第32卷，第2期 气 象 Vol. 32 No. 2
2006年2月 METEOROLOGICAL MONTHLY February, 2006

如何在LINUX下并行运行数值模式

谷湘潜[1] 谷美繁[2]

（1. 中国气象科学研究院，北京 100081；2. 海军潜艇学院）

提 要：针对运行数值模式的目的，着重介绍Linux系统的安装、编译和并行环境的建立，以及相关软件的编译安装。所用软件都可以通过Internet网获得，十分适合基层台站建立并行计算机系统，解决计算机资源和实际业务发展的矛盾，为精细化预报提供基础，同时也适合科学研究。

关键词：微机 并行计算 Linux 数值模式

How to Run a Parallel Numerical Model on Linux

Gu Xiangqian[1] Gu Meifan[2]

(1. Chinese Academy of Meteorological Sciences, Beijing 100081; 2. Naval Submarine College)

Abstract: In order to run a numerical models, the Linux-OS installation, the compiling and parallel environment setup process, and other correlative software are emphatically introduced with less about their principles. All the software can be got from the internet, so it is very suitable for basic meteorological stations to set up their parallel computer systems to solve the contradiction between computer sources and development of practical operation. The systems are the bases to do the finer forecasting, and can also be applied to sciences research.

Key Words: microcomputer paralleling compute Linux numerical model

资助项目：国家自然科学基金（40175027）和科技部公益专项“突发性强灾害天气预警系统的研制”
收稿日期：2005年1月26日； 修定稿日期：2005年11月7日

引　言

精细化预报是气象部门面临的重要课题，精细化预报可以理解为是高分辨率的天气预报的深入应用，高分辨率的数值模式是进行精细化预报的基础。因此，在现有的条件下，开展数值模式的业务运行，就显得十分重要。科学计算特别是像大气数值模式积分这样的大型计算对计算机的容量和速度的要求越来越高，昂贵的巨型机一般供给业务部门使用，其更新的费用也不是一般的部门所能承受的。

目前微处理器的速度发展很快，如 Pentium 4 微处理器的速度已超过了 500 MFLOPS，虽然由于商业化的考虑，面对低端应用的处理器不支持对称多处理器技术（Symmetric Multi Processing，SMP），但采用成熟且价廉的网络和群集（Cluster）技术可以较好解决计算机资源的问题。随着 Internet 应用的日益普及，免费网络操作系统 Linux[1]越来越受人们的关注。我国气象工作者也十分重视利用这些资源[2~5]，进行了微机运行模式和 Vis5d 的工作，但这些主要采用单机进行，对于多台并行计算没有涉及。

1　硬件组建

从安装和运行角度来说，386 以上 PC 机即可运行 Linux，作为运行大型科学计算的系统来说，这显然不够。根据目前的处理器水平和价格状况，以及并行计算的要求，建议如下硬件配置：

服务器（1 台）：PIV3.2G/AMD 64 ＋ 3200 以上的 CPU；主板选择品牌比较好的厂家，如华硕，技嘉等，芯片要和 CPU 配套，最好内置显卡和千兆网卡，另外还要加入一块网卡用来和单位的局域网相连接；内存为 1G 内存（DDR 400），硬盘越大越好，目前 200G 比较合适。服务器既要承担资料储存，网络服务，还要参与计算，因此配置要比结算节点稍高一些。

计算节点（3 台），参考服务器的配置，可以比服务器稍低：PIV3.0G/AMD 64 ＋ 3000 以上的 CPU；主板同服务器内置显卡和千兆网卡；内存为 512G 内存（DDR 400），硬盘能够安装运行操作系统就可以了，容量 40～80G 足够。为测试并行节点数目与并行效率的关系，采用 10 台 PIV 1.6G 运行国家气象中心的 MM5 运行方案，结果发现采用四台的效果较好。综合考虑台站实际情况，建议采用四台微机的运行方案。

2　操作系统[1]

Linux 的发行版本很多，现在用的比较多的是 RedHat Inc. 的 Fedora，目前的版本为 4.0。但这个版本存在的问题还比较多，这里以 Fedora 2.0 为例说明安装中主要步骤和注意事项，安装之前做好下面的准备工作：

