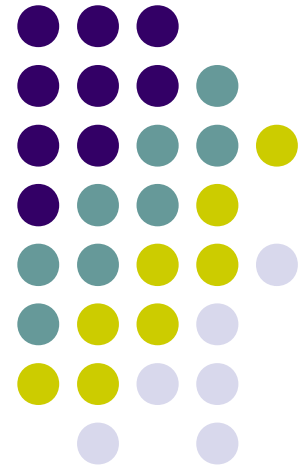


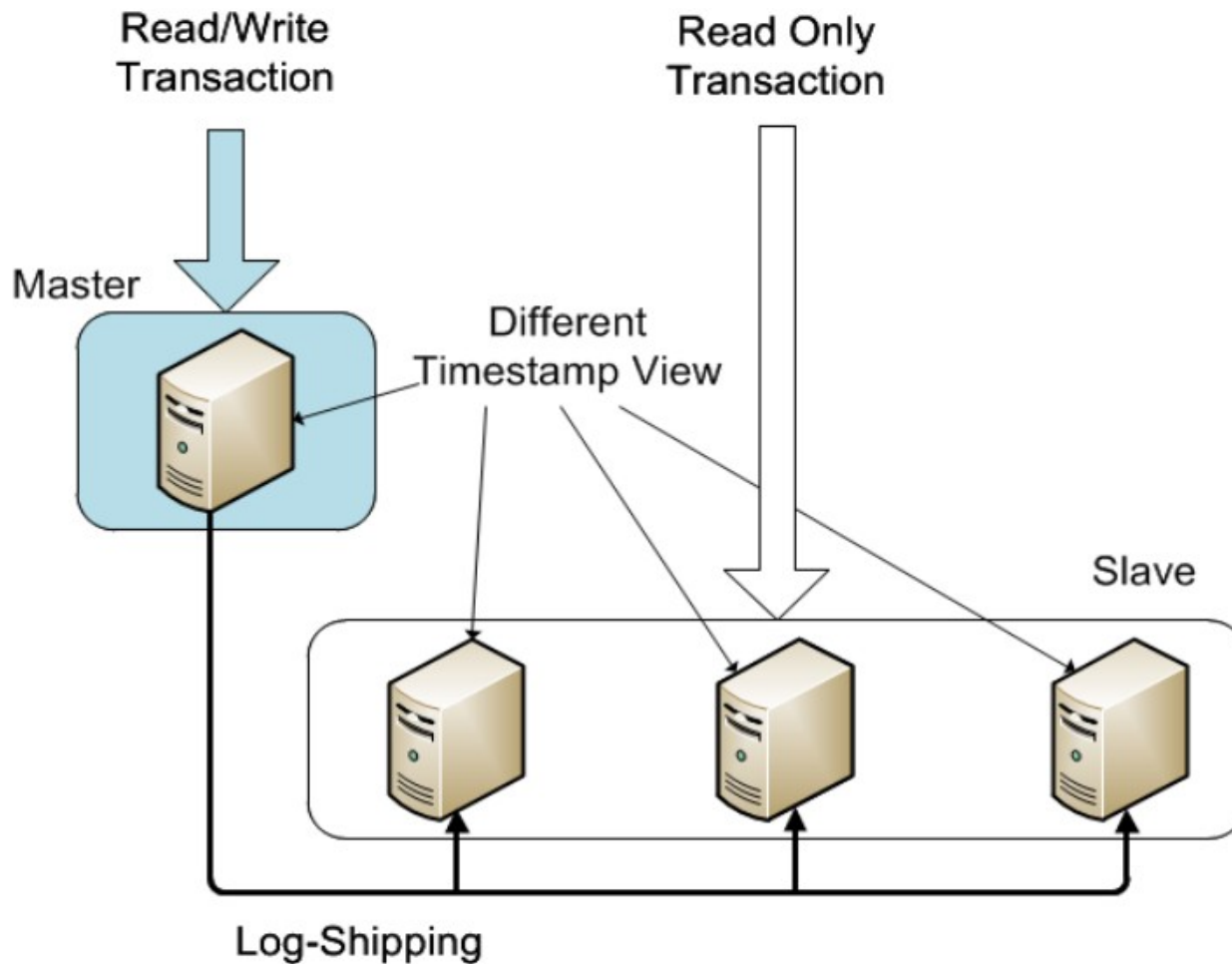
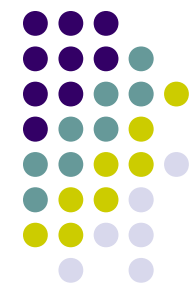
Duplikacja i replikacja w PostgreSQL

