

(/apps/redi
utm_sourc
banner-clic

PostgreSQL持久性优化机制——WAL



ZYJ2016 (/u/d1ef74921117) + 关注

2017.01.10 16:26* 字数 1159 阅读 2115 评论 0 喜欢 1

(/u/d1ef74921117)

关于持久性

持久性是指，事务提交后，对系统的影响必须是永久的，即使系统意外宕机，也必须确保事务提交时的修改已真正永久写入到永久存储中。

最简单的实现方法，当然是在事务提交后立即刷新事务修改后的数据到磁盘。但是磁盘和内存之间的IO操作是最影响数据库系统影响时间的，一有事务提交就去刷新磁盘，会对数据库性能产生不好影响。

WAL机制的引入，即保证了事务持久性和数据完整性，又尽量地避免了频繁IO对性能的影响。

(https://dsp
click.youda
slot=30edc
6d56-45b3
a947b26df
39920776

WAL过程分析

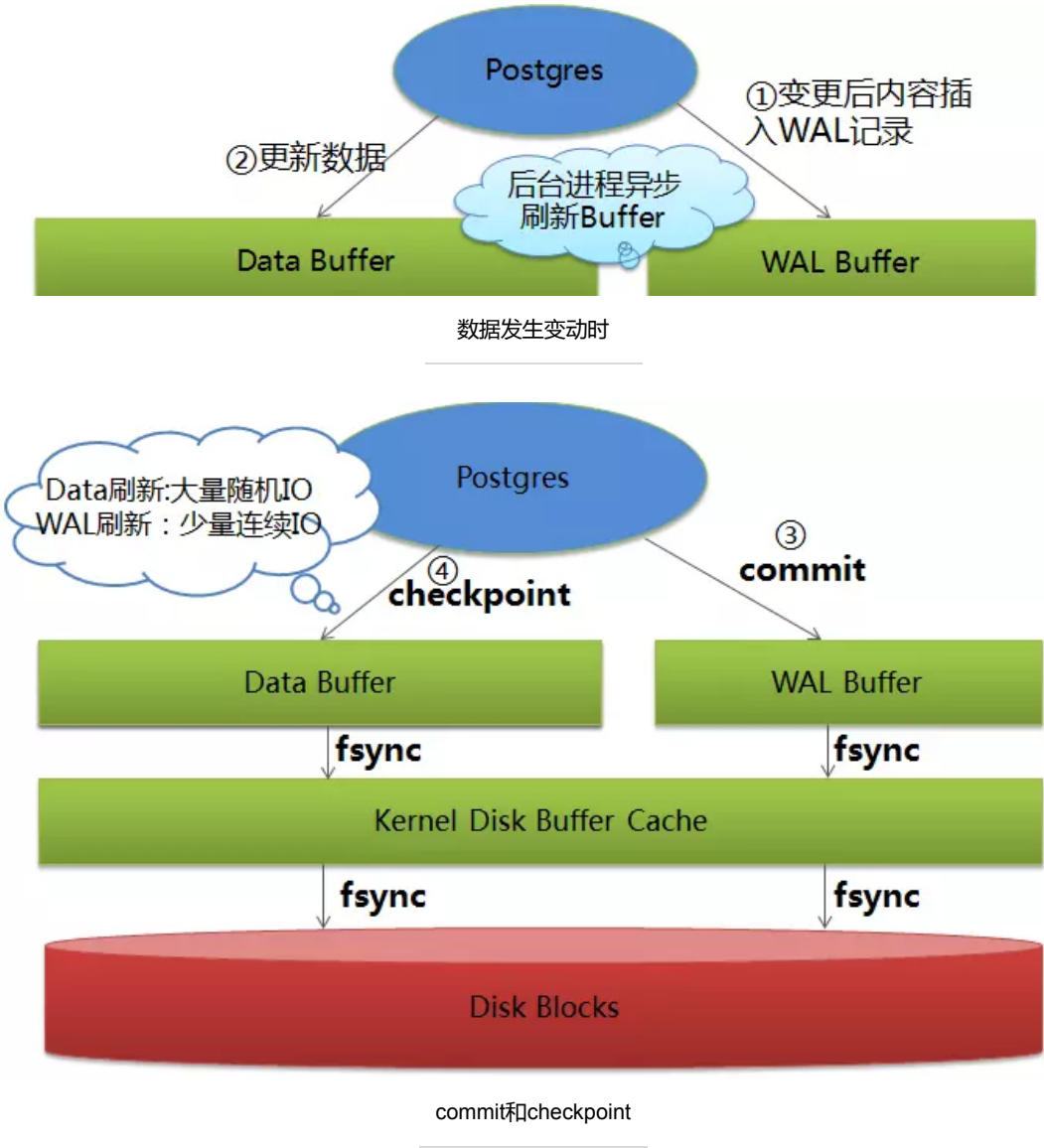
Write-Ahead Logging，前写日志。

在MVCC的部分中，我们已经分析了PostgreSQL的存储结构：元组-文件页-物理段-表；以及写数据的步骤：先写到缓冲区Buffer-再刷新到磁盘Disk。

WAL机制实际是在这个写数据的过程中加入了对应的写wal log的过程，步骤一样是先到Buffer，再刷新到Disk。

- Change发生时：
- 先将变更后内容记入WAL Buffer
- 再将更新后的数据写入Data Buffer
- Commit发生时：
- WAL Buffer刷新到Disk
- Data Buffer写磁盘推迟
- Checkpoint发生时：
- 将所有Data Buffer刷新到磁盘





