**a. Data Sizes**

Provide estimates for the size of various data items. Please explain how you arrived at the estimates for the size of each item by citing references or providing calculations.

| Data Item | Size per Item |
| --- | --- |
| 1. **128 character message.** | **128 Bytes** |
| <https://www.quora.com/How-many-bytes-are-needed-to-store-string>  Total size (in Bytes) = ((Number of bits used to encode a single character) \* Number of characters))/8   * ASCII encoding: 8 bits to encode each character   Total size (in Bytes) = (8\*128)/8 = 128 bytes |  |
| 1. 1024x768 PNG image   <https://4nsi.com/how-do-i-calculate-the-file-size-for-a-digital-image/>  Step 1: Total number of pixels = 1024 \* 768 = 786,432  Step 2: Bit depth of the detector is 16-bit  Total number of bits of data = 786,432 \* 16 = 12,582,912  Step 3: File size in bytes = 12,582,912/8 = 1,572,864  Step 4: File size in MB = 1,572,864/1024 = 1,536/1024 = 1.5 | **1.5 MB** |
| 1. 1024x768 RAW image   <http://preservationtutorial.library.cornell.edu/intro/intro-06.html#:~:text=FILE%20SIZE%20is%20calculated%20by,divide%20this%20figure%20by%208>.  **File Size = (pixel dimensions x bit depth) / 8**  **= (1024 \* 768)/8**  **= 98,304 bytes**  **= 98,304/1024 = 96/1024 = 0.0938 MB** | 0.0938 MB |
| 1. HD (1080p) HEVC Video (15 minutes)  * <https://en.wikipedia.org/wiki/High_Efficiency_Video_Coding> --> compression ratio 🡪 1000: 1 * <https://www.circlehd.com/blog/how-to-calculate-video-file-size> * Frame size = 1080\*1920 pixels * Frame rate = 30 frames per second * 24-bit videos * Formula = 24 (bit) \* 1920 \* 1080 (pixels) \* 30 (frames per second) \* 900 seconds (~15 mintutes) = 1.344e12 bytes * 1.344e12 / 8192 = 164,025,000 / 1024 = 160,180MB * 160,180 MB / 1000 (compression ratio) = 160.18 MB | ? MB |
| 1. HD (1080p) Uncompressed Video (15 minutes)  * Approximate size of each uncompressed frame is 5MB * At 30 frames per second, 5MB \* 30 = 150MB storage space per second * 15 minutes \* 60 (seconds in a minute) = 900 seconds * 150MB \* 900 = 135,000 MB per second * 135,000/60 = 2,250MB per minute * <https://www.circlehd.com/blog/how-to-calculate-video-file-size> | 2,250 MB |
| 1. 4K UHD HEVC Video (15 minutes)  * <https://www.circlehd.com/blog/how-to-calculate-video-file-size> * <https://www.lifewire.com/4k-resolution-overview-and-perspective-1846842> * Pixels = 3840 x 2160 pixel * Formula = 24 (bit) \* 3840 \* 2160 (pixels) \* 30 frames per second \* 900 seconds = 5.37e12 * 5.37e12 / 8192 =656,100,000 / 1024 = 640,723 MB * 640,723 MB / 1000 = 641 MB | ? MB |
| 1. 4k UHD Uncompressed Video (15 minutes)  * HEVC compressed = 18,900 MB a minute * Multiply by 2 to get uncompressed = 18,900\*2 * 37,800MB a minute | ? MB |
| 1. Human Genome (Uncompressed)  * <https://www.biostars.org/p/5514/> * 6e9 bits uncompressed 🡪 6000000000 bytes 🡪 6GB | 6 GB |

* Assume all videos are 30 frames per second
* [HEVC](https://en.wikipedia.org/wiki/High_Efficiency_Video_Coding) stands for High Efficiency Video Coding
* See the Wikipedia article on [display resolution](https://en.wikipedia.org/wiki/Display_resolution) for information on HD (1080p) and 4K UHD resolutions.

**b. Scaling**

Using the estimates for data sizes in the previous part, determine how much storage space you would need for the following items.