Jakub Biernacki
Bartek Lubikowski
Paweł Lesiecki



REPLIKACJA

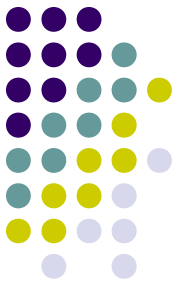
(master-slave replication)



LOG-SHIPPING

◆ file-based log shipping

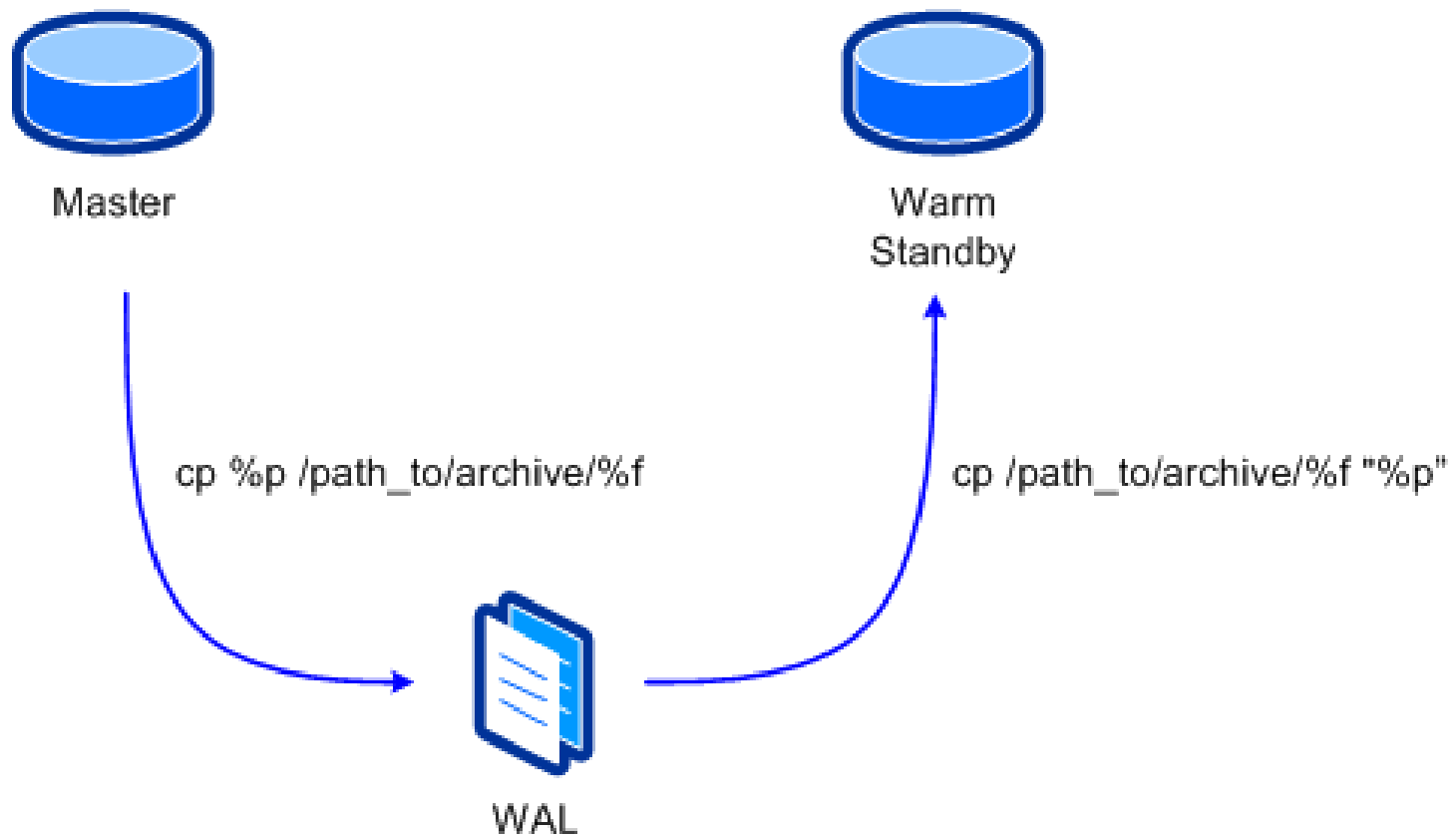
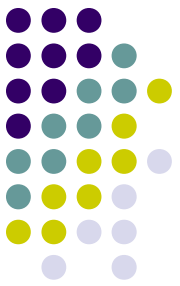
- a.k.a log shipping
- jeden plik WAL (16 MB) po całkowitym zapełnieniu
- minimalne obciążenie serwera głównego
- skalowalne (ze względu na powyższy podpunkt)
- PostgreSQL 8.3 ≤
- Parametry



