

Syllabus: Unix Tools: Data, Software and Production Engineering

1. Course overview

Processing information is the hallmark of all modern organizations, which are increasingly digital: absorbing, processing and generating information is a key element of their business. Being able to interact flexibly and efficiently with the underlying data and software systems is an indispensable skill. Knowledge of the Unix shell and its command-line tools boosts the effectiveness and productivity of software developers, IT professionals, and data analysts.

The Unix tools were designed, written, actively used and refined by the team that defined the modern computing landscape. They allow the performance of almost any imaginable computing task quickly and efficiently by judiciously combining key powerful concepts. The power of Unix tools for exploring, prototyping and implementing big data processing workflows, and software engineering tasks remains unmatched. Unix tools, running on hardware ranging from tiny IoT platforms to supercomputers, uniquely allow an interactive, explorative programming style, which is ideal for the efficient solution of many of the engineering and business analytics problems that we face every day.

The course is uniquely based on carefully-selected, interactive walk-through examples that demonstrate how each command operates in practice. The examples that we use involve problems that engineers and analysts face every day.

2. Learning objectives

At the end of this course, you will be able to:

- Enter and combine commands in the Unix command line
 - Use files, data pipelines, variables and control structures
 - Select the most useful tools and commands for fetching, selecting, generating, processing, summarizing and reporting data
 - Obtain data from databases, cloud-based hosts, version control systems, object files, archives and your desktop files
 - Accomplish diverse processing tasks by putting together suitable commands and configuring their execution parameters
-

3. Course structure

The course is organized in *six modules*. We start by introducing the key ideas and advantages of Unix tools. We then build many one-liners around the tools that follow a sequence of fetching, selection, processing, summarization and reporting. We demonstrate the tools for each phase: the plumbing that joins the parts into a whole, more specialized commands for handling software development and system administration tasks, sound and images and even graph-structures. We end with a discussion of common use patterns to follow and anti-patterns to avoid.

A brief summary of each unit is presented below.

1. Foundations

The first module covers the fundamentals: getting to the command line; introducing the command-line interface; performing input and output redirection; handling files and directories; providing command-line arguments; grouping commands; scripting; execution control; and an overview of the data processing flow covered by the next units.

2. Data Fetching

The second module covers how you can obtain data from diverse sources: remote services, archives, remote hosts, the file system, version control systems, compiled code, graphical desktop systems, system administration entities, and even generators.

3. Selection

The third module covers selecting data with regular expressions and the grep family of tools and other relevant ones.

4. Processing

The fourth module covers processing through sorting and working with sorted data, the mini-languages of the sed and awk tools, obtaining file differences, testing, evaluating expressions, and dealing with characters, lines, graphs, images, and sound.

5. Reporting

The fifth module covers reporting through summarizing, emailing, and text formatting.

6. Professionalism

Finally the last, sixth module covers good shell practices and advice for improving your shell style.

Please note that the estimated workload for this course is between 4 and 6 hours per module.

4. Resources, Tools & Browsers

All educational resources will be available in the course. They consist of short videos and mostly quizzes to support you in the completion of the course topics.

To follow the course, you will need to obtain a Unix command-line prompt. How you get to the Unix command line depends on the system you're using. Please watch video 1.2.1 Getting to the command line to learn about your options under Windows, macOS or the Ubuntu GNU/Linux distribution.

We support the following browsers: Chrome, Firefox and Safari.

5. Licence

The course materials of this course are Copyright Delft University of Technology and are licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License](#).

If you choose to reuse or repost DelftX course materials you must give proper **attribution**. Please utilize the following citation and refer to this MOOC:

"[TITLE OF WORK –with hyperlink to material] by TU Delft / [LECTURER NAME -with hyperlink to lecturers page] is licensed under [CC-BY-NC-SA 4.0](#). This material was created by or adapted from material posted on [TITLE MOOC- with link to MOOC start page]."

Or if it is a derivative please use following citation:

"This work [Your title] by [Your name] is a derivative of "TITLE OF WORK –with hyperlink to material] by TU Delft / [LECTURER NAME -with hyperlink to lecturers page] and (re)licensed under [CC-BY-NC-SA 4.0](#). This material was created by or adapted from material posted on [TITLE MOOC- with link to MOOC page]."

