壹钱包

浅谈PostgreSQL高可用架构

余星
平安壹钱包

PostgreSQL

Agenda

- 关于PG
- 高可用的目标
- 常用的高可用架构
- 我们的高可用方案
- PG 在壹钱包的使用

关于PG

- PostgreSQL是以加州大学伯克利分校计算机系开发的 POSTGRES，现在已经更名为PostgreSQL.最新版本为9.6
- The world’s most advanced open source database
- PG的分布式架构Greenplum大规模运用在OLAP环境

高可用的目标

- 业界常用 N 个9 来量化可用性SLA， 最常说的就是类似 “4个9(也就是99.99%)” 的可用性。

描述	通俗叫法	可用性级别	年度系统不可用时间
基本可用性	2个9	0.99	87.6小时
较高可用性	3个9	0.999	8.8小时
具有故障自动恢复能力的可用性	4个9	0.9999	53分钟
极高可用性	5个9	0.99999	5分钟

“4个9” downtime：0.0001×365×24×60=53 min

影响系统高可用的主要因素

- Unplanned Downtime
  - **System Faults**
  - **Data and Media Errors**
  - Site Outages
- Planned Downtime
  - Routine Operations
  - Periodic Maintenance
  - Upgrades

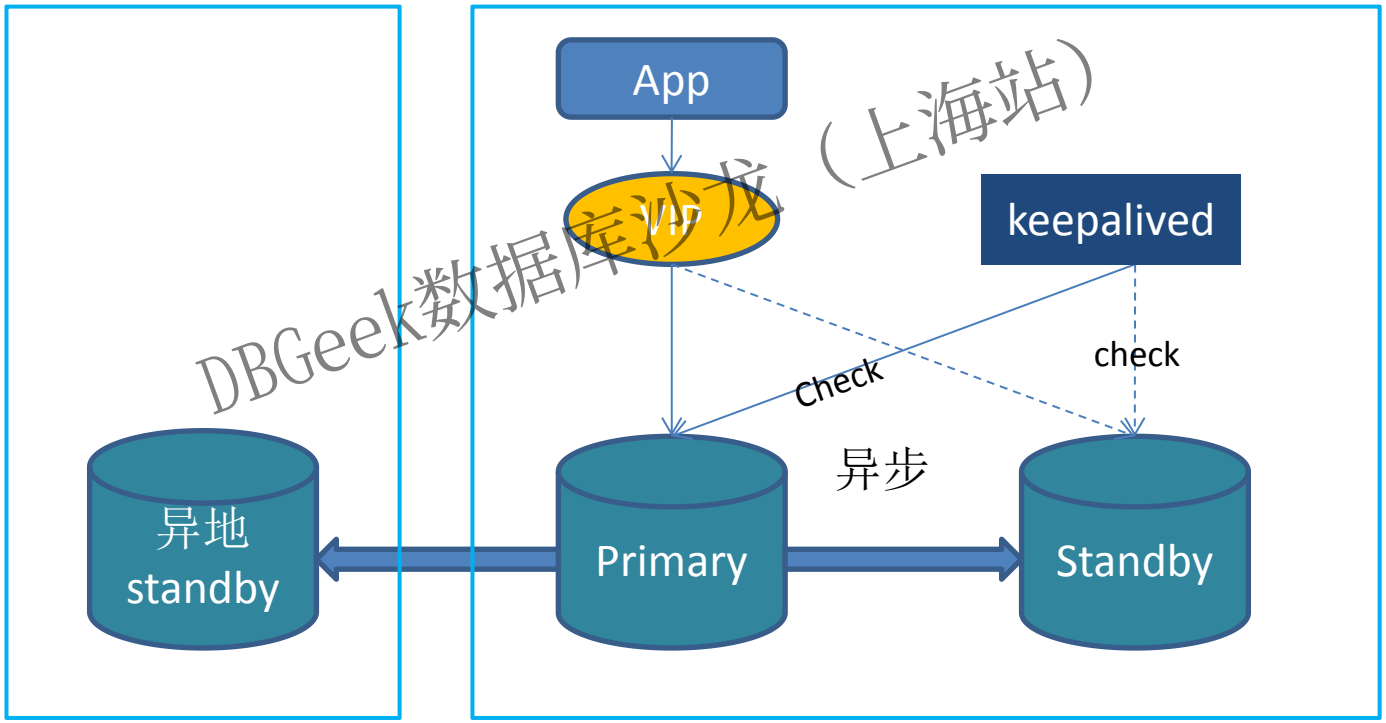
常用的高可用架构

- PostgreSQL 流复制+keepalived
- PostgreSQL 流复制+pgpool
- PostgreSQL 流复制+共享存储
- PG-X系列