MASTER	SLAVE
postgresql.conf: <ul style="list-style-type: none">- wal_level- archive_mode- archive_command- archive_timeout- max_wal_senders	recovery.conf: <ul style="list-style-type: none">- standby_mode- restore_command

LOG-SHIPPING

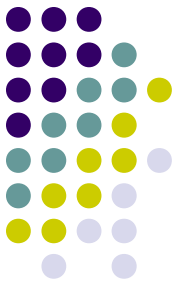
- ◆ file-based log shipping



LOG-SHIPPING

◆ record-based log shipping

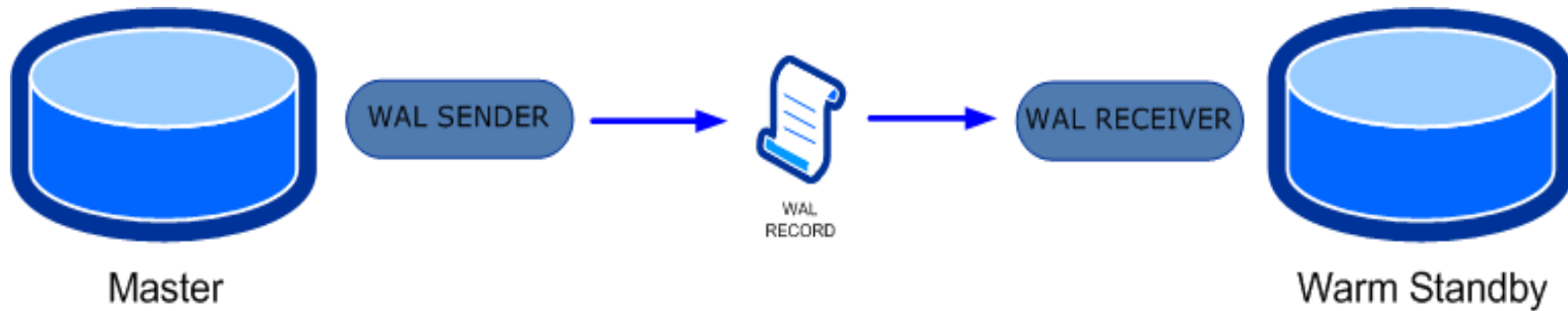
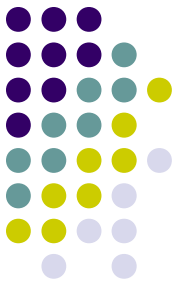
- a.k.a streaming replication
- rekord pliku WAL po każdej zatwierdzonej transakcji
- szybsza aktualizacja niż w file-based shipping
- wymaga aktywnego połączenia (np. TCP)
- procesy walsender, walreciver
- większe obciążenie
- PostgreSQL 9.0 ≤



MASTER	SLAVE
postgresql.conf: <ul style="list-style-type: none">- max_wal_senders- wal_level- archive_command	postgresql.conf: <ul style="list-style-type: none">- hot_standby recovery.conf: <ul style="list-style-type: none">- primary_conninfo- standby_mode

