

# DIGITS\_DevBox 深度学习服务器

Dai Jialun

August 26, 2015

## 1. 硬件配置

**显卡** 4 个 ASUS (华硕) GTX 980Ti-6GD5

**CPU** 1 个 Intel (英特尔) Core i7-5960X

**主板** 1 个 ASUS (华硕) X99-E WS

**内存** 2 个 CORSAIR (海盗船) VENGERNCE (复仇者)LPX 32GB (4 × 8GB) DDR4 2400MHz  
CMK32GX4M4A2400C14R

**硬盘** 3 个 WesternDigital (西部数码) 4TB 7200 转

**固态硬盘** 1 个 Samsung (三星) SSD 850pro 512GB

**固态硬盘** 1 个 Samsung SSD 512GB SM951 cache for RAID

**机箱** 1 个 CORSAIR (海盗船) 900D

**电源** 1 个 CORSAIR (海盗船) AX1500i 1500W

**散热器** 1 个 CORSAIR (海盗船) H110 水冷 CPU 散热器

**风扇** 6 个 CORSAIR (海盗船) AF120 静音版双包装

**光驱** 1 个 AUSU (华硕) DRW-24D1ST

**配件** 1 个 Thermaltake Commander FT 触控式面板风扇控制器, Deepcool FAN HUB (九州风神风扇集线器)

**显示器**

**键盘鼠标**

## 2. 名词解释

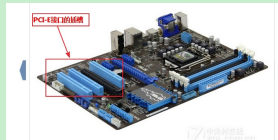
**DVI** Digital Visual Interface, 数字视频接口



**DisplayPort** 高清数字显示接口标准



**PCI-E** PCI Express, 新的总线接口



**SATA Revision 3.0** Serial Advanced Technology Attachment, 串行 ATA 规格第三版,6Gbps



**SATA Express** SATA 3.0 下一代的 SATA 接口, 10Gbps

**M.2** 一种替代 MSATA 新的接口规范, 优势体现在速度和体积。支持 Socket2 和 Socket3 两种接口类型

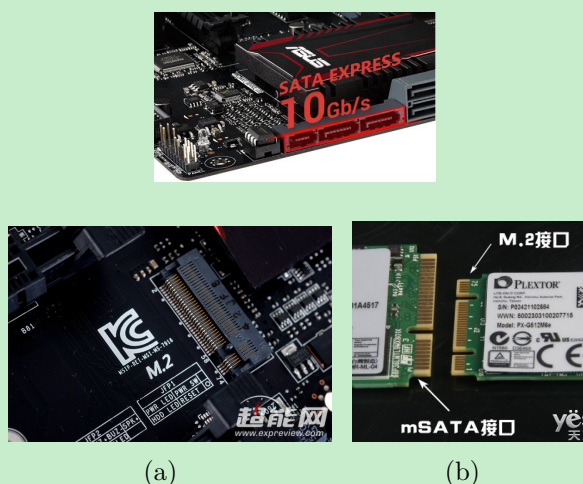


Figure 1:

**RAID** Redundant Arrays of Independent Disks, 磁盘阵列。磁盘阵列是由很多价格较便宜的磁盘，组合成一个容量巨大的磁盘组，利用个别磁盘提供数据所产生加成效果提升整个磁盘系统效能。利用这项技术，将数据切割成许多区段，分别存放在各个硬盘上。

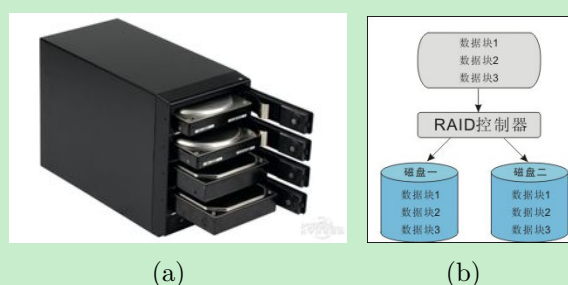
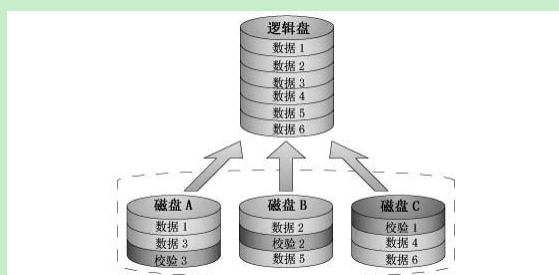
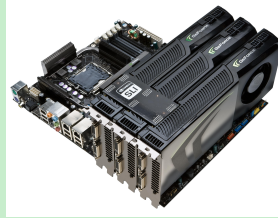


Figure 2:

**RAID5** 一种存储性能、数据安全和存储成本兼顾的存储解决方案。为系统提供数据安全保障，但保障程度要比 Mirror 低而磁盘空间利用率要比 Mirror 高。数据以块为单位分布到各个硬盘上。RAID 5 不对数据进行备份，而是把数据和与其相对应的奇偶校验信息存储到组成 RAID5 的各个磁盘上，并且奇偶校验信息和相对应的数据分别存储于不同的磁盘上。当 RAID5 的一个磁盘数据损坏后，利用剩下的数据和相应的奇偶校验信息去恢复被损坏的数据。