常用的高可用架构

- PostgreSQL 流复制+keepalived
- PostgreSQL 流复制+pgpool
- PostgreSQL 流复制+共享存储
- PG-X系列

PostgreSQL 流复制+keepalived

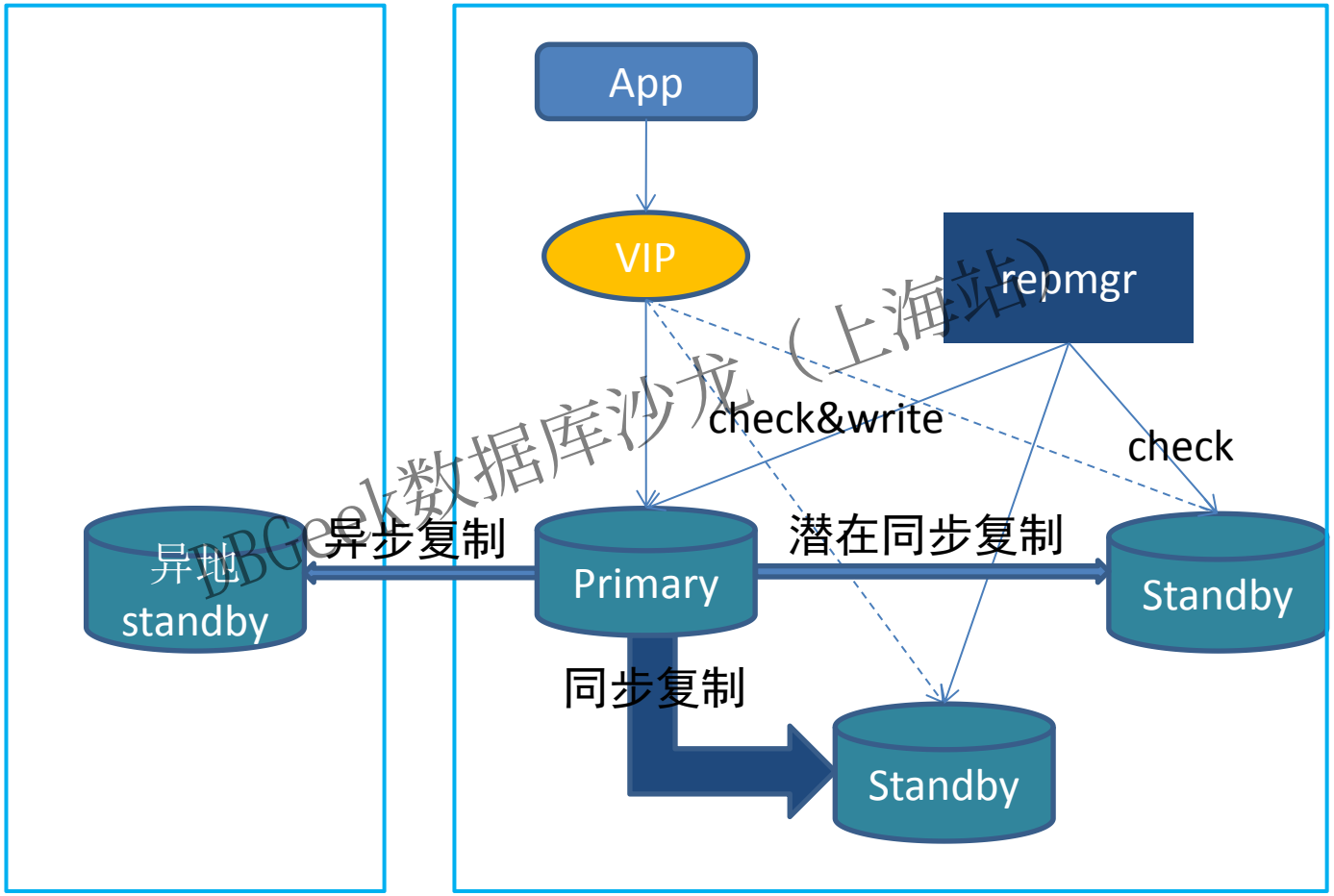


PostgreSQL 流复制+keepalived

- PG异步流复制
- 成本低，使用简单
- 有数据一致性的风险，切换后可能要重搭备库

升级方案：同步流复制+repmgr

同步流复制+repmgr



repmgr升级方案：

- 配置文件repmgr.conf，主备分别配
- 搭建备库非常方便， standby clone
- 注册实例到repmgr的schema中，schema在master register时自动创建
- 检测进程常驻后台repmgrd，可手动或自动failover
- 主库正常关闭，将来可以直接作为备库
- 主库异常关闭，原备库和主库存在lag，则需要根据应用可接受数据丢失的等级来确定恢复方案：

