i.e. 0 = positive, 1 = negative). Don't worry how this works, just know that storing negative numbers as binary can be done.

If you want a ***better*** quality digital sound -- i.e. a digital version that sounds more closely like the actual song -- you use a **faster sampling rate; i.e. take more samples per second.**

For example, if in the song above we took 36 samples (every 10 ms) instead of 12 samples (every 30 ms), it would give a more realistic representation of the audio waveform:



Sampling sound (or *any* **analog** signal) creates a **digital approximation** of that sound, but it's not exactly the same thing; *some* of the quality is lost when converting from an analog to digital signal through sampling (faster sampling rates → higher quality digital approximations → less quality lost).

AP Exam Practice Question

Which of the following are true statements about the data that can be represented using binary sequences?

I. Binary sequences can be used to represent strings of characters.
II. Binary sequences can be used to represent colors.
III. Binary sequences can be used to represent audio recordings.

A) I only

B) I and II only

C) II and III only

D) I, II, and III

My answer:

Check!

AP Exam Practice Question

The position of a runner in a race is a type of analog data. The runner's position is tracked using sensors. Which of the following best describes how the position of the runner is represented digitally?

A. The position of the runner is determined by calculating the time difference between the start and the end of the race and making an estimation based on the runner's average speed.

B. The position of the runner is measured and rounded to either 0 or 1 depending on whether the runner is closer to the starting line or closer to the finish line.

C. The position of the runner is predicted using a model based on performance data captured from previous races.

D. The position of the runner is sampled at regular intervals to approximate the real-word position, and a sequence of bits is used to represent each sample.

My answer:

Check!

AP Exam Practice Question

A student is recording a song on her computer. When the recording is finished, she saves a copy on her computer. The student notices that the saved copy is of lower sound quality than the original recording. Which of the following could be a possible explanation for the difference in sound quality?

A. The song was saved using fewer bits per second than the original song.

B. The song was saved using more bits per second than the original song.

C. The song was saved using a lossless compression technique.

D. Some information is lost every time a file is saved from one location on a computer to another location.

My answer:

Check!

AP Exam Practice Question

A new program was developed for a bank. While testing the program by adding and subtracting decimal values, the developer discovers that some results appear to be imprecise. Which of the following is the most likely cause?

A. The numbers are represented using an unlimited number of bits, resulting in round-off errors.

B. The numbers are represented using an unlimited number of bits, resulting in overflow errors.

C. The numbers are represented using a fixed number of bits, resulting in round-off errors.

D. The numbers are represented using a fixed number of bits, resulting in overflow errors.

My answer:	

Check!

Data Compression

AP EXAM VOCABULARY

- **Data compression**: the process of **reducing the number of bits** that are used to store text, images, sounds, movies, and other kinds of data.
- **Lossless data compression**: a type of data compression in which you can construct the original data *perfectly* from the compressed version.

- We use data compression algorithms to reduce the amount of device storage needed to store a file ("file size").
- Why might we want to do this? The smaller the file size, the more files that can be stored on a device.

Text Compression

An Example of Lossless Data Compression

Lossless Text Compression: Byte Pair Encoding

Let's take this version, and **compress** it to use fewer characters (which means fewer bits = less space) by *encoding* the repeated sequences with other single characters:

```
to be or not to be, that is the question
```
original

```
# % or not # %, @at is @e question
```
compressed = fewer bits!

We would include a "dictionary" in the compressed version to indicate which each symbol in the compressed version stands for:

```
# to
% be
@ is
```

This simple text compression technique is known as **byte pair encoding**

Lossless Text Compression: Run Length Encoding (RLE)

Another text compression scheme is **run length encoding**

This algorithm compresses text by replacing *sequences of consecutively repeated characters* with a single character plus a number; the number represents how many of those characters are in the *original* message

```
AACCCCBBBBBAAAAAAAXFFFFFFFF
```
original message
(27 characters = 216 bits)

```
A2C4B5A7X1F8
```
compressed message using RLE
(12 characters = 96 bits)

- **What type of text compresses *really well*?**

  Text with lots of redundancy in the data (*many* incidents of repeated sequences of two or more characters)

- **Not well (or at all)?**

  Text with little or no redundancy in the data

Text Compression is Lossless Data Compression

- **Lossless data compression** is a type of compression in which you can construct the *original data perfectly* from the compressed version.
  - With a dictionary and compressed message, you can get the original text back **exactly** -- there is no "loss of information"

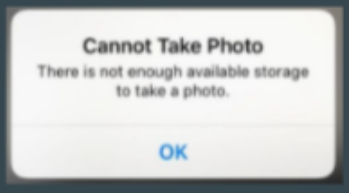
- Lossless data compression algorithms reduce the number of bits stored or transmitted while *guaranteeing complete reconstruction of the original data.*

- **Lossy data compression:** a type of data compression that can significantly reduce the number of bits stored or transmitted but only allow reconstruction of an *approximation* of the original data -- *the original data **cannot** be reconstructed exactly.*

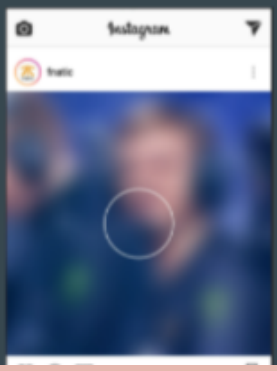
Image Compression

An Example of Lossy Data Compression

- **Why might we want to compress image files (i.e. reduce the file size of image files)?**
- **Compressed image files → more photos can fit on your device!**



- **Compressed image files → websites/apps with lots of images to download (think: your Instagram) load *faster* because smaller image files mean less data to transmit over the internet via your wireless connection**



- Unfortunately, with images, when you compress an image, you will lose a little data (we will talk about why this is shortly)
- Image compression is an example of lossy compression.

How image compression works -- and why we lose data

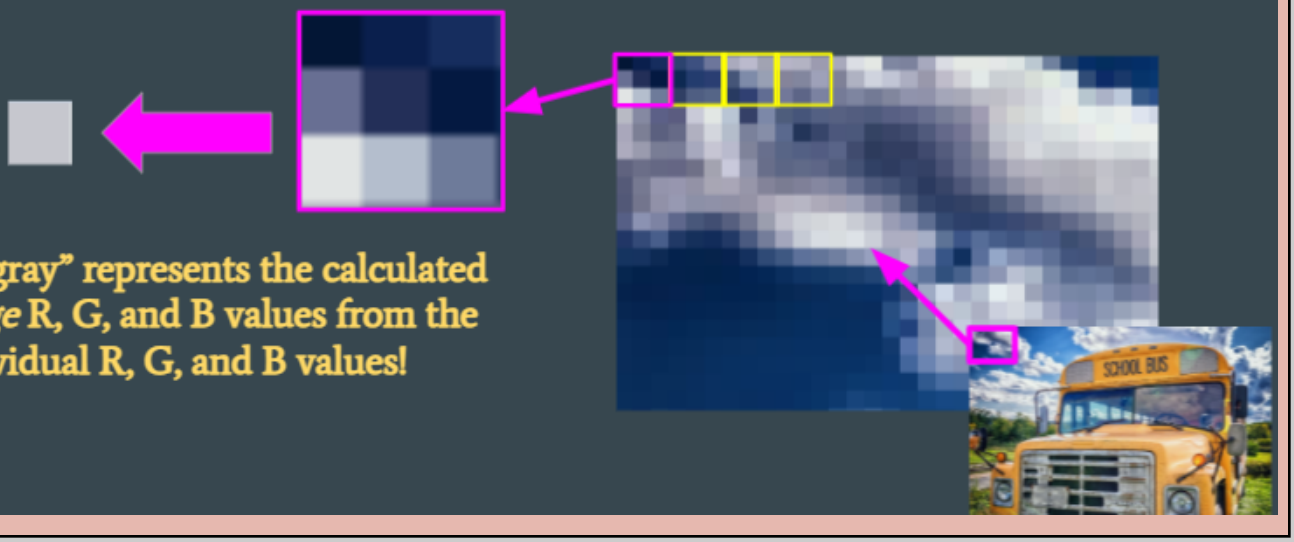
- **Each pixel** has an RGB value (which is the data stored as binary)

RGB: (201, 204, 212)

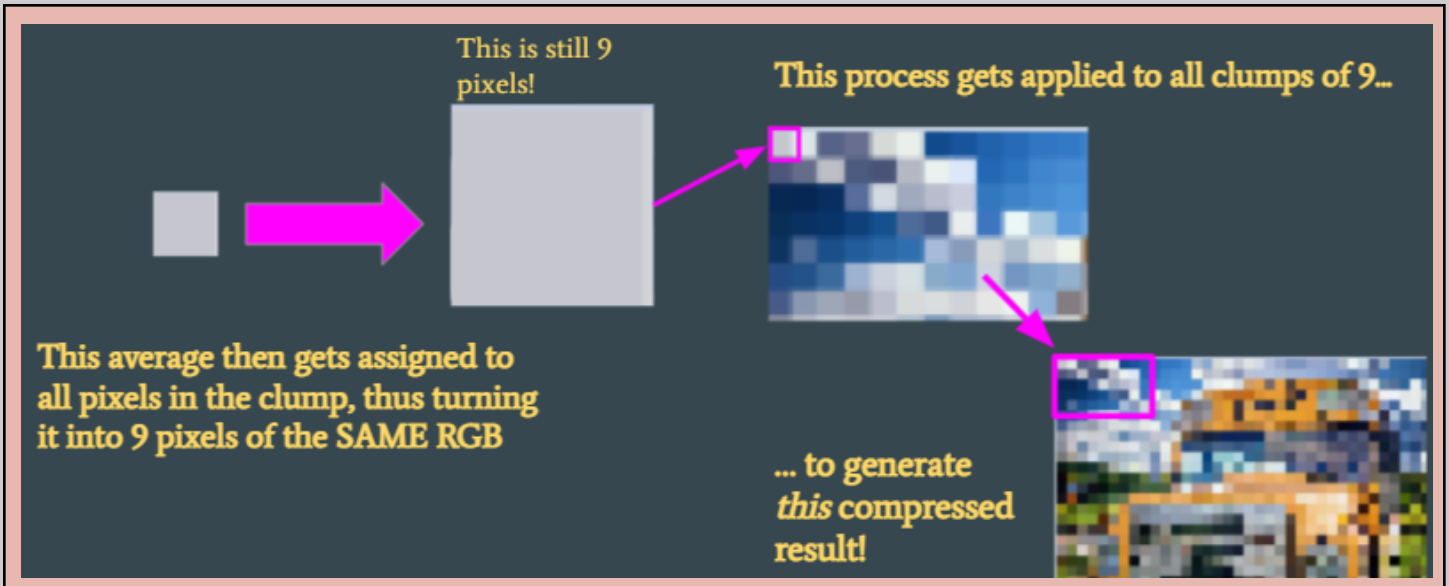
This pixel's RGB in **binary**: (11001001, 11001100, 11010100)

Or simply: 110010011100110011010100

- When an image is **compressed**, the RGB values from the pixels in a group of pixels are **AVERAGED (or "blended")** together, and the average RGB value is assigned to *all pixels in the group!*



This "gray" represents the calculated *average* R, G, and B values from the 9 individual R, G, and B values!



- Instead of including 9 copies of an RGB value in the binary data:

These 9 pixels have identical RGB (due to compression "averaging")

11001001	11001001	11001001
11001100	11001100	11001100
11010100	11010100	11010100
11001001	11001001	11001001
11001100	11001100	11001100
11010100	11010100	11010100
11001001	11001001	11001001
11001100	11001100	11001100
11010100	11010100	11010100

We just need one! Along with a "dictionary" which is how many there are of that color! Cool!

Binary for 9

1001

11001001
11001100
11010100

So why is it *lossy* compression?

- Image compression is an example of **lossy compression** because you can't recreate the original image from a compressed image

If I emailed you this compressed image file:

You could NOT recreate the original image:

- **The reason?** A compressed JPEG image is created using **average** RGB values -- but, given an average, you can't determine which exact numbers were used to create that average! For example, if I said "My test average is 85! What were my 9 test scores?" You couldn't mathematically figure out the 9 test scores that generated that average -- there are too many possibilities.

- Lossy data compression algorithms can significantly reduce the number of bits stored or transmitted *but only allow reconstruction of an* ***approximation*** *of the original data* -- the original data cannot be reconstructed **exactly**.
- JPEG image compression is a great example of lossy image compression -- you can only go one way, some data gets lost, and the original cannot be reproduced exactly from the compressed version

Audio Compression

- The **MP3 audio** format, which you almost certainly use for portable music files, is another example of a **lossy compression format** (a 4 minute song in MP3 format is about 3-4 Mb in file size)
- *Uncompressed* versions of the same audio recording (which exist when they are first recorded at the studio, "CD quality") are *much* bigger in file size (think: 10-15 Mb **PER MINUTE** of song, or about 60 MB for a 4-minute song! Imagine downloading a few of those from Spotify! You would use up your data plan very fast!)
    - You may have heard of the **FLAC** audio format -- look it up if you are interested in learning about a *lossless* audio format (but it comes at an expense: it requires about *6 times* more space per song)

AP Exam Practice Question

Which of the following is an advantage of a lossless compression algorithm over a lossy compression algorithm?

A. A lossless compression algorithm can guarantee that compressed information is kept secure, while a lossy compression algorithm cannot.

B. A lossless compression algorithm can guarantee reconstruction of original data, while a lossy compression algorithm cannot.

C. A lossless compression algorithm typically allows for faster transmission speeds than does a lossy compression algorithm.

D. A lossless compression algorithm typically provides a greater reduction in the number of bits stored or transmitted than does a lossy compression algorithm.

My answer:

Check!

AP Exam Practice Question

A user wants to save a data file on an online storage site. The user wants to reduce the size of the file, if possible, and wants to be able to completely restore the file to its original version. Which of the following actions best supports the user's needs?

A. Compressing the file using a lossless compression algorithm before uploading it

B. Compressing the file using a lossy compression algorithm before uploading it

C. Compressing the file using both lossy and lossless compression algorithms before uploading it

D. Uploading the original file without using any compression algorithm

My answer:

Check!

AP Exam Practice Question

Which of the following is a true statement about data compression?

A. Data compression is only useful for files being transmitted over the Internet.

B. Regardless of the compression technique used, once a data file is compressed, it cannot be restored to its original state.

C. Sending a compressed version of a file ensures that the contents of the file cannot be intercepted by an unauthorized user.

D. There are trade-offs involved in choosing a compression technique for storing and transmitting data.

My answer:

Check!

AP Exam Practice Question

A person wants to transmit an audio file from a device to a second device. Which of the following scenarios best demonstrates the use of lossless compression of the original file?

A. A device compresses the audio file before transmitting it to a second device. The second device restores the compressed file to its original version before playing it.

B. A device compresses the audio file by removing details that are not easily perceived by the human ear. The compressed file is transmitted to a second device, which plays it.

C. A device transmits the original audio file to a second device. The second device removes metadata from the file before playing it.

D. A device transmits the original audio file to a second device. The second device plays the transmitted file as is.

My answer:

Check!

AP Exam Practice Question

Digital images are often represented by the red, green, and blue values (an RGB triplet) of each individual pixel in the image. A photographer is manipulating a digital image and overwriting the original image. Which of the following describes a lossless transformation of the digital image?

A. Compressing the image in a way that may lose information but will suffer only a small loss of image quality.

B. Creating the gray scale of an image by averaging the amounts of red, green, and blue in each pixel and assigning this new value to the corresponding pixel in the new image. The new value of each pixel represents a shade of gray, ranging from white to black.

C. Creating the negative of an image by creating a new RGB triplet for each pixel in which each value is calculated by subtracting the original value from 255. The negative of an image is reversed from the original; light areas appear dark, and colors are reversed.

D. Modifying part of the image by taking the pixels in one part of the picture and copying them to the pixels in another part of the picture.

My answer:	

Check!

AP Exam Practice Question

Byte pair encoding is a text encoding technique that looks for groups of characters that appear in the string more than once and replaces each instance with a corresponding character that does not appear in the string. These replacements are stored as a "dictionary" that gets appended to the compressed text as metadata.

For example, the string WHO_DID_IT_AND_WHY_DID_THEY_DO_IT can be can be encoded as a shorter string $O_&_%_AND_$Y_&_THEY_DO_% by replacing all instances of "WH" with "$", replacing all instances of "DID" with "&", and all instances of "IT" with "%"

For which of the following strings is it NOT possible to use byte pair encoding to shorten its length?

A. TO_BE_OR_NOT_TO_BE

B. ABRACADABRA

C. GIVE_ME_A_BREAK

D. ITSY_BITSY_SPIDER

My answer:	

Check!

AP Exam Practice Question

A programmer is developing software for a social media platform. The programmer is planning to use compression when users send attachments to other users. Which of the following is a true statement about the use of compression?

A. Lossless compression of video files will generally save more space than lossy compression of video files.

B. Lossless compression of an image file will generally result in a file that is equal in size to the original file.

C. Lossy compression of an image file generally provides a greater reduction in transmission time than lossless compression does.

D. Sound clips compressed with lossy compression for storage on the platform can be restored to their original quality when they are played.

My answer:

Check!

Linear & Binary Search

AP EXAM VOCABULARY

- Algorithms that are written to locate a specific value in a list are called **search algorithms**.
- A **linear search** algorithm starts at the first element in a list of values and searches through all the items one by one until it finds the desired value or reaches the end of the list with the value not found.
- A **binary search** algorithm starts in the *middle* of a *sorted* list and *repeatedly eliminates half* the list until either the desired value is found or all elements have been eliminated.
    - It's called a "*binary*" search because each check cuts the remaining list into *two* halves (and eliminates one of the halves each time).
- The **length** of the list being searched refers to *how many elements are in the list* and is often referred to as "**input size**" (or "**N**")
- Each number in the list that is "checked" to see if it's the number you are looking for is a **comparison**, equivalent to a processing "step" in a computer program.
    - The *more* comparisons that need to be made, the longer (in seconds) the search will take to find the value in the list (or to determine the value isn't in the list)

For example, given a list with 15 numbers in it (i.e. its **length = 15**), the **minimum** number of comparisons that a **linear search** might take to find a particular value is **1**, in the case the number you are looking for *happens to be the first element in the list*. The **maximum** number of comparisons is **15**, in the case the number you are looking for isn't in the list *or* is located in the list at the last index (you would have to check the 15th number regardless).

Linear search can be used to search lists that are SORTED or UNSORTED.

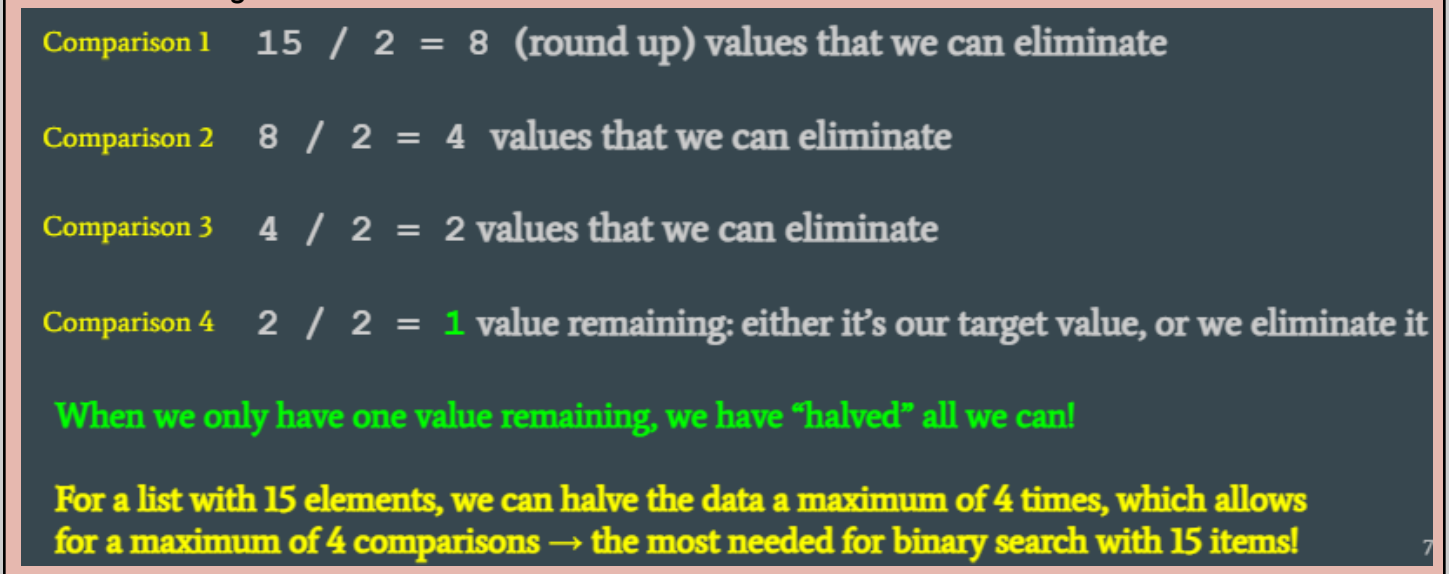
The **binary search** algorithm uses a strategy modeled after a number guessing game strategy in which you start at the middle value of a **sorted** data set of numbers and eliminate half of the data, then choose the middle value of the remaining upper or lower half. This process repeats until the desired value is found or all elements have been eliminated

Binary search can ONLY be used to search SORTED lists

For example, given a **sorted** list with 15 numbers in it, the **minimum** number of comparisons that a **binary search** might take to find a particular value is **1**, in the case the number you are looking for *happens to be the middle element in the list*. The **maximum** number of comparisons is **only 4**.

Because binary search typically takes *fewer* comparisons (= less time) on average than linear search, binary search is considered more efficient than linear search.

The reason a list of 15 has a maximum of 4 comparisons for binary search is illustrated below; it's all about *halving* the dataset:



If we instead had a **sorted** list of 50 elements, we could use linear search or binary search on the list (linear can search sorted or unsorted lists).

If we use ***linear*** search, the *maximum* comparisons it might take is **50** (in the case the value we are looking for is the last value in the list, or not in the list at all).