（1）调整 BIOS 的时间为北京时。

（2）关掉其节能睡眠功能和检查病毒功能。

（3）检查显示卡和网卡是否为 Linux 所支持。

（4）检查网络的连接。

（5）熟悉常用的 UNIX 命令。

对服务器来说，主要步骤：

（1）一般使用安装光盘直接启动即可进入安装界面，如失败，可改用软盘启动方式，在 DOS 下运行安装光盘下的 \ util \ rawrite.exe 文件，输入 \ images \ boot.img 文件；

(2) 注意建立三个 Linux 分区，第一分区为 10G，安装结点为/，这个分区一定要在硬盘 1024 磁道内，用来安装操作系统。第二分区的大小设为和内存一样大，服务器为 1G，计算节点为 512M，作为内存交换区，无需节点。其他的空间都作为一个分区，输入一个合适的安装节点名字，如/YWXT，用来保存数据和进行并行计算。

(3) 选择 custom 安装模式。

(4) 安装程序一般会自动识别显示卡和网卡，去掉 Configure using DHCP 选项后，填入本机与计算节点的相连网卡 IP 地址(如 111.111.111.1) 和与局域网相连网卡的 IP 地址 (如 198.28.13.64)，名称为 node1，如局域网有网关的需加入网关地址(gateway IP，如 198.28.13.1)。购买硬件时，一定要选择 Linux 所支持显示卡和网卡的芯片型号。

(5) Firewall Configure (防火墙设置) 设置中选择 no firewall 或者选择 Medium，

再点击：eth0 eth1 telnet WWW (HTTP) FTP (使选项带“√”)。

(6) Linux 发行版本的应用软件很多，全部选用对运行数值模式没有必要，有时软件之间还存在冲突，一个比较合适的选择如下：

Desktops：

X windows System

GNOME Desktop Enviornmnet

Applications

Editors

Graphical Internet

Graphics

Servers

Server Configuratuion Tools

Windows File Server

FTP Server

Networks Servers (rsh-server，telnet-server)

Development

Development tools

X Softwares Development

System

Administration Tools

Systems Tools

(7) 键入根用户 (root) 的密码，增加一个普通用户 pub 便于进行并行计算。

(8) 选择 GRAPHIC 启动方式

(9) 安装完再启动后使用 root 用户登录 Linux，提示符为 [root@node1 root] #

(10) 建立文件/etc/hosts. equiv，内容为

node1

node2

node3

node4

另外一个替代的方式可以在 pub 主目录下建立 . rhosts，内容同上。

对于计算节点，安装基本与服务器节点的安装相同，不同点主要有：

① 应用软件只需要选择 ftp，rsh，telnet 网络服务，其他的都不选。

② 使用 root 登录

③ 增加用户 pub 并设置好密码

adduser pub

passwd pub

3 编译环境和应用软件

在 Linux 的发行版本中一般都有一个自带的 Fortran 编译器 g77，但只能支持 Fortran 77 标准，作一些简单运算还可以，要实现复杂的模式计算，这个编译器就无能为力了，必须采用第三方的软件支持，目前 PGI 和INTEL编译器比较合适，特别是 INTEL 编译器免费使用，没有时间限制。这

两个编译器都为目前的许多模式系统所支持，如CCSM，WRF，MM5等，由于PGI时间比较长，因而比较完善。

3.1 PGI编译器的安装和使用

由美国Portland公司出品，可在LINUX/NT/工作站运行，目前的版本为5.2。PGI编译器包括C/C++，F77/F90多种编译器，要长期使用必须付费，安装后的期限为15天，但2天后编译一个程序就要等待一定时间的提示，编译后的可执行文件也只能运行15天，因此建议用户10天内重新安装一次PGI编译器，并重新编译一次所有的程序，如果需频繁使用PGI，则应当每两天安装一次。在http://www.pgroup.com上登记注册，或直接在ftp://ftp.pgroup.com/x86/下载。文件名为：linux8.tar.gz，文件大小为65M。

PGI编译器的安装步骤如下：

(1) 使用root权限，把linux86.tar.gz解压到一个临时目录中 如/root/pgi

```
cd /root/
mkdir pgi
cd pgi
mount /dev/cdrom /mnt/cdrom
tar xfvz /mnt/cdrom/linux86.tar.gz
```

