

Postgres-X2介绍

李元佳

自我介绍

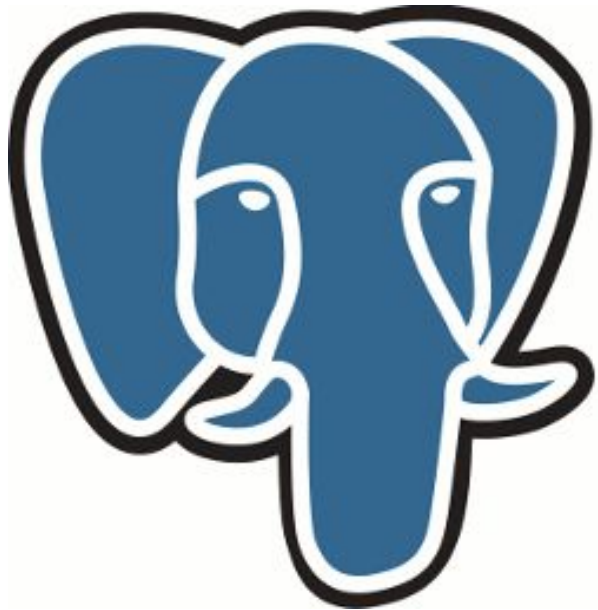
目录

背景介绍

*Postgres-X2*的架构及技术

测试及性能结果

其他



PostgreSQL

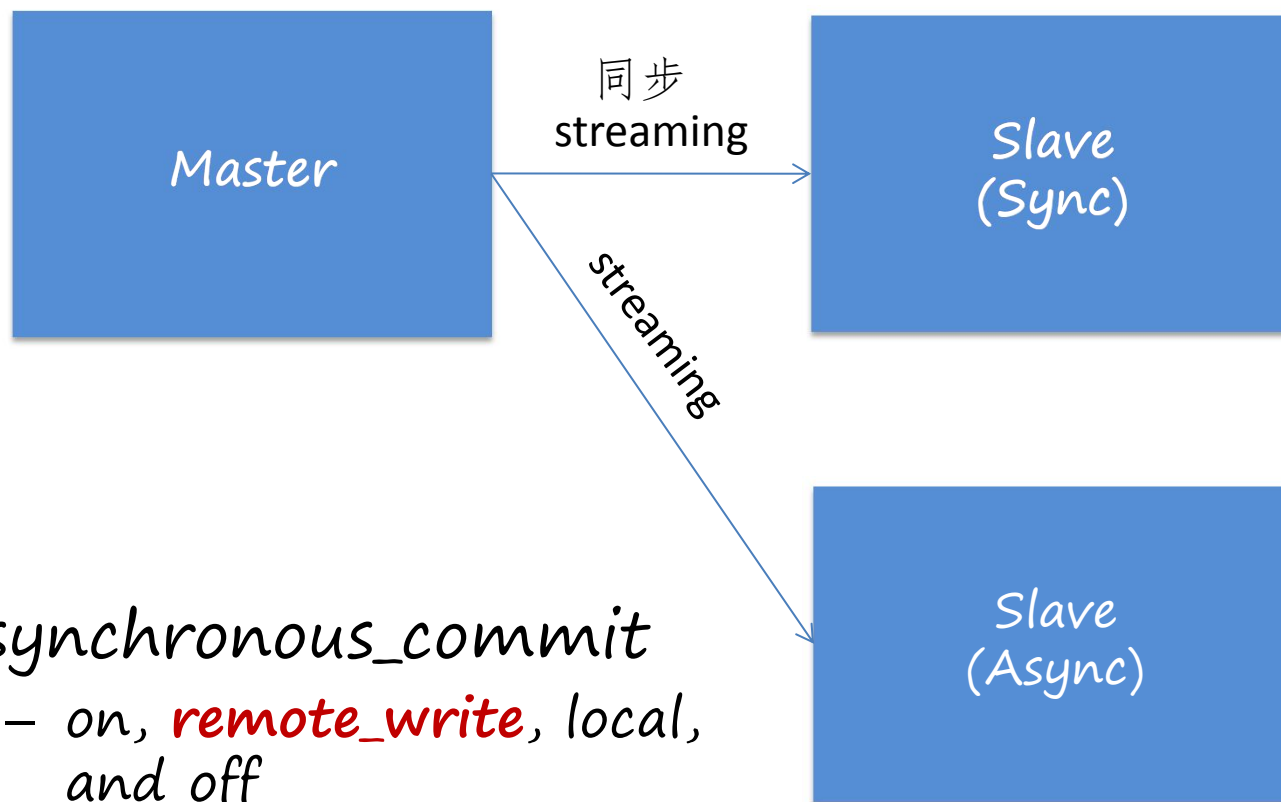
the world's most advanced open source database

