In this talk, I'll describe the SimSQL system, which is a platform for writing and executing data- and compute-intensive codes over large data sets, with a particular emphasis on very large-scale statistical computations. At its heart, SimSQL is really a relational database system, and like other relational systems, SimSQL is designed to support data independence. That is, a single declarative code for a particular statistical inference problem can be used regardless of data set size, compute hardware, and physical data storage and distribution across machines. One concern is that a platform supporting data independence will not perform well. But we've done extensive experimentation, and have found that SimSQL performs as well as other competitive platforms that support writing and executing machine learning codes for large data sets.