

Scientific computing in high-energy physics

Lecture 1, 28.04.2022

Ante Bilandzic
(E62, Dense and Strange Hadronic Matter)



European Research Council

Established by the European Commission

Outline of today's lecture

- Course trivia
- Free adverts
 - Linux
 - Bash
 - ROOT



Course trivia

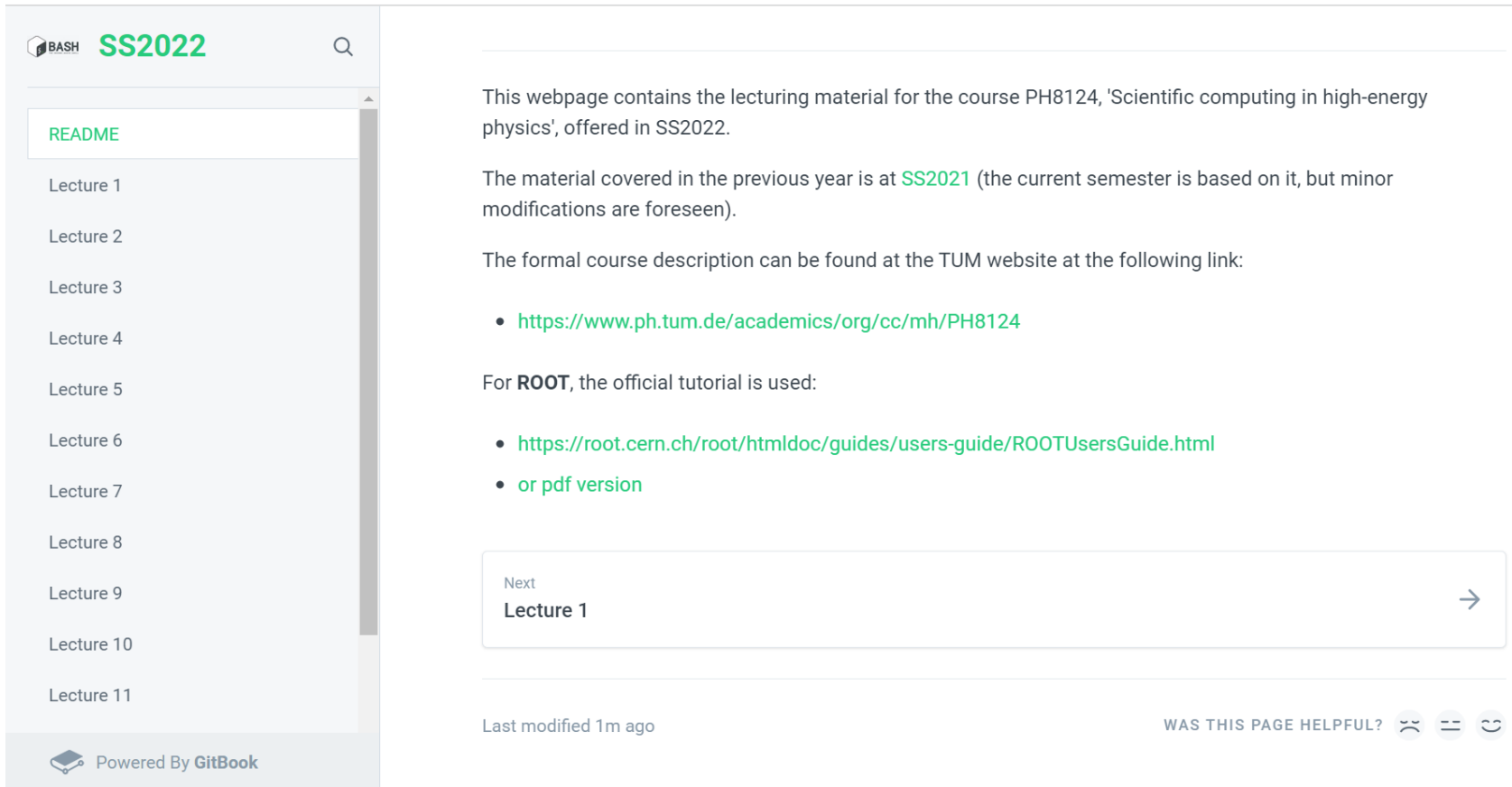
- **PH8124:** ‘Scientific computing in high-energy physics’
 - <https://www.ph.tum.de/academics/org/cc/mh/PH8124/>
- When & where:
 - Thursday: 14:00-16:00 (hybrid mode)
 - Physics Department, E62 seminar room 2024
 - Roomfinder: https://portal.mytum.de/displayRoomMap?roomid=2024@5101&disable_decoration=yes
 - 12 contact days (last lecture is on July 28th)
- Examination:
 - Homeworks (10 in total)
 - Oral examination during the presentation of the final project
- Contact:
 - Ante Bilandzic, ante.bilandzic@tum.de, Office PH 2162

Course trivia

- Given the current situation with pandemic, the presence in person during the lecture is not mandatory, but it's possible
- Parallel online coverage via **Zoom**
 - TUM offers the licensed version at the following link: <https://tum-conf.zoom.us>
 - 'Sign In' with your TUM credentials
- For this lecture, I have created the recurring meetings on Thursdays, from 13:30-16:00
 - **Zoom** coordinates will be distributed regularly via the official mailing list to all registered students for this course
- The lectures will be recorded, and recordings shared immediately afterward

Webpage

- The course has a dedicated webpage:
 - Link: <https://abilandz.gitbook.io/ss2022/>



The screenshot shows a GitBook interface for the 'SS2022' course. The left sidebar contains a table of contents with 'README' at the top, followed by 'Lecture 1' through 'Lecture 11'. The main content area on the right has the following text:

This webpage contains the lecturing material for the course PH8124, 'Scientific computing in high-energy physics', offered in SS2022.