Postgres的简介

- 开源的RDBMS
- 功能丰富
 - 完整SQL、事务、存储过程、同步复制
- 企业应用领域比较多
- 国内案例
 - 去哪儿、平安科技、国家电网等



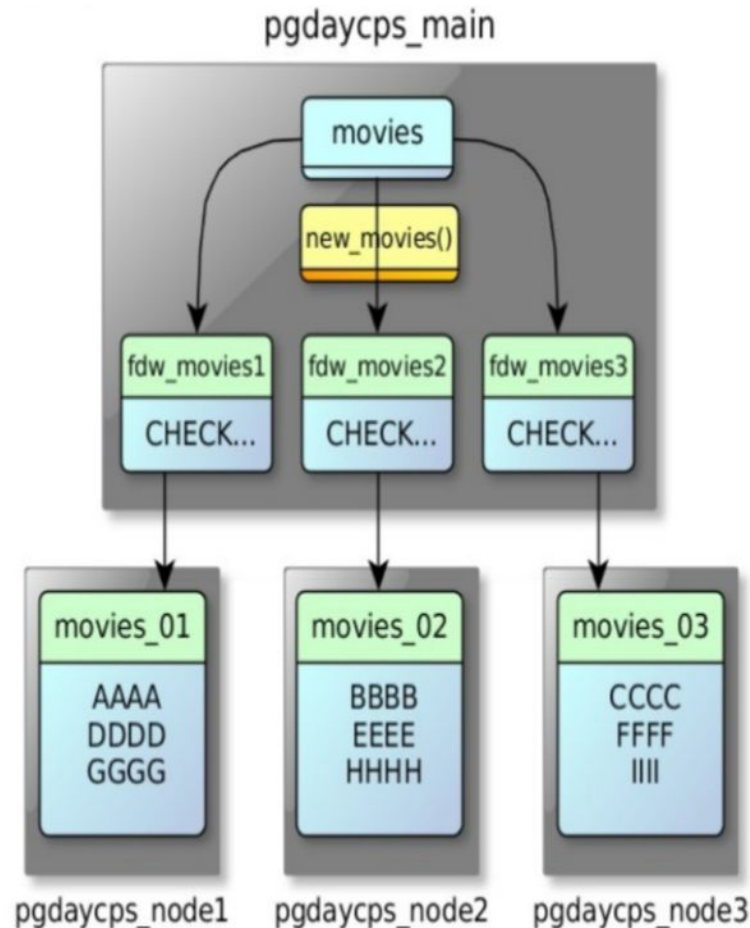
流复制



数据零丢失+高性能

高可用问题已经解决

Sharding: postgres_fdw



高扩展呢？

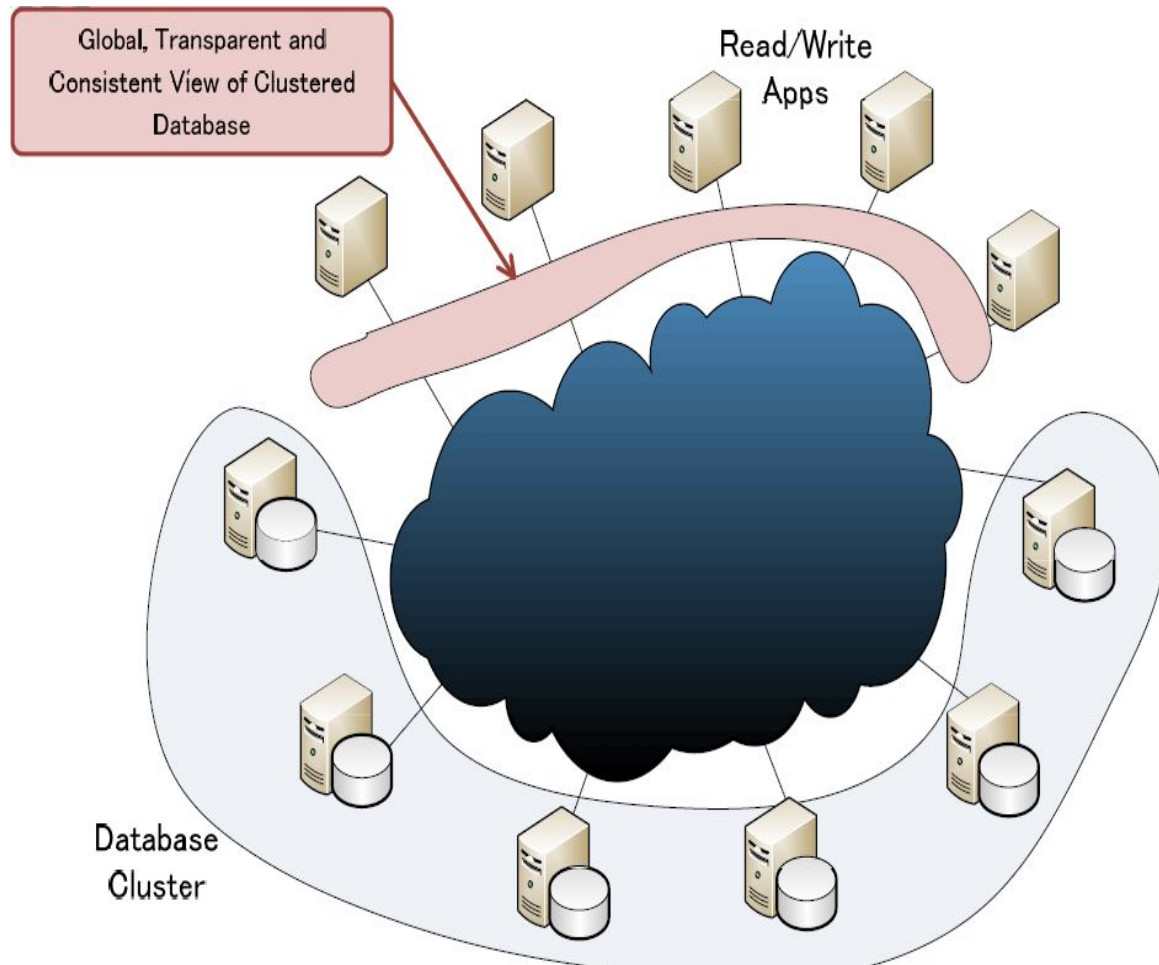


分布式数据库

- 面向大数据量、高并发的`OLTP`场景
- 多主多读、横向扩展
- 全功能关系型数据库(`ACID`、`SQL`几乎无限制)

全功能 + 高扩展

Postgres-XC/XL



Postgres-XC/XL简介

- 开源
 - *Postgres-XC*采用*Postgres*协议(类似*BSD*协议)
 - *Postgres-XL*以前是*Mozilla*协议, 目前已经改为*PostgreSQL*协议
- 面向*OLTP*及*OLAP*场景
- 采用*Share-Nothing*架构、弹性扩展
- 基于*Postgres*改造、功能几乎完全继承

社区发展历史

- 2004~2008 NTT Data构建了模型Rita-DB
- 2009年 NTT Data与EnterpriseDB合作进行社区化开发
- 2012, Postgres-XC 1.0正式发布
- 2012, StormDB在XC基础上增加MPP功能.
- 2013, XC 1.1发布; TransLattice 收购 StormDB
- 2014, XC 1.2发布; StormDB 开源为 Postgres-XL.
- 2015, 两个社区合并为Postgres-X2





目录

背景介绍

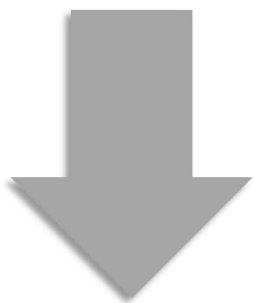
*Postgres-X2*的架构及技术

测试及性能结果

其他

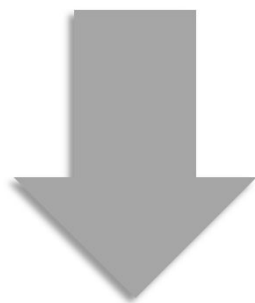
设计理念

高扩展



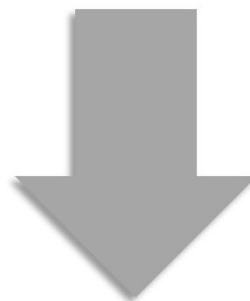
- Share nothing 架构
- 功能解耦、分层扩展
- 数据分散在多个节点

全功能



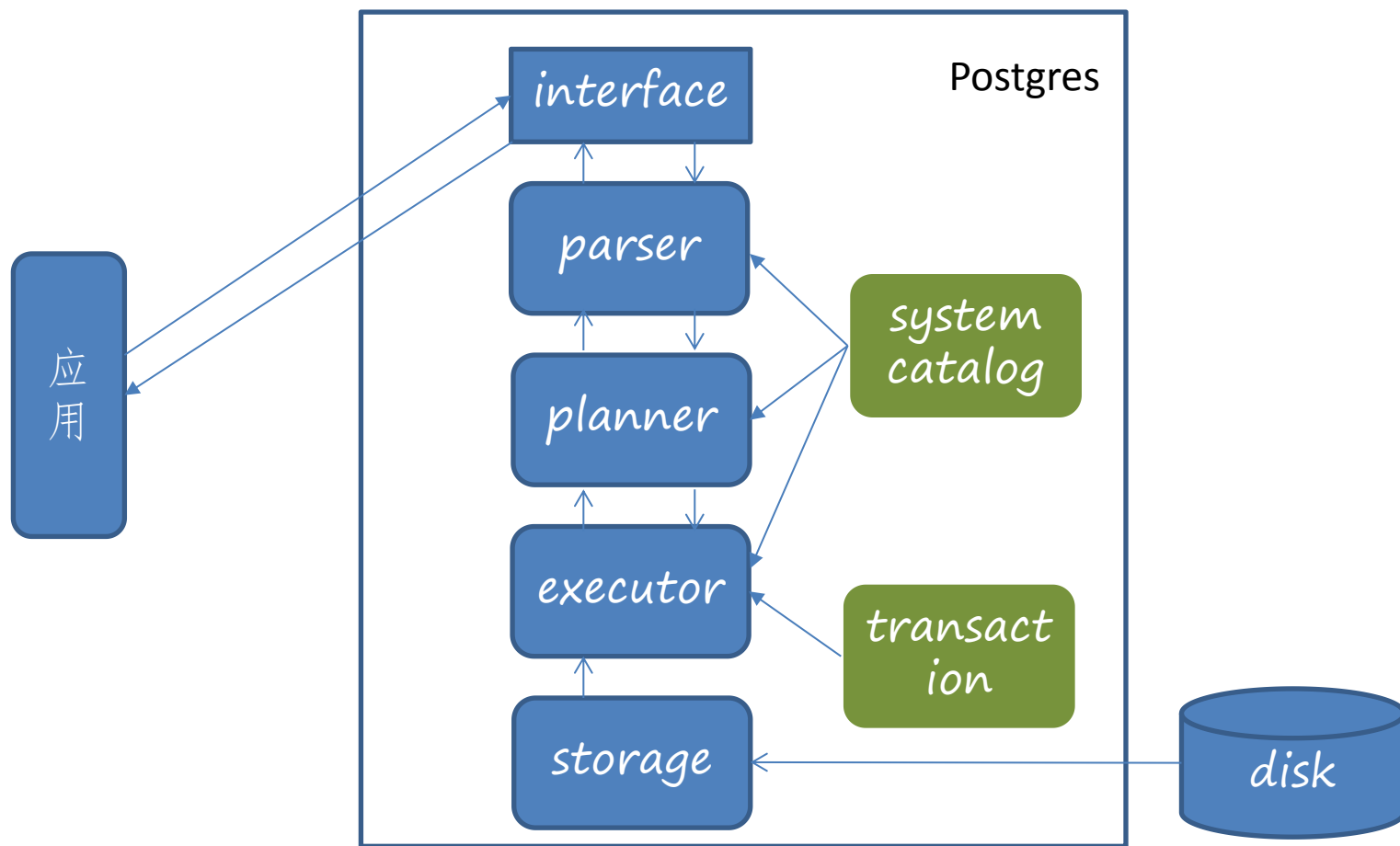
- 继承Postgres功能
- 继承Postgres生态
- SQL能力不受限制
- 支持存储过程

