



中国移动
China Mobile

中国移动基于PostgreSQL系列 开源数据库的实践

单位：中移（苏州）软件技术有限公司

日期：2016年12月

目录页

Contents Page

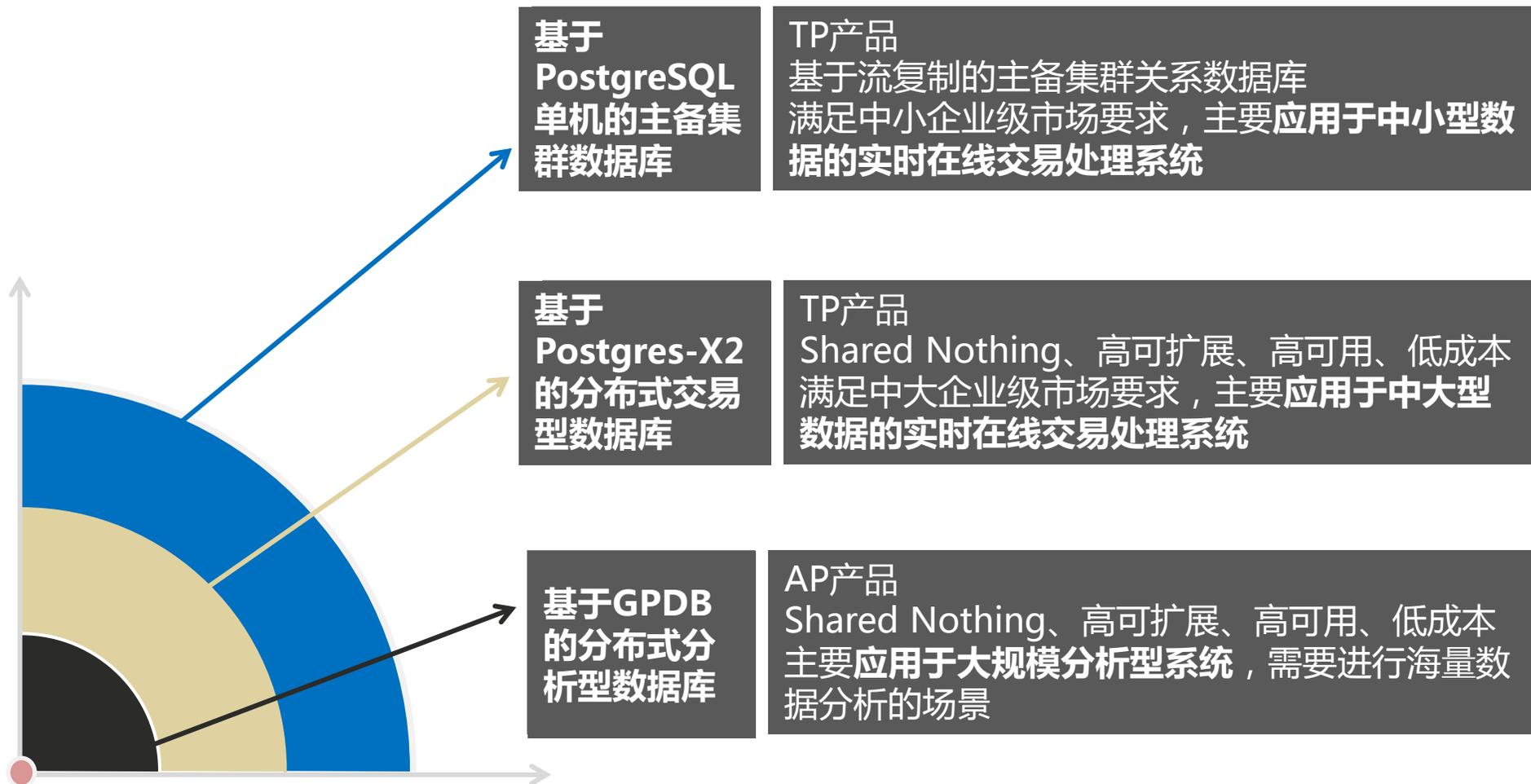


- 一、基于PostgreSQL 系列数据库产品概况
- 二、交易型数据库及其典型案例
- 三、分析型数据库及其典型案例
- 四、工作

OLTP : On-line Transaction Processing 联机事务处理

OLAP : On-Line Analytical Processing 联机分析处理

	OLTP	OLAP
用户	操作人员	业务、管理人员
用户数	数量庞大	数量很小
功能	日常操作处理	分析决策
数据	当前的, 系统中产生的	历史的, 聚集的, 多维集成的
存取	读/写数十条记录	读上百万条记录
工作单位	简单的事务	复杂的查询, 多表join、全表扫描等
SQL数据量	数据量小	数据量大
响应时间	要求高	与具体查询有关系
DB大小	MB-GB级别	TB-PB级别



