LINUX DEEP DIVE FOR DATA SCIENCE

Speaker: Yingquan Li

Date: Thu. 5/19/22 @ 6:00 PM EST

TABLE OF CONTENTS

LINUX OVERVIEW

The 30,000-foot view of Linux.

LINUX COMMAND-LINE INTERFACE (CLI)

The most useful Linux CLI commands and tricks.

ADVANCED CLI TOOLS

Intro to advanced tools such as: Screen/Tmux, Crontabs, Systemctl.

TABLE OF CONTENTS

“All the best people in life seem to like *****.”

— ???

“All the best people in life seem to like Linux.”

— STEVE WOZNIAK 😈

HERE WE GO... 💯

- **01 - LINUX OVERVIEW**
- 02 - LINUX COMMAND-LINE INTERFACE (CLI)
- 03 - ADVANCED CLI TOOLS
- 04 - PYENV, VIRTUALENV, ANACONDA
- 05 - JUPYTER NOTEBOOKS
- 06 - Q&A

THE BEGINNING: BRINGING IT BACK TO THE 1990S

- **Linux** is a free and open-source operating system (OS) that was developed by **Linus Torvalds** in 1991.
- Linus is credited with writing the **Linux Kernel**, which is the main part of the OS that serves as the interface between the system hardware and the user processes (i.e. commands, jobs, etc.).
- To aid the development of Linux within the open-source community, Linus developed the **Git Version Control System** that is used by developers today in a myriad of industries.

From: torvalds@klaava.Helsinki.FI (Linus Benedict Torvalds)
Newsgroups: comp.os.minix
Subject: What would you like to see most in minix?
Summary: small poll for my new operating system
Message-ID:
Date: 25 Aug 91 20:57:08 GMT
Organization: University of Helsinki

Hello everybody out there using minix -

I'm doing a (free) operating system (just a hobby, won't be big and professional like gnu) for 386(486) AT clones. This has been brewing since april, and is starting to get ready. I'd like any feedback on things people like/dislike in minix, as my OS resembles it somewhat (same physical layout of the file-system (due to practical reasons) among other things).

I've currently ported bash(1.08) and gcc(1.40), and things seem to work. This implies that I'll get something practical within a few months, and I'd like to know what features most people would want. Any suggestions are welcome, but I won't promise I'll implement them :-)

Linus (torvalds@kruuna.helsinki.fi)

PS. Yes - it's free of any minix code, and it has a multi-threaded fs. It is NOT protable (uses 386 task switching etc), and it probably never will support anything other than AT-harddisks, as that's all I have :-(.

THE BEGINNING: BRINGING IT BACK TO THE 1990S

- **Linux** is a free and open-source operating system (OS) that was developed by **Linus Torvalds** in 1991.
- Linus is credited with writing the **Linux Kernel**, which is the main part of the OS that serves as the interface between the system hardware and the user processes (i.e. commands, jobs, etc.).
- To aid the development of Linux within the open-source community, Linus developed the **Git Version Control System** that is used by developers today in a myriad of industries.

LINUS!

FAMOUS EMAIL!

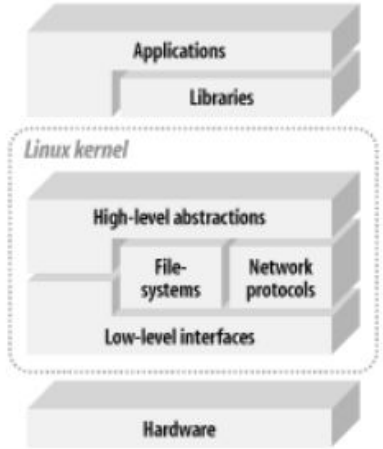
```
From: torvalds@klaava.Helsinki.FI (Linus Benedict Torvalds)
Newsgroups: comp.os.minix
Subject: What would you like to see most in minix?
Summary: small poll for my new operating system
Message-ID:
Date: 25 Aug 91 20:57:08 GMT
Organization: University of Helsinki


Hello everybody out there using minix -

I'm doing a (free) operating system (just a hobby, won't be big and
professional like gnu) for 386(486) AT clones.  This has been brewing
since april, and is starting to get ready.  I'd like any feedback on
things people like/dislike in minix, as my OS resembles it somewhat
(same physical layout of the file-system (due to practical reasons)
among other things).

I've currently ported bash(1.08) and gcc(1.40), and things seem to work.
This implies that I'll get something practical within a few months, and
I'd like to know what features most people would want.  Any suggestions
are welcome, but I won't promise I'll implement them :-)

                Linus (torvalds@kruuna.helsinki.fi)

PS.  Yes - it's free of any minix code, and it has a multi-threaded fs.
It is NOT protable (uses 386 task switching etc), and it probably never
will support anything other than AT-harddisks, as that's all I have :-(.
```

