



varstation

Introdução ao Linux para Bioinformática clínica



ALBERT EINSTEIN

SOCIEDADE BENEFICENTE ISRAELITA BRASILEIRA

QUEM SOU

Biólogo (Graduação)
Bioinformática (PhD)



Pesquisador
Setor de P&D

HERMES PARDINI
Medicina, Saúde e Bem-estar

Pesquisador
Pós-doc



Laboratório
Nacional de
Computação
Científica

Bioinformata
HIAE/Varstation


ALBERT EINSTEIN
SOCIEDADE BENEFICENTE ISRAELITA BRASILEIRA

 varstation

OBJETIVOS

- § Abordar de forma teórico-prática os **conceitos e comandos básicos** na utilização de um sistema **Linux** visando capacitar os alunos a aplicar e analisar dados na prática da bioinformática.
- § Ao final do curso, o aluno deverá ser capaz de compreender a estrutura de um sistema Linux, acesso remoto à servidores, comandos básicos de manipulação de arquivos e pastas, execução de comandos, edição de arquivos e pipeline de análises.

PROGRAMAÇÃO

DIA 1

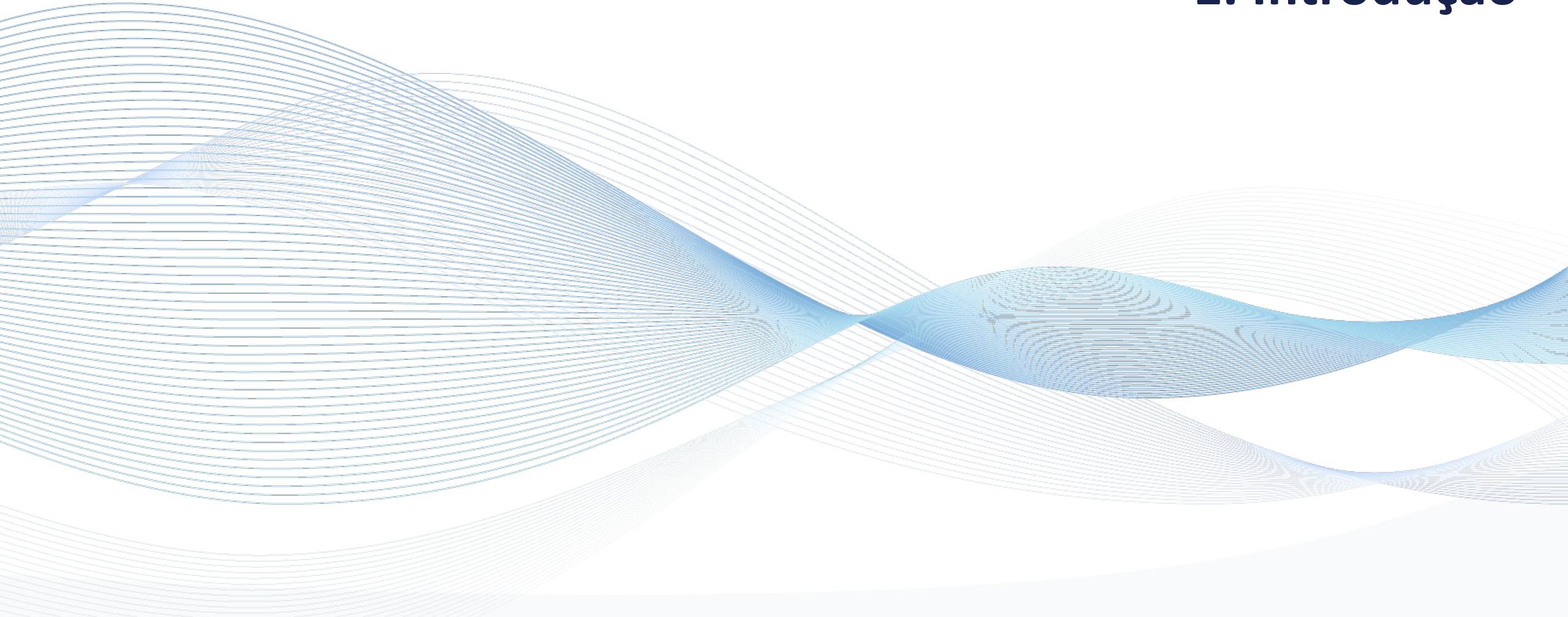
- Introdução à bioinformática e distribuições Linux
- Prática guiada: Acessando servidores e entendendo a estrutura do Linux

DIA 2

- Prática guiada: Os primeiros comandos básicos

DIA 3

- Prática guiada: Da manipulação de arquivos e pastas a execução de comandos e pipelines

