Special Topics:
CSci 8980
Trends in edge computing

Jon B. Weissman (jon@cs.umn.edu)

Department of Computer Science
University of Minnesota

Introduction

- Introductions – all
- Who are you?
- What interests you and why are you here?

Introduction (cont'd)

- What is this course about?
  - New field of Edge Computing
- Is it new? Yes and no.
  - Why no? Examples?
  - New technologies, new applications => "first class" entity

Emergence of Edge Computing

- Satya notes:
  - Centralized in the 60s, 70s
  - Distributed in the 80s, 90s
  - Centralized again in the 2000s
  - Distributed again

Web Site

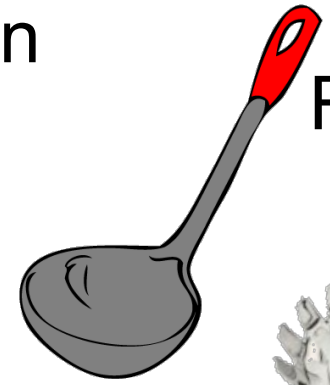
- http://www-users.cselabs.umn.edu/classes/Fall-2021/csci8980-ec/

The “Standard” Cloud

Data in

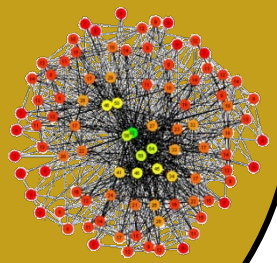
Computation

Results out

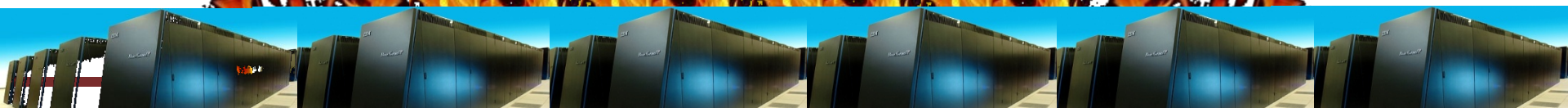


- “No limits”
  - Storage
  - Computing

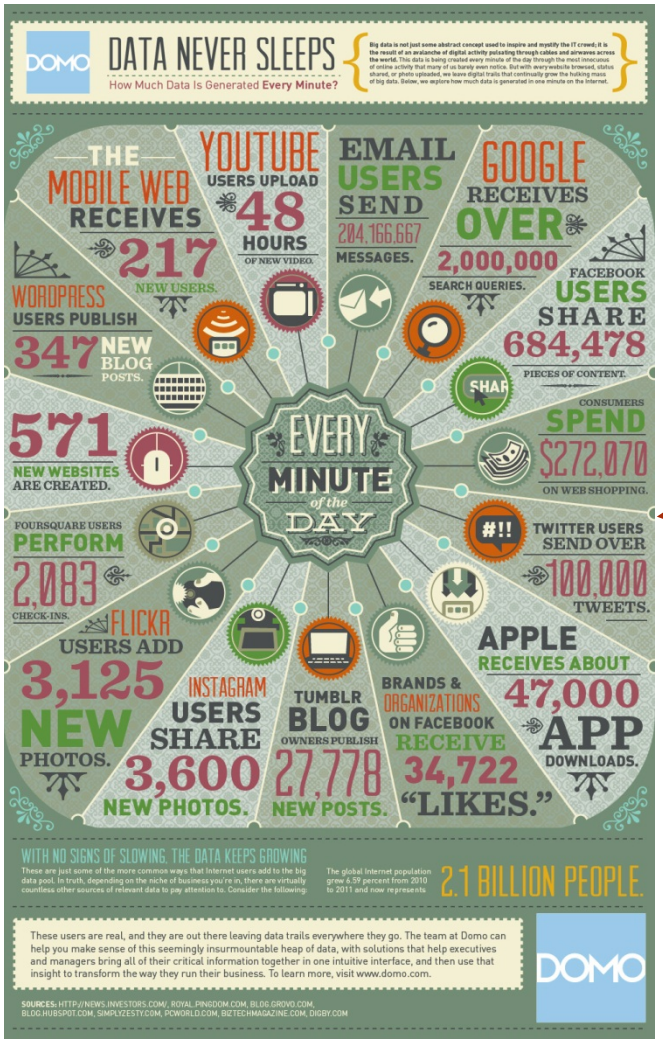
Weighted						
HI	min	max	mean	std_dev	count	median
0	0	108087	22708	15364	923	19943
1	0	34808	19509	9442	10	21786
2	0	25795	17005	8115	7	18656
3	14464	30102	21355	7982	3	19500
4	9427	59183	33118	13457	21	30795
5	4999	42458	21992	8208	39	21176
6	10857	22386	17711	4565	9	20400
7	9418	30347	21344	8679	9	26197
Non-weighted						
HI	min	max	mean	std_dev	count	median
0	0	108087	22979	15364	863	20208
1	0	64962	21247	14022	76	19932.5
2	8770	55315	22046	11355	22	19054
3	0	33886	21825	10062	14	23600.5
4	15383	40588	28753	12672	3	30288
5	8414	42458	21541	8094	29	20833
6	4999	37051	21706	8457	12	20502
7	7927	26821	17374	13360	2	17374