  A:应用可用性优先，可接受少部分数据丢失，备库直接切

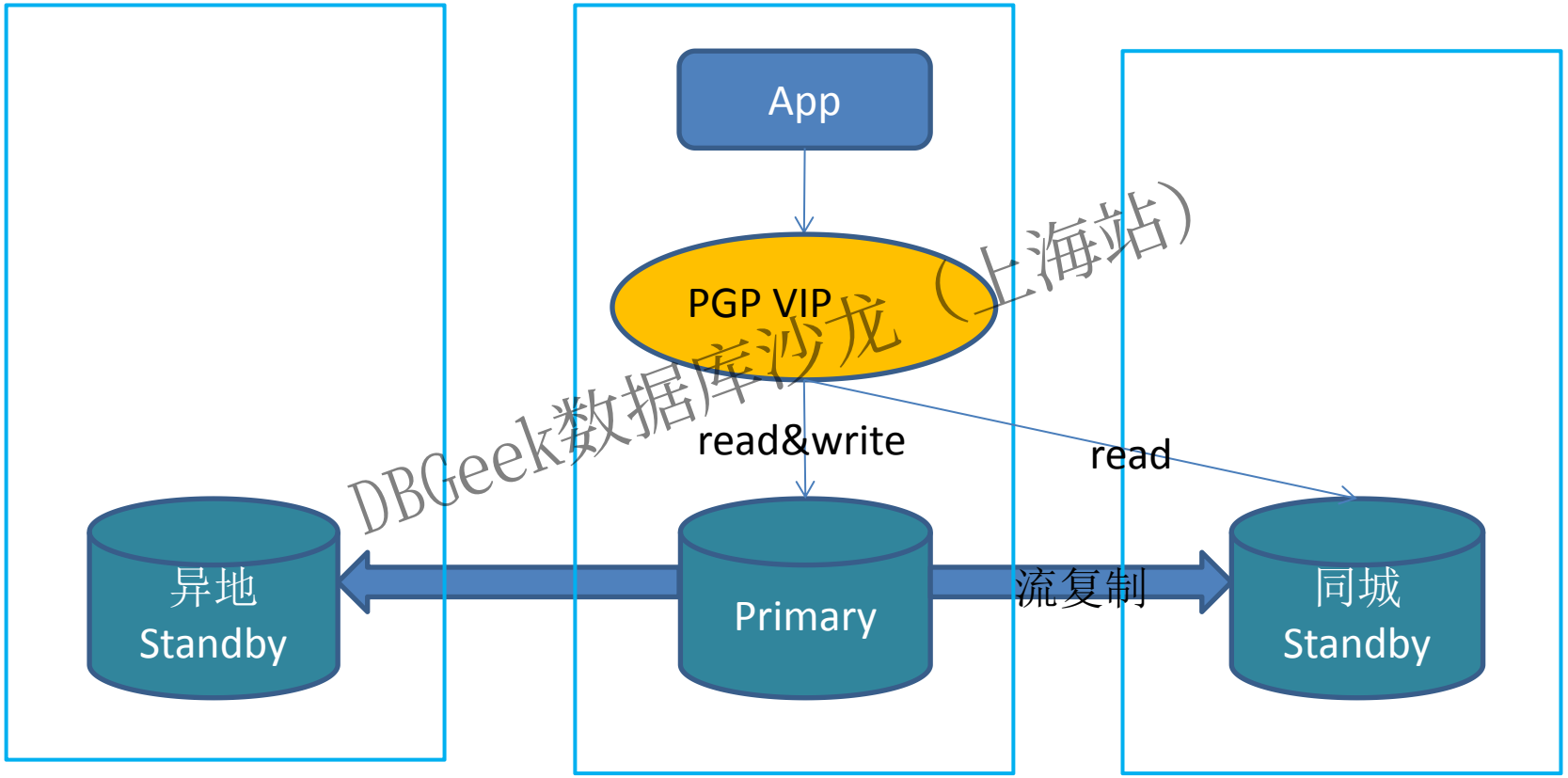
  B:对数据完整性要求高，则要先恢复主库，挽救数据

- 切换后其他备库可以follow 新主库（自动或手动）

常用的高可用架构

- PostgreSQL 流复制+keepalived
- PostgreSQL 流复制+pgpool
- PostgreSQL 流复制+共享存储
- PG-X系列

