



Результаты тестирования СУБД PostgreSQL на Power

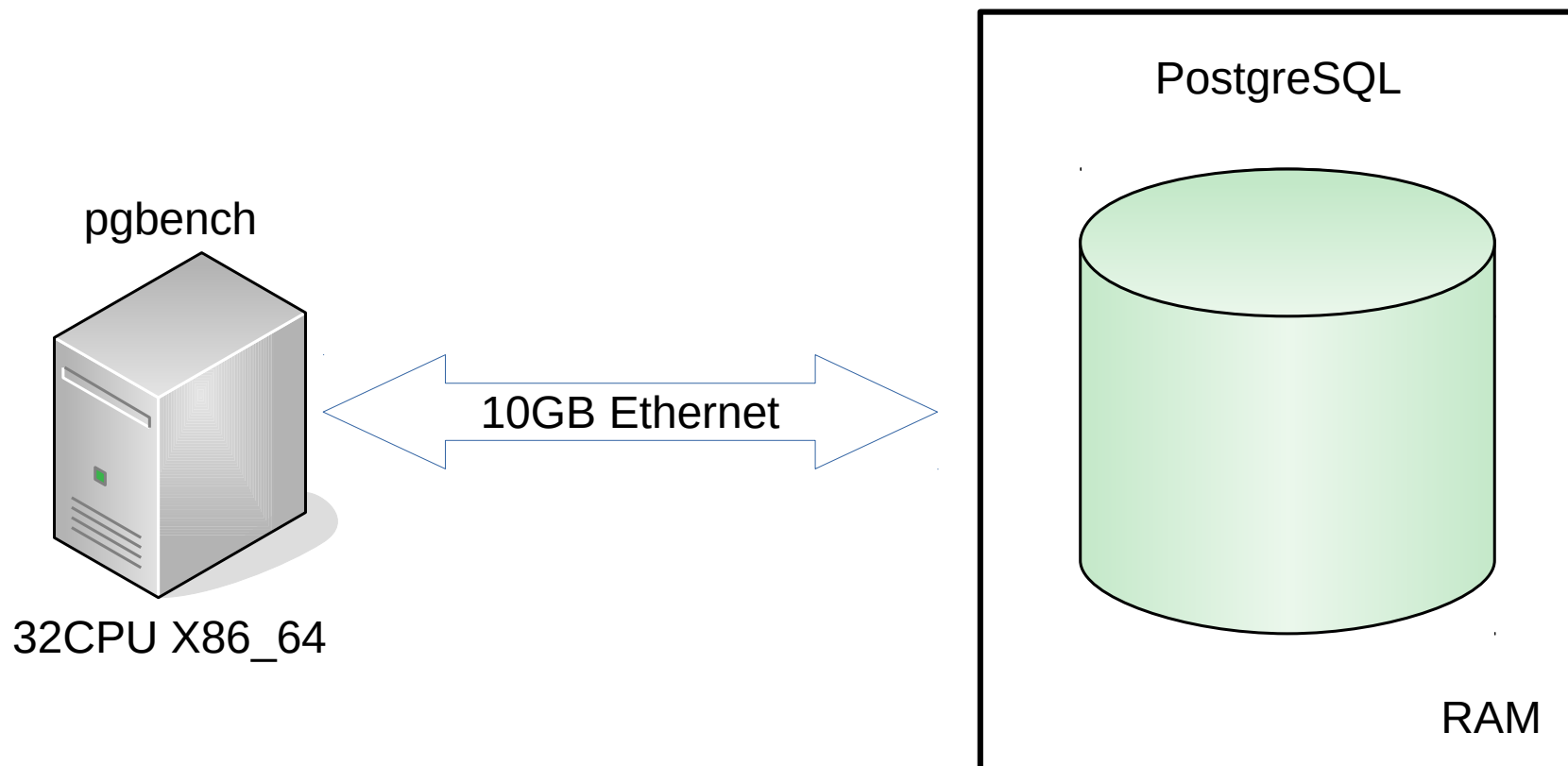
Васильев Дмитрий

IBM E880

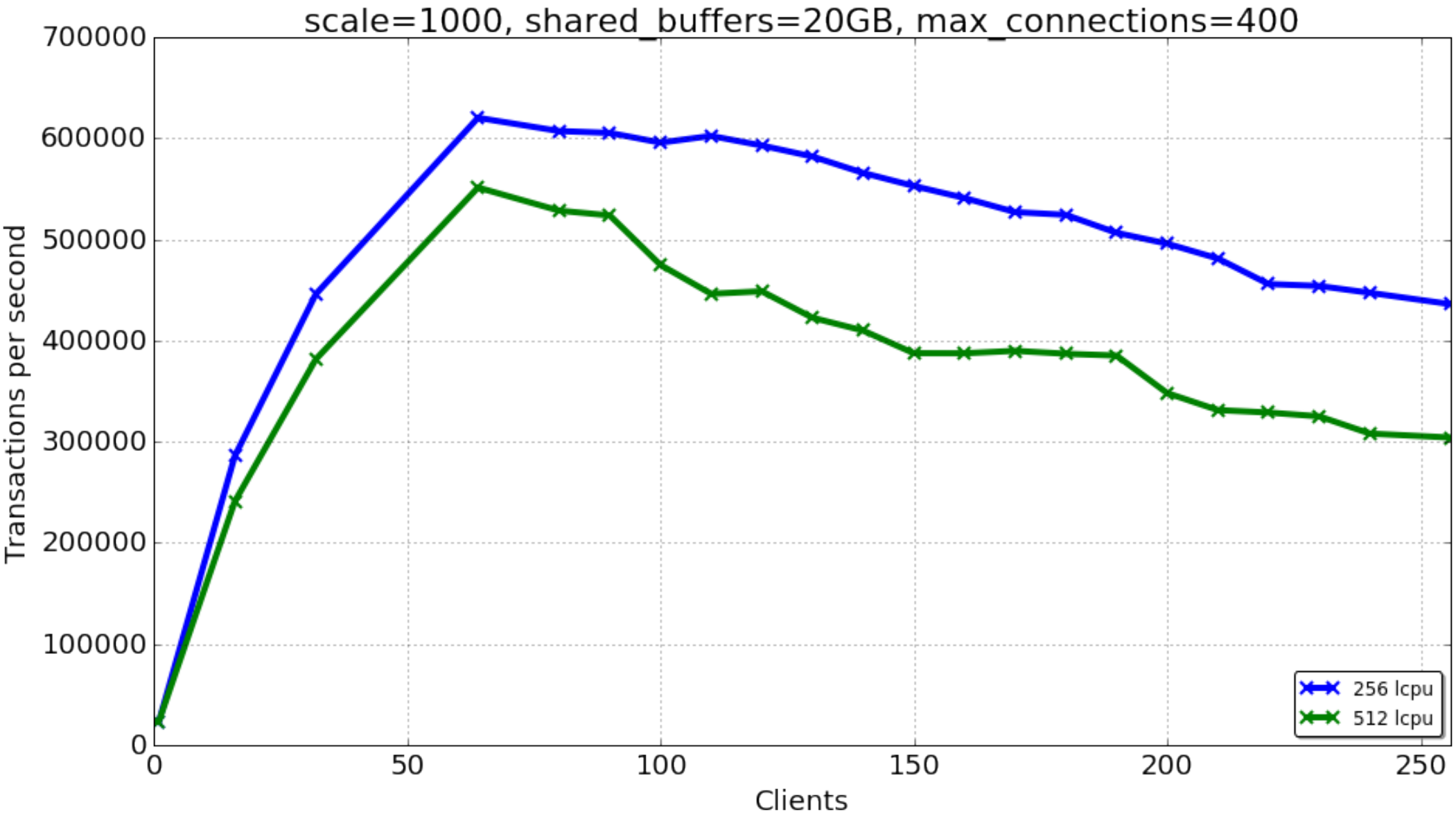


- 2 x 2U Ноды
- Ноды соединены «HyperConnect»
- Одна нода: 4 Сокета с Power8
- Один сокет: 8 Ядер (Core)
- Одно ядро: 8 логических ядер (SMT)

Схема тестирования



256 vs 512 LCPU



Главные враги масштабируемости:

- * Время проведенное в блокировке (линейна от t)
- * Синхронизация событий между исполнителями (N^2)

```
db=# \d+ pg_locks
```

```

View "pg_catalog.pg_locks"
  Column          | Type          | Modifiers | Storage |
-----+-----+-----+-----+
locktype         | text         |           | extended|
database         | oid          |           | plain   |
relation         | oid          |           | plain   |
page             | integer      |           | plain   |
tuple            | smallint     |           | plain   |
virtualxid       | text         |           | extended|
transactionid    | xid          |           | plain   |
classid          | oid          |           | plain   |
objid            | oid          |           | plain   |
objsubid         | smallint     |           | plain   |
virtualtransaction | text        |           | extended|
pid              | integer      |           | plain   |
mode             | text         |           | extended|
granted          | boolean      |           | plain   |
fastpath         | boolean      |           | plain   |

