Title: Assignment 1
Subtitle: Computer performance, reliability, and scalability calculation
Author: Tushar Muley

Assignment 1.2
a. Data Sizes

TM - Summary

Items #	Data Item	Size per Item
1	128 character message.	128 Bytes
2	1024x768 PNG image	0.45704 MB
3	1024x768 RAW image	0.68813 MB
4	HD (1080p) HEVC Video (15 minutes)	641.25MB
5	HD (1080p) Uncompressed Video (15 minutes)	160,180.66MB
6	4K UHD HEVC Video (15 minutes)	2,565 MB
7	4k UHD Uncompressed Video (15 minutes)	640,722.6 MB
8	Human Genome (Uncompressed)	1.5 GB

TM - Detail

1- | 128 character message. | 128 Bytes |

TM- Using ASCII character set each character is 1 byte so 128 bytes

2- | 1024x768 PNG image | 0.45704 MB |

TM- Use an online calculator and doubled using a different one.
Site 1 - 457.04KB or 0.45704MB for a 24bit image with 16.7million colors
Site 2 - 4608kb or 4.608MB for a 48bit

Reference:
Oncline calculator
https://toolstud.io/photo/megapixel.php?width=1024&height=768&compare=video&calculate=compressed#calculate
https://jan.ucc.nau.edu/lrm22/pixels2bytes/calculator.htm

3- | 1024x768 RAW image | 0.68813 MB |

TM - Using a online calculator a 1024x768 RAW image on a Canon sRAW CR2 would be 0.68813MB or 688.13KB compressed
Canon mRAW CR2 1.03MB
If you use uncompressed RAW/DNG 16 bits/pixel 1.57MB

4- | HD (1080p) HEVC Video (15 minutes) | 641.25MB |

TM - Total video file size 900 seconds x 5.7MB/sec
5.13GB or 641.25 (5.13*125)

Reference:
Using a online calculator
https://toolstud.io/video/filesize.php?width=1920&height=1080&framerate=30&timeduration=15&timeduration_unit=minutes&compression=91628&specificbitrate=100&specificbitrate_unit=1000000

5- | HD (1080p) Uncompressed Video (15 minutes) | 160,180.66MB |

TM - Using calculaiton on stackoverflow.com

Part1:
Size of resultion (1920x1080 resolution of 1080P) x 3 bytes per pixel x 30 frames per second / (1024x1024 bytes per megabyte) than

Part2:
Take the total from part 1 which is in MB/s * 60sec*15minutes (lenght of video) =

1920 x 1080 x 3 x 30frames per second / (1024x1024 bytes per megabytes) = 177.98 MB/s
177.98 MB/s * 60 sec*15min = 160,180.66 MB or 156.43GB

Reference:
https://stackoverflow.com/questions/27559103/video-size-calculation

6- | 4K UHD HEVC Video (15 minutes) | 2,565 MB |

TM - Using the UHD screen size of 3840x2160 and HEVC and 24bit RGB you get a compressed size of 20.52GB or 2,565 MB.

Reference:
Online calculator
https://toolstud.io/video/filesize.php?width=3840&height=2160&framerate=30&timeduration=15&timeduration_unit=minutes&compression=91628&specificbitrate=100&specificbitrate_unit=1000000

7- | 4k UHD Uncompressed Video (15 minutes) | 640,722.6 MB |

TM - This got a little complex. Using the below calculation from Slackoverflow.
Part1:
Size of resultion (3840x2160 UHD Video size) x 3 bytes per pixel x 30 frames per second / (1024x1024 bytes per megabyte) than

Part2:
Take the total from part 1 which is in MB/s * 60sec*15minutes (lenght of video) =

3840 x 2160 x 3 x 30frames per second / (1024x1024 bytes per megabytes) = 711.914 MB/s
711.914 MB/s * 60 sec*15min = 640,722.6 MB or 625.71 GB

Reference:
https://stackoverflow.com/questions/27559103/video-size-calculation

8- Human Genome (Uncompressed) | 1.5 GB |

TM - I found this online at medium.com
https://medium.com/precision-medicine/how-big-is-the-human-genome-e90caa3409b0
715MB or 0.715GB for a copy of on genome. Since each human has 2 we double that 1.5GB

Reference:
Robinson, R MD MBA. (June 2014). "How big is the human genome?". From Medium.com.
https://medium.com/precision-medicine/how-big-is-the-human-genome-e90caa3409b0

https://www.dailymail.co.uk/sciencetech/article-2680849/How-gigabytes-does-make-HUMAN-Physicians-works-genetic-code-just-1-5GB-data.html

b. Scaling
TM - Summary

Items #	Data Item	Size per Item	# HD
1	Daily Twitter Tweets (Uncompressed)	64 GB	1
2	Daily Twitter Tweets (Snappy Compressed)	112.8 GB – 127.8 GB	1
3	Daily Instagram Photos	102.86	11
4	Daily YouTube Videos	44,353.97 TB	4,436
5	Yearly Twitter Tweets (Uncompressed)	70.08 TB	7
6	Yearly Twitter Tweets (Snappy Compressed)	41.172 TB – 46.647 TB	5
7	Yearly Instagram Photos	3,754.43 TB	3,755
8	Yearly YouTube Videos	16,189,202.7 TB	1,618,921