THE LINUX OPERATING SYSTEM

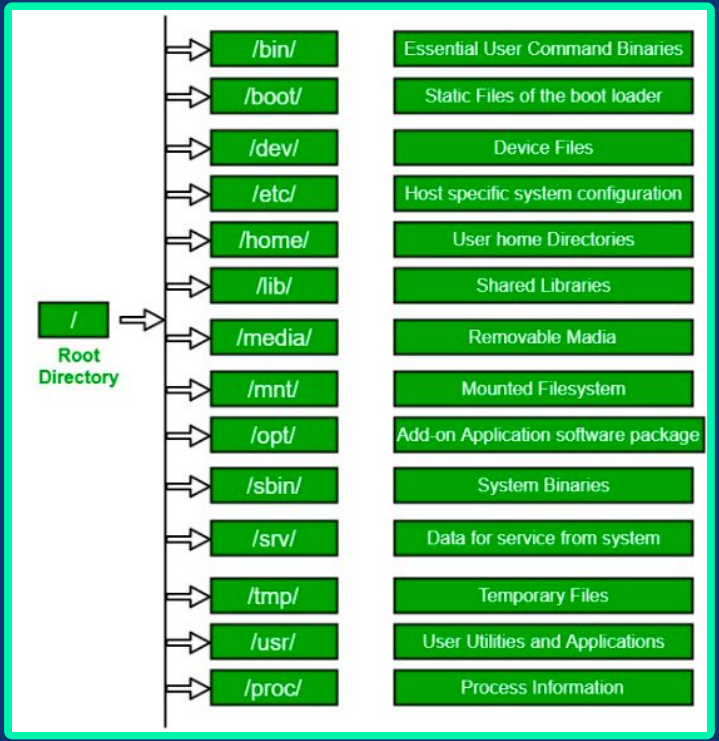
Layers of **abstraction** built on top of each other. We explore each level from highest → lowest.

- (**Application Level**) - The applications and libraries used to build **software applications**.
- (**OS Level**) - The Linux Kernel is the heart of the Linux operating system. It includes:
  a. **Low-level interfaces** talk to the hardware.
  b. The **file-systems** and **network protocols** facilitate storage and communication.
  c. The **high-level abstractions** allow users to work with Linux: processes, files, sockets, signals.
- (**Physical Level**): **Computer hardware** for the Linux Kernel to run on.

