Technical drivers of “cloud” centralization and megacorporate domination

23 March 2019 — LibrePlanet

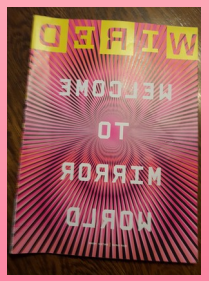
Andy Oram, Editor, O’Reilly Media

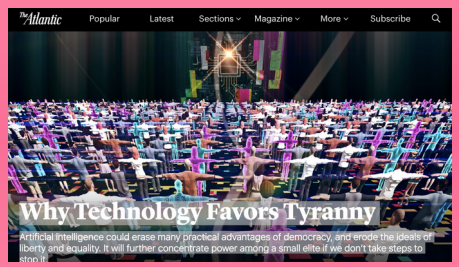
Outline of talk

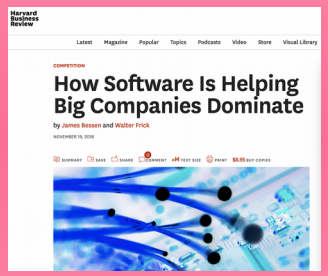
- Description of the problem
- Three technical drivers
- Possible remedies

Problem description

Centralization of computing — decried







https://www.wired.com/story/mirrorworld-ar-next-big-tech-platform/

https://www.nytimes.com/2018/08/31/technology/india-technology-american-giants.html

https://www.theguardian.com/technology/2018/nov/01/tim-berners-lee-says-says-tech-giants-may-have-to-be-broken-up

https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330/

https://hbr.org/2018/11/how-software-is-helping-big-companies-dominate

https://tech.co/news/elizabeth-warren-break-tech-companies-2019-03 (this is not an endorsement)

First say something nice

Some benefits of the cloud:

- Brings advanced computing to people with little training and few resources
- Unleashes creativity in the arts, politics, and elsewhere
- Advances research via the use of data

Underlying factor: monopolization

Many of the criticisms of centralized computing really stem from the more general problem of centralized industries:

- Vendor lock-in
- Regulatory capture
- Lower wages
- Reduced innovation (not clear whether that's happening here)
- Etc.

Social media: a case all its own

Many of the criticisms of centralized computing apply only to a few social media sites, and therefore are not pertinent to this talk:

- That they cheapen relationships (promoting insincerity, leaving people feeling lonely)
- That they degrade information (splitting news stories from publishers and sources, enabling Twitter storms of fake news, etc.)
- That they foster a Panopticon where everything you say can come back to haunt you

Problems of centralized computing

- Drives companies inexorably to collect data (might start innocuously with “improving our service”), leading to dangerous effects on privacy, etc.
- Removes from Individuals and communities the control over data about them (which they may not have generated in the first place) and therefore their autonomy
- Reduces transparency and creates conditions for impunity in the unimpeded exercise of power
- Discriminates against people with low-bandwidth or no connections, who therefore lack both access and representation.
- Leads to asymmetric knowledge: our behavior is manipulated by others who possess data—what happens to free will?
- Creates a vicious cycle in which centralized data and computing allow a large company to crush competition
- Offers a central point of failure and attack, whether through scattered cybercrime or government dictate

"A society of large tools cannot be democratic, egalitarian, socialistic, humane, and just. It must be hierarchical, exploitative, bureaucratic, and authoritarian. If the day comes when all of humanity's wants can be supplied by a few giant tools, the people who tend them will rule the world."

John Holt, in the Planet Economics section of the *Whole Earth Epilog*, 1974. Quoted in WIRED magazine, October 2018.

Three drivers of centralization

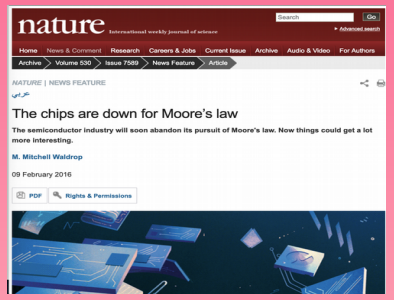
General trends

- Some benefits of “moving to the cloud” are widely touted by the vendors and trade press:
  - Scalability
  - Lower administrative costs
  - Better performance for the money spent
  - Stronger security (probably)
  - Etc.
- The increasing use of sensors, cameras, smart cars, smart homes, etc. (Internet of Things) will further drive analytics to the cloud.
- All cloud computing depends on reliable high-bandwidth networking. Thus, the business drive toward “5G” wireless networks hopes to make everyday computing even more dependent on the cloud.
- These trends are not the focus of this talk.