PostgreSQL 流复制+pgpool



PostgreSQL 流复制+pgpool

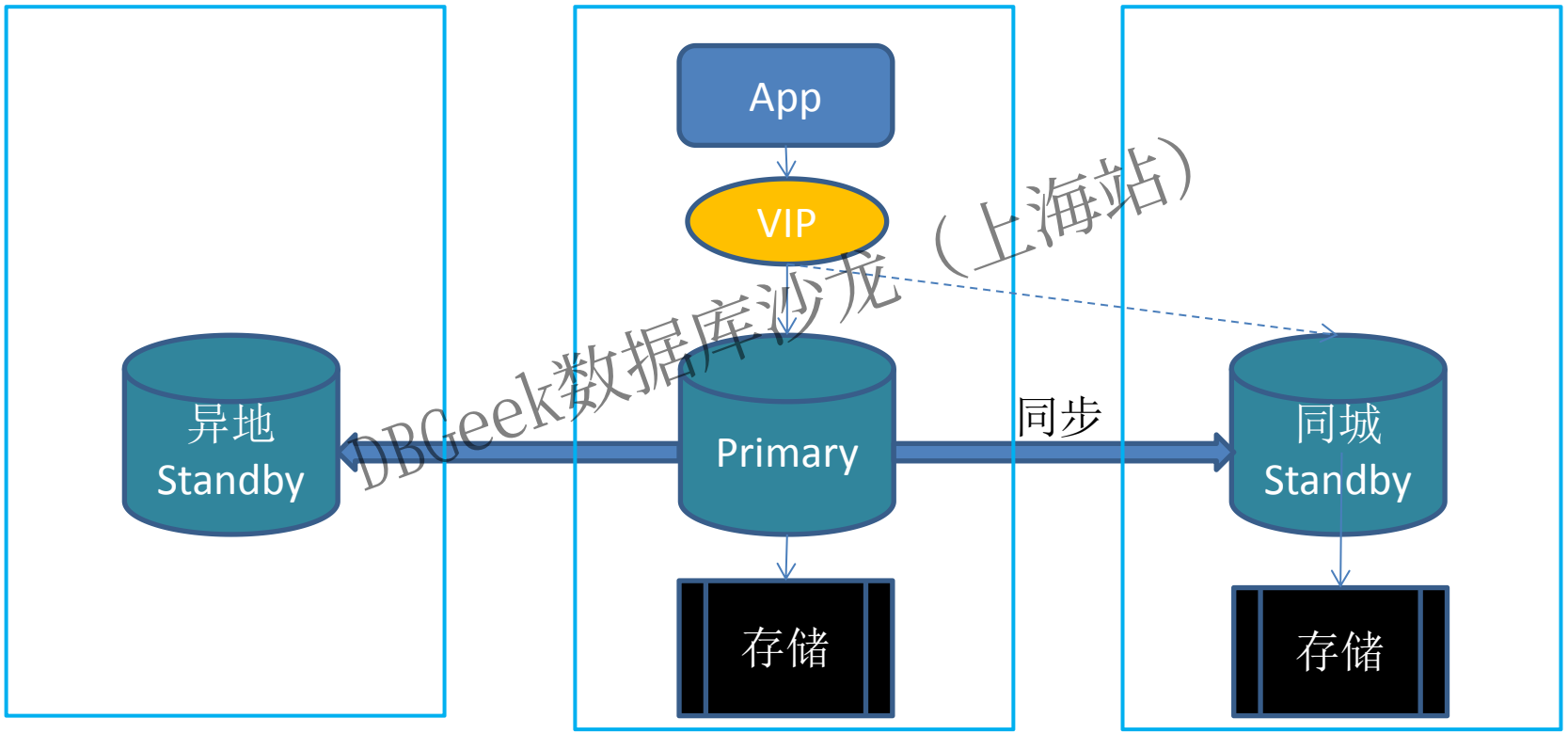
- 读写分离、读的负载均衡
- 配合脚本可以自动failover
- 配置较复杂，相关功能还在完善中
- Pgpool-II：
  http://www.pgpool.net/mediawiki/index.php/Main_Page

常用的高可用架构

- PostgreSQL 流复制+keepalived
- PostgreSQL 流复制+pgpool
- PostgreSQL 流复制+共享存储
- PG-X系列

