

Lecture

CS571 - Course Introduction

This content is protected and may not be shared, uploaded, or distributed.

General Information

- Lectures:
 - **Session 30378D: Tue, Th 5:30PM – 7:20PM (SGM 124)**
 - **Session 30016D: Tue, Th 5:30PM – 7:20PM (DEN@Viterbi)**
 - **Session 30222D: Tue, Th 7:30PM – 9:20PM (SGM 124)**
- Producers: Zoom office hours on course website and Piazza
- Course website: **<https://www.csci571.com>**
- DEN's D2L for Zoom Lectures: **<https://courses.uscdcn.net>**
- Assignments – yes, attendance – up to you
- Two **Exams**, one at half course and one towards the end
- Exams auto-graded using **D2L Quiz Tool**
- Mobile Application – final “comprehensive” project live demo
- Web storage: **GitHub Pages**
- Cloud Storage: **Google Cloud, AWS or Azure**

CS571: Web Technologies

- Instructor: Prof. Marco Papa
- Office hours, in-person and on Zoom:
 - **Thursday 4:00PM-5:00PM PDT**
 - **In-person in PHE 516**
 - **Zoom link and iCalendar (.ics) available on D2L**
- E-mail: `papa@usc.edu`
- Zoom Lectures stored on DEN Desire2Learn (D2L)
- 24/7 Q&A access: Piazza, **30 min. average response time**
TA / Producer access: daily (6 days / week)
- Instructor access: weekly
- Quick way to ask a “personal” question: Private message to Instructors on Piazza

General Rules

NO FACULTY D-CLEARANCE

Unless you are a “superstar” undergrad ☺

COURSE OVERLAP

Will allow, but you agree it is “your” problem

Fixed dates / times:

Exam #1: February 21, 6:30PM PDT

Exam #2: April 25, 6:30PM PDT

Final Mobile Project: May 4

Exams electronically graded using D2L Quiz Tool

Final Mobile project/presentation Zoom video uploaded to D2L

Benefits of Prof. Papa CSCI571

- Top 10 students in each section offered available Grader/Producer positions at end of semester. **BY INVITATION ONLY**
- CSCI571 Graders and Producers that are invited to **MS Honors** Program will be provided with **Recommendation Letter** (20 “recommended” students already awarded MS Honors)
- “Proof of skill” for H1B applications available at:

<https://www.cs.usc.edu/skills-verification/>

- Prof. Papa will act as “reference” in job applications for CSCI571 TAs, Course Producers and Graders

Learning Objectives

- By the end of the course, you should have acquired the following skills:
 - The ability to write RESTful API applications
 - The ability to set up Cloud services
 - The ability to design and code back-end scripts in Python and JavaScript
 - The ability to design and code front-end Web Applications
 - The ability to design and code Mobile Apps in Swift/SwiftUI or Java/Kotlin
 - The ability to design and code web front-end asynchronous applications using AJAX
 - The ability to design and code responsive web apps
 - The ability to read Web Services API documentation and use it in building Web applications
 - The ability to write Microservices and Containers

Course Technologies

- This course focuses on the phenomenon known as the World Wide Web
- Core technologies are:
 - HyperText Markup Language (HTML) and Cascading Style Sheets (CSS)
 - HyperText Transfer Protocol (HTTP)
 - Web servers, their configuration and performance properties
 - Server-Side programming using JavaScript and Python
 - Client-side programming using JavaScript and JS Frameworks
 - Ajax Development Style
- Newer Technologies of Interest
 - Responsive Website Design (Bootstrap, etc.)
 - JS Frameworks (Angular, React and Node.js)
 - Web Services (REST)
 - Web security, TOR, Dark web
 - Native Mobile frameworks (Java / Android and Swift / iOS)
 - React (native)
 - Cloud computing (AWS, GCP, Azure)
 - Serverless Applications, Containers, Docker
 - AWS Lambda, Google Cloud Functions, Azure Functions

Software and Storage

- **Student Disk space on GitHub Pages:**
 - `https://pages.github.com/`
 - **GitHub Student Developer Pack:**
`https://education.github.com/pack`
 - Allows GitHub Pages with “private” repository
 - 1GB of free web space
 - Used for homework 2 and 3 and Table of Exercises
- **Website / Web Services in the cloud**
 - Amazon’s Elastic Compute Cloud (AWS)
 - **Google Cloud Platform (GCP) [recommended]**
 - Microsoft Azure
 - AWS Lambda, Google Cloud Functions, Azure Functions
 - Serverless.com
 - Node.js
 - Docker

Other Issues

- Piazza class news group
 - Activate your membership by self-joining at:
piazza.com/usc/spring2023/csci571
 - Class Access Code: **lafc3252usc**
- Academic Integrity Policy
 - Do NOT submit the same program; you can discuss the project with fellow students, but do not develop code with other students; do not download code online; do not post code online; we use MOSS to check for plagiarism. We scan all the exams. See "Academic Integrity Policy".
- DEN's D2L main site for course materials
 - Course slides (PDF)
 - Exams (Quiz Tool)
 - Assignments (Dropbox Folders)
 - Grades
- Downloading course slides and software
 - Class slides access. Username: **csci571**, password: **notes1**
 - All software and installation instructions can be downloaded from the class website.