If we use ***binary*** search, the *maximum* comparisons it might take is **6.**

Here is where the "6" comes from:

```
Comparison 1   50 / 2 = 25 values that we can eliminate
Comparison 2   25 / 2 = 13 (round up) values that we can eliminate
Comparison 3   13 / 2 = 7 (round up) values that we can eliminate
Comparison 4   7 / 2 = 4 (round up) values that we can eliminate
Comparison 5   4 / 2 = 2 values that we can eliminate
Comparison 6   2 / 2 = 1 value remaining: either it's our target value, or we eliminate it
```

NOTES ON "SORTED" LISTS:

- A **sorted** list is any list that has values arranged in **ascending** or **descending** order.
- Sorted lists can contain **numbers**, in which case "sorted" means values arranged low to high (or high to low), or **text**, in which case "sorted" means arranged *alphabetically*.
- Sorted lists can contain duplicate values.

Examples of **sorted** lists that binary search ***can*** be used on:

- `[1, 4, 6, 8, 10, 15]` (sorted in *ascending* order)
- `[-3, 0, 1, 1, 2, 3, 3, 8]` (sorted in ascending order, with *duplicates*)
- `[12, 5, 2, -1, -6, -18]` (sorted in *descending* order)
- `["apple", "car", "ghost", "kite"]` (sorted in *alphabetical* order)

AP Exam Practice Question

A sorted list of numbers contains 200 elements. Which of the following is closest to the maximum number of list elements that will need to be examined when performing a binary search for a particular value in the list?

A) 5

B) 8

C) 100

D) 200

My answer:

Check!

AP Exam Practice Question

A sorted list of numbers contains 500 elements. Which of the following is closest to the maximum number of list elements that will be examined when performing a binary search for a value in the list?

(A) 10

(B) 50

(C) 250

(D) 500

My answer:

Check!

AP Exam Practice Question

The list `listOne` is a sorted list of numbers that contains 700 elements. The list `listTwo` is a sorted list of numbers that contains 900 elements. Let *x* represent the maximum number of list elements that will need to be examined when performing a binary search for a value in `listOne`, and let *y* represent the maximum number of list elements that will need to be examined when performing a binary search for a value in `listTwo`. Which of the following statements about *x* and *y* is true?

(A) The value of *x* is approximately equal to the value of *y*.

(B) The value of *x* is approximately 10 less than the value of *y*.

(C) The value of *x* is approximately 13 less than the value of *y*.

(D) The value of *x* is approximately 200 less than the value of *y*.

My answer:

Check!

AP Exam Practice Question

For which of the following lists can a binary search be used to search for an item in the list?

I. `["blue", "green", "jade", "mauve", "pink"]`
II. `[5, 5, 5, 5, 6, 7, 8, 8, 8]`
III. `[10, 5, 3, 2, -4, -8, -9, -12]`

A) I only

B) III only

C) I and III only

D) I, II, and III

My answer:

Check!

AP Exam Practice Question

A large number of genetic codes are stored as binary values in a list. Which one of the following conditions must be true in order for a researcher to obtain the correct result when using a binary search algorithm to determine if a given genetic code is in the list?

A) The genetic codes must be converted from binary to decimal numbers.

B) The list must be sorted based on the genetic code values.

C) The number of genetic code values in the list must be a power of 2.

D) The number of genetic code values in the list must be even.

My answer:

Check!

AP Exam Practice Question

Suppose that a list of numbers contains values `[-4, -1, 1, 5, 2, 10, 10, 15, 30]`. Which of the following best explains why a binary search should NOT be used to search for an item in this list?

A. The list contains both positive and negative elements.

B. The elements of the list are not sorted.

C. The list contains an odd number of elements.

D. The list contains duplicate elements.

My answer:

Check

AP Exam Practice Question

A time stamp indicates the date and time that a measurement was taken. A data scientist has a list containing 10,000 time stamps, sorted in chronological order. Which of the following is closest to the maximum number of values that will need to be examined when performing a binary search for a value in the list?

A. 10

B. 15

C. 5,000

D. 10,000

My answer:

Check!

Practice AP Question

The procedure `BinarySearch (numList, target)` correctly implements a binary search algorithm on the list of numbers `numList`. If `target` is found in the list, the procedure returns `true`, otherwise, it reports `false`. Which of the following conditions must be true in order for the procedure to work properly?

(A) `target` must be a non-negative number
(B) `target` cannot be the first or last element in `numList`
(C) There needs to be an odd number of elements in `numList`
(D) The numbers in `numList` must be sorted

My answer:	

Check!

Efficiency & Categorizing Algorithms

AP EXAM VOCABULARY

- **Linear time/efficiency:** An algorithm is said to run in **linear time (linear efficiency)** if the required number of steps (and thus time) is proportional to the input size, in other words, **number of executed steps (processing time) and input size grows linearly, or at the same rate**.

- **Quadratic time/efficiency:** An algorithm is said to run in **quadratic time (quadratic efficiency)** if the required processing steps/time is proportional to the **square** of the input size (n^2)

- **Cubic time/efficiency:** An algorithm is said to run in **cubic time (cubic efficiency)** if the required processing steps/time is proportional to the **cube** of the input size (n^3)

- **Sublinear time/efficiency:** An algorithm that has a **faster than linear efficiency** is said to require **sublinear time (sublinear efficiency)** to run. In an algorithm with sublinear efficiency, the **number of steps (and time) grows more slowly than the size**. Logarithmic relationships are examples of sublinear time/efficiency.

- **Constant time/efficiency:** An algorithm is said to run in **constant time (constant efficiency)** if it takes about the **same number of processing steps (and same amount of time) regardless of input size.** In other words, bigger and bigger lists do **not** require more processing steps or time.

- **Polynomial time/efficiency:** An algorithm runs in **polynomial time (polynomial efficiency)** if the amount of time is less than or equal to *a **power** of the size of the input*.

  The following times are all examples **polynomial times/efficiencies**:
  - **Constant** time/efficiency (n^0)
  - **Sublinear** time/efficiency (log n)

    - **Linear** time/efficiency (n^1)
    - **Quadratic** time/efficiency (n^2)
    - **Cubic** time/efficiency (n^3)

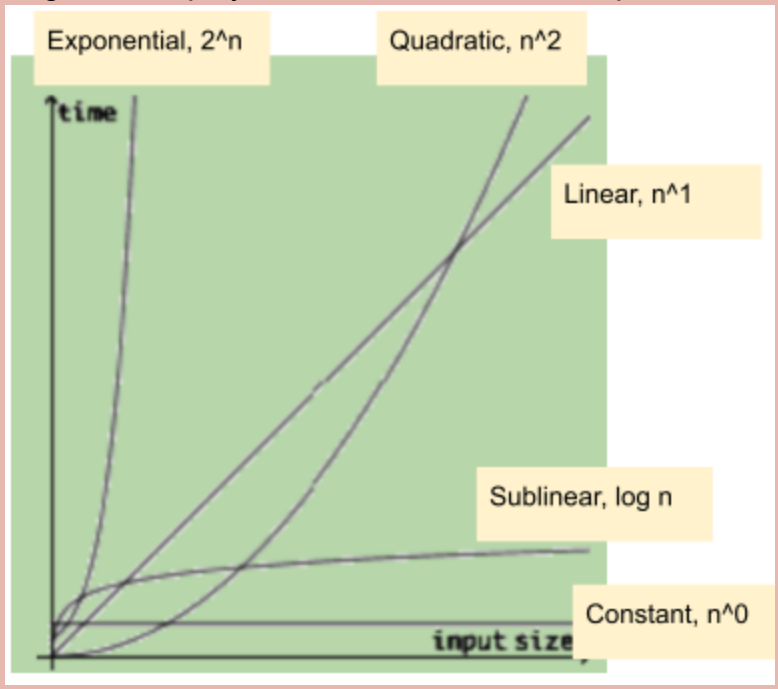
- **Reasonable time:** The phrase "***runs in reasonable time***" describes ***any algorithm that has a polynomial efficiency***. All algorithms with polynomial efficiency are considered "reasonable" (even if they take a seemingly long time!), so **constant, sublinear, linear, quadratic, and cubic** are all **reasonable time** algorithms.

- **Exponential time/efficiency:** An algorithm runs in **exponential time (exponential efficiency)** if the required number of steps grows proportionally to an **exponential function** like 2^n, 10^n, etc.
    - **Note** the difference between *exponential* efficiency, 2^n, and *quadratic* efficiency, n^2

- **Factorial efficiency:** An algorithm runs in **factorial time (factorial efficiency)** if the required number of steps grows proportionally to the factorial of the input size, **n!**

- **Exponential** and **factorial** times/efficiencies are *much slower and less efficient than any polynomial efficiency.* Algorithms with **exponential** or **factorial time/efficiency** are considered "**unreasonable time**" and do **NOT** *run in reasonable time*.

Here is a graph comparing the four polynomial efficiencies with exponential efficiency:



Notice for **exponential** *how quickly time increases for small increases in input size* -- this shows how highly **INEFFICIENT** exponential efficiency is!

Some different `mystery` algorithms have the following run times for inputs of a certain size. Based on the data in the tables, **classify each `mystery` algorithm** as either:

- Polynomial efficiency **or** exponential efficiency
  - *If polynomial,* further classify as linear, sublinear, constant, or quadratic
- Runs in reasonable time **or** does not run in reasonable time

	`mystery1`		`mystery2`		`mystery3`		`mystery4`	
	Input size (n)	Time (ms)	Input size (n)	Time (ms)	Input size (n)	Time (ms)	Input size (n)	Time (ms)
	25	400	50	50	10	50	50	100
	50	800	100	200	11	100	500	200
	100	1600	150	450	12	200	5000	400
	200	3200	200	800	13	400	50000	800
Polynomial or **exponential** efficiency?								
If polynomial efficiency, is it linear, sublinear, constant, or quadratic?								
Does it “run in **reasonable** time”?								

[Check my answers]

Practice AP Exam Question

The table below shows the time a computer system takes to complete a specified task on the customer data of different-sized companies.

Task	Small Company (approximately 100 customers)	Medium Company (approximately 1,000 customers)	Large Company (approximately 10,000 customers)
Backing up data	2 hours	20 hours	200 hours
Deleting entries from data	100 hours	200 hours	300 hours
Searching through data	250 hours	300 hours	350 hours
Sorting data	0.01 hour	1 hour	100 hours

Based on the information in the table, which of the following tasks is likely to take the longest amount of time when scaled up for a very large company of approximately 100,000 customers?

(A) Backing up data

(B) Deleting entries from data

(C) Searching through data

(D) Sorting data

Your answer (A, B, C, D):

[ ]

Check answer!

AP Exam Practice Question

Lina wrote three different algorithms to search a list containing n elements. Assume n is a very large integer.

I. Her first algorithm checks each element in the list five times.
II. Her second algorithm checks each element in the list n times.
III. Her third algorithm checks the first 20 elements only, regardless of the list length.

Which algorithms run in **reasonable** time?

(A) I only
(B) III only
(C) I and III only
(D) I, II, and III

Your answer (A, B, C, D):

[ ]

Check answer!

AP EXAM VOCABULARY

It's important to recognize that **exponential time** or **factorial time algorithms** *still* solve a problem -- *it just takes an **unreasonably** long time to get the answer* (perhaps hours, days, weeks, even years if you would be willing to wait that long!).

Unreasonable-time algorithms (like exponential-time and factorial-time algorithms) can sometimes be replaced by **heuristics.**

- **Heuristics** are polynomial-time algorithms that don't solve the problem exactly, *but give a good enough approximation*. **They are used to *approximate* solutions in *reasonable time* to problems that *have* exact solutions, but require an *unreasonable* time to obtain the exact solution.**

Heuristics are ideal for problems that ***have*** exact solutions, but for which:

- The exact solution *cannot* be determined in reasonable time (i.e. it would take way too long for the algorithm too finish)
- An approximation to the exact solution would be "good enough" for the purposes of the problem (i.e. having the *exact* solution isn't necessary)

An **optimization problem** is a problem with the goal of finding the *best solution among many*; for example, shortest-path problems, like used in Google Maps.

AP Exam Practice Question

Which of the following programs is most likely to benefit from the use of a heuristic?

(A) A program that determines if a number is even or odd
(B) A program that counts the number of words in a list that are 3 or more letters long
(C) A program that calculates the fastest route to get from school to home when accounting for traffic
(D) A program that searches a sorted list using the binary search algorithm

Your answer (A, B, C, D):

[ ]

Check my answer!

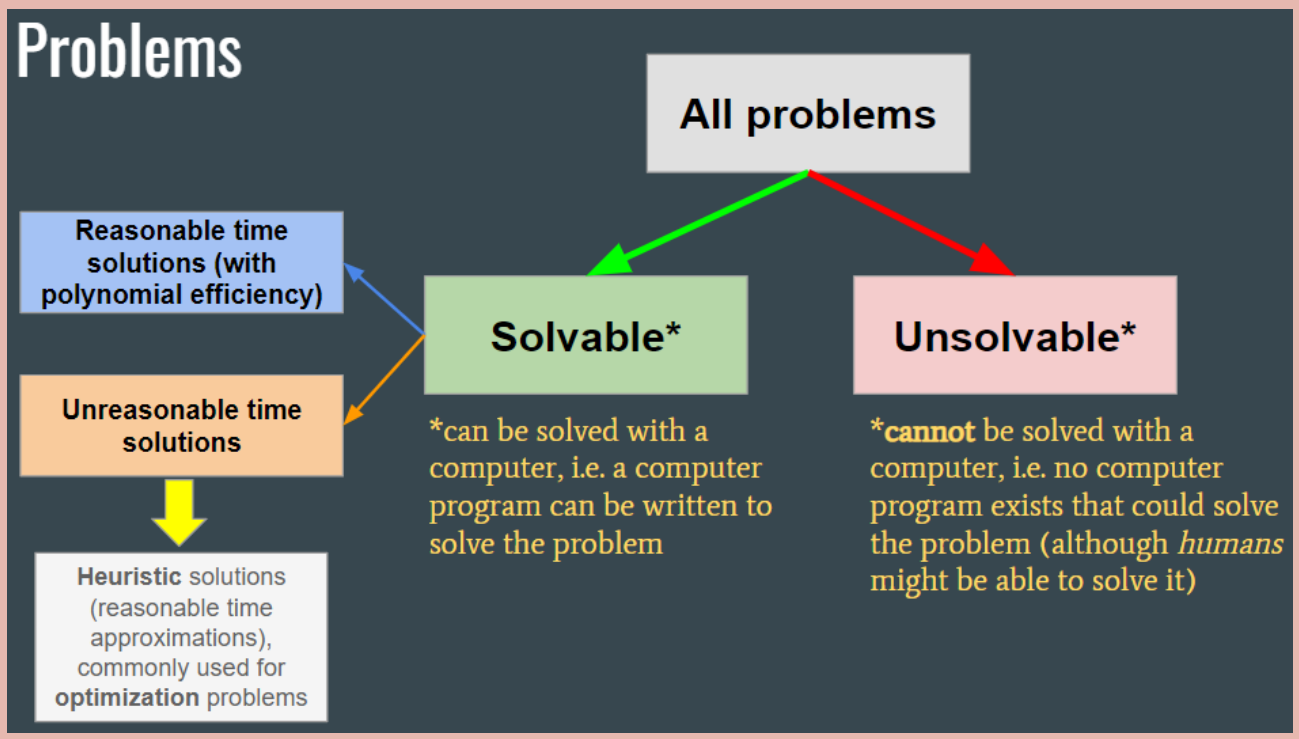
Undecidable Problems & Parallelism

AP EXAM VOCABULARY

- In computer science, a "**problem**" is a task that ***may or may not* be able to be solved using an algorithm** (i.e. by writing a computer program to solve it).
- A problem is considered **solvable** if it **has a solution that *can* be found using an algorithm** (i.e. it *can* be solved using a computer program)
  - Some **solvable** problems use algorithms that take a **reasonable time** to solve (such algorithms have **polynomial efficiencies**, as you learned about in the last lab)
  - Other **solvable** problems take an ***unreasonable*** time to solve -- and a computer could get to the solution *if* it ran long enough -- although this is not practical.
    - **Heuristics** are *reasonable-time* algorithms that are used to *approximate* solutions to problems that would otherwise take an *unreasonable* time.
    - **Optimization** problems are examples of *solvable* problems that take an *unreasonable* time and for which **heuristics** are useful.
- A problem is considered **unsolvable** if **no algorithm can be written to find the solution (provided a solution even exists)**; in other words, it can't be solved by programming a computer to solve it -- although a *human* might be able to solve it!
  - This doesn't necessarily mean the problem doesn't have a solution; it may have a solution, but it is *not possible to write a computer program to find that solution*.

Not all problems can be solved by computers!

Here is a graphic organizing these terms:



EXAMPLE 1: Solvable problem, reasonable time

Here is an example of a **solvable** problem that can be solved with an algorithm in **reasonable** time:

"What is the *maximum* value in this list of numbers: _____?" *(with a list of numbers provided)*

This problem is **solvable** in reasonable time because an *algorithm can be written to find the solution in reasonable time*.

EXAMPLE 2: Solvable problem, *not* reasonable time

Here is an example of a **solvable** problem that *can be solved with an algorithm*, but ***not*** in reasonable time, **and** for which a **heuristic** could be used to approximate:

The Traveling Salesman optimization problem:
"A salesman wants to visit 100 different cities across the country. What is the *shortest* route for the salesperson to visit *every* city and return home?"

This is **solvable** because it *does* have an *exact* solution and an algorithm can be written to find that solution, *but it would take an **unreasonable** amount of time to find the exact answer.* This is because there are 100! (factorial, 100 x 99 x 98 x 97...) possible routes!

- A problem is considered **unsolvable** if **no algorithm can be written to find the solution (provided a solution even exists)**; in other words, it can't be solved by programming a computer to solve it!
    - This doesn't necessarily mean the problem doesn't have a solution; it may have a solution, but it is *not possible to write an algorithm to find that solution*.

EXAMPLE 3: Not solvable

Here is an example of an **unsolvable** problem, one for which *no algorithm could be constructed to find the solution*:

The "debugging programming" problem:
"A student wants to want to write a program that checks itself to determine if there are any bugs in the program"

It turns out that this problem has been proven to be **unsolvable** -- in other words, *it isn't possible to write an algorithm that automatically determines if a program contains bugs or not!*

On the AP Exam, you won't need to determine if a particular problem is solvable or not, you just need to know this: not all problems can be solved with a computer using an algorithm.

- A special type of problem is called a **decision problem**; this is a problem with a **yes or no (or true/false) answer**; for example:
    - Is 496,221 divisible by 3?
    - Does the list [1, 4, 6, 9, 10, 15] contain the value 8?
    - Is there a path from point A to point B?
    - Is the sentence *"there are 7 days in a week"* true?

- The examples above are specific **instances** of the following ***generalized*** decision problems:
  - Is ___ divisible by ___?
  - Does the list _____ contain the value ___?
  - Is there a path from point ___ to point ___?
  - Is the sentence _____ true?

- A **decision problem** is considered **decidable** if it is possible to construct an algorithm that will give a **correct yes/no** (or true/false) result for **every** possible input value(s).

EXAMPLE 4: Decidable problem

The decision problem "**Is ___ divisible by ___?**" ***is*** **decidable** because we could write an algorithm like the one shown below (shown with AP Exam pseudocode) which ***would*** correctly return true or false (equivalent to yes or no) for **any** input values:

```
PROCEDURE divisibleBy numToCheck, divisor
    IF (numToCheck MOD divisor = 0)
        RETURN true
    ELSE
        RETURN false
```

We could then use this algorithm to get a correct answer for **any** inputs:

```
divisibleBy(15, 3) → true
divisibleBy(20, 3) → false
divisibleBy(167, 20) → false
divisibleBy(150, 75) → true
```

- A **decision problem** is considered **undecidable** if it is ***not*** **possible** to construct an algorithm that will always give a correct true/false (yes/no) output for **every** input value.

EXAMPLE 5: Undecidable problem

Here is a decision problem that is actually an **undecidable** problem -- one for which it is **not possible** to answer the question correctly (with T/F or Y/N) for ***all*** possible inputs:

"Is the sentence _____ true?"

This is an **undecidable** problem because you could answer yes or no for **some** inputs, *but* ***not every*** *input!*
For example:

Is the sentence "*6 is equal to 7*" true?
answer: NO, this sentence is not true

Is the sentence "*the sky is blue*" true?
answer: YES, this sentence is true

Is the sentence "*this statement is false*" true?"
answer: ?!?!? 😳😵😱

Cannot correctly answer for this particular input! Logical contradiction!

*All it takes is **ONE** instance of the problem **NOT** working to be undecidable!*

AP Exam Practice Question

Which of the following statements is true?

(A) Every problem is solvable with an algorithm in reasonable time.

(B) Every problem is solvable with an algorithm, but some require an unreasonable amount of time.

(C) Every solvable problem is also a decidable problem.

(D) Some problems can never be solved by a computer using an algorithm.

Your answer:	

Check answer!

AP Exam Practice Question

How can it be shown that a decision problem is **undecidable**?

(A) Show that for one instance of the problem, an algorithm can be written that is always capable of providing a correct yes-or-no answer.

(B) Show that for one instance of the problem, no algorithm can be written that is capable of providing a correct yes-or-no answer.

(C) Show that for one instance of the problem, a heuristic is needed to write an algorithm that is capable of providing a correct yes-or-no answer.

(D) Show that for one instance of the problem, an algorithm that runs in unreasonable time can be written that is capable of providing a correct yes-or-no answer.

Your answer:	

Check answer!

AP Exam Practice Question

A team of programmers is designing software. One portion of the project presents a problem for which there is not an obvious solution. After some research, the team determines that the problem is undecidable. Which of the following best explains the consequence of the problem being undecidable?

A. The problem can be solved algorithmically, but it will require an unreasonably long amount of time.

B. The problem can be solved algorithmically, but it will require an unreasonably large amount of data storage.

C. There is no possible algorithm that can be used to solve all instances of the problem.

D. There are several different possible algorithms that can solve the problem, but there is controversy about which is the most efficient.

My answer:

Check!

AP Exam Practice Question

A student wants to determine whether a certain problem is undecidable. Which of the following will demonstrate that the problem is undecidable?

A. Show that for one instance of the problem, an algorithm can be written that is always capable of providing a correct yes-or-no answer.

B. Show that for one instance of the problem, no algorithm can be written that is capable of providing a correct yes-or-no answer.

C. Show that for one instance of the problem, a heuristic is needed to write an algorithm that is capable of providing a correct yes-or-no answer.