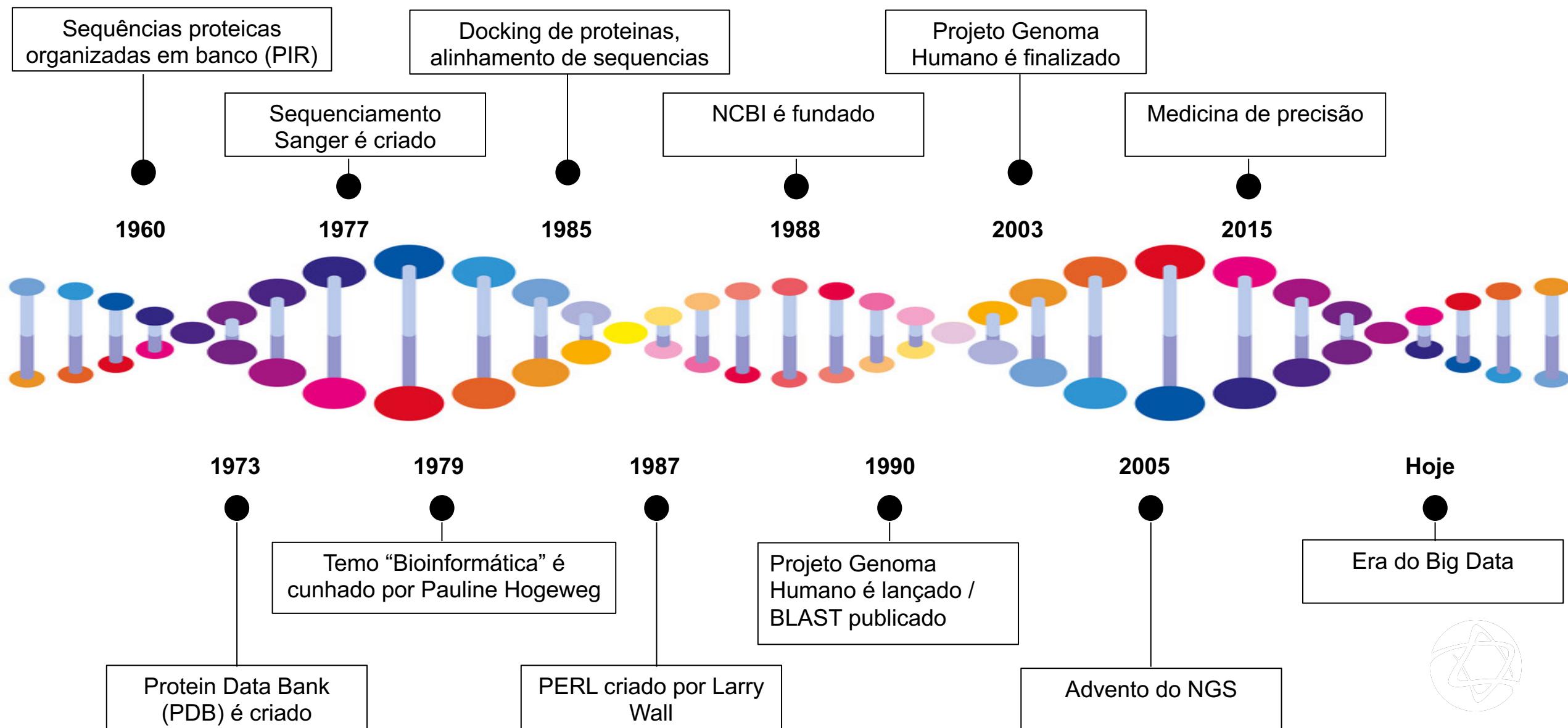
A large, abstract graphic on the left side of the slide features several sets of thin, light blue lines forming waves. These lines are more densely packed and darker blue in the center, creating a sense of depth and motion. They curve from the bottom left towards the top right.

1. Introdução

Conhecendo a classe

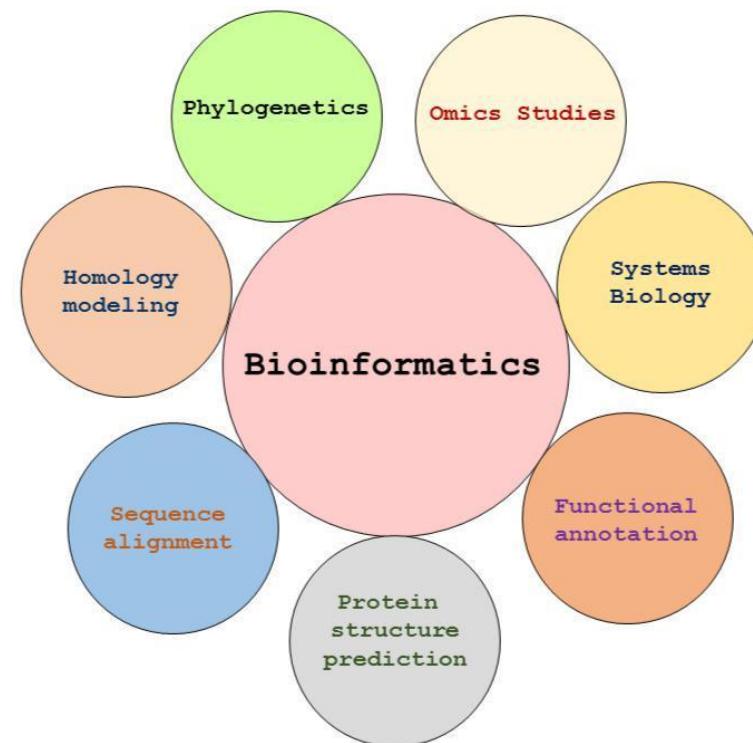
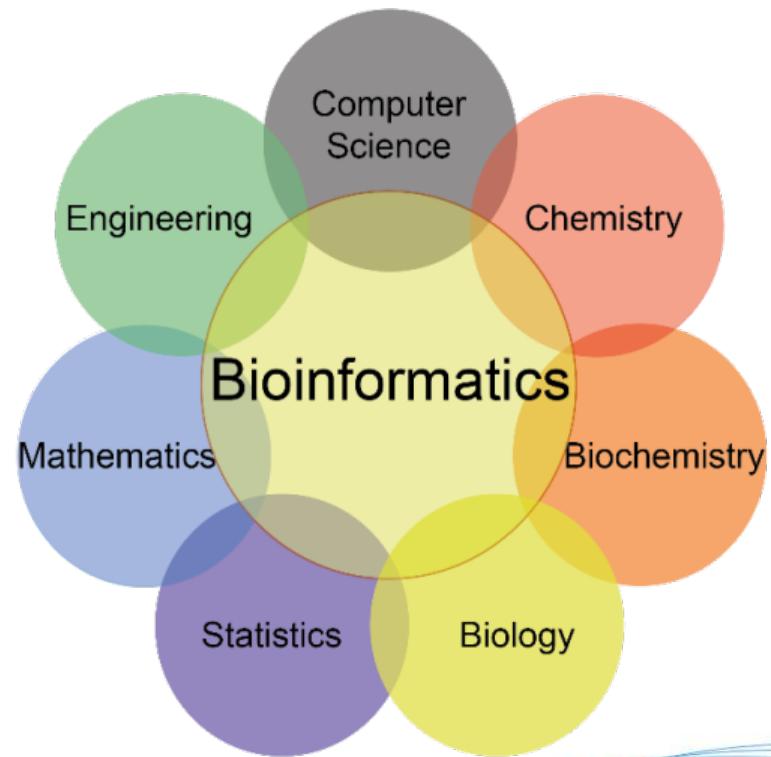
- Você já trabalha com dados biológicos?
- Você analisa dados em programas na web?
- Você já trabalhou com Linux?
- Você já trabalhou com dados de NGS?

