

深腾 7000 上机指导



2009年12月21日

曹宗雁

zycao@sccas.cn

中国科学院计算机网络信息中心
超级计算中心 系统运维部

主要内容

- 一、深腾 7000 超级计算机介绍
- 二、基本操作和常用命令
 - 如何登录深腾 7000
 - 数据的上传和下载
 - 基本操作和命令
- 三、在深腾 7000 上提交和运行作业
 - 提交作业
 - 查看作业状态和运行结果
 - 其它作业操作

主要内容

- 一、深腾 7000 超级计算机介绍
- 二、基本操作和常用命令
 - 如何登录深腾 7000
 - 数据的上传和下载
 - 基本操作和命令
- 三、在深腾 7000 上提交和运行作业
 - 提交作业
 - 查看作业状态和运行结果
 - 其它作业操作



各种应用软件



可视化结点



管理与I/O结点

刀片结点



厚结点



以太网(管理)



光纤存储网络



lenovo

深腾7000超级计算机



胖结点



高速Infiniband网络



Linux操作系统



高速磁盘阵列



磁带库

深腾 7000 超级计算机历史

- 2008 年 8 月 7 日，中科院网络中心与联想公司在京签约，委托联想公司研制并部署一套具备每秒百万亿次运算能力的“深腾 7000”高效能计算机，用于装备中科院超级计算环境总中心和国家网格（北方）主结点。
- 2008 年 10 月，深腾 7000 超级计算机组建基本完毕，联想公司及中科院网络中心超算中心等单位开始在联想公司进行深腾 7000 的调试和测试工作。
- 2008 年 11 月 14 日，新一届全球超级计算机 Top500 排行榜公布，深腾 7000 全球排名第 19 位，深腾 7000 也由此成为入选该排行榜的第五台联想“深腾”系列高端超级计算机。
- 2008 年 12 月 4 日联想集团正式对外宣布，联想已经成功研制出每秒实际性能超过百万亿次的高性能计算机“深腾 7000”。
- 2008 年 12 月底，深腾 7000 超级计算机开始在中科院网络中心超算中心机房进行最终的安装部署工作。
- 经过半年左右的安装调试，深腾 7000 超级计算机于 2009 年 4 月 24 日正式对外提供试运行服务。

深腾 7000 的几大特点

- 异构集群系统。深腾 7000 成功实现了 1240 个 2 路薄结点和 140 个 4 路厚结点的协同计算，实际 Linpack 性能突破每秒 106.5 万亿次。另外深腾 7000 还包括了 2 个胖结点和 12 个带有 GPU 加速卡的可视化结点，满足不同应用对计算机体系结构的需求。
- 无盘启动。深腾 7000 带有 1428 个无盘启动结点，是世界上规模最大的一个结点无盘启动的集群系统。
- 高性能大容量存储。深腾 7000 实现了 PB 级别的三级结构海量存储系统，在线、近线、离线存储总容量超过 PB 级。
- 丰富的软件支持。深腾 7000 不仅安装了众多的系统软件，同时还装有多学科多类别的**商业的或开源的应用软件**，以提高系统的可用性，满足各类用户的需求。

深腾 7000 计算资源 - 刀片结点

- 深腾 7000 共配有 1288 台刀片服务器，其中计算刀片结点 1140 个（采用无盘方案），I/O 结点 120 个，网络启动服务器 12 个，管理、登录等服务结点共 16 个。
- 刀片结点的单机配置如下：
 - 配置 2 个 Intel Xeon E5450 四核处理器（3.0GHz，L1 为 32KB+32KB，L2 为 2*6MB，4 条浮点流水线，单核心性能为 12Gflops，单 CPU 性能为 48Gflops，TDP 为 80W）。
 - 内存 32GB。
 - 配备 4x DDR infiniband 网卡（HCA）。
 - 配备千兆以太网卡。
 - I/O 结点、启动结点等配备 4Gbps 光纤卡（HBA）。
- 所有刀片结点总性能为 123.6Tflops，总内存 41.2TB；其中计算刀片的总性能为 109Tflops，总内存 36.5TB。

深腾 7000 计算资源 - 厚结点

- 深腾 7000 系统中的厚结点采用 IBM X3950M2 服务器，共 38 台。
- 厚结点单机配置为：
 - NUMA 结构，结点内采用 IBM 专用网络连接。
 - 配置 16 个 Intel Xeon X7350 四核处理器（2.93GHz，L1 为 32KB+32KB，L2 为 2*4MB，TDP 为 130W），共 64 个 CPU 核心。
 - 内存 512GB。
 - 配备 4x DDR infiniband 网卡（HCA），每结点 4 块。
 - 配备千兆以太网卡。
- 厚结点总性能 28Tflops，总内存 19TB



