Overview

Cloud Computing is “A large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet.” It has become a driving force for information technology over the past several years, and it is hinting at a future in which we won’t compute on local computers, but on centralized facilities operated by third-party compute and storage utilities. Governments, research institutes, and industry leaders are rushing to adopt Cloud Computing to solve their ever-increasing computing and storage problems arising in the Internet Age. There are three main factors contributing to the surge and interests in Cloud Computing: 1) rapid decrease in hardware cost and increase in computing power and storage capacity, and the advent of multi-core architecture and modern supercomputers consisting of hundreds of thousands of cores; 2) the exponentially growing data size in scientific instrumentation/simulation and Internet publishing and archiving; and 3) the wide-spread adoption of Services Computing and Web 2.0 applications. This course is a tour through various topics and technologies related to Cloud Computing. We will explore solutions and learn design principles for building large network-based systems to support both compute and data intensive computing across geographically distributed infrastructure. Topics include resource management, programming models, application models, system characterizations, and implementations. Our discussions will often be grounded in the context of deployed Cloud Computing systems, such as Amazon EC2 and S3, Microsoft Azure, Google AppEngine, Eucalyptus, Nimbus, OpenStack, Google’s MapReduce, Yahoo’s Hadoop, Microsoft’s Dryad, Sphere/Sector, and many other systems. The course involves lectures, outside invited speakers, discussions of research papers, programming assignments, and a major project (including both a written report and an oral presentation).

We will be using the textbook Distributed and Cloud Computing: Clusters, Grids, Clouds, and the Future Internet by Kai Hwang, Jack Dongarra & Geoffrey C. Fox.