(2) 键入安装程序开始安装

```
./install
```

(3) 首先出现的是有关协议，用空格键翻屏，accept有关协议

(4) 安装方式，选第5项全部安装

(5) 是否要改变目录/pgi/usr，回答为n

(6) 是否要产生evaluation license，y

(7) 出现的是有关协议，用空格键翻屏，accept有关协议

(8) 键入姓名Email等，可随便输入

(9) 是否要安装文件为只读，n

(10) 环境变量的设置

在/etc/profile或用户主目录下的.bashrc文件中加入

```
export PGI=/usr/pgi
export PATH = $PGI/linux86/bin:$PATH
export MANPATH = $MANPATH:$PGI/man
```

建议用户修改/etc/profile文件，这样每个用户都能方便使用pgi编译器。

编译命令为pgf77和pgf90

3.2 INTEL编译器的安装和使用

INTEL公司开发，可以免费自用，但必须注册获得有关协议许可文件，注册网址https://registrationcenter.intel.com，注册方法非常简单，输入自己的邮件地址，就可以获得协议许可文件，在https://premier.intel.com/WhatsNew.aspx下载最新版，文件名为l_fc_pc_8.1.???.tar.gz，大小约为88M，使用ftp把l_fc_pc_8.1.???.tar.gz传送到linux下。主要步骤如下：

(1) 使用root权限：

```
cd /root/
tar xfvz l_fc_pc_8.1.???.tar.gz
cd l_fc_pc_8.1.???
```

(2) 键入安装程序开始安装

```
./install.sh
```

(3) 首先选择安装编译器选项1

(4) 用空格键翻屏，accept有关协议

(5) 键入安装目录，建议输入

/usr/intel.

(6) 环境变量的设置

同样，在/etc/profile或用户主目录下的.bashrc文件中加入

```
export LD_LIBRARY_PATH =/usr/intel/lib: $LD_LIBRARY_PATH
export PATH= $PATH: /usr/intel/
```

bin

export MANPATH = $MANPATH：/usr/intel/man

编译命令为 ifort.

3.3 MPICH 并行环境的安装和使用[6]

MPI（Message Passing Interface）是目前一种比较著名的应用于并行环境的消息传递标准，MPICH 是 MPI1.2 标准的完全实现，也是应用范围最广泛的一种并行及分布式环境，主要包含 MPI 函数库和程序设计运行环境，网站为 http：//www-unix.mcs.anl.gov/mpi/mpich，使用 ftp 工具把 mpich.tar.gz 传送到 linux 下，安装步骤如下：

(1) 使用 root 权限：

cd /root/

tar xfvz mpich.tar.gz

cd mpich-1.2.6

(2) 初始化编译环境：

./configure －c＋＋＝pgCC －cc＝pgcc －fc＝pgf77 －f90＝pgf90 －－prefix＝/usr/local/mpich

(3) 编译：

make

(4) 安装：

install

(5) 环境变量的设置

同样，在/etc/profile 或用户主目录下的 .bashrc 文件中加入

export MPICH＝/usr/local/mpich

export PATH = $MPICH/bin：$PATH

export MANPATH = $MANPATH：/$MPICH/mpich/man

以上是对 PGI 编译器，对于 INTEL 编译器步骤一样，但第 (2) 步应该为：

./configure －cc＝gcc －fc＝ifort －f90＝ifort －－prefix＝/usr/local/mpich－intel

编译命令为 mpif77 和 mpif90.

其他软件的安装，比如 NCARG，NETCDF 的方法差不多，可以参考上面的内容进行。

4 模式的并行调试

现在的模式发展都注意并行版本的开发，模式单机版本的运行比较简单，采用并行方式运行要复杂一些，这里以 mm5v3 为例，最新版为 V3.72，对于模式的描述可以参考文献 [7]。源程序可以在 ftp：//ftp.ucar.edu/mesouser/MM5V3 获得，与并行计算有关的模块主要有 2 个，MM5.TAR.gzr 和 MPP.TAR.gz，这里说明进行并行运算的主要步骤：