(/apps/redi
utm_sourc
banner-clic

(https://dsp
click.youde
slot=30edc
6d56-45b3
a947b26df
39920776

WAL的好处

通过上面的分析，可以看到：

当宕机发生时，

- Data Buffer的内容还没有全部写入到永久存储中，数据丢失；
- 但是WAL Buffer的内容已写入磁盘，根据WAL日志的内容，可以恢复库丢失的内容。

在提交时，仅把WAL刷新到了磁盘，而不是Data刷新：

- 从IO次数来说，WAL刷新是少量IO，Data刷新是大量IO，WAL刷新次数少得多；
- 从IO花销来说，WAL刷新是连续IO，Data刷新是随机IO，WAL刷新花销小得多。

因此WAL机制在保证事务持久性和数据完整性的同时，成功地提升了系统性能。

附：PostgreSQL官网文档关于WAL的翻译

官网翻译



Write-Ahead Logging是一种保证数据完整性的标准方法。简单地说，WAL的概念就是对数据文件的改变（包括表和索引）必须先写入日志，即日志记录刷新到永久储存之后，才能被写。遵循这个过程，就不需要在每个事务提交时都刷新数据页到磁盘，因为在宕机时可以用日志来恢复数据库：任何没有应用到数据页上的改动都可以根据日志记录重做。（即回滚恢复REDO）

(/apps/redi
utm_sourc
banner-clc

因为WAL可以在宕机后恢复数据库文件内容，JFS（journaled file systems）对于数据文件或WAL文件的可靠存储就不是必要的了。实际上，JFS甚至会影响系统性能，尤其当它要把文件系统数据刷新到磁盘的时候。还好，JFS中的数据刷新可以在文件系统挂载选项中禁用。JFS确实提高了宕机后的root速度。

使用WAL可以显著地减少写磁盘的次数，因为只需要把日志文件刷新到磁盘就可以保证事务被提交，而不需要把事务改动过的每一个数据文件都刷新到磁盘。日志文件是连续写的，所以同步log的花销远小于刷新数据页的花销。特别是服务器要处理涉及数据存储不同部分的大量小事务时更是这样。另外，当服务器在处理大量并行小事务时，log文件一次fsync就可以提交多个事务。

(https://dsp
click.youda
slot=30edc
6d56-45b3
a947b26df
39920776

WAL还使得在线备份和时间点恢复成为可能。通过归档WAL数据，我们可以恢复到WAL数据覆盖范围内的任何时间点：只需install一份数据库的物理备份，并恢复WAL日志到所需时间即可。更重要的是，这个物理备份并不必须是一个数据库状态的瞬时快照——如果一段时间的快照，那把WAL日志也恢复成那一段时间的即可。

小礼物走一走，来简书关注我

赞赏支持

postgres (/nb/8832913) 举报文章 © 著作权归作者所有



ZYJ2016 (/u/d1ef74921117)

写了 39525 字，被 53 人关注，获得了 51 个喜欢

(/u/d1ef74921117)

+ 关注


中科院研究生毕业，关注大数据架构，分布式系统。百度入职新人。

喜欢 | 1




更多分享





下载简书 App ▶

随时随地发现和创作内容



(/apps/redirect?utm_source=note-bottom-click)

(/apps/redi
utm_sourc
banner-lic



登录 (/sign_in?utm_source=desktop&utm_medium=not-signed-in-comr

评论

智慧如你，不想发表一点想法 (/sign_in?utm_source=desktop&utm_medium=not-signed-in-nocomments-text)咩~

(https://dsp
click.youda
slot=30edc
6d56-45b3
a947b26df
39920776

《MySQL技术内幕：InnoDB存储引擎(第2版)》书摘 (/p/3eca0b18cf51?ut...

MySQL技术内幕：InnoDB存储引擎(第2版) 姜承尧 第1章 MySQL体系结构和存储引擎 >> 在上述例子中使用了mysqld_safe命令来启动数据库，当然启动MySQL实例的方法还有很多，在各种平台下的方式可能会又...



沉默剑士 (/u/6ca93a173ea2?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendatio

(/p/60f03b16e7ff?

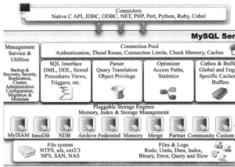


图 1-1 MySQL 体系结构

utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendatio
MySQL InnoDB 技术内幕 (/p/60f03b16e7ff?utm_campaign=maleskine&...

