INVESTIGATING AUDIBLE BEST SELLERS

Stefan Hainzer, August 2017

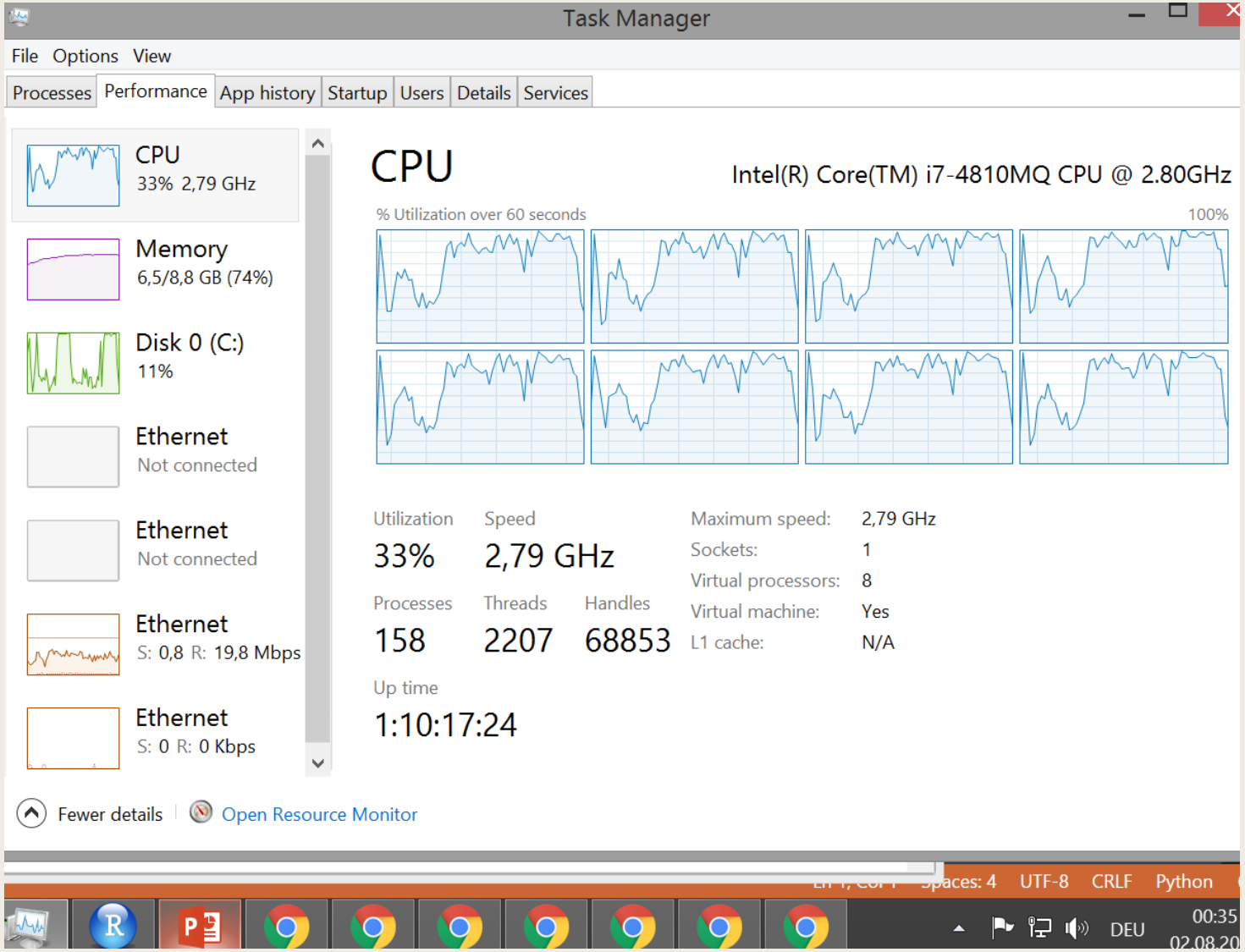
Audible Best Sellers

- Audible.com is an amazon company which is specialized in audio books.
- They provide a best seller list on their homepage.
- The goal was to scrap the best sellers and to analyse them for some trends.

Web Scraping

- Modern websites often use advanced software frameworks to improve the user experience.
- That includes responsive design, client side code (JavaScript), dynamic content loading, asynchrony requests, etc.
- Static requests for scraping are limited due to dynamic site behaviour.
- Selenium provides a solution but has slow performance.
- To overcome this drawback a parallel scraping was developed, using one web browser per processor core.
- The scraper uses the Python package “Parallel Pyhton” http://www.parallelpython.com/
- One Master gets the overview list from www.audible.com/adblbestsellers and delegates the book data collection to the slaves.

Web Scraping



Web Scraping

- To evaluate the performance five serial and parallel scraping runs were performed.
- The best seller list contains 60 books.
- The mean serial duration was: 178 seconds
- The mean parallel duration was: 71 seconds.
- One master and seven slaves have been used.

- The Parallel Python package also support slaves on remote machines.

Data Analysis

- Change to jupyter notebook.