The rise of automatic feature-generation techniques, including deep learning, has the potential to greatly enlarge the pool of machine-learning users. Such methods require large labeled training sets to obtain high-quality results. This raises two related questions: First, how does one scale deep-learning systems? And second, how can one make it easier for users to build training sets? We describe some very recent work on these questions. Our contribution for the first question is a recent result that characterizes asynchronous learning as equivalent to changing the momentum term. Importantly, this result does not depend on convexity and, so, applies to deep learning. For the second question, we describe a new paradigm called Data Programming that enables users to programmatically cheaply generate large but noisy training sets. Nonconvex analysis techniques then allow us to model and denoise these noisy training data sets. We also report on how nonexperts are able to obtain high-quality end-to-end performance using our prototype information extraction framework, DDlite, that implements these ideas.