D. Show that for one instance of the problem, an algorithm that runs in unreasonable time can be written that is capable of providing a correct yes-or-no answer.

My answer:

Check!

AP Exam Practice Question

Which of the following best explains how algorithms that run on a computer can be used to solve problems?

A. All problems can be solved with an algorithm that runs in a reasonable amount of time.

B. All problems can be solved with an algorithm, but some algorithms might need a heuristic to run in a reasonable amount of time.

C. All problems can be solved with an algorithm, but some algorithms might run in an unreasonable amount of time.

D. Some problems cannot be solved by an algorithm.

My answer:

Check!

AP Exam Practice Question

Which of the following statements is true?

A. Every problem can be solved with an algorithm for all possible inputs, in a reasonable amount of time, using a modern computer.

B. Every problem can be solved with an algorithm for all possible inputs, but some will take more than 100 years, even with the fastest possible computer.

C. Every problem can be solved with an algorithm for all possible inputs, but some of these algorithms have not been discovered yet.

D. There exist problems that no algorithm will ever be able to solve for all possible inputs.

My answer:

Check!

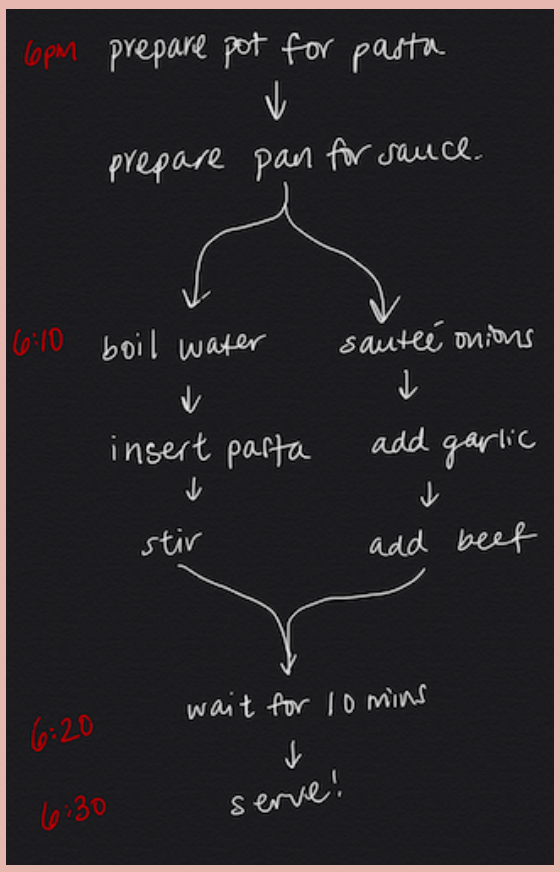
Parallelism & Distributed Computing

AP EXAM VOCABULARY

- **Sequential computing** is a computational model in which operations are performed in order one at a time.

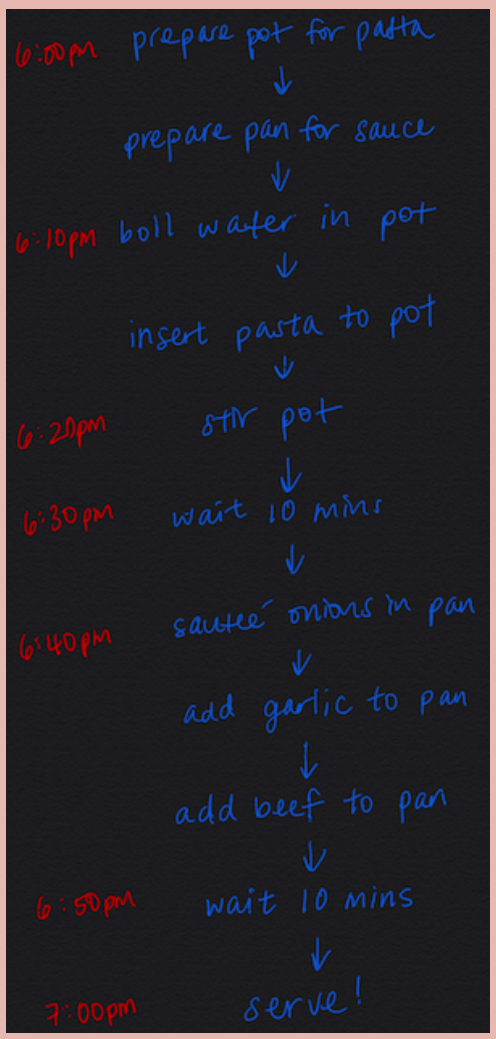
- In the **parallel computing** model, a program is broken into smaller steps, some of which are **performed at the same time (simultaneously)**. Modern computers have *multiple processors* (2, 4, or 8) in a single computer, and programs can be written to have *multiple* tasks executing *simultaneously,* with each processor working on one of the tasks.

Cooking spaghetti with *some* parallel steps:



Doing some tasks in *parallel* improves efficiency by speeding up the overall process!

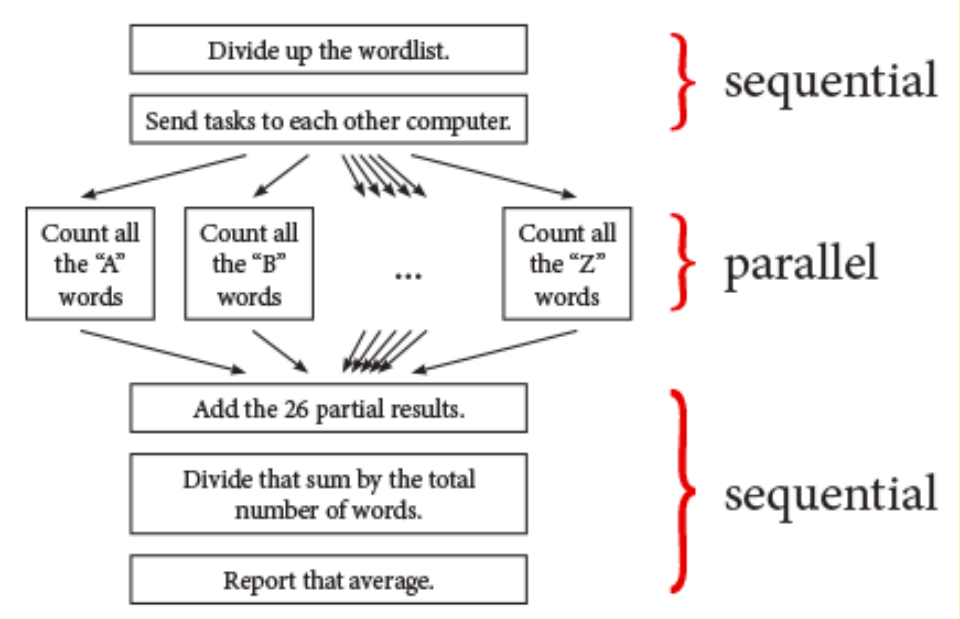
Cooking spaghetti with *only* sequential steps:



In solving most problems, there are often solutions that involve breaking up certain steps and doing them in **parallel**. However, typically not *all* steps can be done in parallel and some computation *does* need to be accomplished **sequentially** (similar to the cooking spaghetti example above).

EXAMPLE

Suppose you want to know the average word length in a list of 100,000 words. You can divide the task among several computers (one for each starting letter).



The **main** computer breaks up the word list by letter into 26 smaller lists and sends them to *other* computers. This is done **sequentially**.

Then, the main computer **distributes** a sublist to each of the **26 computers running in parallel** (one gets all the "A" words, one gets all the "B" words, etc), and each computer adds the lengths of all the words in the sublist assigned to it and sends the results back to the main computer.

Then, the **main** computer adds the 26 partial results and divides by the total number of words to find the average.

To calculate the **TOTAL RUN TIME** this parallel solution takes to run, you would add the run time of the longest parallel portion (the run time for the letter with the most words) to the run times of the sequential portions (adding the 26 partial results and dividing the sum by the total number of words).

This process is called **distributed computing** and lets you ***scale* to very large problems** very efficiently! This is how Google and big tech companies perform their services. At Google, millions of search queries happen every day, and they have huge **server farms** with thousands of computers to process all these requests in **parallel.**

- **Distributed computing** is a computational model in which *multiple devices* are used to run a program.

Suppose the top sequential part (on the main computer) takes **20** seconds, *each* computer working in parallel on an individual letter finishes processing its sublist in **10** seconds, and the bottom sequential part (back on the main computer) takes **15** seconds.

What would be the total **time** required for this solution to the problem of calculating the average word length of a list of 100,000 words?

Check!

AP Exam Practice Question

Which of the following best describes the ability of parallel computing solutions to improve efficiency?

A. Any problem that can be solved sequentially can be solved using a parallel solution in approximately half the time.

B. Any solution can be broken down into smaller and smaller parallel portions, making the improvement in efficiency theoretically limitless as long as there are enough processors available.

C. The efficiency of parallel computing solutions is rarely improved over the efficiency of sequential computing solutions.

D. The efficiency of a solution that can be broken down into parallel portions is still limited by a sequential portion.

My answer:

Check!

EXAMPLE
Maria owns a computer that has two processors that are able to run *in parallel*. The table below indicates the amount of time it would take either processor to execute each of four different processes, let's call them "Process A," Process B, "Process C," and "Process D."

Process	Execution Time
A	30 seconds
B	40 seconds
C	35 seconds
D	20 seconds

Maria wants to assign the four different processes to the two processors on her computer so they can be accomplished in **parallel** (at the same time).

For example, she might assign Process A to one processor, and the other 3 to the second processor, to run in **parallel** like this:

Processor 1's Tasks	Processor 2's Tasks
Process A Process B	Process C Process D

In this case, Processor 1 would take 30 + 40 = 70 seconds to complete both of its tasks, while Processor 2 would take 35 + 20 = 55 seconds to complete its tasks.

Since these are being done in *parallel*, the processors are completing their tasks *at the same time*, so if Maria needed *all four processes* to be completed, she would need to wait **70 seconds**, which is the time processor 1 takes to complete, since it's the longer of the two. Processor 2 would have already finished after the first 55 seconds, but Maria would need to wait an additional 15 seconds for Processor 1 to finish its tasks. During those last 15 seconds, while Processor 1 is finishing up, Processor 2 would be *idle* since it had no more tasks to complete.

She could assign the tasks differently, for example, like this:

Processor 1's Tasks	Processor 2's Tasks
Process B	Process A Process C Process D

How long would it take to complete all 4 tasks if Maria arranged them to run in parallel like this?	

She could even decide to *use only one processor* and run them entirely **sequentially**, like this:

Processor 1's Tasks	Processor 2's Tasks
Process A Process B Process C Process D	*none assigned* *processor idle*

How long would it take to complete all 4 tasks if Maria arranged them to run *sequentially* like this?	

How should Maria arrange the 4 tasks on two processors if she wants to *minimize* the total time she needs to wait for the 4 tasks to be complete? **Complete the chart →**	<table><tr><th>Processor 1's Tasks</th><th>Processor 2's Tasks</th></tr><tr><td></td><td></td></tr></table>
How long will it take to complete all 4 tasks if Maria arranges the tasks optimally like this?	
How long is one of the processors idle while the other finishes?	
What's the difference between the optimal (minimum) execution time when running the 4 tasks in parallel and the time it would take if Maria ran all 4 tasks sequentially on one processor?	

check all answers

AP Exam Practice Question

A certain computer has two identical processors that are able to run in parallel. The table below indicates the amount of time it takes each processor to execute each of two processes. Assume that neither process is dependent on the other.

Process	Execution Time on Either Processor
P	30 seconds
Q	45 seconds

Which of the following best approximates the difference in execution time between running the two processes in parallel instead of running them one after the other on a single processor?

A) 15 seconds

B) 30 seconds

C) 45 seconds

D) 75 seconds

My answer:

Check!

AP Exam Practice Question

A computer has two processors that are able to run in parallel. The table below indicates the amount of time it takes either processor to execute four different processes. Assume that none of the processes is dependent on any of the other processes.

Process	Execution Time
W	20 seconds
X	30 seconds
Y	45 seconds
Z	50 seconds

A program is used to assign processes to each of the processors. Which of the following describes how the program should assign the four processes to optimize execution time?

(A) Processes W and X should be assigned to one processor, and processes Y and Z should be assigned to the other processor.

(B) Processes W and Y should be assigned to one processor, and processes X and Z should be assigned to the other processor.

(C) Processes W and Z should be assigned to one processor, and processes X and Y should be assigned to the other processor.

(D) Process Z should be assigned to one processor, and processes W, X, and Y should be assigned to the other processor.

My answer:

Check!

AP Exam Practice Question

A certain computer has two identical processors that are able to run in parallel. Each processor can run only one process at a time, and each process must be executed on a single processor. The following table indicates the amount of time it takes to execute each of three processes on a single processor. Assume that none of the processes are dependent on any of the other processes.

Process	Execution Time on Either Processor
X	60 seconds
Y	30 seconds
Z	50 seconds

Which of the following best approximates the minimum possible time to execute all three processes when the two processors are run in parallel?

A) 60 seconds

B) 70 seconds

C) 80 seconds

D) 90 seconds

My answer:

Check!

Simulations

AP EXAM VOCABULARY

- **Simulations are computer *representations* of complex things, phenomena, events, or situations that vary over time.**
- Simulations are used to mimic real-world contexts, allowing you to study them to make predictions or draw inferences ***without*** the constraints of running "real world experiments."
- For example, when testing a new airplane design, you might *simulate* wind hitting it at high speeds *using a computer simulation of*

wind hitting the airplane, which could allow you to study the impact of high wind speed over time on airplanes *without the cost, time, or safety concerns of testing it using real airplanes!*

There are many **benefits** to using computer **simulations** *rather than real-world experiments* to prove things about a situation:

- **Simulations can save money**
- **Simulations can save time**
- **Simulations can provide a *safe* way to experiment**
- **Simulations can provide an *ethical* way to experiment**

Despite all the benefits, there are also **drawbacks** to using **simulations**:

- Computer simulations **rarely capture the full complexity of real situations**. For example, simulations that scientists use to predict the impact of global climate change have to account for hundreds of interconnected factors. However, *because of limitations in computer speed, it would be impossible to include all real interdependent factors in a simulation,* and so researchers leave out some of the details to make simulations practical.

- There are **risks in making such simplifications**. For example, there might be an interaction between elements of the real situation that isn't obvious but turns out to be crucial. A simplification might lose the details that would have shown that interaction or lead you to emphasize details that support your ideas (a distortion known as **confirmation bias**).

Simulations often have various input **parameters** that you can control the value of to see the impact. In a simulation, you can *add more parameters* (to make the simulation more complex or realistic) or *take away parameters* (to simplify the simulation) as needed.

AP Exam Practice Question

Supercomputers are machines that, as of 2019, could compute quadrillions of operations per second. These machines are used for a variety of things, including simulating earthquakes, nuclear reactions, and the big bang. Supercomputer simulations can replace the testing of actual nuclear arms. What are two advantages of such simulations?

Select two.

A. Simulations can incorporate all of the complexities of nuclear reactions.

B. Simulations are cheaper than building multiple designs and testing each other in the real world.

C. Simulations produce results that are more accurate than real world experiments.

D. Simulations come without the environmental associated risks with nuclear tests.

Your TWO answers:	and

Check your answer!

AP Exam Practice Question

Before airplane pilots are allowed to fly actual planes, they spend countless hours in a flight simulator. Which of the following is **not** a beneficial reason to use the computer simulation?

A. The simulation can create extreme scenarios to test a pilot's knowledge and ability to work under extreme pressure.

B. Pilots can use the simulation to determine if this is the career for them.

C. If, during the process, the pilot makes a mistake, it will not destroy an actual plane or kill innocent people.

D. Pilots can get familiar with the cockpit controls on a variety of different plans before flying the real plane.

Your answer:	

Check your answer!

AP Exam Practice Question

A plane manufacturer uses simulation software during the design process for a new plane. Which of the following are reasons to use simulation software for this purpose?

I. Using simulation software can save the company money by helping to compare designs early in the process.

II. Using simulation software can help to identify safety concerns with the plane by enabling the testing of the plane's components under various situations.

III. The manufacturer can present simulations to airlines to demonstrate different design options, and develop those options for airlines.

(A) I and II only

(B) I and III only

(C) II and III only

(D) I, II, and III

Your answer:	

Check your answer!

AP Exam Practice Question

Government officials and scientists are designing a simulation to examine how much a plan to convert entirely to electric cars will have on reducing climate change. The simulation uses a model that includes input variables for the plan such as expected growth in car demand, efficiency of electric cars, and consumer sentiment around electric cars. The simulation can then be run many times using different values for the input variables to represent different outcomes on climate change. One challenge they are facing is that the complex simulation takes a very long time to run. They plan to update the model by removing some input variables they consider less important than others. Of the following, which is the *most likely* affect the updated model will have on the simulation's runtime?

(A) The updated model is unlikely to affect the runtime of the simulation because the simulation will remain very complex, regardless of how many input variables the model uses.

(B) The updated model is likely to increase the runtime of the simulation because refining a simulation model requires increased computational time.

(C) The updated model is likely to decrease the runtime of the simulation because it would enable the simulation to be run on a faster computer.

(D) The updated model is likely to decrease the runtime of the simulation because the time required for simulations generally depends on the complexity of the model used.

Your answer:	

Check your answer!

AP Exam Practice Question

A researcher is interested in predicting the number of deaths that will occur in a particular community. She created a computer model that uses data from the past twenty years, including the number of residents and the number of deaths. The model predicted that there would be 500 deaths last year, but the actual number of deaths last year was only 456. Which of the following strategies is LEAST likely to provide a more accurate prediction?

(A) Gathering data for additional years to try to identify patterns in death rates

(B) Refining the computer simulation to reflect new understandings about causes of death

(C) Removing some of the details from the model and simulation so that computations can be performed more quickly

(D) Adding additional parameters to the simulation, such as the ages of residents.

Your answer:	

Check your answer!

AP Exam Practice Question

Meteorologists are tracking a hurricane approaching the East coast of the United States. Which of the following is a limitation of using computer simulations to model this real-world event?

(A) Computer simulations require large amounts of processing speed and memory.

(B) Computer simulations can only be designed and used after a real-world event, in this case a hurricane, has occurred.

(C) Computer simulations often approximate or leave out some details about the real-world event being simulated that may end up being important (such as rainfall, wind, or currents).

(D) Computer simulations require that all input parameters are never changed.

Your answer:	

Check your answer!

AP Exam Practice Question

Which of the following statements describes a limitation of using a computer simulation to model a real-world object or system?

A. Computer simulations can only be built after the real-world object or system has been created.

B. Computer simulations only run on very powerful computers that are not available to the general public.

C. Computer simulations usually make some simplifying assumptions about the real-world object or system being modeled.

D. It is difficult to change input parameters or conditions when using computer simulations.

My answer:	

Check!

AP Exam Practice Question

In which of the following scenarios would a simulation be the LEAST beneficial?

A. An engineering company wants to test whether a change to a car design will negatively affect fuel efficiency.

B. An insurance company wants to study the effect of cold weather patterns on health-care costs.

C. A manufacturing company wants to determine whether using robots in its facility will increase productivity.

D. A retail company wants to determine the most popular item that was sold on the company's Web site last month.

My answer:

Check!

AP Exam Practice Question

A city planner is using simulation software to study crowd flow out of a large arena after an event has ended. The arena is located in an urban city. Which of the following best describes a limitation of using a simulation for this purpose?

A. The model used by the simulation software cannot be modified once the simulation has been used.

B. The model used by the simulation software often omits details so that it is easier to implement.

C. Running a simulation requires more time to generate data from trials than observing the crowd exiting the arena at various events.

D. Running a simulation requires a large number of observations to be collected before it can be used to explore a problem.

My answer:

Check!

AP Exam Practice Question

For which of the following problems is using a simulation LEAST likely to be beneficial?

(A) Determining the longest word in a textbook

(B) Minimizing the customer wait times at a bank

(C) Predicting the outcomes of weather patterns

(D) Studying the formation of a galaxy

My answer:	

Check!

AP Exam Practice Question

In which of the following scenarios is using a simulation more beneficial than performing a calculation?

Select two answers.

[A] Determining the average grade of the students in a certain class

[B] Keeping track of the high score in a game

[C] Investigating ways to reduce the amount of trash in the ocean

[D] Studying the effect of a genetic change in a population

My answers:	and

Check!

AP Exam Practice Question

A new bank plans to make customer convenience a priority by minimizing the amount of time a customer waits in line. The bank is considering two options: a single line where the customer at the front waits for the next available teller, or separate lines for each teller. The bank decides to use a computer simulation of these two options to determine the average wait time for customers.

Which of the following is NOT true about the bank's plan?

A. The bank can incorporate other factors, such as the number of tellers, in the simulation.

B. The bank can use the simulation to investigate these two options without causing inconvenience for customers.

C. The bank may consider new alternatives based on the simulation results.

D. The simulation will not produce usable results because actual customer data are not available.

My answer:

Check!

AP Exam Practice Question

The transportation department plans to build a new high-speed train route between two cities. The transportation department wants to implement a simulation of the train before any construction begins on this project. Which of the following statements is true about the use of a simulation for this project?

A. A simulation cannot be used to test the train under varying weather conditions.

B. Implementing a simulation is likely to increase the overall costs associated with the project.

C. Other high-speed train routes have been constructed in other locations, so a simulation is not needed.

D. Using a simulation may expose potential safety issues that can be corrected before construction begins.

My answer:

Check!

AP Exam Practice Question

Which of the following is a benefit of using a simulation instead of an experiment to make a prediction?

Select two answers.

A A simulation allows investigation of a phenomenon without the real-world limitations on time, safety, or budget.

B A simulation can be used to model real-world events that are impractical for experiments.

C A simulation always produces the same output, so its results can be verified.

D A simulation produces results that are more accurate than experimental results.

My answers: and

Check!

AP Exam Practice Question

A car manufacturer uses simulation software during the design process for a new car. Which of the following are reasons to use simulation software in this context?

I. Using simulation software can save the company money by helping to compare designs early in the process, before prototype cars are built.

II. Using simulation software can help to identify safety issues by providing data about how different mechanical components will interact in a wide variety of situations.

III. The manufacturer can present simulation software to customers to demonstrate different design possibilities.

A I and II only

B I and III only

C II and III only

D I, II, and III

My answer:

Check!