Introdução – A Bioinformática

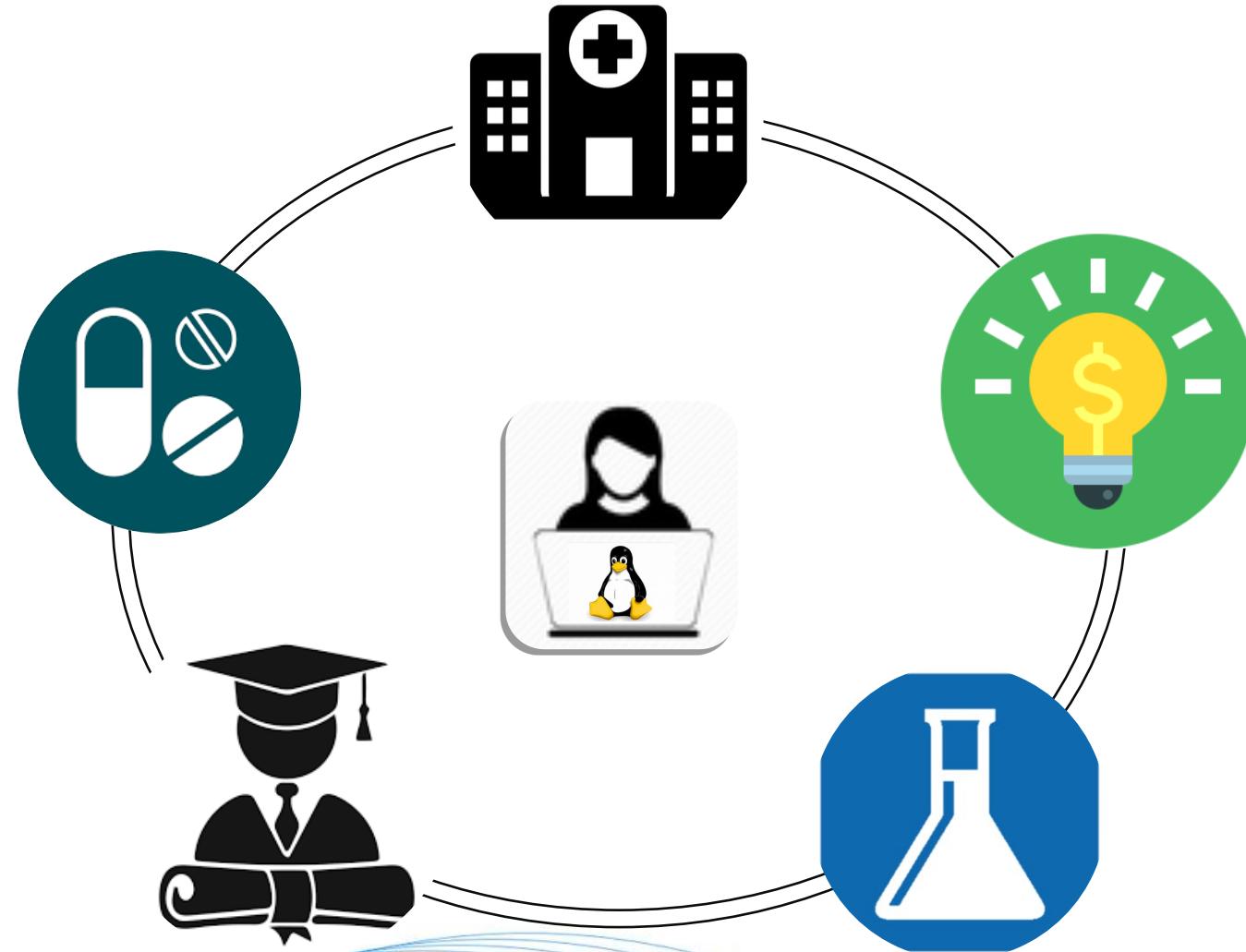


Introdução – A Bioinformática

“Área **interdisciplinar** de estudo e pesquisa dedicada ao desenvolvimento e utilização de ferramentas e métodos computacionais para a análise de dados biológicos”

