If you’re a budding data scientist, a manager who wants to learn, or someone who wants to get started with data science or engineering, there is a chance you might become overwhelmed with all the things you’ll need to learn! In fact, due to the scarcity of actual learning materials, one really needs to cobble together a do-it-yourself curriculum in data science.

Here is a curated list of data science and data engineering links and publications to help get you started. This list is by no means comprehensive; it's just a starting point.

Books, online and downloadable guides:

**Background:**


The Data Analytics Handbook. [https://www.teamleada.com/handbook](https://www.teamleada.com/handbook)

**Programming:**

Treasure Data’s HIVE guide:

Introduction to Spark with Python:

Learn Python the Hard Way:
Infrastructure:

Treasure Data's Amazon Redshift COPY command guide:

Links:

24 Data Science Resources to Keep Your Finger on the Pulse:

Treasure Data's Introduction to Data Science:

Treasure Data's Introduction to Data Ingestion from (Python) Apps:
http://blog.treasuredata.com/blog/2015/04/24/python-for-aspiring-data-nerds/

Treasure Data's Blog: Learn SQL by Calculating Customer Lifetime Value:
http://blog.treasuredata.com/blog/2014/12/05/learn-sql-by-calculating-customer-lifetime-value-part-1/

Algorithmia's Blog on Consuming Algorithms via REST APIs in Python:
http://blog.algorithmia.com/post/115255527924/accessing-algorithmia-through-python

Plus...

Analytics
- Mixpanel Blog
- Amplitude Blog
- KISSMetrics Blog
- RJMetrics Blog
- Mode Analytics Blog
- Periscope Blog
- Google Analytics Blog
- GameAnalytics Blog
Treasure Data's Data Science Reading List

Data Management
  • Curt Monash's blog
  • O'Reilly Radar Data
  • Ad Age Data (adtech)

Databases
  • Aphyr's Call Me Maybe

Infrastructure as a Service
  • AWS Blog
  • AWS Big Data Blog

Platform as a Service
  • Heroku Blog

Software as a Service
  • SaaStr (more to understand SaaS business models)
  • David Skok's blog (again, more to understand how SaaS business work)
  • Marketo Blog (trying to integrate with Marketo)
  • Salesforce Developer Blog (we integrate with them)

Workflow Management
  • Airflow
  • Luigi

Data Quality Management
  • Trifacta Blog
  • Paxata Blog

Data Collection/Ingestion/Stream Processing
  • Fluentd
  • Embulk
  • Elastic (maker of Elasticsearch, Logstash and Kibana)
  • Confluent Blog (maker of Apache Kafka)
  • Databricks Blog (maker of Apache Spark)

Machine Learning
  • ŷhat blog
  • Domino Data Lab
  • Data Tau (aggregator)