|  | Size | # HD |
| --- | --- | --- |
| Daily Twitter Tweets (Uncompressed)   * <https://contingencycoder.wordpress.com/2013/04/10/how-much-space-do-all-those-tweets-take-up/> * <https://drivesaversdatarecovery.com/is-bigger-better-how-to-choose-the-right-hard-drive-sizes/> * Bytes for each tweet = 4 \* 128 characters + 1 (length of characters in the database) = 513 bytes * 513 bytes \* 500 million tweets a day = 2.565e11   + GB = 238.9 GB a day   + TB = 0.23 * Need 10-15% of a hard drive’s storage to function | ?? |  |
| Daily Twitter Tweets (Snappy Compressed)   * https://www.infoq.com/news/2011/04/Snappy/#:~:text=The%20high%20compression%20speed%20is,other%20already%2Dcompressed%20data%E2%80%9D. * For compressed files, we need to remember to multiply by the compression ratio on the total file size * Snappy compresses at about 250 MB/sec   + Decompresses at about 500MB/sec * Compression ratio of about 1.5-1.7x for plain text 🡪 tweets are usually in UTF-8 encoding * Uncompressed file size ~ 238.9 GB aday * Compressed file size = 238.9GB \* 1.6 = 382.2 GB a day | ?? |  |
| Daily Instagram Photos   * 100 million \* .75 = 75 million photos a day * 1024 \* 768 PNG photos * <https://superuser.com/questions/759026/how-big-is-an-800x600-image-file-in-png> * Size = 1024 \* 768 pixel \* 4 bytes per pixel (1 byte per R/G/B + apha) = 3,145,728 bytes * \* 75 million = 214 TB a day | ?? |  |
| Daily YouTube Videos   * <https://teradek.com/blogs/articles/3-reasons-why-hevc-x265-matters-and-how-you-can-start-using-it-now>   + HEVC 🡪 half the bitrate * 500 hours of video uploaded to YouTube every minute * 24 hours in a day \* 60 minutes = 1440 minutes * 500 \* 1440 = 720,000 hours of video uploaded every day * HD quality encoded 🡪 HEVC   + 30 frames per second * <https://www.circlehd.com/blog/how-to-calculate-video-file-size>   + HD (720p)   + Bitrate = 1 Mbps = 0.5 Mbps with HEVC   + 1 minute = 5 MB   + Recording duration per GB = 3.5 hours * File size = Bitrate x duration x compression ratio * Raw HD video = 5 MB (size uncompressed) x 30 (frames per second) = 150 MB \* 0.5 = 75 (HEVC) storage a second * 75 \* 3600 = 270,000 MB storage an hour * 270,000 MB an hour \* 720,000 for YouTube uploads * 181 TB a day | ?? |  |
| Yearly Twitter Tweets (Uncompressed)   * Daily Twitter Tweets uncompressed = 238.9GB a day * Yearly = 238.9 \* 365 = 87198.5GB a year   + 85.2 TB a year | ?? |  |
| Yearly Twitter Tweets (Snappy Compressed) | ?? |  |
| Yearly Instagram Photos | ?? |  |
| Yearly YouTube Videos | ?? |  |

* For estimating the number of hard drives, assume you are using 10 TB and you are storing the data using the Hadoop Distributed File System (HDFS). By default, HDFS stores three copies of each piece of data, so you will need to triple the amount storage required.
* [Twitter statistics](https://www.internetlivestats.com/twitter-statistics/) estimates 500 million tweets are sent each day. For simplicity, assume each tweet is 128 characters.
* See the [Snappy Github repository](https://github.com/google/snappy) for estimates of Snappy's performance.
* [Instagram statistics](https://www.omnicoreagency.com/instagram-statistics/) estimates over 100 million videos and photos are uploaded to Instagram every day. Assume that 75% of those items are 1024x768 PNG photos.
* [YouTube statistics](https://www.omnicoreagency.com/youtube-statistics/) 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.

**c. Reliability**

Using the yearly estimates from the previous part, estimate the number of hard drive failures per year using data from [Backblaze's hard drive statistics](https://www.backblaze.com/b2/hard-drive-test-data.html).

* Since we are using a 10TB for estimating the number of hard sizes, I am going to reference the Seagate 10TB drive for its statistics since it is the only 10TB Backblaze hard drive mentioned in the data.
* AFR = 2.26%

|  | # HD | # Failures |
| --- | --- | --- |
| Twitter Tweets (Uncompressed) | 30 |  |
| Twitter Tweets (Snappy Compressed) | ?? |  |
| Instagram Photos | ?? |  |
| YouTube Videos | ?? |  |

**d. Latency**

Provide estimates of the one way latency for each of the following items. Please explain how you arrived at the estimates for each item by citing references or providing calculations.

##### References  
- For latency between cities, avg ping time was used <https://wondernetwork.com/pings/Los+Angeles>  
- LEOS - <https://www.omniaccess.com/leo/>  
- Geostationary Statellite - <https://www.omniaccess.com/leo/>  
- Earth to Moon/Earth to mars - <https://www.spaceacademy.net.au/spacelink/commdly.htm>

|  | One Way Latency |
| --- | --- |
| Los Angeles to Amsterdam | 133.47 ms |
|  |  |
| Low Earth Orbit Satellite | ? ms |
| Geostationary Satellite   * <https://www.satsig.net/latency.htm> | ? ms |
| Earth to the Moon | ? ms |
| Earth to Mars | ? minutes |