

| |
|---|
| Free Data Engineering Learning Resources from Treasure Data |
| http://get.treasuredata.com/Reading-List.html |
| Book: McKinney, Wes. <i>Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython</i> . O'Reilly, 2013 |
| Book: Milton and Arnold. <i>Introduction to Probability and Statistics</i> . McGraw Hill, 1995 |
| The Data Science Handbook |
| http://www.thedatasciencehandbook.com/#get- |
| Quora: How do I learn big data: |
| https://www.quora.com/How-do-I-learn-big-data-1 |
| Skills of the data architect |
| http://www.capttechconsulting.com/blogs/skills-of-the-data-architect |
| Quora: data architect vs analyst vs engineer vs scientist: |
| https://www.quora.com/Whats-the-difference-between-a-data-architect-data-analyst-data-engineer-and-data-scientist |
| Python for aspiring data nerds |
| https://blog.treasuredata.com/blog/2015/04/24/python-for-aspiring-data-nerds/ |
| Logs are the Lifeblood of your data pipeline: |
| https://www.fluentd.org/ |
| How Luigi works: |
| http://help.mortardata.com/technologies/luigi/how_luigi_works |
| Embulk: |
| https://github.com/embulk/embulk |
| Kafka Guide: |
| http://kafka.apache.org/ |
| What every engineer should know about real time data: |
| https://engineering.linkedin.com/distributed-systems/log-what-every-software-engineer-should-know-about-real-time-datas-unifying |
| Data Lake overview: |
| http://martinfowler.com/bliki/DataLake.html |
| Hadoop overview: |
| https://blog.pivotal.io/pivotal/products/demystifying-hadoop-in-5-pictures |
| NoSQL cheat sheet: |
| http://www.dummies.com/how-to/content/nosql-for-dummies-cheat-sheet.html |
| NoSQL guide: |
| http://nosqlguide.com/ |
| SQL vs NoSQL overview: |
| https://www.digitalocean.com/community/tutorials/understanding-sql-and-nosql-databases-and-different-database-models |
| Public cloud data-store comparison: |
| http://gaul.org/object-store-comparison/ |
| Amazon Redshift guide: |
| https://www.periscopedata.com/amazon-redshift-guide |
| Excel to SQL with lifetime customer value: |
| https://blog.treasuredata.com/blog/2014/12/05/learn-sql-by-calculating-customer-lifetime-value-part-1/ |
| Quora: what's the benefit of SQL: |
| https://www.quora.com/What-is-the-benefit-of-learning-SQL/answer/Kiyoto-Tamura |
| Intro to Pandas: |
| http://www.gregreda.com/2013/10/26/intro-to-pandas-data-structures/ |
| Using SQL and R to do funnel analysis |
| https://blog.treasuredata.com/blog/2016/01/25/using-sql-and-r-to-do-funnel-analysis-part-1/ |
| Processing data with Jupyter and Pandas: |
| https://blog.treasuredata.com/blog/2015/06/23/data-science-101-interactive- |
| Data processing and data pipeline overview: |
| http://flowingdata.com/ |
| Python data visualization libraries: |
| https://www.dataquest.io/blog/python-data-visualization-libraries/ |
| Hadoop vs Elasticsearch: |
| https://blog.treasuredata.com/blog/2015/08/31/hadoop-vs-elasticsearch-for-advanced-analytics/ |
| Python 101 |
| http://learnpythonthehardway.org/book/ |
| Python and Spark intro |
| http://www.kdnuggets.com/2015/11/introduction-spark-python.html |
| Using Java with Hadoop |
| http://www.coreservlets.com/hadoop-tutorial/ |