Title: Assignment 1
Subtitle: Computer performance, reliability, and scalability calculation
Author: Sam Sears

1.2

a. Data Sizes

Data Item	Size per Item
128 character message.	128-512 Bytes
1024x768 PNG image	1.1 MB
1024x768 RAW image	2.25 MB
HD (1080p) HEVC Video (15 minutes)	900 MB
HD (1080p) Uncompressed Video (15 minutes)	66,742 MB
4K UHD HEVC Video (15 minutes)	2,550 MB
4k UHD Uncompressed Video (15 minutes)	320,361 MB
Human Genome (Uncompressed)	1.5 GB

- **128 character message** There are between 0.03125 and 0.125 characters per bit, so 0.25-1 charaters per byte, or 128 bytes to 512 bytes for 128 characters. Source: https://stackoverflow.com/questions/4850241/how-many-bits-or-bytes-are-there-in-a-character#:~:text=An%20ISO%2D8895%2D1%20character,common%20characters%20take%2016%20bits (https://stackoverflow.com/questions/4850241/how-many-bits-or-bytes-are-there-in-a-character#:~:text=An%20ISO%2D8895%2D1%20character,common%20characters%20take%2016%20bits)
- **1024x768 PNG image** Calculated on https://www.scantips.com/basics1d.html (https://www.scantips.com/basics1d.html) using 24-bit RGB
- **1024x768 RAW image** Calculated on https://www.scantips.com/basics1d.html (https://www.scantips.com/basics1d.html)
- **HD (1080p) HEVC Video (15 minutes)** Estimate based on iPhone Video file sizes at 30 frames per second https://larryjordan.com/articles/iphone-video-file-sizes/ (https://larryjordan.com/articles/iphone-video-file-sizes/)
- **HD (1080p) Uncompressed Video (15 minutes)** 30 frames per second x 900 seconds = 27,000 frames. 1080p: 1920 x 1080 x 10 (bit depth) / 8 / 1024 = 2,531.25 Kb. 27,000 frames x 2,531.25 Kb / 1024 = 66,742 MB Source: https://www.youtube.com/watch?v=DDcYvesZsnw (https://www.youtube.com/watch?v=DDcYvesZsnw)
- **4K UHD HEVC Video (15 minutes)** Estimate based on iPhone Video file sizes at 30 frames per second https://larryjordan.com/articles/iphone-video-file-sizes/ (https://larryjordan.com/articles/iphone-video-file-sizes/)
- **4k UHD Uncompressed Video (15 minutes)** 30 frames per second x 900 seconds = 27,000 frames. 4k: 3840 x 2160 x 12 (bit depth) / 8 / 1024 = 12,150 Kb. 27,000 frames x 12,150 Kb / 1024 = 320,361 MB Source: https://www.youtube.com/watch?v=DDcYvesZsnw (https://www.youtube.com/watch?v=DDcYvesZsnw)
- **Human Genome (Uncompressed)** According to https://bitesizebio.com/8378/how-much-information-is-stored-in-the-human-genome/#:~:text=In%20order%20to%20represent%20the,2%20CDs%20worth%20of%20space!

(https://bitesizebio.com/8378/how-much-information-is-stored-in-the-human-genome/#:~:text=In%20order%20to%20represent%20the,2%20CDs%20worth%20of%20space!)

b. Scaling

	Size	# HD
Daily Twitter Tweets (Uncompressed)	238.42 GB	3
Daily Twitter Tweets (Snappy Compressed)	158.95 GB	3
Daily Instagram Photos	78.68 TB	24
Daily YouTube Videos	2.41 PB	742
Yearly Twitter Tweets (Uncompressed)	84.98 TB	26
Yearly Twitter Tweets (Snappy Compressed)	56.65 TB	17
Yearly Instagram Photos	28.04 PB	8,616
Yearly YouTube Videos	626.56 TB	270,676