The material covered in the previous year is at [SS2021](#) (the current semester is based on it, but minor modifications are foreseen).

The formal course description can be found at the TUM website at the following link:

- <https://www.ph.tum.de/academics/org/cc/mh/PH8124>

For **ROOT**, the official tutorial is used:

- <https://root.cern.ch/root/html/doc/guides/users-guide/ROOTUsersGuide.html>
- [or pdf version](#)

At the bottom of the main content area, there is a 'Next' section with a button labeled 'Lecture 1' and a right-pointing arrow.

The footer of the page includes 'Last modified 1m ago' on the left, and 'WAS THIS PAGE HELPFUL?' with three feedback icons (thumbs up, thumbs down, and a refresh icon) on the right. The bottom left corner features the TUM logo, and the bottom right corner has a large blue number '5'.

Course trivia

- After each lecture, executive summary of the covered material will be shared via email
- The course webpage will be updated regularly
- In the same way, I will also share the homework exercises
- Recommended literature:
 - Mendel Cooper: 'Advanced Bash-Scripting Guide' (<http://tldp.org/LDP/abs/abs-guide.pdf>)
 - Cameron Newham and Bill Rosenblatt, 'Learning the bash Shell: Unix Shell Programming (In a Nutshell (O'Reilly))'
 - ROOT User's Guide (<https://root.cern.ch/root/html/doc/guides/users-guide/ROOTUsersGuide.html>)
 - Richard Blum and Christine Bresnahan, 'Linux Command Line and Shell Scripting Bible'

Course trivia

- Grading:
 - 3 ECTS points
 - Final grade = grade at final project examination - '1 unit' if you have completed correctly 75% of all homeworks
 - There will be in total 10 homework exercises, one after each lecture
- Oral examination at the final project presentation:
 - The topic for the final programming project will be offered at some point towards the end of the lecture
 - The oral exam of about 25 minutes consists of presenting:
 1. How your programme was designed/implemented?
 2. Testing the execution of your code (crash-free, bug-free, efficiency in terms of CPU usage and memory consumption)
 3. Testing the code flexibility (e.g. how you would add some new feature in the code?)

Course trivia

- Preliminary list of topics to be covered:
 - **Linux:** filesystem hierarchy and file manipulation, handling processes and jobs, frequently used commands, etc.
 - **Bash:** shell environment, variables, string manipulation, built-in commands, aliases, functions, conditional statements, loops, command substitution, command chain, test constructs, piping, redirections, code blocks, subshells, process substitution, brace expansion, regular expressions, here-strings and here-documents, etc.
 - **ROOT:** using ROOT GUI, plotting, histogramming, functions, fitting, trees, file merging, etc.

Course trivia

- Course classification:
 - At the moment, classified as a ‘Non-physics elective course’
 - Open both to Bachelor and Masters students
 - <https://www.ph.tum.de/academics/msc/physics/nonphys/>
- Course evaluation:
 - At some point during the lecture, you will be asked to evaluate this course: Please, do it! (reminder will be sent later)

Trivia

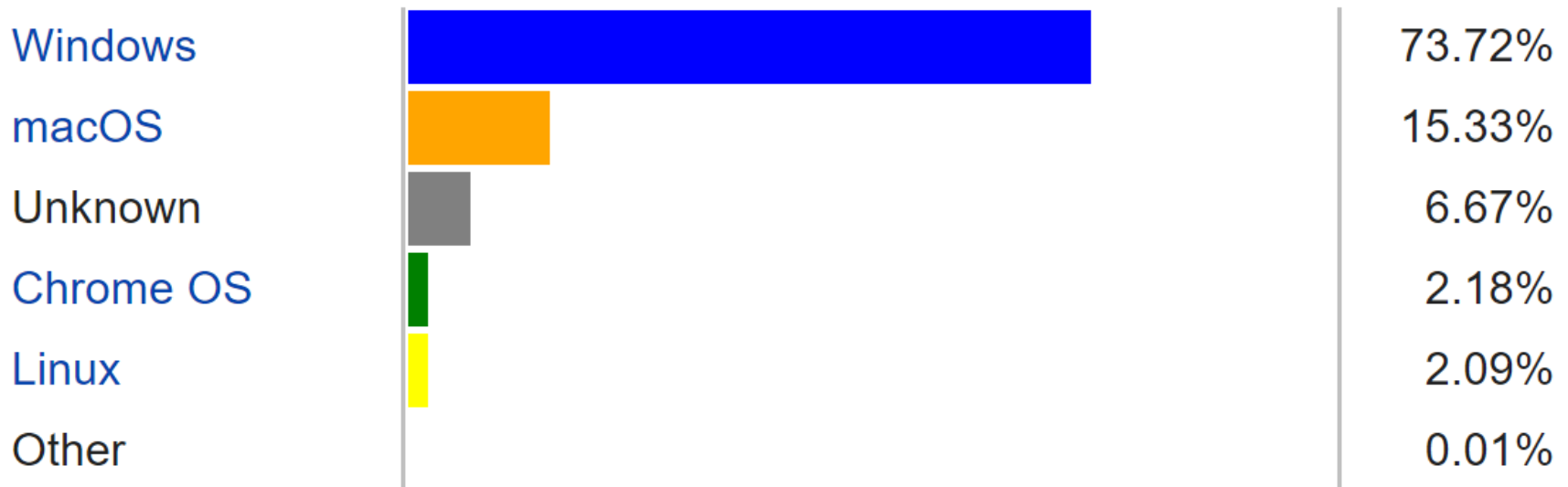
- No lecture on:
 - May 26th – Ascension Day
 - June 16th – Corpus Christi