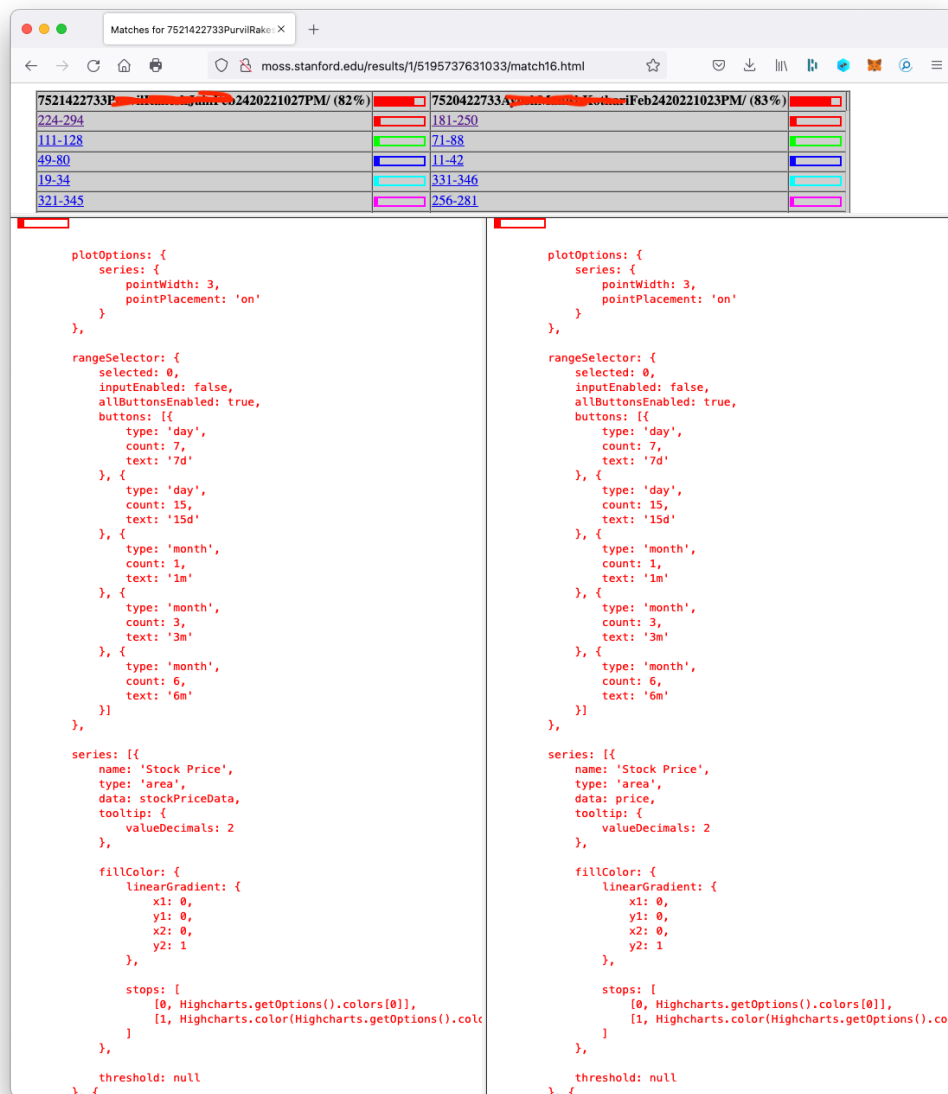
Student Evaluations

- Comments:
 - "Amazing assignments. Learnt a lot on the course."
 - "Projects seemed similar to an actual client for web development would ask for."
 - "Even though the assignments were hard I learnt a lot from them."
 - "It is not a fair game for beginners. I've spent almost 3 weeks to do a homework, and I still can not finish it on time."
 - "Course projects are impressive!"
 - "This class has posted assignments easily x10 times larger than other classes."
 - "HW8 and HW9 take *forever*."
 - "The homework assignments are so difficult."
 - "I had to do so much googling on my own to learn about concepts used in the homework assignments."
 - "Tough class with a lot of valuable assignments."
 - "Massive assignments."
 - "The workload of this course is too much, especially the last two homework."

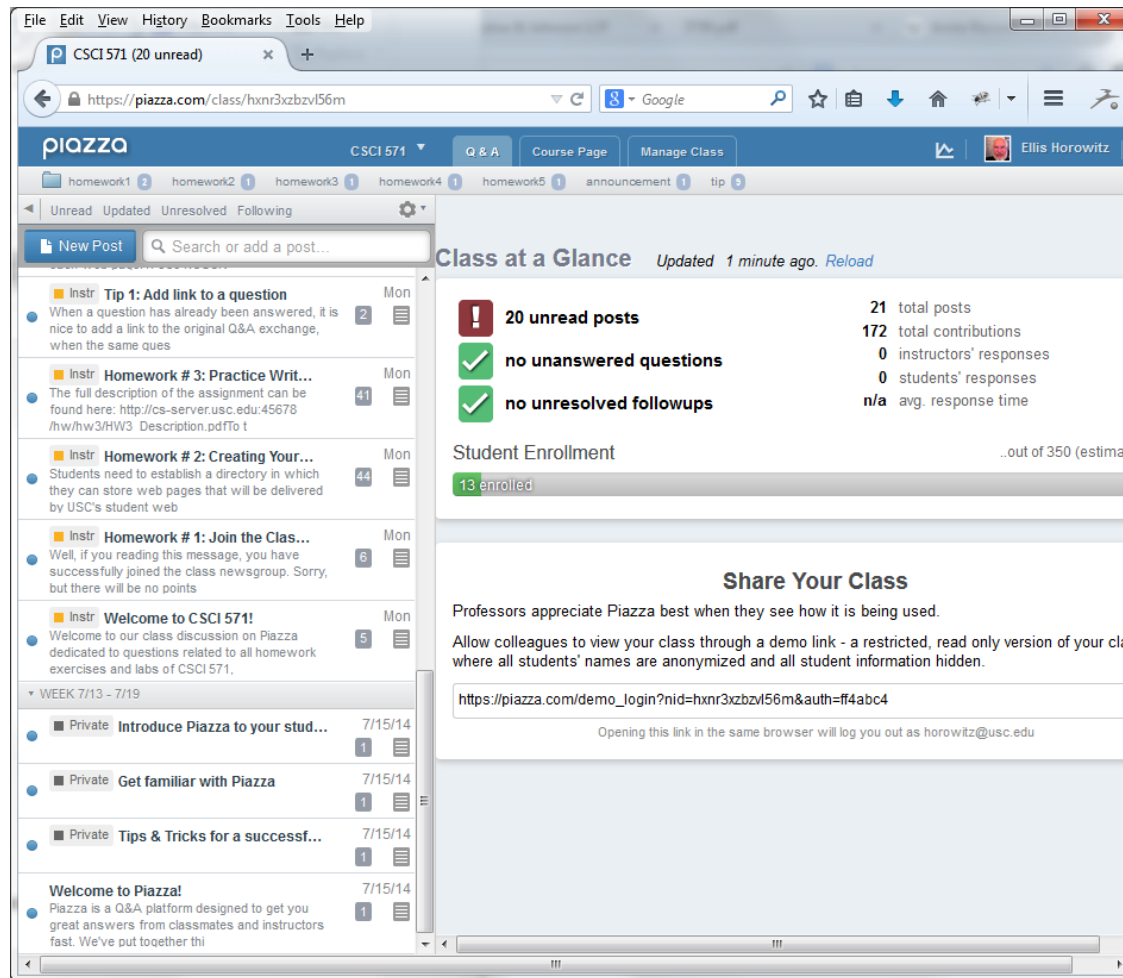
Academic Integrity Violations

- Spring 2018 violations (16):
 - Sanctioned: 16
 - Appeals to Engineering Panel Review: 8, no changes
 - Appeals to Engineering Dean: 2, no changes
 - F in course: 12, 0 + full letter grade reduction: 4 (C-, C-, C-, B-)
- Fall 2018 violations (0):
 - None!
- Spring 2019 violations (4):
 - F in course: 1
 - 0 + full letter grade reduction: 1 (B-)
 - C in the course: 2
- Fall 2019 violations (6):
 - F in course: 1
 - 0 + full letter grade reduction: 5 (B-, C)
- Spring 2020 violations (4):
 - 0 + full letter grade reduction: 1 (A-, C)
- Spring 2021 violations (2)
 - Full letter grade reduction
- Spring 2022 violations (52)
 - F in course: 2, 0 in assignments, C, and D in course
- Fall 2022 violations (5)
 - 0 in assignments, F in course: 1, point deductions



MOSS Plagiarism discovery tool



Piazza



Who am I?

- PhD in CS from USC, class of '88
- Initial career: MS Windows, Commodore Amiga developer
- System Architect -> Team Lead -> IT Project Manager -> VP Engineering
- CTO (Chief Technology Officer) at Luckman Interactive and CareerBuilder
- Part-time Faculty at USC 2003-2022
- **Full-time Faculty at USC starting in August 2022**
- Chief Technologist at LASC (Los Angeles Superior Court) 2002-2022
- Microsoft 365 SharePoint Project Manager 2014-2022
- Active member of LA CTO Forum (Silicon Beach CTOs)
- Season ticket holder of USC Football and  LAFC (Los Angeles Football Club) and  supporter

Characterizing Web Content

There are very few studies that examine the types of content on the web, however . . .

(From IEEE Spectrum, Jan. 2004, pp. 75):

- Claim: 30% of the web is porn
- Claim: 30% of the web is duplicate information
- 50,000,000 pages are either new or changed *each day*
- 65% of the web pages are in English

(From Personal Computer World, Optenet, Sep. 2008):

- Claim: 35% of the web is porn, 11% is e-commerce
- <http://www.optenet.com/en-us/new.asp?id=162>

(From Forbes, Sept. 2011):

- Claim: 4% is porn, 13% are porn Web Searches
- <http://www.forbes.com/sites/julieruvolo/2011/09/07/how-much-of-the-internet-is-actually-for-porn/>

(From BBC, July 2013):

- Claim: is porn 4% or 37%?
- <http://www.bbc.com/news/technology-23030090/>

Sample Web Sites (Modest Size)

- Running a web site can get complicated; here is one example.
- The facts:
 - **www.fogdog.com**, online sale of sporting goods (domain abandoned)
 - Revenues: \$5 million per year
 - 2.2 million-page views per month
 - average of 20,000 unique visitors per day
- The solution (in-house):
 - Commodity hardware
 - Linux server running Apache 2.0 web servers
 - Using MySQL data base
 - They **moved to Ebay!**
 - <https://www.ebay.com/str/fogdog>
 - F5 BIG-IP OS, Apache 2.0.64 web server

Sample Web Sites (Medium size)