深腾 7000 计算资源 - 可视化结点

- 深腾 7000 共配有 12 台可视化结点，配有高性能图形显示卡与商业可视化软件，用于用户计算结果的可视化处理。
- 其单机配置如下：
 - 采用 3U 机架式服务器。
 - 配置 2 个 Intel Xeon E5450 四核处理器。
 - 内存 32GB。
 - 配备 4x DDR infiniband 网卡（HCA）。
 - 配备千兆以太网卡。
 - 配备 Nvidia Geforce 9800GT 显卡，512M 显存，支持 CUDA 加速功能。
- 刀片结点、厚结点与可视化结点的可执行代码是完全兼容的。

深腾 7000 计算资源 - 胖结点

- 深腾 7000 的胖结点采用 SGI Altix4700，包括 2 个计算结点、1 个头结点、2 个文件服务结点，总浮点性能 5Tflops，内存容量 5TB。
- 具体配置如下：
 - 采用 ccNUMA 结构，单个计算结点配备 188 颗处理器，376 个 CPU 核心，共享内存约 2.5TB
 - 头结点，由 1 台计算结点通过逻辑分区实现，用于用户登录、编译与调试程序、提交作业。头结点配 8 颗处理器，共 16 个 CPU 核心，64GB 内存
 - 文件服务结点，作为 CXFS 文件服务器使用，用户不可直接访问
 - 共 384 颗 Intel Itanium2 9140M 双核处理器
 - 1.66GHz 主频，666MHz 总线，TDP 104W
 - L1 Cache: 16KB+16KB
 - L2 Cache: 256KB+1MB
 - L3 Cache: 18MB
 - 互连网络。
 - 结点内采用 NUMALink 互连
 - 结点间用千兆以太网互连，用于管理
 - 结点间用 2 条 20Gbps 的 Infiniband 互连



深腾 7000 网络和存储系统

● 网络系统

● 系统网络：

- 20Gb/s 的 Infiniband 网络，用于计算和文件系统数据
- 千兆以太网，用于管理

● 存储网络

- IO 结点与盘阵之间通过 4Gb/s 光纤相连。

● 存储系统

● 磁盘阵列系统，采用光纤/SAS 盘阵。

- 2 台 HDS USPv 高性能磁盘阵列，总的裸容量 354TB。
- 聚合 I/O 带宽 $\geq 50\text{GB/s}$

● 虚拟磁带库系统 Quantum DX100

- 利用 SATA 磁盘阵列模拟磁带库。
- 裸容量 64TB

● 磁带库系统

- 使用 2 套 ADIC i2000 智能带库，总不压缩容量 1PB
- 使用 LTO4 磁带，不压缩容量 800GB/ 盘
- 配有 8 台磁带驱动器，聚合 I/O 吞吐率 $\geq 800\text{MB/s}$ （不压缩方式）

深腾 7000 软件环境（一）

● 操作系统

● 刀片结点、厚结点、可视化结点操作系统：

- RedHat Enterprise Linux Server release 5.1
- 内核版本为 2.6.18-53.e15□

● 胖结点操作系统

- SUSE Linux 10SP2
- 内核版本为 2.6.16.60-0.21.default ；

● 作业管理系统

- LSF HPC 7.0 。

● 共享文件系统

- Quantum StorNext File System (SNFS)
- SUN Lustre
- IBM General Parallel File System (GPFS)
- SGI Clustered XFS (CXFS)

深腾 7000 软件环境（二） - 集群部分安装

名称	版本	安装目录
Intel C/C++ Compiler	11.0.081	/home_soft/soft/x86_64/compiler/intel/11.0
Intel Fortran Compiler	11.0.081	/home_soft/soft/x86_64/compiler/intel/11.0
Intel Idb Debugger	11.0.081	/home_soft/soft/x86_64/compiler/intel/11.0
Intel Vtune	9.1_002	/home_soft/soft/x86_64/tools/vtune
Intel Clustered Math Kernel Library	10.1.1.0 19	/home_soft/soft/x86_64/lib/mkl
Intel Integrated Performance Primitives	6.0.2- 076	/home_soft/soft/x86_64/lib/ipp
Intel Thread Building Blocks	2.1_015	/home_soft/soft/x86_64/lib/tbb
Intel Trace Analyzer & Collector	7.2.011	/home_soft/soft/x86_64/tools/itac
Intel Thread Checker	3.1_012	/home_soft/soft/x86_64/tools/itt
Intel Thread Profiler	3.1_012	/home_soft/soft/x86_64/tools/itt
Intel MPI	3.2.011	/home_soft/soft/x86_64/mpi/impi
OpenMPI	1.3.2	/home_soft/soft/x86_64/mpi/openmpi
MVAPICH2	1.2p1	/home_soft/soft/x86_64/mpi/mvapich2/
TotalView Debugger	8.6.2-2	/home_soft/soft/x86_64/tools/toolworks/totalview.8.6.2-2
PGI Workstation Complete	8.0-5	/home_soft/soft/x86_64/compiler/pgi

深腾 7000 软件环境（三） - 应用软件

● 商业软件

- Matlab
- VASP
- Gaussian
- Wien2k
- TURBOMOLE
- q_chem
- adf
- crystal
- Molpro
- Fluent
- ansys