Free adverts

Free advert (#1)

- Why Linux? Statistics for all desktop/laptop computers:

Desktop/Laptop operating system browsing statistics

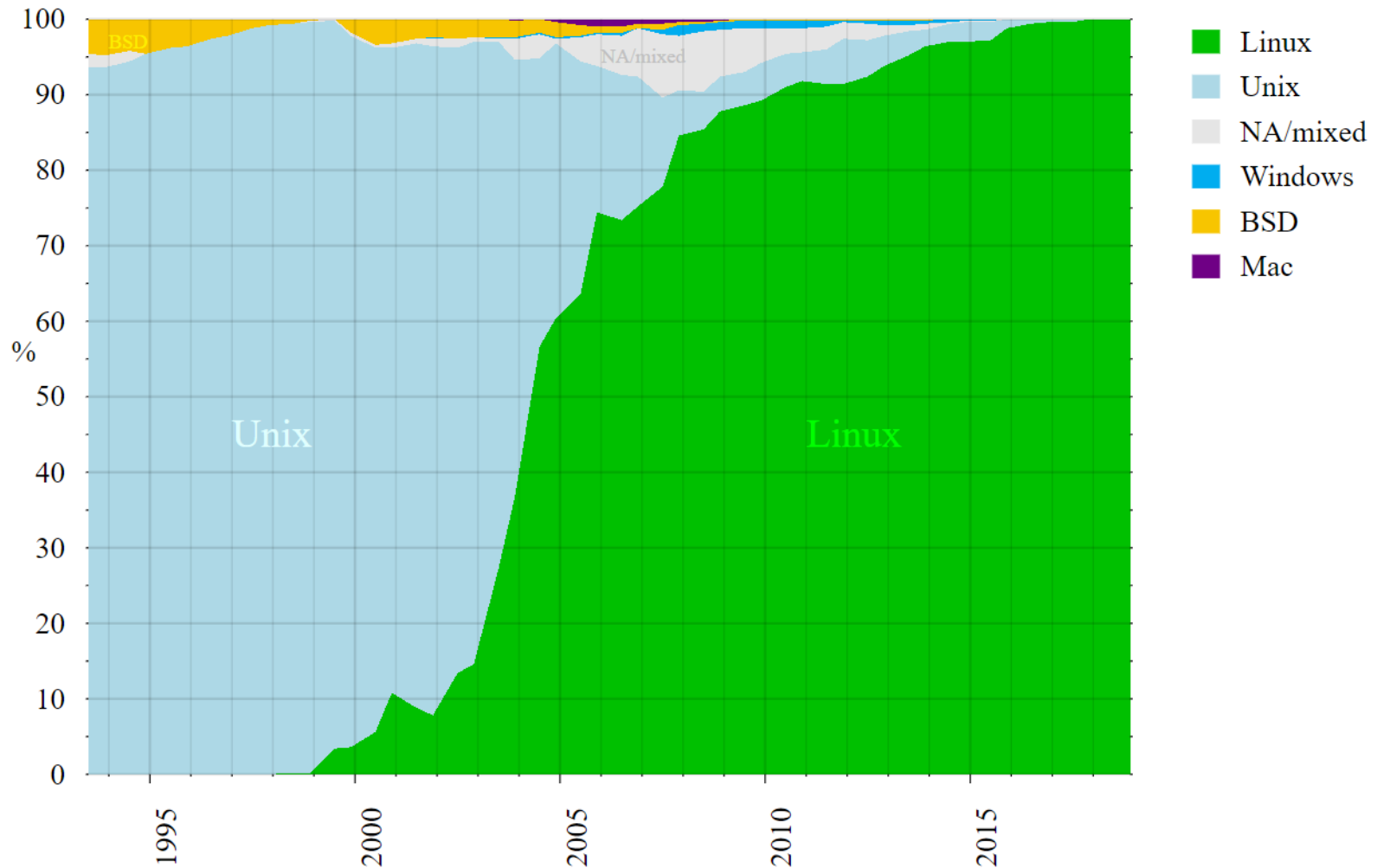


Desktop OS market share according to [StatCounter](#) for December 2021.^[64]

Chrome OS is also based on the [Linux kernel](#).

Free advert (#1)

- Why Linux? Statistics for supercomputers:



Free advert (#1)

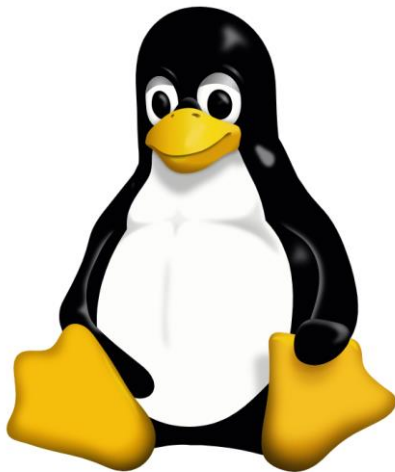
- The most powerful supercomputers (November, 2021):

←	1-100	101-200	201-300	301-400	401-500	→
---	-------	---------	---------	---------	---------	---

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	Supercomputer Fugaku - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442,010.0	537,212.0	29,899
2	Summit - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	200,794.9	10,096
3	Sierra - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438

Free advert (#1)