```

Локи в PostgreSQL

- Тяжелые локи (HWLock)
- Легкие локи (LWLock)
- Spin Lock

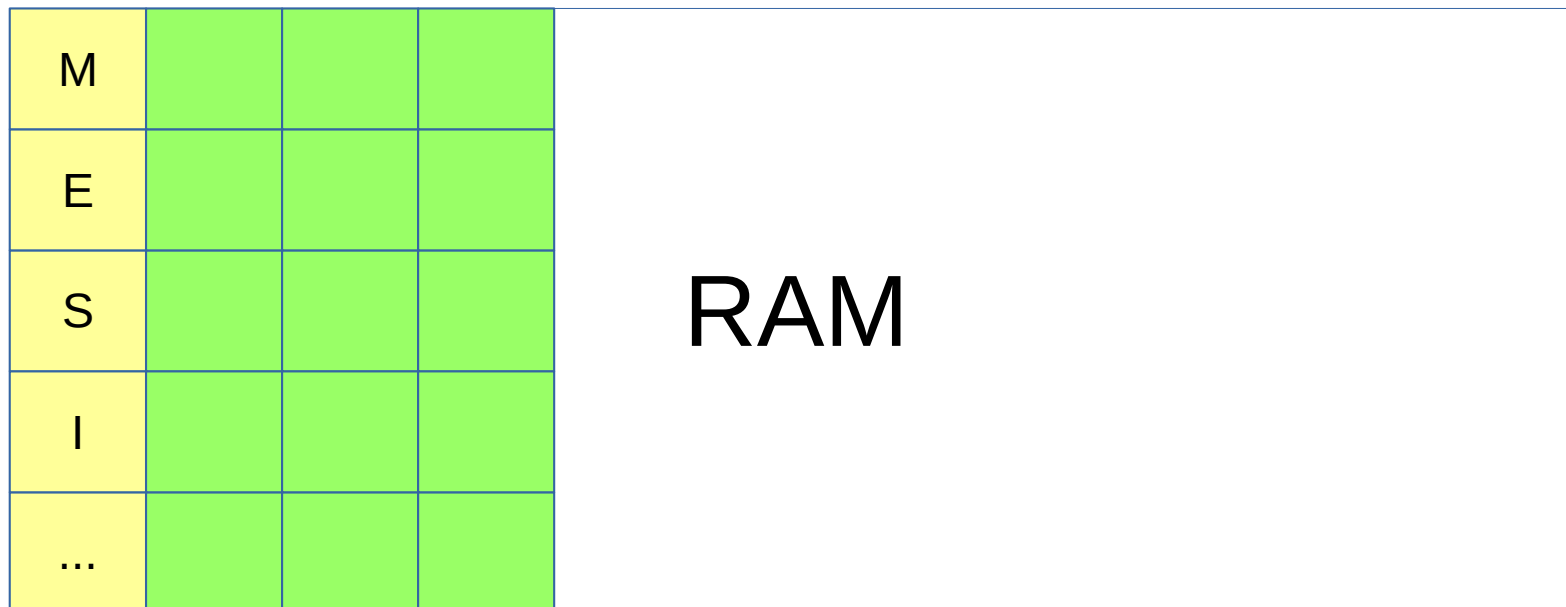
Также: Row-Level, Predicate, Advisory Locks

- Работают на уровне объектов
- Логические блокировки
- Детектор dead-lock

- Очереди: Shared / Exclusive
- Нет dead-lock детектора
- В 9.5 избавились от SpinLock