LOG-SHIPPING

- ◆ record-based log shipping

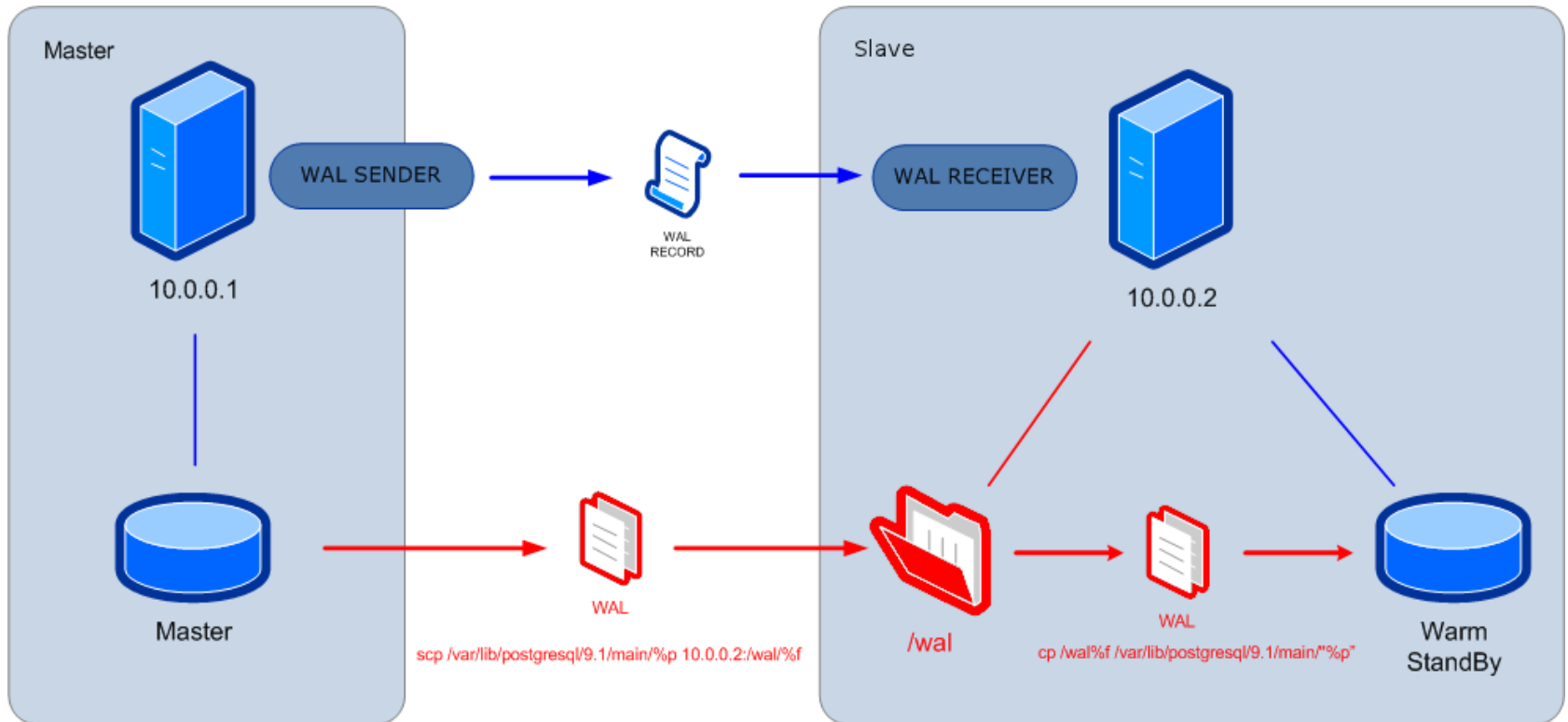
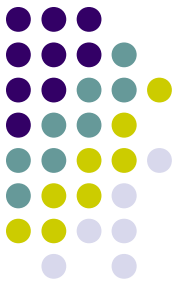


