

Book Review - Due by November 24

Choose a book (which we do not cover in class) related to our course topics: the command line, databases, data warehouses, distributed analytical computing; read it, and tell us about it. Suggestions for titles are below.

- Talks will begin in class on September 29 and continue through November 24.
- Choose a title, and announce it along with your preferred talk date on the Book Reviews discussion board.
- Prepare a (maximum) five minute talk giving a brief summary of the book's contents, its strengths and weaknesses, and perhaps a demo of something you learned.
- Prepare a 300-500 word review of the book; turn this and any slides/notebooks you develop for your talk in via Blackboard when you give your talk.

This project gives you a chance to dig deeper into a topic of your choosing. It also offers you an opportunity to think critically about the text and demonstrate what you've learned. Finally, it gives all of us a chance to get a look at a wide variety of titles related to our topics, and to hear recommendations for which to read for further information.

Unix / Linux

Gancarz, The UNIX Philosophy; Shotts, The Linux Command Line

Databases

Celko, SQL for Smarties; Date, Database in Depth; Hernandez, Database Design for Mere Mortals; Lukaszewski, MySQL for Python

Warehouses

Corr / Stagnitto, Agile Data Warehouse Design; Hughes, Agile Data Warehousing Project Management; Inmon, Building the Data Warehouse; Kimball, The Data Warehouse Toolkit

noSQL

Karau et al., Learning Spark: Lightning-Fast Big Data Analysis; Redmond, Seven Databases in Seven Weeks; Ryza et al., Advanced Analytics with Spark: Patterns for Learning from Data at Scale; Sadalage, Fowler, NoSQL Distilled; Sankar, Karau, Fast Data Processing with Spark 2e

Writing

Dupre, BUGS in Writing; Greene, Writing Science in Plain English; Oliver, A Poetry Handbook; Zinsser, On Writing Well