TM - Detail

1- Daily Twitter Tweets (Uncompressed)

500,000,000 * 128 bytes = 64,000,000,000 or 64,000MB or 64GB
HDFS 64GB*3 = 192GB
1TB = 1000GB so one 1 harddrive should be enough partitioned correctly

2- Daily Twitter Tweets (Snappy Compressed)

500,000,000 * 128bytes = 64,000,000,000 or 64,000MB
compressed @ 1.5 = 42,666.66MB 33.33% compression
compressed @ 1.7 = 37,647.06MB 41.18% compression

Between 42.6GB and 37.6GB in size
HDFS between 127.8GB and 112.8GB
between 0.1278TB and 0.1128TB

Reference:
Typical compression ratios (based on the benchmark suite) are about 1.5-1.7x for plain text, about 2-4x for HTML, and of course 1.0x for JPEGs, PNGs and other already-compressed data.

https://github.com/google/snappy
https://rechneronline.de/transfer/compression.php

3- Daily Instagram Photos

Estimates over 100 million videos and photos are uploaded to Instagram every day. Assume that 75% of those items are 1024x768 PNG photos.
100,000,000 video and photos 75% are 1024x768 = 75,000,000 and remaining 25,000,000 are not 1024x768 PNG

Average video size lenght on Insta about 26 seconds
Your videos should be no more than 60 seconds in length, as that is the upper limit for what the platform will support. Data compiled by HubSpot suggests 30 seconds might be an ideal time frame, as Instagram videos that received the most comments typically averaged about 26 seconds in length.

Assuming 0.45704MB for 1 1024x768 PNG and 75% are 1024x768

75,000,000*0.45704MB = 34,278,000MB or 34,278GB or 34.287TB

34.287*3 (HDFS)=102.861
102.861/10 = 10.2861 Hard drives required per data for photos

Referrence:
https://2060digital.com/blog/right-video-length-platform/#:~:text=Your%20videos%20should%20be%20no%20more%20than%2060%20seconds%20in,about%2026%20seconds%20in%20length.

https://www.omnicoreagency.com/instagram-statistics/

4- Daily YouTube Videos

Estimates 500 hours of video is uploaded to YouTube every minute. For simplicity, assume all videos are HD quality encoded using HEVC at 30 frames per second.

HD (1080p) HEVC Video (15 minutes)-> 641.25MB

1 day has 1,441 minutes (60*24)
500 hrs*1,441=720,500 hours of video a day

20.52GB per hour*720,500 hours of video = 14,784,660GB
14,784,660GB*3 HDFS = 44,353,980GB or 44353.98TB

44,353.98TB/10TB = 4,435.398 Hard drives

Reference:
https://toolstud.io/video/filesize.php?width=1920&height=1080&framerate=30&timeduration=720000&timeduration_unit=hours&compression=91628&specificbitrate=100&specificbitrate_unit=1000000

5- Yearly Twitter Tweets (Uncompressed)

HDFS 64GB*3 = 192GB

192GB from question 1
192*365days = 70,080GB or 70.08TB
70.08TB/10TB = 7 hard drives

6- Yearly Twitter Tweets (Snappy Compressed)

From question 2 HDFS between 127.8GB and 112.8GB

112.8*365=41,172GB or 41.172TB
127.8*365=46,647GB or 46.647TB

41.172/10= 4.5 hard drives
46.647/10= 4.6 hard drives
5 hard drives

7- Yearly Instagram Photos

From question 3
34,278,000MB or 34,278GB or 34.287TB
34.287*3 (HDFS)=102.861TB

102.861*365=37,544.265

37,544.265/10=3,754.4265 or 3,755

8- Yearly YouTube Videos

From question 4
20.52GB per hour*720,500 hours of video = 14,784,660GB
14,784,660GB*3 HDFS = 44,353,980GB or 44353.98TB

44,353.98TB/10TB = 4,435.398 Hard drives

44,353.98*365=16,189,202.7
16,189,202.7/10 = 1,618,920.27 Harddrives

c. Reliability
TM- Summary

Items #	Data Item	# HD	# Failures
	Twitter Tweets (Uncompressed)	7	0
	Twitter Tweets (Snappy Compressed)	5	0
	Instagram Photos	3,755	42
	YouTube Videos	1,618,921	17,808

TM- Detail

1- Twitter Tweets (Uncompressed)

From question above:
HDFS 64GB*3 = 192GB

192GB from question 1
192*365days = 70,080GB or 70.08TB
70.08TB/10TB = 7 hard drives

TM- Based on the reading from the website the annualized failure rate is 1.10% for Q3 2021

7 drives 1.10% AFR = 0.077 or zero drives failure

2- Twitter Tweets (Snappy Compressed)

From question above:
112.8*365=41,172GB or 41.172TB
127.8*365=46,647GB or 46.647TB

41.172/10= 4.5 hard drives
46.647/10= 4.6 hard drives
5 hard drives

5 drives 1.10% AFR = 0.055 or zero drives failure

3- Instagram Photos

From question above:
34,278,000MB or 34,278GB or 34.287TB
34.287*3 (HDFS)=102.861TB