- Here is a popular, alternate strategy for maintaining a web site
- The facts:
 - **www.autobytel.com**, new and used cars (now **AutoWeb**)
 - Market Cap: \$33.92M (Dec. 2019)
 - Quarterly Revenues: \$28.6M (3rd Quarter 2019)
 - Lead traffic: 31.7M visits (3rd Quarter 2019)
 - Mobile version launched in 2012
 - Stock symbol: **AUTO** (Nasdaq)
- Original Microsoft solution:
 - Microsoft Windows Server
 - Microsoft IIS 7.5 web server
 - Microsoft SQL server database
 - Akamai CDN
- Today:
 - Windows Server, Microsoft IIS/7.5 web server

Sample Web Sites (large size)

- The facts:
 - **www.etrade.com**, online investing services and resources
 - Market Cap: \$10.37B (Dec. 2019)
 - Yearly Revenues: \$2.9B (12/2018)
 - 60 million-page views per month
 - average of 53,000 unique visitors per day
 - 4.9 million accounts (Jan. 2015)
 - 25,000 new retail accounts opened (Oct 2015)
 - 1,952,000 customer transactions per month
 - Taken over by Morgan Stanley in 2020 for \$13 billion
- The solution:
 - IBM 90 xSeries running Linux/**Citrix Netscaler, Apache** and Tomcat web servers, AWS Route 53 (DNS)
 - Hardware facility for load balancing and redundancy
 - Oracle database system
 - Proprietary programming systems

Web Server Farms


























- Until recently all serious web sites were maintained using web server farms;
 - A group of computers acting as servers and housed in a single location;
 - Internet Service Providers (ISP's) provide web hosting services using a web server farm
- Hardware and software is used to load balance requests across the machines
- Other issues addressed by web server farms include:
 - **Redundancy** eliminates single point of failure; backup and failover strategy is required
 - **Security**, secure areas are placed behind firewalls which monitor web traffic, network address translation, port translation, SSL

Popular Web Hosting Services

- ***For individuals and small business:***
 - **1&1**
<https://www.ionos.com>
 - **GoDaddy.com**
<http://www.godaddy.com/products/secure-hosting.aspx?ci=72738>
 - **Yahoo**
<http://www.iwebhostingplans.com/yahoo/yahoowebhosting.asp>
- ***For companies willing to pay MUCH higher costs:***
 - **Rackspace**
http://www.rackspace.com/index.php?CMP=Google_hosting
 - **Network Solutions**
<http://www.networksolutions.com/web-hosting/index.jsp>
- ***Reviews and price comparisons:***
 - <https://www.hosting-review.com/best-web-hosting/>
 - See next slide “monthly” prices
 - <http://www.pcmag.com/category2/0,2806,2269,00.asp>

Web Hosting Services

TOP 10 WEB HOSTING PROVIDERS - Updated May 2022

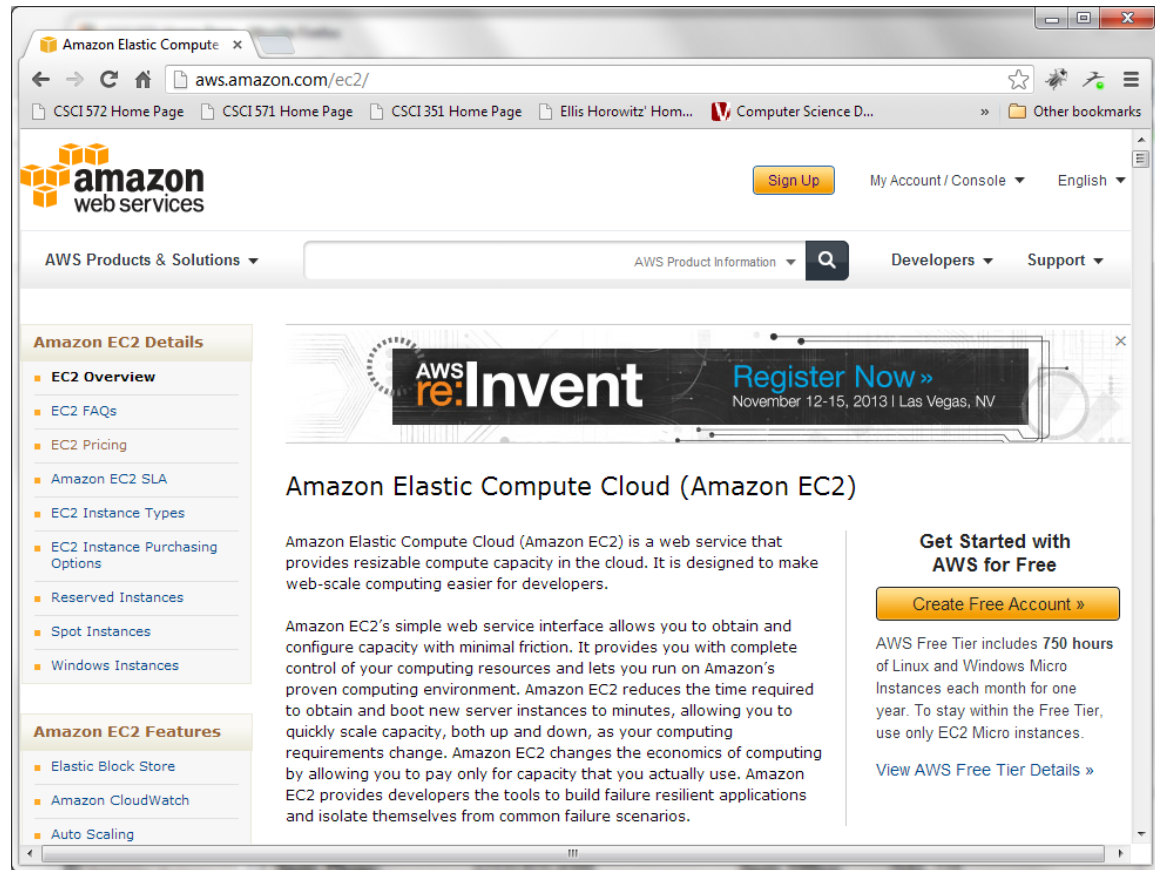
RANK	SHARED WEB HOST	PRICE	SALES INDEX	TREND	UPTIME SPEED	CUSTOMER REVIEWS	EDITOR'S REVIEW	VISIT
1	 MORE ▾	\$3.95		↑		★★★★★	 Read	Visit Site
2	 MORE ▾	\$3.25		↑		★★★★★	 Read	Visit Site
3	 MORE ▾	\$3.95		↑		★★★★★	 Read	Visit Site
4	 MORE ▾	\$4.99		↑		★★★★★	 Read	Visit Site
5	 MORE ▾	\$6.29		↑		★★★★★	 Read	Visit Site
6	 MORE ▾	\$2.95		↑		★★★★★	 Read	Visit Site
7	 MORE ▾	\$3.95		—		★★★★★	 Read	Visit Site
8	 MORE ▾	\$3.99		—		★★★★★	 Read	Visit Site
9	 MORE ▾	\$6.99		↑		★★★★★	 Read	Visit Site
10	 MORE ▾	\$3.92		↑		★★★★★	 Read	Visit Site

Cloud Computing

- **Cloud computing** is Internet-based computing, whereby shared resources, software, and information are provided to computers and other devices **on demand**, like the electricity grid.
- Users no longer have need for expertise in, or control over, the technology infrastructure "in the cloud" that supports them.
- It typically includes web-based tools or applications that users can access and use through a web browser as if it were a program installed locally on their own computer. ¹
- Typical cloud computing providers deliver common business applications online that are accessed from another Web service or software like a Web browser, while the software and data are stored on servers.
- The major cloud service providers include Amazon, Google, Microsoft, Salesforce, Skytap, HP, IBM, Amazon, Google and Apple (iCloud).