Figure 2-4. Architecture of a generic Linux system



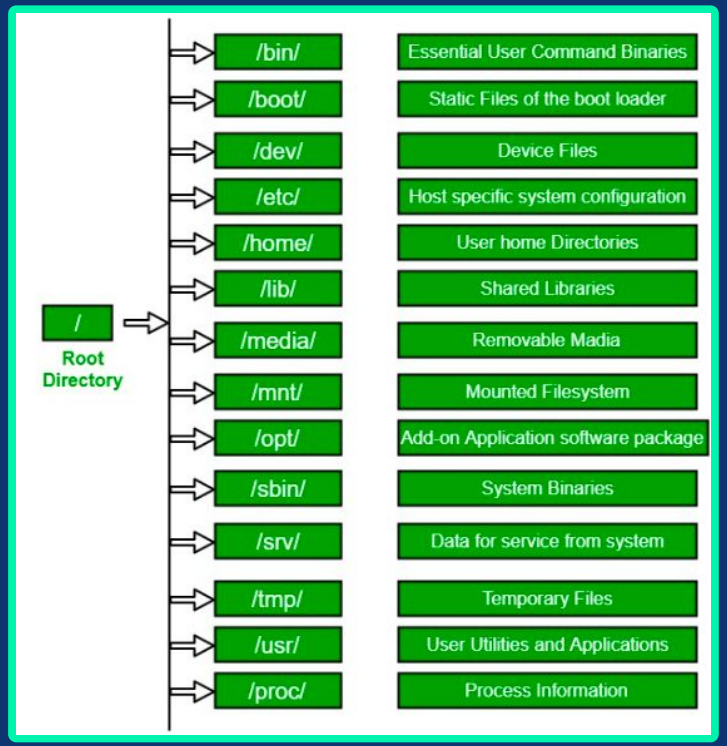
THE LINUX FILE SYSTEM



- The **Linux Filesystem Hierarchy Standard (FHS)** is maintained by the Linux Foundation and delineates the directory structure for the Linux OS.
- The filesystem is highly intimidating for beginners in Linux, and different distributions have slight tweeks on the filesystem.

How will someone interact with the file system?

THE LINUX FILE SYSTEM



- The **Linux Filesystem Hierarchy Standard (FHS)** is maintained by the Linux Foundation and delineates the directory structure for the Linux OS.
- The filesystem is highly intimidating for beginners in Linux, and different distributions have slight tweeks on the filesystem.

Practical Tip #1: **When debugging, please look for advice that pertains to your Linux distribution only**.

Practical Tip #2: **You often only need to go into directories under / to modify config files, that's it!**.

- Please DO NOT **store or delete any files** in any of the directories under **/** except for **/home** and **/tmp**.

DIFFERENT FLAVORS OF LINUX IN 2022

10 Popular Linux Distributions in 2022

1. ArchLinux
2. Debian
3. Fedora
4. Linux Mint
5. Manjaro
6. openSUSE
7. CentOS
8. Tails
9. Ubuntu
10. Zorin OS

The MacOS X terminal is for all intents and purposes a perfect Linux-like clone, but its NOT Linux. However, developers still like Macs.

HERE WE GO... 💯

- 01 - LINUX OVERVIEW
- **02 - LINUX COMMAND-LINE INTERFACE (CLI)**
- 03 - ADVANCED CLI TOOLS
- 04 - PYENV, VIRTUALENV, ANACONDA
- 05 - JUPYTER NOTEBOOKS
- 06 - Q&A

NAVIGATING THE LINUX CLI

- First Principle: **A expert user knows the command-line interface commands and keyboard shortcuts like it's muscle memory**. 💯 To become a master, you have to use the advanced commands.

Advanced Linux Commands:
- **history**
- **grep**
- **find**
- **tree**
- **Symbolic links**
- **alias**
- **!!, !<start of Command>**
- | (Piping)
- **Networking commands**

Advanced Keyboard Shortcuts:
- **CTRL-a**: Go to start of line
- **CTRL-e**: Go to end of line
- **CTRL-u**: Cut from start of line
- **CTRL-k**: Cut to end of line
- **CTRL-w**: Cut word to the left
- **CTRL-y**: Paste text that was cut
- **CTRL-_**: Undo what you just did
- **OPTION Key**: Insert to middle of the command


DEMO!

HERE WE GO... 💯

- 01 - LINUX OVERVIEW
- 02 - LINUX COMMAND-LINE INTERFACE (CLI)
- **03 - ADVANCED CLI TOOLS**
- 04 - PYENV, VIRTUALENV, ANACONDA
- 05 - JUPYTER NOTEBOOKS
- 06 - Q&A

ADVANCED CAPABILITIES

- **Screen**: Terminal multiplexer; user can start screen session and open any number of windows.
- **Tmux**: Alternative terminal multiplexer to Screen.
- **Crontabs**: Daemon (long running process that runs unattended) that executes commands at specific dates and times.
- **Systemctl**: Manages the systemd service manager, which offers on-demand starting of daemons, tracking processes, maintaining mount/automount points, etc. Rather difficult to learn.
- **Bash Scripting**: Scripting language that gives a user the ability to automate tasks. Rather difficult to learn.
- **Configuring SSH**: Being able to configure and administer SSH is an important skill for any administrator.

