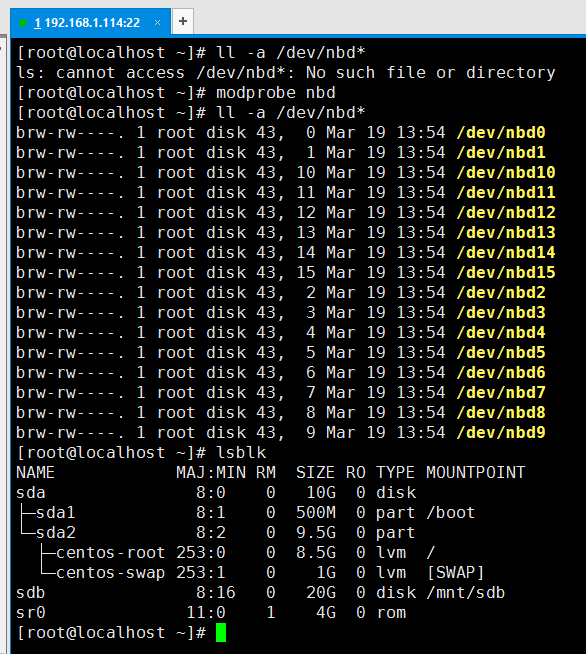
# NBD实践

网络块设备（Network Block Device）提供了一种透明操作远程机器上的块设备的能力，可以像访问本地块设备一样，访问块设备。NBD和网络文件系统（如NFS）不同，文件系统是建立在块设备上，NBD比网络问卷系统更底层，因此可以在上面建立文件系统。NFS只提供一个挂载点供客户端使用，客户端无法改变这个挂载点的分区格式。NBD可以被用作软件RAID使用，提供一种分布式存储。下面对NBD在linux上的使用简单介绍和实践一下，后面对NBD的协议和代码进行解释和走读。

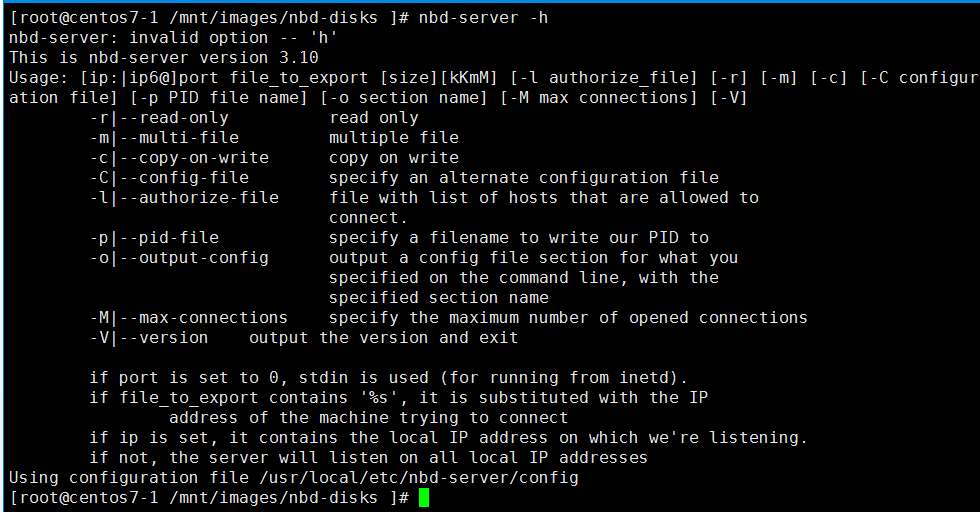
# 安装

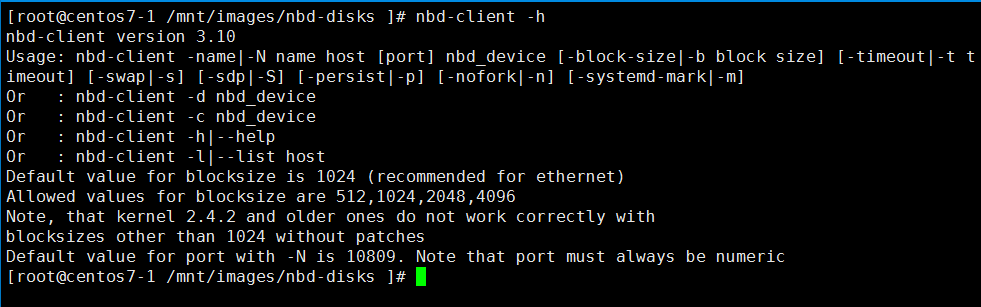
1. 类UNIX系统上，需要一个支持NBD的内核。在编译内核时，从Device Drivers => Block Devices => Network Block Device Support。下面是kvm虚拟化模拟的一台centos 7的虚拟机，加载内核NBD模块后，可以看到NBD预先分配的NBD设备(15个)，可以看到NBD的主设号为43。