VISA



Big Data: hyper-exponential growth



Key observation: much data originates not in the data center but at the "edges"

100s of zettabytes

Mobile web
Sensors
IoT devices
Automobiles
Etc.

Technical Course Goals

- Learn about current state-of-the art in Edge Computing interpreted broadly
- Identify general problems and solutions
- Relate to classic problems and solutions in distributed systems, DB, networking, etc.

Non-Technical Course Goals

- Learn how to read papers and ask technical questions
- Learn how to present papers and lead discussions
- Do a team research project
  - Idea formation
  - Writeup
  - Experiment
  - Present
  - (fingers-crossed) publish a (workshop) paper

Major Topics

- Fault tolerance
- Edge networking
- Machine learning at the edge
- Edge security
- Geo-distributed edge computing
- Edge computing systems
- Edge video and streaming
- Edge applications
- There is some overlap!

Course structure

- Grading …
  - Presentations: 40%
    - 42/N papers; right now N=9
  - Take-home mid-term: 10%
  - Final project: 30%
  - Questions: 10%
    - 42/N (N=9) papers
  - Discussions: 10%

Presentations

- Two presentations per class
- Give paper's context and background
- Key technical ideas
- It's relation to other papers or ideas
- Positive/Negative points (and why)
- 25 min max to leave time for discussion
- Keep it interesting!
  - tough job: don't want gory paper details nor do we want total fluff
  - audience: smart CS/EE students and faculty
  - demos are great!!

Visit web site

Presentations (cont'd)

- Discussion questions for you and questioner
  - go beyond the claims in the paper
  - limitations, extensions, improvements
- You may find .ppt slides online BUT
  - put it in your own words
  - understand everything you are presenting

Projects

- Talk about ideas in a few weeks …
  - present a list of things that are useful, open to other ideas
  - implementation-based
- Work in a team of 1 or 2
- Risk encouraged … and rewarded (even if you fall short)

Projects (cont'd)

- Implementation project done on cloud and/or "edge" – we will interpret edge broadly
  - Project proposal presentation will be at the end of October; can get feedback from me before then
- May present project status to the class later (forcing function)
- Will present final results at the end

Near-term Schedule

- website

- Need volunteers for upcoming papers starting next week (see ? next to papers on the website)
  - I will hand-pick “volunteers” if necessary ☺
  - I will pick questioners

Modality

- First, since we are small you must come to all meetings unless you have a *compelling* reason to miss

- Second, we will experiment with both live and zoom, and see how things go

Admin Questions?