WARM STANDBY vs. HOT STANDBY



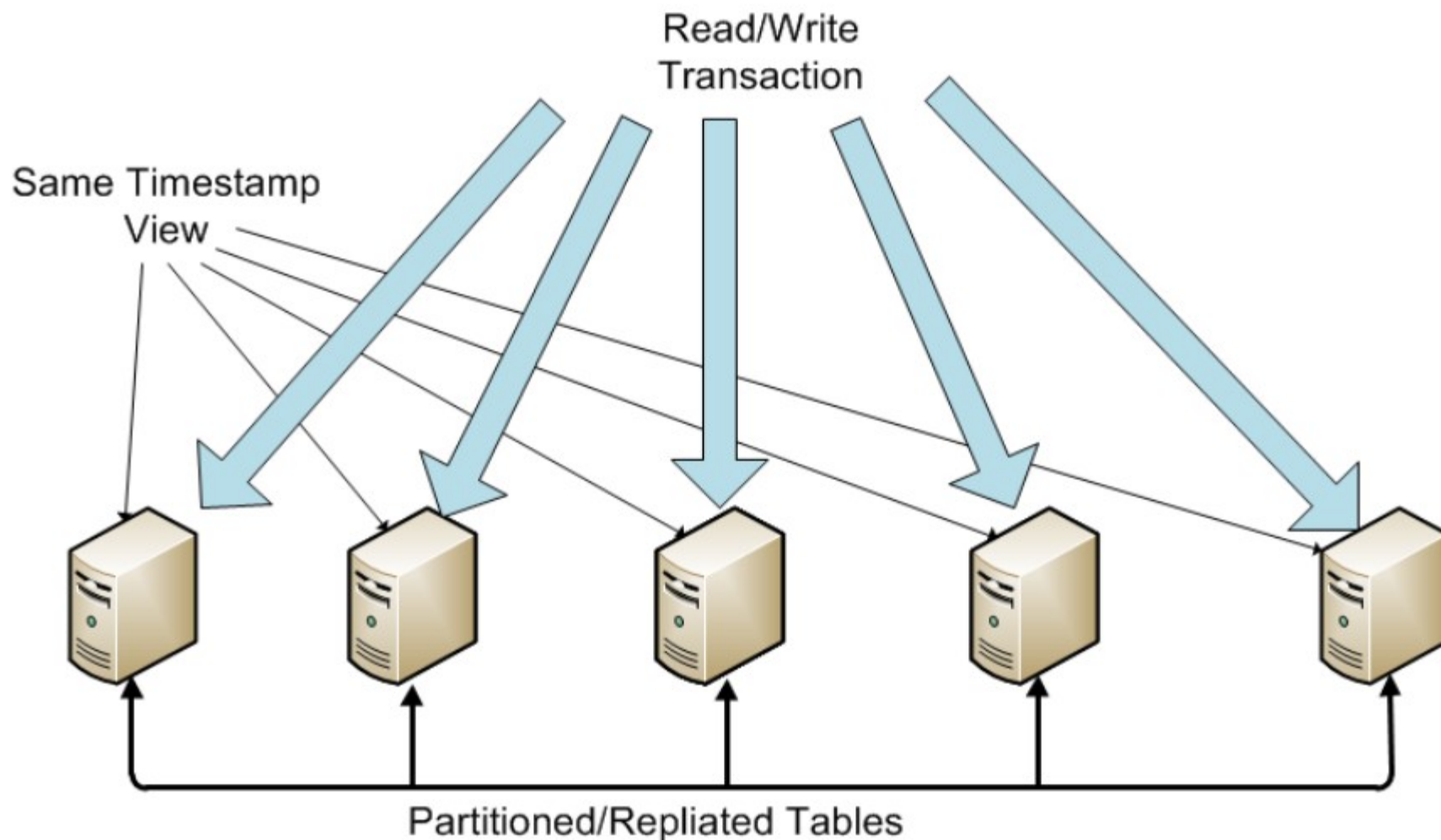
WARM	HOT
<ul style="list-style-type: none">- kojarzone wraz file-based log shipping- nie można wysyłać żadnych zapytań (nawet <i>read-only</i>)	<ul style="list-style-type: none">- kojarzone wraz record-based log shipping- można się połączyć i korzystać z baz w trybie read-only- ze względu na powyższe można zastosować load-balancing
<ul style="list-style-type: none">- replikacja działa dla całych klastrów (nie możemy wybrać poszczególnych baz lub tym bardziej ich obiektów)- maszyny muszą być oparte o tę samą architekturę (32/64bit) oraz tę samą główną wersję PostgreSQL	

WARM STANDBY vs. HOT STANDBY



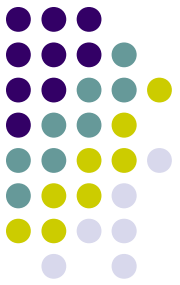
DUPLIKACJA

(master-master/multi-master replication)



