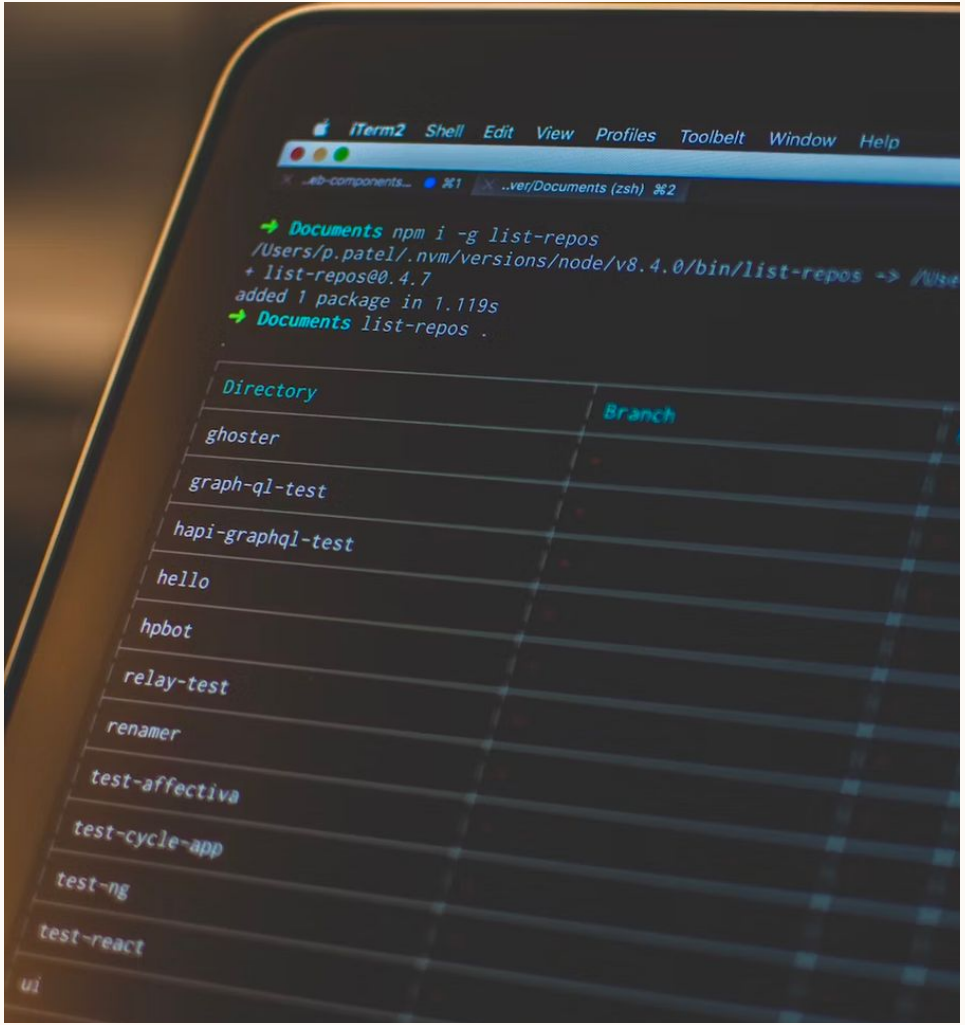
Command Line Interface
Fundamentals

DCA0305
ivanovitch.silva@ufrn.br

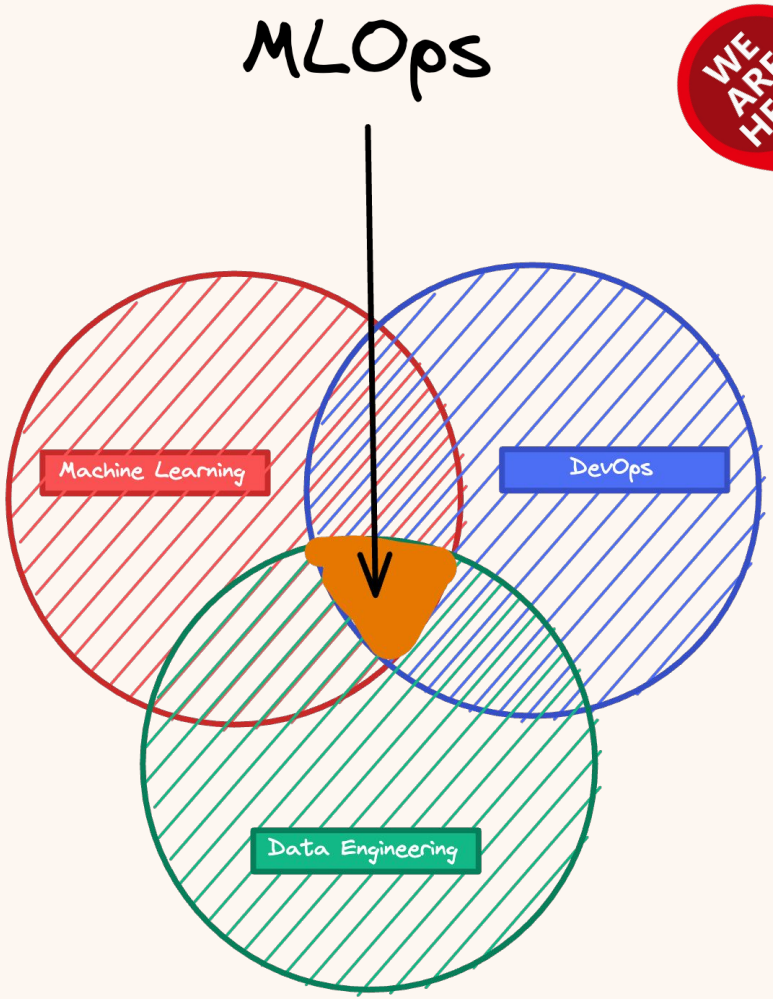


GitHub Codespaces GPU Limited Beta Update

codespaces

August 24, 2023

GitHub is no longer admitting new users or organizations to the limited beta for GPU-powered Codespaces due to limited capacity for this virtual machine type. Existing beta participants will be able to continue using these machine types, however no new users on the current waitlist will be granted access. For any updates on features we're working on and what stage they're in, please follow the GitHub public roadmap.



WE ARE HERE

Python Essentials for MLOps

CLI Fundamentals	Clean Code Principles	Production Ready Code	Programming
Elements of the Command Line	Refactoring Documentation	Catching Errors Logging	Functions Classes Decorators
Infrastructure Github Codespace vscode, colab, terminal	Python Code Quality Authority (PCQA)	Testing	Interact with APIs and SDKs to build command-line tools