An Example – Amazon’s Elastic Compute Cloud

- A web service providing resizable compute capacity
- The “elastic” nature means the service instantly scales to meet demand with no up-front investment
- Users create an Amazon Machine Image (AMI), a virtual computer running your selected operating system (Linux, Windows, etc)
- Users use Amazon’s Simple Storage Service (S3) for large-scale, persistent storage
- You only pay for running AMI
- All accounts are limited to 5 Elastic IPv4 addresses per region
- See: aws.amazon.com/ec2

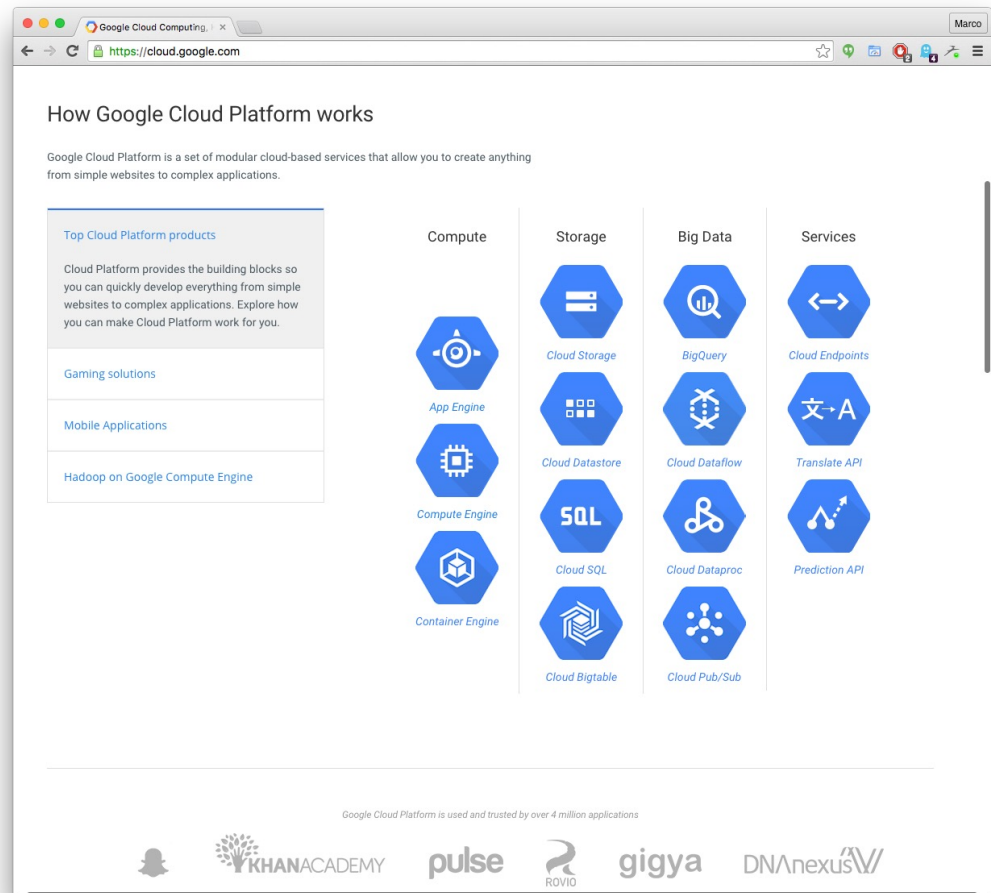


Amazon currently runs in 8 regions: US East, US West (Oregon), US West (Northern CA), Ireland, Asia Pacific (Singapore), Asia Pacific (Tokyo), Asia Pacific (Sydney), South America (Sao Paulo)

An Example - Google Cloud Platform

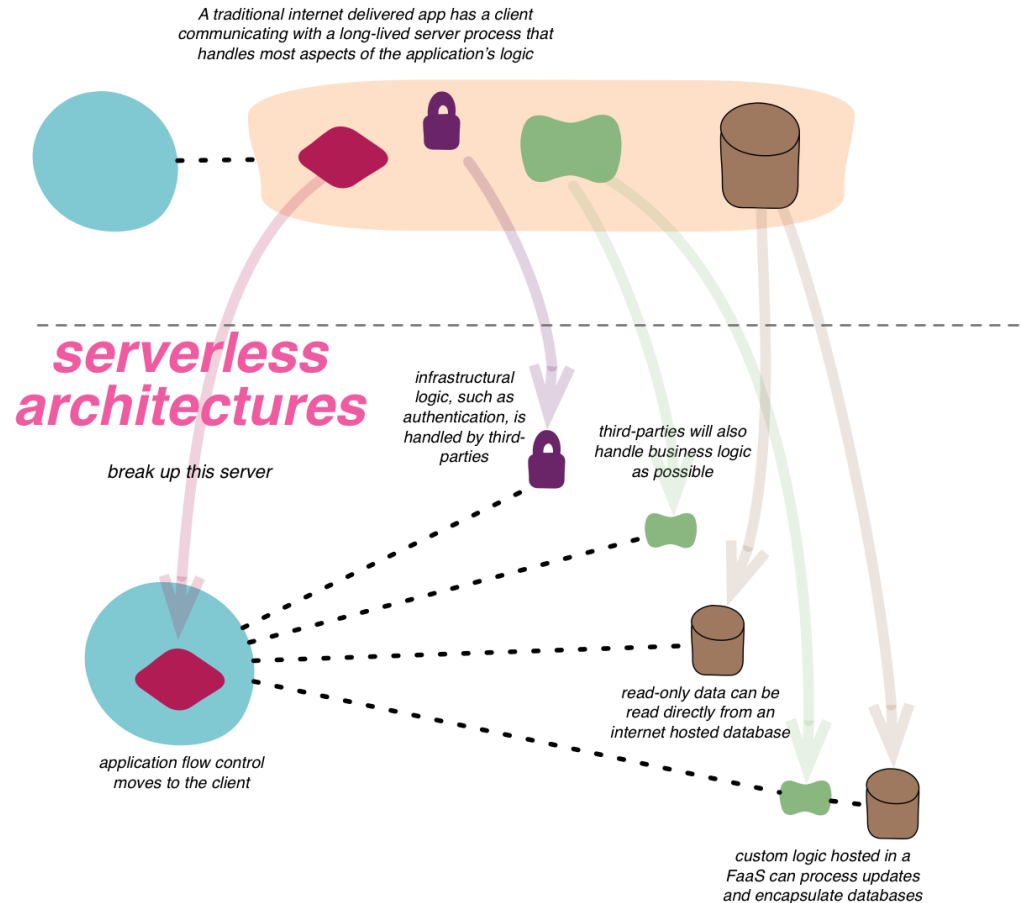
- A web service providing basic Compute, Storage, Big Data and Services.
- Additional services for massively scalable Gaming solutions, Mobile Applications backend, and Apache Hadoop.
- App Engine – A platform for building scalable web applications and mobile backends. App Engine scales applications automatically in response to the amount of traffic it receives.
- Compute Engine - Offers predefined virtual machine configurations: Debian, CentOS, CoreOS, SUSE, Ubuntu, Red Hat, FreeBSD, or Windows 2008/2012.

Google uses software-defined networking technology to route packets across the globe and enable fast edge-caching so that data is where it needs to be to serve users.

