

# Compression

process of reducing the size of data

benefits

- **lower latency and higher throughput while transmitting messages over the network**  
since we have less data to transfer
- **increased storage capacity**  
since we have less data to store
- **decreased costs**  
since the cost of data transfer for many cloud services is based on the total amount of data served

# Compression

compression is everywhere



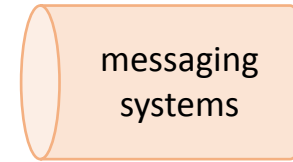
browsers

decompress downloaded data



web  
servers

compress HTTP data for faster  
transfer over the network



messaging  
systems

producers compress,  
consumers decompress



databases

compress data before  
saving to disk

# Compression

some more facts about compression

- Compression becomes generally more effective as the data size increases.  
Compression loves batching.
- Compression and decompression processes consume computational resources (CPU).
- There are two kinds of compression: lossless and lossy.
- Lossy compression loses data.  
Most commonly used to compress audio, video, images.
- No information is lost in lossless compression.

# Compression

compression algorithms trade-off between the following three areas

## compression speed

how fast the compression algorithm  
compresses data

important for write-heavy applications

## decompression speed

how fast the compression algorithm  
decompresses data

important for read-heavy applications

## compression ratio

ratio the uncompressed data is reduced by

important for applications that store  
a lot of data on disk

# Compression

algorithm	compression speed	decompression speed	compression ratio
<b>Deflate (gzip)</b> Standard format for HTTP compression.	B	B	A
<b>Snappy</b> Created by Google. Used extensively in Google projects like Bigtable and MapReduce. Many different NoSQL databases support Snappy.	A -	A	B
<b>Zstandard</b> Created by Facebook. Widely used in file systems and databases.	A -	A -	A +