

Quick Intro to Unix

Filip Buric
January 2020



Big Data

(1 EB = 1 million TB = 1 billion GB)

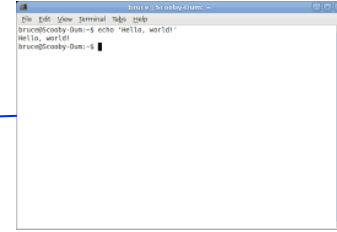
<u>Data Phase</u>	<u>Astronomy</u>	<u>Twitter</u>	<u>YouTube</u>	<u>Genomics</u>
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year	500–900 million hours/year	1 zetta-bases/year
Storage	1 EB/year	1–17 PB/year	1–2 EB/year	2–40 EB/year
Analysis	In situ data reduction	Topic and sentiment mining	Limited requirements	Heterogeneous data and analysis
	Real-time processing	Metadata analysis		Variant calling, ~2 trillion central processing unit (CPU) hours
	Massive volumes			All-pairs genome alignments, ~10,000 trillion CPU hours
Distribution	Dedicated lines from antennae to server (600 TB/s)	Small units of distribution	Major component of modern user's bandwidth (10 MB/s)	Many small (10 MB/s) and fewer massive (10 TB/s) data movement

doi:10.1371/journal.pbio.1002195.t001

Year 2025 projection of annual storage and computing needs

(from <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002195>)

Mainframes (1950s – 1970s)



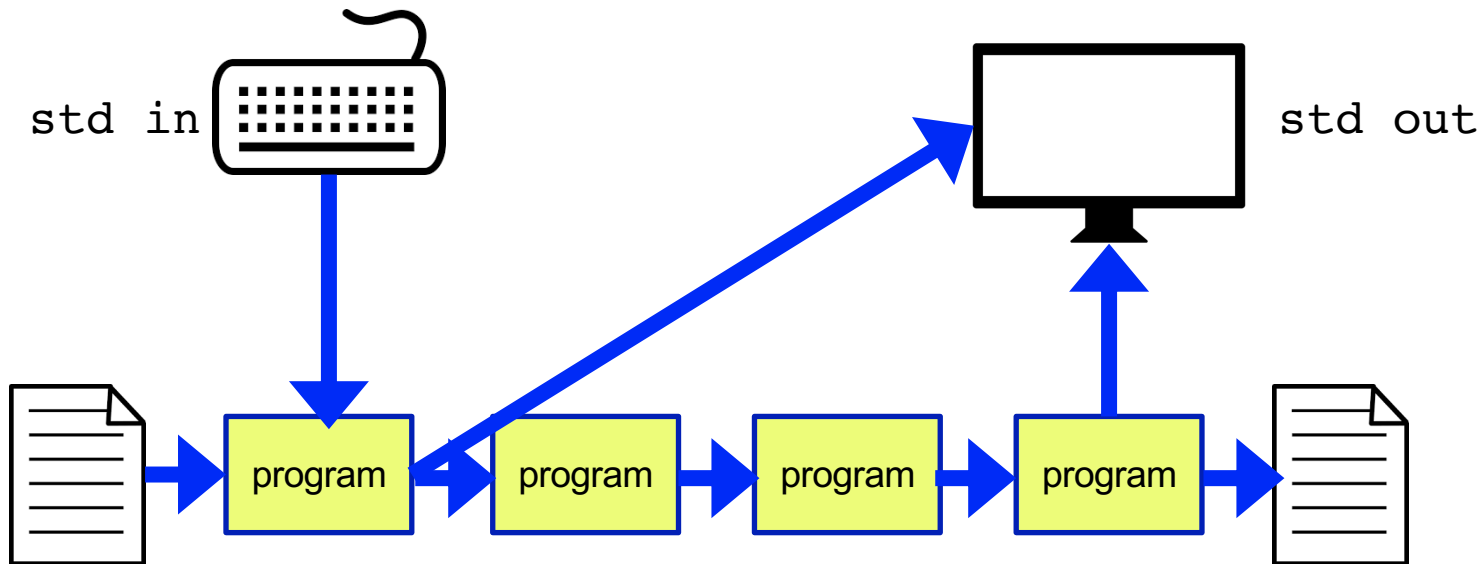
- Much like clusters today (high performance and capacity)
- Users connected from *terminal* computers (glorified keyboard + screen/paper output)
- Behaviour and terminology originates from back then

Unix systems 1970s - present

- Many versions across the years: Bell Labs, Berkeley, IBM, hobbyists
- “Unix” is actually a trademark but became commonplace
- Sometimes people use **nix*
- Official documentation uses the term **POSIX** = the behavioural standard that anything calling itself a Unix system has to conform to
- Examples: macOS, Linux, Android, iOS
- the vast majority of commands work the same way, but some versions have additional functionality (e.g. BSD grep vs GNU grep)

Philosophy

A collection of **specialized** programs (or “commands”) that are inter-**connectable** through **text**



Some terms

- command line = terminal = shell \approx bash
- prompt = text after which you may enter commands (usually ends with \$)
- ssh = secure shell : creates an encrypted connection to another computer
- command = program
- directory = folder
- (compressed) archive = (compressed) files joined into single file (e.g. zip, tar.gz, tar.bz2)
- pipe = data redirector that takes output from one command and inputs it to another command

Course material: <https://mpbio-bbt015.github.io/>

Appendix: Clipart Licences

- AlisonW (https://commons.wikimedia.org/wiki/File:Teletype_with_papertape_punch_and_reader.jpg), „Teletype with papertape punch and reader“, crop, <https://creativecommons.org/licenses/by-sa/3.0/legalcode>
- Jason Scott (https://commons.wikimedia.org/wiki/File:DEC_VT100_terminal.jpg), „DEC VT100 terminal“, <https://creativecommons.org/licenses/by/2.0/legalcode>
- <https://www.flickr.com/people/janitors/> (https://commons.wikimedia.org/wiki/File:Samsung_Galaxy_S6_edge+.jpg), „Samsung Galaxy S6 edge+“, <https://creativecommons.org/licenses/by/2.0/legalcode>
- Stéphanie Walter (https://commons.wikimedia.org/wiki/File:Devicetemplates_computer-02.png), „Devicetemplates computer-02“, <https://creativecommons.org/licenses/by-sa/3.0/legalcode>