第一章 MySQL 体系架构和存储引擎 mysql是数据库也是数据库实例 mysql 是一个单进程多线程架构的数据
库 daemon 守护进程 当启动实例时，MySQL数据库会去读取配置文件，根据配置文件的参数来启动数据...



snail_knight (/u/b418a3410f2d?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendatio


(/p/f053e70f9b18?



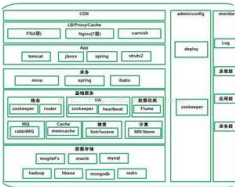
utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendatio
MongoDB新存储引擎WiredTiger实现(事务篇) (/p/f053e70f9b18?utm_ca...

导语：计算机硬件在飞速发展，数据规模在急速膨胀，但是数据库仍然使用是十年以前的架构体系，

WiredTiger 尝试打破这一切，充分利用多核与大内存时代，开发一种真正满足未来大数据管理所需的数据...


 isgiker (/u/162bb3e4453d?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendatio (/apps/redi
utm_sourc
banner-clic
(/p/f65886eb269f?

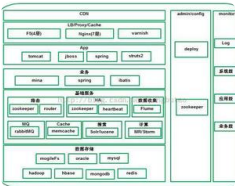


utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendatio
构建高并发高可用的电商平台架构实践 (/p/f65886eb269f?utm_campaign=...

原文出处：杨步涛的博客 一、设计理念 1. 空间换时间1) 多级缓存，静态化客户端页面缓存 (http header中
包含Expires/Cache of Control，last modified(304，server不返回body，客户端可以继续用cache，减少流...

 CookieziSui (/u/0b6d3c2f4836?


utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendatio
(/p/09a29507fd4e?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendatio
构建高并发高可用的电商平台架构实践 (/p/09a29507fd4e?utm_campaign=...

问题导读： 1.如何构建高并发电商平台架构 2.哈希、B树、倒排、bitmap的作用是什么？ 3.作为软件工程
师，该如何实现读写？ 4.如何实现负载均衡、反向代理？ 5.电商业务是什么？ 6.基础中间件该如何设计？ ...

(https://dsp
click.youda
recommendatio
slot=30edc
6d56-45b3
a947b26df
39920776%


 MaLiang (/u/e83ce3312e83?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendatio
(/p/15bd335e6aa0?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendatio
在西安~哪凉快哪呆着去 (/p/15bd335e6aa0?utm_campaign=maleskine&...


躲进冰箱？太年轻了！ 躺进浴缸？呵呵，傻白甜！ 空调制冷？浪费电不？ 避暑新姿势，还在费尽心思想
么？ 躲避酷热高温，还在想方设法么？ 哼！是时候来一发新的姿势了！ 防空洞！怕不怕？ 是的，你没想...

 小端纸 (/u/8c07ea1a7b25?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendatio

怎么老是海星？ (/p/37143efe7005?utm_campaign=maleskine&utm_con...

最近海星出现地越来越频繁了。我曾经换过了很多设备来测试如何摇一摇最轻松，最后还是选中了iPad：
iPad摇起来最累，但是输入太方便了，速度快，能同时HOLD住多个妹子，还能同时摇一摇，这样一旦有...

 66分先生 (/u/2c1e5ab5e534?


utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendatio
(/p/a76c66c7fd84?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendatio

春花毒舌 | 郭德纲不回骂是认怂了？没准憋大招呢 (/p/...

昨天春花 in 节目最后问：你遭受过道德绑架吗？可能这个问题有点难啊，大家都在说别的，只有赵丹很乖的在下方留言回答了问题，他说：道德绑架不就是圣...

 今日排行榜 (/u/5505c7bd95ca?



(/apps/redi
utm_sourc
recommendatio
banner-clic

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendatio


(/p/0ff2e1fc9f7a?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendatio

中国最美稻城亚丁..... (/p/0ff2e1fc9f7a?utm_campaign=maleskine&utm_...


四川的美景似乎总是数不完，看不厌！▽一条九寨沟，勾住了多少人的灵魂 一个泸沽湖，缭绕憧憬着多少人的梦中仙境 一个若尔盖，策马奔腾了多少人心中的那匹野马 一座峨眉山，惊叹了多少登山人的双眸 一村...

 东西风尚 (/u/20b005adb855?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendatio

20170113 周五 多云 (/p/e849d03efde2?utm_campaign=maleskine&ut...

“生命中的任何事物都有保鲜期，那些美好了愿望如果只能珍重的供奉在理想的桌面上，那么只能让它在岁月里积满灰尘。” 每日一文的文章不能复制，一开始觉得不方便，但转而又认为这也挺好，毕竟对于喜欢的文...

 夜阑风静 (u/67d5a28eeb11?

(https://dsp
click.youde
slot=30edc
6d56-45b3
a947b26df
39920776%

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendatio

