Unix Shell - Introduction

About the Instructor

About the Instructor

▷ Researcher at the [Advanced Institute for Artificial Intelligence](#);

▷ Associate Researcher at the [São Paulo State University - UNESP](#);

▷ Member of the [SPRACE project](#) (a [CMS LHC collaboration](#));

▷ Chair of the CODATA-RDA Data School groups since 2018;

▷ RDA Technical Advisory Board since 2022;

▷ Background on Artificial Intelligence and High Performance Computing;

Background

▷ Human - Computer Interaction
▷ Graphical User Interface - *GUI*
  ○ (Personal Computers);
  ○ Heavy use of the Mouse;
  ○ "Menu-driven interactions";
  ○ Very intuitive to learn;
  ○ Scales very poorly;

Background (continued)

Imagine the following task:

For a literature search, you have to:

1. copy the third line
2. of one thousand text files
3. in one thousand different directories and
4. paste it into a single file

Background (continued)

One thing we know for sure:

The Shell

▷ Both a **command-line interface** (CLI) and a **scripting language**;

▷ Allowing repetitive tasks to be done automatically... and fast;

▷ Bourne Again SHell - BASH
  ○ Default shell on most modern implementations of Unix;

▷ Maybe not as easy as the GUI 😒;

▷ Easiest way to interact with supercomputers

The Shell (continued)

▷ Essential to use a variety of specialized tools and resources including high-performance computing systems;
  ○ Clusters, clouds, etc.
▷ Need to learn a few commands;
  ○ Like a vocabulary of a new language;
  ○ A small number of "words" gets you a long way;
▷ Let's get started!

SW Carpentry Materials on Bash

▷ ~~Setup~~

▷ ~~Working with Files and Directories~~

▷ Working with Pipes and Filters

▷ For Loops

▷ Shell Scripts

▷ File Permissions

▷ Job Control

Nelle's Pipeline: A Typical Problem

▷ Example that we will use throughout the lessons;
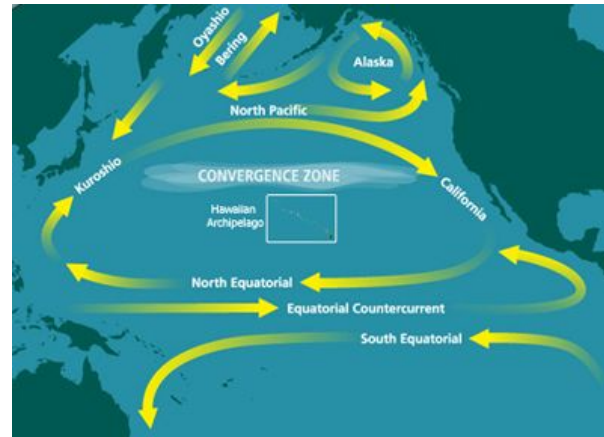▷ Six-month survey data containing samples of gelatinous marine life in the [Great Pacific Garbage Patch](#);

Image from [Wikipedia](#)

Nelle's Pipeline: A Typical Problem (continued)

▷ 1520 samples;

▷ Measure the relative abundance of 300 proteins;

▷ Perform the analysis using imaginary program called goostats.sh (GUI);

  ○ Select and open a file **1520 times**;

  ○ 30 seconds to run each file;

    ■ **More than 12 hours** of Nelle's attention

▷ Assign a computer this task while she focuses on writing a paper about this results;