ADVANCED CAPABILITIES

- **Screen**: Terminal multiplexer; user can start screen session and open any number of windows.
- **Tmux**: Alternative terminal multiplexer to Screen.
- **Crontabs**: Daemon (long running process that runs unattended) that executes commands at specific dates and times.
- **Systemctl**: Manages the systemd service manager, which offers on-demand starting of daemons, tracking processes, maintaining mount/automount points, etc. Rather difficult to learn.
- **Bash Scripting**: Scripting language that gives a user the ability to automate tasks. Rather difficult to learn.
- **Configuring SSH**: Being able to configure and administer SSH is an important skill for any administrator.

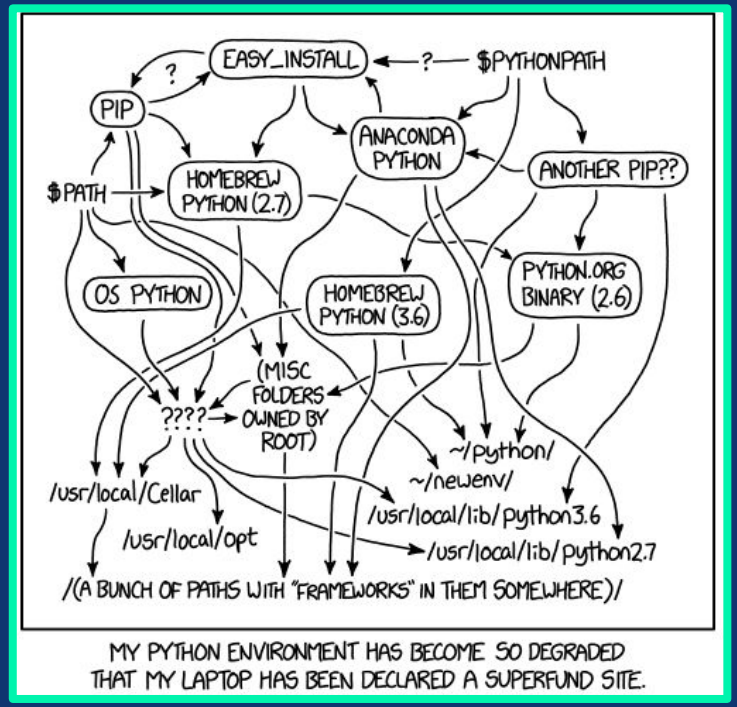

DEMO!

HERE WE GO... 💯

- 01 - LINUX OVERVIEW
- 02 - LINUX COMMAND-LINE INTERFACE (CLI)
- 03 - ADVANCED CLI TOOLS
- **04 - PYENV, VIRTUALENV, ANACONDA**
- 05 - JUPYTER NOTEBOOKS
- 06 - Q&A

PYTHON ENV. MANAGEMENT IS A NIGHTMARE!

- **pyenv**: Tool used to isolate different Python versions.
- **virtualenv**: Command line tool used to isolate different Python environments (Python versions + associated libraries). Highly recommend.
- **Anaconda**: Package manager for scientific computing, but has a lot of bloat. It's ready to use out-of-the-box, but is intensive. A less intensive version is **Miniconda**.




DEMO!

HERE WE GO... 💯

- 01 - LINUX OVERVIEW
- 02 - LINUX COMMAND-LINE INTERFACE (CLI)
- 03 - ADVANCED CLI TOOLS
- 04 - PYENV, VIRTUALENV, ANACONDA
- **05 - JUPYTER NOTEBOOKS**
- 06 - Q&A

EVERYONE KNOWS JUPYTER NOTEBOOKS...

- The most used Integrated Development Environment (IDE) for data science.
- Alternatives include: Google Colab, Amazon SageMaker Studio Lab, Zeppelin, and others.
- Given how hard it is to install Python and manage libraries, I suggest **Google Colab** as a faster place to start prototyping new ideas.

HERE WE GO... 💯

- 01 - LINUX OVERVIEW
- 02 - LINUX COMMAND-LINE INTERFACE (CLI)
- 03 - ADVANCED CLI TOOLS
- 04 - PYENV, VIRTUALENV, ANACONDA
- 05 - JUPYTER NOTEBOOKS
- **06 - Q&A**

THANK YOU!

ANY QUESTIONS?

Email: yli12313@umd.edu
Phone: 301-204-3957