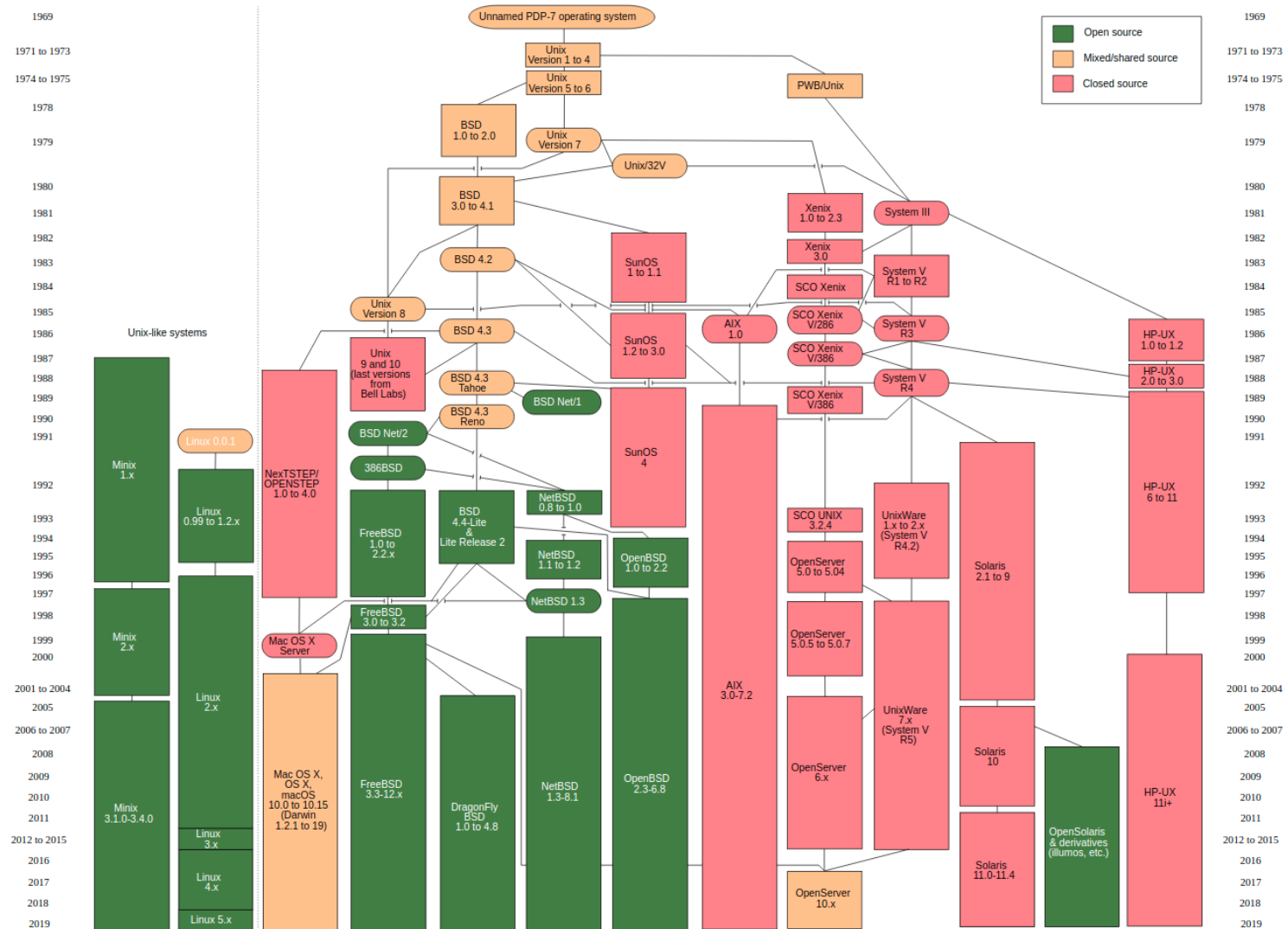
- Why Linux?
- Linux is by far the leading operating system in computers used in scientific research (CERN, NASA, etc.)
- **Linux kernel:** developed by Linus Torvalds in the early 90s
- **GNU ('GNU's not Unix'):** open source utilities to perform standard actions on the computer - no licencing
 - Linux kernel + GNU utilities = Linux operating system
- Initially, the Linux OS was basically a free Unix



Penguin named Tux is the most commonly used logo for Linux

Free advert (#1)

- Unix family tree – Linux is one of its descendants



Free advert (#1)

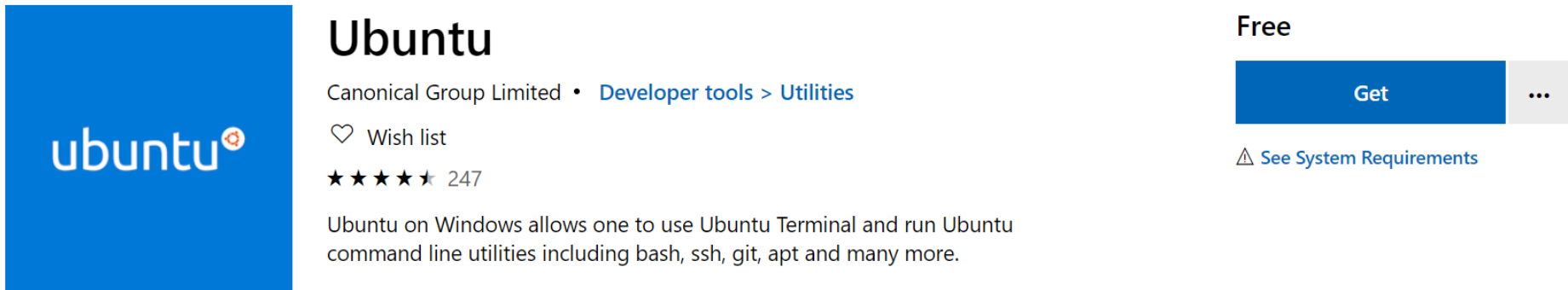
- Linux family tree: Common Linux kernel and plethora of different Linux distributions built on top of it
 - Full-core: Debian, Red Hat Enterprise, openSUSE, etc.
 - Specific: Ubuntu, Fedora, CentOS, Scientific Linux, Linux Mint, etc.
- Specific distributions are derivatives of full-core distributions:
 - Ubuntu is derivative of Debian, Fedora of Red Hat, etc. (see complete tree at <https://distrowatch.com/images/other/distro-family-tree.png>)
- The material presented in this course will be demonstrated on Ubuntu, but it applies also to any other Linux distribution
- Regular updates on all distributions: <https://distrowatch.com>