● 开源软件

- abinit
- dl_poly
- grads
- Life
- mm5
- netcdf
- WRF
- Chem
- gmt
- GRAPES
- LIS
- NETCDF
- vmd



主要内容

- 一、深腾 7000 超级计算机介绍
- 二、基本操作和常用命令
 - 如何登录深腾 7000
 - 数据的上传和下载
 - 系统基本操作和常用命令
- 三、在深腾 7000 上提交和运行作业
 - 提交作业
 - 查看作业状态和运行结果
 - 其它作业操作

登录深腾 7000 - Windows 下登录

- 在 Windows 下使用 PuTTY 进行登录

- 下载 PuTTY 软件包并解压：

- <http://the.earth.li/~sgtatham/putty/latest/x86/putty.zip>

- 启动 PuTTY，屏幕上将弹出右图所示窗口。

- 在“Host Name (or IP address)”栏中填入“shenteng.sccas.cn”。

- 点击右下角的“Open”即可。

- 成功后窗口提示输出用户名，在此输入深腾 7000 的用户名与密码。

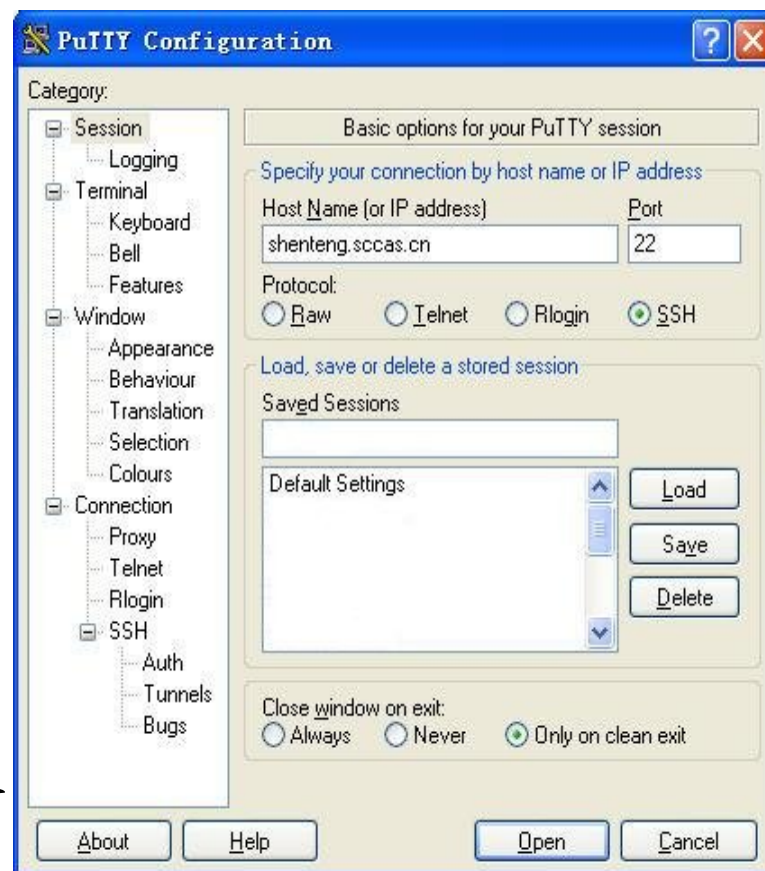
- 用户名：user1100-05

- 密码：chem0900-05

- 密码需要输入两遍。

- 之后窗口显示系统提示信息，并进入到深腾 7000 命令行界面。

- 其它的如 SSH Secure Shell 以及 Secure CRT 等软件，设置和登录方法类似。



登录深腾 7000 - Linux 下登录

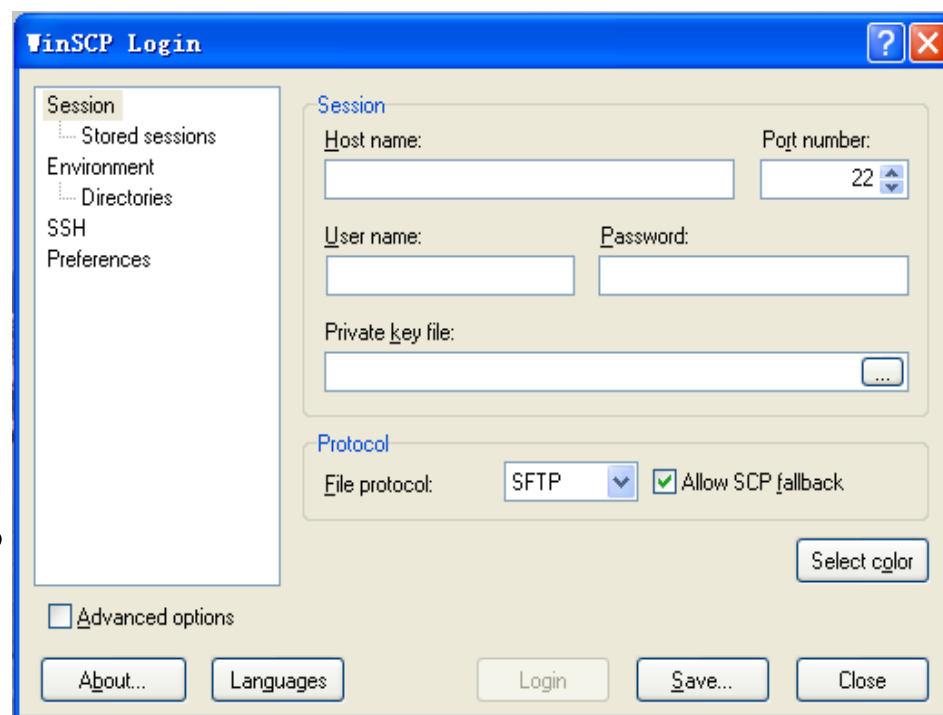
- Linux 系统一般都默认安装了 SSH 客户端软件。登录深腾 7000 的方法如下：
 - 在命令行下直接执行命令 “ssh username@shenteng.sccas.cn”，其中 username 是深腾 7000 上的用户名。
 - 注：用 ssh -l username shenteng.sccas.cn 效果相同
 - 用户名：user1100-05
 - 密码：chem0900-05
 - 如果是第一次使用 SSH 命令连接深腾 7000，有可能看到：

```
The authenticity of host 'xxxxx(shenteng.sccas.cn)' can't be established.  
RSA key fingerprint is xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx.  
Are you sure you want to continue connecting (yes/no)?
```

这里必须输入 yes
 - 输入密码（需要输入两遍），完成 SSH 连接过程，屏幕显示出系统提示信息之后，进入到深腾 7000 命令行界面。

深腾 7000 数据的上传与下载 - Windows 环境

- 推荐使用 WinSCP 客户端软件进行数据传输。其具体操作方法如下。