AP Exam Practice Question

In an upcoming election, 80 percent of New York voters are expected to vote for Mike Jones in a local election. The computer program below is intended to simulate the result of the election with `n` voters, and display the number of votes received by Mike Jones.

```
Line 1:  voteTotal ← 0
Line 2:  REPEAT n TIMES
Line 3:  {
Line 4:     IF (<MISSING CONDITION>)
Line 5:     {
Line 6:        voteTotal ← voteTotal + 1
Line 7:     }
Line 8:   }
Line 9:   DISPLAY (voteTotal)
```

Which two of the following can be used to replace `<MISSING CONDITION>` in `Line 4` so that the program works as intended?

Select two.

(A) `RANDOM (1, 5) = 4`

(B) `RANDOM (1, 5) > 1`

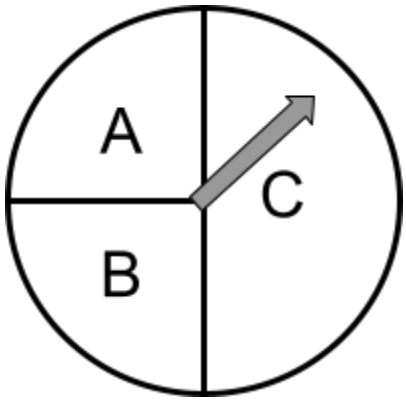
(C) `RANDOM (1, 10) = 8`

(D) `RANDOM (1, 10) ≤ 8`

Your answers:	**and**

Check your answer!

A spinner is divided into three sections, A, B, and C, as shown below. The area labeled C is twice as large as areas A and B, which are equal in size.



Which of the following Spinner procedures correctly simulates the spinner?

(A)

```
PROCEDURE spinner
    spin ← RANDOM 1, 3
    IF spin = 1
        RETURN "A"
    IF spin = 2
        RETURN "B"
    RETURN "C"
```

(B)

```
PROCEDURE spinner
    spin ← RANDOM 1, 3
    IF spin = 1
        RETURN "C"
    IF spin = 2
        RETURN "B"
    RETURN "A"
```

(C)

```
PROCEDURE spinner
    spin ← RANDOM 1, 4
    IF spin = 1
        RETURN "A"
    IF spin = 2
        RETURN "C"
    RETURN "B"
```

(D)

```
PROCEDURE spinner
    spin ← RANDOM 1, 4
    IF spin = 1
        RETURN "A"
    IF spin = 2
        RETURN "B"
    RETURN "C"
```

Your answer:	

Check your answer!

Data Analysis & Metadata

AP EXAM VOCABULARY

- **Data** are the raw values that computers receive from various sources.
  - A **data set** is a *set* of data.
  - **"Big data"** is a common term that describes *extremely large* data sets**.**
- By itself, **data** doesn't *mean anything*; people ***make information out of data.***
- **Data** provides *opportunities* for identifying trends, making connections, drawing conclusions, and addressing problems.
- **Information** is the *result* of analyzing that data and *making sense of what the data is showing*.
- A **spreadsheet** is a software tool that is used to efficiently organize data and help users turn the data into information.
- The power of **spreadsheets** and other similar computing tools really lies in their ability to help you turn massive amounts of raw ***data*** *into useful* ***information*** which allows you to answer questions about the data, uncover **trends**, and make predictions. This is often done using **graphs** of the data to visualize it, or doing things like **filtering** or **sorting** the data.
- Often, a single data source (i.e. one spreadsheet) does **not** contain the data needed to draw a conclusion or answer a particular question about the data. *It may be necessary to combine data from a variety of sources to formulate a conclusion or answer a particular question.*

AP Exam Practice Question

A social media app stores the following data for each post that is made as separate records in a data set:

- The full text of the post
- The date and time the post was made
- The location of the user when the post was made

Assuming that a company has access to this information, which of the following could be determined using the social media post data described above?

I. The number of posts that were taken at a particular location

II. The post made in the last year that had the most number of words in its text

III. The user who made the most recent post

(A) I only

(B) II only
(C) I and II only
(D) I, II, and III

Your answer:	

Check answer

AP Exam Practice Question

A grocery store maintains a single database that contains records with the following information about each item currently for sale in the grocery store:

- Grocery UPC code (the number on the barcode: 

)
- Type of item (meat, bakery, household goods, cereal/breads, milk/juices, produce)
- Selling price (in dollars)
- Size (weight, in ounces)
- Quantity available in the store

Using only the data in the database, which two of the following can be determined?

Select <u>two</u>.

(A) Which type of grocery (meat, bakery, etc.) sell the most frequently
(B) Which grocery item has the lowest price per ounce of weight
(C) Which specific cereal brand is the most popular
(D) Which grocery items are not currently in stock

Your answers:	and

Check answer

AP Exam Practice Question

Computers are often used to search through large data sets to find useful trends and patterns in the data. Which of the following tasks is NOT an example where searching for trends and patterns produces useful information?

(A) A bank analyzing deposit and withdrawal data in customer accounts to identify potential fraudulent account activity

(B) An search engine analyzing users' search history to target advertisements and new products the user may be interested in purchasing

(C) A high school student analyzes her test scores to determine which tests she performed the best on.

(D) A video game publisher analyzing user review data for all released video games to determine which types of games users are most interested in purchasing.

Your answer:	

Check answer

AP EXAM VOCABULARY

- Depending on how data were collected (e.g. with a Google form), the data may *not* be **uniform** (all in the same format). For example, if users enter data into an open field, the way they choose to abbreviate, spell, or capitalize something may vary from user to user.

- The process of *manually* making non-uniform data uniform is called **cleaning** data; this involves a human *manually* going in, looking at the data, and fixing it.

For example, here is some **non-uniform** data ***before*** **cleaning** it:

	A	B	C	D	E
1	year	state	avg math score	avg verbal score	total test takers
47	2006	California	520	501	191740
48	2007	California	518	four hundred	195181
49	2008	CA	517	499	205083
50	2009	California	515	500	207301
51	2010	California	518		210926
52	2011	California		499	222658
53	2012	Calliforrnia	514	496	231964
54	2013	California	514	49 8	234767
55	2014	236923	512	498	California
56	1/1/15	California	507	496	241553

Here is the data ***after* cleaning** it to make it **uniform**:

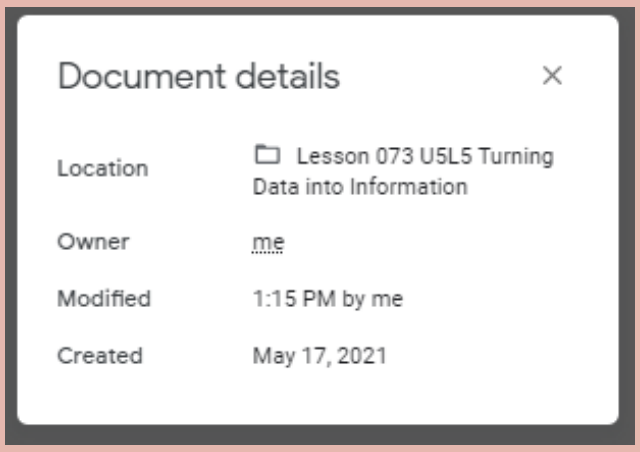
1	year	state	avg math score	avg verbal score	total test takers
47	2006	California	520	501	191740
48	2007	California	518	400	195181
49	2008	California	517	499	205083
50	2009	California	515	500	207301
51	2010	California	518	511	210926
52	2011	California	520	499	222658
53	2012	California	514	496	231964
54	2013	California	514	498	234767
55	2014	California	512	498	236923
56	2015	California	507	496	241553

Data can even include *data about itself*. For example, a music file might contain data describing the song title, artist, album, genre, record label, etc. Or an image file might include data for the date that the picture was taken, the location where it was taken, information about the camera settings, the serial number of the specific camera, etc. This kind of information about data is called **metadata**.

- **Metadata is descriptive data *about* data**. For example, the *piece of data* may be an *image*, while the ***metadata*** may include the *date of creation, the file size of the image, or the geographical location of where the image was photographed.*

Metadata tells you additional information about the data you are working with and can be used to organize and manage your data.

As we saw in the U5L5 lab, some information you might see if you looked at the **metadata** for a Google doc include location of the document in Google Drive, the owner of the document, and when the document was last modified and created:



Metadata of a Google Doc document

AP Exam Practice Question

An online music streaming service contains metadata for each song. The metadata is intended to help a search feature find songs that users are interested in listening to. Which of the following is LEAST likely to be contained in the metadata of each song?

(A) The title and artist of the song

(B) The date the song was released

(C) The audio data of the song itself

(D) The genre of the song (e.g., dance, rock, pop, rap, etc.)

Your answer:	

Check your answer

AP Exam Practice Question

A certain social media website allows users to post messages and to comment on other messages that have been posted. When a user posts a message, the message itself is considered *data*. In addition to the data, the site stores the following *metadata*:

- The time the message was posted
- The name of the user who posted the message
- The names of any users who comment on the message and the times the comments were made

For which of the following would it be more useful to analyze the *data* instead of the *metadata*?

(A) To determine the users who post messages most frequently

(B) To determine the topics that many users are posting about

(C) To determine the time of day that the site is most active

(D) To determine which posts have received the greatest number of comments

Your answer:	

Check your answer

AP Exam Practice Question

A student is creating a Web site that is intended to display information about a city based on a city name that a user enters in a text field. Which of the following are likely to be challenges associated with processing city names that users might provide as input?

Select two answers.

A	Users might attempt to use the Web site to search for multiple cities.
B	Users might enter abbreviations for the names of cities.
C	Users might misspell the name of the city.
D	Users might be slow at typing a city name in the text field.

My answer:

Check!

DONE WITH PACKET 4

Great job completing the AP Exam Reviews!

You are ready to ace the exam! 😎

Answer (back)

Correct Answer: D

A video game character can face toward one of four directions: north, south, east, and west. Each direction is stored in memory as a sequence of four bits. A new version of the game is created in which the character can face toward one of eight directions, adding northwest, northeast, southwest, and southeast to the original four possibilities. Which of the following statements is true about how the eight directions must be stored in memory?

A. Four bits are not enough to store the eight directions. Five bits are needed for the new version of the game.

B. Four bits are not enough to store the eight directions. Eight bits are needed for the new version of the game.

C. Four bits are not enough to store the eight directions. Sixteen bits are needed for the new version of the game.

D. Four bits are enough to store the eight directions. ✔

Explanations:

Answer A

Incorrect. Four bits can represent 2^4, or 16 pieces of information, so it is not necessary to add an extra bit.

Answer B

Incorrect. Four bits can represent 2^4, or 16 pieces of information, so it is not necessary to add extra bits.

Answer C

Incorrect. Four bits can represent 2^4, or 16 pieces of information, so it is not necessary to add extra bits.

Answer D

Correct. Four bits can represent 2^4, or 16 pieces of information.

Answers (back)

True or false: 3 bits would be enough to store 9 unique values in memory. *If false, correct it to make it true!*	**False! 3 bits gives 2^3 = 8 possible values, so this would not be enough to store 9 values** To make this true: "**4** bits would be enough to store 8 unique values in memory" 4 bits could store 2^4 = 16 values, and since we only need 9, there would be 7 values unused.
What is the fewest number of bits you would need to represent 80 different values?	**7 bits** Look for the first power of 2 past 80: 2^1 = 2 2^2 = 4 2^3 = 8 2^4 = 16 2^5 = 32 2^6 = 64 **2^7 = 128** 2^7 = 128 is the first power of 2 that is bigger than 80, so **7** is the fewest number of bits needed to store 80 unique values.
How many *different* values can be represented with: A. 3 bits B. 4 bits C. 8 bits D. 9 bits E. 16 bits	A. 2^3 = **8** different values B. 2^4 = **16** different values C. 2^8 = **256** different values D. 2^9 = **512** different values E. 2^16 = **65536** different values
How does the number of possible values change with the addition of *one* more bit?	Each additional bit **doubles** the number of possible values (e.g. 3 bits is 8 values and 4 bits is 16 values, which is double the values of 3 bits)
How many times *more* values can be represented with: A. 8 bits than with 4 bits? B. 16 bits than with 15 bits? C. 32 bits than with 31 bits	Each additional bit doubles the number of possible values, so: **A**. 8 bits is 4 bits more than 4 bits, so you can represent 2 x 2 x 2 x 2 = **16 times more values** with 8 bits than 4 bits **B**. 16 bits is 1 bit more than 15 bits, so you can represent **2 times more values** with 16 bits than 15 bits (don't believe it? Work it out: 2^15 = 32768 and 2^16 = 65536, which is 2 times more) **C**. 32 bits is 1 bit more than 31 bits, so you can represent **2 times more values** with 32 bits than 31 bits

a. How many **bits** does this binary sequence of bits represent? `100101100011010101100101` **b.** How many **bytes** does it represent?	a. Each digit is 1 bit, so the binary sequence represents **24 bits** (since there are 24 digits) b. A byte is 8 bits, so 24 bits is **3 bytes**

(back)

Answer (back)

Answers: B & C

Practice AP Exam Question

A. The position of the minute hand of a clock

B. The remainder when dividing a whole number by 2

C. The value of a Boolean variable

D. The volume of a car radio

Correct Answers

B	C

EXPLANATION

Choice A is incorrect because the position of a minute hand on a clock can have many possible values (60). Binary digits can only store two possible values: 0 or 1.

Choice B is correct because when dividing a whole number (0, 1, 2, 3, …) by 2, the remainder will always be 0 or 1. A binary digit, by its definition, stores 0 or 1.

Choice C is correct because the value of a Boolean variable is either “true” or “false.” These two possible values can be represented by the binary digits 0 or 1.

Choice D is incorrect because the volume of a car radio can have many possible values between the lowest possible setting and the highest possible setting. Binary digits can only store two possible values: 0 or 1.

Answer (back)

Answer: B

Practice AP Exam Question #2

A store uses binary numbers to assign a unique binary sequence to each item in its inventory. What is the minimum number of bits required for each binary sequence if the store has 600 items in its inventory?

A. 9 bits

B. 10 bits

C. 600 bits

D. 1024 bits

Your answer:

B

EXPLANATION

In order to store 600 different values, you would need a minimum of 10 bits:

2^8 = 256 values
2^9 = 512 values
2^10 = 1,024 values

Since 2^10 is the *first power of 2 greater than 600*, 10 is the number of bits you would need.

Note that 9 bits can only accommodate 512 values, so that isn't enough!

And 600 bits would be 2^600 = some huuuuuuuuuggee number so that's wrong. Same with 1024 bits: calculating 2^1024 would probably crash your computer!

Answer (back)

Correct Answers: B & C

Which of the following can be represented by a single binary digit?

Select two answers.

A The position of the minute hand of a clock

B The remainder when dividing a whole number by 2

C The value of a Boolean variable

D The volume of a car radio

Explanations:

Answer A

This option is incorrect. The position of a minute hand on a clock can have many possible values. Binary digits can only store two possible values: 0 or 1.

Answer B

This option is correct. When dividing a whole number (0, 1, 2, 3, ...) by 2, the remainder will always be 0 or 1. A binary digit, by its definition, stores 0 or 1.

Answer C

This option is correct. The value of a Boolean variable is either "true" or "false." These two possible values can be represented by the binary digits 0 or 1.

Answer D

This option is incorrect. The volume of a car radio can have many possible values between the lowest possible setting and the highest
possible setting. Binary digits can only store two possible values: 0 or 1.

Answer (back)

Correct Answer: C

An online store uses 6-bit binary sequences to identify each unique item for sale. The store plans to increase the number of items it sells and is considering using 7-bit binary sequences. Which of the following best describes the result of using 7-bit sequences instead of 6-bit sequences?

A 2 more items can be uniquely identified.

B 10 more items can be uniquely identified.

C 2 times as many items can be uniquely identified.

D 10 times as many items can be uniquely identified.

Explanations:

Answer A

This option is incorrect. Adding an extra binary digit allows for two times as many items to be identified, not two more.

Answer B

This option is incorrect. Adding an extra binary digit allows for two times as many items to be identified, not 10 more.

Answer C

This option is correct. Using 6-bit binary sequences allows for 26 or 64 different items to be identified. Using 7-bit binary sequences allows for 27 or 128 different items to be identified. Thus there are two times as many items that can be uniquely identified.

(Note: in the answer above, 26 should read 2^6 and 27 should read 2^7 **)**

Answer D

This option is incorrect. Adding an extra binary digit allows for two times as many items to be identified, not 10 times as many.

Answer (back)

Correct Answer: C

Consider the 4-bit binary numbers 0011, 0110, and 1111. Which of the following decimal values is NOT equal to one of these binary numbers?

A 3

B 6

C 9 ✔

D 15

Explanations:

Answer A

Incorrect. Binary 0011 is equivalent to $2^1 + 2^0$, or decimal 3.

Answer B

Incorrect. Binary 0110 is equivalent to $2^2 + 2^1$, or decimal 6.

Answer C

Correct. Binary 0011 is equivalent to $2^1 + 2^0$, or decimal 3. Binary 0110 is equivalent to $2^2 + 2^1$, or decimal 6. Binary 1111 is equivalent to $2^3 + 2^2 + 2^1 + 2^0$, or decimal 15. Decimal 9 is not equivalent to any of the given binary numbers.

Answer D

Incorrect. Binary 1111 is equivalent to $2^3 + 2^2 + 2^1 + 2^0$, or decimal 15.

Answers (back)

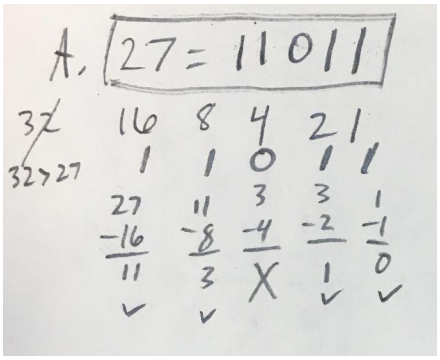
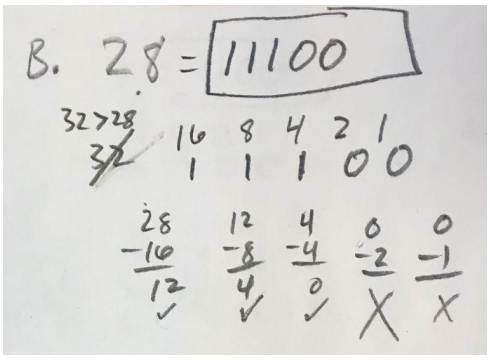
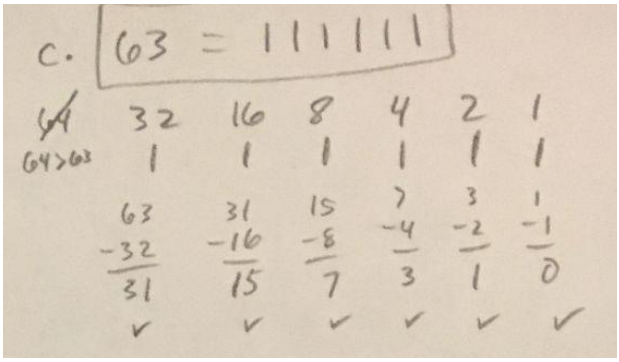
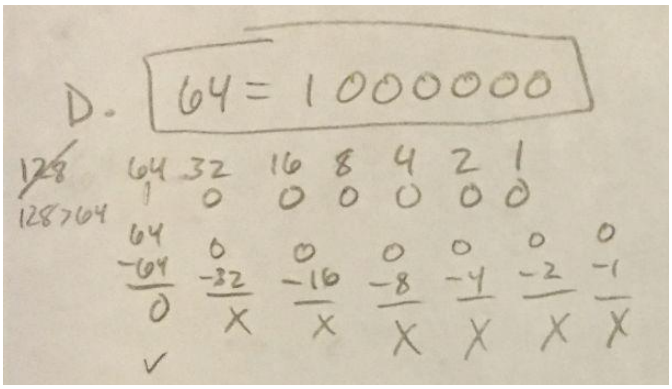
Binary	Decimal
A. `10`	10 = (1 x 2) + (0 x 1) = 2 + 0 = 2
B. `101`	101 = (1 x 4) + (0 x 2) + (1 x 1) = 4 + 0 + 1 = 5
C. `111`	111 = (1 x 4) + (1 x 2) + (1 x 1) = 4 + 2 + 1 = 7
D. `1000`	1000 = (1 x 8) + (0 x 4) + (0 x 2) + (0 x 1) = 8 + 0 + 0 + 0 = 8 (note that 1000 -- which is 8 -- is one more than 111, which is 7)
E. `1011`	1011 = (1 x 8) + (0 x 4) + (1 x 2) + (1 x 1) = 8 + 0 + 2 + 1 = 11
F. `11000`	11000 = (1 x 16) + (1 x 8) + (0 x 4) + (0 x 2) + (0 x 1) = 16 + 8 + 0 + 0 + 0 = 24

Answers (back)

Decimal	Binary		Decimal	Binary
0	0000		**8**	1000
1	0001		**9**	1001
2	0010		**10**	1010
3	0011		**11**	1011
4	0100		**12**	1100
5	0101		**13**	1101
6	0110		**14**	1110
7	0111		**15**	1111

True or false: The decimal value 16 can be represented in binary using 4 bits.	**False!** 15 (which is 1111 in binary) is the largest possible value that can be represented with 4 bits 16 in binary is 10000, which requires **5 bits**.

Answers (back)

Translate/convert each **decimal** (base 10) representation below *into* **binary** (base 2)		
Decimal	**Binary**	**Scrap Paper**
A. 27	**11011**	
B. 28	**11100**	
C. 63	**111111**	
D. 64	**1000000**	

Answer (back)

Correct Answer: C

Practice AP Exam Question

A particular program uses 3 bits to represent whole numbers. When that program adds the numbers 3 and 6, the result is given as 0. Identify the best explanation of the result.

A. The result was incorrect due to data being corrupted during the operation due to a technical glitch.

B. The result was incorrect due to a round-off error.

C. The result was incorrect due to an overflow error.

D. The result is correct when all values are converted to binary.

Correct answer:

C

EXPLANATION:

Overflow errors occur when there aren't enough bits to hold a number.

3 + 6 = 9, which equals 1001 in binary; 1001 takes up *4 bits* and the program only uses *3 bits* for numbers, so an overflow error occurs which causes 0 to be displayed rather than the expected value of 9.