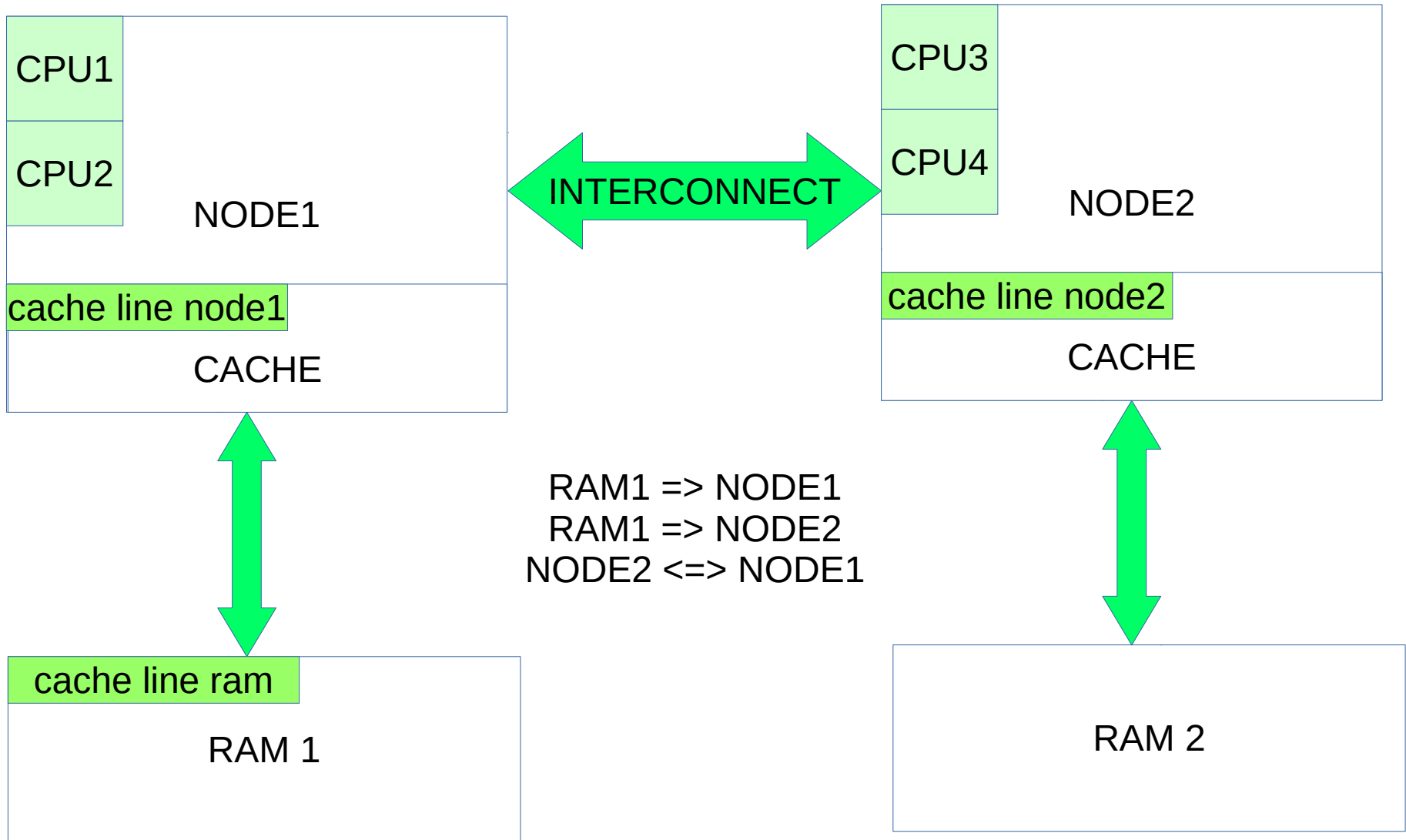
- Для коротких операций
- Только эксклюзивный режим
- Нет детектора dead-locks
- Нет очередей

Доступ к RAM (несколько CPU)



- * CPU читает из основной памяти страницами cache-line
- * Каждая cache-line кроме данных имеет тэг хранящее ее состояние
- * Тэг меняет/читает CPU при обращении или записи
- * Тэг может принимать значения: Modified, Exclusive, Shared, Invalid ...