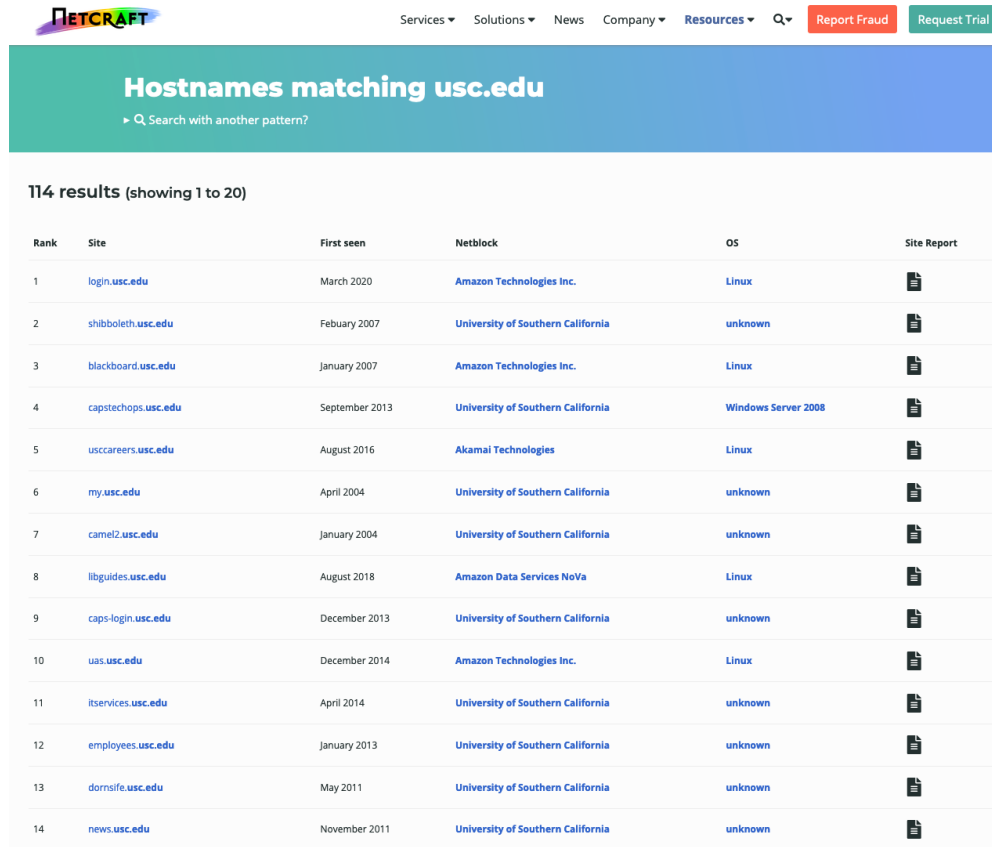


Serverless Architecture

- Internet based systems where the application development does not use the usual server process.
- They rely solely on a combination of:
 - third-party services, or Backend as a Service (BaaS)
 - client-side logic
 - service hosted remote procedure calls, or Function as a Service (FaaS).
- AWS Lambda is one of the most popular implementations of FaaS at present, but there are others. See: <https://aws.amazon.com/lambda/>
- Serverless and containers will be covered later in the course



USC Has Many Web Servers Running



The screenshot shows the Netcraft website interface. At the top, there's a navigation bar with links: Services, Solutions, News, Company, Resources, a search icon, and buttons for 'Report Fraud' and 'Request Trial'. Below this is a blue header section titled 'Hostnames matching usc.edu' with a sub-link 'Search with another pattern?'. The main content area shows '114 results (showing 1 to 20)'. A table lists the first 14 results, including columns for Rank, Site, First seen, Netblock, OS, and a Site Report icon.

Rank	Site	First seen	Netblock	OS	Site Report
1	login.usc.edu	March 2020	Amazon Technologies Inc.	Linux	
2	shibboleth.usc.edu	February 2007	University of Southern California	unknown	
3	blackboard.usc.edu	January 2007	Amazon Technologies Inc.	Linux	
4	capstechops.usc.edu	September 2013	University of Southern California	Windows Server 2008	
5	usccareers.usc.edu	August 2016	Akamai Technologies	Linux	
6	my.usc.edu	April 2004	University of Southern California	unknown	
7	camel2.usc.edu	January 2004	University of Southern California	unknown	
8	libguides.usc.edu	August 2018	Amazon Data Services NoVa	Linux	
9	caps-login.usc.edu	December 2013	University of Southern California	unknown	
10	uas.usc.edu	December 2014	Amazon Technologies Inc.	Linux	
11	itservices.usc.edu	April 2014	University of Southern California	unknown	
12	employees.usc.edu	January 2013	University of Southern California	unknown	
13	dornsife.usc.edu	May 2011	University of Southern California	unknown	
14	news.usc.edu	November 2011	University of Southern California	unknown	

- Netcraft lists **112** separate sites / web servers with usc.edu in their name, e.g.
- `www.usc.edu`
- `mat.usc.edu`
- `www.cs.usc.edu`
- `dornsife.usc.edu`
- `web-applusc.edu`
- `www-scf.usc.edu`
- However, some may not be connected to USC, e.g.
- `www.usc.edu.au`
- Check at:
`https://searchdns.netcraft.com`

Web Browsers Use Standard Layout Engines

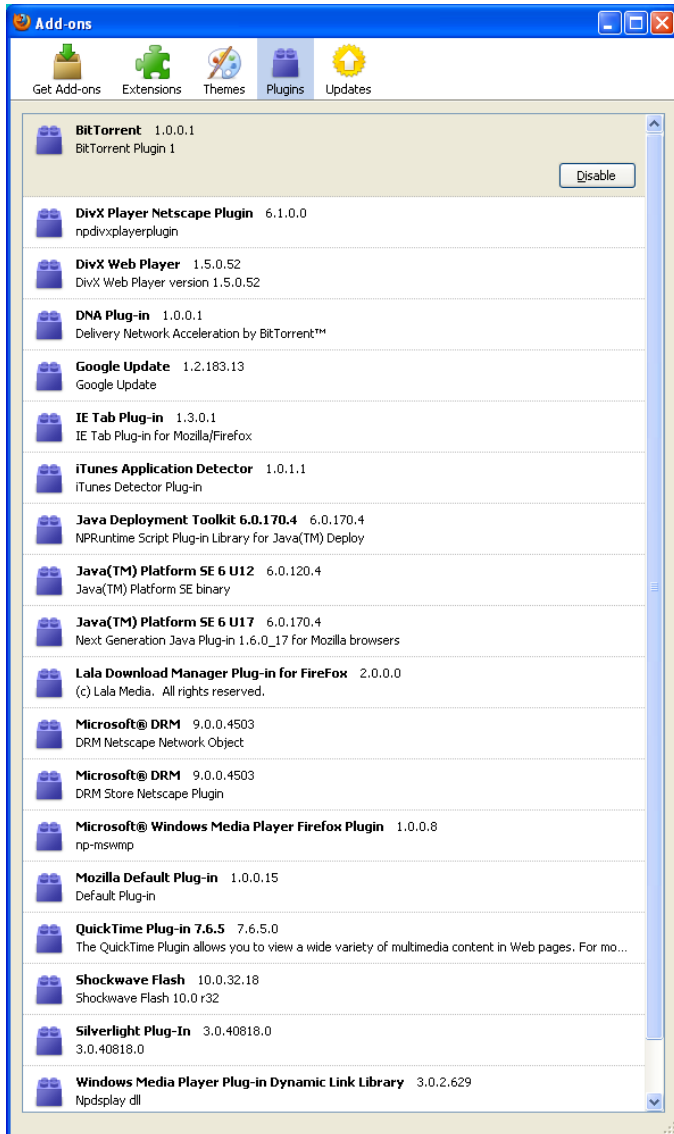
- **WebKit** is a software component used to render web pages; it is open source.
 - It is used by Google's Chrome and Apple's Safari web browsers
 - WebKit is also the name of the Mac OS X system framework version of the engine that's used by Safari, Dashboard, Mail, and many other OS X applications;
- **Gecko** is a layout engine developed by Mozilla Corporation, known as the layout engine of the Firefox web browser.
 - It is used to display web pages and, in some cases, an application's user interface.
 - It offers a rich programming API that makes it suitable for a wide variety of roles in Internet-enabled applications, such as web browsers
 - Its development originated with Netscape Communications Corporation
- Some web kits and the browsers that use them
 - **Gecko-based:** FireFox (Mozilla), Flock, Netscape
 - **Trident-shells:** Internet Explorer (Microsoft, retired)
 - **Edge:** Edge Legacy (Microsoft), fork of Trident 7, retired in Aug. 2021
 - **WebKit-based:** Chrome and Android (Google), Midori, Safari and Mobile Safari (Apple), Symbian^3 (Nokia) and many others
 - **Chromium-based:** Chrome, Microsoft Edge (2021)
 - **Presto-based:** Opera, Nintendo DS, Opera Mini, Opera Mobile
 - **Java-based:** HotJava, Lobo