ubuntu®

Getting Ubuntu for Windows users

- If you have on your laptop only Windows, you could either:
 - install Ubuntu on Windows from the following link
<https://www.microsoft.com/en-us/p/ubuntu/9nblggh4msv6?activetab=pivot:overviewtab>

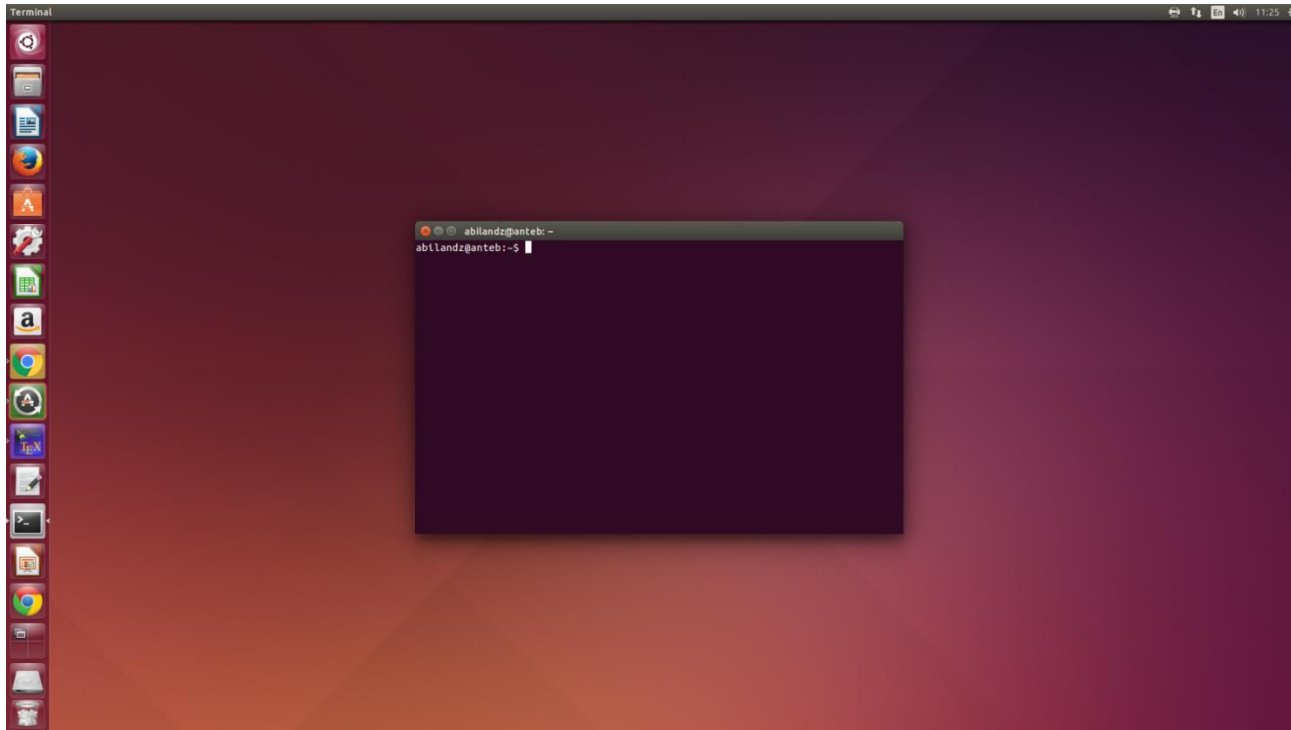


The screenshot shows the Microsoft Store page for Ubuntu. On the left is the Ubuntu logo on a blue background. To its right, the title 'Ubuntu' is displayed, followed by 'Canonical Group Limited' and a link to 'Developer tools > Utilities'. Below this is a 'Wish list' icon and a star rating of 247. A description states: 'Ubuntu on Windows allows one to use Ubuntu Terminal and run Ubuntu command line utilities including bash, ssh, git, apt and many more.' On the right side of the card, the word 'Free' is shown above a blue 'Get' button and a grey button with three dots. Below the buttons is a link: 'See System Requirements'.

- install PuTTY (<https://www.putty.org/>) and then use it to connect to some computer running Linux, on which you have an account with access rights

Free advert (#2)

- Is this a right course for you?
 - If, after you have opened a terminal in Linux ...



... you have asked yourself: 'What now?', then this is the right course for you!