Quickstart

Create Codespace

Connect to a foward port

View your running application

Publish your codespace

Personalize your setup

1. Hello World
2. Create a Codespace using an existing repository

https://docs.github.com/en/codespaces/getting-started/quickstart

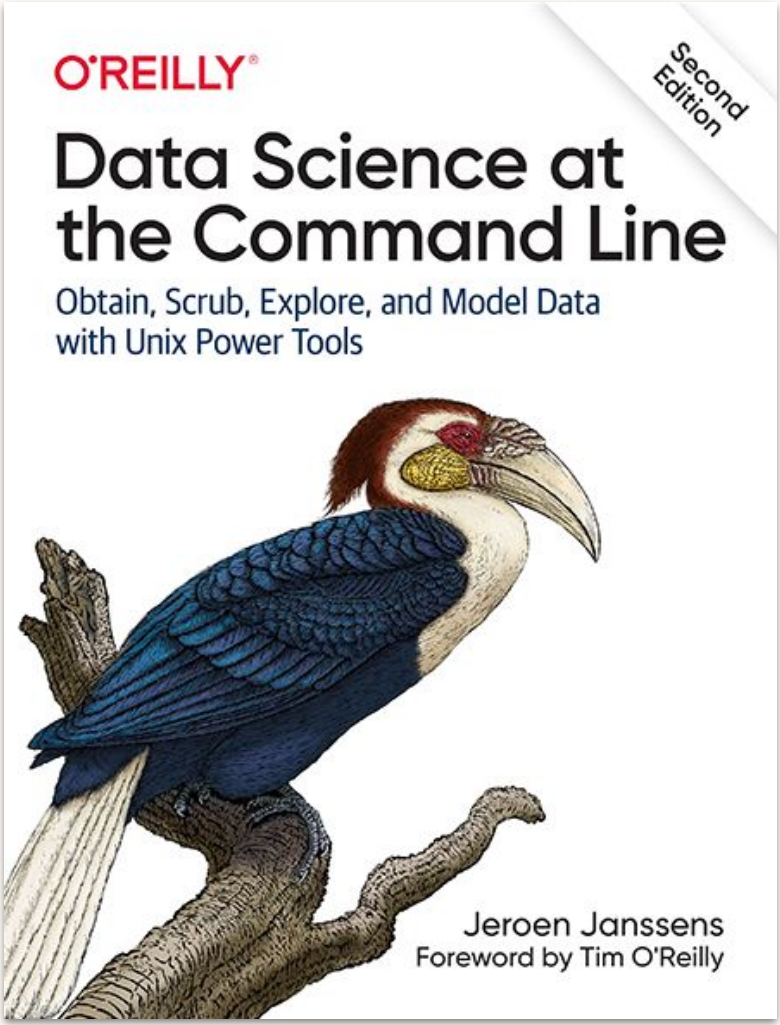
Deep dive into GitHub Codespaces

Understand how GitHub Codespaces works.

GitHub Codespaces is an instant, cloud-based development environment that uses a container to provide you with common languages, tools, and utilities for development. GitHub Codespaces is also configurable, allowing you to create a customized development environment for your project. By configuring a custom development environment for your project, you can have a repeatable codespace configuration for all users of your project.