DUPLIKACJA

PostgresXC

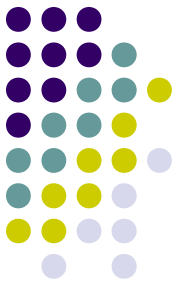


Multi-master PostgreSQL extensible cluster

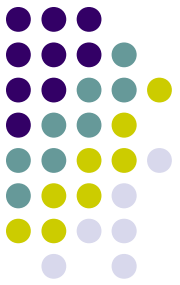
DUPLIKACJA

PostgresXC

- Klaster bazodanowy oparty na PostgreSQL
- Działa z najnowszymi wersjami PostgreSQL
- Więcej niż replikacja
- Odczyt / zapis do dowolnego serwera
- Równoległe transakcje/obsługa instrukcji



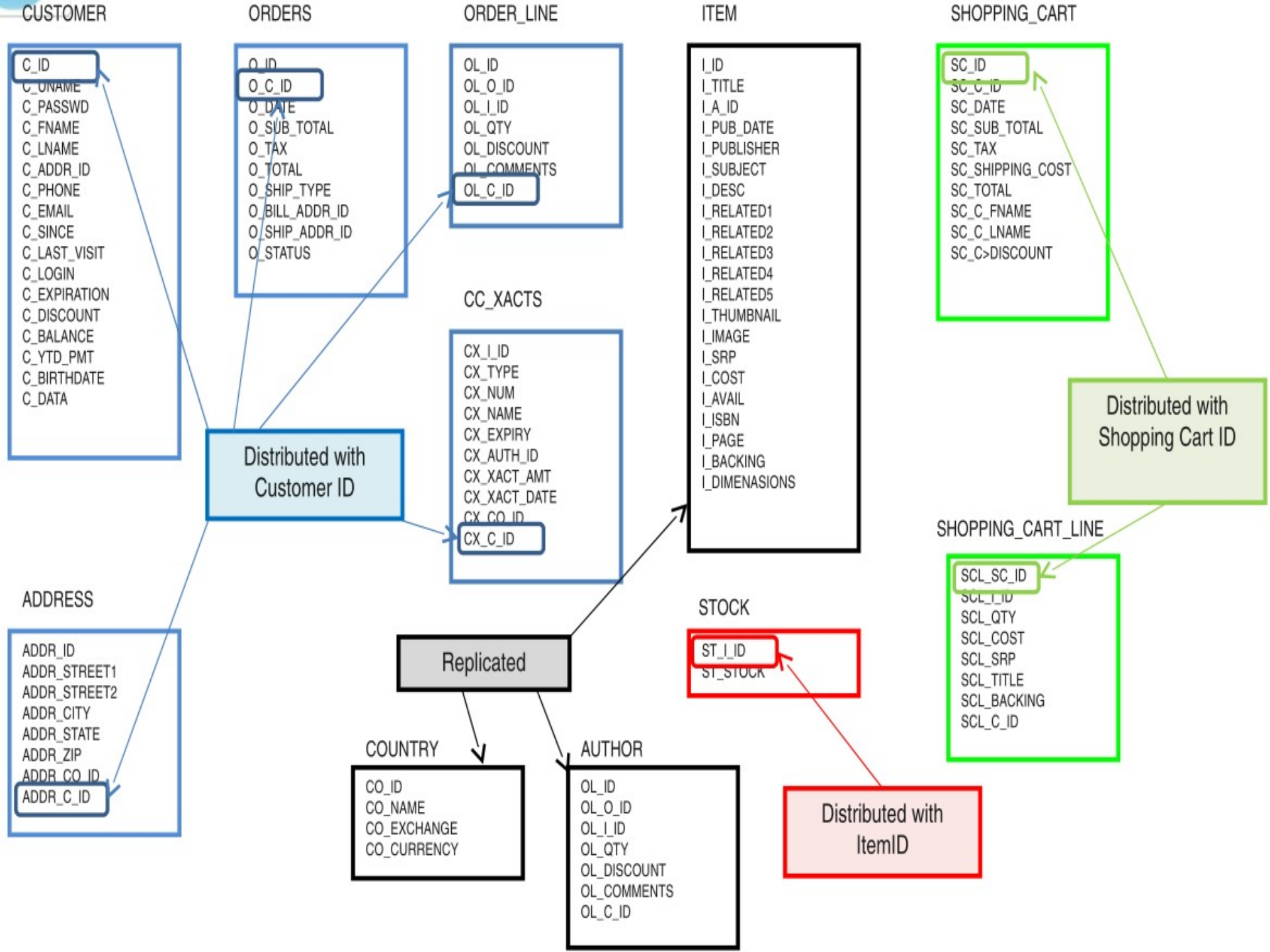
Dystrybucja i replikacja tabel

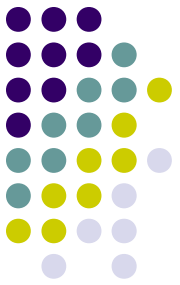


Każda tabela może być rozproszona/replikowana

Nie jest prosta replikacja danych

- Dobrze nadaje się do dystrybucji bazy danych o konstrukcji o schemacie gwiazdy (data warehouse)
 - Tabele transakcyjne → rozpraszane
 - Tabele nadrzędne → replikowane
- Join ze stosem
- Klauzula Where ze stosem
- Równoległa agregacja

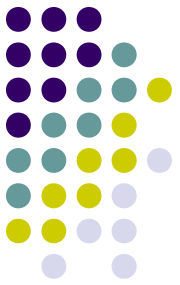




Tworzenie tabel

```
CREATE TABLE tab (...) DISTRIBUTE BY  
HASH(col) | MODULO(col) | REPLICATE
```

Dystrybucja danych



Replikowane tabele

- Każdy wiersz w tabeli jest replikowany do datanodes
- Replikacja oparta na logowaniu instrukcji modyfikujących dane

Writes



write write write



val	val2
1	2
2	10
3	4



val	val2
1	2
2	10
3	4



val	val2
1	2
2	10
3	4

Reads



read



val	val2
1	2
2	10
3	4

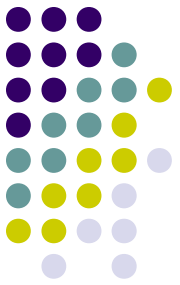


val	val2
1	2
2	10
3	4



val	val2
1	2
2	10
3	4

Dystrybucja danych



