Q: What is caching?

...have we talked about caching before?

Q: Why is caching important on a cloud-scale?

Caching can occur at multiple different places – and we have different strategies and techniques for each “layer” of the full caching strategy.

Caching Technique #1: Age-Based Caching

Idea + Purpose of Age-Based Caching:

Age-Based Caching Use Case:
- The CS 240 course websites the Bootstrap CSS library from stylizing the front-end content.
- When we load the page from a browser that has never visited the CS 240 website, we see the following requests:

```
GET https://stackpath.bootstrapcdn.com/bootstrap/4.2.1/js/bootstrap.min.js
```

HTTP Response:

```
HTTP/1.1 200 OK
date: Wed, 06 Apr 2022 18:37:00 GMT
age: 1883002
cache-control: public, max-age=31919000
cdn-cache: HIT
cdn-cachedat: 2021-06-08 14:31:34
```

[Line 3]: The `age` header:

[Line 4]: The `cache-control["max-age"]` header:

Q: What happens when we request a page using this resource again before the content expires?

....how much traffic is saved?

....how much time is saved?

...what are disadvantages of age-based caching?
Caching Technique #2: Entity Tag Caching
Idea + Purpose of Entity Tag (ETag) Caching:

**ETag-Based Caching Use Case:**
- Some content is likely to be the same for a period of time, but change occasionally and at unpredictable times.
- Ex: My Institute for Interactive Visualization Systems site has data updates to the visualizations each day. Here’s a request for the snow data in Champaign, IL:

```plaintext
GET https://vis.cs.illinois.edu/data/weather-snow/USC00118740.csv
```

**HTTP Response:**

```
HTTP/1.1 200 OK
age: 12278
date: Wed, 06 Apr 2022 15:30:13 GMT
etag: "aa7cc2cf9f22e2dc3bc9e5e1afcf150f"
x-cache: Hit from cloudfront
```

[Line 4]: The etag header:

**Q:** What happens when we request a page again?

```plaintext
GET https://vis.cs.illinois.edu/data/weather-snow/USC00118740.csv
1 GET /data/weather-snow/USC00118740.csv HTTP/1.1
2 [....]
```

**HTTP Response:**

```
HTTP/1.1 304 Not Modified
```

...how much traffic is saved?

...how much time is saved?

**Q:** What are the disadvantages of Entity Tag caching?

**One Additional, Modern Drawback:**
In recent years, regulations have prohibited various tracking using a specific technology called HTTP Cookies. If a developer can’t use a cookie, what else can they use to track you?

Example:

```plaintext
GET /trackingImage.png
```

**HTTP Response:**

```
HTTP/1.1 200 OK
etag: "unique-identifier-just-for-you"
```

When you visit again:

```plaintext
GET /trackingImage.png
1 GET /trackingImage.png HTTP/1.1
2 [....]
```

**Q1:** You need to set up a cloud-scale system that handles 1,000 requests/second. How many EC2 instances do you need if you use no caching? What is the cost?

**Q2:** If you have a cache-hit rate of 70%, how does that change things?

...how does ETag and Age-based differ here?