hpc4you toolkit v2 for Cluster

零基础、零配置组建SLURM调度并行计算集群

用户手册

English Version https://hpc4you.github.io

v2: 管理和使用, 均通过指令模式

ask@hpc4you.top

二〇二五年八月十九日

目录

1 格式约定

为了便于查看, 主要排版约定如下:

- 文件名或路径: /path/file
- 变量名: MKLROOT
- 命令: command parameters
- 需按顺序逐行执行的指令:

```
export OPENMPI=/opt/openmpi/1.8.2_intel-compiler-2015.1.133
export PATH=$OPENMPI/bin:$PATH
export MANPATH=$MANPATH:$OPENMPI/share/man
```

- 命令输出或者文件内容:

```
QUEUE_NAME      PRIO STATUS          MAX JL/U JL/P JL/H NJOBS  PEND   RUN  SUSP
serial            50  Open:Active      -   16    -    -     0     0     0     0
long              40  Open:Active      -    -    -    -     0     0     0     0
normal            30  Open:Active      -    -    -    -     0     0     0     0
```

特别强调:

1. 指令, 是逐行执行. 也就是敲完一行或者复制粘贴一行内容, 就按Enter. 不是粘贴所有指令一起贴到命令行终端.
2. 指令严格区分大小写.
3. 所谓脚本文件, 就是把按顺序逐行依次执行的指令, 写在一个文档中.
4. 除非特别强调, 所有的操作, 均是采用 root 用户来完成.
5. 所有的操作指令, 字符以及标点符号, 都是关闭输入法, 在纯英文状态下输入的.
6. 一个常识, *#开头的内容*, 都是注释. 无论#出现在指令的任何地方, 包括*#在内及其右侧所有内容*, 都是注释.
7. “在master机器操作”, 即可以是通过网络ssh远程登录目标机器进行操作; 也可以是通过键盘、显示器, 直接操作master机器.

2 Quick Start

选一台机器做master机器,

1. 准备获取软件, 拔掉服务器上所有的U盘、移动硬盘等移动存储, 运行:

```
curl http://tophpc.top:1080/getInfo.sh | bash
```

   视频教程BV1NY4y1C7ya

2. 修改/etc/hosts文件, 录入所有机器的IP和机器名; root可登录, 密码都相同
   视频教程bv19A4y1U7uX

3. 上传压缩包hpc4you_toolkit*.zip
   视频教程BV1fJ411n7uV

4. 解压, 而后执行指令:
   source code (输入一次密码)
   后续均是复制屏幕提示的绿色指令, 粘贴后按回车键. 耐心等待, 集群就绪.

以微软用户视角, 操作演示, bilibili短视频BV1GY411w7ZV.

3 写在前面

高性能计算(High Performance Computing, HPC)是改造世界的第三大科学研究方法, 是大规模科学计算和工程计算的必备基础设施, 是科技创新的重要手段, 在信息服务、工业仿真、科学研究、生物信息、基因测序、石油勘探、航天航空等众多领域发挥着不可替代的作用, 是研究和解决各领域挑战性问题的重要手段, 是国家综合国力和科技创新力的重要标志, 也是世界大国投入巨资争夺科学技术制高点的领域之一.

高性能并行计算集群, 是实施高性能计算的平台. 在科学计算领域, 集群模式可以将多个分散、孤立的服务器组合为一台服务器, 告别手动找资源、找数据、多次编译安装软件的苦恼, 让多台计算设备自动服务于计算任务, 简化工作流程. 亦可拓展并行计算规模, 得以计算更大体系(取决于网络规格和应用场景).

当前, 很多高性能计算集群都采用Linux操作系统, 其运维工作比较烦琐, 尤其对科研一线的小伙伴非常困难, 更别说自己搭建并行计算集群.

本人从事计算材料、计算化学、计算生物学领域高性能计算集群系统管理、运维、集群系统调优等工作15年. 结合工作经验, 创造了一种傻瓜式快速构建slurm调度并行计算集群的方法. 组建得到的并行计算集群, 免维护、免管理, 可以最大限度解决一线科研科作者在自建高性能并行计算集群方面的实际痛点.

本人非计算机专业从业人员, 也未曾研修过任何计算机专业基础理论知识. 本手册以及hpc4you toolkit方案中, 除了提及科学计算中能涉及到的CPU核心、内存容量、GPU卡、磁盘空间概念之外, 不涉及任何其他Linux平台运维术语. 使用hpc4you toolkit方案, 用户仅需认识简单英文, 会把hpc4you toolkit放在目标机器上, 会使用tar解压缩, 会使用cd进入目录, 会使用vi或者cat修改一次/etc/hosts文档即可. 全程需要使用root权限, 用户需要晓得如何使用root用户登录机器, 或者如何切换为真正的root用户. [1]

[1] CentOS系列, 安装时候会设定root密码. 以普通用户登录后, 输入su -, 而后输入root用户密码, 即可成为真正root用户. Ubuntu系列, 以普通用户登录后, 输入sudo passwd, 会要求先输入当前用户密码给sudo授权, 而后输入两次新密码, 此新密码就是设定的root密码, 而后root用户就开启成功了. 当然, Ubuntu系列, 也是使用这个脚本https://gitee.com/hpc4you/linux/blob/master/enableRootLogin-ubuntu.sh, 会自动设定root密码为123456 (仅仅红色字符).

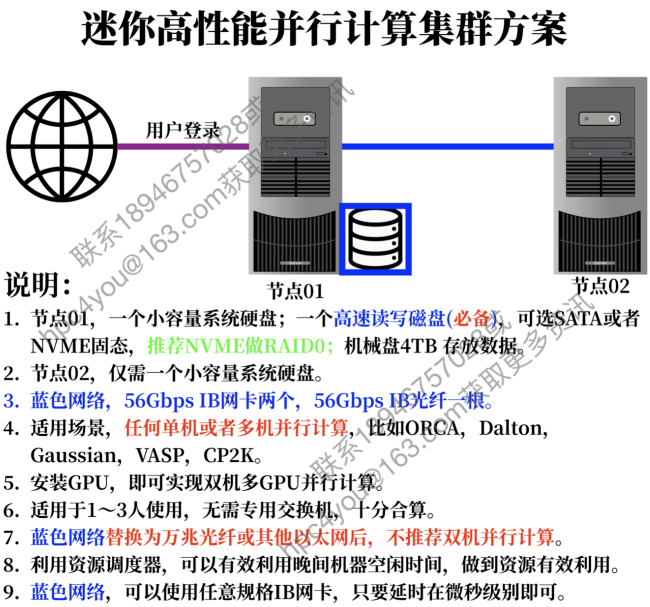

图 1: 双机并行迷你集群, 无需交换机.

在过去15年中, 我不仅做计算模拟, 也自己组建并行计算集群, 用于自己的计算研究, 主要跑VASP, Gaussian, ADF, ORCA, Dalton. 目前正常服役集群2套, 合计12个节点; 也为同事组建和维护集群, 目前正常服役4套, 合计16节点, 主要运行VASP, CP2K. 也协助多个课题组搭建过集群, 小至两节点的迷你集群, 大至16节点的集群, 或者是采用IB网络的20节点的并行计算集群, 十来套, 目前都在正常运行. 当然也曾参与管理过曙光、浪潮的集群, 计算节点高达200个. 目前也在为课题组管理一套24节点的浪潮的集群.

hpc4you toolkit, 由从事计算化学、计算材料专业的“计算机专业外行”结合工作实际开发制作, 充分理解需要HPC来实施的科学计算任务要做什么、需要什么资源、以及如何充分利用资源. 本工具套件, 将轻松规避科技工作者在组建高性能计算集群方面遇到的Linux相关的所有痛点和难点. 你所需要做的仅仅是, 复制粘贴指令, 按Enter键; 等待重启完成, 重新登录, 除此之外, 无他.