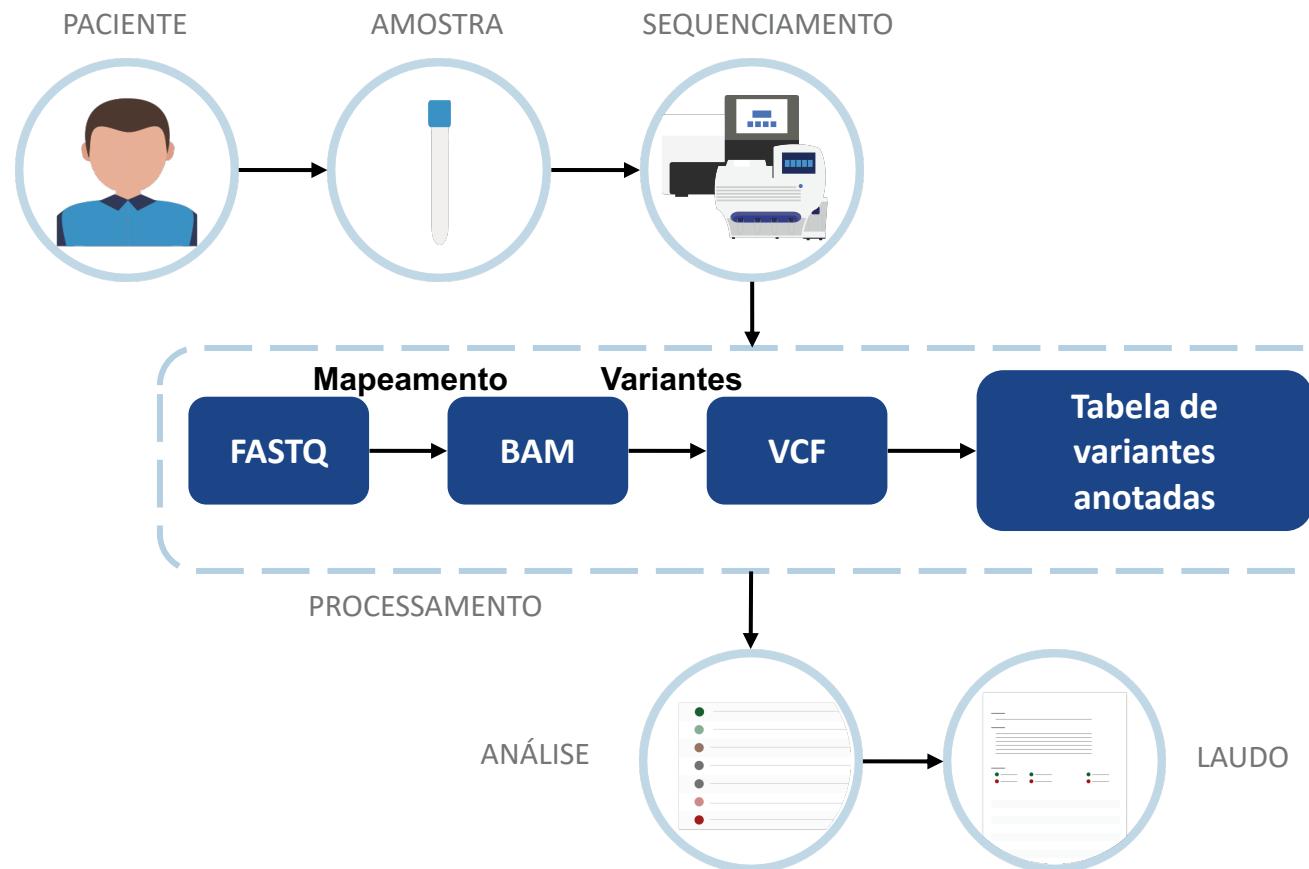


Introdução – O Bioinformata



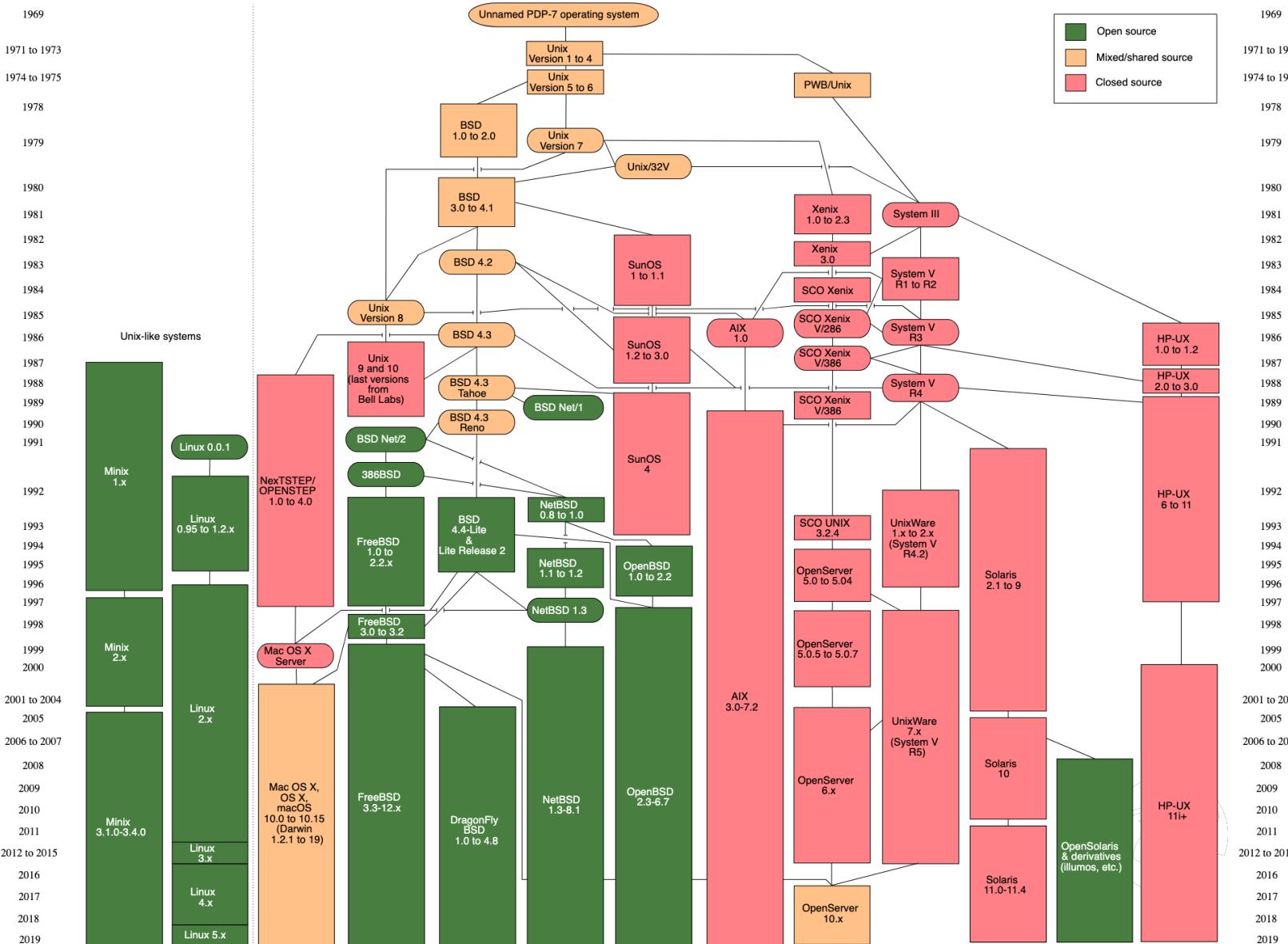
Introdução – Uma visão clínica

Um *pipeline* de NGS (*Next Generation Sequencing*)



