

Миграция на СУБД PostgreSQL/Postgres Pro с многоядерными серверами Bull

Реальный опыт

06-02-2019

Trusted partner for your Digital Journey



Atos

1

**Серверы Bull – решения для
больших внедрений
PostgreSQL/Postgres Pro**

Bull - технологический бренд Atos

Более 80 лет инноваций

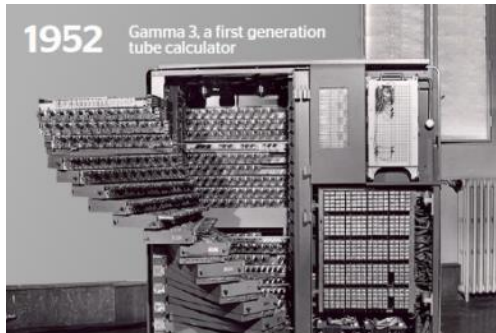


Honeywell Bull

1923 Fredrik Rosing Bull tabulating machine



1952 Gamma 3, a first generation tube calculator



Bull
atos technologies

1963 Bull Gamma 10 electronic data processing system



1990 Gcos, combining the best of the mainframe and the open world



2009 bullx, the world's best HPC server product



2014 bullion, the emblematic big data product



Серверы Bullion S

Большой модельный ряд серверов на базе процессоров семейства Intel® Xeon® E7-v4



Серверы Bullion S на базе процессоров семейства Intel® Xeon® E7

S2



S4



S8



S16



2 ЦПУ

4 ЦПУ

8 ЦПУ

16 ЦПУ

- До 48 ядер
- До 3 ТБ памяти
- 7 разъемов PCI-e

- До 96 ядер
- До 6 ТБ памяти
- 14 разъемов PCI-e

- До 192 ядер
- До 12 ТБ памяти
- 28 разъемов PCI-e

- До 384 ядер
- До 24 ТБ памяти
- 56 разъемов PCI-e

BullSequana S: линейка мощных серверов

На новейших процессорах Intel® Xeon® Scalable



Расширение до 32 ЦПУ



До 320 дисков



До 48 ТБ



До 96 NVMe



До 64 ТБ



До 32 графических процессоров

АЕР

- Модуль с 2 ЦПУ в стойке 2U
- До 3 ТБ ОЗУ на каждые 2 ЦПУ
- Упрощенное обслуживание

2

**Тестирование
СУБД PostgreSQL
на серверах Bullion S**

Bullion S против аналогичных NUMA-серверов

Сравнение производительности при стрессовом тестировании СУБД PostgreSQL

	Bullion S8	Bullion S16	Сервер с архитектурой NUMA
ЦПУ	8 x Intel Xeon E7-8880 v4 с частотой 2.20 ГГц	16 x Intel Xeon E7-8880 v4 с частотой 2.20 ГГц	8 x Intel Xeon E7-8880 v4 с частотой 2.20 ГГц
ОЗУ	4 ТБ	4 ТБ	4 ТБ

Процедура тестирования:

- Postgres запускался на различных нодах с использованием команды numactl.
- Тест pgbench – на нодах, на которых не работал postgres. (исключалось влияние процессов БД на процессы теста pgbench)
- В случае теста All CPU тесты Pgbench и Postgres разделяли ЦПУ между собой

Тестирование

- ▶ Необходимо помнить о numa. Interleave между numa-нодами
- ▶ Большой shared buffer может быть проблемой
- ▶ Большое количество процессоров=> большой max_connection
- ▶ Не очень честный LWLock

Нагрузочные тесты

Подход 1

- ▶ Pgbench (insert into table(1000)+select from table(1000))
- ▶ Количество ht ядер 704
- ▶ Max_connection ~ 2000. А кто запретит на такой машинке?
- ▶ Insert по 10 минут и больше %)
- ▶ Tps insert 300 tps, select 500k tps
- ▶ Читатели задавили писателей

Нагрузочные тесты

Подход 2

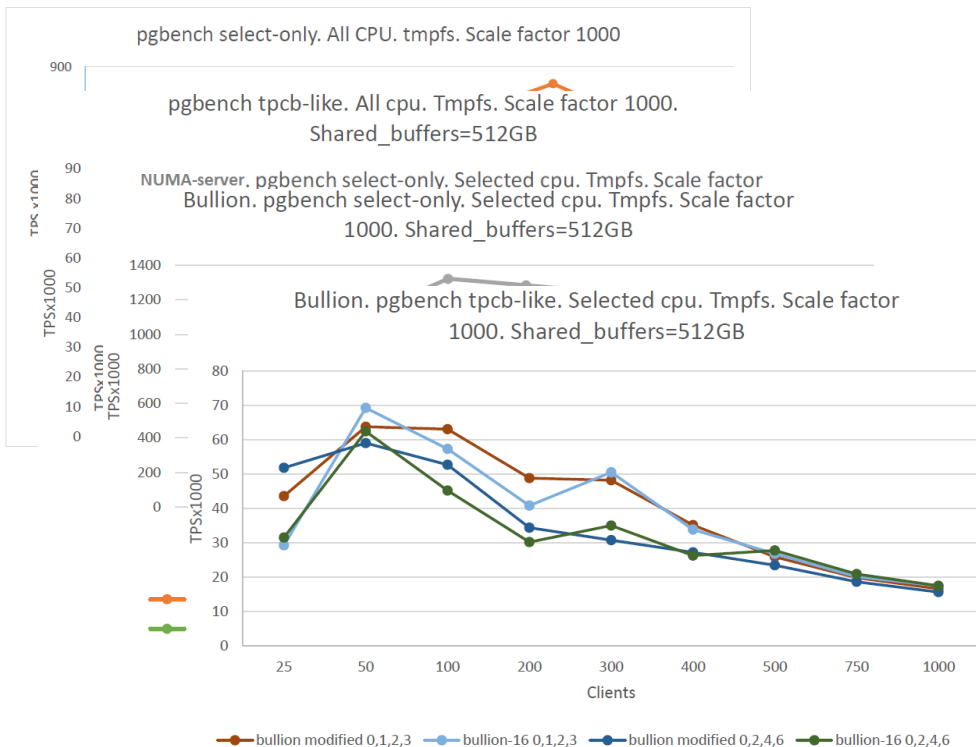
- ▶ Нужен патч чтобы за Postgres стыдно не было
- ▶ Патч от Александра Короткова
- ▶ Tps по insert выровнялся
- ▶ Благодарные читатели