Free advert (#2)

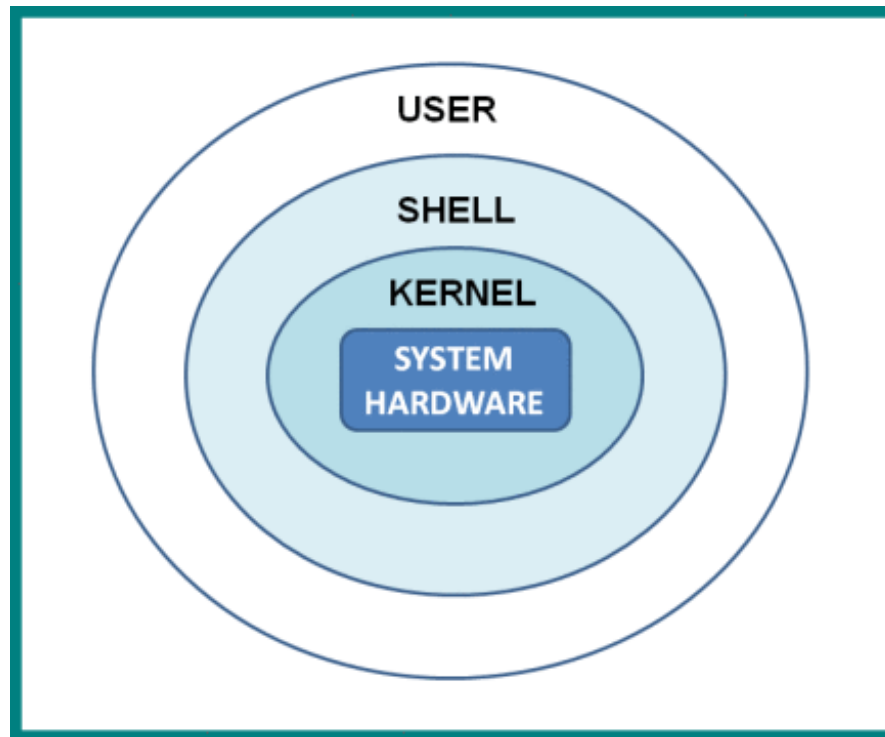
- What can we do in the terminal?
 - Not that much with the mouse...
- Next, you can start typing and pressing 'Enter', but especially if you do it for the first time most likely whatever you have typed in the terminal will produce only the error messages
 - Still, that is something, as it clearly means that there is some secret/magic language which is trying to respond to, or to interpret, your command input, as soon as you have typed something in the terminal and pressed 'Enter'. What is that secret built-in language available in the terminal?

Linux shells

- Loosely speaking, shell is any program that user employs to type commands in the terminal (text window)
- Example shells:
 - sh
 - bash
 - zsh
 - ksh
 - csh
 - fish
 - PowerShell (developed by Microsoft!)
- Since Bash is the default shell on most Linux distributions nowadays, we focus on it
 - If not set by default, just type **bash** in the terminal, and you are in the Bash wonderland

Why shell?

- The shell translates the commands you type into a format which the computer can understand
 - It works both ways: User is shielded from Linux kernel, and Linux kernel is shielded from user



A bit of Bash history

- Initial development by Brian Fox in 1989 as a part of GNU project... And it's still alive!
- Bash is an acronym for 'Bourne-again shell'
 - The original shell 'sh' was written in 1977 by Stephen Bourne
- Written entirely in C
 - Linux kernel is mostly written in C
- Executable: /bin/bash
- File extension: .sh
- Command processor / interpreted / scripting language



The current status of Bash

- Bash is well maintained and is under regular development
 - Webpage: <https://www.gnu.org/software/bash/>
 - Source code: <http://git.savannah.gnu.org/cgiit/bash.git>
- Latest release: version 5.1.16 (January 5, 2022)
 - The current maintainer: Chet Ramey



Testing Bash code online?

- In the case you do not have currently the access to the computer running Linux, you can test your Bash code online
 - For instance: https://www.tutorialspoint.com/execute_bash_online.php
- Use this link only as a temporary solution, as this is not a development environment













Interpreted vs. compiled languages

- Interpreted:
 - write code & execute line-by-line
 - less reliable (there is no compiler to catch the errors!)
 - source code can be easily read and copied
 - examples: Bash, Python, Mathematica, JavaScript
- Compiled:
 - write code & compile & execute the compiled file ('binaries')
 - generally runs faster than interpreted code
 - examples: C, C++, Java, Go











Intermezzo: Popularity of languages

- As of recently, Python is the most popular programming language

Apr 2022	Apr 2021	Change	Programming Language		Ratings	Change
1	3	▲		Python	13.92%	+2.88%
2	1	▼		C	12.71%	-1.61%
3	2	▼		Java	10.82%	-0.41%
4	4			C++	8.28%	+1.14%
5	5			C#	6.82%	+1.91%
6	6			Visual Basic	5.40%	+0.85%
7	7			JavaScript	2.41%	-0.03%
8	8			Assembly language	2.35%	+0.03%
9	10	▲		SQL	2.28%	+0.45%
10	9	▼		PHP	1.64%	-0.19%