4 支持的集群架构

集群硬件架构方案可以查看图 1, 图 2和图 3. 各应用场景和注意事项, 请看图片上的文字, 或者查询页面https://hpc4you.github.io/. SLURM调度器轻松管理上万个处理器核心和各种加速卡. 实际支持的机器数量, 取决于交换机的实际容量. [2]

特别留意:

1. 为保证机器网络安全, 管理节点必需具备至少两个网卡, 用以隔离用户登录网络和集群运算/管理网络.
2. 如果管理节点仅仅负责存储空间和调度管理, 且计算节点不超过10台, 则管理节点可以使用具有两个网卡的普通PC机器来承担.
3. 不必要求所有机器处理器、内存规格一样, 只要求所有机器运行相同的操作系统版本.
4. 调度器本身支持CPU、GPU以及各类加速卡的混合调用管控.

[2]具有8个接口的交换机, 最多连接8台机器; 同理, 具有16个接口交换机, 最多连接16台机器, 这就是交换机容量.

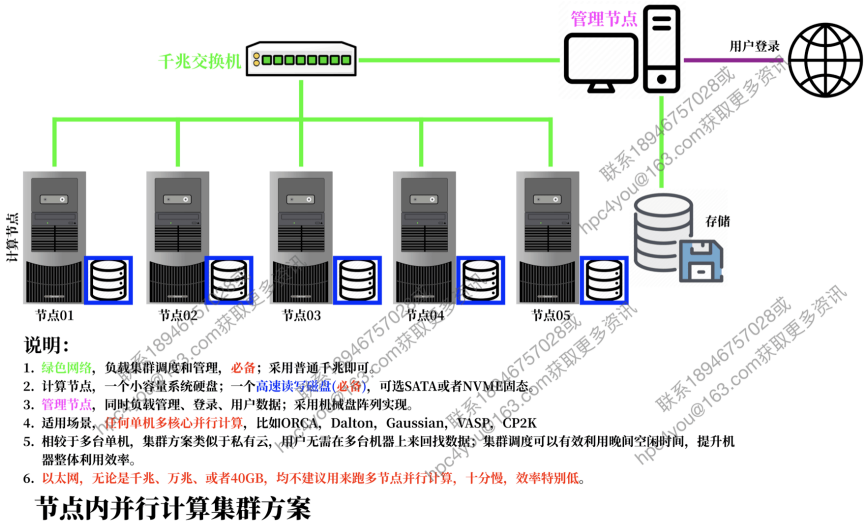

图 2: 节点内多核心并行集群.

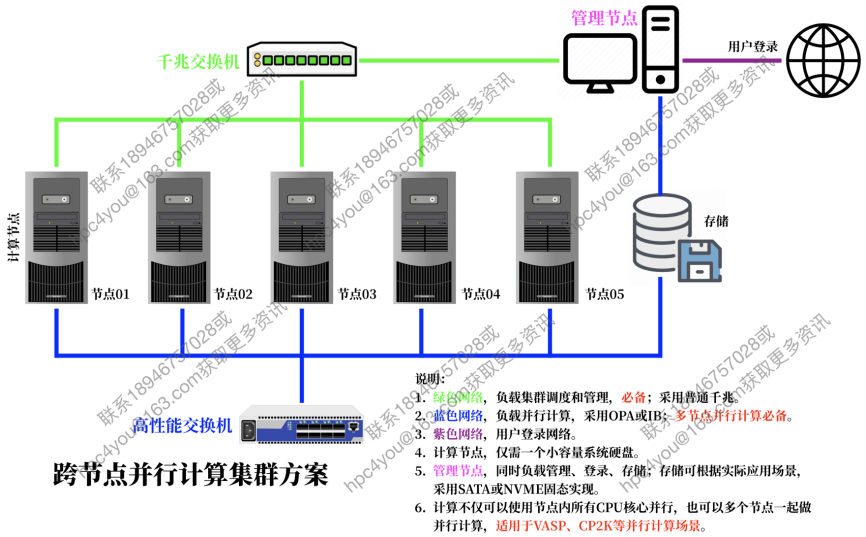

图 3: 跨节点并行集群.

温馨提示, 合理升级网络规格, 使用单独的服务器分别承担管理、登录; 使用商用存储服务器或者存储集群负责所有读写操作, 上文描述的**跨节点并行集群架构方案**, 即可适用于大规模集群. 高性能存储(或者并行文件系统)需要硬件支持, 不是耗费两三百元使用四五个硬盘就可以实现的.

5 功能模块与版本

5.1 原版本信息

根据使用场景差异, 本工具套件(hpc4you toolkit v2 for Cluster)分为基础版, 进阶版和专业版三个版本. 任何版本均支持GPU+CPU混合调度. 根据SLURM手册描述, GPU依赖于驱动, 如果无法自动识别, 则需手动调试.

进阶版本, 增加SLURM记账模块.

专业版, 提供更多模块. 尤其是用户管控, 可以杜绝任何形式的资源盗用. 如您的集群登录节点, 暴露在互联网, 安全设定模块可以协助您抵挡99%的安全威胁, 剩下那0.9999%, 请管理好你手里的微软机器. 系统调优模块, 可以让你的系统运行更加稳定. 动态监控和历史监控模式, 适合放在大屏幕, 给莅临视察的领导欣赏.

各版本模块差异请看表 1.

各版本包含的模块明细如下:

基础版 hpc4you_toolkit-basic**.zip, 包含如下文件:

```
enable_master-to-run-calculation.sh  getFileFingerPrint.sh
step1.sh  step2.sh  step3.sh  step4.sh
```

进阶版 hpc4you_toolkit-adv**.zip, 包含如下文件:

```
enable_master-to-run-calculation.sh  getFileFingerPrint.sh
step1.sh  step2.sh  step3.sh  step4.sh
enable_slurmLog-step1.sh enable_slurmLog-step2.sh
```

专业版 hpc4you_toolkit-pro**.zip, 包含如下文件:

```
enable_master-to-run-calculation.sh  getFileFingerPrint.sh
step1.sh  step2.sh  step3.sh  step4.sh
enable_slurmLog-step1.sh enable_slurmLog-step2.sh
enable_UserControl.sh enable_netdata.sh enable_ganglia.sh
```

其中**是系统版本标注, 比如是el7, el9等.

5.2 重要更新

原进阶版和专业版功能合并至hpc4you toolkit v3, HPC via Web for Cluster, 在hpc4you toolkit v2 for Cluster中, 不再提供.

6 系统与网络

如上所述, 这里讨论的并行计算集群, 是基于Linux操作系统和SLURM调度器调试而成的贝奥武夫架构的并行计算集群. 先得有机器, 机器必须安装同一个版本的Linux操作系统.

表 1: hpc4you toolkit各版本功能对比

功能模块	基础版/basic	进阶版/adv	专业版/pro
CPU调度	yes	yes	yes
GPU调度[1]	yes	yes	yes
记账功能	no	yes	yes
用户管控[2]	no	no	yes
历史监控	no	no	yes
实时监控	no	no	yes
安全设定[3]	no	no	yes
系统调优[3]	no	no	yes
定价	398元起[4]	电邮联系ask@hpc4you.top	电邮联系ask@hpc4you.top

[1] slurm天生支持GPU调度. 但是该功能依赖于GPU驱动以及其他相关库.
[2] 用户管控, 如无计算任务, 拒绝访问任何计算节点; 限定用户仅可使用已经申请到的计算资源. 更多讯息, 请查阅https://slurm.schedmd.com/pam_slurm_adopt.html.
[3] 安全设定和系统调优, 基于本人工作经验积累以及RHEL、Ubuntu操作手册相关章节.
[4] 教育优惠价, 基于“无技术支持的自助模式”. 更多资讯, 访问https://gitee.com/hpc4you/hpc查阅“定价与优惠”部分.

须知1: 在集群搭建过程中, 任何一台机器, 都需要访问互联网. 任何缺失的组件, 都需要通过dnf/yum/apt/pip来安装. 集群调试完毕后, 可以切断互联网链接.

须知2: 所有机器必须时间同步, 系统时间误差小于5秒钟. 如系统时间误差较大, 请先调整所有机器系统时间.

须知3: 如果需要安装任何第三方程序、驱动或软件(系统默认软件源中未包含的), 请务必在集群系统部署完毕后, 再单独进行安装.

6.1 支援系统版本

hpc4you toolkit支持RHEL7, RHEL8, RHEL9及其兼容系统, 比如CentOS 7.x, CentOS 8.x, CentOS Stream 8, CentOS Stream 9, Rocky Linux 8.x, AlmaLinux 8.x, AlmaLinux 9.x; Ubuntu 20.04 Focal、Ubuntu 22.04 Jammy、Ubuntu 24.04 Noble及其兼容系. CentOS 7.x, CentOS 8.x系列需要用户先行确认repo是否有效.

6.2 利旧前置条件

如果机器已经在运行如上任何一种Linux版本, 请跳过系统安装环节.

满足以下条件时, 可以直接开始组建集群:

1. 机器名唯一, 不重复.
   hostname看到的输出结果不同. 如需修改机器名, 使用nmtui指令.
2. 任何两台机器, 均可以通过IP登录.
   使用ip a查询到IP地址, 并通过ping测试网络.
3. 任何一台机器, 都可以访问互联网. 比如apt update或者yum update可以工作.
4. 未创建UID是500和501的用户和用户组.

此时, 因为没有私有独享交换机和网络, 只能构建为节点内多核心并行计算集群. 此使用场景, 集群模式能提高资源利用效率, 所有计算任务都由调度器自动分配计算服务器, 算完后, 数据自动回传, 避免来回找机器和找数据的烦恼.

考虑到实际负载, 不推荐借用公用交换机和公用网络来组建集群.

6.3 支援系统版本变更

结合现实情况, hpc4you toolkit v2 for cluster, 截至2025年后半年, 推荐使用RHEL8, RHEL9及其兼容系统; Debian 13.x; Ubuntu 24.04 Noble.

6.4 安装系统

新购机器, 安装Linux操作系统.

如今, Linux系统安装十分简单, 操作得当, 大约10分钟搞定. 以下教程有对应视频教程, 如有必要, 请自行查看.

step1 U盘安装系统必备工具Ventoy, https://www.ventoy.net/cn/index.html.
使用流程, 插入一个容量至少16GB的空白U盘, 点击运行Ventoy工具, 根据提示选中刚刚插入的U盘, 按照提示, 点击确认, 约20秒钟, 搞定. 而后下载系统镜像, 拖入该U盘里面即可.

step2 下载系统镜像
CentOS 7.9 https://mirrors.nju.edu.cn/centos/7.9.2009/isos/x86_64/CentOS-7-x86_64-Everything-2207-02.iso
Rocky Linux 8.9 https://mirrors.nju.edu.cn/rocky-vault/8.9/isos/x86_64/Rocky-8.9-x86_64-dvd1.iso
Ubuntu 20.04 https://mirrors.tuna.tsinghua.edu.cn/ubuntu-releases/20.04.4/ubuntu-20.04.4-desktop-amd64.iso
Ubuntu 22.04 https://mirrors.tuna.tsinghua.edu.cn/ubuntu-releases/22.04/ubuntu-22.04-desktop-amd64.iso
AlmaLinux 9 http://mirrors.nju.edu.cn/almalinux/9.0/isos/x86_64/AlmaLinux-9.0-x86_64-dvd.iso
下载后, 请勿解压, 直接拖入到step1处理完毕的U盘中.

注意, Rocky Linux, CentOS以及AlmaLinux都是RedHat企业版(RHEL)的衍生版本, 在操作配置层面, 方案都是通用的.

step3 安装系统
centOS 7.x和Rocky Linux是类似的, *系统安装教学视频*https://www.bilibili.com/video/BV11Z4y1M7xZ/. 系统安装过程中, 请使用英文语言和纽约时区.

step4 配置网络
CentOS 7.x、Rocky Linux、Ubuntu是类似的, *网络配置教学视频*https://space.bilibili.com/470332016/channel/collectiondetail?sid=268334.
如果是图形界面, 请打开终端(鼠标右键桌面选择Open Terminal; 或者按微软徽标键而后输入terminal), 在终端里面输入nmtui即可.
建议, master设定为192.168.XX.254, 网关是192.168.XX.254; 所有计算节点设定为192.168.XX.Y, 其中Y可以从1到253, 网关是192.168.XX.254.
管理节点和所有计算节点, 是通过交换机连接, 请勿使用路由器. 并且, 该专属交换机上, 只能接入 192.168.XX.Y线缆, 其他线缆一概不能接入. 网络设定和命名规范, 详见表 2.

表 2: 网络地址规范(推荐)

网络地址	机器名/hostname
192.168.251.254	master
192.168.251.1	node1
192.168.251.2	node2
192.168.251.3	node3
…	…

XX取50仅仅是一个示例. 实际中, 要确保选用的IP地址和办公室/实验室现有IP地址不重复.

特别留意, 管理节点必需标注为master, 计算节点必需标注为nodeY格式, Y为数字, 不可以使用其他字符. 此时master需要充当NAT网关, https://gitee.com/hpc4you/linux/blob/master/nat.sh, 脚本在此, 自己动手哦.

7 集群部署

重要注意事项:

文件操作规范 禁止在Windows系统解压hpc4you_toolkit*.zip压缩包; 所有操作必须通过SSH以root用户身份在指定的登录节点执行; 必须使用unzip ; source code方式运行工具套件.

硬件准备要求 确保所有硬件设备(包括附加存储设备)已完成物理安装; 必须完成硬件安装后再进行集群系统调试; 必须保持与制作hardwareXXX.dat文件时硬件完全一致.

驱动安装警告 若使用指定镜像安装系统后出现存储设备无法识别的情况, 禁止安装任何第三方驱动, 待集群系统完成安装调试后, 再处理驱动安装事宜. 系统镜像未包含的驱动均视为第三方驱动.

7.1 准备工作

7.1.1 在微软上操作(非必需)

如果晓得在Linux中如何使用vi编辑器, 可以跳过所有在微软上的操作.

在微软新建一个记事本, 文件名是net-info.txt. 如果是参考表 2做的网络设定, 那么该文件的内容应该如下:

```
192.168.251.254 master
192.168.251.1 node01
192.168.251.2 node02
192.168.251.3 node03
```

其中, 数字末尾和英文之间, 可以采用一个或者多个空格, 或者使用一个或者多个Tab键. 当前示例, 采用两个Tab键.

或者, 打开微软电子表格(Excel), 录入IP地址和机器名信息, 样式参考图 4, 而后复制电子表格中的内容, 粘贴到记事本, 并保存为文件名net-info.txt.