Introdução – Linux

- UNIX: Sistema Operacional que começou a ser desenvolvido nos 1960s
- Linux, derivado do UNIX, criado por Linus Torvalds em 1991
- Open source / free
- Seguro, estável, confiável



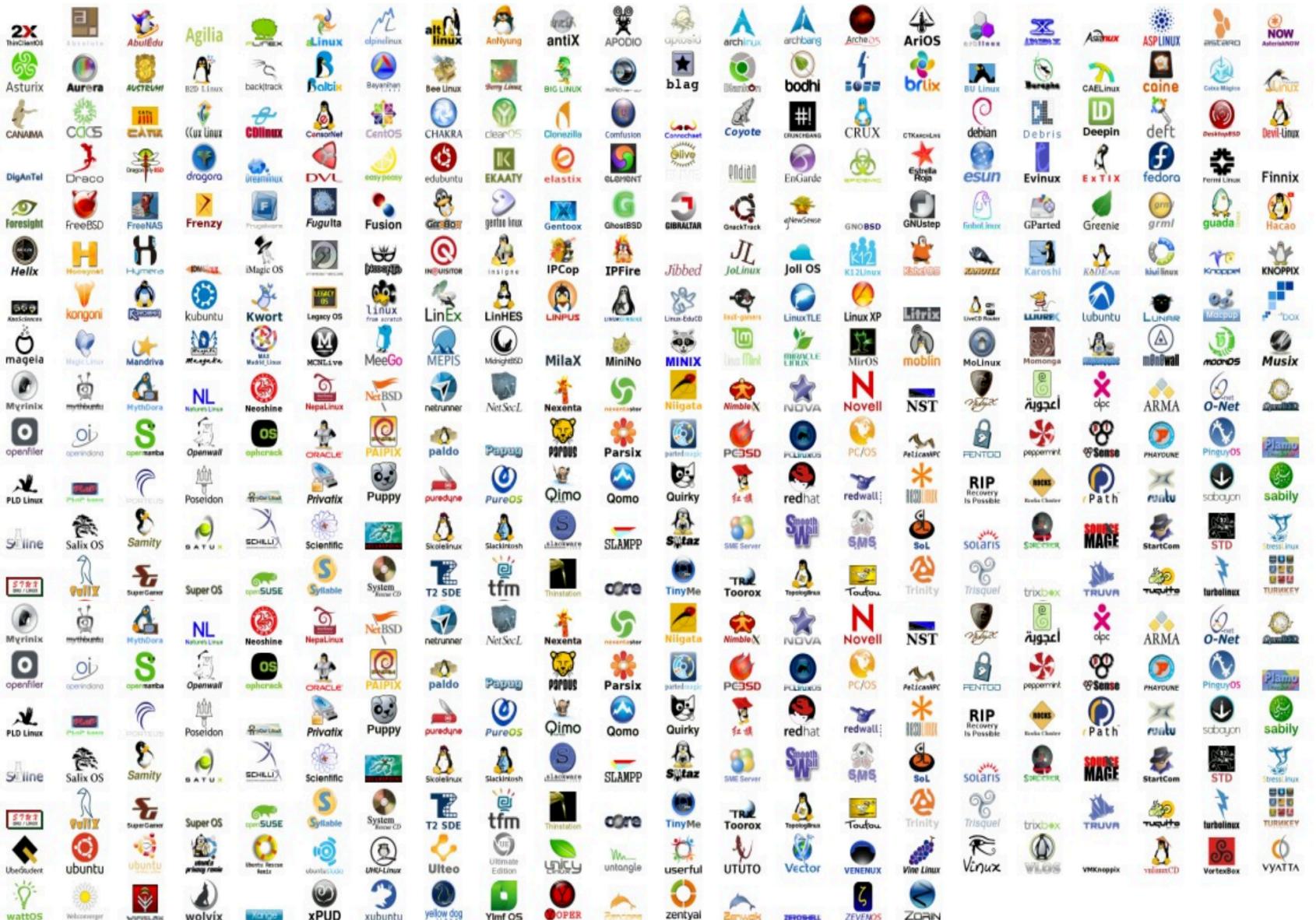
Introdução – Linux

O sistema operacional do Linux é composto de 3 partes principais: o **kernel**, o **shell** e os programas.

