Linux for Data Engineers - Detailed Revision Notes

# 1. Introduction to Linux for Data Engineers

* Walmart uses Linux for backend infrastructure, CLI automation, and system log analysis.
* Linux is preferred due to being open-source, customizable, secure, and efficient for automation.

# 2. Operating Systems: Linux vs macOS vs Windows

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Linux | macOS | Windows |
| Source | Open | Proprietary | Proprietary |
| Security | Strong, community-audited | Strong | Reactive |
| Customization | High | Low | Moderate |
| User Base | Developers, servers | Creative pros | General users |

Linux offers freedom and control with community-driven development.

# 3. Linux Fundamentals

* Created by Linus Torvalds in 1991.
* All Linux distros share the Linux kernel.
* Uses GNU Utilities.



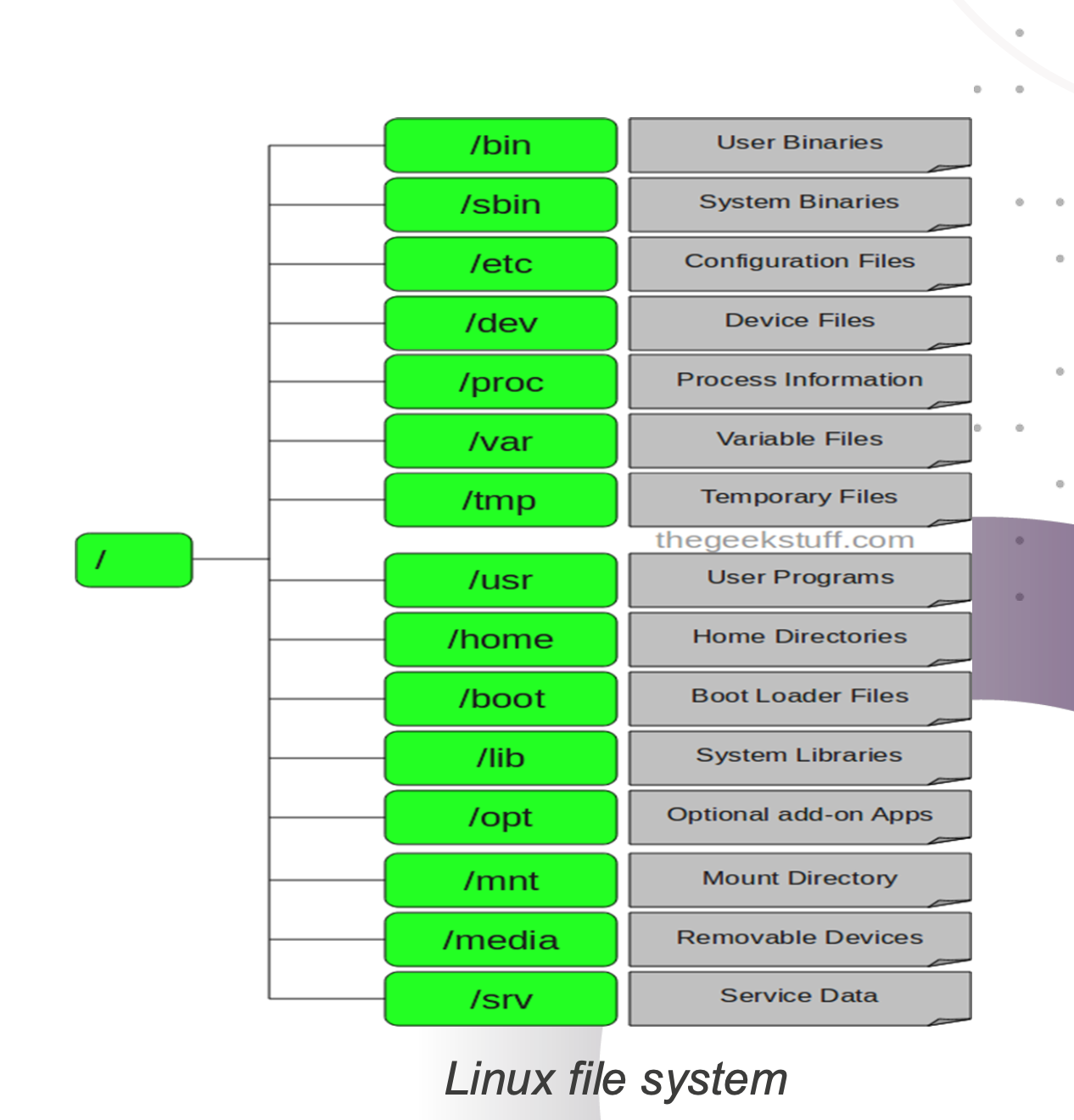
# 4. Linux File System Structure

* Unified under root directory /
* Everything is a file including memory, devices, directories, etc.

## Key Directories

|  |  |
| --- | --- |
| Directory | Purpose |
| /etc | Configuration files |
| /dev | Device files |
| /bin | Basic user commands |
| /sbin | System commands (admin only) |
| /var | Logs, mail, spool |
| /usr | User programs and libraries |
| /home | User home directories |
| /boot | Boot files (GRUB, kernel images) |
| /opt | Optional software packages |
| /proc, /sys | System & process info |
| /tmp, /var/tmp | Temp files |

# 



# 5. Linux Security and Permissions

* Root (superuser) has full privileges.
* Users and groups managed via /etc/passwd and /etc/group.
* Permissions: Read (r), Write (w), Execute (x).

chmod 755 filename

# 6. Basic Linux Commands

|  |  |
| --- | --- |
| Command | Description |
| ls | List files |
| cd | Change directory |
| pwd | Show current directory |
| cat | View file contents |
| echo | Print to screen |
| clear | Clear terminal |
| man | Manual pages |
| su | Switch user |
| passwd | Change password |
| df -h | Disk space |
| du -sh | File size summary |
| uname -a | System info |
| mkdir | Create directory |
| rm -rf | Delete directory and contents |
| touch | Create empty file |
| date | View/set date & time |
| ps | Process info |

# 7. Shell Scripting Basics

#!/bin/sh  
echo “Welcome, $USER”  
echo “Today is: $(date)”  
pwd

# 8. File Searching & Scheduling

* find /path -name "file.txt" – search for file
* grep "text" filename – search inside file
* crontab -e – Edit cron jobs

0 1 \* \* \* /home/user/backup.sh

# 9. Practical Lab Tasks

* System Health Monitor Script – write, execute, schedule.
* Use find, grep, chmod, at, cron.

# 10. Post-Session Reflection

* How is Linux used in your company?
* Which distro is used?
* Use cases and user feedback?
* How many servers run Linux?

# 11. Knowledge Check Answers

* Which directory controls devices? → /dev
* Command to delete non-empty directory? → rm -r