https://docs.github.com/en/codespaces/getting-started/deep-dive

Instant Cloud-Based Normalized Development



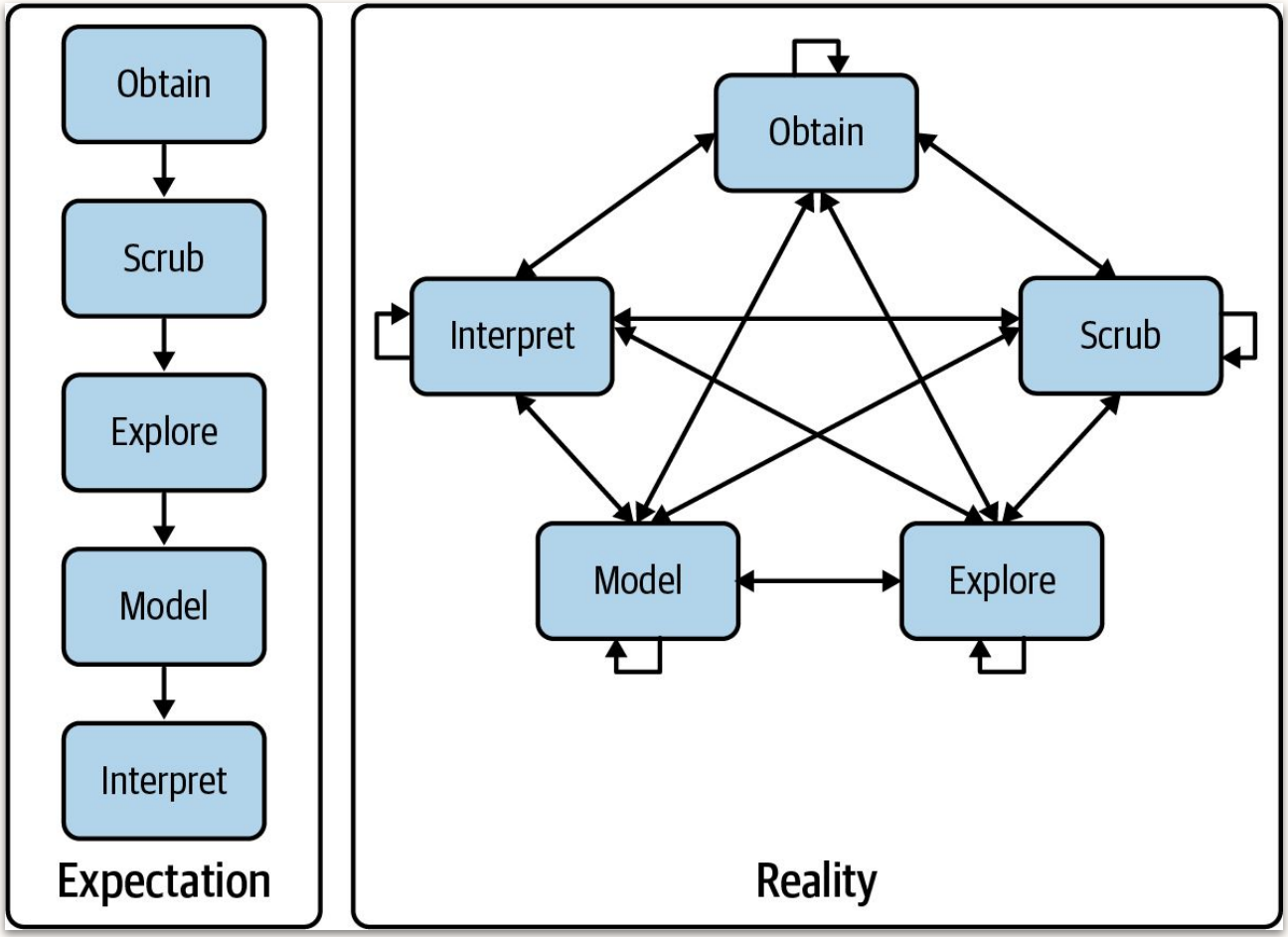
This thoroughly revised guide demonstrates how the flexibility of the command line can help you become a more efficient and productive data scientist. You'll learn how to combine small yet powerful command-line tools to quickly obtain, scrub, explore, and model your data. To get you started, author Jeroen Janssens provides a Docker image packed with over 100 Unix power tools—useful whether you work with Windows, macOS, or Linux.

Traditional computer and data science curricula all too often mistake the command line as an obsolete relic instead of teaching it as the modern and vital toolset that it is. Only well into my career did I come to grasp the elegance and power of the command line for easily exploring messy datasets and even creating reproducible data pipelines for work. The first edition of Data Science at the Command Line was one of the most comprehensive and clear references when I was a novice in the art, and now with the second edition, I'm again learning new tools and applications from it.

This book explains how to integrate common data science tasks into a coherent workflow. It's not just about tactics for breaking down problems, it's also about strategies for assembling the pieces of the solution.

https://jeroenjanssens.com/dsatcl

Doing data science is an iterative and non-linear process



https://jeroenjanssens.com/dsatcl

What does each of these commands do?

```
$ pip install scikit-learn
Requirement already satisfied: scikit-learn in /usr/lib/python3.6/site-packages

$ cd ~/.ssh
$ ssh-keygen
$ cat ~/.ssh/id_rsa.pub | pbcopy

$ curl -s 'http://api.citybik.es/v2/networks/citi-bike-nyc' |
> jq '.network.stations[].free_bikes' |
> paste -sd+ | bc
7572
```

Before ChatGPT ...

https://explainshell.com

Why Data Science at the Command Line?

