



Weisong Shi <weisong.shi@gmail.com>

异步io实现讨论

1 message

褚霸 <chuba@taobao.com>

Thu, Jun 7, 2012 at 6:25 AM

To: "weisong.shi@gmail.com" <weisong.shi@gmail.com>

Hi教授,

在《《Foundations of Computer Systems Research》》一书的第12章第三节的async io中，教授介绍了aio_系列函数的工作原理和演示程序，很好的为我们演示了异步io的原理和使用流程。

但是可能由于历史的原因，不同的平台aio据我所知有3-5种不同的实现。在linux下aio_*系列是由glibc实现的，它是用多线程来模拟的，简单的说就是在aio_read/write的时候，让背景线程用pread/pwrite来发起真正的读写操作，在IO完成后，让背景线程通知发起线程操作完成的情况。所以它不是真正的aio。

Linux下真正的aio是oracle的人贡献到社区的，它的函数是iosetup, io_destroy(2), io_getevents(2), io_submit(2), io_cancel(2)系列，透过libaio导出的,使用的时候包含libaio.h就好。这个库是使用真正的系统调用，由操作系统的fs/aio.c完成异步操作的。

测试工具fio可以用到libaio，Mysql的innodb引擎也是用的这个aio,我们平常在项目中包括OB，图像搜索中也用这个库来提升IO的能力，降低延迟。

具体参考：

1. <http://www.ibm.com/developerworks/cn/linux/l-async/>
2. <http://blog.yufeng.info/archives/tag/libaio>
3. <http://lse.sourceforge.net/io/aio.html>
4. <http://lxr.linux.no/#linux+v3.4.1/fs/aio.c>
5. http://www.koders.com/c/fidC99B1C2D010B48C5EDB8D4FA728097A29E5FBCBC.aspx?s=my*+-mysql
6. `git clone git://git.kernel.dk/fio.git`

不当之处，请教授指点，谢谢！

褚霸

褚霸 简单就是美！ <http://blog.yufeng.info>

This email (including any attachments) is confidential and may be legally privileged. If you received this email in error, please delete it immediately and do not copy it or use it for any purpose or disclose its contents to any other person. Thank you.

本电邮(包括任何附件)可能含有机密资料并受法律保护。如您不是正确的收件人，请您立即删除本邮件。请不要将本电邮进行复制并用作任何其他用途、或透露本邮件之内容。谢谢。