Доступ к нескольким RAM через несколько Numa Nodes

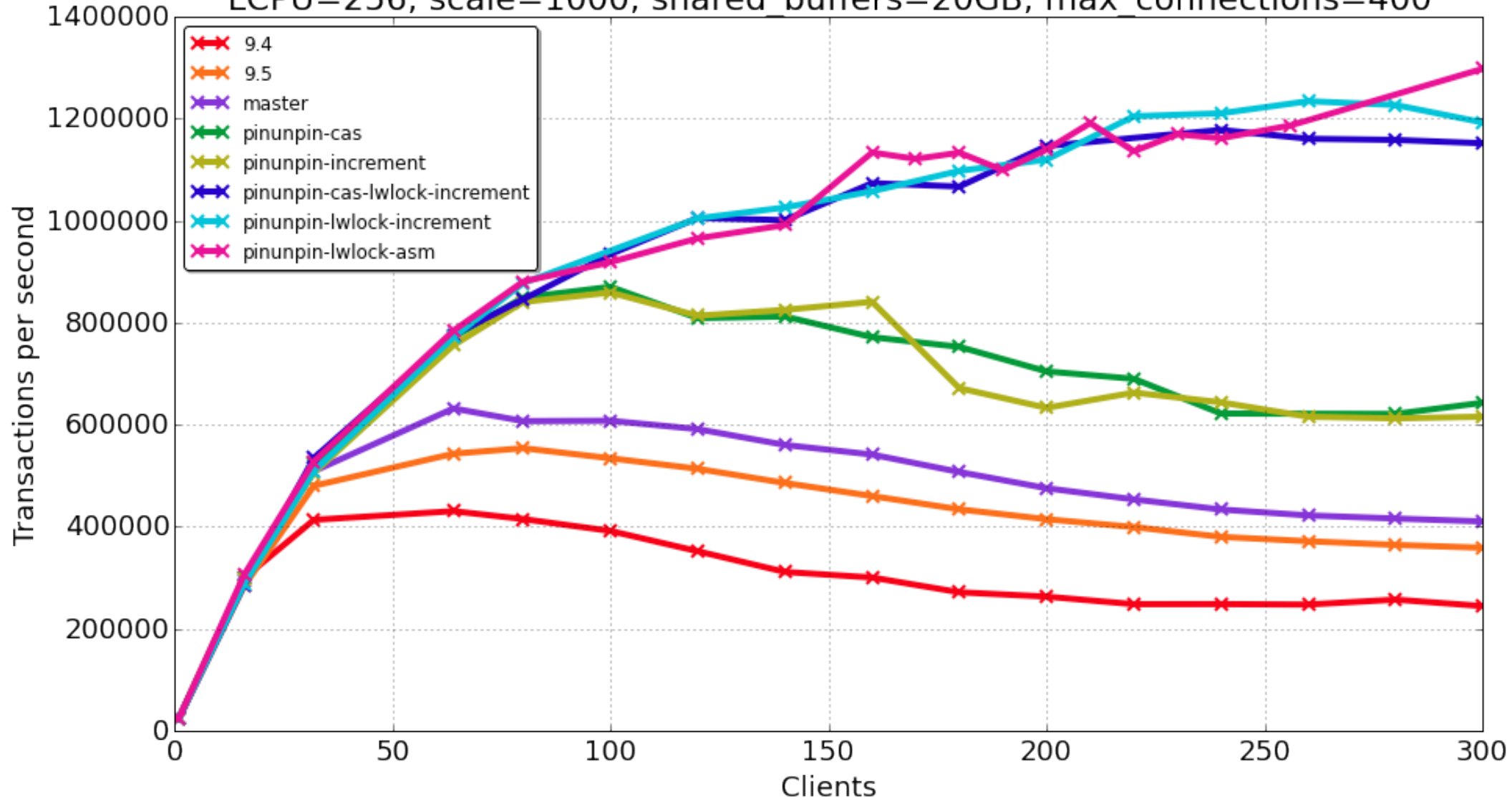


Плюсы Power8

- Быстрый “interconnect”
- Больше кол-во атомарных операций чем у конкурентов на сопоставимом по цене железе
- Уменьшение издержек, связанное на синхронизацию памяти
- Лучшая масштабируемость

PostgresPro патчи для 9.5

LCPU=256, scale=1000, shared buffers=20GB, max connections=400



pinunpin-cas – PinBuffer: CAS

pinunpin-increment – PinBuffer: оптимистичный AtomicAdd

pinunpin-cas-lwlock-increment – PinBuffer: CAS, LWLockAttemptLock оптимистичный AtomicAdd

pinunpin-lwlock-increment – PinBuffer и LWLockAttemptLock: оптимистичный AtomicAdd

pinunpin-lwlock-asm – PinBuffer и LWLockAttemptLock: оптимизированный asm

- LPAR – мощная и прозрачная виртуализация
- Эффективная работа на Power8 с PgPro 9.5
- Время не стоит на месте и есть место для маневра на Power (уже сейчас транзакционная память, в будущем NVLink, ...)



Спасибо за внимание!