**SLI** Scalable Link Interface, 可灵活伸缩的连接接口 (支持多显卡技术)。这是一种可把两张或以上的显卡连在一起, 作单一输出使用的技术, 从而达至绘图处理效能加强的效果。



**DDR4** Dual Data Rate SDRAM, 是一种高速 CMOS 动态随即访问的内存。DDR4 支持 2133MHz, 32GB DDR4-2133 达到 48.4GB/s。

**GDDR5** Graphics Double Data Rate SDRAM version5, 是一种高性能显卡用内存, 需搭配支持 PCI-E 以上规格的显卡, 高频率达 4GHZ, 低功耗。

### 3. 软件配置名词

**UEFI** Unified Extensible Firmware Interface, 统一的可扩展固件接口, 是一种详细描述类型接口的标准。这种接口用于操作系统自动从预启动的操作环境, 加载到一种操作系统上。

**BIOS** Basic Input/Output System, 基本输入/输出系统。

**固件** Firmware, 固定软件 (自己理解), 写入 EROM 或 EEPROM 中的程序。固件担任着一个系统最基础最底层工作的软件。初期, 这些硬件内所保存的程序是无法被用户直接读出或修改的, 如今这些是可以重复刷写的, 让固件得以修改和升级。

## 4. 环境配置

### 4.1 显卡驱动安装

#### 4.1.1 驱动来源

- 开源驱动 nouveau (livecd 安装时用的驱动)
- 源 (受限制驱动列表)
- PPA 源 (一般是私人建的, 方便群众用)
- 自己下载编译的驱动 (我们使用的方法)

## 4.2 安装 NVIDIA 显卡驱动

1. 受限制驱动列表（源）`sudo apt-get install nvidia-current nvidia-settings`
2. 编译驱动
  - (a) 下载驱动 Nvidia 中文官网是 <http://www.nvidia.cn/page/home.html>
  - (b) 将下载的 NVIDIA-Linux-x86-185.18.14-pkg1.run 驱动文件, 放到 /home/用户名/ 目录下。
  - (c) 编译依赖, `sudo apt-get install build-essential pkg-config xserver-xorg-dev linux-headers-$(uname -r)`
3. 屏蔽开源驱动 nouveau
  - blacklist（推荐）
    - (a) 打开终端, 输入 `sudo vim /etc/modprobe.d/blacklist.conf`
    - (b) 添加 blacklist nouveau
  - grub2
    - (a) 打开终端, 输入 `sudo vim /etc/modprobe.d/blacklist.conf`
    - (b) 修改 `GRUB_CMDLINE_LINUX=""` 为 `GRUB_CMDLINE_LINUX="nomodeset"`
    - (c) 输入 `sudo update-grub`
4. 安装装备
  - (a) 清除之前与 nvidia 相关的驱动程序, `sudo apt-get -purge remove nvidia-*`
  - (b) 编译依赖, `sudo apt-get install build-essential pkg-config xserver-xorg-dev linux-headers-$(uname -r)`
  - (c) 切换到虚拟终端 tty1, `ctl+alt+F1` (如果不屏蔽 nouveau, 可能会出现黑屏现象); 黑屏则 `sudo reboot`, 然后重启后, 按下 `Ese` 或者选择 `low-quality`, 进入 tty1, 进行驱动的安装
5. 注销系统, 关闭图形环境 `sudo stop lightdm` (Ubuntu15.04 下, 运行 `sudo systemctl stop lightdm`)
6. 安装过程
  - (a) 在驱动文件目录下, `sudo ./NVIDIA*.run`
7. 启动图形环境, `sudo start lightdm`

## 4.3 创建 RAID5

### 4.3.1 RAID 的优点

- 可高效恢复磁盘
- 增强了速度
- 扩容了存储能力

### 4.3.2 RAID 的分类

**硬 RAID** hardware raid。通过用硬件来实现 RAID 功能的就是硬 RAID，比如：各种 RAID 卡，还有主板集成能够做的 RAID 都是硬 RAID。全硬的 RAID 则全面具备了自己的 RAID 控制/处理与 I/O 处理芯片，甚至还有阵列缓冲（Array Buffer），对 CPU 的占用率以及整体性能是这三种类型中最优势的，但设备成本也是三种类型中最高的。hardRAID 自成一个单元，由自己的 firmware 硬件和软件，与主板和操作系统无关，即 Ubuntu 不需要额外的程序来管理。

**软 RAID** software raid。通过用操作系统来完成 RAID 功能的就是软 RAID，比如：在 Linux 操作系统下，用 3 块硬盘做的 RAID5。全软 RAID 就是指 RAID 的所有功能都是操作系统（OS）与 CPU 来完成，没有第三方的控制/处理（业界称其为 RAID 协处理器——RAID Co-Processor）与 I/O 芯片。这样，有关 RAID 的所有任务的处理都由 CPU 来完成，可想而知这是效率最低的一种 RAID。