目录页

Contents Page

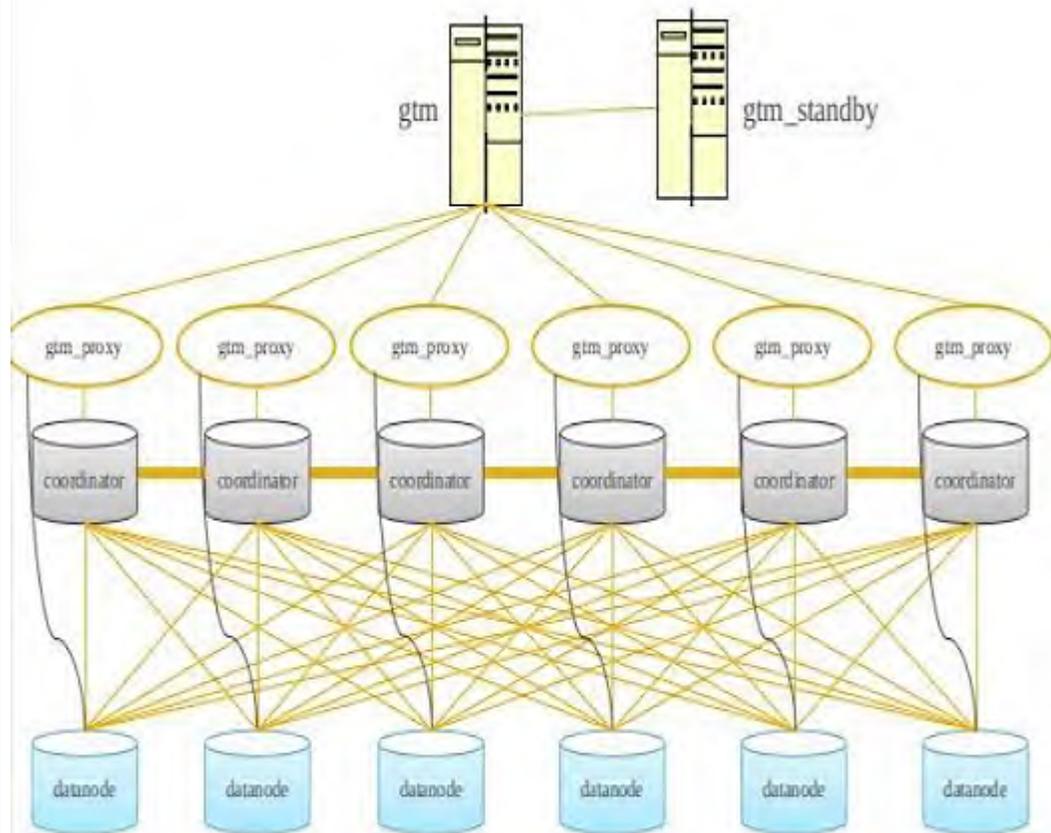


- 一、基于PostgreSQL 系列数据库产品概况
- 二、交易型数据库及其典型案例
- 三、分析型数据库及其典型案例
- 四、工作

苏研基于pgx2研发的一款**分布式交易型关系数据库**，主要应用于海量数据的实时在线交易处理系统。

支持SQL92标准，使得传统交易型应用可实现平滑迁移，目标是为CRM、客服等核心OLTP系统提供数据库服务。

Postgres-X2的架构



核心组件

GTM：gtm用于全局事务控制，在一个CLUSTER中只能有一台主gtm。

Gtm-standby：Gtm-standby属于高可用组件的一部分。在gtm出现故障时会自动替代gtm的位置，对外提供服务。

Coordinator：协调节点是数据节点与应用之间的接口，协调节点并不物理上存储表数据。当应用发起SQL时，会先到达协调节点，然后协调节点将SQL分发到各个数据节点，汇总数据。