The Edge: Gentle Intro

- The Emergence of Edge Computing

  Satya

- Edge Computing: Vision and Challenges

  Shi, and others

Taxonomy-1

- Far edge: sensor/IoT, human
  - very limited networking
- The “edge”
  - local compute, storage
  - 1 hop to far edge, Internet connected
- Local cloud
  - Collection of edge nodes
- Centralized cloud

Taxonomy-2

`far-edge` (data gen: sensors, actuators, cars, robots, human)
→ `near-edge` (carried: phones, tablets, wearables)
→ `localized` (infra: one-hop server)
→ `micro-DC` (infra: close-by resources, fog)
→ `geo-distributed-edge` (infra: WAN)
→ `central-cloud` (infra: WAN)

- Notion of far/near edge, localized/micro-DC may be blurry

Why the shift again?

- Centralization => Dispersion

- Reasons
  - Proximity/latency: highly responsive cloud services/applications (e.g. AR, VR, *cognitive assistance*)
    - Low latency, high b/w, low jitter
  - Scalability via edge analytics
    - Local processing of high b/w sensors (e.g. cameras, cars)
  - Privacy enforcement
    - First point of contact between far edge and system
  - Masking cloud outages

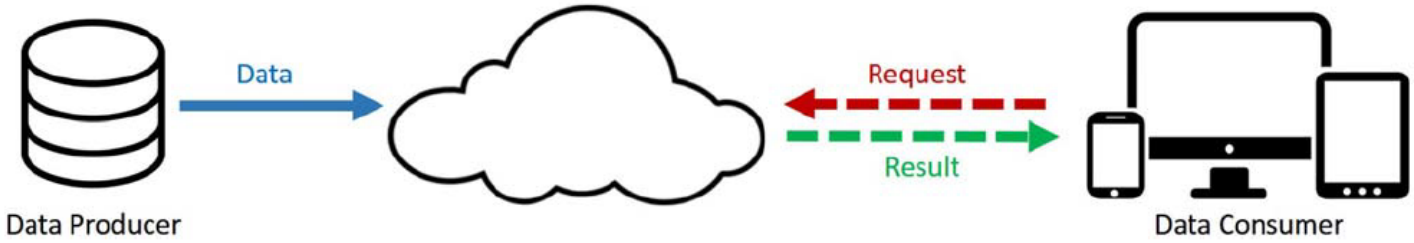
  - Sheer volume of edge resources and far edge IoT devices (~ 50 billion things)

Why Now?

- Networking: SDN/NFV, Ultra-low-latency, 5/6G
- Computing power: smartphones, wearables, etc.
- Explosion of data at the edge

Edge Evolution

- Just a data producer to the cloud



- Now, it is a data consumer and producer

(Personal) Edge Computing Models

- Mobile offloading: face recog (lat and energy)
- Cloud offloading: shopping cart updating (lat)
- Edge data processing: localized data analytics
  - Local search (e.g. lost child) => lat
  - Filtering (e.g. remove faces) => privacy, b/w
    Aggregation (e.g. combine data) =>b/w

Edge Scales

- Personal (Satya: AR/VR, cog assistance)
- Smart Home
- Smart City
- Collaborative Edge

Technical Challenges

- All the usual problems with dispersion
  - reliability, naming, programming, naming, hetereogentity, scalability
- Algorithms, systems for collective control and sharing of edge resources
- Runtime infrastructure: edge services
- Complexity management
- Weaker security perimeter

Edge-centric

- Cloud = loss of privacy requires unilateral trust
- Key Points
  - Proximity is in the edge
  - Intelligence is in the edge
  - Trust …
  - Control … but cloud may play a role
  - Humans …

Non-technical Challenges

- Edge infrastructure: who provides it?
- Edge business case: who funds it?

On Thursday

Early Edge: cloudlets

Cloudlets: at the Leading Edge of Mobile-Cloud Convergence

Just-in-Time Provisioning for Cyber Foraging