Introduction to Linux and Cluster Computing Environments for Bioinformatics

Abstract:

Many of the software tools created for bioinformatics run optimally on a Linux or Linux cluster environment. Linux and clusters permit the simultaneous execution of many programs running on many CPUs spread across many computers. This means an analysis that may have taken weeks to complete with a single desktop computer could finish in hours on a Linux cluster. This ability to scale up with many simultaneous programs running at once on many computers is especially valuable when dealing with big data, where analysis on a single computer is impractical. This multi-part talk provides an introduction to Linux computing environments, to Linux cluster environments, and to methods for parallel execution of bioinformatics tools. These Linux computational environments provide the basis for the bioinformatics tools introduced in future talks.