Abstract

For many organizations dealing with genealogical data, the increasing amount of data available can quickly outstrip capacity to adequately process that data.

On the other side of the coin, a huge number of individuals interested in the results of that processing have computers that sit completely idle for a vast majority of each day.

We propose to bring these two groups together for mutual benefit.

Sharing Resources

HTCondor is a system designed with the goal to “develop, implement, deploy, and evaluate mechanisms and policies that support High Throughput Computing (HTC) on large collections of distributively owned computing resources.”

It can be installed on Window, Linux, and OS X machines, and is currently running on well over 488,272 machines in 2901 “pools” worldwide.

An HTCondor pool named “AllCyclesCommon” has been created specifically for genealogy and will be managed by GeneSys Foundation. Contact info@allcyclescommon.org to connect to the “central manager”.

Consuming Shared Resources

The needed processing should be broken into distributable tasks. Using cooperative computing tools developed at the University of Notre Dame, a “Work Queue” can be started to manage the completion of a large number of these tasks. Computing resources can be directed to the Work Queue opportunistically, and the computed results will be returned to the consumer upon completion.

The consumer can develop the software on a laptop with a subset of the data, and then run that same software on the shared resources with the full dataset.

For more information contact: info@AllCyclesCommon.org