

KDnuggets

[Subscribe to KDnuggets News](#)

[Contact](#)
 search KDnuggets



- [SOFTWARE](#)
- [News/Blog](#)
- [Top stories](#)
- [Opinions](#)
- [Tutorials](#)
- [JOBS](#)
- [Companies](#)
- [Courses](#)
- [Datasets](#)
- [EDUCATION](#)
- [Certificates](#)
- [Meetings](#)
- [Webinars](#)



Live Online Training GraphDB for DevOps

[Live Online Training GraphDB for DevOps - Register Now](#)

[KDnuggets Home](#) » [News](#) » [2016](#) » [Jun](#) » [Tutorials, Overviews](#) » 7 Steps to Mastering SQL for Data Science ([16:n22](#))

7 Steps to Mastering SQL for Data Science

[◀ Previous post](#)

[Next post ▶](#)

http likes 430



Like 6



Share 6



Share 264

Tweet



Share

85

Tags: [7 Steps](#), [Data Science](#), [Database](#), [Relational Databases](#), [SQL](#)

Follow these 7 steps to go from SQL data science newbie to seasoned practitioner quickly. No nonsense, just the necessities.



[Distinguish Your Career With Data](#)
[NYU MS in Business Analytics](#)

By [Matthew Mayo](#), KDnuggets.

[Structured Query Language](#) (forever and always referred to henceforth as SQL, and usually pronounced *sequel*) is **the** language for managing and querying data in [relational database management systems](#) (RDBMS).

So intertwined are the terms SQL and RDBMS that they are often conflated, sometimes by the uninitiated, but often simply out of convenience, and the term SQL is used adversarially to distinguish relational systems from non-relational database systems, which are often categorized by the umbrella term "NoSQL." Yet, SQL skills aren't wasted on non-RDBMS systems; the top data processing frameworks all have some implementation of SQL that sits atop their architecture, including Apache Big Data behemoths [Spark](#) and [Hadoop](#).



Structured Query Language is usually pronounced *sequel*.

Together with R and Python, SQL forms the triumvirate of data science programming languages (liberal usage of the term *programming languages*); in fact, the most recent KDNuggets data science and analytics [software poll results](#) indicate that these are also the 3 most-used tools by respondents, when **all** software is considered.

Clearly, SQL is important in data science. As such, this post aims to take a reader from SQL newbie to competent practitioner in a short time, using freely-available online resources. Lots of such resources exist on the internet, but mapping out a path from start to finish, using items which complement each other, is not always as straightforward as it may seem. Hopefully this post can be of assistance in this manner.

Step 1: Relational Database Basics

First off, since SQL is used to manage and query data in relational databases, having an understanding of what, exactly, a relational database is would be a good start.

Jennifer Widom is a renowned database researcher and professor at Stanford. She also put together one of the very first MOOCs on Coursera. Videos from this course are a good place to start:

- [Relational Databases - Relational Model](#)
- [Relational Databases - Querying](#)

A good introductory read on the relational database model is presented by UMass Boston's Dan Simovici here:

- [Database Management Notes Chapter 1 - Introduction \(pages 1-7\)](#)

Extra credit: For a better theoretical understanding of the foundations upon which relational database systems are built, have a look at Widom's videos on relational algebra:

- [Relational Algebra - Select, Project, Join](#)
- [Relational Algebra - Set Operators, Renaming, Notation](#)

Step 2: SQL Overview

Now that we're all familiar with the relational paradigm, we shift focus to SQL.

First, get a high-level overview, again from Widom:

- [SQL- Introduction to SQL](#)

Read the first few pages of Simovici's SQL notes, or until it stops making intuitive sense:

- [Database Management Notes Chapter 5 - SQL](#)

For a great extended introductory overview of basic SQL commands, watch the following video from Khan Academy:

- [Basic SQL Commands](#)

Once finished with the video, the following is a good treatment of SQL commands, with explanations, and can also serve as a useful reference moving

SHARES

- [SQL Examples](#)

Finally, in preparation of the of next step, get an SQL environment up and running. You may not want to enter every SQL statement you encounter, but having an SQL interpreter up and running just makes sense. I suggest installing SQLite locally; it is a simple, but capable, SQL installation:

- [Get SQLite](#)
- [SQLite Introductory Tutorial](#)



Constructing SQL statements.

Step 3: Selecting, Inserting, Updating

While SQL can be used to perform a number of varied data management tasks, querying databases with SELECT, inserting records with INSERT, and updating existing records with UPDATE are some of the most heavily-used commands, and are good places to get started in practical SQL. Read and go through the following preliminary exercises:

Select

- [Selecting Data \(from SQL Course\)](#)
- [SELECT Basics \(from SQL Zoo\)](#)
- [SQL SELECT \(from Mode Analytics\)](#)

Insert

- [Inserting Data \(from SQL Course\)](#)
- [INSERT Values \(from SQL Zoo\)](#)
- [INSERT Select \(from SQL Zoo\)](#)

Update

- [Updating Records \(from SQL Course\)](#)
- [UPDATE \(from SQL Zoo\)](#)

This tutorial covers SQL basics and beginner to intermediate queries, and is a solid review of what we have seen to this point, and then some. It is presented from a [MySQL](#) point of view, but nothing in here is exclusive to that platform's SQL implementation.