强一致性



- MVCC
- 全局事务支持强一致

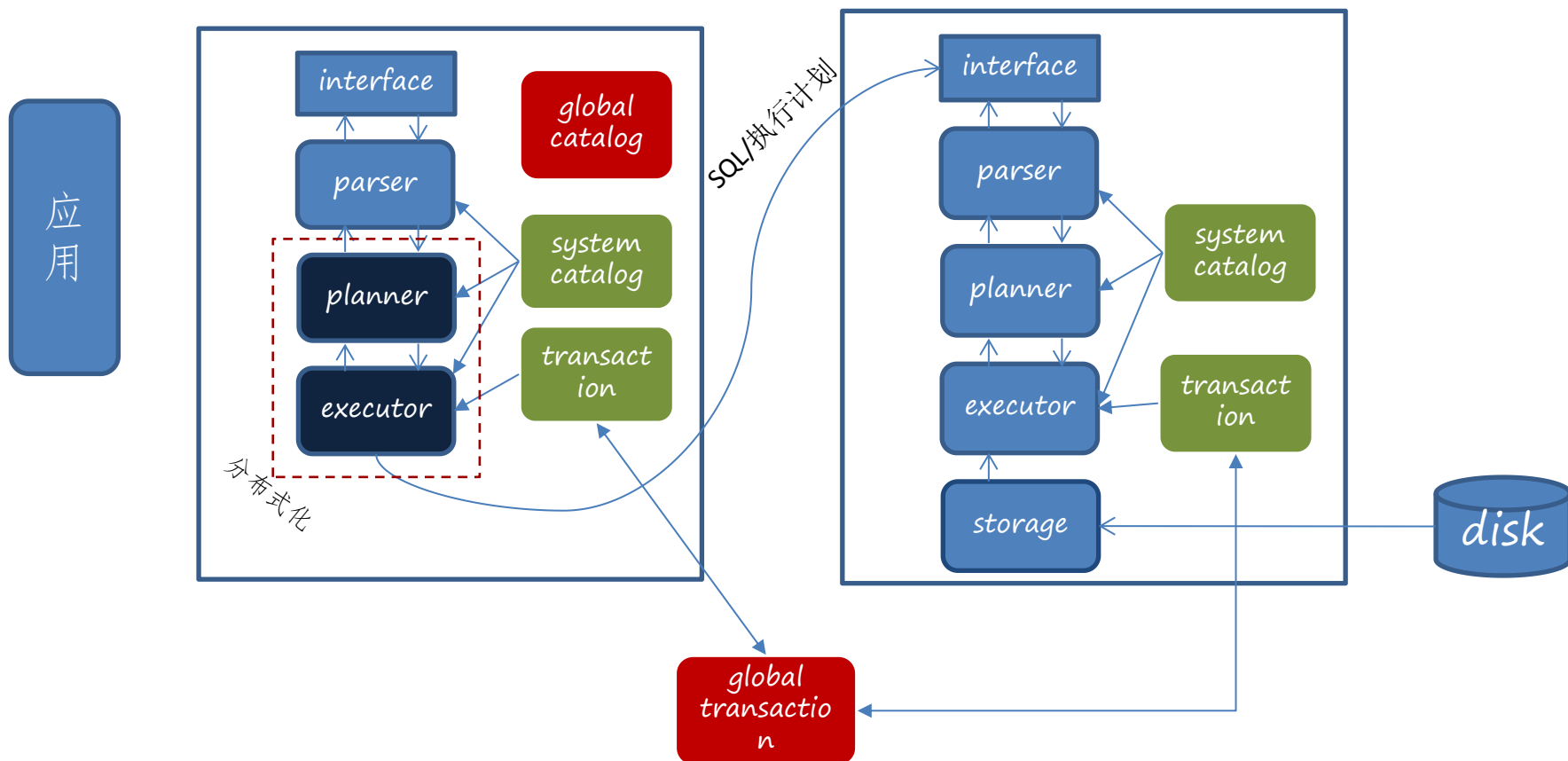
Postgres的架构



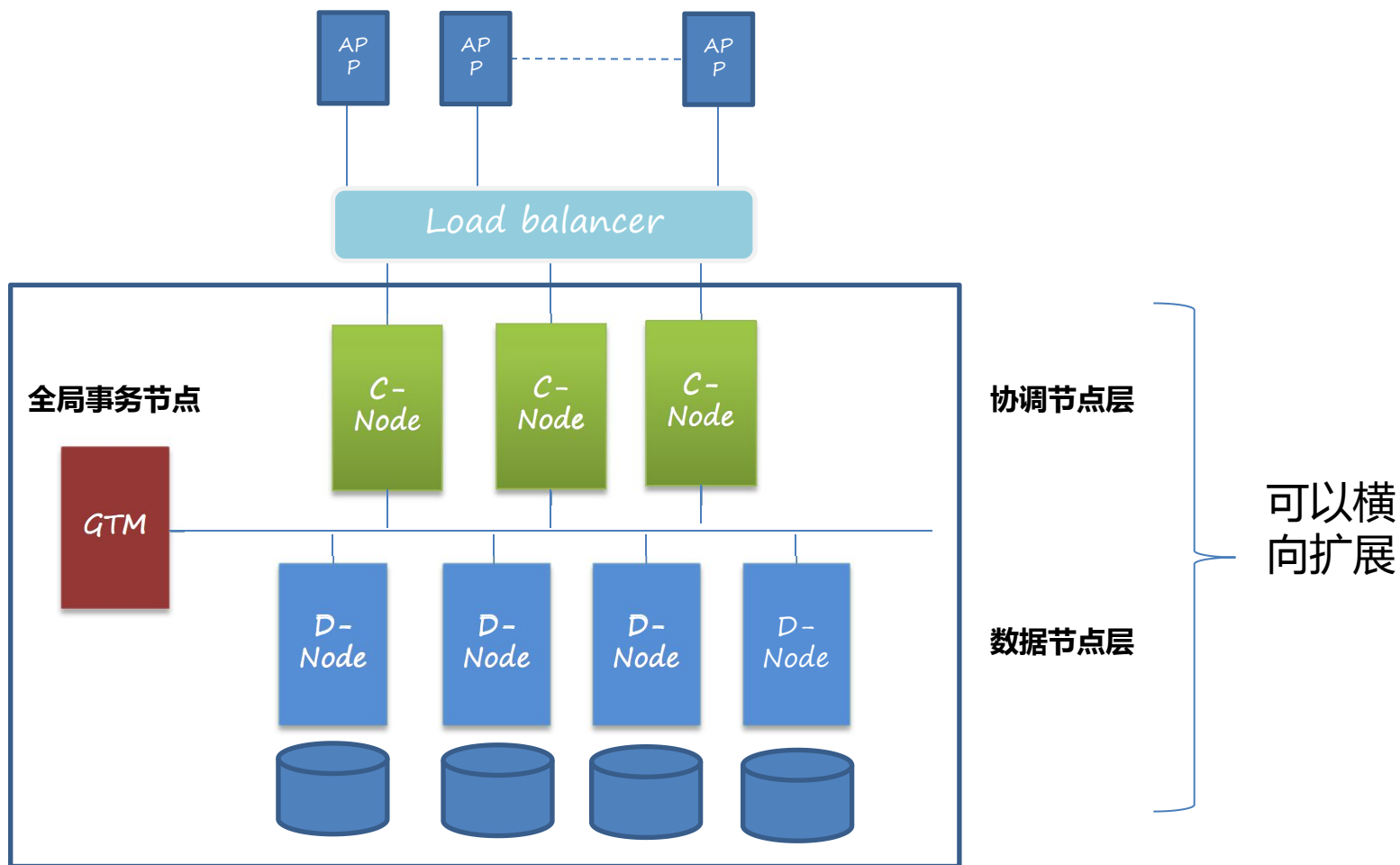
架构的解耦及分布式化

分布式执行层

数据存储层



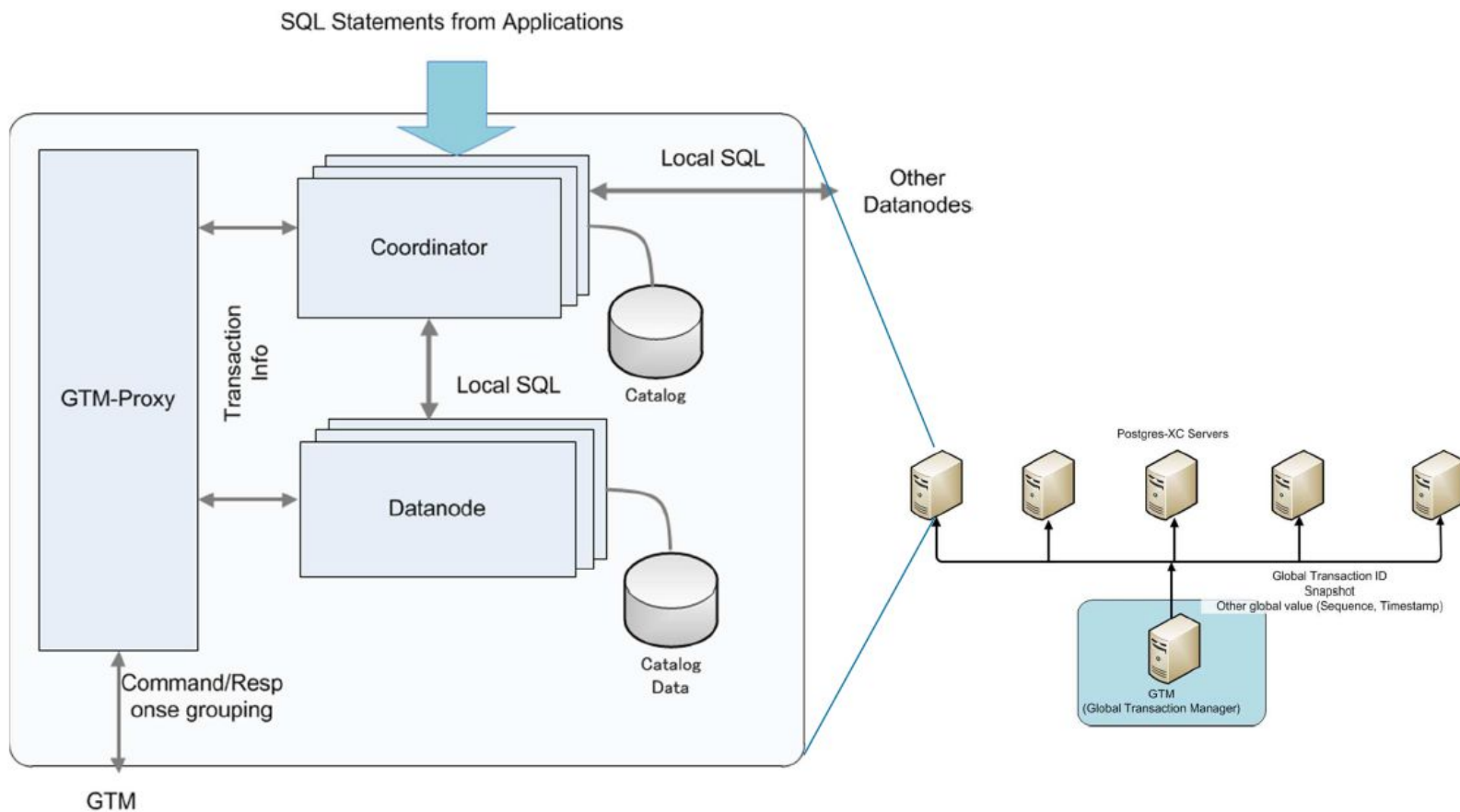
Postgres-XC的整体架构



