

知 A10 GPU卡直通给虚拟机CenOS7.6使用，安装GPU驱动报错

GPU 金占勇 2024-05-31 发表

组网及说明

R4900 G3服务器安装物理机，然后起虚拟机CentOS7.6

告警信息

1.安装525驱动从报错来说有一直反馈Unable to load the kernel modul 'nvidia.ko'

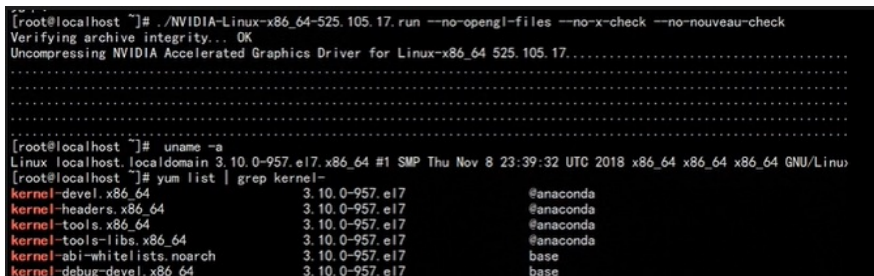


问题描述

某局点A10 GPU卡直通给虚拟机CenOS7.6使用，安装GPU驱动报错

1.客户手上有2个驱动版本，一个4xx版本一个550版本，前者安装报错，后者安装没有问题，但是安装后nsmi无回显

2.uname -a确认当前内核版本是3.10.0-957，标准7.6内核，cc等其他文件版本与内核完全一样



3.按照525驱动提供的版本说明书以驱动安装指南操作，按照步骤均是正确的，但问题依旧存在 <https://zhiliaoh3c.com/Theme/details/222914>

过程分析

4.错误一直指向Unable to load the kernel modul 'nvidia.ko'，和虚拟机系统强关系

5.查询vm官网关于gpu直通一些设置，比对客户环境，目前根据以下文章

<https://docs.vmware.com/en/VMware-Edge-Compute-Stack/1.0/ecs-enterprise-edge-ref-arc/h/GUID-412AD9B3-6B9B-4BE0-B833-9205ACBCF956.html>

发现与客户环境差别是虚拟机设置时BIOS引导为EFI模式，而客户是默认的BIOS

Create the virtual machine that will be using the GPU and configure EFI Boot Mode under Boot Options. The default boot option is BIOS.

Configure EFI boot mode

▼ Boot Options	
Firmware	EFI
Secure Boot	<input type="checkbox"/> Enabled
Boot Delay	When powering on or resetting, delay boot order by 0 milliseconds
Force EFI setup	<input type="checkbox"/> During the next boot, force entry into the EFI setup screen
Failed Boot Recovery	<input type="checkbox"/> If the VM fails to find boot device, automatically retry after 10 seconds

从已有的报错看是系统的前置条件有问题，与客户沟通建立个新的虚拟机，全包安装系统，拉黑相关驱动、init3后再安装GPU驱动

解决方法

虚拟机设置层面Firmware应该选为EFI模式而非Legacy导致，修改EFI后安装成功