End of Moore’s Law — background

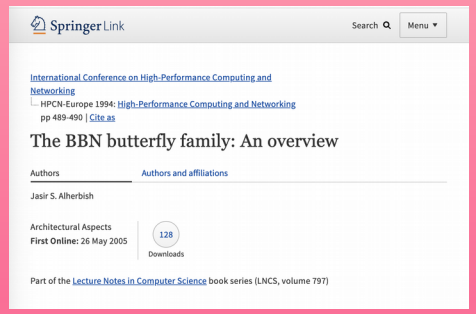
- We used to run all the calculations we needed on our own computers.
- And if our computers weren't powerful enough, we went out and bought bigger ones.

- The end of exponential growth in the mid-2000s challenged that path to progress.

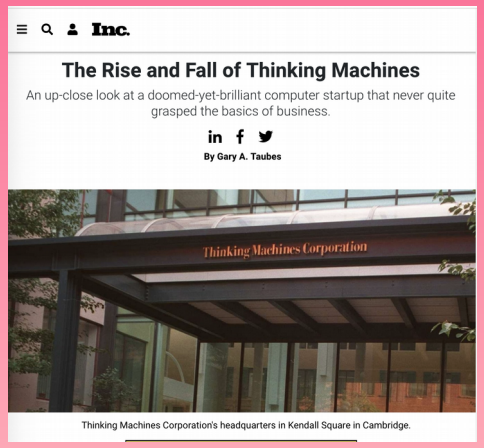


End of Moore’s Law — fumbled transition

- Supercomputers evolved by combining processors in complex fabrics to provide increasingly critical calculations.
  - BBN Butterfly

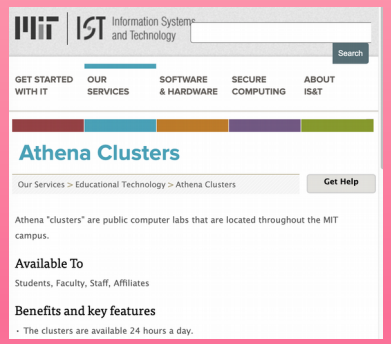


  - Thinking Machines

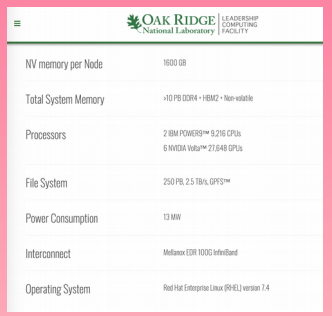


End of Moore's Law — toward the data center

- But a more flexible architecture, attuned to open standards and free software, proved to be the path forward.
  - Example:
    MIT's pioneering Project Athena (1983)



- Even supercomputers mimic the open standards and free software of data centers.



End of Moore’s Law—
Why do we need advanced analytics?

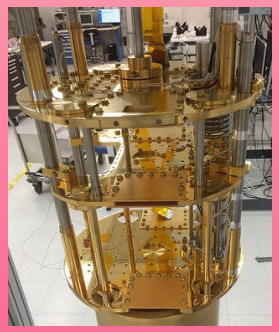
- Compute-intensive applications (weather forecasting, financial fraud detection, etc.)
- Voice interfaces (from "Turn on the light" to “How do you get lunch near the Stata Center?”)
- More human-computer interactions: haptic interfaces, AR/VR, capturing eye movements, etc.
- Intelligent agents, affective computing (requires "a brain the size of a planet")

