This thesis seeks to propose and evaluate a scheduler that can leverage full distributions (e.g., the histogram of observed runtimes or resource usage) rather than single point estimates. Knowing point estimates, such as how long each job will execute, enables a scheduler to more effectively pack jobs with diverse time concerns (e.g., deadline vs. the-sooner-the-better) and placement preferences on heterogeneous cluster resources. But, existing schedulers use single-point estimates (e.g., mean or median of a relevant subset of historical runtimes), and we show that they are fragile in the face of real-world estimate error profiles. In particular, analysis of job traces from three different large-scale cluster environments shows that, while the runtimes of many jobs can be predicted well, even state-of-the-art predictors have wide error profiles with 8-23% of predictions off by a factor of two or more. Instead of reducing relevant history to a single point, a distribution provides much more information (e.g., variance, possible multi-modal behaviors, etc.) and allows the scheduler to make more robust decisions. By considering the range of possible runtimes and resource usage for a job, and their likelihoods, the scheduler can explicitly consider various potential outcomes from each possible scheduling option and select an option based on optimizing the expected outcome.