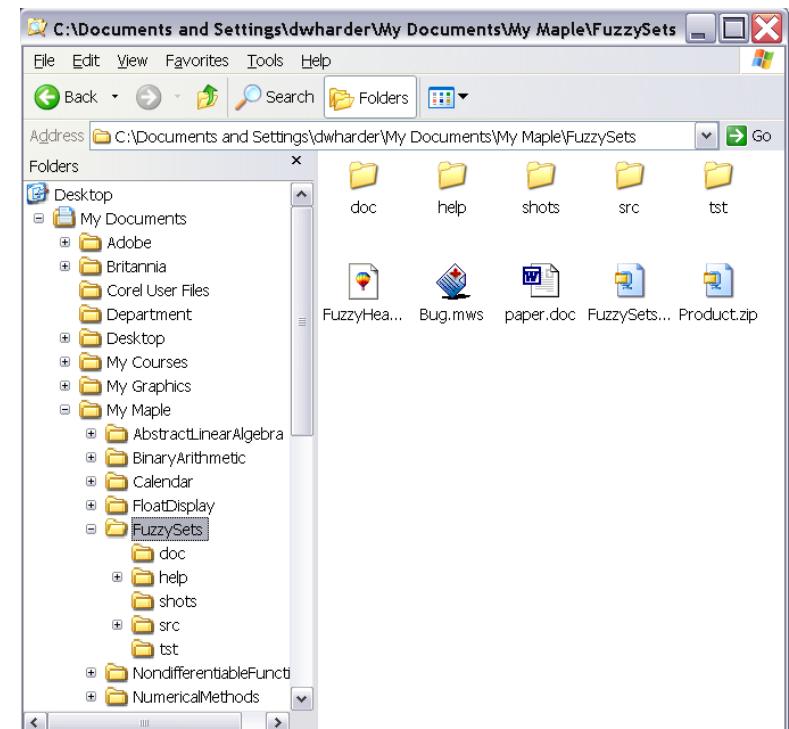
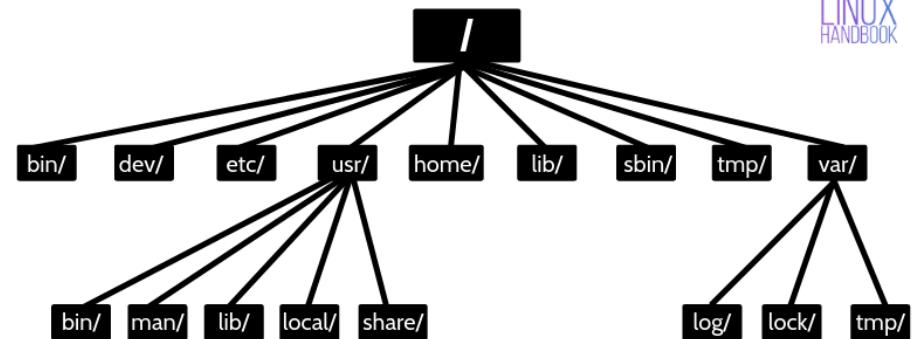
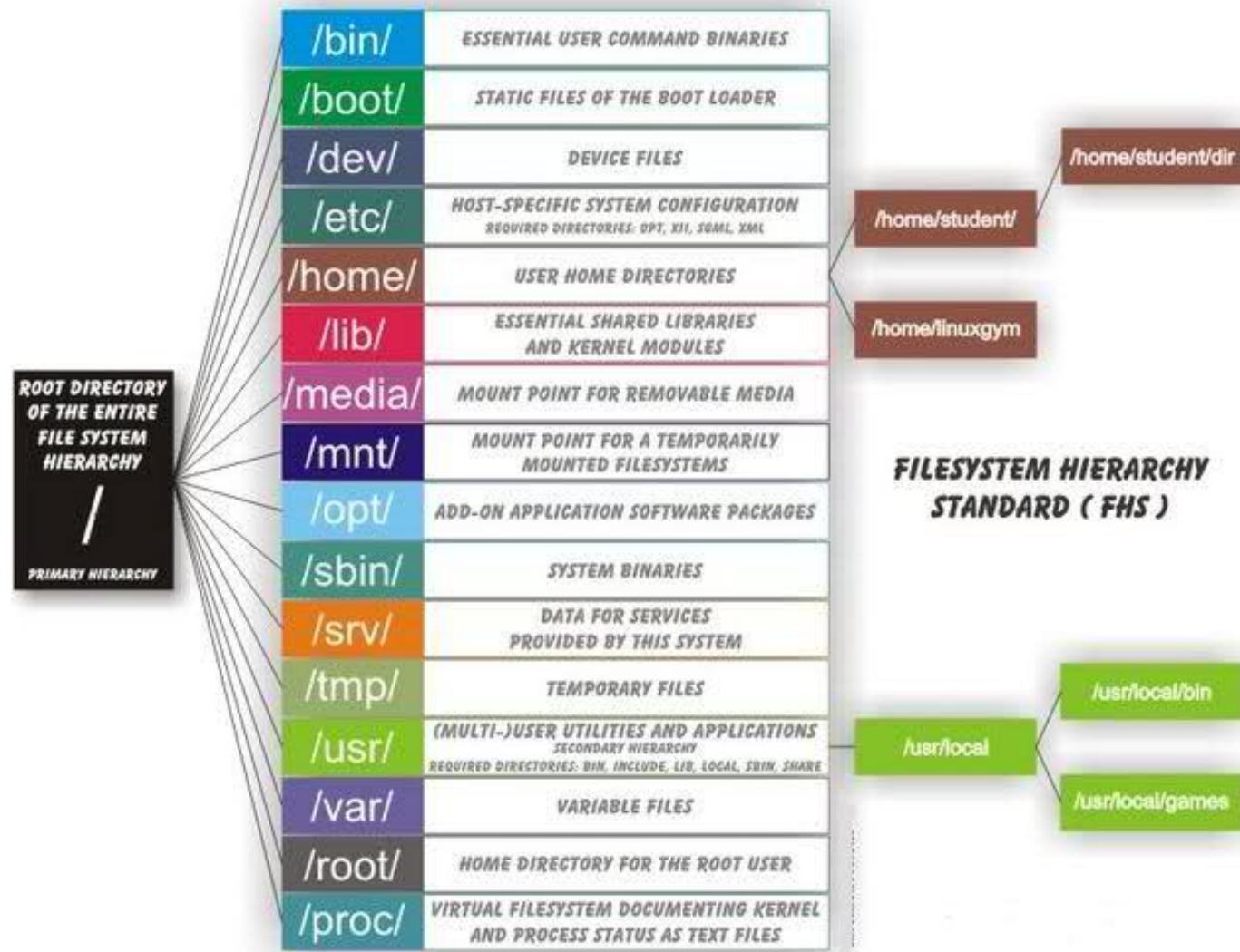
- **Kernel:** o centro do SO. Responsável por alocar processamento e memória aos programas e lidar com arquivos.
- **Shell:** É um *command line interpreter* (CLI) e funciona como interface entre o usuário e o kernel. Os comandos também são programas.
 - Tudo no UNIX ou é um processo (PID) ou um arquivo com dados.