Answer (back)

Correct Answer: A

Each student that enrolls at a school is assigned a unique ID number, which is stored as a binary number. The ID numbers increase sequentially by 1 with each newly enrolled student. If the ID number assigned to the last student who enrolled was the binary number 1001 0011, what binary number will be assigned to the next student who enrolls?

- A 1001 0100 ✔
- B 1001 0111
- C 1101 0100
- D 1101 0111

Explanation:

Adding 1 to 10010**011** will result in 10010**100**; the digits in bold are the ones that are changed when adding 1. You could check by converting both binary numbers to decimal: 10010011 = 147 and 10010100 = 148 (which is 1 more than 147).

Answer (back)

Correct Answer: C

ASCII is a character-encoding scheme that uses 7 bits to represent each character. The decimal (base 10) values 65 through 90 represent the capital letters A through Z, as shown in the table below.

Decimal	ASCII Character
65	A
66	B
67	C
68	D
69	E
70	F
71	G
72	H
73	I
74	J
75	K
76	L
77	M

Decimal	ASCII Character
78	N
79	O
80	P
81	Q
82	R
83	S
84	T
85	U
86	V
87	W
88	X
89	Y
90	Z

What ASCII character is represented by the binary (base 2) number 1001010 ?

A H

B I

C J

D K

Explanations:

Answer A

This option is incorrect. The table shows that the letter H is represented by the decimal value 72, which in binary (base 2) is 1001000.

Answer B

This option is incorrect. The table shows that the letter I is represented by the decimal value 73, which in binary (base 2) is 1001001.

Answer C

This option is correct. The table shows that the letter J is represented by the decimal value 74, which in binary (base 2) is 1001010.

Answer D

This option is incorrect. The table shows that the letter K is represented by the decimal value 75, which in binary (base 2) is 1001011.

Answer (back)

Correct Answer: A

Consider the following numeric values.

- Binary 1011
- Binary 1101
- Decimal 5
- Decimal 12

Which of the following lists the values in order from least to greatest?

A Decimal 5, binary 1011, decimal 12, binary 1101

B Decimal 5, decimal 12, binary 1011, binary 1101

C Decimal 5, binary 1011, binary 1101, decimal 12

D Binary 1011, binary 1101, decimal 5, decimal 12

Explanations:

Answer A

Correct. Binary 1011 is equivalent to $2^3 + 2^1 + 2^0$, or decimal 11, and binary 1101 is equivalent to $2^3 + 2^2 + 2^0$, or decimal 13. The order of the numbers (written in their equivalent decimal format) is 5, 11, 12, 13.

Answer B

Incorrect. Binary 1011 is equivalent to decimal 11 and should be placed before decimal 12.

Answer C

Incorrect. Binary 1101 is equivalent to decimal 13 and should be placed after decimal 12.

Answer D

Incorrect. Binary 1011 is equivalent to decimal 11 and should be placed after decimal 5. Binary 1101 is equivalent to decimal 13 and should be placed after decimal 12.

Answer (back)

Correct Answer: C

A certain programming language uses 4-bit binary sequences to represent nonnegative integers. For example, the binary sequence `0101` represents the corresponding decimal value `5`. Using this programming language, a programmer attempts to add the decimal values `14` and `15` and assign the sum to the variable `total`. Which of the following best describes the result of this operation?

A. The correct sum of `29` will be assigned to the variable `total`.

B. An overflow error will occur because 4 bits is not large enough to represent either of the values `14` or `15`.

C. An overflow error will occur because 4 bits is not large enough to represent `29`, the sum of `14` and `15`. ✓

D. A round-off error will occur because the decimal values `14` and `15` are represented as approximations due to the fixed number of bits used to represent numbers.

Explanations:

Answer A

Incorrect. The sum is too large to be represented with a 4-bit representation.

Answer B

Incorrect. The values `14` and `15` can each be represented using a 4-bit representation. However, their sum is too large to be represented with a 4-bit representation.

Answer C

Correct. The largest binary value that can be represented using 4 bits is `1111`, which is equal to the decimal value `15`. Since the sum is larger than the largest representable value, an overflow error will occur.

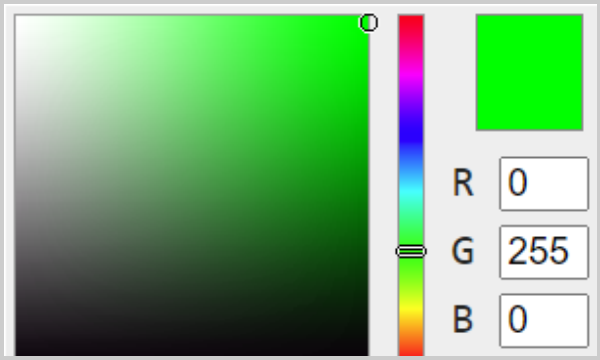
Answer D

Incorrect. A round-off error occurs when there is a lack of precision in the representation of real numbers. In this case, an overflow error occurs because the sum exceeds the largest representable value.

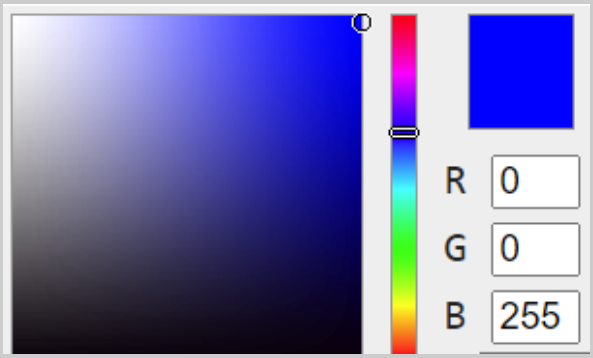
Examples of colors in RGB (back)

All red (R = 255, G = 0, B = 0):

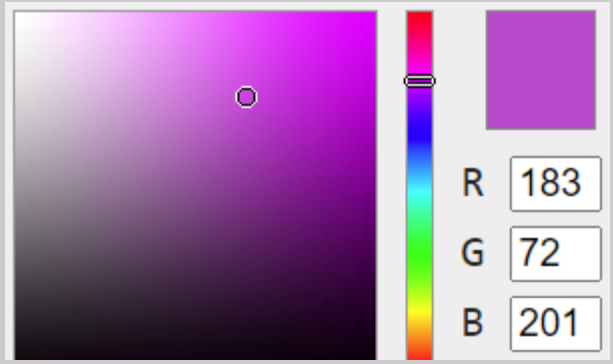
All green (R = 0, G = 255, B = 0):



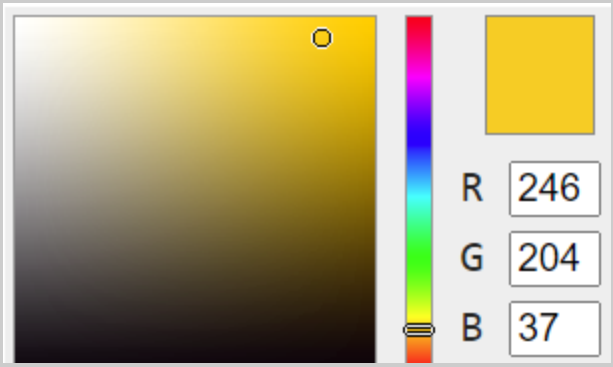
All blue (R = 0, G = 0, B = 255):



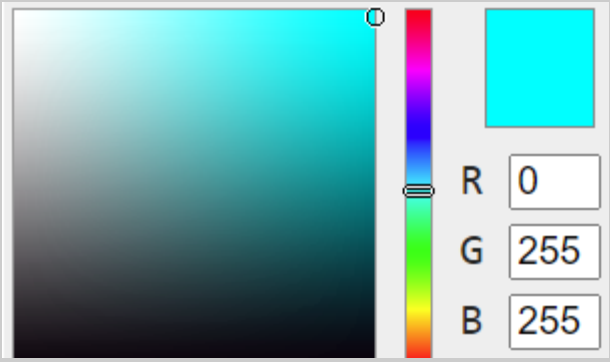
Random purple (R = 183, G = 72, B = 201):



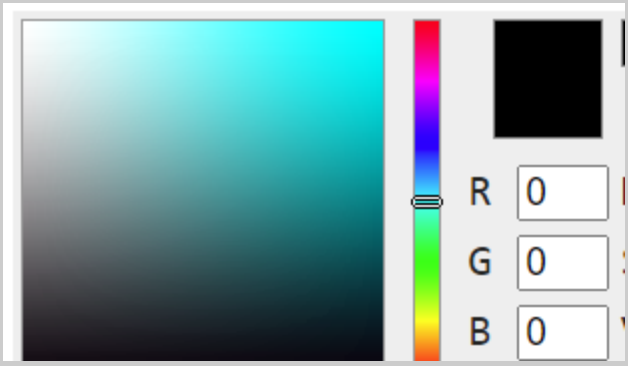
Random yellow (R = 246, G = 204, B = 37)



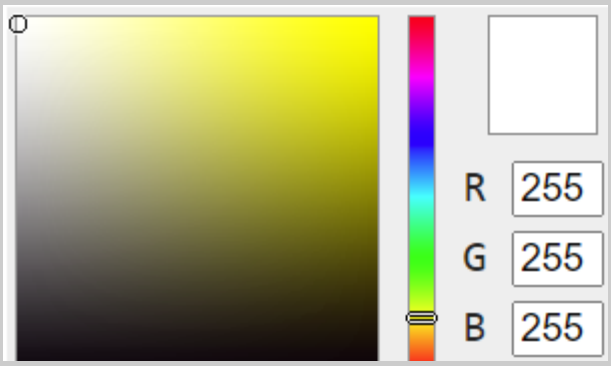
Cyan (R = 0, G = 255, B = 255)



Black (R = 0, G = 0, B = 0):



White (R = 255, G = 255, B = 255):



Answer (back)

Correct Answer: D

Which of the following are true statements about the data that can be represented using binary sequences?

I. Binary sequences can be used to represent strings of characters.
II. Binary sequences can be used to represent colors.
III. Binary sequences can be used to represent audio recordings.

A I only

B I and II only

C II and III only

D I, II, and III

Explanations:

Answer A

Incorrect. Statement II is true because colors can be encoded as sequences of bits. Statement III is true because sequences of bits can be used to represent sound.

Answer B

Incorrect. Statement III is true because sequences of bits can be used to represent sound.

Answer C

Incorrect. Statement I is true because strings of characters can be represented by sequences of bits.

Answer D

Correct. All digital data is represented at the lowest level as sequences of bits. Statement I is true because strings of characters can be represented by sequences of bits. Statement II is true because colors can be encoded as sequences of bits. Statement III is true because sequences of bits can be used to represent sound.

Answer (back)

Correct Answer: D

The position of a runner in a race is a type of analog data. The runner's position is tracked using sensors. Which of the following best describes how the position of the runner is represented digitally?

A. The position of the runner is determined by calculating the time difference between the start and the end of the race and making an estimation based on the runner's average speed.

B. The position of the runner is measured and rounded to either 0 or 1 depending on whether the runner is closer to the starting line or closer to the finish line.

C. The position of the runner is predicted using a model based on performance data captured from previous races.

D. The position of the runner is sampled at regular intervals to approximate the real-word position, and a sequence of bits is used to represent each sample. ✔

Explanations:

Answer A

Incorrect. While a runner's position could be estimated using the runner's average speed, a more accurate representation of the position over time can be achieved using a sampling technique.

Answer B

Incorrect. Analog data such as a runner's position is typically represented with a sequence of bits rather than with a single bit. Using a sequence of bits allows for greater accuracy in each measurement.

Answer C

Incorrect. While a runner's position could be predicted using a model, a more accurate representation of the position over time can be achieved using a sampling technique.

Answer D

Correct. Analog data, like the runner's position, have values that change smoothly, rather than in discrete intervals. Analog data can be approximated digitally by measuring values of the analog signal at regular intervals called samples. The samples are represented digitally as sequences of bits.

Answer (back)

Correct Answer: A

A student is recording a song on her computer. When the recording is finished, she saves a copy on her computer. The student notices that the saved copy is of lower sound quality than the original recording. Which of the following could be a possible explanation for the difference in sound quality?

A. The song was saved using fewer bits per second than the original song.

B. The song was saved using more bits per second than the original song.

C. The song was saved using a lossless compression technique.

D. Some information is lost every time a file is saved from one location on a computer to another location.

Explanations:

Answer A

This option is correct. The representation of sound as data involves the computational manipulation of information. For one copy of a song to have a lower sound quality than another copy, a lower ratio of bits per second must have been used.

Answer B

This option is incorrect. The representation of sound as data involves the computational manipulation of information. For one copy of a song to have a lower sound quality than another copy, a lower ratio of bits per second must have been used. Because the student noticed that the saved copy is of lower sound quality, it was not saved using more bits per second than the original song.

Answer C

This option is incorrect. If the song was saved using a lossless compression technique, there should be no change in sound quality.

Answer D

This option is incorrect. Digital information that is saved in one location on a computer can be saved to another location exactly, with no change in sound quality.

Answer (back)

Correct Answer: C

A new program was developed for a bank. While testing the program by adding and subtracting decimal values, the developer discovers that some results appear to be imprecise. Which of the following is the most likely cause?

A. The numbers are represented using an unlimited number of bits, resulting in round-off errors.

B. The numbers are represented using an unlimited number of bits, resulting in overflow errors.

C. The numbers are represented using a fixed number of bits, resulting in round-off errors.

D. The numbers are represented using a fixed number of bits, resulting in overflow errors.

Explanation:

The answer is C because a *fixed* number of bits used to represent real numbers limits the range of the values that can be represented, and this limitation can result in **round-off** errors when you try to work with them or display them. Numbers like decimals and fractions are particularly prone to round-off errors! Note that the very reason these types of numbers are problematic is because they can't be represented exactly in binary, due to having a **fixed** number of bits to work with -- although *if* a computer did have an *infinite* number of bits (which it *doesn't*), then round-off errors wouldn't exist!

Note that an **overflow error** occurs when a value exceeds the *maximum* representable value. This type of error does not typically lead to *imprecision* of values, but rather to weird and unpredictable results (like displaying 1 when adding 7 and 9).

Answer (back)

Correct Answer: B

Which of the following is an advantage of a lossless compression algorithm over a lossy compression algorithm?

A. A lossless compression algorithm can guarantee that compressed information is kept secure, while a lossy compression algorithm cannot.

B. A lossless compression algorithm can guarantee reconstruction of original data, while a lossy compression algorithm cannot.

C. A lossless compression algorithm typically allows for faster transmission speeds than does a lossy compression algorithm.

D. A lossless compression algorithm typically provides a greater reduction in the number of bits stored or transmitted than does a lossy compression algorithm.

Explanations:

Answer A

Incorrect. The ability to keep data secure is not a primary function of a compression algorithm.

Answer B

Correct. Lossless compression algorithms are guaranteed to be able to reconstruct the original data, while lossy compression algorithms are not.

Answer C

Incorrect. In situations where transmission time is maximally important, lossy compression algorithms are typically chosen, as lossy compression typically provides a greater reduction in file size.

Answer D

Incorrect. Lossless compression algorithms usually achieve less reduction in the number of bits stored or transmitted than do lossy compression algorithms.

Answer (back)

Correct Answer: A

A user wants to save a data file on an online storage site. The user wants to reduce the size of the file, if possible, and wants to be able to completely restore the file to its original version. Which of the following actions best supports the user's needs?

A Compressing the file using a lossless compression algorithm before uploading it

B Compressing the file using a lossy compression algorithm before uploading it

C Compressing the file using both lossy and lossless compression algorithms before uploading it

D Uploading the original file without using any compression algorithm

Explanations:

Answer A

Correct. Lossless compression algorithms allow for complete reconstruction of the original data and typically reduce the size of the data.

Answer B

Incorrect. While a lossy compression algorithm will reduce the size of the data, it does not allow for complete reconstruction of the original data.

Answer C

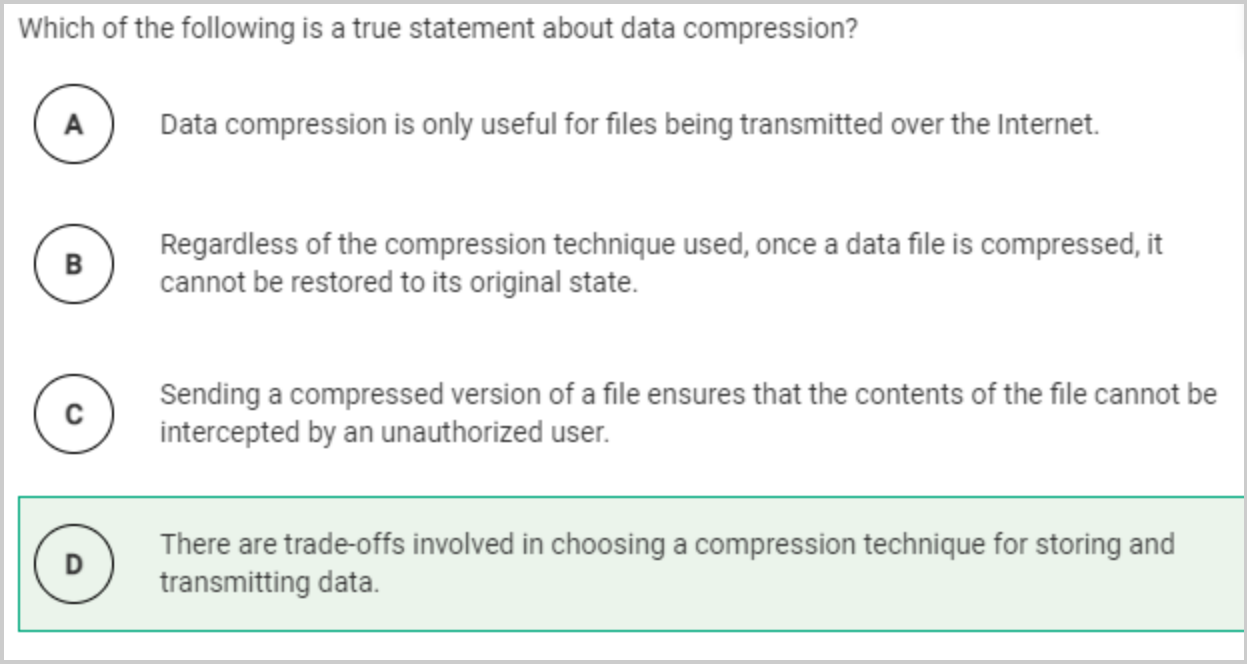
Incorrect. Applying lossy compression the to file will prevent the user from restoring it to its original version.

Answer D

Incorrect. Uploading the original file allows complete reconstruction of the original data but does not reduce the size of the file.

Answer (back)

Correct Answer: D



Explanations:

A is incorrect because data compression can be used for reducing file sizes on your computer or phone -- not just those you send on the internet.

B is incorrect because if you use a *lossless* data compression technique, you *can* restore it to its original state (only lossy compression cannot be restored).

C is incorrect because data compression has nothing to do with security of transmission (that is where encryption and decryption come into play -- not compression).

D is correct because you need to decide how much compression to use versus how much quality you are OK losing. It's a tradeoff! High compression can reduce the size of a file by a *lot* but you will get worse quality.

Answer (back)

Correct Answer: A

A person wants to transmit an audio file from a device to a second device. Which of the following scenarios best demonstrates the use of lossless compression of the original file?

A. A device compresses the audio file before transmitting it to a second device. The second device restores the compressed file to its original version before playing it. ✔

B. A device compresses the audio file by removing details that are not easily perceived by the human ear. The compressed file is transmitted to a second device, which plays it.

C. A device transmits the original audio file to a second device. The second device removes metadata from the file before playing it.

D. A device transmits the original audio file to a second device. The second device plays the transmitted file as is.

Explanations:

Answer A

Correct. Lossless compression is a technique that allows for complete reconstruction of the original data.

Answer B

Incorrect. This technique is an example of lossy audio compression.

Answer C

Incorrect. Removing a file's metadata does not allow for reconstruction of the original data.

Answer D

Incorrect. This situation does not make use of any compression.

Answer (back)

Correct Answer: C

Digital images are often represented by the red, green, and blue values (an RGB triplet) of each individual pixel in the image. A photographer is manipulating a digital image and overwriting the original image. Which of the following describes a lossless transformation of the digital image?

A. Compressing the image in a way that may lose information but will suffer only a small loss of image quality.

B. Creating the gray scale of an image by averaging the amounts of red, green, and blue in each pixel and assigning this new value to the corresponding pixel in the new image. The new value of each pixel represents a shade of gray, ranging from white to black.

C. Creating the negative of an image by creating a new RGB triplet for each pixel in which each value is calculated by subtracting the original value from 255. The negative of an image is reversed from the original; light areas appear dark, and colors are reversed. ✔

D. Modifying part of the image by taking the pixels in one part of the picture and copying them to the pixels in another part of the picture.

Explanations:

Answer A

This option is incorrect. Even a small loss of information is considered lossy. This is not a lossless transformation.

Answer B

This option is incorrect. By converting the red, green, and blue values in each pixel into a single number, information has been lost that cannot be restored.

Answer C

This option is correct. If a negative of the original image is made, each RGB triplet value will be computed by subtracting the original value from 255. The original value can then be restored by subtracting the new value from 255. This process is lossless because the exact original can be restored.

Answer D

This option is incorrect. By copying pixels from one part of the image to another part of the image, the old RGB values of the pixels have been lost.

Answer (back)

Correct Answer: C

Byte pair encoding is a text encoding technique that looks for groups of characters that appear in the string more than once and replaces each instance with a corresponding character that does not appear in the string. These replacements are stored as a "dictionary" that gets appended to the compressed text as metadata.

For example, the string WHO_DID_IT_AND_WHY_DID_THEY_DO_IT can be can be encoded as a shorter string $O_&_%_AND_$Y_&_THEY_DO_% by replacing all instances of "WH" with "$", replacing all instances of "DID" with "&", and all instances of "IT" with "%"

For which of the following strings is it NOT possible to use byte pair encoding to shorten its length?

A. TO_BE_OR_NOT_TO_BE

B. ABRACADABRA

C. GIVE_ME_A_BREAK

D. ITSY_BITSY_SPIDER

Explanations:

A is incorrect because there is at least one group of characters that is repeated ("TO" and "BE") and so it *would* be possible to use byte-pair encoding on this string.