2、使用NBD设备时，还需要一个用户态的daemon程序（why ? 请参考后面对NBD协议的分析）。这里我选用nbd-3.10的实现，下载源码后，进行编译和安装（安装步骤自行google）。

安装过后，系统上就有了nbd-server和nbd-client两个daemon程序，用来启动nbd服务端导出nbd设备和启动客户端请求nbd。





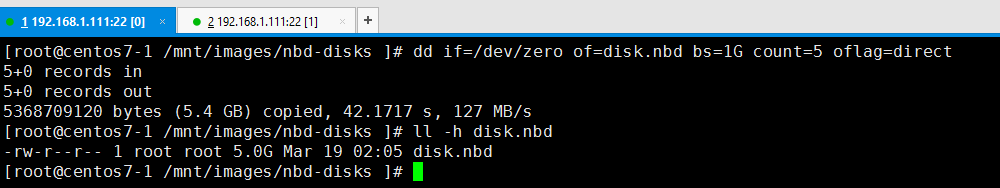
# 小试牛刀

下面，对nbd小试牛刀，做以下的实践：

1. 将一台linux服务器（主机），作为nbd服务端。将主机的一个raw文件导出（export）成为一个网络块设备。
2. 在服务器（主机）测，使用kvm虚拟化，启动一台linux虚拟机，作为nbd客户端，连接并请求nbd服务端的网络块设备。
3. nbd客户端（虚拟机）对远程挂载的nbd块设备进行格式化和挂载。

# 2.1 服务端（主机测）

1. 首先在主机测，创建dd一个5G大小的普通文件



1. 启动nbd服务端

这里我们安装配置文件的方式启动，根据官方配置文档，参考如下：

**启动nbd用户态服务器。**

**（1）配置一个配置文件【可选，也可以直接通过命令行传入参数启动】**

参考配置如下：

#The first section in the configuration file must be called generic, and is used for global options that apply to more than one export.

#This section must always be present, even if it holds no options.

**[generic]**

**user = root**

**group = root**

#Whether to allow the client to fetch a list of exports from this server. If enabled, the client can run nbd-client -l to get a list of exports on this server.

**allowlist = true**

#In versions of nbd-server between 2.9.17 and 3.9.1, when this option was set to true, nbd-server would export all exports on a separate port

#with the old (pre-2.9.17) handshake protocol. In that case, the 'port' option for individual exports was mandatory.

#oldstyle = true

#The port on which to listen for new-style nbd-client connections. If not specified, the IANA-assigned port of 10809 is used.

**port = 10809**

**authfile = /etc/nbd-server/allow**

#Every other section defines one export; the names of these sections are not important, except that you should take care to make sure that each section name is unique.

#

**[export1]**

#The name of the file (or block device) that will be exported.

#This must be a fully-qualified path and filename; relative paths are not allowed.

#If used in conjunction with the temporary, this specifies a template for the temporary file concerned,

#and thus can be used to control the directory it is created in.

#If the file does not exist, but filesize is set, then the file will be created.

**exportname = /mnt/images/nbd-disks/disk.nbd**

# authfile = /etc/nbd-server/allow

#Disable autodetection of file or block device size, and forcibly specify a size. Sizes must be specified in bytes.

#If the multifile option is in effect, this option specifies the size of the entire export, not of individual files.

#If the file is not present, a single file is created of this size.

#filesize =

#gnored, kept for compatibility with the obsolete 'oldstyle' global parameter.

#listenaddr =

#If specified, then it limits the number of opened connections for this export.