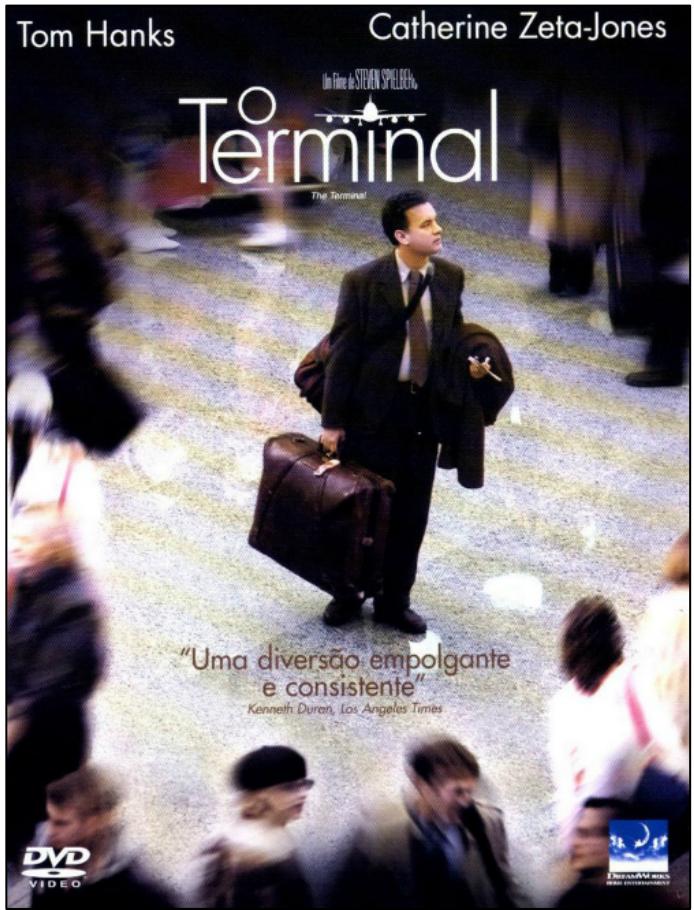
Introdução – Distribuições Linux



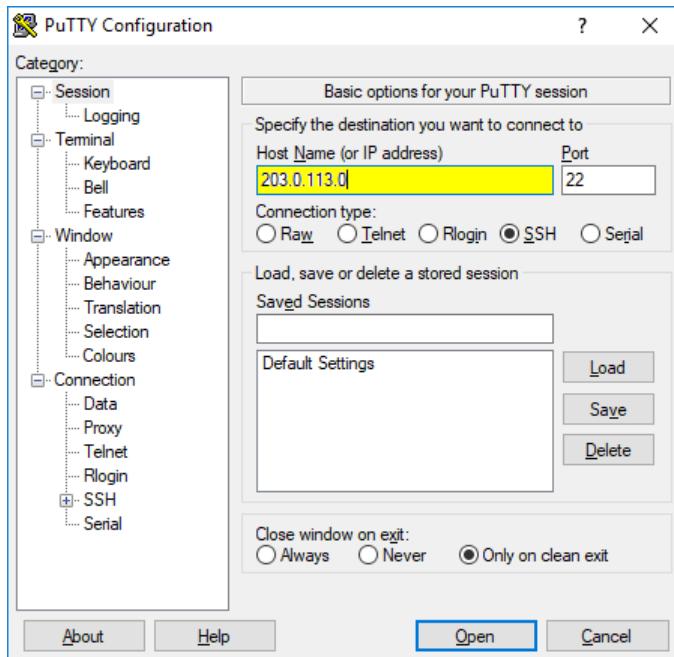
Introdução – Estrutura do Linux



Introdução – O Terminal



Introdução – O Terminal



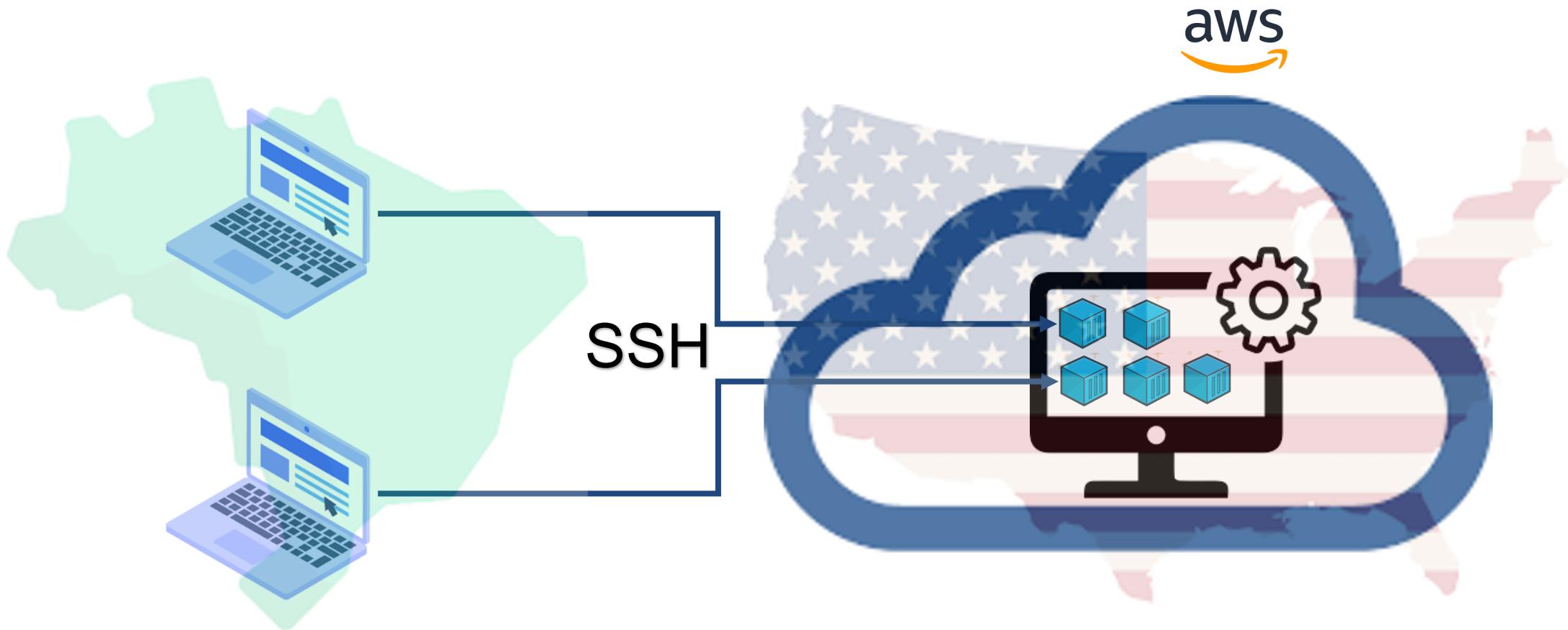
Finder->Applications->Terminal



Ctrl + Shift + T



Introdução – O servidor



Introdução – SSH



- SSH ("Secure SHell") é um protocolo para acessar com segurança outros computadores/servidores

```
$ ssh user@hostname
```

```
$ ssh joaosilva@111.11.1.11
```

```
$ ssh -i /path/to/keyfile user@hostname
```

```
$ ssh -p 1111 user@hostname
```

Introdução – SSH

- **IP (Internet Protocol)**: É o endereço da máquina na internet, onde ela pode ser localizada.
- **Porta**: É o endereço do serviço que será acessado em uma máquina
 - 80 → HTTP Hypertext Transfer Protocol
 - 443 → HTTPS Hypertext Transfer Protocol Secure
 - 21 → FTP (File Transfer Protocol)
 - 22 → SSH (Secure Shell)

Introdução – Terminologia

Termo	Significado
command	Nome do comando ou programa a ser executado
option	Opções fornecidas aos programas, geralmente com “-letra” sem espaços: ls -a
argument	Objetos passados para os comandos que sofrerão as ações previstas pelo mesmo.
prompt	Texto visível na linha de comando do terminal
Output	Resultado/saída de um comando
Pipeline	Junção de comandos em sequência, onde a saída de um serve de entrada para o próximo

Introdução – Terminologia

```
[severin@andrewseverin ~$ ssh bridges.psc.edu  
[severin@bridges.psc.edu's password:  
Last login: Wed Jul 17 17:21:55 2019 from gif-server.biotech.iastate.edu  
***** W A R N I N G *****  
You have connected to br006.pvt.bridges.psc.edu
```

This computing resource is the property of the Pittsburgh Supercomputing Center. It is for authorized use only. By using this system, all users acknowledge notice of, and agree to comply with, PSC policies including the Resource Use Policy, available at <http://www.psc.edu/index.php/policies>. Unauthorized or improper use of this system may result in administrative disciplinary action, civil charges/criminal penalties, and/or other sanctions as set forth in PSC policies. By continuing to use this system you indicate your awareness of and consent to these terms and conditions of use.