PostgreSQL流复制+共享存储

- 同步复制



同步复制+共享存储

- 要有共享存储设备：SAN交换机和存储
- PostgreSQL同步复制

  a、设置 postgresql.conf (on primary)

  synchronous_standby_names = 'pg_sync_1'

  b、配置 recovery.conf (on standby)

  primary_conninfo = 'host=192.168.1.1 port=5432 user=repluser' application_name=pg_sync_1'

同步复制+共享存储

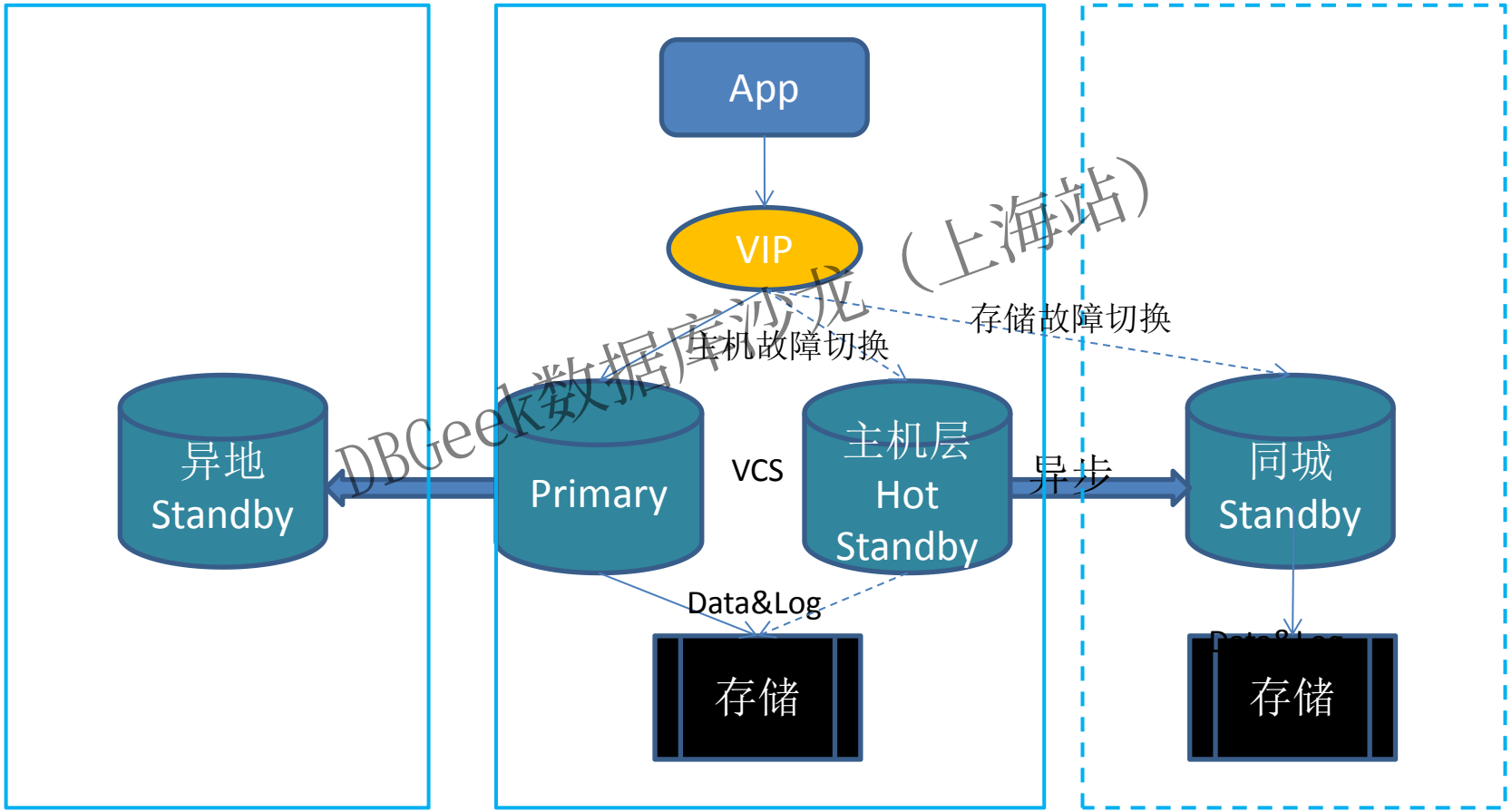
- PG 同步流复制
- 数据，日志都放在共享存储上
- 满足强一致性和高可靠性
- 性能有所下降，价格高

