Documentation for paralleled *DHSMAP* software

May 9, 2002

1 Parallel version of DHSMAP

The parallel version of *DHSMAP* is designed for a single Unix/Linux machine with multiple processors.
To run the program, type following in command console

dhsmap_p datafile pedfile [parallel_degree]

The parameters *datafile* and *pedfile* are the configuration file and pedigree file used by *DHSMAP* program. The parallel_degree parameter (optional) decides the number of maximum simultaneous *DHSMAP* processes run in the system.
The default parallel degree is set to 2. To achieve best performance (least running time),this parameter should be set to the number of all available CPUs in the system.

2 PVM version of DHSMAP

The PVM version of *DHSMAP* is intended for the computer clusters (as opposed to a single machine with multiple processors) running PVM, the Parallel Virtual Machine software from Oak Ridge National Labs.
Before running PVM version of dhsmap, PVM environment must be installed and configured properly.

2.1 Installation of PVM version of DHSMAP

There are 2 pieces of binary executions in the PVM version of *DHSMAP* package: *dhsmap_pvm* and *dhsmap_slave*. Install *dhsmap_pvm* in master host of the cluster and *dhsmap_slave* in each execution (slave) host.

2.2 Set up PVM environment

The environment variable $PVM_ROOT is required to be defined in each node of the cluster. If it is not defined system widely by administrator, user should define this variable in his/her own shell start-up script.
For example, if PVM package is installed in /usr/local/pvm3. The csh/tcsh users could define $PVM_ROOT by

setenv PVM_ROOT /usr/local/pvm3

For bash users, the command is

export PVM_ROOT=/usr/local/pvm3

PVM needs permission to start processes on remote machines, it relies on the existence of .rhsots in user's home directory. The format of this file is

node1 username
node2 username
...

Each line should be the name of a host which user connects from, followed by a space, followed by the username

2.3 Prepare PVM host file

PVM host file is used for giving PVM information of participating machines in a DHSMAP computing. A sample PVM host file looks like:

```
#filename: pvm_hostfile
*ep /usr/local/bin
node1
node2
*ep /usr/local/dhsmap/
node3
```

node1, node2, node3 are all 3 machines participate the parallel computing. the **ep* option tells PVM where to find the slave program dhsmap_slave on each machine. In this example, dhsmap_slave is installed in the directory /usr/local/bin on machine node1 and node2, and in the directory /usr/local/dhsmap/ on machine node3.

2.4 Start PVM

User could start PVM by typing

pvm pvm_hostfile

A pvm console is brought up signifying that PVM is now running on the host.

To see current PVM configuration, type

pvm>*conf*

The output should be similar as the following:

```
pvm>conf
conf
3 hosts, 1 data format
   HOST     DTID     ARCH   SPEED    DSIG
  node1     3000   SUNSOL2    1000   0x68e59
  node2     6000   SUNSOL2    1000   0x68e59
  node3     9000   SUNSOL2    1000   0x68e59
```

To leave PVM console, type

pvm>*quit*

This command leaves PVM in a running state.
To stop PVM, type

pvm>*quit*

2.5 Run DHSMAP

Once PVM is up, type the following in the command console to start PVM version of *DHSMAP* program:

dhsmap_pvm datafile pedfile [parallel_degree]

Same as the parallel version, the parameters *datafile* and *pedfile* refer to the configuration file and pedigree file. The *parallel_degree* is optional, but it is highly recommended to specify this parameter for better performace.