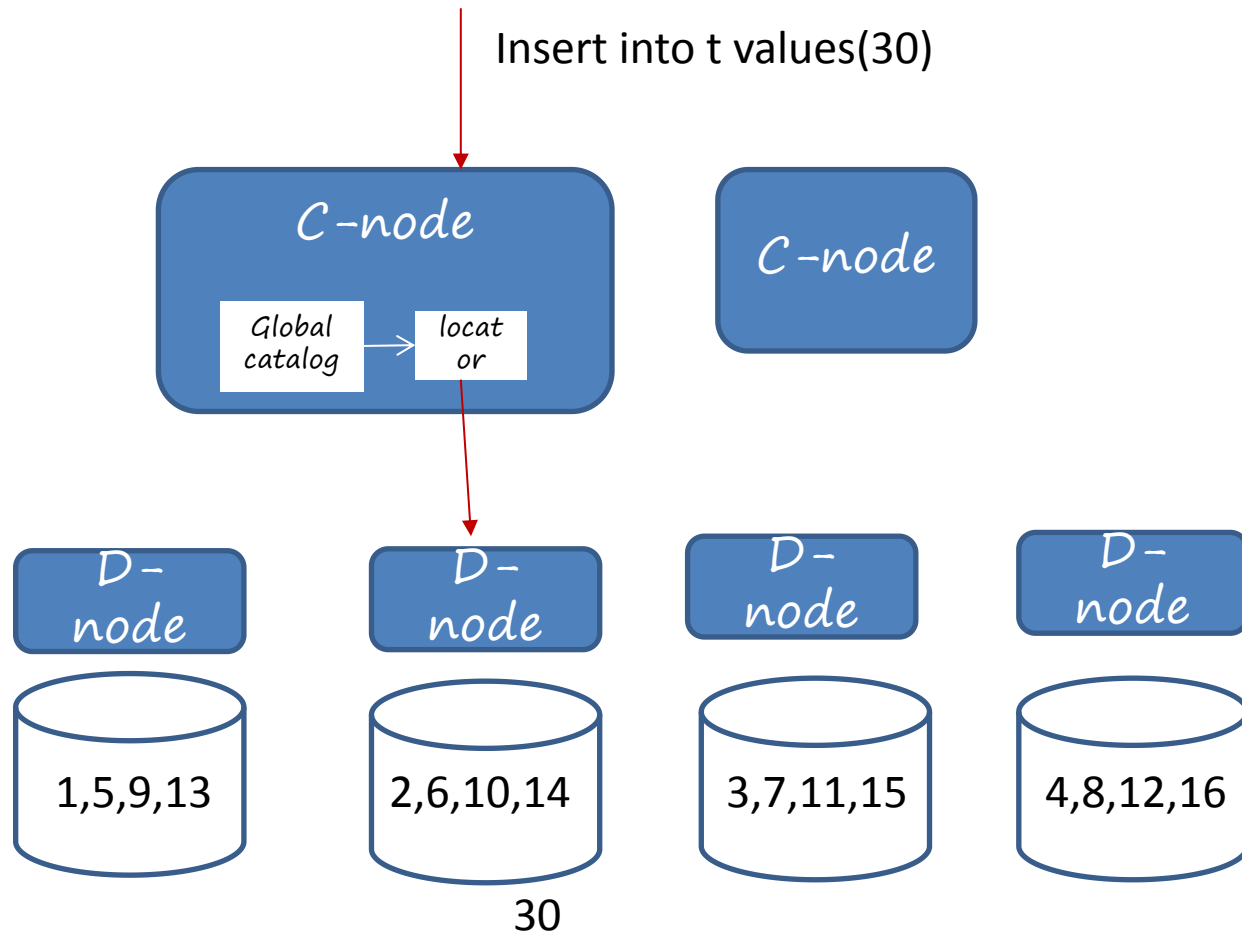
Postgres-XC主要的模块

- **Coordinator node: 协调节点**
 - 负责接收用户请求、生成并执行分布式查询、把SQL语句发给相应的数据节点
- **Data node: 数据节点**
 - 实际数据存储节点
- **GTM: 全局事务节点**
 - 生成全局唯一的事务ID
 - 全局的事务的状态
 - 序列等全局信息

