

前言

PostgreSQL 15 如约和各位正式见面啦！包括改进的排序性能，流行的 MERGE 命令，更多的压缩算法支持，以及更多用于观察/调整数据库状态的功能，PGer又可以过年了。随着 PostgreSQL 热度的不断提升，市场上岗位也在随之增多，相较于笔者 18 年辞职的时候，那简直是一个天上一个地下，赶着金九银十的尾巴，在此也整理一些常见的面试题（以我的角度，我会关心的点，可以了解一个人对 PostgreSQL 掌握程度），让各位能在 PG 的面试中如鱼得水，一起头脑风暴一下吧！

- MVCC 实现机制以及和 Oracle 的差异
- 为什么会有表膨胀及表膨胀的危害
- 长事务的危害以及如何溯源长事务
- 子事务的危害和注意事项
- 表结构变更哪些操作是非 online 的
- 物理备份需要注意什么 (pg_start_backup)
- 逻辑备份是如何确保一致性的
- WAL 堆积的原因有哪些
- 长连接的危害是什么
- infomask 标志位的作用是什么
- 空值是如何存储的以及索引是否存储空值
- 为什么需要有全页写 (full_page_write)
- 索引失效的各种原因
- commit log 的作用
- 数据库的连接方式以及各自适用的场景
- 各种索引的适用场景 (HASH/GIN/BTREE/GIST/BLOOM/BRIN)
- 行锁是如何实现的，行锁是否会存储在共享内存中
- 流复制和逻辑复制的区别以及各自适用的场景
- 流复制冲突是什么以及为什么会产生复制冲突
- 简述 PostgreSQL 中的权限体系
- 常见的高可用方案以及高可用选型及优缺点
- synchronous_commit 五种级别的区别，为什么备库的查询不能立马看到主库插入的数据
- 事务 ID 回卷的原因以及如何维护优化
- vacuum / autovacuum 的作用以及如何调优
- 函数三态以及函数为什么需要有 execute
- 为什么要使用 create index concurrently 以及 CIC 的危害
- HOT 原理
- PostgreSQL 中是否有锁升级
- 复制槽的作用以及复制槽的危害
- 为什么会有死锁以及死锁检测机制
- SQL 慢能从哪些方面入手排查
- 为什么需要使用分区表以及分区表的优势和劣势
- 软硬解析的概念
- vm / fsm / init 文件是什么

这些是关于数据库自身的一些比较典型的问题和原理，需要掌握。另外由于数据库和操作系统息息相关，关于操作系统相关的知识也得多多了解，比如

- 内存回收机制：kswapd/direct memory reclaim/pdflush

- 进程调度，D进程的危害和形成原因
- 抓包解包，分析 PostgreSQL 协议
- 存储，SAN/NAS/DAS
- 一条 IO 请求的生命周期
- ...

后面一期我再抽空将这些问题解答一下。

xiongcc@PostgreSQL学徒