

LSF 作业管理系统使用方法

上海交通大学高性能计算中心

<http://hpc.sjtu.edu.cn>

2014 年 3 月 3 日更新

目录

作业管理系统是高性能计算机的“指挥部”，它接受用户的作业请求，将作业分配到合适的节点上运行，最后将各节点的计算结果汇总给用户。作业管理系统能够提高计算资源的利用率、降低集群维护难度，因此高性能计算系统大都配备了作业管理系统。

IBM Platform LSF 是一个被广泛使用的作业管理系统，具有高吞吐、配置灵活的优点。上海交通大学 π 集群也使用了 LSF 作业管理系统。这份文档将指导您通过 LSF 提交和管理高性能计算作业。

遵循文档的操作规范和反馈方法，将帮助您顺利完成工作。也欢迎大家对文档内容[提出建议](#)，谢谢！

1 查看计算队列**bqueues**

作业队列是一系列可用的计算资源池，不同的队列在软硬件配置上有侧重，适合不同性质的作业。用户可以使用**bqueues**查看 π 集群可用的计算队列：

\$ bqueues											
QUEUE_NAME	PRIO	STATUS	MAX	JL/U	JL/P	JL/H	NJOBS	PEND	RUN	SUSP	
cpu	40	Open:Active	-	-	-	-	4135	0	4135	0	
fat	40	Open:Active	-	-	-	-	32	0	32	0	
gpu	40	Open:Active	-	-	-	-	560	0	560	0	
mic	40	Open:Active	-	-	-	-	0	0	0	0	

π 集群可用的计算队列有四个，分别是**cpu**、**fat**、**gpu**和**mic**。各队列的硬件配置简要说明如下：

- **cpu**: 采用双路 8 核服务器，64GB 内存，共 332 台服务器，合计 5312 个 CPU 核心、约 21TB 内存。这个队列容量大，适合处理大型计算任务。
- **fat**: 采用双路 8 服务器，256GB 内存，共 20 台服务器，合计 320 个 CPU 核心、约 5TB 内存。这个队列适合进行大内存计算。
- **gpu**: 采用双路 8 核服务器，64GB 内存，每节点配备 2 块 NVIDIA K20M 加速卡，共 50 台服务器。合计 800 个 CPU 核心、约 3TB 内存。这个队列适合进行 CUDA 通用 GPU 计算。

- **mic**: 采用双路 8 核服务器, 64GB 内存, 每节点配备 2 块 Intel Xeon Phi 加速卡, 共 5 台服务器。合计 80 核 CPU、约 300GB 内存。这个队列适合执行需要 MIC 加速的程序。

2 作业提交 bsub

bsub 命令用于向 LSF 作业管理系统提交作业请求。**bsub** 可接收的参数很多, 通过指定不同的运行参数, 可以精细地设定作业运行需求。

```
$ bsub -h
```

下面分别介绍 **bsub** 命令调用、提交作业的方法和额外的资源控制参数。

2.1 bsub 调用方法

在命令行中, 用户可以通过如下三种方法使用 **bsub** 命令, 三种方法各有优点。

1. 直接在命令行中输入完整参数;
2. 进入 **bsub** 环境交互提交;
3. 编写作业提交脚本供 **bsub** 处理;

2.1.1 直接输入完整参数

直接输入 **bsub** 完整参数, 可以方便地提交单线程作业。下面这条命令提交了一个需要一个 CPU 核运行的单线程作业:

```
$ bsub -n 1 -q cpu -o job.out ./myprog "-input data.txt"
```

主要参数说明如下:

- **-n** 指定所需的计算核心数。
- **-q** 指定作业运行的队列, 在 π 集群上可用的计算队列有 **cpu**、**fat**、**gpu** 和 **mic**。