	A	B
1	192.168.251.254	master
2	192.168.251.1	node01
3	192.168.251.2	node02
4	192.168.251.3	node03
5	192.168.251.4	node04
6		

图 4: 电子表格内容示例

```
To protect your rights and ensure your eligibility for the paid hpc4you toolkit,
please send the following blue text via WeChat/WeiXin or email.
版权软件付费提供.
为验证您的付费资格并保护您的权益,请通过微信或者电邮发送以下两行蓝色内容.

586357bdbbc3aeb8e3b03c486ef75ba3  hardware379821.dat
SN: VL4i6MWU

- WeChat/Weixin/微信请联系: hpc4you
- Email/电邮,请发送至: ask@hpc4you.top

Good Luck.
```

图 5: 运行curl指令后, 屏幕输出内容示例(背景颜色可能不同).

7.1.2 在master操作获取授权许可

移除master机器上所有的移动硬盘、U盘等所有外置硬盘.

登录master机器(亦可以普通用户执行如下指令), 在终端中, 执行指令,[3]输出示例参看图 5.

```
bash <(curl -k -Ss https://gitee.com/hpc4you/hpc/raw/master/getInfo.sh)
或者
curl http://tophpc.top:1080/getInfo.sh | bash
```

如果运行该指令后, 在屏幕有看到任何错误提示, 请使用root用户运行如上指令.

而后根据屏幕提示, 发送电子邮件. 如有疑问, 请查阅B站视频BV1NY4y1C7ya.

如果担心操作失败, 请使用如下方式运行:

```
su -
cd /tmp
bash <(curl -k -Ss https://gitee.com/hpc4you/hpc/raw/master/getInfo.sh)
或者
curl http://tophpc.top:1080/getInfo.sh | bash
```

此时, 得到的文件在是/tmp/XXXX.dat. 如果硬件不变, 多次进行如上操作, 得到的文件指纹(digital fingerprint)必定相同.

curl指令后面是否有参数, 不影响结果.

hpc4you toolkit套件通过电邮提供, 请查询您的电邮并检查附件. [4]

[3]如果质疑脚本的安全性, 请下载后, 先打开看看. 直接在微软, 使用浏览器访问https://gitee.com/hpc4you/hpc/raw/master/getInfo.sh即可打开.

[4]Windows <–>Linux互传文件, 请查询教学视频, https://www.bilibili.com/video/BV1fJ411n7uV

```
Sorry.
You are NOT licensed to run this app.
Please contact ask@hpc4you.top via email to request a valid license file.
License files are only available upon payment.
Contact ask@hpc4you.top for details.
Bye.
```

图 6: 无效授权许可会看到的提示信息(背景颜色可能不同).

授权许可基于上述指令抓取到的硬件信息生成. [5]

如果提示“Command not found ...”, 请优先运行如下指令, 而后再次尝试.

```
yum -y install unzip zip tar wget curl # RHEL 7及其兼容系统
dnf -y install unzip zip tar wget curl # RHEL 8, 9及其兼容系统; OpenEuler
apt update && apt -y install unzip zip tar wget curl # Ubuntu及其兼容系统
```

7.1.3 无效授权处置

如果没有授权许可, 运行任何模块, 都会看到警告信息, 详细查阅图 6.

授权失效将导致以下影响:

1. 您无法使用本工具套件部署集群;
2. 如果集群已经安装完成, 变更硬件会导致授权失效, 您将无法使用任何_hpc结尾的指令, 包含添加/删除用户, 重启/关闭集群(除非手动操作); 也无法再次使用本工具套件的任何模块;
3. 已经添加的用户仍可继续使用集群, 集群系统会继续正常运行.

重装系统会导致现有许可失效. 请自行重置machineID条目即可恢复许可. 或者同时提供新旧dat文件, 电邮联系重新获取工具. 更多信息, 请查阅https://gitee.com/hpc4you/hpc/blob/master/FAQ.md以及 https://gitee.com/hpc4you/hpc/blob/master/TOS.md.

7.1.4 上传hpc4you toolkit压缩包

为避免不必要的麻烦, 请使用root账户操作文件上传.

上传压缩包hpc4you_toolkit-*.zip到master机器/root目录.

上传net-info.txt文件到master机器/root目录(如直接在Linux中使用vi编辑器, 可以跳过本步骤).

根据教程B站视频号BV1GY411w7ZV操作, 文件会自动出现在master机器/root目录.

7.1.5 继续在master机器操作

请采用root用户通过ssh登录到master机器, 继续操作.

修改网络信息, 请依次执行指令(仅适用于在微软系统上创建了net-info.txt文件的情形):

```
cp /etc/hosts /etc/hosts.original
```

[5] 具体细节, 可以查阅getInfo.sh脚本.

```
dos2unix /root/net-info.txt
```

```
cat /root/net-info.txt >> /etc/hosts
```

录入网络信息操作完毕. 请根据实际情况, 替换net-info.txt为实际的文件名.

如果你晓得使用vi, 请直接修改master机器上的/etc/hosts文件, 额外添加如下的内容并保存:

```
192.168.251.254 master
192.168.251.1 node01
192.168.251.2 node02
192.168.251.3 node03
```

7.1.6 主机名与IP地址配置规范

命名规则

禁止使用的名称 不得使用localhost或null作为主机名.

标准命名格式 必须采用node+数字的命名方式(如node123); 非标准命名将导致部分集群功能不可用.

地址映射规则

别名配置 允许单个IP地址对应多个主机名(私有别名); 私有别名可与hostname命令返回值不同; 所有别名必须指向同一物理节点

配置文件管理 新增记录将追加至/etc/hosts文件末尾; 无需修改现有文件内容; 必须确保所有主机名唯一不重复

严格限制

禁止单个主机名/别名对应多个物理节点.

7.1.7 自定义机器名(非必须)

在本方案中, 机器名或者别名是master的机器, 被认定为主控节点, 负载存储、登录、和管理, 也可以兼做计算节点. 机器名或者别名是nodeXX的机器, 都是计算节点.

假设有四台机器, 在同一个楼宇内(未必是同一个局域网), 四台机器网络互通. 显然这四台机器均可访问互联网. 机器名和IP信息如下:

```
192.39.40.12 work88
192.44.77.88 dell
99.77.98.11 zhang11
34.98.11.69 gov
```

如果保持现有机器名不变, (机器名, 就是输入hostname看到的输出结果; 修改hostname, 当然是输入nmtui), 那么必须在标注为master的机器上, 保持/etc/hosts中现有内容的基础之上, 额外添加如下内容:

```
192.39.40.12 work88
192.44.77.88 dell
99.77.98.11 zhang11
34.98.11.69 gov

# for HPC
192.39.40.12 master
192.44.77.88 node80
99.77.98.11 node3
34.98.11.69 node1
```

计划当作计算节点的机器, 别名必须是node开头, 但是node后面的数字不必是连续的; 这些机器中, 必须有一个, 标注为master.

本示例中, 任何一台机器, 都有一个真正的机器名, 以及一个私有别名. 比如机器192.44.77.88, 真正机器名是dell; 私有别名是node80. 如果要给机器192.44.77.88再来一个别名prof, 那就是再添加一行192.44.77.88 prof到文件/etc/hosts.

集群调试完毕后, 仅可删减/添加node, 但是不可替换/更换master机器; 以上网络信息录入, 仅可在集群调试之前进行.

7.1.8 确保软件源工作

本工具套件, 依赖既有软件源, 从开源可靠的软件源安装必要的组件, 务必确保*.repo或者source.list文件有效, 保证apt/dnf/yum update工作正常.