B is incorrect because there is at least one group of characters that is repeated ("ABRA") and so it *would* be possible to use byte-pair encoding on this string.

C is correct because no group of characters is repeated in the string, and so it would NOT be possible to use byte pair encoding.

D is incorrect because there is at least one group of characters that is repeated ("ITSY") and so it *would* be possible to use byte-pair encoding on this string.

Answer (back)

Correct Answer: C

A programmer is developing software for a social media platform. The programmer is planning to use compression when users send attachments to other users. Which of the following is a true statement about the use of compression?

A. Lossless compression of video files will generally save more space than lossy compression of video files.

B. Lossless compression of an image file will generally result in a file that is equal in size to the original file.

C. Lossy compression of an image file generally provides a greater reduction in transmission time than lossless compression does.

D. Sound clips compressed with lossy compression for storage on the platform can be restored to their original quality when they are played.

Explanations:

Answer A

Incorrect. Lossy data compression algorithms can usually provide a greater reduction in the space required than lossless compression algorithms can.

Answer B

Incorrect. Lossless compression usually results in a file that is smaller in size than the original file.

Answer C

Correct. Although fewer bits may be stored, information is not necessarily lost when lossy compression is applied to an image.

Answer D

Incorrect. Lossy compression algorithms allow only an approximation of the original data to be reconstructed.

Answer (back)

Correct Answer: B

A sorted list of numbers contains 200 elements. Which of the following is closest to the maximum number of list elements that will need to be examined when performing a binary search for a particular value in the list?

A 5

B 8 ✔

C 100

D 200

Explanations:

Answer A

Incorrect. After 5 iterations of the binary search, there will still be 6 elements remaining.

Answer B

Correct. The binary search algorithm starts at the middle of the sorted list and repeatedly eliminates half the elements until the desired value is found or all the elements have been eliminated. For a list with 200 elements, the list will be cut in half a maximum of 7 times (with a total of 8 elements examined). The list starts with 200 elements, then is reduced to 100 elements, then to 50 elements, then to 25 elements, then to 12 elements, then to 6 elements, then to 3 elements, and then, finally, to 1 element.

Answer C

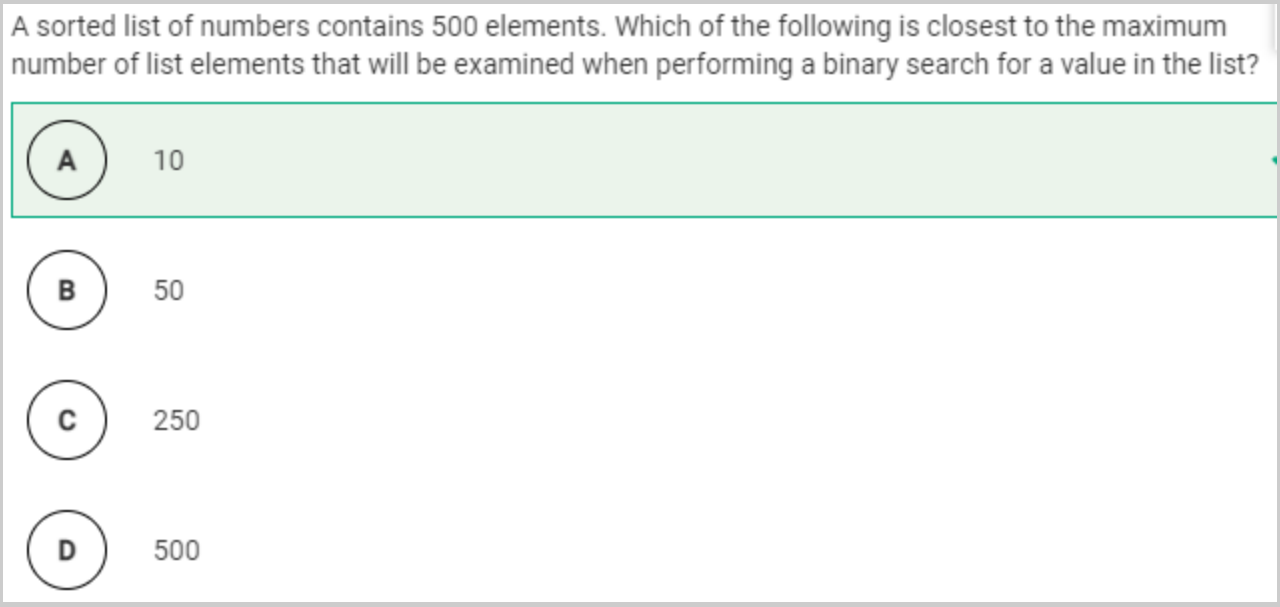
Incorrect. While a binary search on a list of length 200 will eliminate 100 elements on the first iteration, fewer than 100 iterations are needed to eliminate all elements.

Answer D

Incorrect. A linear search of a list of length 200 may examine all 200 elements. A binary search requires fewer elements to be examined.

Answer (back)

Correct Answer: A



Explanation:

Similar to the previous problem, to get the maximum number of comparisons needed for binary search, halve the data until you get to 1:

500 / 2 = 250
250 / 2 = 125
125 / 2 = 63 (round up)
63 / 2 = 32
32 / 2 = 16
16 / 2 = 8
8 / 2 = 4
4 / 2 = 2
2 / 2 = 1

This is 9 halvings, which is **closest to 10**.

Answer (back)

Correct Answer: A

The list `listOne` is a sorted list of numbers that contains 700 elements. The list `listTwo` is a sorted list of numbers that contains 900 elements. Let *x* represent the maximum number of list elements that will need to be examined when performing a binary search for a value in `listOne`, and let *y* represent the maximum number of list elements that will need to be examined when performing a binary search for a value in `listTwo`. Which of the following statements about *x* and *y* is true?

A. The value of *x* is approximately equal to the value of *y*. ✔

B. The value of *x* is approximately 10 less than the value of *y*.

C. The value of *x* is approximately 13 less than the value of *y*.

D. The value of *x* is approximately 200 less than the value of *y*.

Explanations:

Answer A

Correct. The binary search algorithm starts at the middle of the list and repeatedly eliminates half of the elements until the desired value is found or all elements have been eliminated. For `listOne`, the list will be cut in half a maximum of 9 times, with a total of 10 elements needing to be examined. The list will start with 700 elements, then will be reduced to 350 elements, then to 175 elements, then to 87 elements, then to 43 elements, then to 21 elements, then to 10 elements, then to 5 elements, then to 2 elements, and then, finally, to 1 element. For `listTwo`, the list will also be cut in half a maximum of 9 times, with a total of 10 elements needing to be examined. The list will start with 900 elements, then will be reduced to 450 elements, then to 225 elements, then to 112 elements, then to 56 elements, then to 28 elements, then to 14 elements, then to 7 elements, then to 3 elements, and then, finally, to 1 element.

Answer B

Incorrect. The value 10 is the maximum number of elements each binary search will examine, not the difference between the maximum number of elements examined.

Answer C

Incorrect. The binary search on `listOne` will examine a maximum of 10 elements. Even though `listTwo` contains about 30% more elements than `listOne`, the binary search on `listTwo` does not require 30% more elements to be examined.

Answer D

Incorrect. The binary search on `listOne` will examine a maximum of 10 elements. Even though `listTwo` contains 200 more elements than `listOne`, the binary search on `listTwo` does not require 200 more elements to be examined.

Answer (back)

Correct Answer: D

For which of the following lists can a binary search be used to search for an item in the list?

I. `["blue", "green", "jade", "mauve", "pink"]`
II. `[5, 5, 5, 5, 6, 7, 8, 8, 8]`
III. `[10, 5, 3, 2, -4, -8, -9, -12]`

A. I only

B. III only

C. I and III only

D. I, II, and III

Explanations:

Answer A

Incorrect. The second and third lists are sorted (the third one in reverse order), so a binary search is also appropriate for these lists.

Answer B

Incorrect. The first and second lists are sorted (the first one in alphabetical order), so a binary search is also appropriate for these lists.

Answer C

Incorrect. The second list is sorted, so a binary search is also appropriate for this list.

Answer D

Correct. In order for a binary search on a list to work, the list must be sorted. All three lists are sorted (the first one in alphabetical order, the third one in reverse order), so a binary search is appropriate for all of them.

Answer (back)

Correct Answer: B

A large number of genetic codes are stored as binary values in a list. Which one of the following conditions must be true in order for a researcher to obtain the correct result when using a binary search algorithm to determine if a given genetic code is in the list?

A The genetic codes must be converted from binary to decimal numbers.

B The list must be sorted based on the genetic code values. ✓

C The number of genetic code values in the list must be a power of 2.

D The number of genetic code values in the list must be even.

Explanations:

Answer A

Incorrect. The values can be stored in any numerical base as long as they are sorted.

Answer B

Correct. In order for a binary search on a list to work as intended, the list must be sorted.

Answer C

Incorrect. A binary search will work on any sorted list, regardless of whether the number of elements in the list is even.

Answer D

Incorrect. A binary search will work on any sorted list, regardless of whether the number of elements in the list is a power of 2.

Answer (back)

Correct Answer: B

Suppose that a list of numbers contains values `[-4, -1, 1, 5, 2, 10, 10, 15, 30]`. Which of the following best explains why a binary search should NOT be used to search for an item in this list?

A. The list contains both positive and negative elements.

B. The elements of the list are not sorted.

C. The list contains an odd number of elements.

D. The list contains duplicate elements.

Explanations:

Answer A

Incorrect. A binary search will work on any sorted list, regardless of whether the list contains both positive and negative values.

Answer B

Correct. In order for a binary search on a list to work, the elements of the list must be sorted.

NOTE: Although the list is *mostly* sorted, the 5 and 2 are **NOT** sorted properly, and so the list is considered “unsorted”.

Answer C

Incorrect. A binary search will work on any sorted list, regardless of whether the list contains an even or odd number of elements.

Answer D

Incorrect. A binary search will work on any sorted list, regardless of whether the list contains duplicate elements.

Answer (back)

Correct Answer: B

A time stamp indicates the date and time that a measurement was taken. A data scientist has a list containing 10,000 time stamps, sorted in chronological order. Which of the following is closest to the maximum number of values that will need to be examined when performing a binary search for a value in the list?

A 10

B 15 ✔

C 5,000

D 10,000

Explanations:

Answer A

Incorrect. After 10 binary search iterations on a list of length 10,000, there will still be approximately 9 elements remaining.

Answer B

Correct. The binary search algorithm starts at the middle of the list and repeatedly eliminates half the elements until the desired value is found or all the elements have been eliminated. For a list with 10,000 elements, the list will be cut in half a maximum of 13 times, with the total of 14 elements needing to be examined. The list will start with 10,000 elements, then will be reduced to 5,000 elements, then to 2,500 elements, then to 1,250 elements, then to 625 elements, then to 312 elements, then to 156 elements, then to 78 elements, then to 39 elements, then to 19 elements, then to 9 elements, then to 4 elements, then to 2 elements, and then, finally, to 1 element. Of the given values, 15 is closest to 14.

Answer C

Incorrect. While a binary search on a list of length 10,000 will eliminate 5,000 elements on the first iteration, fewer than 5,000 iterations are needed to find the target value or to determine that it is not in the list.

Answer D

Incorrect. A linear search of a list of length 10,000 may examine all 10,000 elements. A binary search requires fewer elements to be examined.

Answer (back)

Correct Answer: D

Practice AP Question

The procedure `BinarySearch (numList, target)` correctly implements a binary search algorithm on the list of numbers `numList`. If `target` is found in the list, the procedure returns `true`, otherwise, it reports `false`. Which of the following conditions must be true in order for the procedure to work properly?

(A) `target` must be a non-negative number
(B) `target` cannot be the first or last element in `numList`
(C) There needs to be an odd number of elements in `numList`
(D) The numbers in `numList` must be sorted

Explanations:

A is incorrect because you *can* search a list for a negative number if you want to.

B is incorrect because binary search (and linear search, for that matter) can successfully find the target value if it's the first or last element in the list.

C is incorrect because binary search is able to work on a list with an even *or* odd number of elements -- just as long as the list is sorted.

D is correct because binary search can *only* be used on *sorted* lists.

Answer (back)

In U5L2, you discovered the following four algorithms had the following efficiencies:

has **linear** efficiency

has **sublinear** efficiency

has **constant** efficiency

has **quadratic** efficiency

A. Which of the four algorithm(s) have a **polynomial efficiency**?

Here are all the ***polynomial*** efficiencies:

- **Constant** efficiency (n^0)
- **Sublinear** efficiency (log n)
- **Linear** efficiency (n^1)
- **Quadratic** efficiency (n^2)
- **Cubic** efficiency (n^3)

So ALL FOUR algorithms have an efficiency that is considered a polynomial efficiency!

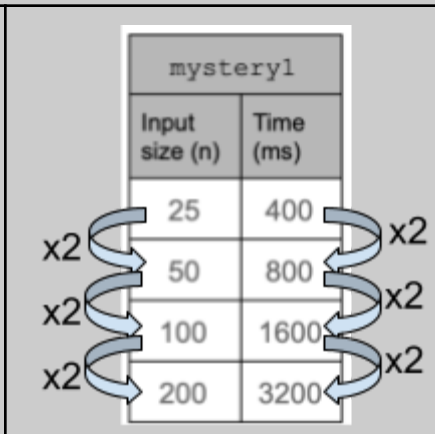
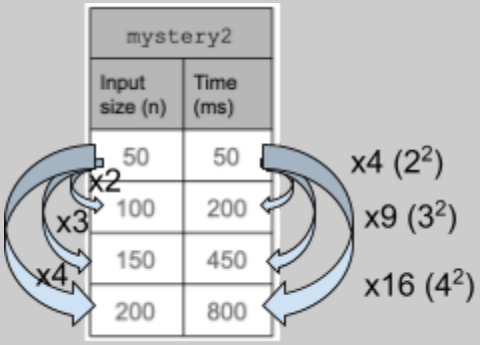
B. Which of the four algorithm(s) "**run in a reasonable time**"?

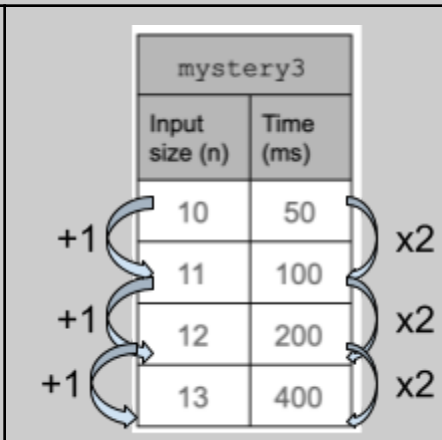
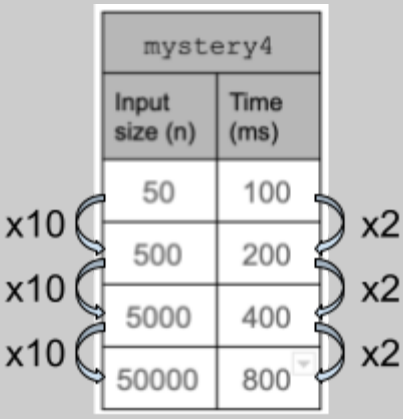
ALL of them "run in reasonable time" because ALL of them have polynomial efficiencies! All polynomial efficiencies are considered "reasonable time" efficiencies.

Answer (back)

In U5L2, you discovered the following algorithm had the following efficiency:  has an **exponential efficiency (2^n)** because adding just **1** to the input size **doubled** the amount of time it takes.	**A.** Is this a **polynomial efficiency**? **No!** Exponential efficiency is **NOT** one of the polynomial efficiencies, which are these: • **Constant** efficiency (n^0) • **Sublinear** efficiency (log n) • **Linear** efficiency (n^1) • **Quadratic** efficiency (n^2) • **Cubic** efficiency (n^3) **B.** Does this algorithm "**run in a reasonable time"**? **No!** Only *polynomial* efficiencies are considered to run in reasonable time; exponential efficiency is *not* polynomial, therefore it does **NOT** run in reasonable time.

Answers (back)

		
Polynomial or **exponential** efficiency?	**Polynomial**	**Polynomial**
If polynomial efficiency, is it linear, sublinear, constant, or quadratic?	**Linear**, because when you multiply input size by 2, time also goes up by 2x	**Quadratic**, because when you multiply input size by 2, time goes up by $2^2 = 4$x; when you multiply input size by 3, time goes up by $3^2 = 9$x; when you multiply input size by 4, time goes up by $4^2 = 16$x
Does it run in **reasonable** time?	**Reasonable**, since all polynomial efficiencies are reasonable	**Reasonable**, since all polynomial efficiencies are reasonable

		
Polynomial or **exponential** time?	**Exponential** Because *adding just one more element doubles the required time*	**Polynomial**
If polynomial time, is it linear, sublinear, constant, or quadratic?	*N/A (not polynomial!)*	**Sublinear**, because when you multiply input size by 10, time goes up by 2x, which means that time is growing *slower than linear*, which makes it sublinear (it this were linear, time would also grow by 10x)

Does it run in **reasonable** time?	**NO! Exponential time is *not* reasonable time!**	**Reasonable**, since all polynomial efficiencies are reasonable

Answer (back)

Correct Answer: D

The table below shows the time a computer system takes to complete a specified task on the customer data of different-sized companies.

Task	Small Company (approximately 100 customers)	Medium Company (approximately 1,000 customers)	Large Company (approximately 10,000 customers)
Backing up data	2 hours	20 hours	200 hours
Deleting entries from data	100 hours	200 hours	300 hours
Searching through data	250 hours	300 hours	350 hours
Sorting data	0.01 hour	1 hour	100 hours

Based on the information in the table, which of the following tasks is likely to take the longest amount of time when scaled up for a very large company of approximately 100,000 customers?

(A) Backing up data

(B) Deleting entries from data

(C) Searching through data

(D) Sorting data

EXPLANATIONS

The correct choice is D because as the input size increases by 10x, the time it takes to sort that data goes up by 100x (this is *quadratic* time); scaled up to 100,000, you could expect sorting to take 100 x 100 = **10,000 hours**; compare to the other answers below, this is the **longest** amount of time.

The other options would take *less time*:

- Choice A shows that as input size increases by 10x, the time required also increases by 10x (this is *linear* time); scaled up to 100,000, you could expect backing up data to take 200 x 10 = 2,000 hours.

- Choice B shows that as input size increases by 10x, the time required increases by 100 (not 100x!); since this time is growing much more slowly than input size, this is *sublinear* time; scaled up to 100,000, you could expect deleting entries to take 400 hours.

- Choice C shows that as input size increases by 10x, the time required increases by 50 (not 50x), which is also *sublinear* time like choice B; scaled up to 100,000, you could expect searching through data to also take 400 hours.

Answer (back)

Correct Answer: D

Lina wrote three different algorithms to search a list containing *n* elements. Assume *n* is a very large integer.

I. Her first algorithm checks each element in the list five times.
II. Her second algorithm checks each element in the list *n* times.
III. Her third algorithm checks the first 20 elements only, regardless of the list length.

Which algorithms run in **reasonable** time?

(A) I only
(B) III only
(C) I and III only
(D) I, II, and III

EXPLANATIONS

Algorithm I checks each of the *n* element 5 times, which means that there are **5n** comparisons per element; if you were to create a table of input size vs. comparisons (and thus time) required, this situation, it might look something like this:

Input size, **n**	Comparisons required, **5n**
5	25
10	50
20	100

You can see here that this is a **linear** relationship because doubling the input would lead to doubling the number of comparisons required (which would double the time). Linear time algorithms all run in polynomial time (which are all reasonable time), so **this algorithm runs in reasonable time**.

Algorithm II checks each of the *n* element *n* times, which means that there are $\mathbf{n^2}$ comparisons per element; if you were to create a table of input size vs. comparisons (and thus time) required, this situation, it might look something like this:

Input size, **n**	Comparisons required, $\mathbf{n^2}$
5	25
10	100
20	400

You can see here that this is a **quadratic** relationship because doubling the input would lead to a 4x increase (2^2) in the number of comparisons required (which would 4x the time). Quadratic time algorithms all run in polynomial time (which are all reasonable time), so **this algorithm runs in reasonable time**.

Algorithm III similarly checks exactly 20 elements every time, which means that there are a **constant 20** comparisons per element; if you were to create a table of input size vs. comparisons (and thus time) required, this situation, it might look something like this:

Input size, **n**	Comparisons required
5	20
10	20
20	20

You can see here that this is a **constant** relationship because changing the input size does *not* impact the number of required comparisons (or time), and the comparisons required stays constant regardless of input size. Constant time algorithms all run in polynomial time (which are all reasonable time), so **this algorithm runs in reasonable time**.

Added note!

The three examples in this question are all examples of reasonable time algorithms; the **"non-reasonable"** time algorithm you would be most likely to see would be ***exponential***, which might look something like a 2^n relationship (or 3^n, or any x^n).

Here is an example of an exponential, non-reasonable time algorithm:

Input size, **n**	Comparisons required, 2^n
5	32
10	1,024
20	1,048,576

Answer (back)

Correct Answer: C

Which of the following programs is most likely to benefit from the use of a heuristic?

(A) A program that determines if a number is even or odd
(B) A program that counts the number of words in a list that are 3 or more letters long
(C) A program that calculates the fastest route to get from school to home when accounting for traffic
(D) A program that searches a sorted list using the binary search algorithm

EXPLANATION:

C is correct because a heuristic is best for problems that *have exact* solutions, but for which:

- The exact solution cannot be determined in reasonable time
- An approximation would be "good enough"

The situation that best fits this type of problem is Choice C, which is similar to the travelling salesman problem in which a fastest path *does* exist, but it would likely take a computer an unreasonable amount of time to check all possible ways to get from points A to B. Finding the fastest driving route is an **optimization** problem that cannot be solved in a reasonable time, and a heuristic is a technique that can find an approximate solution more quickly when exact methods are too slow.

Choice A is an example of a program that is a true/false decision problem and one that can be solved exactly and very quickly in reasonable time by a computer.

Choice B is also an example of a program that solves a problem exactly and can be done very quickly by a computer (we did this in class and it was fast!).

Choice D, searching a list, is also one that can be done exactly and in reasonable time.

Answer (back)

Correct answer: D

Which of the following statements is true?

(A) Every problem is solvable with an algorithm in reasonable time.