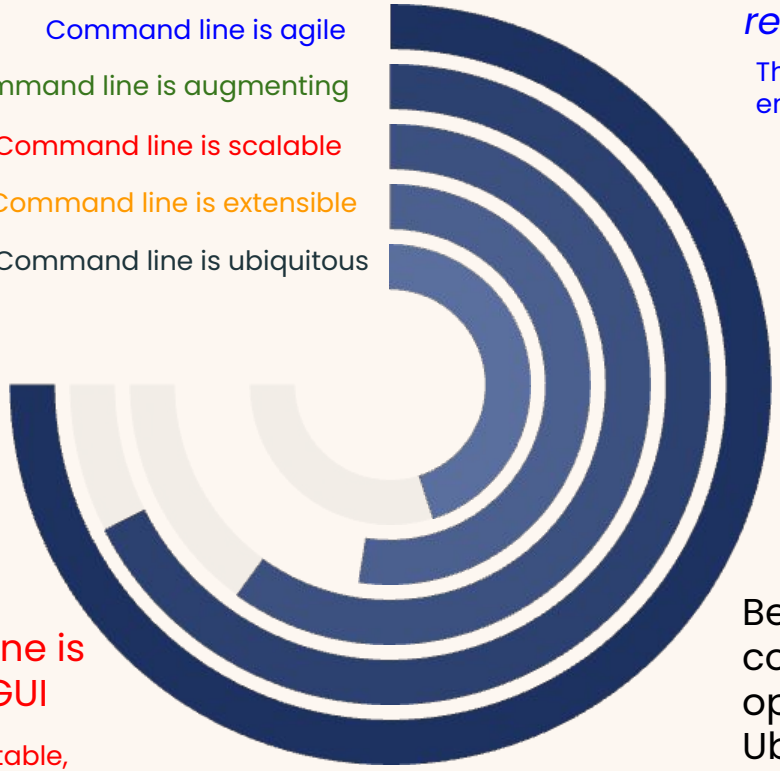
Command line is agile

Command line is augmenting

Command line is scalable

Command line is extensible

Command line is ubiquitous



read-eval-print-loop (REPL)

The shell, with its read-eval-print-loop, enables you to play with your data.

The command line integrates well with other technologies.

One way to glue tools is by connecting the output from the first tool to the input of the second tool

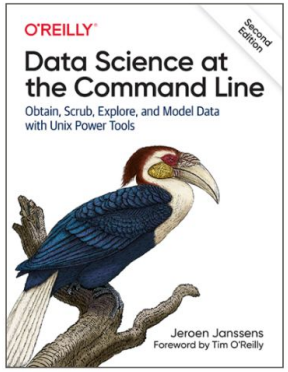
The command line itself is language-agnostic

The open source community is producing many free and high-quality command-line tools that we can use for data science

Working on the command line is very different from using a GUI

Because the command line is automatable, it becomes scalable and repeatable.

Because the command line comes with any Unix-like operating system, including Ubuntu Linux and macOS, it can be found in many places



Search

Data Science at the Command Line

Obtain, Scrub, Explore, and Model Data with Unix Power Tools

Welcome to the website of the second edition of *Data Science at the Command Line* by Jeroen Janssens, published by O'Reilly Media in October 2021. This website is free to use. The contents is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. You can order a physical copy at Amazon.

If you find this book helpful, consider spreading the word! You could, for instance, share it on Twitter, write a review on Amazon, or star the Github repository. Much appreciated!

Description

This thoroughly revised guide demonstrates how the flexibility of the command line can help you become a more efficient and productive data scientist. You'll learn how to combine small yet powerful command-line tools to quickly obtain, scrub, explore, and model your data. To get you started, author Jeroen Janssens provides a Docker image packed with over 100 Unix power tools—useful whether you work with Windows, macOS, or Linux.

You'll quickly discover why the command line is an agile, scalable, and extensible technology. Even if you're comfortable processing data with Python or R, you'll learn how to greatly improve your data science workflow by leveraging the command line's power. This book is ideal for data scientists, analysts, engineers, system administrators, and researchers.

- Obtain data from websites, APIs, databases, and spreadsheets