关键模块之间的关系



数据如何分布



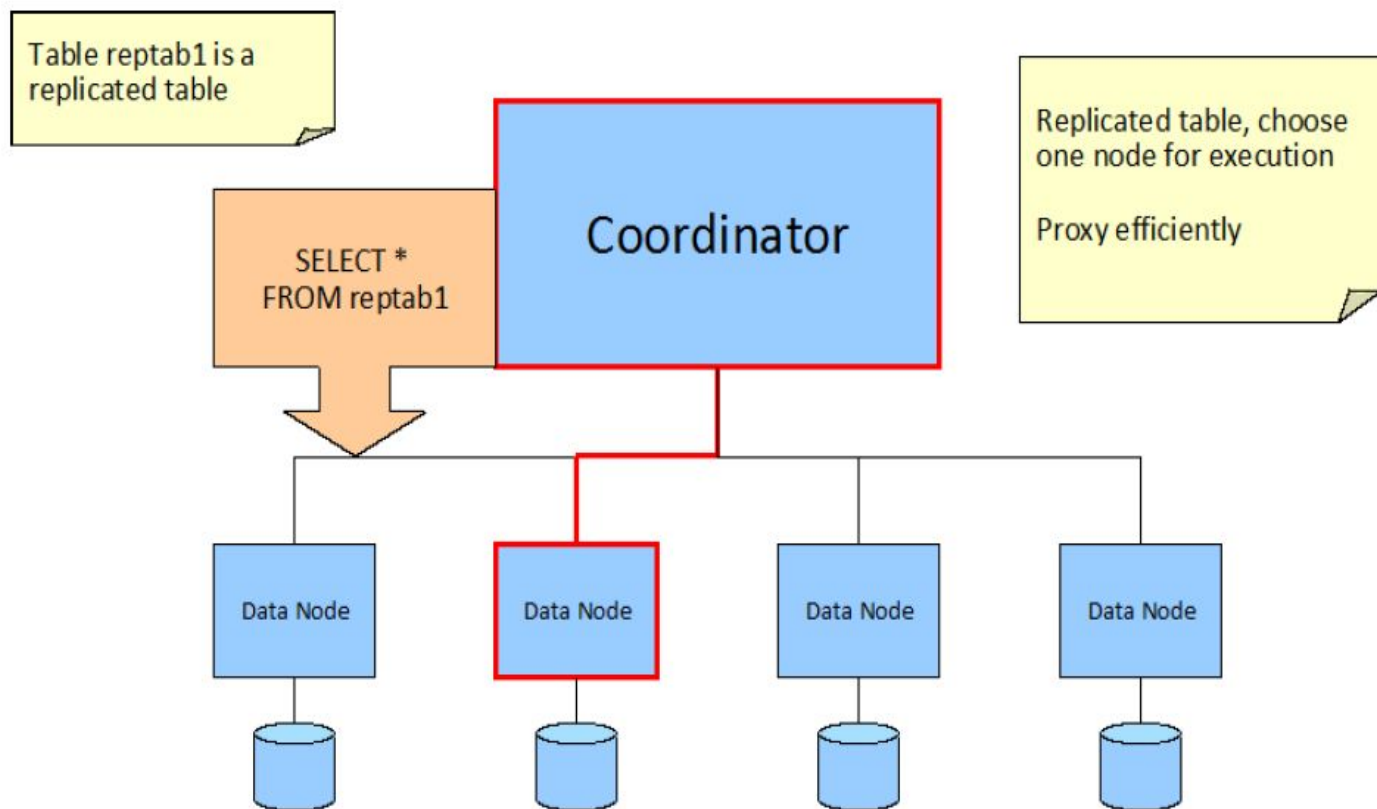
数据分布方式

- *replicated table* 复制表
 - 表在多个节点复制
- *distributed table* 分布式表
 - Hash
 - Round robin
 - Range(未实现)
 - User define (未实现)

分布式查询优化

- Parallel query
- Where pushdown
- Join pushdown
- Expression pushdown
- Order by
- Two phase Aggregate
- Fast query shipping

单一节点处理



单一节点处理

```
EXPLAIN VERBOSE DELETE FROM test WHERE a = 100;
```

```
QUERY PLAN
```

```
-----  
Data Node Scan on "__REMOTE_FQS_QUERY__" (cost=0.00..0.00 rows=0 width=0)
```

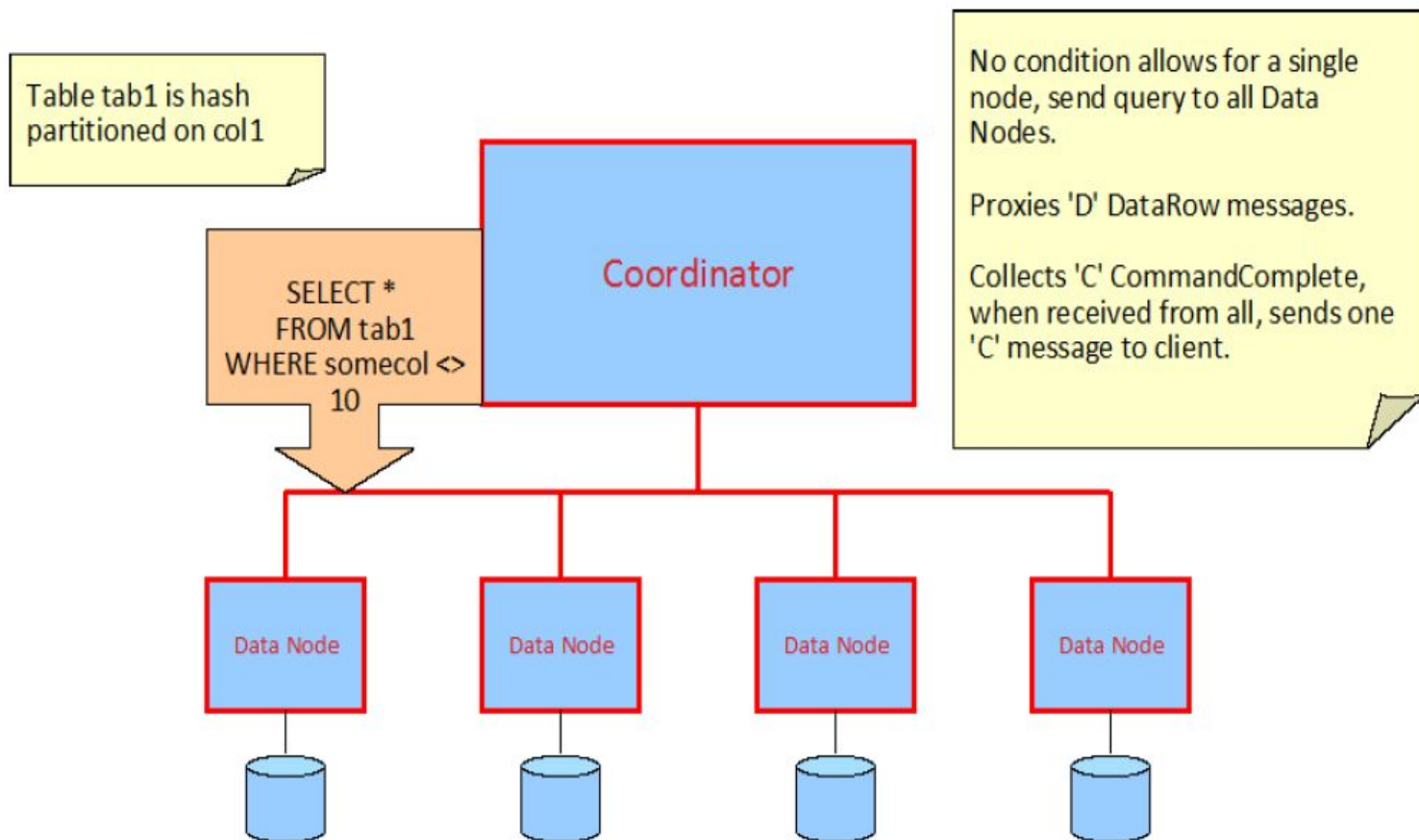
```
Output: test.a, test.ctid, xc_node_id
```

```
Node/s: node_dn2
```

```
Remote query: DELETE FROM public.test WHERE (a = 100)
```

```
(4 rows)
```