(B) Every problem is solvable with an algorithm, but some require an unreasonable amount of time.

(C) Every solvable problem is also a decidable problem.

(D) Some problems can never be solved by a computer using an algorithm.

Explanations:

(A) Every problem is solvable with an algorithm in reasonable time.

This is **false** because not every problem can be solved with an algorithm! And for those problems that *are* solvable with an algorithm, some require *unreasonable* time (such as optimization problems, e.g. the travelling salesman).

(B) Every problem is solvable with an algorithm, but some require an unreasonable amount of time.

This is **false** because the first part of the statement is false: not every problem can be solved with an algorithm!

(C) Every solvable problem is also a decidable problem.

This is **false** because decidable problems require "yes/no" answers -- not every solvable problem requires yes/no answers (Examples A, B and C in the lab are three examples of solvable problems that are not decidable problems). It is important to note that the *converse* of this statement *is* true: "Every decidable problem is also a solvable problem" -- this is true because decidable problems are a *subcategory* of solvable problems.

(D) Some problems can never be solved by a computer using an algorithm.

This is **true**! And these problems are called **unsolvable** problems!

Answer (back)

Correct answer: B

How can it be shown that a decision problem is **undecidable**?

(A) Show that for one instance of the problem, an algorithm can be written that is always capable of providing a correct yes-or-no answer.

(B) Show that for one instance of the problem, no algorithm can be written that is capable of providing a correct yes-or-no answer.

(C) Show that for one instance of the problem, a heuristic is needed to write an algorithm that is capable of providing a correct yes-or-no answer.

(D) Show that for one instance of the problem, an algorithm that runs in unreasonable time can be written that is capable of providing a correct yes-or-no answer.

Explanations:

(A) Incorrect because this statement does not provide enough information to conclude that the problem is undecidable. If it could be shown that an algorithm can be constructed that is always capable of providing a correct yes-or-no answer for all other instances of this problem, then this problem would be decidable.

(B) This is correct because a decidable problem is one for which an algorithm can be constructed to produce a correct output for all inputs. If, for one instance of the problem, this is not possible, then this problem cannot be decidable. Therefore, it must be undecidable.

(C) Incorrect because this statement does not provide enough information to conclude that the problem is undecidable. This states that a heuristic is used for one instance of the problem. If it could be shown that an algorithm can be constructed that is always capable of providing a correct yes-or-no answer for all other instances of this problem, then this problem would be decidable

(D) Incorrect because this statement does not provide enough information to conclude that the problem is undecidable. This states that for one instance of the problem, the algorithm runs in an unreasonable amount of time. If it could be shown that an algorithm can be constructed that is always capable of providing a correct yes-or-no answer for all other instances of this problem, whether the algorithms run in a reasonable or unreasonable amount of time, then this problem would be decidable.

Answer (back)

Correct Answer: C

A team of programmers is designing software. One portion of the project presents a problem for which there is not an obvious solution. After some research, the team determines that the problem is undecidable. Which of the following best explains the consequence of the problem being undecidable?

A. The problem can be solved algorithmically, but it will require an unreasonably long amount of time.

B. The problem can be solved algorithmically, but it will require an unreasonably large amount of data storage.

C. There is no possible algorithm that can be used to solve all instances of the problem. ✔

D. There are several different possible algorithms that can solve the problem, but there is controversy about which is the most efficient.

Explanations:

Answer A

Incorrect. There is no possible algorithm for solving an undecidable problem.

Answer B

Incorrect. There is no possible algorithm for solving an undecidable problem.

Answer C

Correct. An undecidable problem is one for which no algorithm can be constructed that is always capable of providing a correct yes-or-no answer. There is no algorithmic solution that can solve all instances of the problem.

Answer D

Incorrect. There is no possible algorithm for solving an undecidable problem.

Answer (back)

Correct Answer: B

A student wants to determine whether a certain problem is undecidable. Which of the following demonstrate that the problem is undecidable?

A. Show that for one instance of the problem, an algorithm can be written that is always capable of providing a correct yes-or-no answer.

B. Show that for one instance of the problem, no algorithm can be written that is capable of providing a correct yes-or-no answer.

C. Show that for one instance of the problem, a heuristic is needed to write an algorithm that is capable of providing a correct yes-or-no answer.

D. Show that for one instance of the problem, an algorithm that runs in unreasonable time can be written that is capable of providing a correct yes-or-no answer.

Explanations:

Answer A

Incorrect. This statement does not provide enough information to conclude that the problem is undecidable. If it could be shown that an algorithm can be constructed that is always capable of providing a correct yes-or-no answer for all other instances of this problem, then this problem would be decidable.

Answer B

Correct. A decidable problem is one for which an algorithm can be constructed to produce a correct output for all inputs. If, for one instance of the problem, this is not possible, then this problem cannot be decidable. Therefore, it must be undecidable.

Answer C

Incorrect. This statement does not provide enough information to conclude that the problem is undecidable. This states that a heuristic is used for one instance of the problem. If it could be shown that an algorithm can be constructed that is always capable of providing a correct yes-or-no answer for all other instances of this problem, then this problem would be decidable.

Answer D

Incorrect. This statement does not provide enough information to conclude that the problem is undecidable. This states that for one instance of the problem, the algorithm runs in an unreasonable amount of time. If it could be shown that an algorithm can be constructed that is always capable of providing a correct yes-or-no answer for all other instances of this problem, whether the algorithms run in a reasonable or unreasonable amount of time, then this problem would be decidable.

Answer (back)

Correct Answer: D

Which of the following best explains how algorithms that run on a computer can be used to solve problems?

A. All problems can be solved with an algorithm that runs in a reasonable amount of time.

B. All problems can be solved with an algorithm, but some algorithms might need a heuristic to run in a reasonable amount of time.

C. All problems can be solved with an algorithm, but some algorithms might run in an unreasonable amount of time.

D. Some problems cannot be solved by an algorithm.

Explanations:

Answer A

Incorrect. Not all algorithms run in a reasonable amount of time, and undecidable problems cannot be solved with an algorithm.

Answer B

Incorrect. A heuristic can be used to approximate a solution to some problems in order to run in a reasonable amount of time; however, undecidable problems cannot be solved with an algorithm.

Answer C

Incorrect. Some problems can be solved with an algorithm that runs in an unreasonable amount of time; however, undecidable problems cannot be solved with an algorithm.

Answer D

Correct. Undecidable problems cannot be solved with an algorithm.

Answer (back)

Correct Answer: D

Which of the following statements is true?

A. Every problem can be solved with an algorithm for all possible inputs, in a reasonable amount of time, using a modern computer.

B. Every problem can be solved with an algorithm for all possible inputs, but some will take more than 100 years, even with the fastest possible computer.

C. Every problem can be solved with an algorithm for all possible inputs, but some of these algorithms have not been discovered yet.

D. There exist problems that no algorithm will ever be able to solve for all possible inputs.

Explanations:

Answer A

This option is incorrect. Some problems, such as attempting to break a strongly encrypted message, cannot be solved on modern computers in a reasonable amount of time.

Answer B

This option is incorrect. Although some problems will take a very long time to solve even with the fastest computers, some problems cannot be solved with an algorithm for all possible inputs.

Answer C

This option is incorrect. Some problems cannot be solved with an algorithm for all possible inputs.

Answer D

This option is correct. Some problems, such as determining if any program will eventually stop, cannot be solved by an algorithm.

Answer (back)

Suppose the top sequential part (on the main computer) takes **20** seconds, *each* computer working in parallel on an individual letter finishes processing its sublist in **10** seconds, and the bottom sequential part (back on the main computer) takes **15** seconds. What would be the total **time** required for this solution to the problem of calculating the average word length of a list of 100,000 words?	20 + 10 + 15= **45** seconds Note that you only have to add the 10 ***once*** since all 26 computers are running their parts **in PARALLEL** and so are all doing their tasks during the *same* 10 seconds!

Answer (back)

Correct Answer: D

Which of the following best describes the ability of parallel computing solutions to improve efficiency?

A. Any problem that can be solved sequentially can be solved using a parallel solution in approximately half the time.

B. Any solution can be broken down into smaller and smaller parallel portions, making the improvement in efficiency theoretically limitless as long as there are enough processors available.

C. The efficiency of parallel computing solutions is rarely improved over the efficiency of sequential computing solutions.

D. The efficiency of a solution that can be broken down into parallel portions is still limited by a sequential portion.

Explanations:

Answer A

Incorrect. Some sequential solutions cannot be easily broken into two parallel portions that require an equal workload. For example, some steps of an algorithm may not be able to be executed until a previous step is completed.

Answer B

Incorrect. The efficiency of a parallel solution is limited by a sequential portion. Eventually, adding parallel portions will no longer meaningfully increase efficiency.

Answer C

Incorrect. Solutions that use parallel computing can scale more effectively than solutions that use sequential computing, sometimes providing significant improvements in efficiency.

Answer D

Correct. Parallel computing solutions consist of a parallel portion and a sequential portion. When increasing the use of parallel computing in a solution, the efficiency of the solution is limited by the sequential portion. This means that at some point, adding parallel portions will no longer meaningfully increase efficiency.

Answer (back)

Correct Answer: B

A certain computer has two identical processors that are able to run in parallel. The table below indicates the amount of time it takes each processor to execute each of two processes. Assume that neither process is dependent on the other.

Process	Execution Time on Either Processor
P	30 seconds
Q	45 seconds

Which of the following best approximates the difference in execution time between running the two processes in parallel instead of running them one after the other on a single processor?

A 15 seconds

B 30 seconds ✔

C 45 seconds

D 75 seconds

Explanations:

Answer A

Incorrect. This is the difference between the amount of time it takes processes P and Q to execute, not the amount of time saved by running the processes in parallel.

Answer B

Correct. Running the two processes one after the other on a single processor requires $30 + 45 = 75$ seconds. Running the two processes in parallel requires 45 seconds because if both processes start at the same time on different processors, process P will finish in 30 seconds and process Q will finish 15 seconds later at 45 seconds. Therefore, the amount of time saved by running the two processes in parallel is $75 - 45 = 30$ seconds.

Answer C

Incorrect. This is the amount of time it takes to run the two processes in parallel, not the amount of time saved by running the two processes in parallel.

Answer D

Incorrect. This is the amount of time it takes to run the two processes one after the other on a single processor, not the amount of time saved by running the two processes in parallel.

Answer (back)

Correct Answer: C

A computer has two processors that are able to run in parallel. The table below indicat amount of time it takes either processor to execute four different processes. Assume that none of the processes is dependent on any of the other processes.

Process	Execution Time
W	20 seconds
X	30 seconds
Y	45 seconds
Z	50 seconds

A program is used to assign processes to each of the processors. Which of the following describes how the program should assign the four processes to optimize execution time?

A. Processes W and X should be assigned to one processor, and processes Y and Z should be assigned to the other processor.

B. Processes W and Y should be assigned to one processor, and processes X and Z should be assigned to the other processor.

C. Processes W and Z should be assigned to one processor, and processes X and Y should be assigned to the other processor. ✔

D. Process Z should be assigned to one processor, and processes W, X, and Y should be assigned to the other processor.

Explanations:

Answer A

Incorrect. Using this solution, all four processes will finish executing in 95 seconds. Processes W and X will finish executing in 50 seconds on one processor, and processes Y and Z will finish executing in 95 seconds on the other processor. It is possible to assign the processors in such a way that all four processes will finish executing in 75 seconds.

Answer B

Incorrect. Using this solution, all four processes will finish executing in 80 seconds. Processes W and Y will finish executing in 65 seconds on one processor, and processes X and Z will finish executing in 80 seconds on the other processor. It is possible to assign the processors in such a way that all four processes will finish executing in 75 seconds.

Answer C

Correct. Execution time is optimized when the workload of the two processors is as close to equal as possible, so that one processor does not finish too early and have to wait for the other processor to finish. The closest-to-equal workloads are achieved by assigning processes W and Z to one processor (taking 70 seconds) and assigning processes X and Y to the other processor (taking 75 seconds). Using this solution, all four processes will finish executing in 75 seconds.

Answer D

Incorrect. Using this solution, all four processes will finish executing in 95 seconds. Process Z will finish executing in 50 seconds on one processor, and processes W, X, and Y will finish executing in 95 seconds on the other processor. It is possible to assign the processors in such a way that all four processes will finish executing in 75 seconds.

Answer (back)

Correct Answer: C

A certain computer has two identical processors that are able to run in parallel. Each processor can run only one process at a time, and each process must be executed on a single processor. The following table indicates the amount of time it takes to execute each of three processes on a single processor. Assume that none of the processes are dependent on any of the other processes.

Process	Execution Time on Either Processor
X	60 seconds
Y	30 seconds
Z	50 seconds

Which of the following best approximates the minimum possible time to execute all three processes when the two processors are run in parallel?

A 60 seconds

B 70 seconds

C 80 seconds ✔

D 90 seconds

Explanation:
Since you have only two processors and three Processes to complete, you can minimize the time by placing one Process on one of the processors and the other two Processes on the other processor, and run them in parallel.

To figure out minimum possible time, find the two Processes that sum up to the smallest value:
X + Y = 90 seconds
X + Z = 110 seconds
Y + Z = 80 seconds

Processes Y + Z = 80 seconds is the *smallest*, so run these two Processes in sequence on one of the processors.

Place the other process, X, which takes 60 seconds on the other processor. If both processes run in parallel, the processor with Y + Z will be done in 80 seconds and the second processor will be done in 60 seconds, and since they are running simultaneously, the overall process for all three to get complete is **80 seconds** (since the 60 second Process is happening during the same time that the 50 + 30 second processes are running!)

Answer (back)

Correct Answers: B & D

Supercomputers are machines that, as of 2019, could compute quadrillions of operations per second. These machines are used for a variety of things, including simulating earthquakes, nuclear reactions, and the big bang. Supercomputer simulations can replace the testing of actual nuclear arms. What are two advantages of such simulations?

Select two.

A. Simulations can incorporate all of the complexities of nuclear reactions.

B. Simulations are cheaper than building multiple designs and testing each other in the real world.

C. Simulations produce results that are more accurate than real world experiments.

D. Simulations come without the environmental associated risks with nuclear tests.

Correct answer is B, D:

Computer simulations are **cheaper** than building actual nuclear devices (**SAVE MONEY**), and they **come without the dangers** associated with detonating such devices (**SAFER** and **MORE ETHICAL**).

A is NOT correct because simulations generally do NOT capture all the complexities of the system being simulated (this is one of the drawbacks of simulations)

C is NOT correct because simulations often leave out details (to make them possible to run on computers) that result in less precise results than actually doing an experiment in the real world (but good enough results that can be used to make predictions).

Answer (back)

Correct Answer: B

Before airplane pilots are allowed to fly actual planes, they spend countless hours in a flight simulator. Which of the following is **not** a beneficial reason to use the computer simulation?

A. The simulation can create extreme scenarios to test a pilot's knowledge and ability to work under extreme pressure.

B. Pilots can use the simulation to determine if this is the career for them.

C. If, during the process, the pilot makes a mistake, it will not destroy an actual plane or kill innocent people.

D. Pilots can get familiar with the cockpit controls on a variety of different plans before flying the real plane.

Correct answer is B. While a simulation allows you to get the feeling of flying a plane, the pilot may simply want to do it as a hobby. Also, working in a simulation does not provide the full experience of being a pilot - traveling often and being responsible for other's lives. The stress of being a pilot is not fully understood by a person only using a flight simulator!

Answer (back)

Correct Answer: D

A plane manufacturer uses simulation software during the design process for a new plane. Which of the following are reasons to use simulation software for this purpose?

I. Using simulation software can save the company money by helping to compare designs early in the process.

II. Using simulation software can help to identify safety concerns with the plane by enabling the testing of the plane's components under various situations.

III. The manufacturer can present simulations to airlines to demonstrate different design options, and develop those options for airlines.

(A) I and II only

(B) I and III only

(C) II and III only

(D) I, II, and III

Correct answer is D

(A) **This option is incorrect.** While statements I and II are correct, statement III is also correct because the results of computer simulations may be useful in presenting design options to customers.

(B) **This option is incorrect.** While statements I and III are correct, statement II is also correct because simulations can be made that model components with the purpose of improving safety.

(C) **This option is incorrect.** While statements II and III are correct, statement I is also correct because software simulations can usually be built before an actual plane.

(D) **This option is correct.** Statement I is correct because software simulations can usually be built before a real plane. Statement II is correct because simulations can be made that model components in various circumstances. Statement III is correct because the results of computer simulations may be useful in presenting design options to customers.

Answer (back)

Correct Answer: D

Government officials and scientists are designing a simulation to examine how much a plan to convert entirely to electric cars will have on reducing climate change. The simulation uses a model that includes input variables for the plan such as expected growth in car demand, efficiency of electric cars, and consumer sentiment around electric cars. The simulation can then be run many times using different values for the input variables to represent different outcomes on climate change. One challenge they are facing is that the complex simulation takes a very long time to run. They plan to update the model by removing some input variables they consider less important than others. Of the following, which is the most likely effect the updated model will have on the simulation's runtime?

(A) The updated model is unlikely to affect the runtime of the simulation because the simulation will remain very complex, regardless of how many input variables the model uses.

(B) The updated model is likely to increase the runtime of the simulation because refining a simulation model requires increased computational time.

(C) The updated model is likely to decrease the runtime of the simulation because it would enable the simulation to be run on a faster computer.

(D) The updated model is likely to decrease the runtime of the simulation because the time required for simulations generally depends on the complexity of the model used.

Explanation

(A) Incorrect because removing details from the model can often speed up the runtime of a simulation because it reduces computational complexity, thus requiring less time.

(B) Incorrect because removing details from the model can often speed up the runtime of a simulation because it reduces computational complexity, thus requiring less time.

(C) Incorrect because, although the model is likely to decrease the runtime of the simulation, it isn't for the stated reason; in fact, a faster computer should be able to handle a model with *more* input variables, not fewer.

(D) Correct because the time required for a simulation to run is impacted by the level of detail used in the model. Generally, a simulation based on a less detailed model will require less time to run.

Answer (back)

Correct Answer: C

A researcher is interested in predicting the number of deaths that will occur in a particular community. She created a computer model that uses data from the past twenty years, including number of residents and the number of deaths. The model predicted that there would be 500 deaths last year, but the actual number of deaths last year was only 456. Which of the following strategies is LEAST likely to provide a more accurate prediction?

(A) Gathering data for additional years to try to identify patterns in death rates

(B) Refining the computer simulation to reflect new understandings about causes of death

(C) Removing some of the details from the model and simulation so that computations can be performed more quickly

(D) Adding additional parameters to the simulation, such as the ages of residents.

Explanations:

(A) This option is incorrect. Gathering additional data may help the simulation to provide a more accurate prediction.

(B) This option is incorrect. Refining the model used in the simulation may help provide a more accurate prediction.

(C) This option is **correct**. Removing details from the model may help it run more quickly, but is unlikely to provide more accurate results (in fact, it would likely have the opposite effect and provide less accurate results).

(D) This option is incorrect. Adding more parameters to the simulation may help the simulation provide a more accurate prediction.

Answer (back)

Correct Answer: C

Meteorologists are tracking a hurricane approaching the East coast of the United States. Which of the following is a limitation of using computer simulations to model this real-world event?

(A) Computer simulations require large amounts of processing speed and memory.

(B) Computer simulations can only be designed and used after the real-world event, in this case a hurricane, has occurred.

(C) Computer simulations often approximate or leave out some details about the real-world event being simulated that may end up being important (such as rainfall, wind, or currents).

(D) Computer simulations require that all input parameters are never changed.

Explanation

Choice 'C' is the correct answer.

In computer simulations, in order to decrease the run time, we make certain generalization or assumptions to limit the number of variables. For example, if you are tracking a hurricane, you assume that when the hurricane reaches land the hurricane will travel more slowly. You assume that, when over the ocean, the hurricane will gain speed and strength. Factors that are variables would be wind patterns and precipitation in specific areas.

Explanation of Distractors:

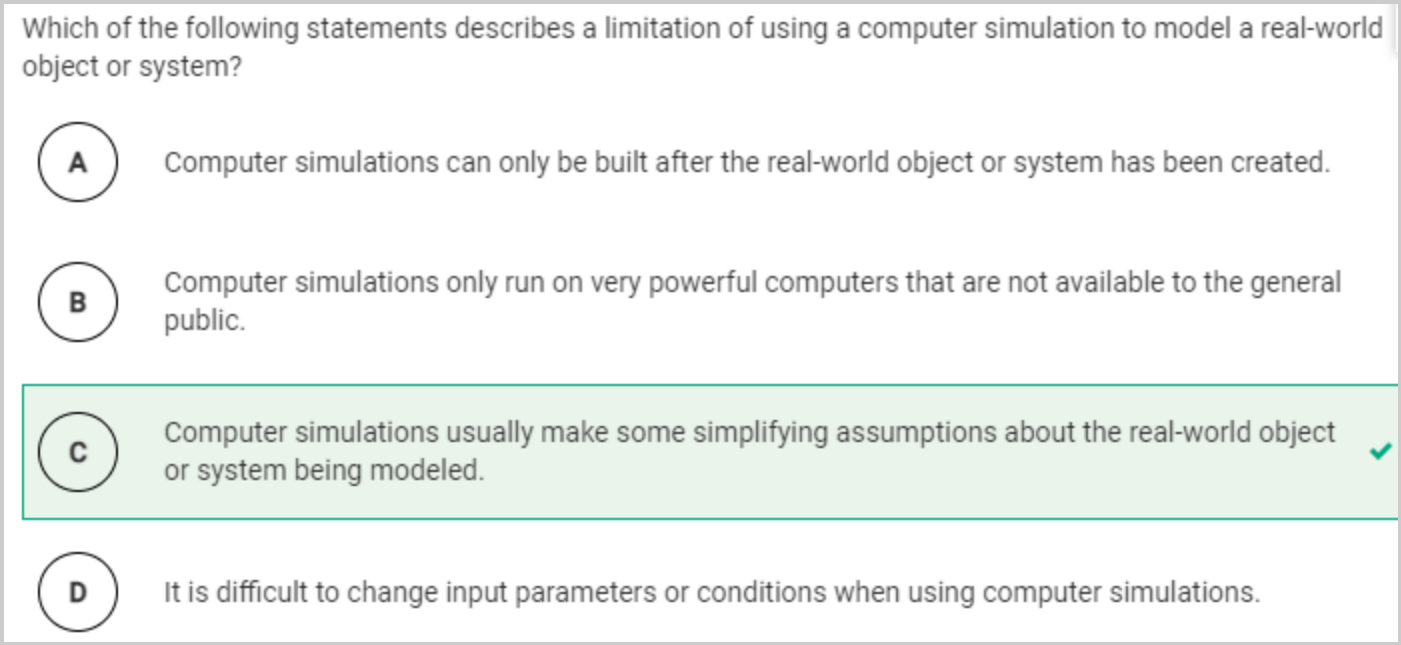
Choice 'A' is incorrect. Some simulations require a lot of memory and processing power, however, most computers today can handle the requirements. You might have even played a simulation game before!

