Outlook

Meeting assets for GeoComp & ML 2025 course are ready!

From Zoom <no-reply@zoom.us>
Date Thu 9/18/2025 10:04 AM
To Amatulli, Giuseppe <giuseppe.amatulli@yale.edu>

zoom

Meeting assets for GeoComp & ML 2025 course are ready!

Meeting summary

Quick recap

The meeting focused on setting up and verifying the virtual machine environment for a blockchain course in auction environments, with Giuseppe providing guidance on Linux terminal usage and command line navigation. The session covered essential Linux commands, file manipulation techniques, and data processing approaches, including demonstrations of text manipulation, sorting methods, and variable usage in Bash scripting. The class was introduced to JupyterLab for data manipulation and presentation, with discussions about file handling, debugging techniques, and guidance for upcoming lessons on advanced operations and machine learning applications.

Next steps

- All students: Ensure they have the virtual machine running with the terminal and the "/media/sf_mysedata" directory accessible.
- All students: Familiarize themselves with basic Bash commands for text file manipulation.
- All students: Practice using the "man" command to search for and understand command documentation.
- All students: Learn the difference between text files and binary files and when to use each format.
- All students: Practice using special characters for file queries in Bash.
- All students: Utilize tab auto-completion to avoid path mistakes when navigating directories.
- All students: Learn and practice Bash keyboard shortcuts for more efficient command line navigation.

Summary

Blockchain Auction Course VM Setup

The meeting focused on setting up and verifying the virtual machine environment for a course on blockchain in auction environments. Giuseppe confirmed that participants should have the virtual machine running with the necessary terminal access and media files, including Git clone and my SC data. He noted that recording was active for the session, and the team was preparing to begin the first lecture on blockchain usage in auction environments.

Linux Terminal and Bash Commands

Giuseppe explained the structure of the Linux terminal prompt and Bash command syntax, including the user, hostname, and current directory display. He demonstrated how to use the manual (man) command to search for and view help documentation for specific commands, particularly showing how to search for and navigate through command options using the slash key. Giuseppe emphasized that Bash remains a fast and efficient tool for text manipulation, especially for large files, and can be used as a preprocessing step before working with Python or R.

Linux Text and Binary Files

Giuseppe explained the difference between text and binary files in Linux, noting that text files are human-readable and can be created by saving documents in formats like CSV or rich text, while binary files require specialized software for reading. He demonstrated how to redirect terminal output to text files using commands like "redirect" and showed various methods to open and navigate text files, including using commands like "cat", "head", "tail", "more", and "less". The discussion concluded with an explanation of special characters in file queries, including the use of asterisks, question marks, and brackets for pattern matching, and how to properly use quotation marks to avoid special character interpretation.

Linux Command Line Navigation Training

Giuseppe led a training session on Linux command line navigation and piping, covering essential commands like LS, CD, and directory navigation using special characters. He demonstrated features like autocomplete (using tab completion), keyboard shortcuts (including Ctrl-A, Ctrl-E, and Ctrl-R for reverse search), and the concept of piping commands. The session included a discussion about the differences between piping directly versus writing to a temporary file first, with Olha correctly identifying that the first method processes data in RAM while the second method involves file I/O.

Programming Performance and Memory Management

Giuseppe discussed the performance trade-offs between different programming approaches, explaining that while one method is faster, it has limitations based on RAM size. He emphasized the importance of understanding how different programming languages handle memory and recommended using software like GDAL that works with smaller data chunks rather than loading entire files into RAM. The meeting included a technical discussion about using Jupyter, where Alejandro experienced some loading issues due to high resource usage from multiple running applications, but was able to resolve the problem.

JupyterLab Data Manipulation Tutorial

Giuseppe guided the class on using JupyterLab for data manipulation and presentation. He explained how to access and use the local file system, including the sharing folder and MIC data, through JupyterLab. Giuseppe demonstrated how to create and manage cells in JupyterLab, switching between code and markdown formats, and introduced basic text manipulation techniques

using the txt command. He also explained the use of the /tmp directory for storing temporary files during trials.

Linux Aliases and Data Sorting

Giuseppe explained the concept of aliases in Linux commands and demonstrated how to use the LL alias for listing files with detailed information including size and creation details. He showed how to query and sort data in text files, emphasizing the importance of using the correct separator (space by default) and the key parameter for sorting specific columns. The discussion included tips for handling data imports into Python and R, particularly regarding consistent spacing in datasets.

Bash Sorting and Variable Techniques

Giuseppe explained sorting techniques in bash, including how to sort specific columns using the "-k" option and the importance of using the "-G" flag for general numerical sorting instead of the default binary sorting. He demonstrated how to assign and print variable values in bash, emphasizing the use of meaningful variable names and avoiding spaces in file and folder names. The discussion concluded with an introduction to using variables in loops for repeating processes, though the transcript ended before this topic was fully explored.

Bash Text Processing Techniques

Giuseppe demonstrated different approaches to processing text files using Bash, including for loops with variable iteration and conditional statements. He explained the computational differences between processing data at the end of a loop versus intermediate processing, noting that for small files the difference is negligible but becomes significant with large datasets. Several students, including Magda and Nafeesa, shared their experiences working with large text files, with Nafeesa describing her work with environmental data from tree monitoring devices. Giuseppe also covered debugging techniques in Bash, emphasizing the importance of using print statements to track variable values and process progress, and introduced the use of the "time" command to measure computational efficiency.

Data Migration and Analysis Instructions

Giuseppe instructed the class to start using the Pasha system and migrate text files to the virtual machine, suggesting they use well-structured files for efficient processing. He explained that next week's lesson will cover more advanced operations like column manipulation using AWK. Alejandro asked about working with LA/LAS files, and Giuseppe advised using PK tools for binary files, though conversion to text files might be necessary. Hager discussed his genetic population analysis project involving 30 samples from three locations, with plans to use machine learning techniques, and Giuseppe provided guidance on how to structure the data for machine learning applications.