Capabilities of a Browser

- Web browsers fetch and display documents from other WWW sites; their capabilities include:
 - A mouse-driven graphical user interface
 - Display of
 - Hypertext documents conforming to latest HTML standard
 - Text with fonts, styles, and varying point sizes
 - Foreign-language character sets conforming to ISO-8859
 - Forms composed of edit boxes, check boxes, radio boxes, lists, text areas, etc.
 - Graphics in different formats (GIF, JPEG, MPEG, PNG, XBM) including monochrome, color

GIF = graphic interchange format, MPEG = Motion Picture Experts Group, JPEG = Joint Photographic Experts Group, PNG = Portable Network Graphics, XBM = x bitmap

Capabilities of a Browser



- Ability to invoke helper applications and plug-ins, (**Obsoleted in HTML5**) e.g.
 - *Adobe Acrobat* - used to view pdf files
 - *Windows Media Player* to play digital sound files
 - *Adobe Flash Player*, used to display video. **Retired in 2020.**)
- Ability to communicate over a secure channel, using SSL
- Ability to maintain and exchange digital certificates
- Ability to run scripts in JavaScript
- Ability to run *Java applets* and *Active X* components (**also obsoleted in HTML5**)

The Browser Wars - Desktop Statistics

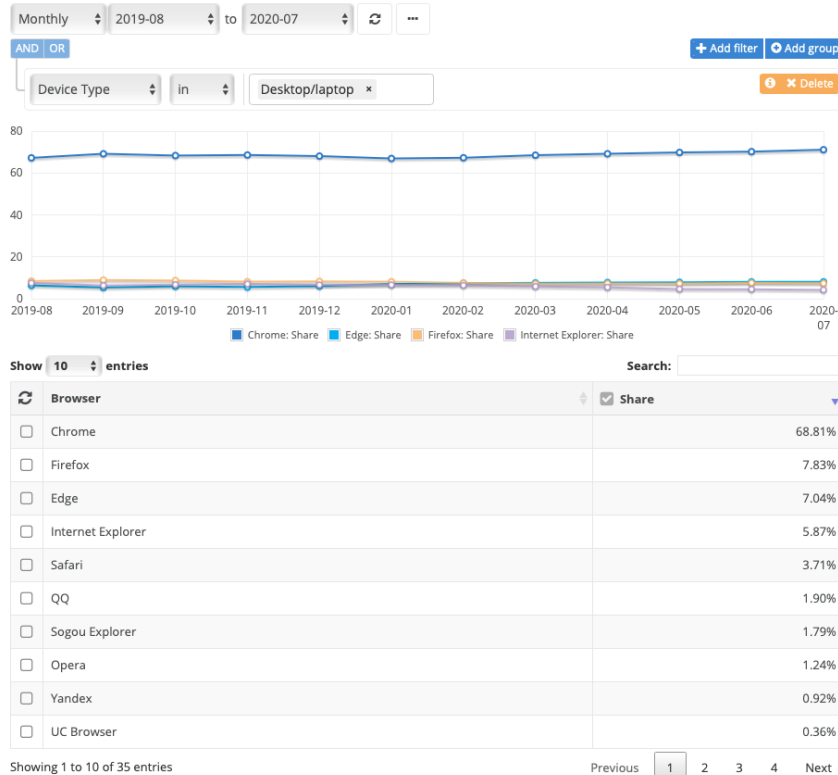
2022	<u>Chrome</u>	<u>Edge</u>	<u>Firefox</u>	<u>Safari</u>	<u>Opera</u>
October	79.9 %	8.1 %	5.2 %	4.2 %	1.7 %
September	80.9 %	7.8 %	5.2 %	3.7 %	1.5 %
August	81.1 %	7.6 %	5.2 %	3.4 %	1.7 %
July	81.1 %	7.5 %	5.0 %	3.4 %	2.1 %
June	76.3 %	7.4 %	5.1 %	3.6 %	2.3 %
May	79.9 %	7.3 %	5.3 %	3.8 %	2.4 %
April	80.3 %	7.2 %	5.3 %	3.8 %	2.4 %
March	80.3 %	7.5 %	5.3 %	3.7 %	2.3 %
February	79.9 %	7.5 %	5.4 %	4.0 %	2.3 %
January	80.1 %	7.3 %	5.5 %	3.9 %	2.3 %
2021	Chrome	Edge	Firefox	Safari	Opera
December	81.0 %	6.6 %	5.5 %	3.7 %	2.3 %

Conclusion of the above study:
- Chrome is the clear winner
- Firefox comes second, but losing ground
- Edge next
- Safari and Opera having small percentages
- WebKit total over 84%

- See http://www.w3schools.com/browsers/browsers_stats.asp
- See also <http://www.upsdell.com/BrowserNews/stat.htm>

Desktop/laptop Browser Market Share Statistics

Browser Market Share

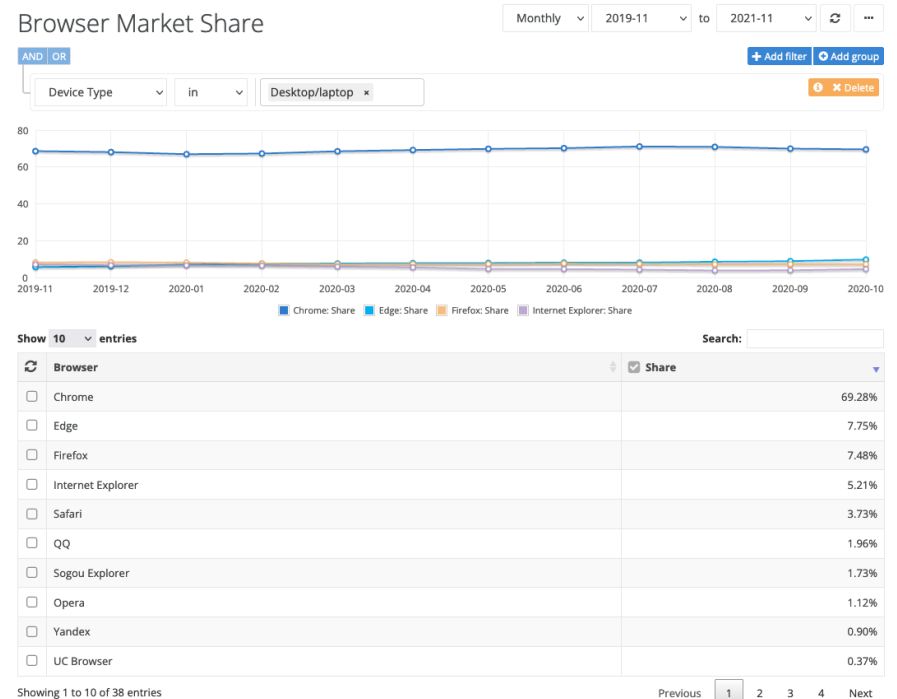


August 2020

<http://www.netmarketshare.com/> [retired]

