AllCyclesCommon.org – Opportunistic Computing Resources for Genealogy

Peter Ivie, Evan Ivie

GENESYS FOUNDATION

Peter Ivie, Douglas Thain
Department of Computer Science and Engineering

UNIVERSITY OF NOTRE DAME

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.

Individuals willing to allow opportunistic computing to be performed on their computers need assurance that they are fully protected from potentially malicious consumers. They also need to know that they have control over the full capabilities of their computer whenever they need them.

Resource consumers may or may not be concerned that the results they get are free from malicious resource donors. If precise results are required, it is not unreasonable to place the burden of verification on the resource consumer.
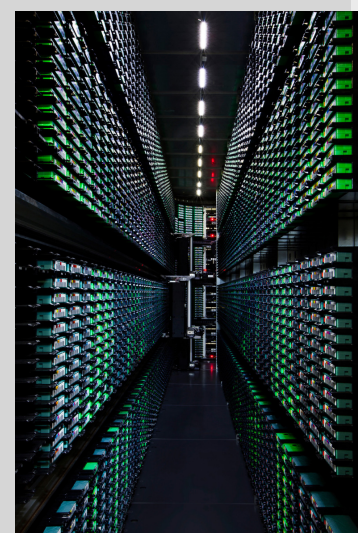
Sharing Resources

HTCondor
High Throughput Computing

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".

The owner of a machine has control over when resources are shared, but often resources are set to become available after 15 minutes with no keyboard or mouse activity, and sharing ceases when such owner activity resumes. The consumer must handle restarting of tasks that were cancelled when an owner resumed control of a machine's resources.

Consuming Shared Resources

Work Queue
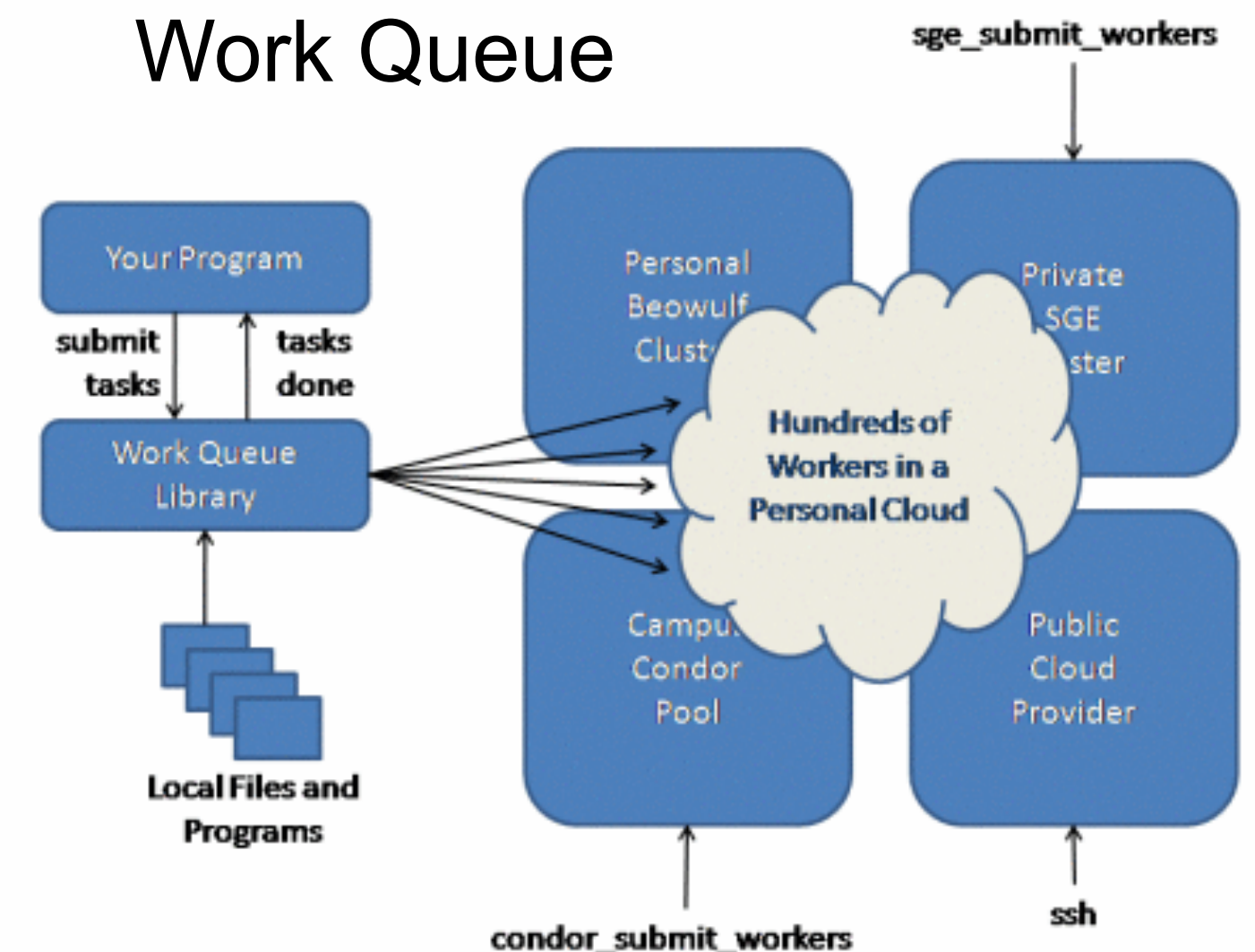
The needed processing should be broken into distributable tasks. Using cooperative computing tools developed at the University of Notre Dame, a "Work Queue" can be started to manage the completion of a large number of these tasks. Computing resources can be directed to the Work Queue opportunistically, 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