Advances in processors — compiling into hardware

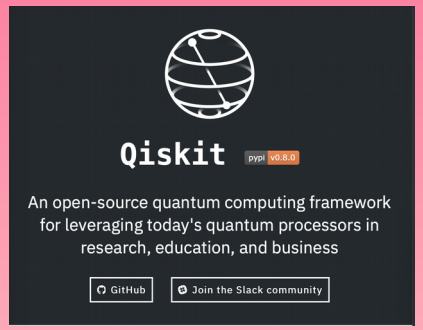
- CISC versus RISC: an amiable resolution.
- CISC still inspires specialized chips: network processors, GPUs, hardware-based encryption devices, field-programmable gate arrays (FPGAs), etc.
- Data centers have revived the CISC philosophy.
  - Google's Cloud Tensor Processing Unit (TPU) can perform analytics 30 to 80 times faster than garden-variety computers, even those using GPUs.
  - The concept is simple: compile common functions such as matrix multiplies into hardware.
- Using specialized hardware, centralized data centers will stay far ahead of commodity computers in processing power (but a surprise will come later).

Advances in processors — does this end the argument?

- Quantum computing promises to provide previously unachievable results in critical areas such as encryption.
- Requires temperatures under 2 degrees Kelvin—colder than outer space. Not something to run out of your basement.



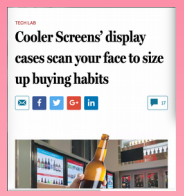
- If the hype is true, the fight over the cloud will be over.
- But quantum computing can be provided through cloud services.



Increasing returns on collecting data

- Two correlated data sets generate far more value than one.
- Example: combining the list of people who attended a political rally with a voter registration list.

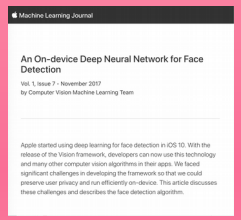
- Benefits grow exponentially with a hundred lists, a thousand, a million…
- This explains both the power of modern analytics and the addictive tug of data on large institutions.
- Now even retail cases and airline entertainment systems contain cameras (to improve service, of course)



Possible remedies

Local processing — current efforts

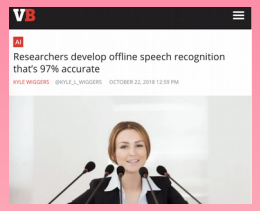
- Mobile devices today can perform sophisticated analytics.
- Apple has moved face recognition to the local device as a privacy-preserving measure.



- Face recognition can be done on a Raspberry Pi.

- You can make 3D scans of objects on a mobile phone.

- Advances are bringing speech recognition to mobile devices.



Local processing — division of labor

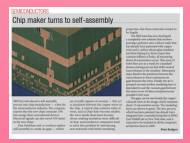
Localized analytics represent a collaboration between centralized engines and local devices:

- A central hub still collects data.
- The central hub runs machine learning over the data and other available information to create an analytical model.
- Edge devices download the model and apply it to local data.

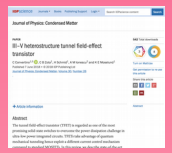
Local processing— potential hardware advances

We may overcome Moore's Law and carry out even more work on our devices through breakthroughs in hardware. Possibilities:

- Directed sub-assembly



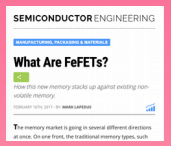
- Tunnel field-effect transistors (TFETs)



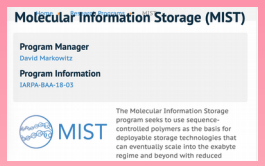
- Spintronics

- Gallium nitride

- Ferroelectric FET storage

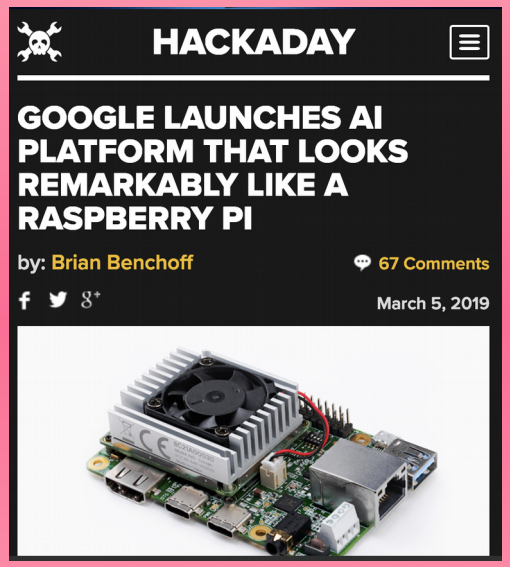


- DNA



Local processing—an actual hardware advance

Google has released its TPU on both an embedded device and a dongle.