多节点处理

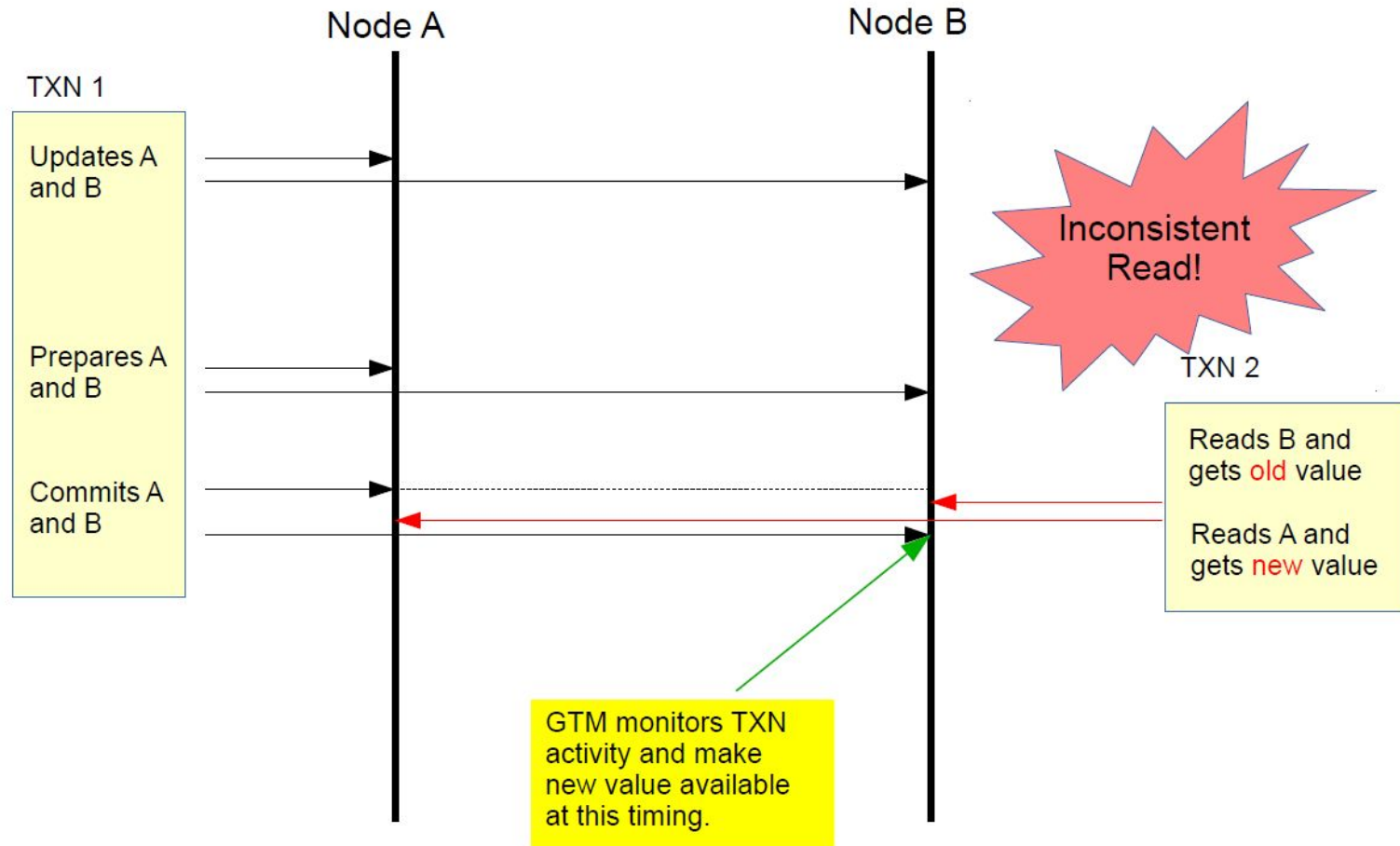


MPP方面的优化

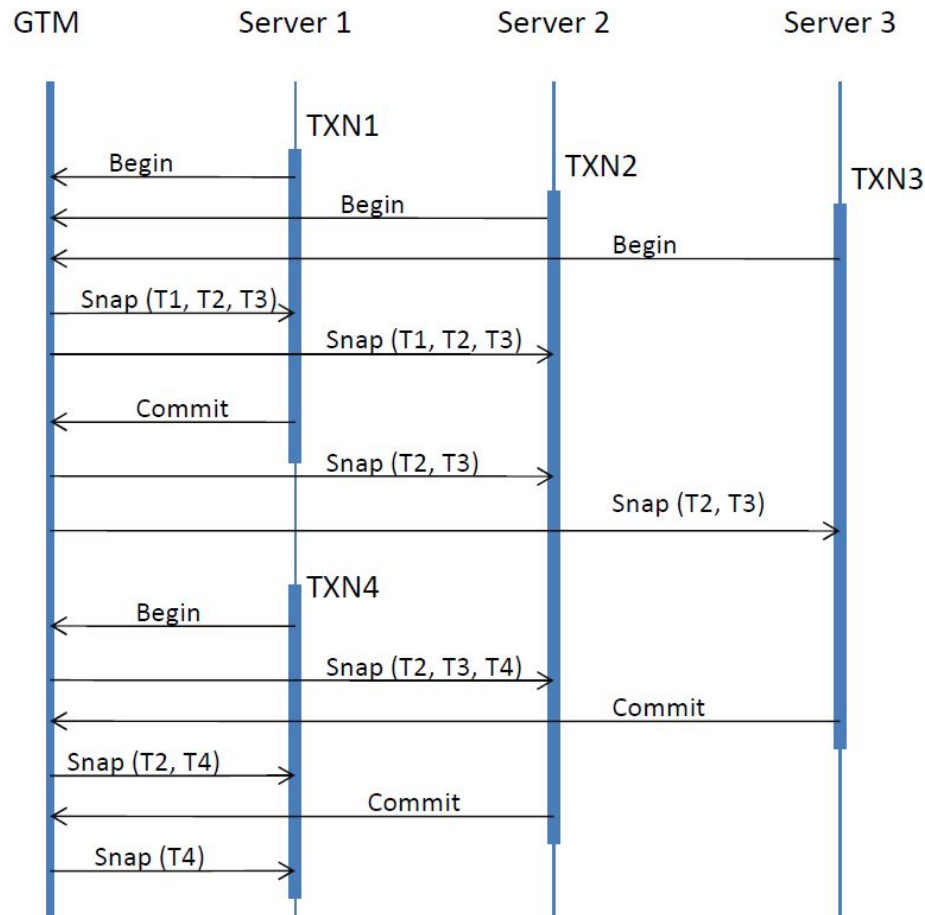
- 数据节点间的数据传输
 - 原来的*Postgres-XC*的*D-Node*间不能传数据
 - 数据需要汇聚到*C*节点进行处理
 - *Postgres-XL*允许*D-Node*间进行数据传输
- 执行计划.VS. SQL语句
 - *C-node* → *D-node* *XC*发送SQL语句, *XL*发送执行计划