- [MySQL Tutorial 1: Overview, Tables, Queries \(from arachnoid.com\)](#)

Pages: 1 [2](#)

[◀ Previous post](#)

[Next post ▶](#)

Top Stories Past 30 Days

Most Popular

1. [Top 10 Machine Learning Algorithms for Beginners](#)
2. [Top 10 TED Talks for Data Scientists and Machine Learning Engineers](#)
3. [Quantum Machine Learning: An Overview](#)
4. [Comparing Machine Learning as a Service: Amazon,](#)

Most Shared

1. [Top 10 TED Talks for Data Scientists and Machine Learning Engineers](#)
2. [Comparing Machine Learning as a Service: Amazon, Microsoft Azure, Google Cloud AI](#)
3. [The Art of Learning Data Science](#)

SHARES

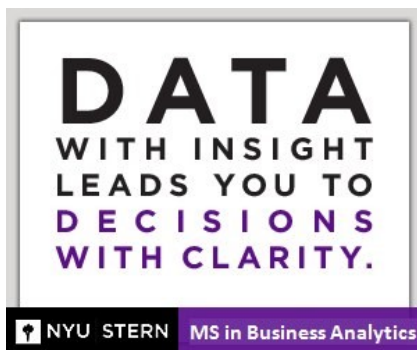
5. [The Art of Learning Data Science](#)
6. [Want to Become a Data Scientist? Try Feynman Technique](#)
7. [Docker for Data Science](#)

[Scientist](#)

5. [A Beginners Guide to Data Engineering – Part I](#)
6. [How To Grow As A Data Scientist](#)
7. [Web Scraping Tutorial with Python: Tips and Tricks](#)

[Latest News](#)

- [Expanding Self-Service BI with AI-Powered Analytics, Fe...](#)
- [Which Machine Learning Algorithm be used in year 2118?](#)
- [Introduction to Python Ensembles](#)
- [Top 15 Scala Libraries for Data Science in 2018](#)
- [Pray.com: Sr Data Engineer](#)
- [Join RE•WORK & AI experts in London, Boston and ...](#)



[Data with Insight Leads you to Decisions with Clarity](#)
[NYU MS in Business Analytics](#)







[Distinguish Your Career With Data](#)
[NYU MS in Business Analytics](#)

Top Stories
Last Week

SHARES

Most Popular

1.  [Comparing Machine Learning as a Service: Amazon, Microsoft Azure, Google Cloud AI](#)
2.  [Top 10 Machine Learning Algorithms for Beginners](#)
3.  [Want to Become a Data Scientist? Try Feynman Technique](#)
4.  [Data Structures Related to Machine Learning Algorithms](#)



5.  [The 8 Neural Network Architectures Machine Learning Researchers Need to Learn](#)
6.  [How To Grow As A Data Scientist](#)
7.  [Top 10 TED Talks for Data Scientists and Machine Learning Engineers](#)

Most Shared

1. [Web Scraping Tutorial with Python: Tips and Tricks](#)



2. [Data Structures Related to Machine Learning Algorithms](#)
3. [The 8 Neural Network Architectures Machine Learning Researchers Need to Learn](#)
4. [Avoid Overfitting with Regularization](#)
5. [Understanding Learning Rates and How It Improves Performance in Deep Learning](#)
6. [My Journey into Deep Learning](#)
7. [Automated Text Classification Using Machine Learning](#)

More Recent Stories

- [Join RE•WORK & AI experts in London, Boston and Hong Ko...](#)
- [Why Data Scientists Must Know About Change Management](#)
- [5 Machine Learning Projects You Should Not Overlook](#)
- [Fast.ai Lesson 1 on Google Colab \(Free GPU\)](#)
- [Top tweets, Jan 31 – Feb 6: #DeepLearning for Natural...](#)
- [AI & Machine Learning: the key skills every software engi...](#)
- [Building a Daily Bitcoin Price Tracker with Coindesk and Shin...](#)
- [Deep Feature Synthesis: How Automated Feature Engineering Works](#)
- [KDNuggets 18:n06, Feb 7: 5 Fantastic Practical Machine Lear...](#)
- [NYU Stern MS in Business Analytics](#)
- [Register for DataScience: Elevate Livestream, Feb 22](#)
- [Top January Stories: Docker for Data Science; Top 10 TED Talks...](#)
- [2018 Predictions for the Analytics & Data Science Hiring ...](#)
- [5 Fantastic Practical Machine Learning Resources](#)
- [The Doing Part of Learning Data Science](#)
- [The Voleon Group: Machine learning research, software developm...](#)
- [Future Trends in Biometrics](#)
- [Challenge Yourself to Think, Mar 19-22, Las Vegas](#)
- [Top Stories, Jan 29 – Feb 4: Web Scraping Tutorial with ...](#)
- [A Simple Starter Guide to Build a Neural Network](#)

[KDNuggets Home](#) » [News](#) » [2016](#) » [Jun](#) » [Tutorials, Overviews](#) » 7 Steps to Mastering SQL for Data Science ([16:n22](#))

© 2018 KDNuggets. [About KDNuggets](#)

Subscribe: Email Name



X

SHARES