RHEL及其兼容系统, 如何判定yum/dnf工作正常? 在终端以root用户输入:

```
yum -y install epel-release
yum clean all
sed -i "s/enabled=0/enabled=1/g" /etc/yum.repos.d/*repo
yum makecache
```

如果没有看到任何错误提示, 那就是yum/dnf工作正常呀(RHEL8, RHEL9及其兼容系统中, 对于当前场景, yum与dnf等价).

如果看到有类似:

```
Errors during downloading metadata for repository ...
```

那就说明yum/dnf无法正常工作. 或者是因为, 你的软件源设定有问题, 或者网络有问题. 请自行解决, 或者寻求技术支持. [6]

Ubuntu系统, 如何判定apt工作正常? 在终端以root用户输入:

```
apt clean all
apt update
```

[6]一般而言, RHEL系统默认的软件源无需修改, 但是要确保你所在的网络环境下, 是否限制访问某些软件镜像. 对于RHEL及其兼容系统, 需要epel软件源(如不存在, 本工具套件会自动安装), 请确保你所在网络环境, 可以正常访问epel软件源. 如对访问国际互联网没有信心, 请参考https://mirrors.tuna.tsinghua.edu.cn/help/epel/, 修改为中国大陆教育网镜像.

如果没有看到任何错误提示, 那就是apt工作正常呀(在当前使用场景中, apt和apt-get是等价的). [7]

罕见情形是, 单独测试yum update或者apt update不报错, 但是在运行hpc4you_toolkit过程中, 又看不到部分yum/dnf/apt报错. 如果发生, 说明网络不稳定, 无法保持和既定软件源的稳定链接. 请检查网络, 重新运行本工具套件.

7.1.9 手动设定软件源

如下配置软件源配置, 针对教育网优化. 可以根据自己的网络环境, 自行判定是否使用.

如何使用? 请使用root用户, 仅仅在master机器上逐行复制粘贴按回车即可.

Rocky9.x, 请逐行复制粘贴按回车:

```
rm -fr /etc/yum.repos.d/*repo
curl https://gitee.com/hpc4you/linux/raw/master/repos/rocky9/rocky-edu-auto.repo > /etc/yum.repos.d/rocky9-hpc4you.repo
dnf makecache
```

更多讯息, 请查阅https://gitee.com/hpc4you/linux/tree/master/repos/rocky9.

Rocky8.x, 请逐行复制粘贴按回车:

```
rm -fr /etc/yum.repos.d/*repo
curl https://gitee.com/hpc4you/linux/raw/master/repos/rocky8/rocky-edu-auto.repo > /etc/yum.repos.d/rocky8-hpc4you.repo
dnf makecache
```

更多讯息, 请查阅https://gitee.com/hpc4you/linux/tree/master/repos/rocky8.

CentOS8.x, 请逐行复制粘贴按回车:

```
rm -fr /etc/yum.repos.d/*repo
curl https://gitee.com/hpc4you/linux/raw/master/repos/centOS8/centOS8-hpc4you.repo > /etc/yum.repos.d/centOS8-hpc4you.repo
dnf makecache
```

更多讯息, 请查阅https://gitee.com/hpc4you/linux/tree/master/repos/centOS8.

AnolisOS 8.x, 请逐行复制粘贴按回车:

```
rm -fr /etc/yum.repos.d/*repo
curl https://gitee.com/hpc4you/linux/raw/master/repos/anolisOS8/AnolisOS-hpc4you.repo > /etc/yum.repos.d/AnolisOS-hpc4you.repo
dnf makecache
```

更多讯息, 请查阅https://gitee.com/hpc4you/linux/tree/master/repos/anolisOS8.

其他Linux发行版系统, 请自行查阅手册设定软件源. 如果对自己的网络环境有信心, 直接使用系统默认的软件源配置即可.

7.2 运行hpc4you toolkit, 部署集群系统

本工具套件历经充分测试, 可在红帽企业版诸如RHEL7、RHEL8、RHEL9及其兼容系统; Ubuntu20.04, 22.04, 24.04及其兼容系统正常工作; Debian 13.x及其兼容系统. 在华为OpenEuler 22.03 LTS也工作. CentOS, RockyLinux, AlmaLinux, Oracle Linux等均是RHEL兼容系统. 不支持CentOS 8.x系列.

当前集群方案, 所有机器共享/home和/opt目录. 任何后续安装的软件以及各种MPI程序, 必需安装到/opt目录, 否则无法工作. [8]

[7] 一般而言, Ubuntu系统默认的软件源无需修改, 但是要确保你所在的网络环境下, 是否限制访问某些软件镜像. 如对访问国际互联网没有信心, 请参考https://mirrors.tuna.tsinghua.edu.cn/help/ubuntu/, 修改为中国大陆教育网镜像.

[8] 如有其他路径要添加, 请在集群配置完毕后, 首先在master修改/etc/exports文件; 而后再修改所有nodeXX上的/etc/fstab文件. 也许setup_hpc --sync_do XXX能帮到您. 但是您需要具备一定的NFS以及fstab修改经验哦, 否则机器可能无法启动. 或者联系 hpc4you@163.com 获取帮助.

必须采用root用户直接登录主控机器完成以下各操作. [9]

请勿在微软系统解压hpc4you_toolkit*.zip压缩包.

运行hpc4you toolkit, 以下两种模式, 任选其一.

7.2.1 自助模式

第一步 确认已经完成小节 7.1描述的准备工作.

第二步 确认所有机器均已开启root登录, 并且密码相同.

第三步 解压hpc4you_toolkit*.zip, 而后输入指令:

```
source code
```

后续所有的操作指令, 都会自动在屏幕上以绿色显示, 直接复制粘贴按回车键即可完成集群组建.

7.2.2 技术协助模式

第一步 确认已经完成小节 7.1描述的准备工作. 提供有偿技术协助.

第二步 确认所有机器均已开启root登录, 并且密码相同.

第三步 查阅来自ask@hpc4you.top的电子邮件, 拷贝邮件中标注的两行指令, 粘贴到终端, 按回车键, 耐心等待(取决于网络), 即可完成集群组建. 提供便捷指令, 粘贴复制即可, 无需下载、上传操作.

7.3 集群就绪

运行完毕屏幕上提示的所有绿色指令之后, 在master节点执行指令sinfo -lNe, 应该看到类似如下的信息, 说明集群已经正常工作了. [10]

```
[root@master ~]# sinfo -lNe
Wed Nov 03 19:43:39 2021
NODELIST   NODES PARTITION       STATE CPUS    S:C:T MEMORY TMP_DISK WEIGHT AVAIL_FE REASON
node01         1    workq*        idle 4       1:4:1   7982        0      1   (null) none
node02         1    workq*        idle 4       1:4:1   7982        0      1   (null) none
[root@master ~]#
```

至此, 集群系统组建完毕.

如果未看到如上类似信息, 原因是机器系统时间不一致. 可以使用指令date;pdsh date来确认.

同步所有机器系统时间, 既可以恢复工作. 可以使用如下指令同步集群时间:

```
setup_hpc --sync_time
```

[9] Ubuntu系统, 如果安装的是桌面版本, 默认禁止root用户登录. 可以使用脚本https://gitee.com/hpc4you/linux/raw/master/enableRootLogin-ubuntu.sh 开启root账户和root远程登录.

[10] 这仅仅是一个示例信息. 你看到的实际输出样式类似,但是具体内容肯定不同.