- **升级方案1.0：异步复制+共享存储+VCS集群**

异步复制+共享存储

- PG异步流复制
- 数据，日志都放在共享存储上
- VCS集群，冗余一台主机
- 本地机房冗余，异地可视需求而定

异步复制+共享存储

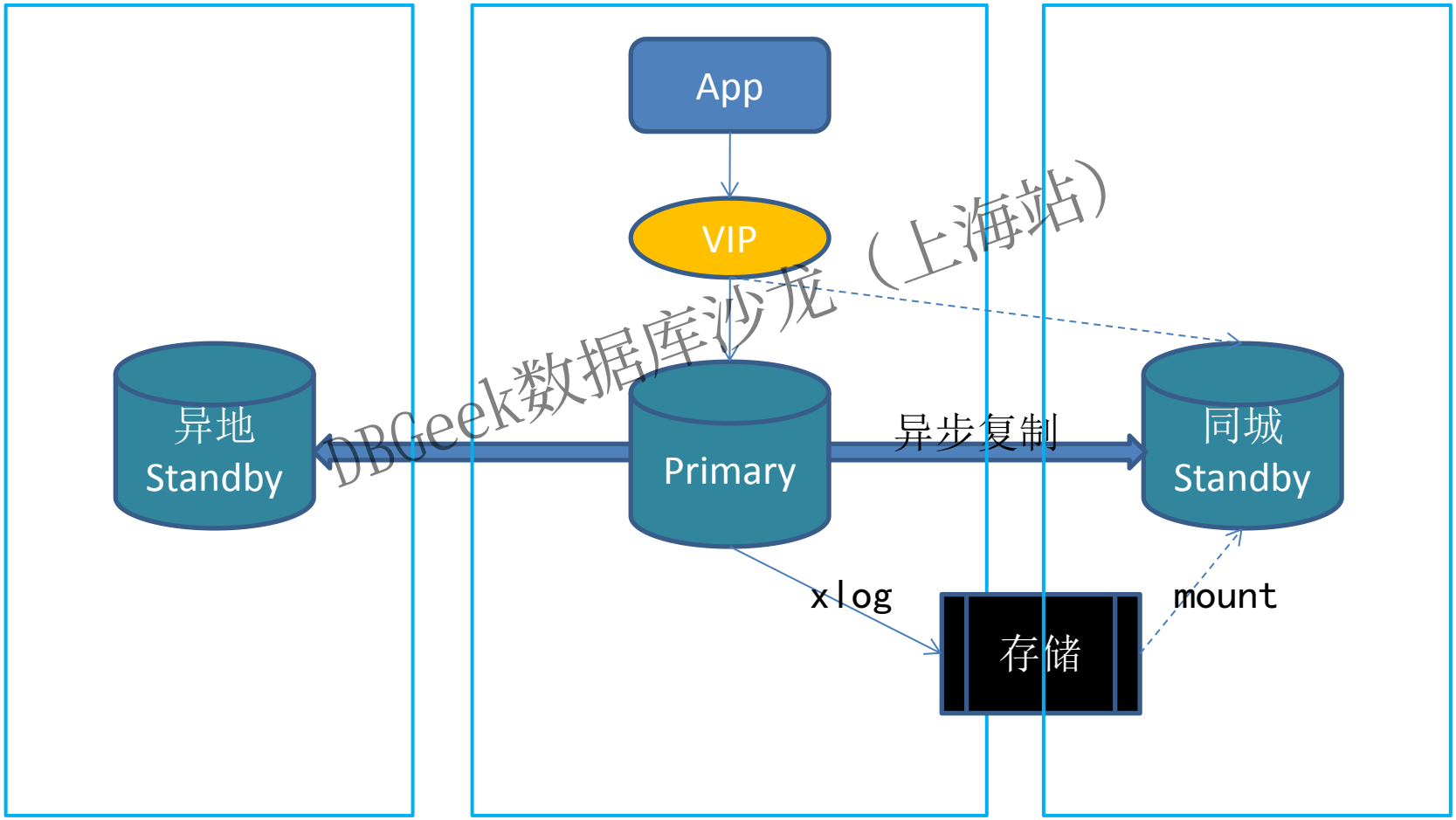


异步复制+共享存储

升级方案2.0:

