



Hyrje në shkencën e të dhënave

Pjesa 4 – UNIX / LINUX

Prof. Asoc. Dr. Ermir Rogova



What is UNIX?

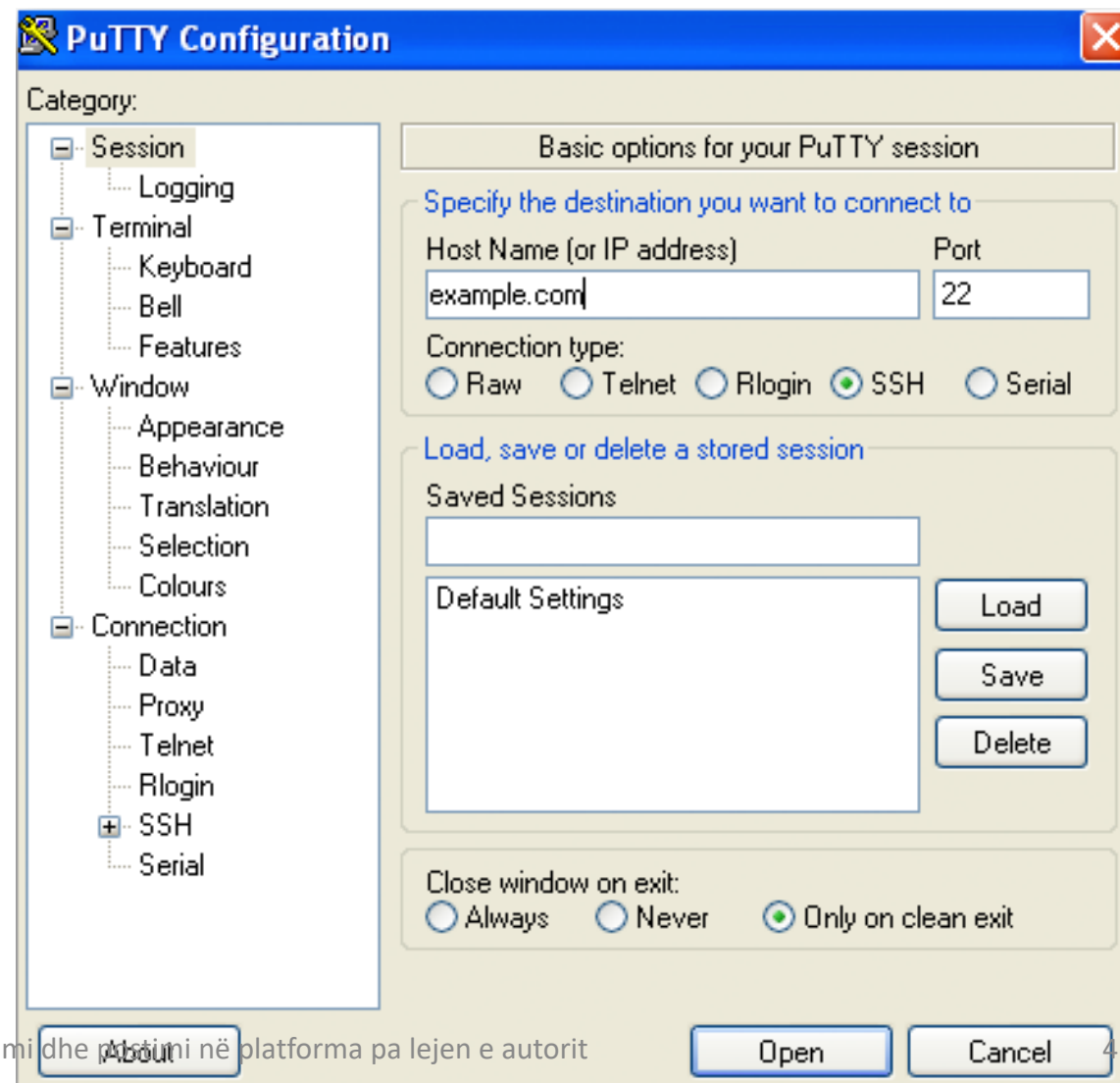
- The oldest (and perhaps the strongest) networked operating system
- Development platform
- Fine-tuned, very efficient programs and utilities to process data
- Most common variants: Linux, macOS