7.4 特殊案例双机集群

双机集群或双节点集群, 顾名思义, 一共有两台机器. 其中一台, 标注为master, 承担管理、存储、计算; 另一台, 标注为nodeX, 作为纯计算节点. 直接用网线(或其他光纤)互联, 无需交换机.

完成集群搭建后, 在master机器, 额外运行:

```
./enable_master-to-run-calculation.sh
```

即可完成对master节点的配置. 如果技能不熟练, 请重启整个集群以确保配置生效.

8 操作hpc4you toolkit组建集群, 实况视频

以上文字版描述, 也可以查看如下的实况视频.

RHEL/CentOS平台 B站视频BV19A4y1U7uX或BV1GY411w7ZV, 有旁白讲解

Ubuntu平台 B站视频BV1j44y1u7YR

原始录像 阿里云盘 https://www.aliyundrive.com/s/GrcXoWrccTP

其实无论哪一个系统, 操作流程和屏幕提示信息都是一样的. 升级后的版本, 所有操作都是复制屏幕提示的绿色指令, 无需用户输入任何cd, ls之类的操作指令.

9 集群系统管理

集群系统免维护, 免管理. 唯一的管理工作是添加/删除用户信息.

显然, 所有的管理操作, 必须在master节点进行.

机器硬件故障, 不属于运维管理范畴. 硬件坏了, 找商家走售后流程.

9.1 修改root密码

集群组建完毕后, 可以直接在master机器, 修改root密码, 直接输入:

```
passwd
```

按照屏幕提示操作输入新密码即可.

最简单的安全设定是, 集群配置完毕后, 仅仅允许特定地址登录root用户. 所有计算节点, 拔掉互联网线缆即可.

其他高级的安全设定, 请查阅Linux SSH安全设定.

9.2 添加用户

在master机器上, 采用默认用户组,

```
useradd_hpc tom # 默认添加用户tom到用户组users
```

或者, 指定用户组:

```
useradd_hpc chem tom # 将添加用户tom到用户组chem, 如果chem不存在, 则会自动创建.
```

注意1, 在添加用户时候, 需要给用户设定新密码. 设定密码环节, 需要输入两次, 但是显示器不会有任何提示符, 输入完毕后, 按回车即可.

9.3 删除用户

在master机器上, 执行如下指令:

```
userdel_hpc tom
```

当前示例中, 用户名是tom.

9.4 集群开机

无论何种情形, 开机, 是先开启交换机电源, 再开启主控/管理节点. 待管理节点启动完毕后, 再开启其他计算节点.

9.5 集群关机

关机, 请在master机器执行:

```
poweroff_hpc
```

9.6 集群重启

重启, 请在master机器执行:

```
reboot_hpc
```

如果机器具有IPMI功能, 请联系硬件供货商, 配置IPMI, 并查询手册, 学习如何使用IPMI. 如果不想使用指令, 那么去现场按电源吧.

10 添加新机器

10.1 硬件克隆

最便捷的操作是, 硬件物理克隆.

1. 克隆当前任何一个计算节点的系统盘;
2. 安装到新机器, 接好网络线缆. 开机, 修改hostname, 修改网络设定.
3. 去master上面修改/etc/hosts文档, 添加新机器的信息(IP和机器名).
4. 同步/etc/hosts, 执行setup_hpc --sync_file /etc/hosts
5. 新机器上执行slurmd -C, 获取必要信息; 比如看到类似:

   ```
   abbott@e5node1:~ $ slurmd -C
   NodeName=e5node1 CPUs=36 Boards=1 SocketsPerBoard=2 CoresPerSocket=18 ThreadsPerCore=1 RealMemory=128838
   UpTime=106-06:53:07
   abbott@e5node1:~ $
   ```

   仅仅需要复制其中的

   ```
   NodeName=e5node1 CPUs=36 Boards=1 SocketsPerBoard=2 CoresPerSocket=18 ThreadsPerCore=1 RealMemory=128838
   ```

   并添加到管理节点的/etc/slurm/slurm.conf文档中类似区域即可.
6. 全集群重启slurmd, 执行setup_hpc --sync_do 'systemctl restart slurmd'

7. master重启控制端, 执行systemctl restart slurmctld

如果不具备在Linux平台dd指令克隆技能, 请购买一个绿联的脱机硬盘拷贝盒. 按照硬盘盒说明书, 插入源盘和新盘, 按一下克隆按钮, 耐心等待, 即可搞定硬盘克隆. 硬盘容量越大, 克隆所需要的时间越久. 使用120GB SATA接口固态是明智的选择. 当然, 绿联的硬盘盒, 也支持NVMe硬盘克隆.

10.2 在线自动配置

此模块仅在hpc4you toolkit v3, HPC via Web for Cluster提供.

执行addNewComputeNode.sh, 自动配置, 无需拆卸硬盘. 不影响既有集群和正在运行、排队的任务.

操作流程如下. 按照屏幕提示, 输入:

1. 新机器的IP地址
2. 输入新节点的名字, 比如是node8

然后等待. 等待时间和网络快慢相关, 大约30分钟起步.

全程需要互联网连接, 同时也需要事先给新机器安装好Linux系统, 并配置完毕集群私有网络和互联网.

专业版默认提供此模块. 亦可单独提供此模块, 费用和绿联脱机硬盘克隆盒子相当.

警告: 不支持以多线程方式运行addNewComputeNode.sh模块同时添加多个新节点到现有集群系统. You cannot add multiple new nodes to a cluster simultaneously using the addNewComputeNode.sh module.

11 高阶功能(部分模块不再提供)

当前, hpc4you toolkit支持的四个高阶功能如下:

1. 解禁算力, 让master负担双角色. 适用于登录节点配置较好的场景, 或者双节点迷你集群场景.
   免费赠送
2. 开启记账功能, 便于查询作业信息记录, 便于统计计时消耗. 不仅仅是记账, 核心功能是Accounting and Resource Limits.
   高级版和专业版支持/adv and pro
   根据屏幕提示, 运行enable_slurmLog-step1.sh和enable_slurmLog-step2.sh即可.
   此模块仅在hpc4you toolkit v3, HPC via Web for Cluster提供.
3. 资源管控, 让资源盗用者无处遁形, 保证所有计算均通过调度器进行.
   专业版支持/pro only
   此模块仅在hpc4you toolkit v3, HPC via Web for Cluster提供. 根据屏幕提示, 运行enable_userControl.sh.
   功能演示 https://www.bilibili.com/video/BV1j3411e7xz/
   此模块仅在hpc4you toolkit v3, HPC via Web for Cluster提供.
4. 监控模块, netdata + ganglia, 可以给出漂亮的动态监测图表, 视察的领导绝对喜欢.
   专业版支持/pro only

功能演示 https://www.bilibili.com/video/BV1Sb4y167Nw/
https://www.bilibili.com/video/BV12J411V7op
此模块仅在hpc4you toolkit v3, HPC via Web for Cluster提供.

5. 安全+系统调优, SSH安全加固, 内核参数调整等.
   专业版支持/pro only
   根据屏幕提示, 运行enhance_security.sh.
   基于本人工作经验积累以及RHEL、Ubuntu管理员手册相关章节.
   此模块仅在hpc4you toolkit v3, HPC via Web for Cluster提供.

11.1 ganglia负载监控

本配置, 需要重启整个集群一次. 如果计算节点不在同一个局域网, 可能导致ganglia运行失败.

This section describes installing and testing Ganglia, a system for monitoring and capturing metrics from services and components of the cluster.

需要在所有节点上通过yum/dnf安装ganglia和相关依赖, 需要互联网畅通. 配置完毕后, 即可在本地网络工作, 无需互联网.