**主板 RAID** 通过主板内建 raid 控制器创建阵列，由操作系统驱动识别。这个在 Intel Desktop 的主板上表现的比较明显。主要缺乏自己的 I/O 处理芯片，所以这方面的工作仍要由 CPU 与驱动程序来完成。而且，半软半硬 RAID 所采用的 RAID 控制/处理芯片的能力一般都比较弱，不能支持高的 RAID 等级。FakeRaid 又称 BiosRaid，是由主板的 Bios 程序组成，与 Ubuntu 系统无关。但是 Ubuntu 提供 dmraid 命令与 BIOS 进行沟通，由 FakeRAID 帮助管理。基本上，ubuntu 也把 FakeRAID 当成单一硬盘使用。

### 4.3.3 主板集成 RAID 与外插 RAID 卡区别

**性能** 主板集成的 RAID，它的性能以及它的速度是通过主板的 CPU 与内存来实现的，它会占有主板一定的带宽，会影响整机的性能；外插 RAID 卡，它本身由自己的 CPU 和内存，所以它的数据处理大部分都会由自己处理，不会影响主板上的 CPU 与内存速度，总体看来，外插的 RAID 卡的 RAID 要比主板集成的 RAID 快得多。格式化、挂载、写入与重建全部由 mdadm 负责。



**安全性** 主板集成的 RAID 它的安全性不能够得到保证，比如：我们用 P8SCT 主板做一个 SATA RAID，不论你做 RAID 几，它是通过更改主板的 BIOS 选项做成的，所以一旦主板损坏、主板的 CMOS 电池掉电、无意更改了主板 BIOS 的设置都会带来 RAID 的丢失，通过主板做成的 RAID，一旦丢失，将会不能恢复，后果是非常严重；而外插的 RAID 卡做成的 RAID 就不会因为主板损坏、主板的 CMOS 电池掉电等现象对数据造成影响，所以外插的 RAID 卡，它的安全性远远大于主板集成的。SoftRaid 与主板 BIOS 程序无关，完全由 Ubuntu 的 mdadm 命令管理。

#### 4.3.4 创建 RAID5 步骤 (主板 BISO)

1. 对各个磁盘删除分区，且进行格式化。大分区硬盘使用 GPT 分区，MRB 分区最大只支持 2TB
2. 创建，create
3. 分配，
4. 格式化 `sudo mkfs.ext4 /dev/sdb`
5. 挂载 `sudo /dev/mapper/isw_dfafd_Volume1 /deep`
6. 自动挂载 `sudo vim /etc/fstab`

可能出现问题

- 在装系统过程中，在选择系统分区时，`ctrl+alt+F1`，输入 `dmraid -ay` 激活 RAID (Ubuntu server) 这种方式其实还是属于介于软 raid 与硬 raid 之间。在启动时候，由硬件 raid 驱动，当载入 linux 内核之后，由 linux 接手管理，还是会消耗 cpu 等资源。与更传统的软件 raid 比较，就是启动的时候 (linux 内核未介入之前) 系统看到的仍然是一个 raid 的虚拟硬盘，所以两块硬盘完全一样，要恢复重建之类的更加简单。另外的话，dmraid 映射了底层的硬件 raid 驱硬件控制器，raid 控制可能也能帮助处理一些操作，可能对性能也会有一定提高。
- dmraid 将硬件的 raid 映射成 `/dev/mapper/` 下面的设备，例如 `/dev/mapper/isw_dfadcda_Volume1`，其中 isw 为 intel 的硬件名字，Volume1 为 RAID 名称。Ubuntu 在安装过程中，已经将 RAID 显示为 `/dev/mapper/isw_dfadcda_Volume1`，已经显示正确了，只不过容量出现问题，理论上的容量应该为 7.2TB，但是实际情况只有 3.6TB (至今未解开)
- dmraid 可参考 <http://www.cnblogs.com/linuxer/archive/2012/03/07/2441224.html> <http://book.51cto.com/art/200902/110754.htm>

- Ubuntu 的软 RAID 相关命令为 mdadm, 其配置、测试、删除参考 <http://blog.itpub.net/27771627/viewspace-1246416/>
- 目前使用的 RAID 为主板的 Intel Rapid Storage Technology, 目前驱动只支持 Window, 对 Windows 的兼容性不好, 为 fakeraid
- fake raid 仅提供廉价的控制器, raid 处理开销仍由 CPU 负责, 因此性能与 CPU 占用基本与 software raid 持平。如果只有单个 linux 系统, 使用 software raid 一般比 fake raid 更健壮, 但是, 在多启动环境中 (例如 windows 与 linux 双系统), 为了使各个系统都能正确操作相同的 raid 分区, 就必须使用 fake raid 了。