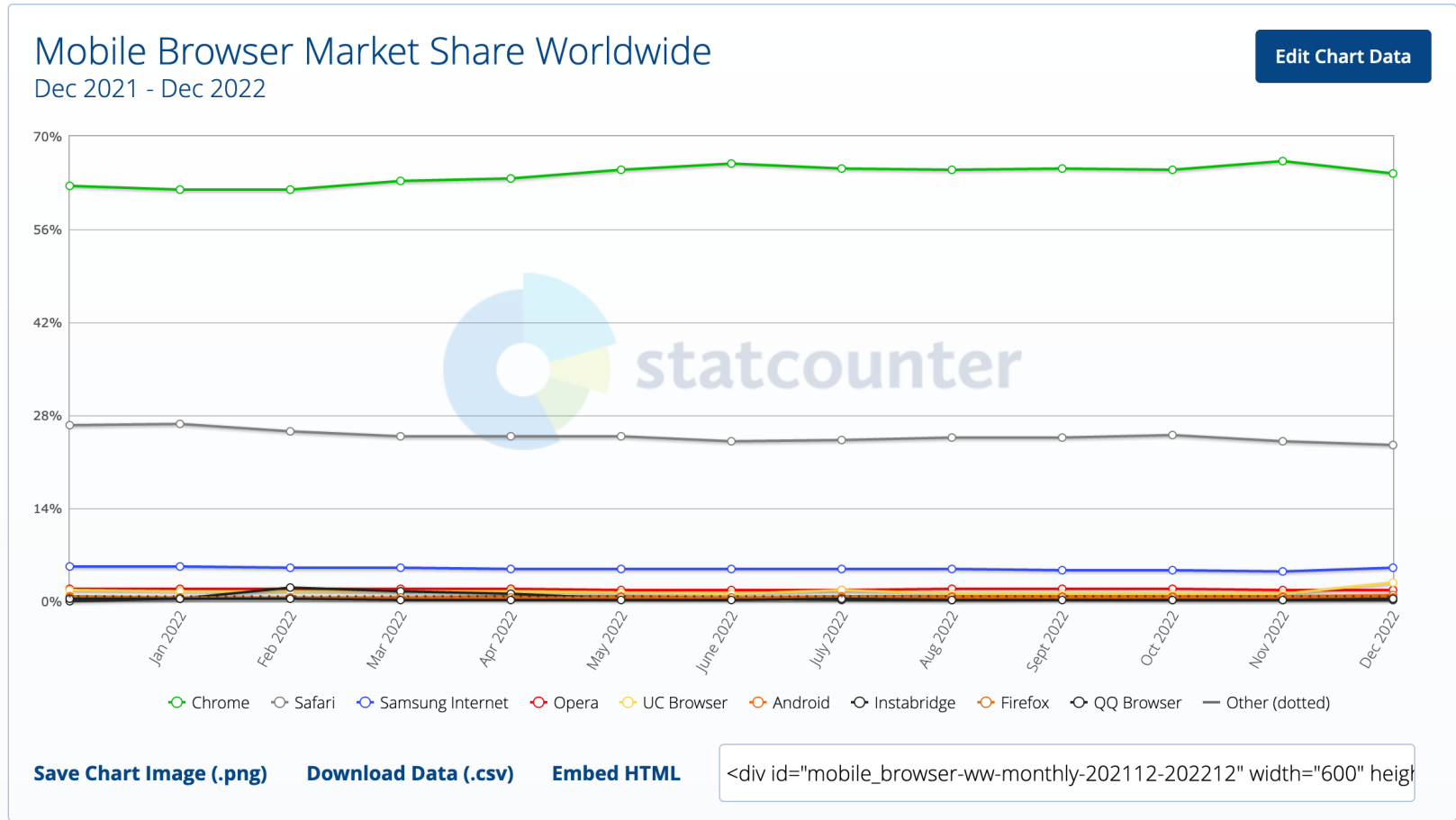
Chrome leads with 69% market share.

Browser Market Share



December 2021

The Browser Wars Comparison(cont'd)



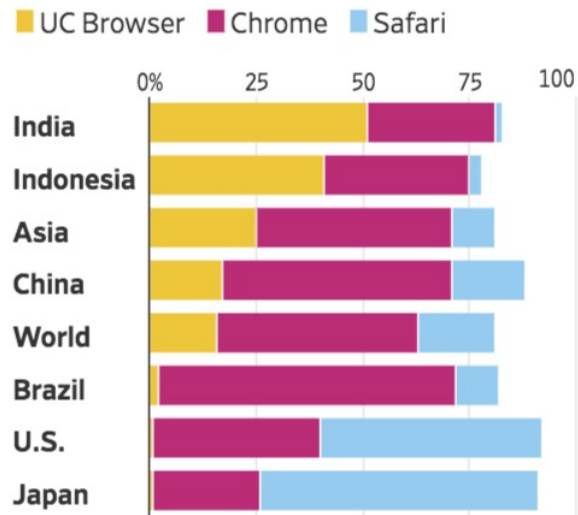
StatCounter Global Stats, Dec 2021 - Dec 2022, See <http://gs.statcounter.com>
Chrome has the lead with about 60%, followed by Safari at 24%, Samsung at 6% and UC Browser from UCWeb of Alibaba Group of China at 3%.

The Browser Wars Comparison (cont'd)

Browser for the Next Billion

Alibaba's mobile browser, UC Browser, has a larger market share than Google's Chrome in India and Indonesia, where many of the world's 'next billion users' are getting online for the first time.

Mobile web browser market share

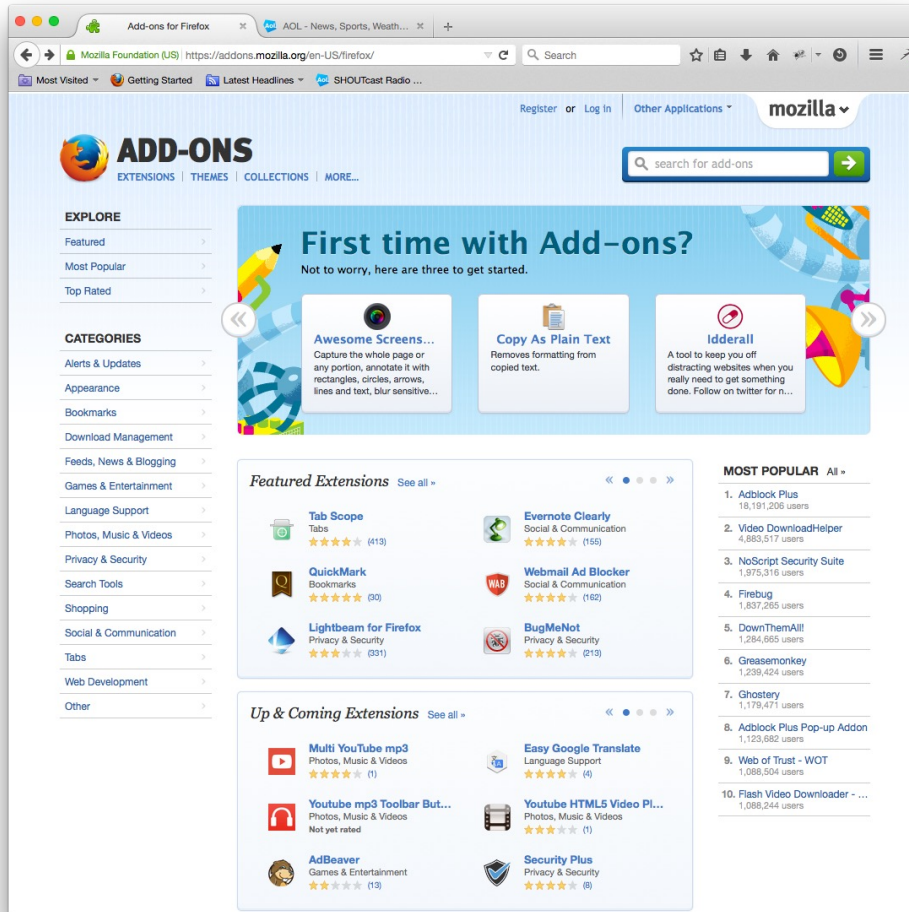


Note: Data Oct.-Dec. 2016 through Oct.-Dec. 2017.