在master节点, 复制粘贴屏幕提示的指令, 或者运行如下指令:

```
./enable_ganglia.sh
```

按照屏幕提示操作即可.

ganglia监控通过浏览器界面呈现.

RHEL7, RHEL8机器兼容系统, 支持以下三种登录鉴权模式, 请根据屏幕提示, 选择其中一个.

1. 不使用密码保护, 即所有知道登录节点域名或者IP的人, 都可以通过浏览器看到监控信息.
2. 使用默认用户名和密码, 用户名是hpc_monitor, 密码是8566262.
3. 创建新的用户名和密码.

最后, 打开浏览器, 地址栏输入指定地址, http://IP-Address-of-master-node/hpc4you, 即可看到整个集群的历史监控数据. 默认, 每分钟刷新一次. 主要涉及CPU负载, 内存负载, 磁盘空间, 网络负载等. 运行示例, 请看图 7, 图 8和图 9.

特别留意:

1. Ubuntu 20.04 Focal, Ubuntu 22.04 Jammy, Ubuntu 24.04 Noble, RHEL9.x, 不支持密码验证.
2. OpenEuler 22.03 LTS, 不支持ganglia监控.

11.2 netdata集群负载实时监控

采用netdata采集实时监控数据, 默认仅仅存储1小时历史数据. 自动汇集所有计算节点监控数据到登录节点或者管理节点, 实时显示.

图 7: ganglia集群监控运行示例. 第一步, 打开浏览器登录.

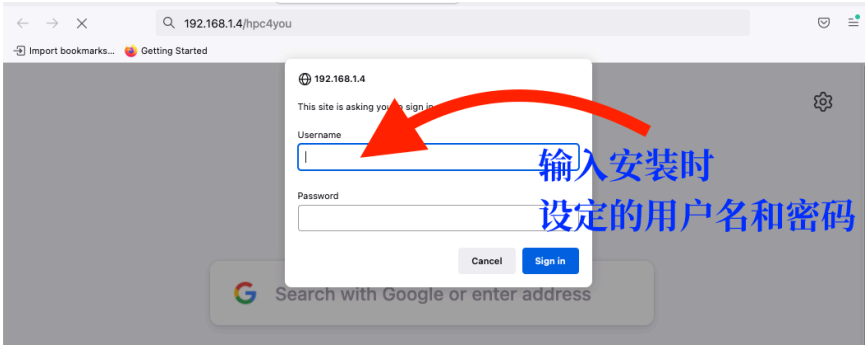

图 8: ganglia集群监控运行示例. 第二步, 输入用户名和密码(如有设定).

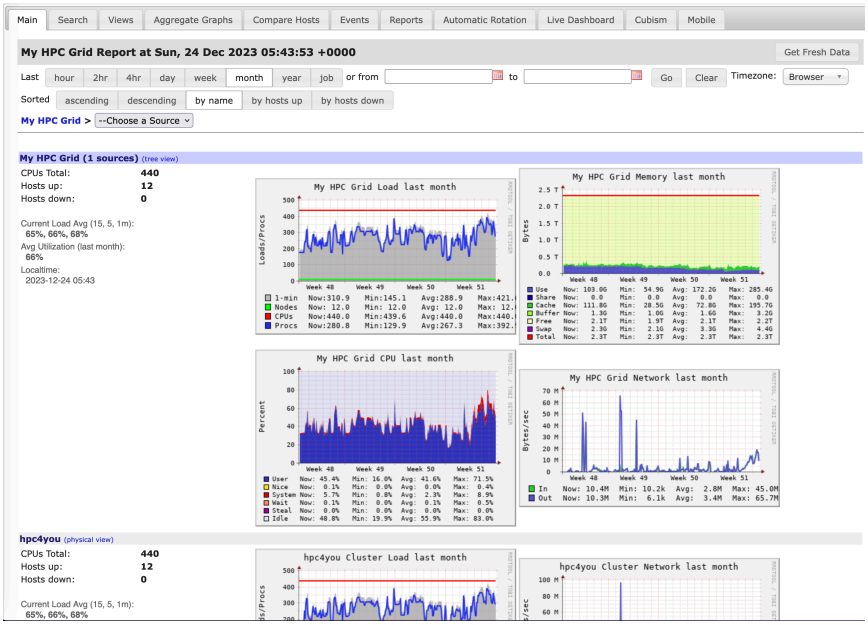

图 9: ganglia集群监控运行示例. 监控信息演示.

```
All Done.



Please open Chrome/Firefox,
then type htttp://IP-Address-Master-Node:19999 into the address bar.
In the opened webpage, click on the leftmost white > to check all nodes.


[root@master slurm-c7]# date
Sun Oct 17 13:37:01 CST 2021
[root@master slurm-c7]#
```

图 10: 安装netdata完毕后, 指令行看到的提示信息.

图 11: 查看netdata实时监控信息, 第一步, 打开浏览器.

当前部署方案, 会自动化配置, 除了执行一键脚本之外, 无需任何其他配置操作. 请留意, 安装过程需要互联网支持. 所需组件采用yum/dnf从官方软件源安装.

在master节点, 复制粘贴屏幕提示的指令, 或者运行如下指令:

```
./enable_netdata.sh
```

注意, 脚本名称一样, 但是脚本内容不同. centOS7和centOS8操作系统中, netdata配置使用方式不太相同. 安装输出请看图 10.

在能访问登录节点或者管理节点的网络环境中, 打开浏览器, 推荐使用真英文版Firefox, 可以查看实时监控信息. 使用示例, 请看图 11, 图 12, 图 13.

特别留意, 动态实时监控暂不兼容OpenEuler系统.

12 SLURM技能自我修养

本手册主要介绍hpc4you toolkit工具套件的使用流程. hpc4you toolkit工具套件是一款功能强大、易于使用的HPC工具包, 可帮助您轻松搭建高性能计算集群.

对于使用集群进行计算的用户而言, 运行计算任务的流程与使用超算平台一致. 具体步骤包括: 提交作业脚本、监控作业状态、查看作业结果等.

图 12: 默认显示登录节点实时监控信息. 查看其他节点信息.

图 13: netdata显示的节点列表示例.

Slurm相关的管理策略和使用技巧需要由管理员自行学习掌握. 管理员可以参考官方文档、社区论坛等资源进行学习, 或者参加Slurm商业培训.

本章节收集了一些Slurm相关参考资料, 请管理员自行学习掌握.

12.1 选一个浏览一下

1. https://slurm.schedmd.com/
   原版英文手册最佳
2. https://docs.slurm.cn/users/
   中文资料
3. http://hmli.ustc.edu.cn/doc/userguide/slurm-userguide.pdf
   中国科大超级计算中心出品
4. https://bicmr.pku.edu.cn/~wenzw/pages/slurm.html
   北京大学某研究组出品
5. https://www.cloudam.cn/help/docs/cloudE10
   查看slurm作业管理系统部分北鲲云编写
6. https://leo.leung.xyz/wiki/Slurm
   centOS8 PAM Slurm Adopt Modulem | Very nice wiki.

12.2 快速制作slurm脚本

参考这个站点
https://www.hpc.iastate.edu/guides/classroom-hpc-cluster/slurm-job-script-generator.
一个简单的slurm脚本设定, 只需在这个页面上填写:

Number of compute nodes 1 数字1代表使用一个节点.

Number of processor cores per node 16 数字16表示一个节点上使用16个CPU核心.

Walltime 18 数字18代表18个小时. 如果计算没有在18小时内完成, 会被调度器杀掉.