Postgres-XL

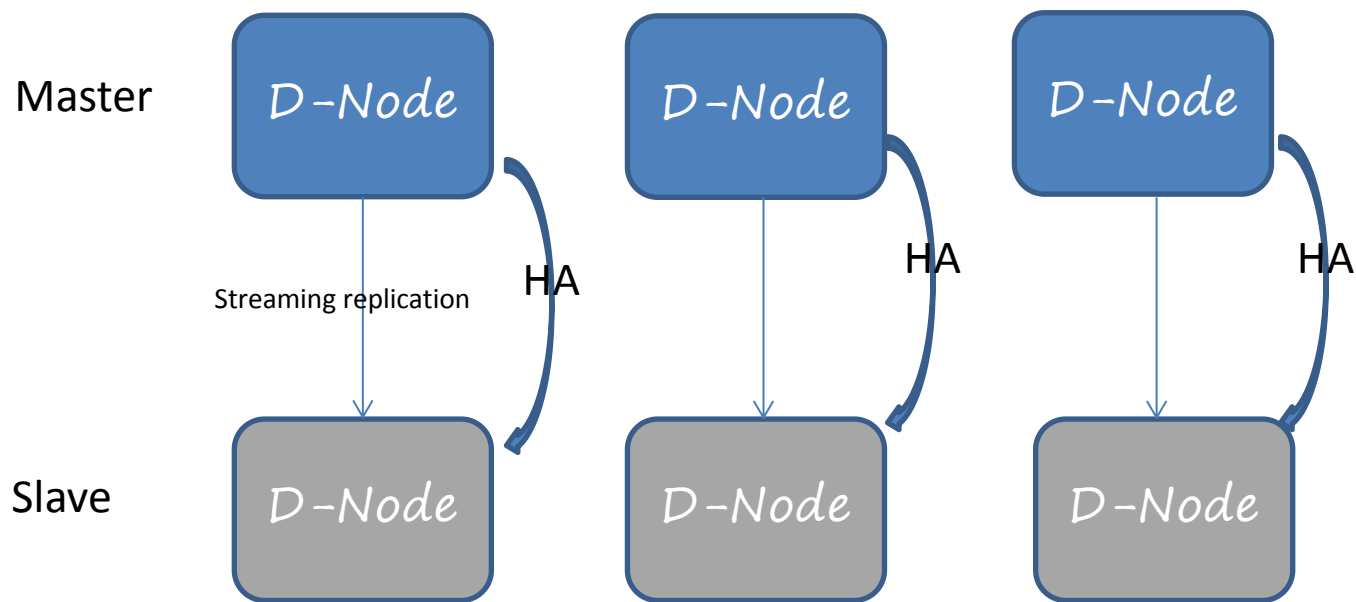
事务处理：2PC就可以了嗎？



GTM全局事务状态的处理

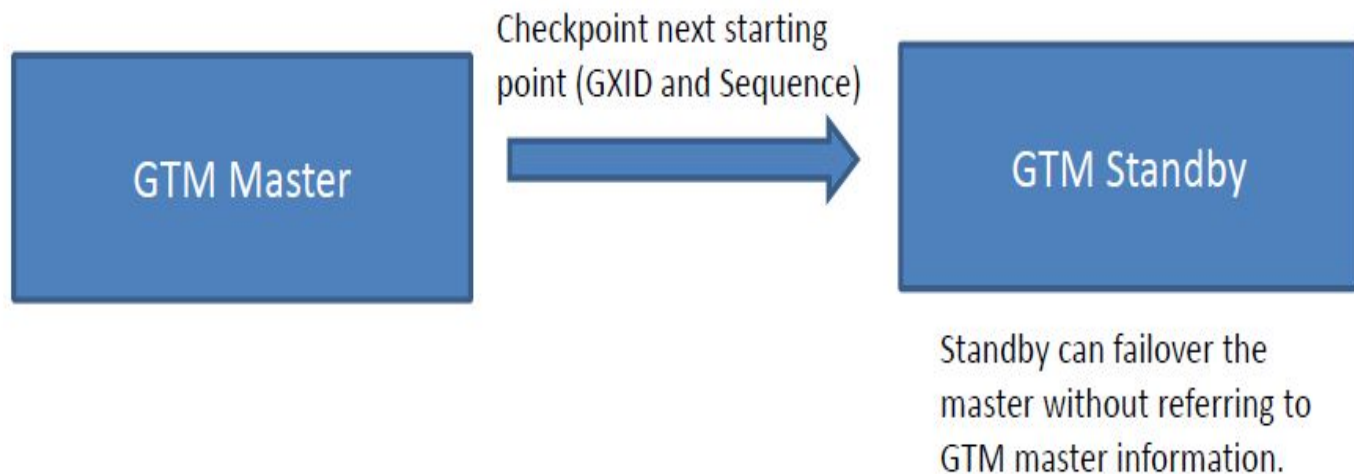


高可用的设计 - *D-Node*



高可用的设计 - *GTM*

- Simple to implement GTM standby



为什么可以处理混合负载

- *OLTP*

- 强事务一致性
- 多主节点可以应付高并发
- 全功能

- *OLAP*

- 多节点并行处理、*Share nothing*
- *MPP*

目录

背景介绍

*Postgres-X2*的架构及技术

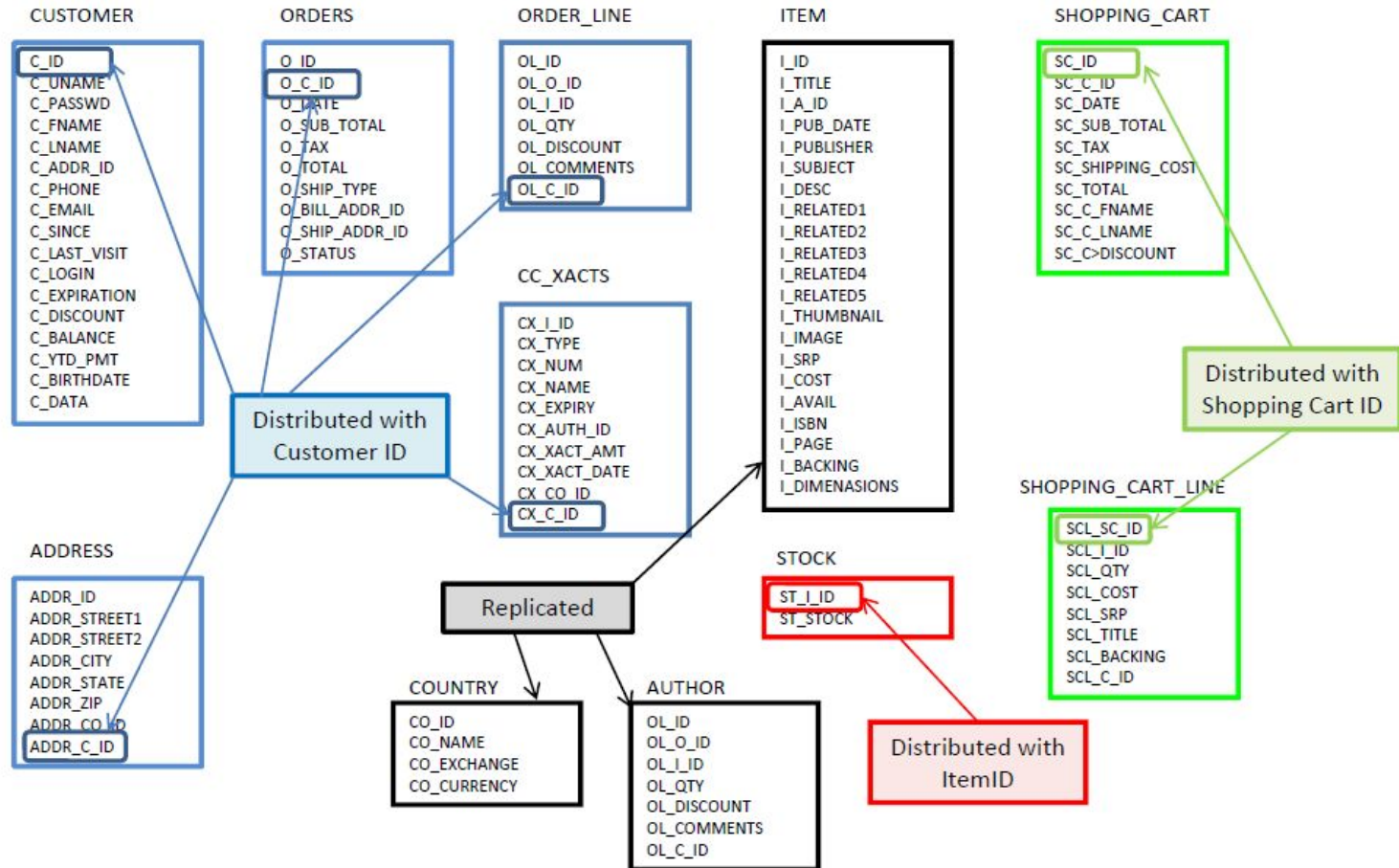
测试及性能结果

其他

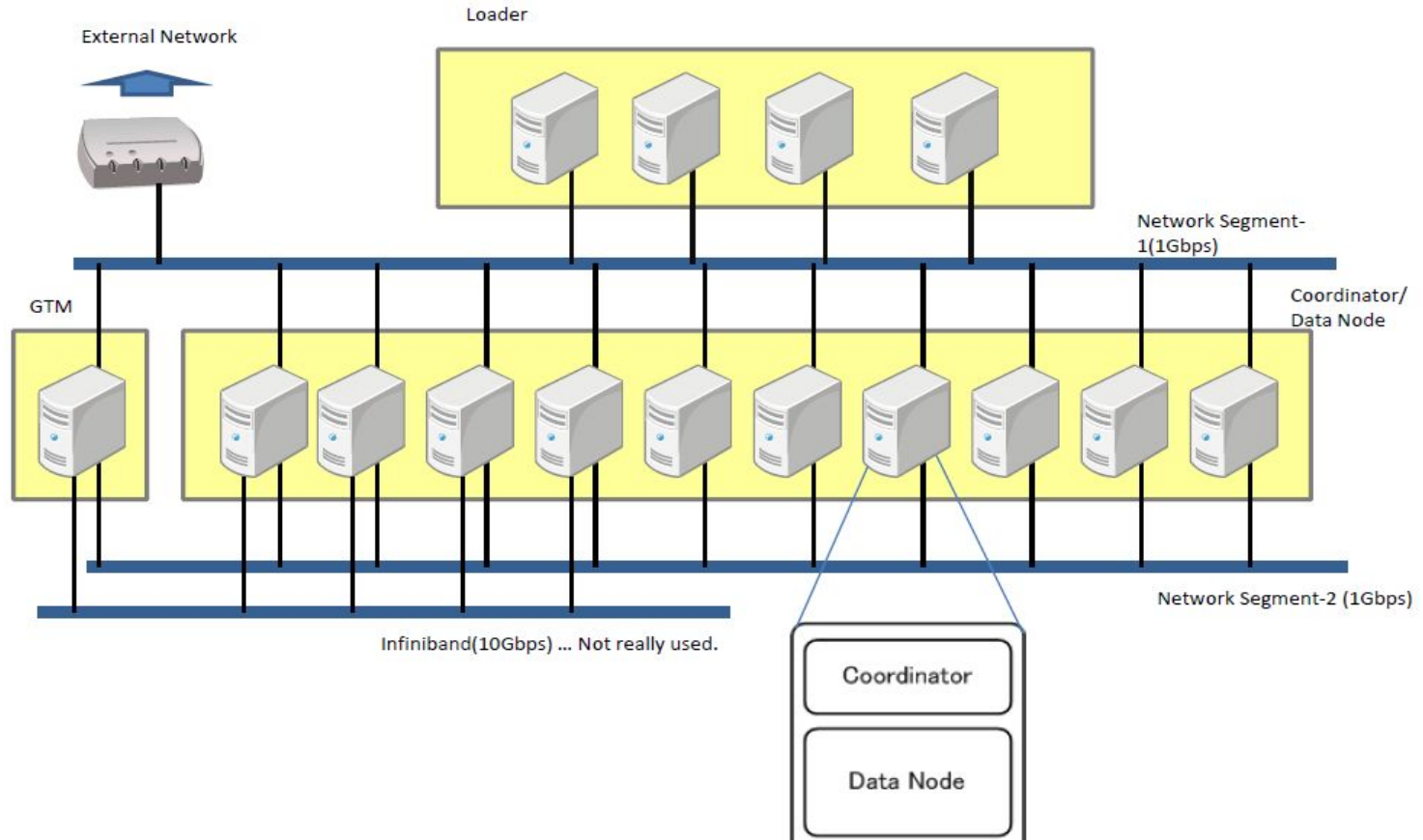
基准测试模型

- *DBT-1* 是 *TPC-W* 的开源版
 - 模拟网店的在线交易
 - 主要行为为用户浏览网站、网上购物：
primarily shopping (WIPS), browsing (WIPsb) 以及 web-based ordering (WIPSo).