#maxconnections =

#[expor]

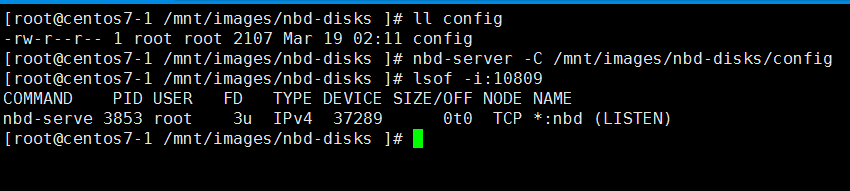
# exportname = /mnt/images/nbd-disks/disk2.nbd

**（2）通过命令行启动nbd-server，输出nbd设备（export a device）**

示例命令行：

**nbd-server -C /mnt/images/nbd-disks/config**

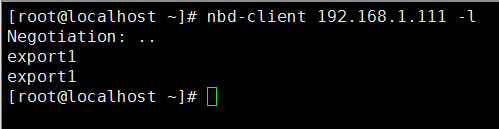
执行完上述步骤，就启动了nbd服务端，并配置导出了一个nbd设备（这里我们将一个普通文件模拟成块设备导出），此处nbd服务端使用了nbd新协议的监听端口号：10809。



# 客户端（虚拟机）

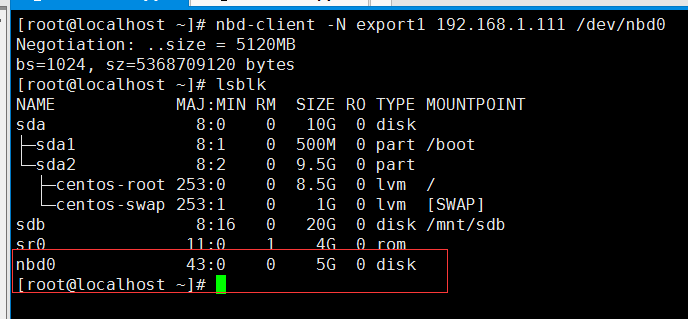
nbd客户端，直接使用nbd-client进行操作：

（1）首先可以先查下一下服务端导出的nbd设备名称（这需要nbd服务端支持该命令，且开放权限。配置了“**allowlist = true**”）。



可以看到，nbd服务端导出设备的名称”export1”，这里输出两遍与nbd daemon程序的实现有关。

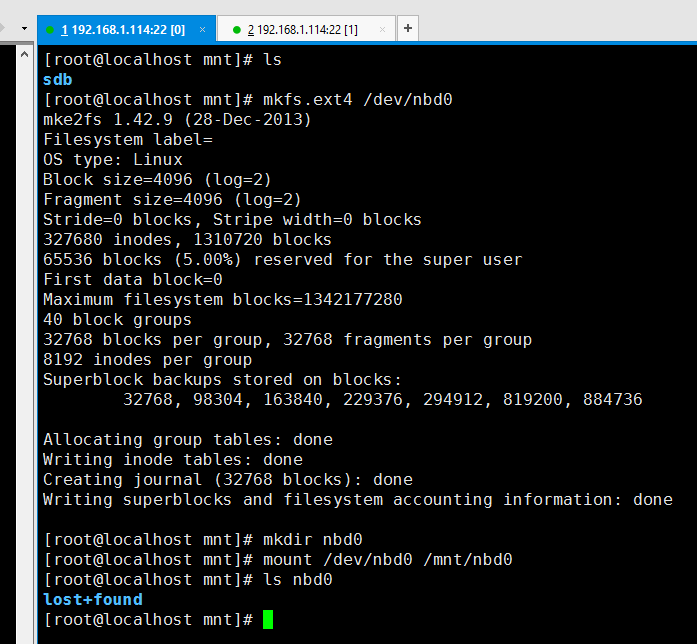
（2）拿到服务端导出设备的名称，就可以连接nbd服务端，请求这个nbd设备，并关联到本地/dev下面的一个nbd设备，如下：



至此，虚拟机测（nbd客户端）与主机侧（nbd服务端）协商完成。lsblk命令可以查看到一个5G 的nbd0设备了。

# 虚拟机使用设备

在虚拟机测对设备进行格式化并mount到目录，如下：



在QEMU虚拟化组件中，提供了类似的用户态工具qemu-nbd，这个工具不仅仅实现了NBD协议的用户态部分，加上QEMU底层镜像格式驱动的支持，也适配了对一定镜像格式文件的读写，功能更加丰富，有兴趣可以自行研究。