Max memory per compute node 12 数字12表示, 需要这个节点给12GB内存.

其他项目可以不填写. 会得到一个类似的内容:

```
#!/bin/bash

# Copy/paste this job script into a text file and submit with the command:
#    sbatch thefilename

#SBATCH --time=18:00:00   # walltime limit (HH:MM:SS)
#SBATCH --nodes=1   # number of nodes
#SBATCH --ntasks-per-node=16   # 16 processor core(s) per node
#SBATCH --mem=12G   # maximum memory per node
#SBATCH --job-name="test"

# LOAD MODULES, INSERT CODE, AND RUN YOUR PROGRAMS HERE
```

如果, 没有调度器的时候, 作业运行指令是:

```
module load vasp_mpi
mpirun -np 16 vasp_std
```

那么, 创建一个文件, 比如job01.pbs, 内容如下:

```
#!/bin/bash

# Copy/paste this job script into a text file and submit with the command:
#    sbatch thefilename

#SBATCH --time=18:00:00   # walltime limit (HH:MM:SS)
#SBATCH --nodes=1   # number of nodes
#SBATCH --ntasks-per-node=16   # 16 processor core(s) per node
#SBATCH --mem=12G   # maximum memory per node
#SBATCH --job-name="test"

# LOAD MODULES, INSERT CODE, AND RUN YOUR PROGRAMS HERE

module load vasp_mpi
mpirun -np $SLURM_NTASKS vasp_std
```

一句话, 就是把之前的运行指令, 附在脚本的最后面; 把原来的-np XX中的XX修改为$SLURM_NTASKS, 仅此而已.

如何提交:

```
qsub job01.pbs
或者
sbatch job01.pbs
```

12.3 SLURM调度器内置参数

slurm调度器中更多控制参数, 请看图 14.

13 自定义

本工具套件在安装过程中, 必须严格按照此手册描述之流程操作, 不可变更.

安装完毕后, 默认创建一个队列, workq, 所有的节点都在该队列, 如果用户不通过脚本主动申明, 那么默认一个cpu_core搭配512MB或者1024MB内存.

集群系统部署完毕后, 整个Linux系统依旧属于自由软件范畴, 如果您拥有root权限, 您可以自由变更任何配置文件, 但需承担所有责任.

13.1 声明与警告

请仔细阅读以下内容, 并在进行任何操作之前充分理解.

13.1.1 配置文件说明

1. 本集群系统的所有相关组件均依照各Linux发行版要求进行配置.
2. 相关配置文件均位于默认路径; 源码编译组件(如有)在/opt/hpc4you路径下.
3. 所有配置文件均采用明文存放.
4. 所有*_hpc组件, 由慧计算开发, 采用二进制提供.

13.1.2 配置文件修改

1. Linux系统是开源自由的, 您可以根据自己的喜好修改任何现有配置.
2. 必要的配置文件已采用 chattr 设置了只读权限, 请自行解锁.

SLURM Variables	Torque/MOAB	Description
SLURM_ARRAY_TASK_COUNT		Total number of tasks in a job array
SLURM_ARRAY_TASK_ID	PBS_ARRAYID	Job array ID (index) number
SLURM_ARRAY_TASK_MAX		Job array's maximum ID (index) number
SLURM_ARRAY_TASK_MIN		Job array's minimum ID (index) number
SLURM_ARRAY_TASK_STEP		Job array's index step size
SLURM_ARRAY_JOB_ID	PBS_JOBID	Job array's master job ID number
SLURM_CLUSTER_NAME		Name of the cluster on which the job is executing
SLURM_CPUS_ON_NODE		Number of CPUS on the allocated node
SLURM_CPUS_PER_TASK	PBS_VNODENUM	Number of cpus requested per task. Only set if the **--cpus-per-task** option is specified.
SLURM_JOB_ACCOUNT		Account name associated of the job allocation
SLURM_JOBID **SLURM_JOB_ID**	PBS_JOBID	The ID of the job allocation
SLURM_JOB_CPUS_PER_NODE	PBS_NUM_PPN	Count of processors available to the job on this node.
SLURM_JOB_DEPENDENCY		Set to value of the --dependency option
SLURM_JOB_NAME	PBS_JOBNAME	Name of the job
SLURM_NODELIST **SLURM_JOB_NODELIST**	PBS_NODEFILE	List of nodes allocated to the job
SLURM_NNODES **SLURM_JOB_NUM_NODES**		Total number of different nodes in the job's resource allocation
SLURM_MEM_PER_NODE		Same as **--mem**
SLURM_MEM_PER_CPU		Same as **--mem-per-cpu**
SLURM_NTASKS **SLURM_NPROCS**	PBS_NUM_NODES	Same as **-n, --ntasks**
SLURM_NTASKS_PER_NODE		Number of tasks requested per node. Only set if the **--ntasks-per-node** option is specified.
SLURM_NTASKS_PER_SOCKET		Number of tasks requested per socket. Only set if the **--ntasks-per-socket** option is specified.
SLURM_SUBMIT_DIR	PBS_O_WORKDIR	The directory from which **sbatch** was invoked
SLURM_SUBMIT_HOST	PBS_O_HOST	The hostname of the computer from which **sbatch** was invoked
SLURM_TASK_PID		The process ID of the task being started
SLURMD_NODENAME		Name of the node running the job script
SLURM_JOB_GPUS		GPU IDs allocated to the job (if any).

图 14: slurm内置变量.

13.1.3 责任声明

1. 不正确修改配置文件可能导致集群工作异常或无法工作.
2. 对于因修改配置文件而导致的任何问题, 用户需自行承担全部责任.
3. 变更/修改配置文件后, 慧计算不再提供质保和免费技术支持.

再次提醒: 请仔细阅读以上内容, 并在进行任何操作之前充分理解.

13.2 用户信息

本集群方案中, 用户信息采用Linux系统/etc/passwd文件控制. 推荐使用useradd_hpc来添加用户.

如果坚持使用useradd, 请务必确保在所有机器上执行同样的操作.

13.3 资源调度管理

hpc4you toolkit v2 for Cluster版本提供:

共享队列 所有机器共享同一个队列, 所有用户均可使用该默认队列.

资源共享 所有用户均可访问所有资源.

调度策略 默认采用先来后到原则, 所有用户优先级相同.

进阶版, 专业版, 开启了slurm accounting, 需要管理员借助sacctmgr管理slurm accounting, slurm QoS等. 具体细节, 请查阅`https://slurm.schedmd.com/accounting.html`.

13.4 MySQL数据库

安装版本, 均来自apt/yum/dnf推荐版本. 默认root密码是hpc4you.

在进阶版和专业版中, 调度器以用户名slurm访问数据库slurm_acct_db, 具体配置/etc/slurm/slurmdbd.conf.

13.5 集群名称

默认集群名称是hpc4you, 写在文件/etc/slurm/slurm.conf. 如需修改, 请修改如下必要配置:

- /etc/slurm/slurm.conf
- 执行指令sacctmgr修改集群名称, 具体查阅slurm手册.

13.6 NFS共享

默认, 本集群方案, 通过NFS共享/home和/opt. 采用各Linux发行版中提供的NFS进行配置.

您可以新增挂载点, 但是不可以移除/opt目录的NFS共享.

所有计算节点, NFS挂载通过/etc/fstab控制.

集群部署完毕后, 你可以增加专业存储, 请确保将您的专业存储挂载给所有的机器. 比如使用专业存储系统负载/home分区. 专业存储如何使用, 请咨询你的存储供应商.

14 图片目录

List of Figures