LOG OFF IMMEDIATELY if you do not agree to the conditions stated in this warning

Please contact remarks@psc.edu with any comments/concerns.

```
[severin@br006 ~$ ls  
boa edirect isugif jupyter_runtime_dir ncbi notebook.txt perl5 raw test_v1.dict  
[severin@br006 ~$ ls -l  
total 8  
lrwxrwxrwx 1 severin mc48o5p 24 Oct 19 2018 boa -> /pylon5/mc5fr5p/severin/  
drwxr-xr-x 2 severin mc48o5p 4096 Jan 10 2019 edirect  
lrwxrwxrwx 1 severin mc48o5p 30 Jun 21 2018 isugif -> /pylon5/mc48o5p/severin/isugif  
lrwxrwxrwx 1 severin mc48o5p 43 Mar 14 15:57 jupyter_runtime_dir -> /pylon5/mc48o5p/severin/jupyter_runtime_dir  
lrwxrwxrwx 1 severin mc48o5p 29 Apr 16 15:49 ncbi -> /pylon5/mc48o5p/severin/ncbi/  
-rw-r--r-- 1 severin mc48o5p 2221 Nov 28 2018 notebook.txt  
drwxr-xr-x 4 severin mc48o5p 40 Apr 13 2016 perl5  
lrwxrwxrwx 1 severin mc48o5p 23 Mar 14 2018 raw -> /pylon5/mc48o5p/severin  
-rw-r--r-- 1 severin mc48o5p 0 Aug 13 2018 test_v1.dict  
[severin@br006 ~$ not-a-valid-command  
-bash: not-a-valid-command: command not found
```

Welcome Message

Prompt

Command

Argument

Standard Out

Standard Error

Introdução – Dicas úteis (para a vida)

- **Tecla TAB:** utilizada para completar comandos e nomes de pastas e arquivos automaticamente. É recomendável SEMPRE utilizar, pois evita erros de digitação.
- 99% dos erros são **erros de digitação** ou pastas/arquivos não encontrados! Verifique seus comandos e localize-se na estrutura de pastas.
- **Histórico de comandos:** Seus últimos comandos ficam armazenados e podem ser acessados apertando a flechinha para cima do teclado. Evita digitar novamente e economiza muito tempo.
- Os **espaços** são importantes! Certifique-se que eles estão nos lugares certos após cada comando/argumento. Espaço demais OK, espaço de menos NAO!
- Trabalho em equipe: verifique sempre a utilização do servido na hora do trabalho. Tem outros processos ativos?
- Quer praticar em ambiente Linux? <https://bellard.org/jslinux/>



varstation

Vamos praticar com comandos?

<https://github.com/Varstation/curso-basico-linux>



ALBERT EINSTEIN

SOCIEDADE BENEFICENTE ISRAELITA BRASILEIRA

rafael.muniz@varstation.com