- **Daily Twitter Tweets (Uncompressed)** Let's assume using unicode characters in UTF-8 which would be between 1 and 4 Bytes per character. I'll use the high end, 4 Bytes per character. 128 characters x 4 Bytes = 512 Bytes per tweet. Other estimates I found were an average of 200 Bytes per tweet and 560 Bytes per 140 character tweet. https://stackoverflow.com/questions/5999821/how-many-bytes-of-memory-is-a-tweet/5999852 (https://stackoverflow.com/questions/5999821/how-many-bytes-of-memory-is-a-tweet/5999852). 512 Bytes per tweet x 500 million tweets per day = 256 billion Bytes / 1024 / 1024 / 1024 = 238.42 GB in size. For HD, multiply by 3 since we are using HDFS, 715.26 GB, which is less than 10 TB. But we need three machines for Hadoop because the copies can't be on the same machine, so we'd need 3 hard drives.
- **Daily Twitter Tweets (Snappy Compressed)** Snappy compression ratio is 1.5-1.7x for plain text. I will take the prior for my estimate. 238.42 GB compressed with a ratio of 1.5 would be 158.95 GB in size. 158.95 x 3 copies = 476.84 GB. However, we would still need 3 hard drives because of we need at least 3 machines for Hadoop.
- **Daily Instagram Photos** Based on the assumption that 75% of the the 100 million videos and photos are photos, and those photos are 1024x768 PNG photos. 75 million photos x 1.1 MB per photo = 82.5 million MB / 1024 = 80,566.41 GB / 1024 = 78.68 TB in size. 78.68 TB x 3 copies = 236.03 TB / 10 TB per HD = 23.603 HD ~ 24 hard drives.
- **Daily YouTube Videos** 900 MB per 15 minutes (from example above). 900 MB x 4 = 3,600 MB per hour x 500 hours = 1.8 million MB per minute x 60 minutes x 24 hours = 2,592,000,000 MB per day/1024 = 2,531,250 GB per day/1024 = 2,471.92 TB size x 3 copies = 7,415.77 TB storage needs / 10 TB per hard drive = 741.577 ~ 742 hard drives
- **Yearly Twitter Tweets (Uncompressed)** 238.4 GB per day x 365 days per year = 87,016 GB per year / 1024 = 84.98 TB in size. 84.98 x 3 copies = 254.94 TB of storage needed. 254.94 / 10 TB per hard drive = 25.494 ~ 26 hard drives.
- **Yearly Twitter Tweets (Snappy Compressed)** Similar to the the other twitter problem, I will use 1.5x compression ratio for plain text. 84.98 TB compressed with a ratio of 1.5 would be 56.65 TB in size. 56.65 TB x 3 copies = 169.96 TB total storage needed. 169.96 TB / 10 TB per hard drive = 16.996 ~ 17 hard drives.
- **Yearly Instagram Photos** 78.68 TB per day x 365 day per year = 28,717.52 TB or 28.04 PB in size. 28,717.52 TB x 3 copies = 86,152.55 TB / 10 TB per HD = 8,615.26 ~ 8,616 hard drives.

- **Yearly YouTube Videos** 2.41 PB per day x 365 days per year = 881.11 PB per year size x 3 copies = 2,643.32 PB x 1024 = 2,706,756.59 TB / 10 TB per HD = 270,675.695 ~ 270,676 hard drives.

c. Reliability

	# HD	# Failures
Twitter Tweets (Uncompressed)	26	0
Twitter Tweets (Snappy Compressed)	17	0
Instagram Photos	8616	80
YouTube Videos	270,676	2,517

- **Twitter Tweets (Uncompressed)** 26 hard drives x 0.0093 annualized failure rate = 0.2418 failures, can't have 0.2418 failures, so best guess would be rounding to 0.
- **Twitter Tweets (Snappy Compressed)** 17 hard drives x 0.0093 annualized failure rate = 0.1581 failures
- **Instagram Photos** 8616 hard drives x 0.0093 annualized failure rate = 80.1288 failures
- **YouTube Videos** 270,676 hard drives x 0.0093 annualized failure rate = 2,517.28 failures

d. Latency

	One Way Latency
Los Angeles to Amsterdam	51.175 ms
Low Earth Orbit Satellite	20 ms
Geostationary Satellite	300 ms
Earth to the Moon	1.3 s
Earth to Mars	3-21 minutes

- **Los Angeles to Amsterdam** 5565 miles from LA to Amsterdam - 102.35 ms RTT / 2 = 51.175 ms one way. All assuming 10% Fibre path adjustment, 100 miles for Metro fiber and local loop length, 124 mi/millisecond speed of light in fibre, and 1 millisecond equipment latency, using https://wintelguy.com/wanlat.html (https://wintelguy.com/wanlat.html).
- **Low Earth Orbit Satellite** 40 ms RTT (20 ms one way) according to https://www.omniaccess.com/leo/ (https://www.omniaccess.com/leo/).
- **Geostationary Satellite**600 ms RTT (300 ms one way) according to https://www.omniaccess.com/leo/ (https://www.omniaccess.com/leo/)
- **Earth to the Moon**1.3 s according to https://www.spaceacademy.net.au/spacelink/commdly.htm (https://www.spaceacademy.net.au/spacelink/commdly.htm)
- **Earth to Mars** 3-21 minutes according to https://www.spaceacademy.net.au/spacelink/commdly.htm (https://www.spaceacademy.net.au/spacelink/commdly.htm)

In [ ]: