

Hydra Cluster at Indiana University

The Hydra cluster consists of a Condor pool of computers, running MS Windows spread across the Bloomington campus of Indiana University, that currently lets Bioinformatics researchers to run specific computational Biology software in parallel. The availability of computing cycles is on an opportunistic basis -- Hydra users can run jobs if a machine is untouched (by students in the lab where the machine is located at) for a specified period of time. The cluster is accessible through a web portal that uses Kerberos authentication.

The primary components of the Hydra cluster include:

- * A Condor server - To take care of scheduling, match-making, so forth.
- * Simple Message Brokering Library (SMBL) - A library used to enable parallel computing on sporadically available desktop systems by introducing a server to keep track of the processing nodes and route messages between them.
- * Process and Port Manager (PPM) -- A program that manages resources (processes, ports, so forth) consumed by multiple parallel sessions on the server machine.
- * A PHP based web-portal used for authentication and job submission

Future Direction:

Among others, one of the current objectives, that the Hydra project team is working on, is to bring the cluster onto the Teragrid. Efforts are being made to use Globus certificates for authentication apart from Kerberos. The project team is also looking into the possibility of deploying virtual machines/networks in an effort to increase the range of applications that Hydra users can use (given that the worker machines are MS Windows based). A JSR-168 compliant portal is another thing that the project team is working on.

Answers to Questions given in "Suggested Topics for Discussion in the Four Key Areas":

USERS: As mentioned above, currently the user base is restricted to Bioinformatics researchers though that is expected to change in the near future. Researchers in Bioinformatics include experts in the field of Biology and/or Computer Science. Since Hydra is accessible through a web portal -- home grown, requires authentication -- the learning curve is not steep.

APPLICATIONS: Currently the Hydra cluster lets users to run Blast, Meme and FastDNAmI in parallel on multiple CPUs at the same time. The project team is looking into various possibilities including deployment of virtual machines/networks to increase the range of applications that can run on the Hydra's grid. Jobs are scheduled by a Condor server on an opportunistic basis -- on a computer that is not being used by a student at a given time.

INFRASTRUCTURE: As mentioned earlier, the Hydra cluster uses Condor, SMBL and PPM along with the software that the end-user runs (currently Blast, Meme and FastDNAmI. Account management is done on a per-user basis at this point of time. Users who have accounts authenticate to a Kerberos server. Usage tracking is done via Condor

logs. The network connectivity ranges from 10 Mbps to 100 Mbps speeds while the server has Gigabit network connectivity.

MANAGEMENT: Currently, the Condor server is managed by a group of experts from the Unix Systems Support Group (USSG) at Indiana University. The Windows based worker machines are managed by the Student Technology Cluster group at Indiana University. The restriction on software that runs on the Hydra cluster arises mostly out of security concerns – running arbitrary code on Windows machines -- and Windows compatibility issues. Once again, the project team is trying to address this issue (Virtual machines, so forth).