- 异步复制
- 数据放在本地，日志放在共享存储上
- 备库可以打开查询
- XLOG需要空间较小，成本较低

升级方案2.0



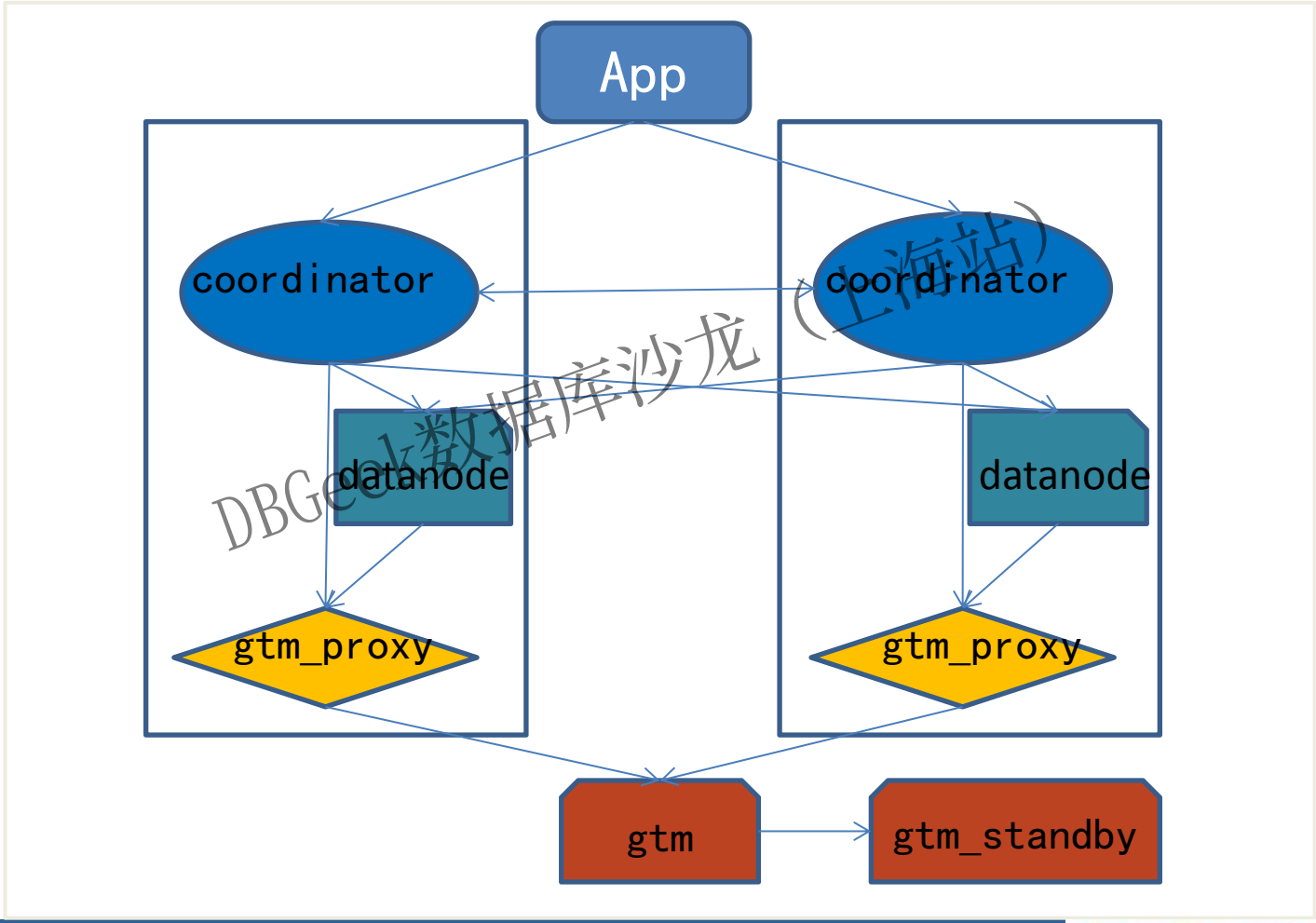
常用的高可用架构

- PostgreSQL 流复制+keepalived
- PostgreSQL 流复制+pgpool
- PostgreSQL 流复制+共享存储
- PG-X系列