 - 从网上下载 WinSCP 客户端。
 - <http://winscp.net/download/winscp424.exe>
 - 启动 WinSCP，屏幕上弹出以下窗口。
 - 在“Host Name”栏填入“shenteng.sccas.cn”，在“User name”栏中填入用户名，点击 Login 按钮，就可以按照提示完成 SFTP/SCP 登录，并通过拖曳操作上传和下载数据。
- 用 Filezilla 等支持 SFTP 的客户端工具也可以完成数据的上传与下载。



深腾 7000 数据的上传与下载 - Linux 环境

- 在 Linux 下，可以在本地机器上使用系统自带的 scp 命令进行数据上传和下载操作。
- scp < 参数 > < 源文件或路径 > < 目标文件或路径 >
 - 文件或路径可以是本地，也可以是远程。如果是远程文件或路径，格式为：
 - [用户名 @] < 主机名 | IP 地址 > : 目录或文件名
如：username@shenteng.sccas.cn:mydir/myfile
 - 常用参数
 - -r 递归复制
 - -p 保留源文件属性
 - 当源文件或路径是本地机器，目标文件或路径是远程机器（深腾 7000）时，则完成的是上传操作。
 - 当源文件或路径是远程机器（深腾 7000），目标文件或路径是本地机器时，则完成的是下载操作。

基本目录和文件操作（一）

● 改变当前所在目录

● cd [目录名]

- 不加参数将回到用户主目录
- 相对路径 (./)dirname
 - ./ 当前目录
 - ../ 上一层目录
- 绝对路径 /home/user/dirname

● 创建目录

● mkdir <dirname>

● 删除空目录

● rmdir <dirname>

- rmdir 不能删除非空目录

基本目录和文件操作（二）

- 显示当前所在目录
 - pwd
- 输出文件列表
 - ls [参数] [目录或文件名]
 - 不加目录或文件名表示列表当前目录下的文件
 - 常用参数
 - -l 详细文件详细信息
 - -a 显示所有文件
 - -t/-S (-r) 按时间 / 大小排序 (逆排序)
- 统计目录中所有文件大小总和 (包含子目录)
 - du [参数] [目录名]
 - -s 直接列出最终统计值
 - -h 自动加上数学单位 (G/M/K)
 - --max-depth=n 最多统计到指定目录下第 n 层

基本目录和文件操作（三）

- 复制文件
 - cp [参数] <源文件> <目标文件或路径>
 - -r 递归复制所有子文件夹和其中的文件
 - -a 保留源文件的属性
- 文件或目录移动
 - mv <源文件或目录> <目标文件或路径>
 - 常用来对文件和目录进行重命名
- 删除文件或目录
 - rm [参数] <文件名>
 - -r 递归删除
 - -f 强制删除
- 特别提醒：对于覆盖和删除操作，默认情况下系统不会进行提示并确认。
 - 设置命令转义防误操作（加入 .bashrc 当中）：
 - alias rm=" rm -i"
 - alias mv=" mv -i"
 - alias cp=" cp -i"