Source: StatCounter

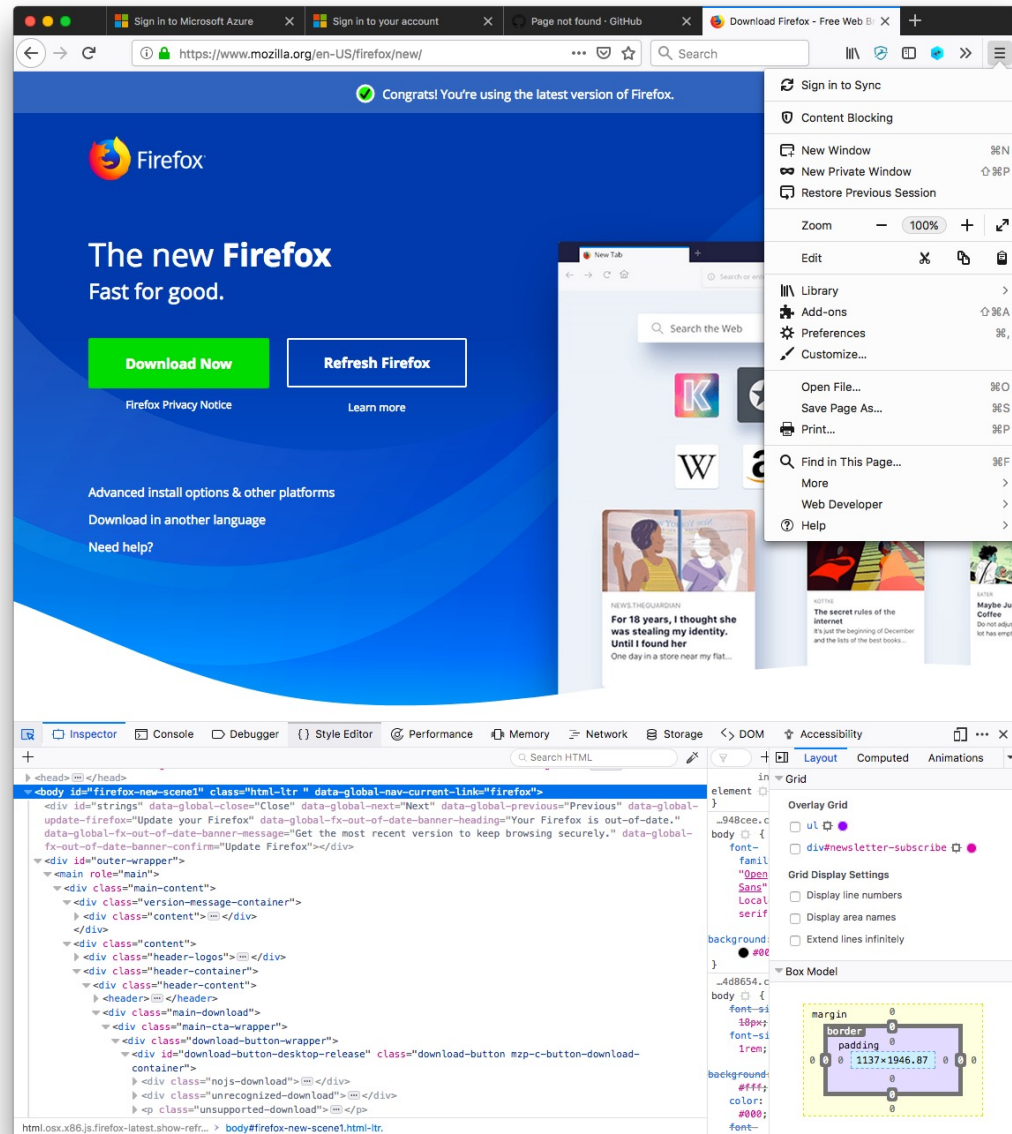
StatCounter Global Stats, Oct.-Dec. 2016 through Oct.-Dec. 2017, See <https://www.wsj.com/articles/a-browser-youve-never-heard-of-is-dethroning-google-in-asia-1514808002>

Browsers Have Many Plugins Available

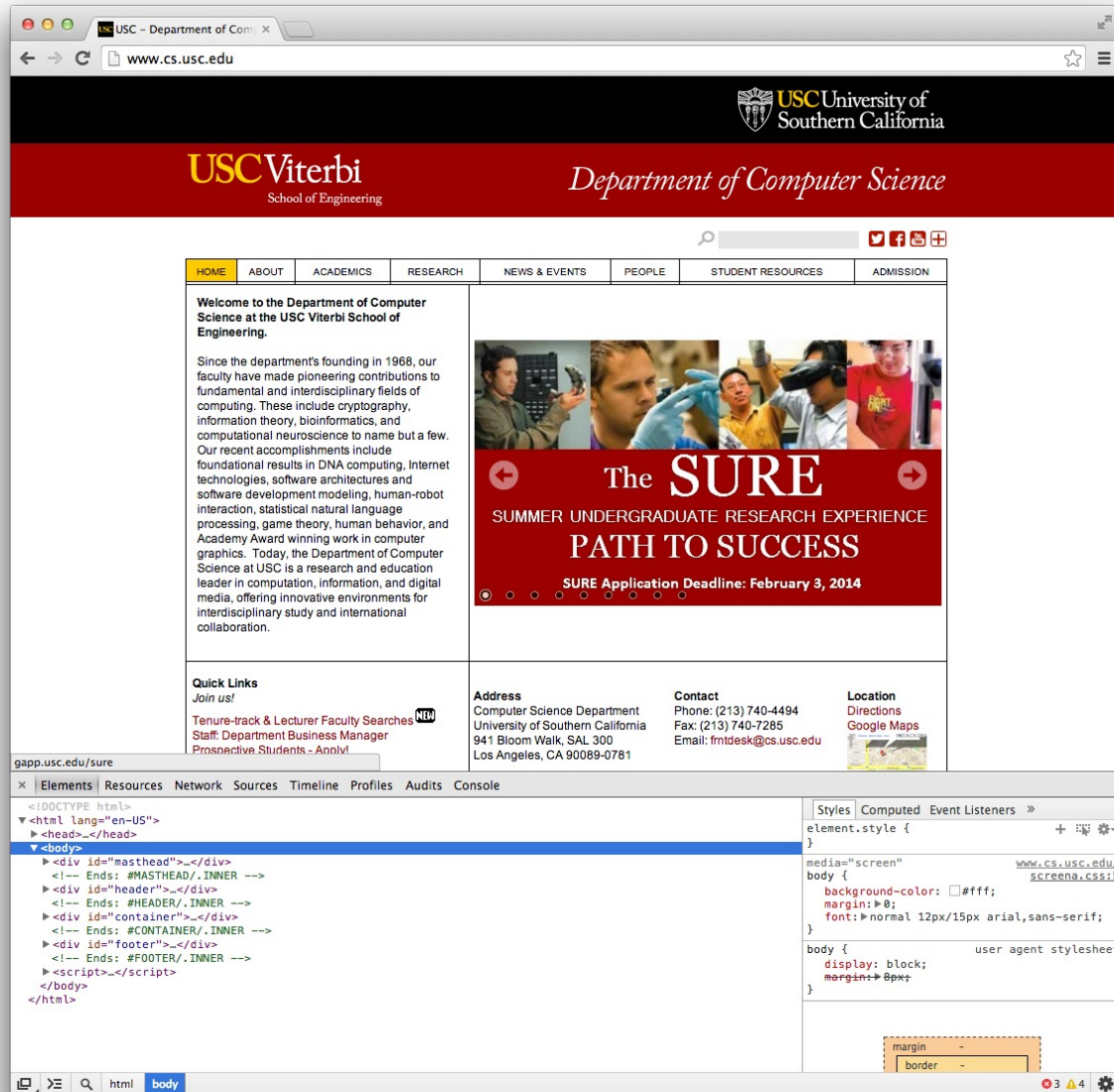


- Plug-ins are also listed as “Extensions”
- Firefox plug-ins that will be especially useful in this course are:
 - HTTP Header Live
- HTML5 does away with most video / audio plug-ins

Firefox: Tools | Web Developer



Chrome: Menu | More Tools | Developer Tools



Evolution of Web Sites

Client-centric Static	Server Applications Databases Dynamic web pages	Web services Multiple layers Business and service Integration	Service Oriented Arch. (SOA) Client-centric	Multi-platform (desktop, tablet, phone) Client-centric	IoT, Wearables, Cloud computing, Serverless Arch. (BaaS, FaaS) Docker
HTML Scripts CGI	ODBC, JDBC ASP Applets, ActiveX	XML, WML, SQL, .NET COM+, Beans	Ajax, Web 2.0, JSON	HTML5, CSS3, JS gestures navigation	JS Frameworks AWS, GCP, Azure Microservices containers
1 st gen	2 nd gen	3 rd gen	4 th gen	5 th gen	6 th gen
1991	1997	2000	2005	2008	2014