# Mellanox Infiniband安装&使用方法总结&问题解决

https://zhuanlan.zhihu.com/p/450151891

## 1. 基础知识

[RDMA技术详解（一）：RDMA概述](https://zhuanlan.zhihu.com/p/55142557)

[RDMA技术详解（二）：RDMA Send Receive操作](https://zhuanlan.zhihu.com/p/55142547)

[深入浅出全面解析RDMA](https://link.zhihu.com/?target=https%3A//tjcug.github.io/blog/2018/06/04/%25E6%25B7%25B1%25E5%2585%25A5%25E6%25B5%2585%25E5%2587%25BA%25E5%2585%25A8%25E9%259D%25A2%25E8%25A7%25A3%25E6%259E%2590RDMA/)

## 2. 安装指南

驱动下载链接:

https://www.mellanox.com/products/infiniband-drivers/linux/mlnx\_ofed

官方安装指导文档:

下载[驱动](https://link.zhihu.com/?target=http%3A//www.mellanox.com/page/mlnx_ofed_eula%3Fmtag%3Dlinux_sw_drivers%26mrequest%3Ddownloads%26mtype%3Dofed%26mver%3DMLNX_OFED-5.5-1.0.3.2%26mname%3DMLNX_OFED_LINUX-5.5-1.0.3.2-ubuntu18.04-x86_64.tgz),以Ubuntu18.04为例：

2.1检查PCIE 是否识别IB卡

lspci | grep -i Mellanox\*

2.2安装驱动

1. tar -zxvf MLNX\_OFED\_LINUX-5.5-1.0.3.2-ubuntu18.04-x86\_64.tgz

2. cd MLNX\_OFED\_LINUX-5.5-1.0.3.2-ubuntu18.04-x86\_64

3. sudo ./mlnxofedinstall --force

4. sudo /etc/init.d/openibd restart

5. sudo /etc/init.d/opensmd restart

2.3检查IB状态

ibstat // ib卡State为 active 并且 Link Layer 为: InfiniBand 则正常，如果Link Layer 为 Ethernet 模式，请见本文第三章节的FAQ

2.4配置临时IP

sudo ifconfig ib0 11.1.1.15 up //ib0为第一块IB卡， ip地址自己定义

2.5测试读带宽

前提: Server端和 Client 网络通常，

ib卡已配置ip地址

- mlx5 代表的iib卡的型号，具体型号根据请根据 ibstat 中显示的为准

- ib\_read\_bw 是ib卡安装包自带的命令

- 其他测试命令如下：

ib\_atomic\_bw

ib\_atomic\_lat

ib\_read\_bw

ib\_read\_lat ib\_send\_bw ib\_send\_lat ib\_write\_bw ib\_write\_lat

1.ib\_read\_bw //server 端执行

2.ib\_read\_bw -d mlx5\_0 -a -F -i 1 11.1.1.15 // ip地址可以是网卡的ip地址,也可以是IB卡的 IP地址

3.等待结果输出

## 3. RDMA 编程参考

[https://github.com/tarickb/the-geek-in-the-corner](https://link.zhihu.com/?target=https%3A//github.com/tarickb/the-geek-in-the-corner)

[https://github.com/jcxue/RDMA-Tutorial](https://link.zhihu.com/?target=https%3A//github.com/jcxue/RDMA-Tutorial)

[RDMA编程：事件通知机制](https://link.zhihu.com/?target=https%3A//www.jianshu.com/p/4d71f1c8e77c)

[RDMA read and write with IB verbs](https://link.zhihu.com/?target=https%3A//thegeekinthecorner.wordpress.com/2010/09/28/rdma-read-and-write-with-ib-verbs/)

[RDMA Aware Networks Programming User Manual](https://link.zhihu.com/?target=https%3A//www.mellanox.com/sites/default/files/related-docs/prod_software/RDMA_Aware_Programming_user_manual.pdf)

## 4. FAQ

4.1切换IB卡模式为 InfiniBand

1. sudo mst start

2. sudo mlxconfig -y -d /dev/mst/mt4119\_pciconf0 set LINK\_TYPE\_P1=1

3. sudo reboot

3. ibstat // 查看修改后的IB卡模式

4.2 查看IB 卡硬件型号信息

sudo mlxvpd -d mlx5\_0 // -d 为 ib hca\_id, 可以通过ibstat中查看

4.3 NUMA 架构下IB卡带宽不稳定解决方法

// server 端执行

$ cat /sys/class/net/ib0/device/numa\_node

$ 3

$ numactl --cpunodebind=3 --membind=3 ib\_read\_bw -d mlx5\_0

// client 端执行

$ cat /sys/class/net/ib0/device/numa\_node

$ 7

$ numactl --cpunodebind=7 --membind=7 ib\_read\_bw -a -F -d mlx5\_0 10.0.0.1 --run\_infinitely