GTM_Proxy：gtm_proxy为降低gtm压力而诞生的，用于对coordinator节点提交的任務进行分组等操作，机器中可以存在多个gtm_proxy。

Datanode：数据节点物理上存储表数据，表数据存储方式分为分片和完全复制两种。

Q：高并发下触发bug，hang所有用户连接

A：pooler组件问题，此bug已修复

Q：数据不一致

A：经过三个月的努力，终于解决了这个bug，修复后再没有出现数据丢失问题

老夫聊发少年狂，写程序，到天亮。

千行代码，bug何处藏。

纵使上线又怎样。朝令改，夕断肠。

数据每天不重样，天天修，日日忙。

又要加班，惟有泪千行。

每晚灯火阑珊处，程序员，工作狂！

稳定性：

内存溢出 coredump

功能：

支持更多并发
sequence

性能：

减少磁盘I/O
减少网络I/O

新功能：

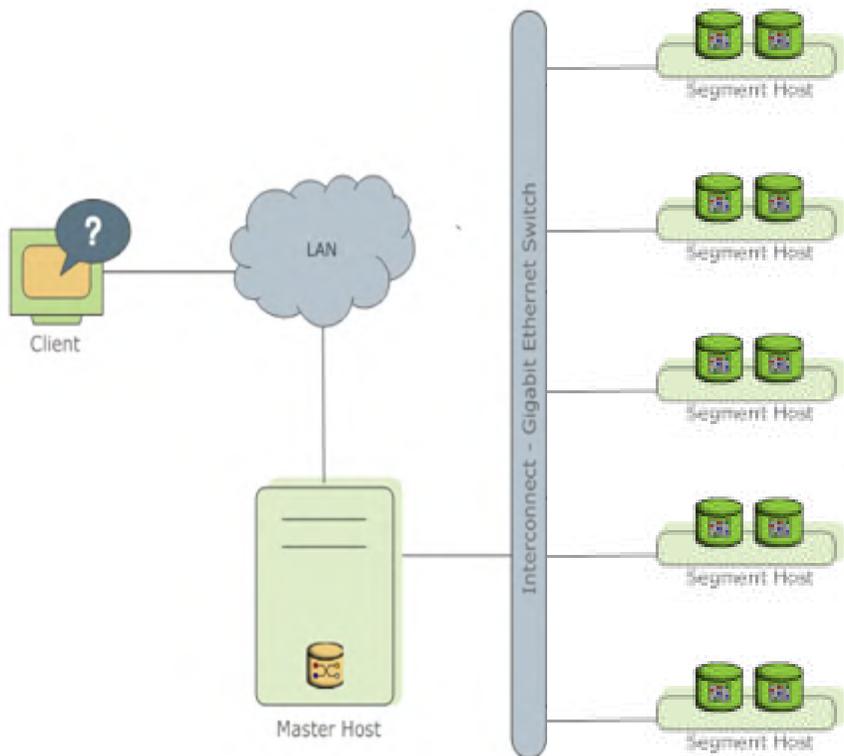
丰富集群管理工具功能
增加分片规则(range, list)

目录页

Contents Page

- 一、基于PostgreSQL 系列数据库产品概况
- 二、交易型数据库及其典型案例
-  三、分析型数据库及其典型案例
- 四、工作

苏研基于gpdb研发的一款**分布式分析型关系数据库**，主要用来处理大规模的数据分析任务，包括数据仓库（DW）、商务智能（BI/OLAP）和数据挖掘（DM）等。支持ANSI SQL 2008标准和SQL OLAP 2003 扩展，目标是提供大数据的存储引擎、计算引擎和分析引擎。



支持psql、pgadmin、ODBC、JDBC、PDBI、Python、libpq

Master节点负责协调整个集群、配备standby、分布式优化器

Interconnect负责集群中各个节点间的数据传输

数据节点Segment用于存放数据、配备mirror

支持多样化的外部数据源导入/导出，采用并行数据加载，性能达TB/小时

Q：大表查询时，所有数据丢失

Q：大并发时出现部分事务执行失败的情况

Q：工具在多线程环境下容易触发一个问题，造成软件夯住。

锁冲突关系

Requested Lock Mode	Current Lock Mode							
	ACCESS SHARE	ROW SHARE	ROW EXCLUSIVE	SHARE UPDATE EXCLUSIVE	SHARE	SHARE ROW EXCLUSIVE	EXCLUSIVE	ACCESS EXCLUSIVE
ACCESS SHARE								X
ROW SHARE							X	X
ROW EXCLUSIVE					X	X	X	X
SHARE UPDATE EXCLUSIVE				X	X	X	X	X
SHARE			X	X		X	X	X
SHARE ROW EXCLUSIVE			X	X	X	X	X	X
EXCLUSIVE		X	X	X	X	X	X	X
ACCESS EXCLUSIVE	X	X	X	X	X	X	X	X