102.861*365=37,544.265
37,544.265/10=3,754.4265 or 3,755
3,755 drives 1.10% AFR = 41.3 or 42 drives failure

4- YouTube Videos

From question above:
20.52GB per hour*720,500 hours of video = 14,784,660GB
14,784,660GB*3 HDFS = 44,353,980GB or 44353.98TB

44,353.98TB/10TB = 4,435.398 Hard drives

44,353.98*365=16,189,202.7
16,189,202.7/10 = 1,618,920.27 Harddrives

1,618,921 drives 1.10% AFR = 17,808.131 or 17,808 drives failure

Reference for all questions:
https://www.backblaze.com/b2/hard-drive-test-data.html
https://en.wikipedia.org/wiki/Annualized_failure_rate

d. Latency

TM - Summary

	One Way Latency
1- Los Angeles to Amsterdam	53 ms
2- Low Earth Orbit Satellite	45 ms
3- Geostationary Satellite	500 ms
4- Earth to the Moon	1,300 ms
5- Earth to Mars	20 minutes

Item #	Description	On Way Latency
1	Los Angeles to Amsterdam	53 ms
2	Low Earth Orbit Satellite	45 ms
3	Geostationary Satellite	500 ms
4	Earth to the Moon	1,300 ms
5	Earth to Mars	20 minutes

TM - Detail

1- | Los Angeles to Amsterdam

5,571 total miles one way

5,571*110% = 6,128.1
+ 100 miles (LA local fiber) = 6,228.1
+ 100 miles (AMS local fiber) = 6,328.1

/124 msec/miles =51.03 msec
+ 2msec = 53.03 msec

Reference for distance:
https://www.airmilescalculator.com/distance/lax-to-ams/

2- | Low Earth Orbit Satellite

Distance to low earth orbit = 1,200 miles

The latency depends on the number of hops the signal has to make before getting to its destination. If the satellite was directly overhead 45ms is the best. Assume calculation provided by Starlink are accurate.

Reference:
1-Rutkowski R. (Jun 2021). "5 FAQs About Low Earth Orbit (LEO) Satellite Constellations".from bliley Technonolgy Blog
https://blog.bliley.com/5-faq-answers-new-space-leo-satellite-constellations#:~:text=A%20Low%20Earth%20Orbit%20is,altitude%20loss%20and%20orbital%20decay.

2-https://en.wikipedia.org/wiki/Satellite_Internet_access

3-Vaughan-Nichols. S (August 2021). "Starlink is better than its satellite competition but not as fast as landline internet".
From ZDNet.com
https://www.zdnet.com/article/starlink-is-better-than-its-satellite-competition-but-not-as-fast-as-landline-internet/

3- | Geostationary Satellite

Distance to GOE = 22,282 miles

500 ms
For a single hop:
speed = distance / time

time = distance / speed
time = 35,786 km / 299,792 km/s
0.11937 second for a Radio Frequency

Reference:
1-Rutkowski R. (Jun 2021). "5 FAQs About Low Earth Orbit (LEO) Satellite Constellations".from bliley Technonolgy Blog
https://blog.bliley.com/5-faq-answers-new-space-leo-satellite-constellations#:~:text=A%20Low%20Earth%20Orbit%20is,altitude%20loss%20and%20orbital%20decay.

2-https://en.wikipedia.org/wiki/Satellite_Internet_access

3-https://satoms.com/satellite-latency/

4- | Earth to the Moon

Distance from Earth to Moon = 238,606.538miles
latency = 1.3sec = 1,300 ms

Reference:
https://www.spaceacademy.net.au/spacelink/commdly.htm

5- | Earth to Mars

Distance from Earth to Mars = 231.55 million miles
20 minutes

Reference:
https://mars.nasa.gov/mer/mission/timeline/surfaceops/navigation/
https://www.spaceacademy.net.au/spacelink/commdly.htm

Reference for all questions:
1- "Estimating WAN Latency".(October 2020). from Oneneck IT Solutions.com.
https://www.oneneck.com/blog/estimating-wan-latency-requirements

To estimate latency of a wavelength between metro areas, take a look at your carrier's proposed fiber route. Their fiber map is probably available on their website. An online mapping website can provide a rough estimate of highway distance between two cities, along the path used by your carrier. I'd recommend padding that distance by 10% to account for a few twists and turns at river crossings, freeway junctions, etc. To that estimate, add 100 miles for the local metro fiber path at each end, double the result (to get the round-trip distance), divide by 124 miles/msec, and add 2 milliseconds (for the electronics). For example, a Chicago/Dallas wavelength latency estimate might be:

925 miles x 110% = 1017.5 miles
+ 100 miles (Chicago local fiber) = 1117.5 miles
+ 100 miles (Dallas fiber) = 1217.5 miles
x 2 = 2435 miles (round-trip)
/ 124 msec/mile = 19.6 msec
+ 2 msec = 22 msec (electronics)

2- Online calculator used to confirm numbers
https://wintelguy.com/wanlat.html