Choice 'B' is incorrect. The purpose of computer simulations is to anticipate and predict how certain events will unfold ***before*** they happen. For example, if you are tracking a hurricane, a computer simulation will predict the place where the hurricane will land. There is no point in predicting ***after*** the hurricane has landed.

Choice 'D' is incorrect. In order to predict real-world events, some parameters must be assumed. However, simulations allow for users to test possible scenarios by changing some of the input parameters. For example, you may change the direction and strength of air currents to predict different paths they may push the hurricane. Another example, if you are trying to predict population growth of a frog, you may change the parameter for percent deforestation to see what impact deforestation has on frog population.

Answer (back)

Correct Answer: C



Explanations:

Choice A is incorrect because it is **not true**; simulations are in fact often used to design and test things (like NASA rovers) before they are built -- saving money on a project that may not be viable.

Choice B is incorrect because it is **not true**; simulations can be simple, like coin flipping apps or Snap programs designed to simulate disease spread. You don't necessarily need powerful computers to run simulations (although the most powerful simulations may require powerful computers).

Choice C is correct because simulations often make simplifying assumptions about the subject being modeled, which can lead to certain crucial details being missed or left out inadvertently.

Choice D is incorrect because it is **not true**; you can add or remove parameters as needed, just like you did in the lab on simulations and disease spread!

Answer (back)

Correct Answer: D

In which of the following scenarios would a simulation be the LEAST beneficial?

A. An engineering company wants to test whether a change to a car design will negatively affect fuel efficiency.

B. An insurance company wants to study the effect of cold weather patterns on health-care costs.

C. A manufacturing company wants to determine whether using robots in its facility will increase productivity.

D. A retail company wants to determine the most popular item that was sold on the company's Web site last month.

Explanations:

Answer A

Incorrect. This is a good candidate for simulation. It will be less expensive to simulate the results rather than spending the money and using the time to manufacture and test the new design.

Answer B

Incorrect. This is a good candidate for simulation. Specific types of weather patterns are impractical to track.

Answer C

Incorrect. This is a good candidate for simulation. It will be less expensive to simulate the results rather than spending the money and using the time to test actual robots in the manufacturing facility.

Answer D

Correct. The quantity of each item sold on the Web site can be counted. Therefore, the most popular item can be determined without the use of a simulation.

Answer (back)

Correct Answer: B

A city planner is using simulation software to study crowd flow out of a large arena after an event has ended. The arena is located in an urban city. Which of the following best describes a limitation of using a simulation for this purpose?

A. The model used by the simulation software cannot be modified once the simulation has been used.

B. The model used by the simulation software often omits details so that it is easier to implement. ✔

C. Running a simulation requires more time to generate data from trials than observing the crowd exiting the arena at various events.

D. Running a simulation requires a large number of observations to be collected before it can be used to explore a problem.

Explanations:

Answer A

Incorrect. The model used by a simulation can be modified, both before and after running the simulation. In fact, one of the benefits of using a simulation is the ease of modification.

Answer B

Correct. Simulations are limited by the model that is used. There may be many reasons for using a simplified model, including ease of implementation.

Answer C

Incorrect. A simulation is often used when a real-world scenario is too impractical to test. A benefit of using a simulation is that it allows for rapid testing.

Answer D

Incorrect. In order to validate the predicted results of a simulation, observed data may be needed for comparison purposes. However, running a simulation does not require observations before it can be used.

Answer (back)

Correct Answer: A

For which of the following problems is using a simulation LEAST likely to be beneficial?

A Determining the longest word in a textbook

B Minimizing the customer wait times at a bank

C Predicting the outcomes of weather patterns

D Studying the formation of a galaxy

Explanations:

Answer A

Correct. This calculation does not need a simulation. Using standard algorithms that find a maximum value from a list of values would be more appropriate.

Answer B

Incorrect. The cost and time to test various scenarios to see which combination of changes would reduce wait times could be prohibitive. Thus, this scenario is well suited for a using a simulation.

Answer C

Incorrect. A particular weather pattern may be difficult to test due to its infrequency and unpredictability, and simulations may be the only safe way to study some weather patterns. Thus, this scenario is well suited for using a simulation.

Answer D

Incorrect. Galaxy formation cannot be practically studied due to extreme time and distance constraints. Thus, this scenario is well suited for using a simulation.

Answer (back)

Correct Answers: C & D (must have both!)

In which of the following scenarios is using a simulation more beneficial than performing a calculation?

Select two answers.

A Determining the average grade of the students in a certain class

B Keeping track of the high score in a game

C Investigating ways to reduce the amount of trash in the ocean

D Studying the effect of a genetic change in a population

Explanations:

Answer A

Incorrect. Determining the average grade can be computed using mathematical calculations. In this scenario, it would not be beneficial to use a simulation.

Answer B

Incorrect. Determining whether a score exceeds an existing high score can be determined using a comparison. In this scenario, it would not be beneficial to use a simulation.

Answer C

Correct. Investigating ways to reduce the amount of trash in the ocean is too impractical for an experiment. In this scenario, it would be more beneficial to use a simulation.

Answer D

Correct. Studying the effect of a genetic change in a population is too impractical for an experiment. In this scenario, it would be more beneficial to use a simulation.

Answer (back)

Correct Answer: D

A new bank plans to make customer convenience a priority by minimizing the amount of time a customer waits in line. The bank is considering two options: a single line where the customer at the front waits for the next available teller, or separate lines for each teller. The bank decides to use a computer simulation of these two options to determine the average wait time for customers.

Which of the following is NOT true about the bank's plan?

A. The bank can incorporate other factors, such as the number of tellers, in the simulation.

B. The bank can use the simulation to investigate these two options without causing inconvenience for customers.

C. The bank may consider new alternatives based on the simulation results.

D. The simulation will not produce usable results because actual customer data are not available. ✔

Explanations:

Answer A

This option is incorrect. The bank can add parameters, such as the number of tellers, to the simulation.

Answer B

This option is incorrect. By writing a computer simulation, the bank can investigate options without involving the customers.

Answer C

This option is incorrect. By running a computer simulation with additional or different features, the bank can decide upon other options to meet their goal of minimizing customer wait time.

Answer D

This option is correct. Although actual customer data is not available in the bank's simulation, it can give information that relates the average customer wait time for each option.

Answer (back)

Correct Answer: D

The transportation department plans to build a new high-speed train route between two cities. The transportation department wants to implement a simulation of the train before any construction begins on this project. Which of the following statements is true about the use of a simulation for this project?

A. A simulation cannot be used to test the train under varying weather conditions.

B. Implementing a simulation is likely to increase the overall costs associated with the project.

C. Other high-speed train routes have been constructed in other locations, so a simulation is not needed.

D. Using a simulation may expose potential safety issues that can be corrected before construction begins. ✔

Explanations:

Answer A

Incorrect. Simulations facilitate the formation and refinement of hypotheses. By varying the input data to the simulation, the operation of the train can be tested under a variety of circumstances.

Answer B

Incorrect. Simulations generally lower overall costs by finding design flaws or other issues before construction begins. Fixing these issues after construction has started is much more costly.

Answer C

Incorrect. While there may be some lessons learned from the construction of the other train routes, there may be issues unique to this location. The simulation for this project can use a model that is specifically designed for this particular train route.

Answer D

Correct. Simulations are most useful when real-world events are impractical for experiments. Discovering and correcting potential safety issues using a simulation before the train route is constructed is much safer and less expensive than discovering them after the train route has been constructed.

Answer (back)

Correct Answers: A & B (must have both!)

Which of the following is a benefit of using a simulation instead of an experiment to make a prediction?

Select two answers.

A A simulation allows investigation of a phenomenon without the real-world limitations on time, safety, or budget.

B A simulation can be used to model real-world events that are impractical for experiments.

C A simulation always produces the same output, so its results can be verified.

D A simulation produces results that are more accurate than experimental results.

Explanations:

Answer A

Correct. Simulations often mimic real-world events with the purpose of drawing inferences, allowing investigation of a phenomenon without having to consider real-world limitations.

Answer B

Correct. Simulations are most useful when real-world events are impractical for experiments.

Answer C

Incorrect. Random number generators can be used to simulate the variability that exists in the world, so results are not always the same. Validation of the results is usually accomplished by comparing the outcomes with those based on experimentation.

Answer D

Incorrect. The process of developing an abstract simulation involves removing specific details or simplifying functionality.

Answer (back)

Correct Answer: D

A car manufacturer uses simulation software during the design process for a new car. Which of the following are reasons to use simulation software in this context?

I. Using simulation software can save the company money by helping to compare designs early in the process, before prototype cars are built.

II. Using simulation software can help to identify safety issues by providing data about how different mechanical components will interact in a wide variety of situations.

III. The manufacturer can present simulation software to customers to demonstrate different design possibilities.

A I and II only

B I and III only

C II and III only

D I, II, and III ✔

Explanations:

Answer A

This option is incorrect. While statements I and II are correct, statement III is also correct because the results of computer simulations may be useful in presenting design possibilities to customers.

Answer B

This option is incorrect. While statements I and III are correct, statement II is also correct because simulations can be made that model components and their interactions.

Answer C

This option is incorrect. While statements II and III are correct, statement I is also correct because software simulations can usually be built before a prototype car.

Answer D

This option is correct. Statement I is correct because software simulations can usually be built before a prototype car. Statement II is correct because simulations can be made that model components and their interactions. Statement III is correct because the results of computer simulations may be useful in presenting design possibilities to customers.

Answer (back)

Correct answers: B and D (must have both!)

In an upcoming election, 80 percent of New York voters are expected to vote for Mike Jones in a local election. The computer program below is intended to simulate the result of the election with `n` voters, and display the number of votes received by Mike Jones.

```
Line 1:  voteTotal ← 0
Line 2:  REPEAT n TIMES
Line 3:  {
Line 4:     IF (<MISSING CONDITION>)
Line 5:     {
Line 6:        voteTotal ← voteTotal + 1
Line 7:     }
Line 8:   }
Line 9:   DISPLAY (voteTotal
```

Which two of the following can be used to replace `<MISSING CONDITION>` in `Line 4` so that the program works as intended?

Select two.

(A) `RANDOM (1, 5) = 4`

(B) `RANDOM (1, 5) > 1`

(C) `RANDOM (1, 10) = 8`

(D) `RANDOM (1, 10) ≤ 8`

EXPLANATIONS

- Recall RANDOM (1, 5) produces a random number between 1 and 5 (so either 1, 2, 3, 4, or 5)
- RANDOM (1, 10) produces a random number between 1 and 10 (1, 2, 3, 4,5, 6, 7, 8, 9, or 10)

A is incorrect because the likelihood of RANDOM (1, 5) producing 4 is 1/5, which is 20% (not 80%)

B is correct because the likelihood of RANDOM (1, 5) producing a number > 1 is 4/5 which is 80%

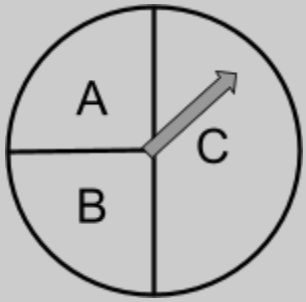
C is incorrect because the likelihood of RANDOM (1, 10) producing exactly 8 is 1/8, which is 12.5% (not 80%)

D is correct because the likelihood of RANDOM (1, 10) producing a number ≤ 8 is 8/10 (1, 2, 3, 4, 5, 6, 7, 8) which is 80%

Answer (back)

Correct answer is D

A spinner is divided into three sections, A, B, and C, as shown below. The area labeled C is twice as large as areas A and B, which are equal in size.



Which of the following Spinner procedures correctly simulates the spinner?

(A)

```
PROCEDURE spinner
    spin ← RANDOM 1, 3
    IF spin = 1
        RETURN "A"
    IF spin = 2
        RETURN "B"
    RETURN "C"
```

(B)

```
PROCEDURE spinner
    spin ← RANDOM 1, 3
    IF spin = 1
        RETURN "C"
    IF spin = 2
        RETURN "B"
    RETURN "A"
```

(C)

```
PROCEDURE spinner
    spin ← RANDOM 1, 4
    IF spin = 1
        RETURN "A"
    IF spin = 2
        RETURN "C"
    RETURN "B"
```

(D)

```
PROCEDURE spinner
    spin ← RANDOM 1, 4
    IF spin = 1
        RETURN "A"
    IF spin = 2
        RETURN "B"
    RETURN "C"
```

Correct answer is D because area C takes up ½ of the space (equivalent to 2 parts out of 4), and procedure D gets a random number between 1 and 4, and if it's a 1 -- which happens ¼ of the time -- then returns A, if it's a 2 -- which also happens ¼ of the time -- then returns B, and if neither of those are true (in other words, the number was a 3 or 4), then it returns C, which happens ½ of the time. The ¼, ¼, and ½ likelihoods for A, B, and C match the area model of the spinner!

IMPORTANT REMINDER! As soon as a "RETURN" is reached in a procedure, it ends, and all other code below it in the procedure is *skipped*!

Answer (back)

Correct Answer: C

A social media app stores the following data for each post that is made as separate records in a data set:

- The full text of the post
- The date and time the post was made
- The location of the user when the post was made

Assuming that a company has access to this information, which of the following could be determined using the social media post data described above?

I. The number of posts that were taken at a particular location

II. The post made in the last year that had the most number of words in its text

III. The user who made the most recent post

(A) I only

(B) II only

(C) I and II only

(D) I, II, and III

Correct answer:	C

Explanation

Option I is CORRECT because it would be possible to count up the number of records in the data set (how many rows) that had a specific geographic location in the record; this is similar to when you determined how many video games were published by a particular publisher (the "count").

Option II is CORRECT because it would be possible to count up the number of words in each post, since the full text of each post is known, and find the post that had the most words.

Option III is INCORRECT because *which user* made each post is **not** part of this data set, as described; while each record has location data, no identifying information about the user is included. (Note that *if* the user associated with each post *was* included in this data set -- which it's not -- then this option would be correct because it would be possible to determine which user made the most recent post, since date and time of the post is included in the data set).

Answer (back)

Correct Answers: B & D (must have both)

A grocery store maintains a single database that contains records with the following information about each item for sale in the grocery store.

- Grocery UPC code (the number on the barcode: 

)
- Type of item (meat, bakery, household goods, cereal/breads, milk/juices, produce)
- Selling price (in dollars)
- Size (weight, in ounces)
- Quantity available in the store

Using only the data in the database, which two of the following can be determined?

Select two.

(A) Which type of grocery (meat, bakery, etc.) sell the most frequently

(B) Which grocery item has the lowest price per ounce of weight

(C) Which specific cereal brand is the most popular

(D) Which grocery items are not currently in stock

Correct answers:	**B** and **D**

Explanation

A is INCORRECT because *how many of each item sold* is ***not*** a piece of data in this data set!

B is CORRECT because it is possible to determine price per ounce of weight using the "selling price" and "size" values (these two numbers could be divided to yield the "price per ounce of weight" and then choosing the item with the lowest)

C is INCORRECT because "most popular" means that you are looking for "which cereal is sold the most," but similar to choice A, how many of each item is sold is *not* a part of this data set. Also, which specific cereal a particular bar code item refers to is *also* not included in this data set!

D is CORRECT because it is possible to determine which are "not in stock" by filtering the data set of which records have "0" in the "Quantity available in the store"

Answer (back)

Correct Answer: C

Computers are often used to search through large data sets to find useful trends and patterns in the data. Which of the following tasks is NOT an example where searching for trends and patterns produces useful information?

(A) A bank analyzing deposit and withdrawal data in customer accounts to identify potential fraudulent account activity

(B) An search engine analyzing users' search history to target advertisements and new products the user may be interested in purchasing

(C) A high school student analyzes her test scores to determine which tests she performed the best on.

(D) A video game publisher analyzing user review data for all released video games to determine which types of games users are most interested in purchasing.

Correct answer:	C

Explanation

A is INCORRECT because this **IS** an example of searching a large data set for trends and patterns to produce useful information.

B is INCORRECT because this **IS** an example of searching a large data set for trends and patterns to produce useful information.

C is CORRECT -- this is **NOT** an example of searching a large data set for trends and patterns to produce useful information for two reasons:

- A high school student's set of individual test scores is not a "large data set"
- Determining which tests she did best on is *not* considered a "pattern" or a "trend"

D is INCORRECT because this IS an example of searching a large data set for trends and patterns to produce useful information.

Answer (back)

Correct Answer: C

An online music streaming service contains metadata for each song. The metadata is intended to help a search feature find songs that users are interested in listening to. Which of the following is LEAST likely to be contained in the metadata of each song?

(A) The title and artist of the song

(B) The date the song was added to the streaming service

(C) The audio data of the song itself

(D) The genre of the song (e.g., dance, rock, pop, rap, etc.)

Correct answer:	C

Explanation

A is INCORRECT because title and artist are examples of descriptive metadata that would help support the search feature (for example, if the user wanted to search for songs with a particular title or by a particular artist).

B is INCORRECT because the date the song was added to the streaming service is descriptive metadata that would help support the search feature (for example, if a user wanted to search for "New Releases").

C is CORRECT because metadata for a song would typically be used to provide *descriptive information about the song*. The audio data of the song itself is the *data* of the song file, not metadata.

D is INCORRECT because the genre of the song is descriptive metadata that would help support the search feature (for example, if a user wanted to search for popular dance songs).

Answer (back)

Correct Answer: B

A certain social media website allows users to post messages and to comment on other messages that have been posted. When a user posts a message, the message itself is considered *data*. In addition to the data, the site stores the following *metadata*:

- The time the message was posted
- The name of the user who posted the message
- The names of any users who comment on the message and the times the comments were made

For which of the following would it be more useful to analyze the *data* instead of the *metadata*?

(A) To determine the users who post messages most frequently

(B) To determine the topics that many users are posting about

(C) To determine the time of day that the site is most active

(D) To determine which posts have received the greatest number of comments

Correct answer:	**B**

Explanation

A is INCORRECT because to determine the users who post most frequently, it would be best to analyze "The name of the user who posted the message," which is stated to be a part of the *metadata* (not the data).

B is CORRECT because to determine the topics that many users are posting about, it would be best to analyze the *message itself* (i.e. the text contained in the message), which *is* stated to be a part of the data.

C is INCORRECT because to determine the time of day that the site is most active, it would be best to analyze "The time the message was posted," which is stated to be a part of the *metadata* (not the data).

D is INCORRECT because to determine which posts have received the greatest number of comments, it would be best to analyze "The names of any users who comment on the message and the times the comments were made," which is stated to be a part of the *metadata* (not the data).

Answer (back)

Correct Answers: B & C

A student is creating a Web site that is intended to display information about a city based on a city name that a user enters in a text field. Which of the following are likely to be challenges associated with processing city names that users might provide as input?

Select two answers.

A Users might attempt to use the Web site to search for multiple cities.

B Users might enter abbreviations for the names of cities.

C Users might misspell the name of the city.

D Users might be slow at typing a city name in the text field.

Explanations:

Answer A

Incorrect. A user entering data into the Web site to search for multiple cities does not directly affect the quality of the data. If the Web site is working as intended, users should be able to use it as many times as they want.

Answer B

Correct. Different users may abbreviate city names differently. This may require the student to clean the data to make it uniform before it can be processed.

Answer C

Correct. Misspelled city names will not be an exact match to information stored by the Web site. This may require the student to clean the data to make it uniform before it can be processed.

Answer D

Incorrect. A user's typing speed does not directly affect the quality of the data. Until a city name is entered, the Web site cannot search for information.

Answers (back)

Process	Execution Time
A	30 seconds
B	40 seconds
C	35 seconds
D	20 seconds

How long would it take to complete all 4 tasks if Maria arranged them to run in parallel like this:

Processor 1's Tasks	Processor 2's Tasks
Process B	Process A Process C Process D

85 seconds

Explanation: Processor 1 would take 40 seconds. Processor 2 would take 30 + 35 + 20 = 85 seconds. Since the processors are running in parallel (simultaneously), it would take **85 seconds** (the longer of the two) for all 4 processes to complete. Note that processor 1, which finished in 40 seconds, would be idle for the remaining 45 seconds while processor 2 finishes.

How long would it take to complete all 4 tasks if Maria arranged them to run *sequentially* like this:

Processor 1's Tasks	Processor 2's Tasks
Process A Process B Process C Process D	*none assigned* *processor idle*

It would take 30 + 40 + 35 + 20 = **125 seconds** for all 4 tasks to complete when running sequentially on one processor.

How should Maria arrange the 4 tasks on two processors if she wants to *minimize* the total time she needs to wait for the 4 tasks to be complete?

Complete the chart →

The optimal arrangement of tasks that minimizes the time it would take would be to put Processes A and C on one processor, and Processes B and D on another:

Processor 1's Tasks	Processor 2's Tasks
Process A Process C	Process B Process D

OR

Processor 1's Tasks	Processor 2's Tasks
Process B Process D	Process A Process C

How long will it take to complete all 4 tasks if Maria arranges them optimally like this?	**65 seconds** The processor with A and C would take 30 + 35 = 65 seconds, and the processor with B and D would take 40 + 20 = 60 seconds. So the longer of the two is 65 seconds.
How long is one of the processors idle while the other finishes?	**5 seconds** One processor finishes in 60 seconds and the other in 65, so the one that finishes in 60 is idle for 5 seconds while the other processor finishes.
What's the difference between the optimal (minimum) execution time when running the 4 tasks in parallel and the time it would take if Maria ran all 4 tasks sequentially on one processor?	The minimum time is 65 seconds using parallel processors. Using one processor to do the tasks sequentially takes 125 seconds. So the difference is 125 - 65 = **60 seconds**

(back)