基本目录和文件操作（四）

- 输出文件全部内容
 - `cat <文件名>`
- 输出文件开头部分内容
 - `head [-n 行数] <文件名>`
- 输出文件末尾部分内容
 - `tail [-n 行数] <文件名>`
- 分屏查看文件内容
 - `more <文件名>`
 - `less <文件名>`
 - 执行其它命令的输出信息较多时，通常使用 `command | more` 或是 `command | less` 来分屏查看输出结果。

基本目录和文件操作（五）

● 改变文件属性

- chmod <参数> <文件或目录名>
 - (+/-)rwx (赋予/去除)读/写/可执行属性
 - -R 递归更改文件或目录属性

● 搜索符合条件的文件

- find <路径> [参数]
 - 常用 find <路径> -name “<文件名>” 查找文件

● 从文件中搜索匹配字符串

- grep [参数] “<字符串>” <文件或目录名>
 - 常用参数：
 - -i 忽略大小写
 - -r 子目录递归搜索
 - -v 反匹配模式
 - -n 显示匹配行的行号

文件打包和解包

● tar 打包工具

● 解包：

● .tar:

- tar xvf <压缩包>

● .tar.gz/.tgz: gzip 压缩包

- tar zxvf <压缩包>

● .tar.Z: compress 压缩包

- tar Zxvf <压缩包>

● .tar.bz2: bzip2 压缩包

- tar jxvf <压缩包>

亦可使用相应解压工具先解压成 tar 再用 xvf 命令解压

● 打包：

- tar cvf/zcvf/Zcvf/jcvf <目标文件> <源文件或路径>

配置环境变量

- 定义环境变量

- export < 变量名 = 值 >

- 如：export GAUSS_SCRDIR=/home_soft/home/username/mydir

- 查看环境变量

- export
- env

- 取消环境变量

- unset < 变量名 >

- 对于使用 Bash 环境的用户，主要通过以下两个文件设置自己的环境变量（**将对应语句写在文件里即可**）：

- .bash_profile
- .bashrc

vim 文本编辑器

- 打开 vim 编辑器
 - vim [文件名 1,2...n]
- vim 的模式
 - 一般模式
 - 编辑模式
 - 命令行模式
 - 可视模式
- vim 的环境变量设置 (~/.vimrc)
 - set nocompatible 关闭 vi 兼容模式，使光标键有效
 - set ruler 在编辑过程中一直显示当前位置
 - set smartindent 开启编辑时的自动缩进
 - set number 显示行号
 - syntax on 自动语法高亮

vim 常用命令 (一)

● 移动命令

- h/Left j/Down
k/Up l/Right 光标上下左右移动
- Ctrl+f/PageUp 向前翻页
- Ctrl+b/PageDown 向后翻页
- 0/^/Home 移至行首
- \$/End 移至行末
- n+Space 向右移动 n 个字符
- H 移至屏幕最上一行
- M 移至屏幕正中一行
- L 移至屏幕最下一行
- gg 移至文件第一行
- G 移至文件的最后一行
- nG 移至文件的第 n 行
- n+Enter 光标向下移动 n 行

● 注：命令通常在一般模式下输入才能生效

vim 常用命令 (二)

● 编辑命令

- x/De1 删除当前字符
- r+c 用输入的字符 c 替换光标所在字符
- i/R 以插入 / 替换状态进入编辑模式
- o/O 在当前行之前 / 之后插入新行，并进入编辑模式
- u 逐级撤销之前的操作
- Ctrl+r 逐级恢复被撤消的操作
- . 重复上一次的操作
- d\$/y\$ 将光标至行尾的内容删除 / 不删除并存入缓冲区
- dd/yy 删除 / 不删除当前行并存入缓冲区
- n+dd/n+yy 删除 / 不删除当前行往下 n 行并存于缓冲区
- p/P 在光标位置后 / 前插入缓冲区的内容
- /string 向后搜索第一个为 string 的字符串
- n/N 向后 / 向前继续搜索上次搜索的字符串

● 注：命令通常在一般模式下输入才能生效

vim 常用命令 (三)

● 文件操作命令

- `:e filename` 关闭当前文件并打开另一个文件
- `:n` 多文件编辑时跳至下一个文件
- `:N` 多文件编辑时跳至上一个文件
- `:sp` 横向切分窗口进行编辑
- `:vs` 纵向切分窗口进行编辑
- `:w` 保存当前文件
- `:w filename` 另存为其它文件
- `:q` 退出
- `:wq` 存盘并退出
- `:q!` 不存盘退出

● 注：命令通常在在一般模式下输入才能生效

命令帮助和软件文档

- 命令简单用法
 - `command --help`
- 命令详细说明
 - `man [-a] command`
 - `info command`
- 软件详细文档
 - `/usr/share/doc`
 - STFW.....
- 深腾 7000 的使用手册以及其它应用程序的详细文档都会放在 `/home_soft/doc` 目录中，管理员会随时更新和补充。