(1) 解包建立相关目录

tar MM5.TAR.gz

cd MM5

tar MPP.TAR.gz

(2) 修改 configure.user

vi configure.user

把相关的编译器的选项前的“＃”去掉，使其产生作用。如 INTEL 编译器和 MPICH 组成的编译参数为：

```
#
#  7g2. Linux PCs. Need INTEL and MPICH.
#
RUNTIME_SYSTEM = "linux"
MPP_TARGET= $ (RUNTIME_SYSTEM)
```

```
## edit the following definition for your system
LINUX_MPIHOME = /usr/local/mpich-intel
MFC = $ (LINUX_MPIHOME) /bin/mpif77
MCC = $ (LINUX_MPIHOME) /bin/mpicc
MLD = $ (LINUX_MPIHOME) /bin/mpif77
FCFLAGS = -O2 -convert big_endian -pc32
LDOPTIONS = -O2 -convert big_endian -pc32
LOCAL_LIBRARIES = -L $ (LINUX_MPIHOME) /build/LINUX/ch_p4/lib -
lfmpich -lmpich
MAKE = make -i -r
AWK = awk
SED = sed
CAT = cat
CUT = cut
EXPAND = /usr/bin/expand
M4 = m4
CPP = /lib/cpp -C -P
CPPFLAGS = -traditional -DMPI -Dlinux
CFLAGS = -DMPI -I/usr/local/mpi/include
ARCH_OBJS = milliclock.o
IWORDSIZE = 4
RWORDSIZE = 4
LWORDSIZE = 4
```

(3) 编译产生执行文件

make mpp

如果成功，在 MM5/Run 目录下产生执行文件 mm5.mpp

如果不成功，可以修改编译参数，必要时也要修改程序中的语句，重新编译：

make uninstall

make mpp

(4) 在计算节点建立同样的目录以及文件副本。

为确保计算顺利，建议在计算节点建立相同的并行环境，当然也可以采用 NFS 的方式来实现。

(5) 控制参与计算的节点

可以使用两个文件控制参与计算的节点：

一个文件为/usr/loca/mpich-intel/share/machines.LINUX

有关内容为：

node1

node2

…

node24

另一个可在 MM5/Run 目录下建立文件 pgfile，内容为：

```
node1 0 /home/pub/MM5/Run/mm5.mpp
node2 1 /home/pub/MM5/Run/mm5.mpp
node3 1 /home/pub/MM5/Run/mm5.mpp
…
```

(4) 运行

make mm5. deck
mm5. deck
cd Run
mpirun －np 4 mm5. mpp
或 mpirun －p4pg pgfile mm5. mpp

WRF 的运行类似 MM5，甚至比 MM5 的步骤更简单些，参考上面的方法，完全可以顺利运行 WRF，WRF 目前的版本为 2. 0. 3. 1，主要有三个模块：WRF-SI. TAR. gz、WRFV2. TAR. gz 以及一个 3DVAR 模块。下载网址为：http：//www. mmm. ucar. edu/wrf/users/download/get _ source. html

5 结 语

目前，微机及其并行可以胜任只有大型机才能完成的部分工作，从而解决精细化预报所需要的计算机软件资源，十分适合我国国情。巨大的 INTERNET 自由软件库也可提供软件基础，但这需要具备一定的计算机知识和技巧，对于气象部门、特别是对省级至县的基层完全可以在财力所及的情况下，建立运行中尺度模式硬件和软件环境，进而开展精细化天气预报的业务和服务。

参考文献

1 Michael Jane 著，邱仲潘等译. 红帽 Linux9 从入门到精通. 北京：电子工业出版社，2003：20.
2 贝刚. 在微机上运行 MM5V3 模式系统. 气象，2001，27 (2)：16～20.
3 贝刚. 用 Vis5D 软件包在 PC 上实现模式预报输出结果的可视化. 气象，2000，26 (11)：14～18.
4 周小珊，杨森，张立祥. 中尺度数值模式 (MM5V3) 在沈阳区域气象中心的试用. 气象，2001，27 (8)：28～32.
5 谷湘潜，谷美繁. 大型数值模式的移植与计算结果. 气象科技，1999，16 (4)：30～33.
6 William Gropp and Ewing Luck，Installation and User's Guide to MPI (a Portable Implementation of MPI Version 1. 2. 6 The ch _ p4 device for workstaion Networks)，Mathmatics and Computer Science Devision，ANL/MCS-TM-ANL-01/X Rev x，2002：pp120.
7 Georg A. Grell，Jimy Dudhia and David R. Stauffer. A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). NCAR/TN－398＋STR NCAR technical note. 1995：pp117. (Grell G. A，D. Jimy and D. R. Stauffer).