表锁 : `gp_toolkit.gp_locks_on_relation`

资源队列锁 : `gp_toolkit.gp_locks_on_resqueue`

资源队列

限制用户可以使用的系统资源：并发的sql数量、每个sql的cost、sql的优先级

新建用户默认属于pg_default 超级用户不受资源队列限制

```
create resource queue queue_name with (xxx)
```

xxx参数

active_statements：同时执行的sql数量

max_cost：每个sql的cost最大值，超过则报错无法执行

cost_overcommit：没有其他语句执行时，超过cost值也可以执行

priority：资源队列的优先级

memory_limit：该资源队列的最大内存限制

查看方法

```
gp_toolkit.gp_resq_activity
```

```
gp_toolkit.gp_resq_priority_statement
```

数据倾斜

数据按指定的分布键插入到表后，数据集中存储某些数据节点（segment）上，没有均匀分布

数据倾斜后果

- 1.计算倾斜
- 2.sql执行效率降低
- 3.部分节点磁盘使用率不高

检查方法

普通表：select pg_relation_size('table') from gp_dist_random('gp_id');

分区表：自定义函数

解决方法

- 1.重新设计分布键，尽可能保证数据唯一性
- 2.分布方式设置为 random

表膨胀(系统表、普通表)

存储使用MVCC实现，表空间不能被vacuum回收

表膨胀后果

导致表数据文件越来越大，增大表扫描时间

检查方法

gp_toolkit.gp_bloat_diag

解决方法

定期执行vacuum analyze

目录膨胀(pgsql_tmp、pg_log)

解决方法

查杀长sql

磁盘使用率、目录大小监控

一些限制

数据库是否有长事务，异常长连接

数据库年龄

数据库组件角色的变化

数据库、主要表数据倾斜情况

表（系统表、普通表）膨胀检查

慢sql监控

目录页

Contents Page

一、基于PostgreSQL 系列数据库产品概况

二、交易型数据库及其典型案例

三、分析型数据库及其典型案例



四、工作

已接收patch

- (1) <https://github.com/postgres-x2/postgres-x2/pull/124>
- (2) <https://github.com/postgres-x2/postgres-x2/pull/132>
- (3) <https://github.com/postgres-x2/postgres-x2/pull/138>
- (4) <https://github.com/postgres-x2/postgres-x2/pull/145>
- (5) <https://github.com/postgres-x2/postgres-x2/pull/147>
- (6) <https://github.com/postgres-x2/postgres-x2/pull/149>
- (7) <https://github.com/postgres-x2/postgres-x2/pull/151>
- (8) <https://github.com/postgres-x2/postgres-x2/pull/156>
- (9) <https://github.com/postgres-x2/postgres-x2/pull/164>
- (10) <https://github.com/postgres-x2/postgres-x2/pull/284>
- (11) <https://github.com/postgres-x2/postgres-x2/pull/285>
- (12) <https://github.com/postgres-x2/postgres-x2/pull/325>
- (13) <https://github.com/greenplum-db/gpdb/pull/867>
- (14) <https://github.com/postgres-x2/postgres-x2/pull/329>
- (15) <https://github.com/postgres-x2/postgres-x2/pull/330>
- (16) <https://github.com/greenplum-db/gpdb/pull/895>
- (17) <https://github.com/postgres-x2/postgres-x2/pull/332>
- (18) <https://github.com/postgres-x2/postgres-x2/pull/333>
- (19) <https://github.com/postgres-x2/postgres-x2/pull/334>
- (20) <https://github.com/greenplum-db/gpdb/pull/1035>

我们的团队：一个年轻又充满激情的团队

我们的口号：搞数据库我们是认真的

我们的招聘：数据库开发工程师，DBA

如果你有兴趣，欢迎加入我们

yanchaozhong@cmss.chinamobile.com