Getting access to UNIX

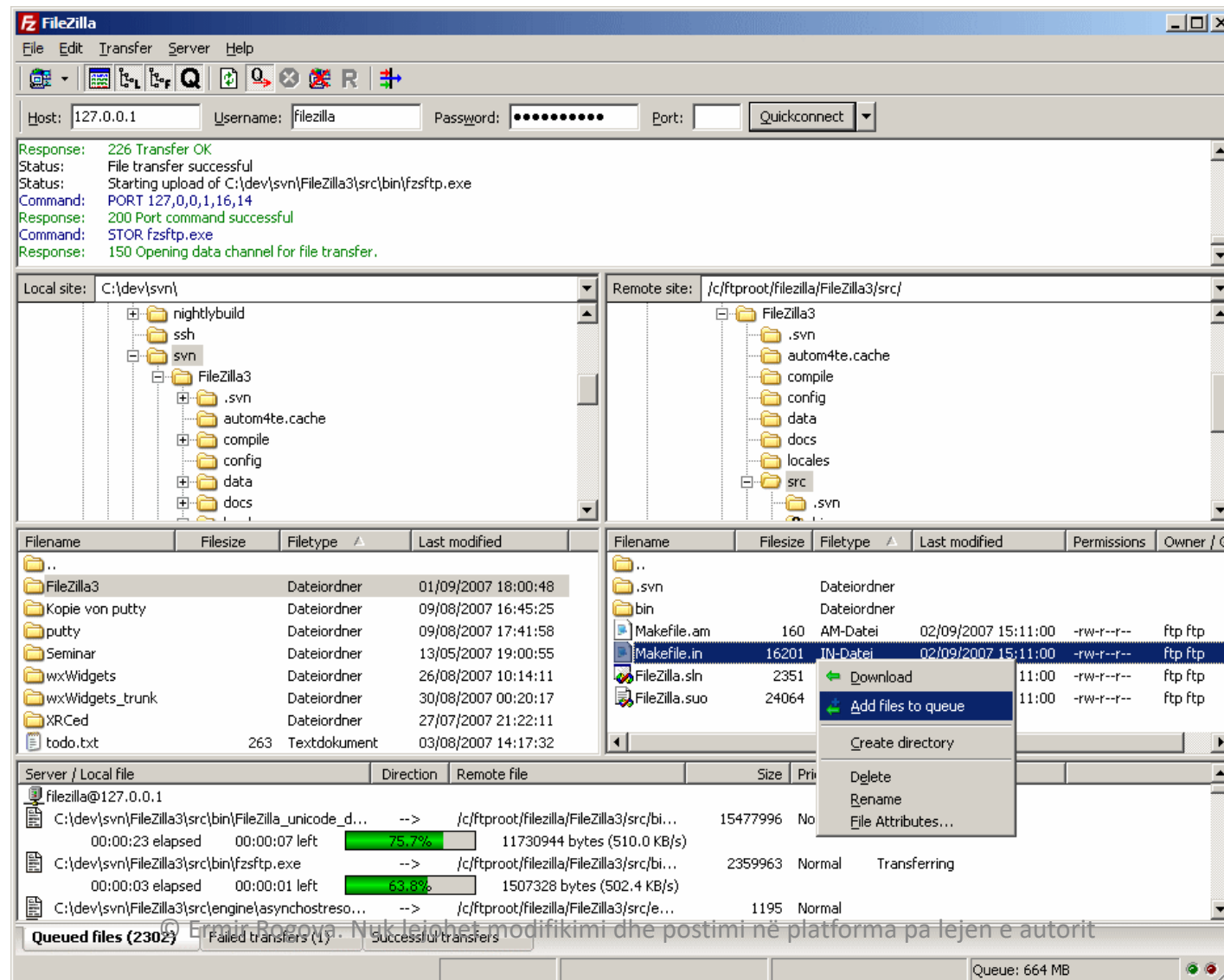
- Locally on a Linux or macOS machine
- Remotely through SSH and FTP
- On the cloud through AWS, Google Cloud, and Microsoft Azure

Connecting to a UNIX server

- Mac: run ssh command in the console/terminal app
- Windows: use software like PuTTY or WinSCP



File transfer with a UNIX server



Hands-on: basic UNIX commands

- pwd
- rm
- rmdir
- cd
- ls
- du
- wc
- cat
- more
- head
- tail
- man
- who

Hands-on: process-related UNIX commands

- Ctrl+c
- Ctrl+d
- ps
- ps aux
- ps aux | grep daffy
- top
- kill

Shortcuts

- Auto-complete
- Recall
- Search and recall
- Auto-path

Hands-on: editing on UNIX

- vi
- emacs
- (Alternatively) edit on your machine and FTP to the server

Hands-on: redirection and piping

- $\>$
- $\<$
- $\>\>$
- $|$

Hands-on: solving data problems on UNIX

- Small problems involving sorting, filtering, searching, merging, etc.
- Arithmetic operations
- File operations
- Working with a large CSV file (housing data)

Summary

- Many simple and frequent tasks with data can be done effectively and efficiently with UNIX commands.
- Most useful commands while dealing with data files: `wc`, `sort`, `uniq`
- Going forward, shell scripting/programming.

Temat e javës së pestë

- Python
- Getting Access to Python
- Basic Examples
- Control Structures
- Statistics Essentials
- Introduction to Machine Learning



Pyetje???