- *DBT-3* 是 *TPC-H* 的开源版
 - 数据仓库

DBT-1的表结构



硬件拓扑结构



OLTP的测试性能

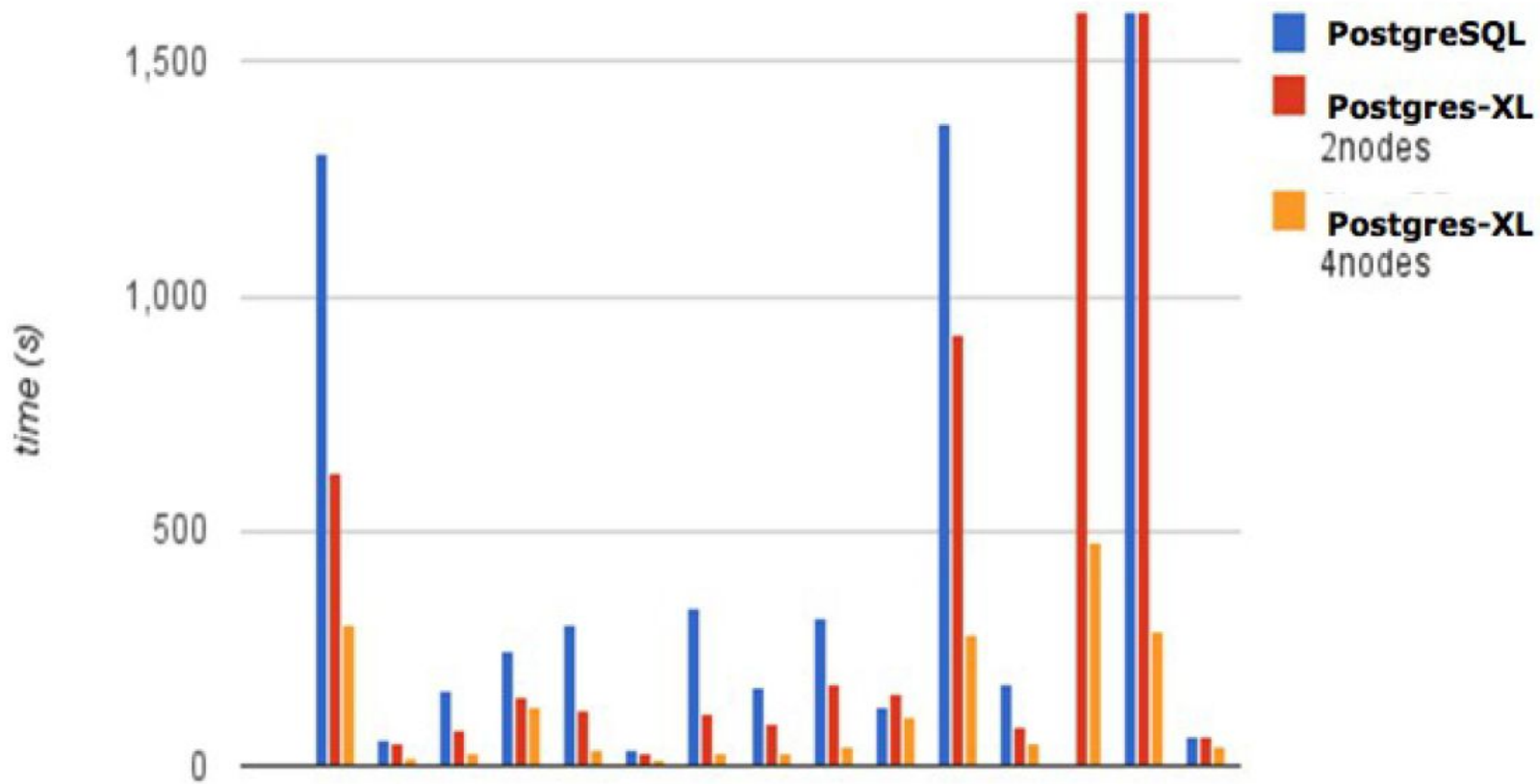
Full Load Throughput

Database	Num. of Servers	Throughput (TPS)	Scale Factor
PostgreSQL	1	2,617	1.0
Postgres-XC	1	1,869	0.71
Postgres-XC	2	3,646	1.39
Postgres-XC	3	5,379	2.06
Postgres-XC	5	8,473	3.24
Postgres-XC	10	15,380	5.88

性能基本可以线性扩展



TPC-H的性能



应用场景

- 大规模 $OLTP$ 应用，尤其是企业领域
- 云环境下的弹性伸缩
- $OLTP$ 及 $OLAP$ 混合负载
- 详单查询
- ODS
- ...

目录

背景介绍

*Postgres-X2*的架构及技术

测试及性能结果

其他

相关的资料

- <https://github.com/postgres-x2/>
- <http://www.postgres-xl.org/>

谢谢

galylee@gmail.com