Google's explicit goal is to promote machine learning on edge devices.

Local processing—potential software advances

- The flip side of improving efficiency through faster hardware.
- Technical advances are constantly trimming the computational burdens of algorithms. Examples:
  - Shift from convolutional neural networks (CNNs) or recurrent neural networks (RNNs) to long short-term memory (LSTM) or gated recurrent unit (GRU) algorithms.
  - Never-ending learning incrementally incorporates new information, somewhat as a Bayesian process updates a spam filter.
- Streaming data has prepared the ground for local, privacy-protecting analytics (Lambda and Kappa architectures).

Local processing—relationship to the cloud

- While these advances will support edge computing, they will also improve the efficiency and reach of cloud computing.

- We will be better off not pursuing an arms race between cloud companies and individuals, but looking for ways to empower individuals in a cloud environment.

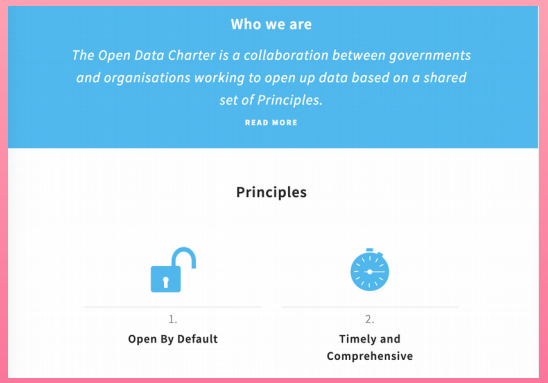
Local processing—for networks too

- When someone turns on the light in their home by asking their Amazon Echo device, the request goes to Amazon's servers and back to the light.

- An architectural design for a man-in-the-middle attack, snooping and collecting data.
- Ecologically harmful, because cloud computing is more energy-intensive than local computing on average.
- By sending information unnecessarily over the Internet, the architecture opens up new security threats.
- Therefore, local communications should be limited to local networks under the control of the individual.

More public data

- Governments around the world are releasing data to the public.

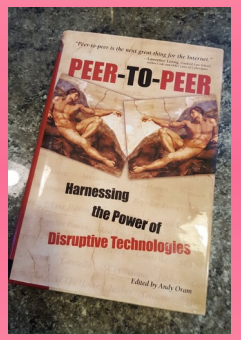


- Small companies and non-profits as well as large firms can build tools on open data.
- To avoid abuses such as those done by Psy-Group and Cambridge Analytica, the data should be released under a license that limits its uses.

Distributed services — status

- Many available: Jabber (based on XMPP), Riot (based on Matrix), Diaspora, etc.

- Current era started around 2000, and quickly revealed problems in addressing, coordination, and trust.



- Never achieved the stupendous reach of centralized services, but with some help they might.

Distributed services — prospects

- Why not create data analytics algorithms for distributed services?
  - Goal: achieve the benefits of data aggregation and analysis without centralizing the data.
  - Programmers are always responding creatively to new requirements.
  - For instance, a whole toolset and architecture for streaming data (Apache Flink and Storm, etc.).
  - Apply this creativity to algorithms that work across thousands of personal computer servers (federated learning).
  - Could we implement a service like differential privacy over distributed data?
- Distributed services would also be energized by open data.

Legal limits on data collection

- Despite complaints, the General Data Protection Regulation (GDPR) has a positive effect.
- Paying individuals for data—a very popular proposal.
  - Problems:
  - Legitimizes data collection
  - Doesn't address "data exhaust" that is collected without asking us
  - Creates two-tier privacy for the affluent versus others
- Meaningful consent is not only hard to achieve, but in many ways inapplicable.

An inspiration

- Cloud computing presents challenges to people who value freedom, autonomy and privacy.
- The challenges must be addressed through both technology and policy.
- People with the skills to understand and work on these issues should employ that power to carve out safe spaces for individuals to flourish.

@praxagora on Twitter Related article My web site: https://praxagora.com/