University of California, Berkeley
Master of Information and Data Science (MIDS)
W205 – Fundamentals of Data Engineering

Week 11 – Web APIs, Part I

Agenda for Today's Class

- Attendance and Participation
- Announcements
- Schedule and Due Dates
- Work / Life / School Balance
- Asynch High Level Review in a Nutshell
- Breakouts
- Summary

Attendance and Participation

Please record your attendance and participation for today's class:

GitHub => ucb_mids_w205_repo => README.md => Attendance and Participation

Announcements

- Upcoming holidays and/or breaks
- Makeup classes for holidays
- Upcoming events
- Student evaluations
- Etc.

Schedule and Due Dates

Take a quick look at the next couple of weeks' due dates:

GitHub => ucb_mids_w205_repo => README.md => Schedule and Due Dates

Work / Life / School Balance
Open Discussion

Student feedback

- About 5 minutes
- How are things going related to work / life / school balance?
- How is w205 going? Difficulty? Time?
- Impact of any natural and/or man-made disasters
- Etc.

Asynch High Level Review in a Nutshell

Each week we will spend about 15 minutes reviewing the most important high level concepts from the asynch

Web Servers

- Simple case
  - User requests a static web page
  - Web server returns a static web page
- Advanced cases
  - Images, audio, video
  - Dynamic content
  - Client-side scripts
  - Client-side scripts which make web API calls

Web Servers (continued)

- URL
  - Uniform Resource Locator
  - `protocol://username:password@hostname:port/directory/file?param1=value1&param2=value2`
- DNS
  - Domain Name System
  - Translates domain name (hostname) to IP address
- Static Content
  - User request file on web server
  - Files changed at server level, not at user request level
- Dynamic Content
  - Content that is dynamically generated when user requests it

Static versus Dynamic

- Static content
  - Very low demands in terms of memory and CPU
  - Single thread can serve thousands of user connections
  - CDN (content delivery networks) are easy to scale out by replicating and pushing static content out to edge servers all around the world
- Dynamic content
  - Extremely high demands in terms of memory and CPU
  - Each dynamic request requires a separate thread of execution
  - Cannot use CDN to scale out
    - We will study techniques to scale out dynamic content later
  - Weak link

APIs

- API – Application Programming Interface
  - APIs allow programmers to write programs to access software systems programmatically
  - Web servers support client-side scripts which can make web API calls
  - A program can make the same API calls as client-side scripts
  - Possible to write web API that is never called by a client-side script, such as phone apps, tablet apps, IoT apps, etc.

Web API Servers

- Web Servers make great scale up Web API Servers
- HTTP – Hypertext Transfer Protocol
  - Runs on top of TCP/IP
  - Common Methods
    - GET – request a resource, possibly with parameters
    - HEAD – get only resource headers
    - POST – like GET but we can also send a payload
      - JSON is typically used for payloads and returned message bodies
      - Binary data, such as images, videos, etc. must be encoded such as MIME using Base64 (uuencoding)
- HTTPS – secure (encrypted) version of HTTP

Breakouts

GitHub => ucb_mids_w205_repo => breakouts

(time permitting, we may not get to all of them)

Summary

Instructor will give a brief (about 2 minute) summary of today’s class.