PGX2主要特点：

- PG-X2是未来技术发展的方向
- 分布式集群化处理
- 保障数据强一致性，可横向扩展
- 高可用，无单点故障
- 目前还未完全成熟，集群备份方案待完善

PGX2架构



PGX2

- Replication or Distribution

  Replication：每个数据节点都有完整的表数据

  Distribution：即每个数据节点仅保留表的部分数据

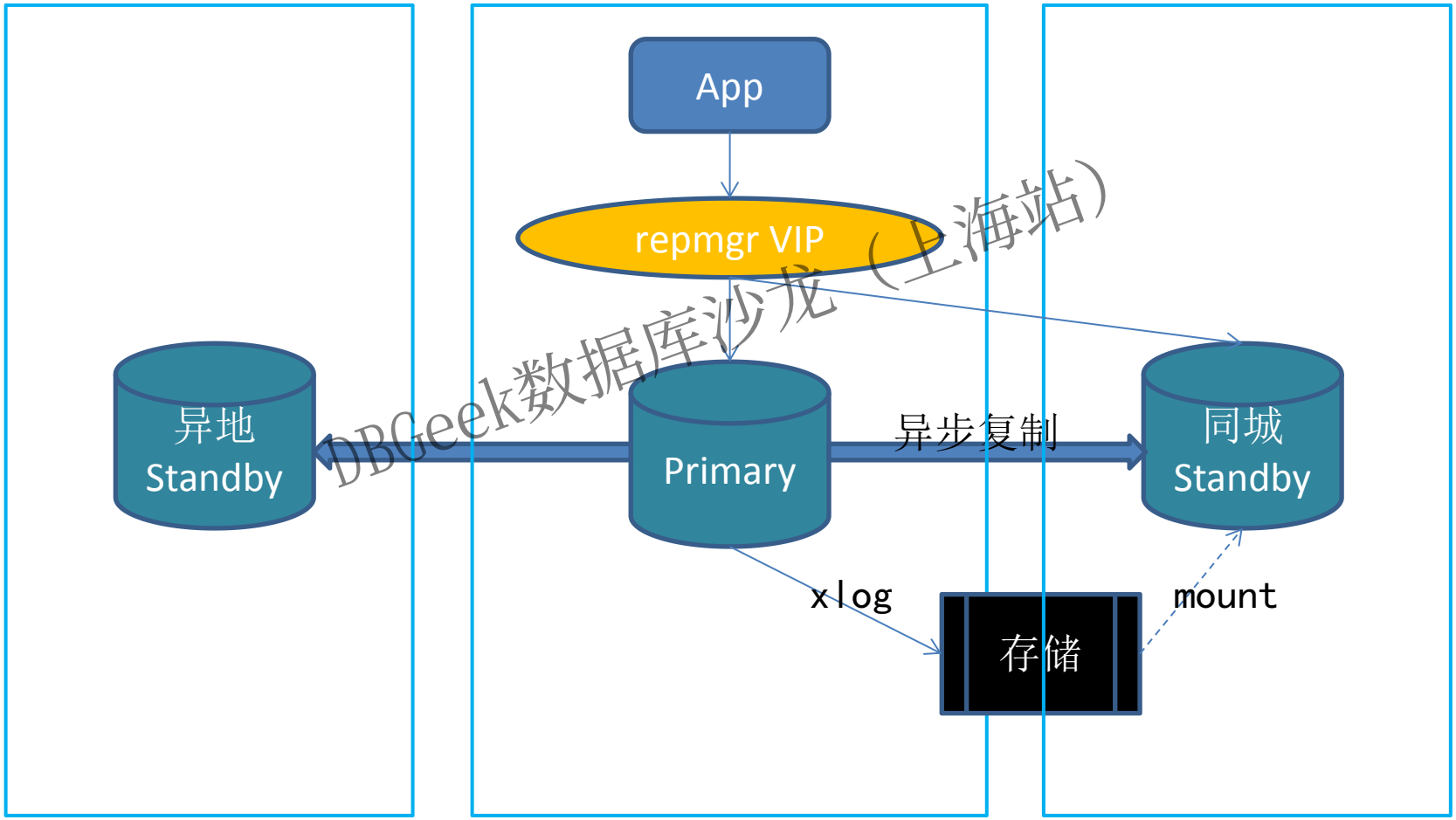
- Write-scalable PostgreSQL cluster
  水平的扩展写能力
- Synchronous multi-master configuration

  在任意master的操作立即为其它节点可见
- Table location transparent

  应用无需更改，在外界看来事务处理没有变化
- 本身还没有大型成熟的应用
- 在高可用上还有很多事情要做
- 备库难做

我们的高可用方案

repmgr+异步复制+共享存储



PG在壹钱包的使用

- ✓PostgreSQL 9.4
- ✓repmgr + 1 primary+2 standby
- ✓repmgr+共享存储+异步流复制
- ✓pg_rman

Thanks!

Q & A