主要内容

- 一、深腾 7000 超级计算机介绍
- 二、基本操作和常用命令
 - 如何登录深腾 7000
 - 数据的上传和下载
 - 基本操作和命令
- 三、在深腾 7000 上提交和运行作业
 - 提交作业
 - 查看作业状态和运行结果
 - 其它作业操作

作业管理系统的基本概念

- 什么是作业
 - 对于深腾 7000 这样的并行环境来说，用户在系统中执行一次某个串行或并行的过程就是一个作业。
 - 一般说来，作业有如下属性：
 - 作业需要执行的程序
 - 作业执行的方式，如串行、并发执行、MPI 并行等
 - 对系统资源的需求，如 CPU、内存、虚拟内存、硬盘等
 - 对执行时间的需求
- 什么是作业管理系统
 - 存在于操作系统和用户之间，能够接收各类型用户作业，并对其进行处理的软件。
- 什么是队列 (Queue)
 - 在大规模系统（集群）中通过作业管理系统人为的划分出的用来接收不同作业计算资源的逻辑分块就是队列。
 - 通过对队列进行设置，使得队列能接收相应用户或满足相应资源需求的作业。

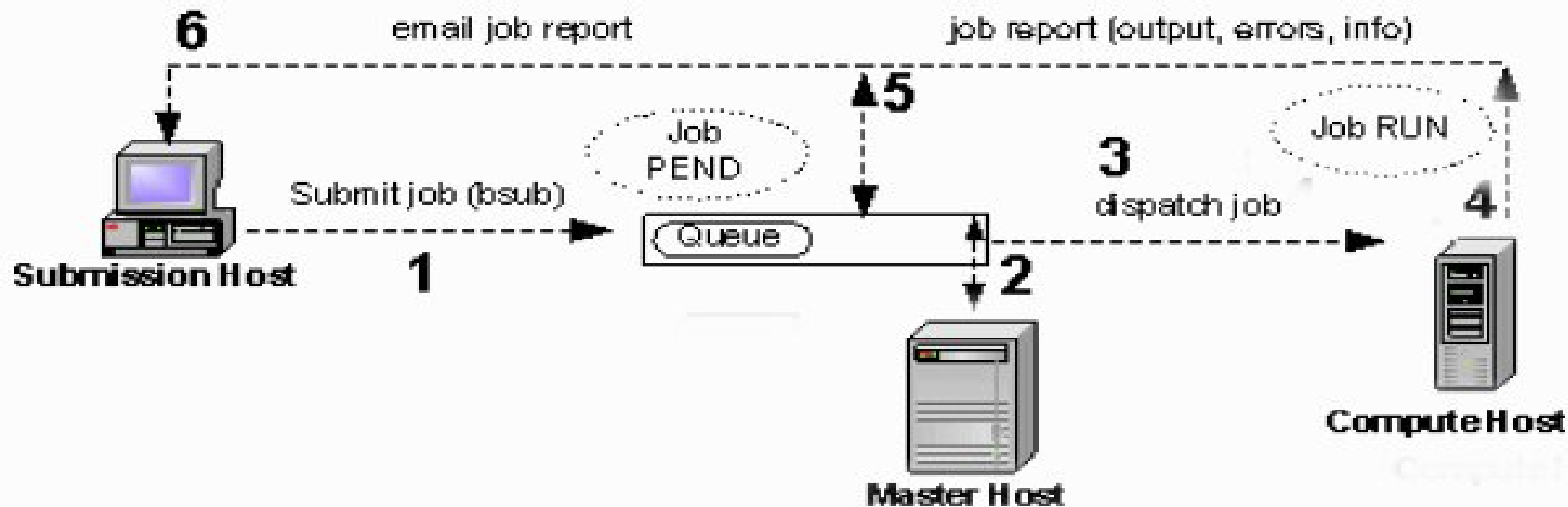
作业管理系统的功能

- 提供单一系统映像
 - 用户在任何结点提交的作业都被放入作业队列中等待调度，并保证运行结果相同。而用户不知道提供服务的计算结点的具体位置。
- 作业监控、调度
 - 通过调度策略，选取合适的作业准备运行。
 - 根据作业的请求，为其分配必要的系统资源。
- 负载均衡
 - 对系统的负载情况进行评估，如 CPU 利用率、页交换频率、内存使用率等，通过负载均衡算法为作业选取合适的运行结点（主要选取空闲结点）。
- 资源管理
 - 能够增加、修改、删除系统资源，并提供资源使用状况的动态信息。

几种常见的作业管理系统

- 商业级集群作业管理系统
 - Altair PBS Pro/Gridworks
 - Platform LSF/LSFHPC
 - IBM LoadLeveler
- 开源作业管理系统
 - OpenPBS (Torque)
 - Condor
 - SGE (Sun Grid Engine)
 - SLURM

LSF 作业生命周期



- 1、提交作业 — 用户提交作业，得到作业号，作业在队列中等待执行
- 2、作业调度 — 主服务器向调度器发调度请求，调度器选择最适合结点来运行作业
- 3、分发作业 — 主服务器收到调度器提供的结点列表，将作业分发至这些结点
- 4、作业运行 — 计算结点收到主服务器的指令，创建运行环境，执行作业
- 5、结果返回 — 作业完成，计算结点将输出和运行信息发回给主服务器
- 6、作业报告 — 主服务器向用户返回作业输出、退出信息