Rozproszone tabele

- Każdy wiersz tabeli jest przechowywany na jednym datanode, O tym gdzie, decyduje jedna z następujących strategii
 - Hash
 - Round Robin
 - Modulo
 - Range, UDF

Write



write



val	val2
1	2
2	10
3	4

val	val2
11	21
21	101
31	41

val	val2
10	20
20	100
30	40

Read



read

read

read

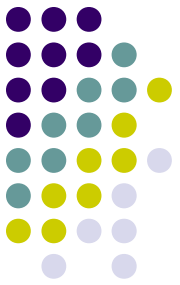


val	val2
1	2
2	10
3	4

val	val2
11	21
21	101
31	41

val	val2
10	20
20	100
30	40

Komponenty



1. Coordinator

- punkt połączenia z aplikacjami
- analiza SQL i globalne planowanie
- globalna realizacja SQL

2. Datanode

faktyczny magazyn bazy danych
lokalna realizacja SQL

3. GTM (Global Transaction Manager)

Zapewnia transakcjom spójny obraz danych

- GXID (Globalne ID transakcji)
- snapshot (lista aktywnych transakcji)
- inne globalne wartości, takie jak SEQUENCE

4. GTM Proxy

- Grupuje połączenia i interakcje między GTM i innymi komponentami Postgres-XC, aby zmniejszyć liczbę interakcji i rozmiar komunikatów

Działa jak PostgreSQL
Określa, do których Datanodes przejść
Planuje rozproszone kwerendy
Dostarcza GXID i Snapshoty do Datanodes

Obsługuje lokalne instrukcje od koordynatorów
Więcej niż jeden PostgreSQL?

Zmniejsza interakcję z GTM poprzez grupowanie żądań/odpowiedzi od GTM
Dba o połączeniu do GTM

Dostarcza globalny identyfikator transakcji oraz Snapshoty
Dostarcza sekwencję