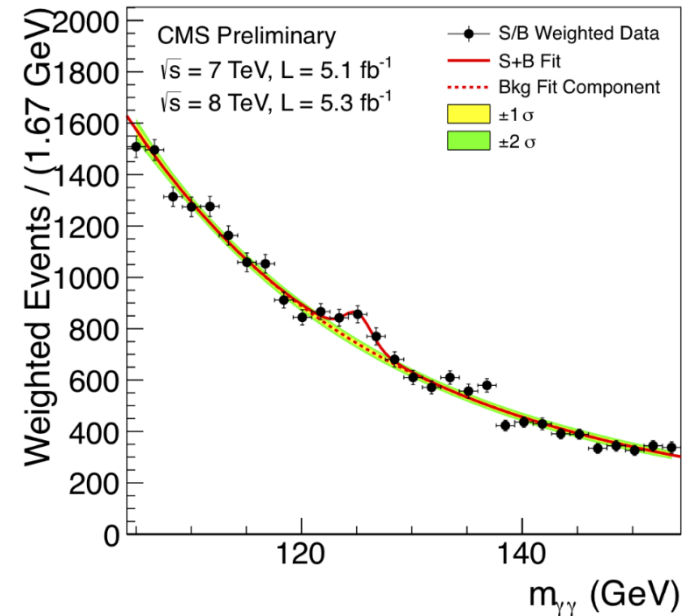
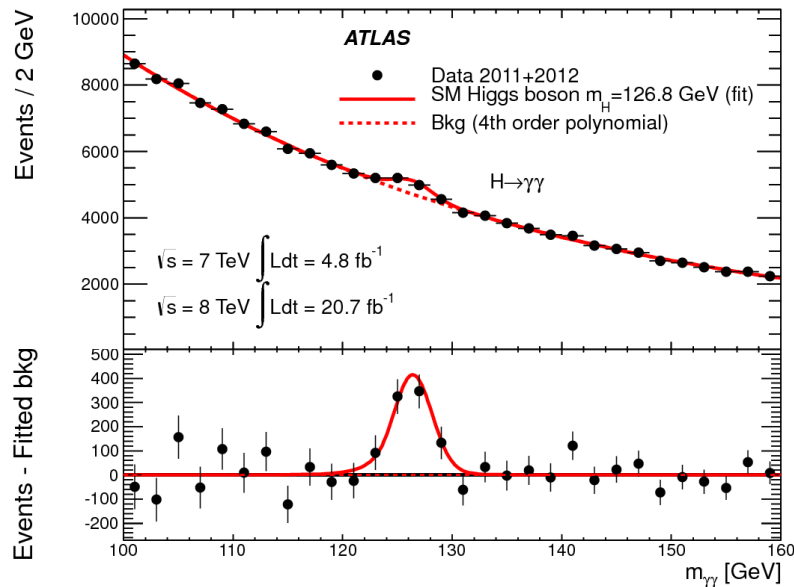
Intermezzo: Popularity of languages

- As of recently, Python is the most popular programming language

11	16	⬆️		R	1.55%	+0.44%
12	12			Delphi/Object Pascal	1.18%	-0.29%
13	14	⬆️		Go	1.09%	-0.14%
14	15	⬆️		Swift	1.00%	-0.19%
15	13	⬇️		Ruby	0.88%	-0.35%
16	11	⬇️		Classic Visual Basic	0.83%	-0.71%
17	23	⬆️		Objective-C	0.82%	+0.15%
18	18			Perl	0.79%	-0.21%
19	37	⬆️		Lua	0.78%	+0.45%
20	19	⬇️		MATLAB	0.74%	-0.25%

Free advert (#3)

- Is this a right course for you? ROOT, what's that?
 - If you have ever asked yourself what is the software in which the Higgs discovery plots were actually made ...