用 bsub 在深腾 7000 上提交作业（一）

- 此次培训会涉及以下两种类型作业的提交：
 - 使用单处理器核心运行的串行作业
 - `bsub -q Q_chem -n 1 -W [hour:]minute -o out.%J -e err.%J PROGRAM`
 - 使用多处理器核心运行的并行作业
 - `bsub -q Q_chem -n N -W [hour:]minute -R "span[hosts=1]" -o out.%J -e err.%J PROGRAM`
- 必须指定的参数：
 - `-q` 指定作业运行的队列名称（此次培训专用队列为 `Q_chem`）。
 - `-n` 指定使用 `N` 个 CPU 核心来运行作业（最大为 4，默认为 1）。
 - `-W` 指定作业的最大运行时间，需要由用户根据经验或实际情况进行估算，而且不能超出队列限定值（最大为 2 小时）。
 - `PROGRAM` 是需要运行的程序，如程序有运行参数，在后面依次写上。
 - 对于 `qchem` 作业，格式为 “`qchem.lsf infile outfile`”。
 - 对于 `Gaussian` 作业，格式为 “`g03.lsf infile`”。

用 bsub 在深腾 7000 上提交作业（二）

● 作业输出参数：

- -o 指定用于存放标准输出（即执行程序时原本显示在屏幕的正常输出信息）的文件，参数中可以使用 %J 代表作业的作业号。
- -e 指定用于存放标准错误输出（即执行程序时原本显示在屏幕的警告和报错信息）的文件，参数中可以使用 %J 代表作业的作业号。

● 作业号：

- 用 bsub 命令**成功提交一个作业**之后，系统会返回一条类似于：

“Job <599028> is submitted to queue <Q_chem>.”

的信息，这条信息显示了被提交作业的作业号（第一个括号里面的内容即 599028）以及作业提交到的队列（第二个括号中的内容即 Q_chem）。

- **作业号是每个作业的唯一标识**，建议每次提交作业后将对应的作业名及作业号记录下来。在提交作业之后对作业进行操作或是在作业退出之后查看作业历史信息时，都必须用到这个作业号。

● 作业提交示例：

- `bsub -q Q_chem -n 1 -W 2 -o outfile -e errfile echo "hello."`
- `bsub -q Q_chem -n 4 -W 1:30 -R "span[hosts=1]" -o out.%J -e err.%J qchem.lsf ftc_alanine10.in ftc_alanine10.out`

深腾 7000 上常见 bsub 提交失败信息

- You must specify your own runtime limit for this job through -W option.

Usage: bsub -W [hour:]minute -o outfile -e errfile -n X -q queue_name mpijob job_name

Comment: The more runtime limit you specified, the longer time you will wait for that job dispatched.

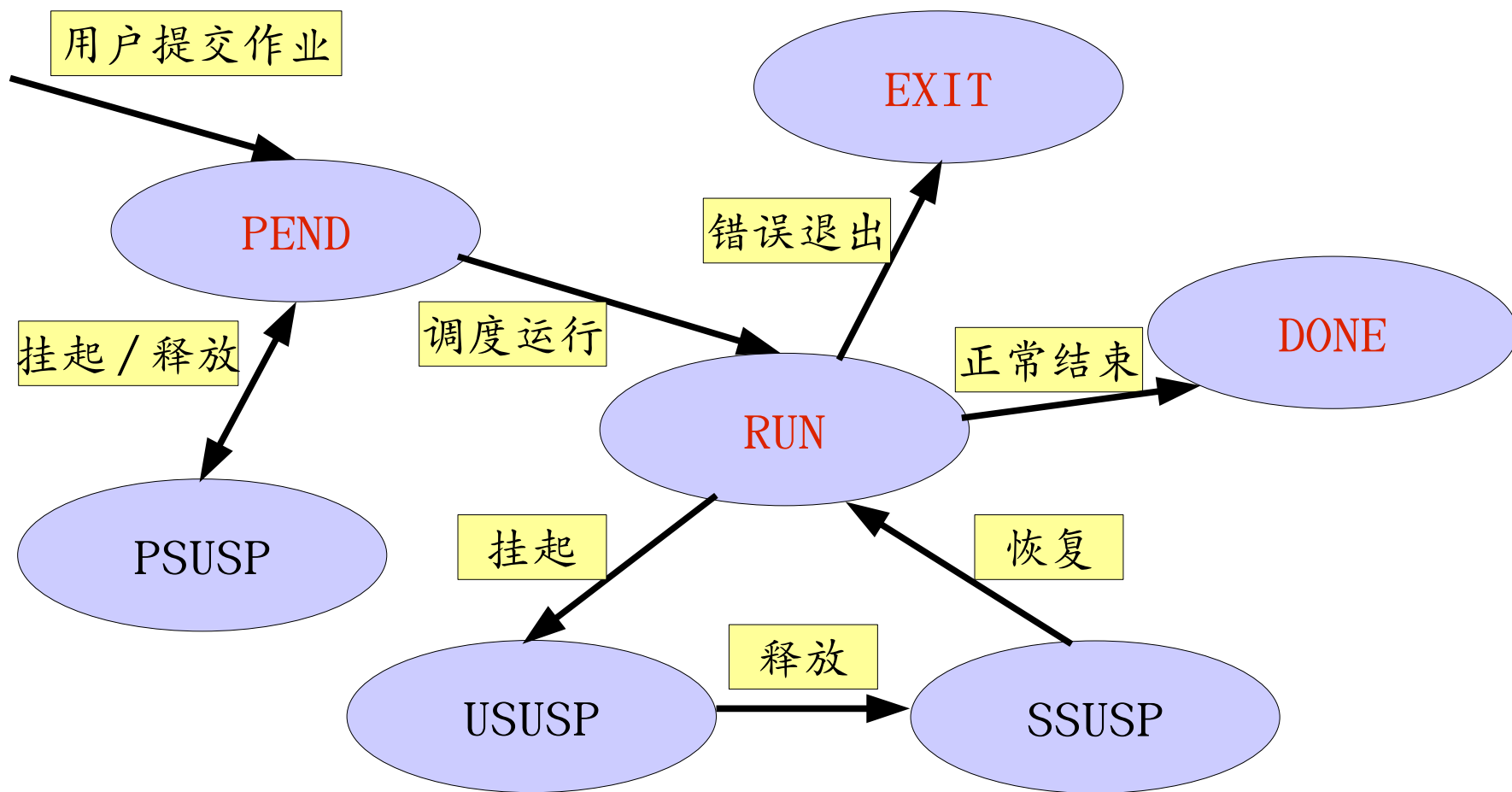
Request aborted by esub. Job not submitted.