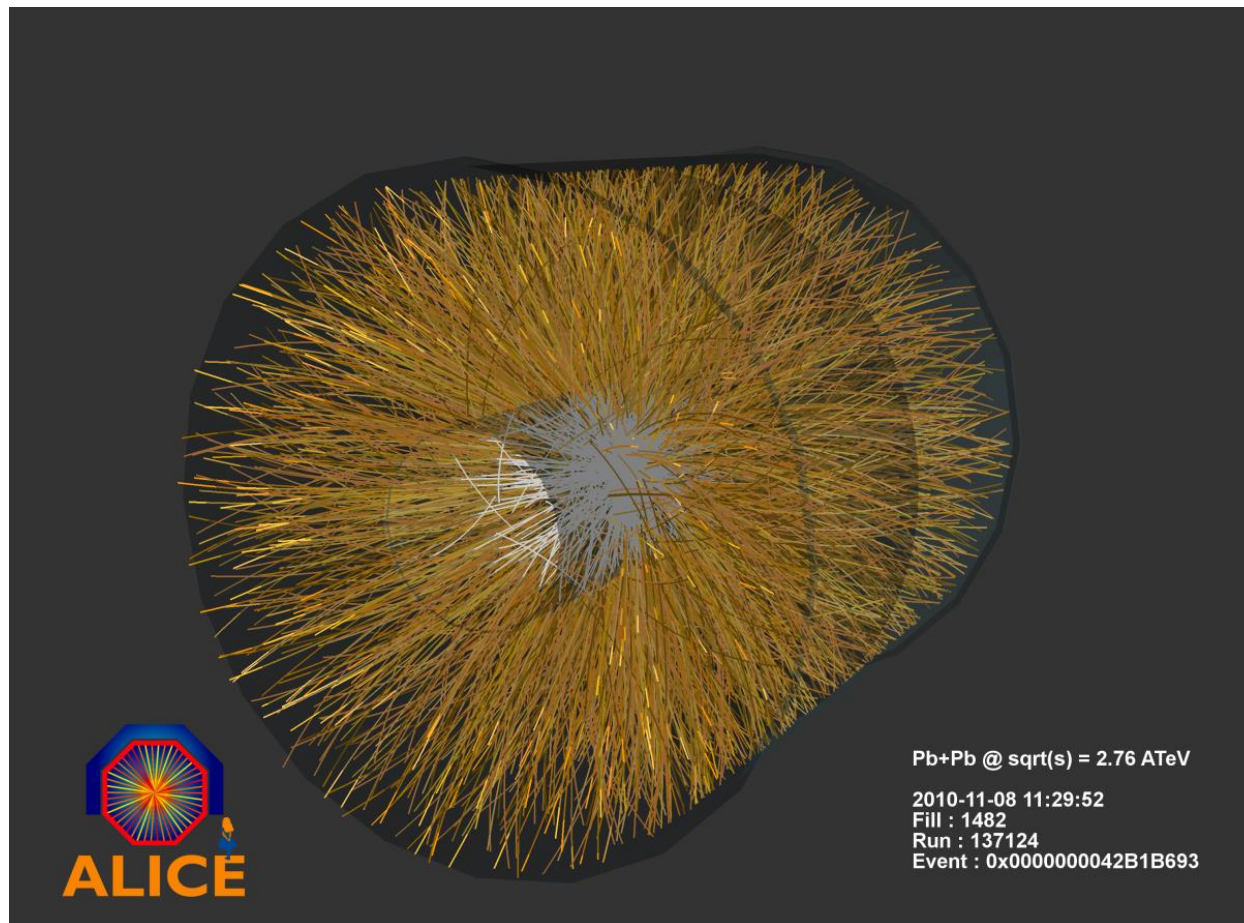


... then this is the right course for you!

- At the very basic level, we can use ROOT for plotting, histogramming, fitting, etc.

Free advert (#3)

- This is the typical heavy-ion event at Large Hadron Collider reconstructed by ALICE Collaboration

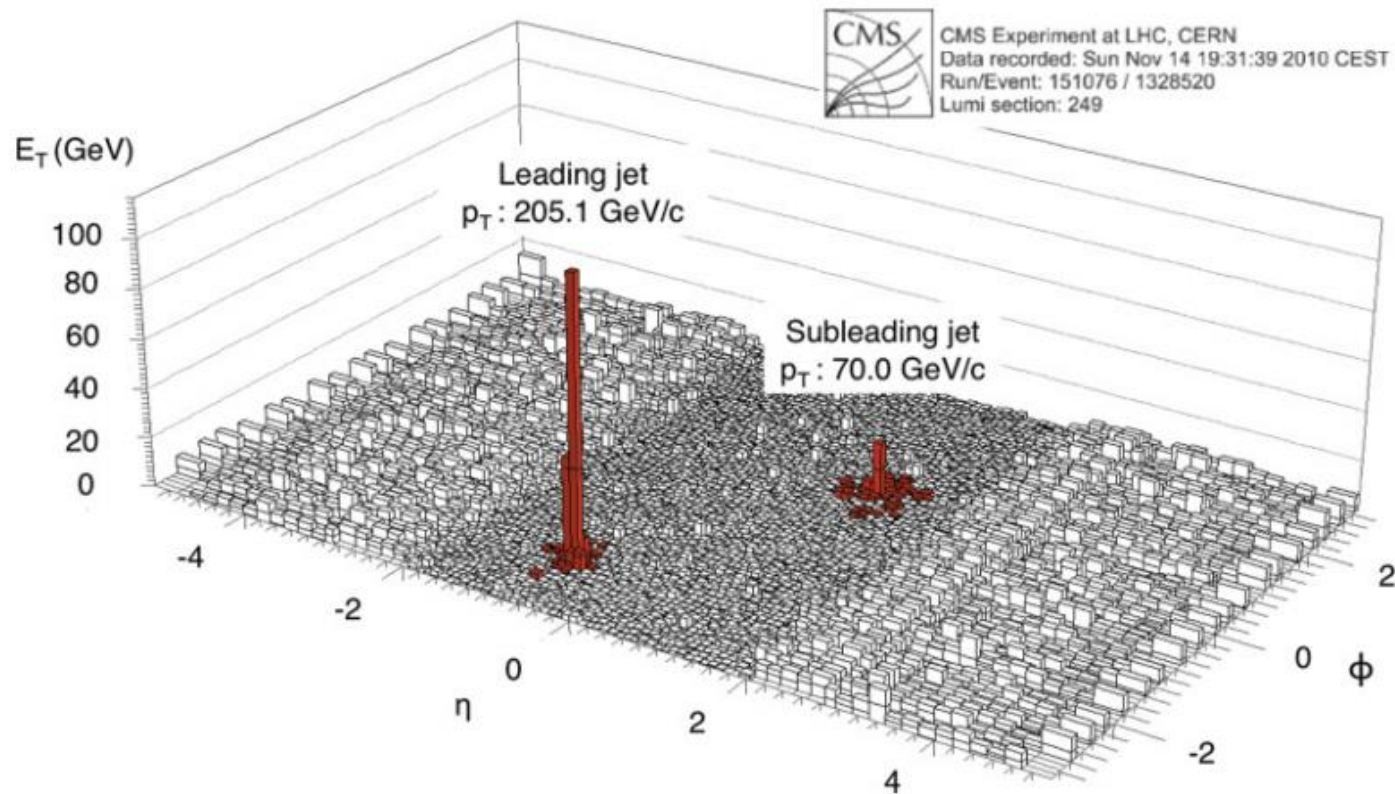


Free advert (#3)

- Trajectories of more than 10000 particles are reconstructed by AliROOT (C++ code specific to ALICE Collaboration built on top of ROOT)
- Most important major collaborations worldwide in high-energy physics currently use ROOT
 - Also the future ones (e.g. CBM at GSI, which will start data taking in 2025, is developing CbmRoot)

Free advert (#4)

- In terms of histogramming quality and performance, it's difficult to beat ROOT...



ROOT

- Object-oriented framework, written in C++ and developed at CERN, for data analysis in high-energy physics
- The development was initiated by René Brun and Fons Rademakers in 1994, and is still under active development
 - Latest release: Version 6.26/00 (March 3, 2022)
- Webpage: <https://root.cern.ch/>
- Root forum: <https://root-forum.cern.ch/>
- Source code: <https://github.com/root-project/root>



Thanks!