- 没有通过 -W 参数设定作业最大运行时间。
- Too many processors requested. Job not submitted.
 - 作业申请的处理器数超出了队列的限制。
- RUNLIMIT: Cannot exceed queue's hard limit(s). Job not submitted.
 - 作业申请的运行时间超出了队列的限制。

LSF 作业状态

- 每个作业在提交成功到运行结束，在任一时刻都拥有一个作业状态，该状态表明这个作业在该时刻的基本情况。
- 常见的 LSF 作业状态有以下几种：
 - PENDING
 - 作业在队列中等待调度运行
 - RUNNING
 - 作业正在运行
 - DONE
 - 作业“正常”退出（退出码为零）
 - EXITED
 - 作业“异常”退出（退出码非零）
 - PSUSP
 - 作业在队列中等待调度时被其所有者或管理员挂起
 - USUSP
 - 作业在运行时被其所有者或管理员挂起
 - SSUSP
 - 作业被 LSF 系统挂起

LSF 作业状态转换示意图



用 bjobs 查看作业状态

- bjobs 的功能是查看系统中作业的情况。
- 直接执行“bjobs”命令会得到当前用户正在排队和正在运行的作业列表。
- bjobs 命令的常用参数如下：
 - -a 在不加任何参数的情况下，看到的只是自己提交的并且尚未结束的作业。如果使用了“-a”参数，除了未完成的作业之外，还能看到一些刚结束不久的作业的信息。
 - -u 如果需要查看系统中别的用户的作业情况的话，只需加上“-u”参数，比如想查看用户“user1100”的作业情况，那么执行“bjobs -u user1100”即可。如果执行了“bjobs -u all”的话，将会看到所有用户的作业信息。
 - -l 加上 -l 参数可以查看某个作业的详细信息，具体格式是“bjobs -l JOBID”。
- 当直接执行 bjobs 看不到之前处于运行状态的作业了时，表明该作业已运行完毕或异点退出，可以用 bjobs JOBID 查看其退出状态和等其它信息。

深腾 7000 上的其它作业操作

● 查看作业运行结果

- 作业结束后，运行时输出到屏幕上信息会保存在 bsub 提交作业时用 -o 和 -e 指定的用于存放输出的文件中，直接查看即可。
- 如作业运行时另有输出文件产生，则一般会在作业提交时所在目录下生成相应的输出文件。

● 查看运行中作业的输出 (bpeek)

- bpeek [jobID]
 - 不加 jobID 显示当前用户所有正在运行的作业的输出
 - 或是直接查看在 bsub 提交作业时通过 -o 或 -e 指定的输出文件。

● 删除作业 (bkill)

- bkill jobID

● 查看历史作业信息 (bhist)

- bhist [参数] [jobID]
 - -a 显示 1 小时内的作业状态，包括完成的、挂起的等
 - -u 显示指定用户作业的历史信息
 - -l 显示历史作业详细信息



谢谢大家！



曹宗雁

zycao@sccas.cn

2009年12月21日

中国科学院计算机网络信息中心
超级计算中心 系统运维部