Практика

- ▶ Загрузка 10%
- ▶ “In memory” Postgres – 4 ТВ памяти 4 ТВ база %)
- ▶ Тысячи пользователей системы

Bullion S против аналогичных NUMA-серверов

Стадии тестирования



- 1) Производительность серверов в тесте только на чтение в зависимости от количества клиентов
- 2) Производительность серверов в тесте tpcb-like в зависимости от количества клиентов
- 3) Производительность конкурентного сервера в тесте select-only в зависимости от количества клиентов
- 4) Производительность конкурентного сервера в тесте tpcb-like в зависимости от количества клиентов
- 5) Производительность сервера Bullion S (+конфигурация с измененными аппаратными настройками) в тесте select-only в зависимости от количества клиентов
- 6) Производительность сервера Bullion S (+конфигурация с измененными аппаратными настройками) в тесте tpcb-like в зависимости от количества клиентов

* по запросу доступны полноразмерные графики с результатами тестов

Bullion S против аналогичных NUMA-серверов

Результаты тестирования

- 1) У сервера Bullion S видна меньшая в сравнении с конкурентом зависимость производительности от выборки процессоров, на которых запускается БД
- 2) В тестах экстремальной производительности сервер Bullion S8 в некоторых тестах ведет себя несколько лучше конкурентных аппаратных решений, что связано с различной реализацией в построении аппаратного комплекса. Bullion S16 обеспечивает значительный прирост производительности в сравнении с результатами тестирований конкурентного сервера
- 3) Производительность БД на серверах Bullion с 16 ЦПУ практически не деградирует при увеличении количества процессоров в системе.

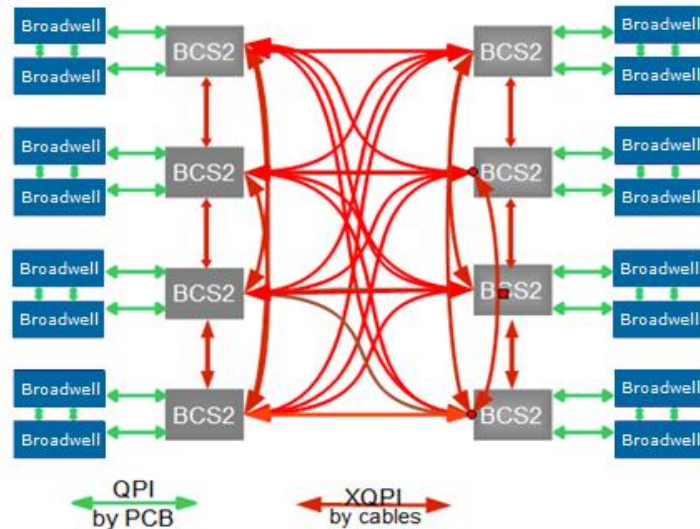
Таким образом, БД в данной конфигурации работоспособна и пригодна для промышленной эксплуатации на серверах Bullion S

Glued-архитектура в серверах Bullion S

Описание и характеристики BCS2

В конфигурациях Bullion S для межсоединения процессоров применяется архитектура Glued:

- Узловой контроллер BCS2
- Межсоединение контроллеров BCS2 через линки XQPI
- Межсоединение ЦПУ и контроллеров через линки QPI
- Обеспечивает рост до 16 ЦПУ
- Один контроллер BCS2 на каждый двухсокетный модуль



Топология сервера Bullion S16

Glued-архитектура в серверах BullSequana S800+

Описание и характеристики UNC

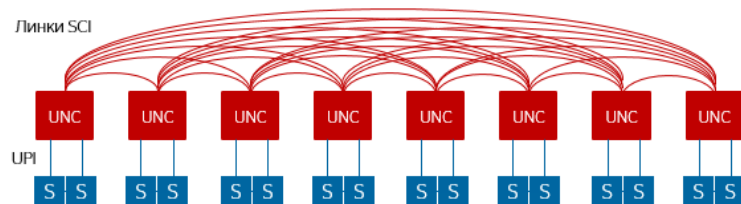
В конфигурациях BullSequana S свыше 8 ЦПУ применяется архитектура Glued (опционально при 8 ЦПУ)

- Узловой контроллер UNC
- Межсоединение контроллеров через линки SCI
- Межсоединение ЦПУ и контроллеров через линки UPI
- Обеспечивает рост до 32 ЦПУ

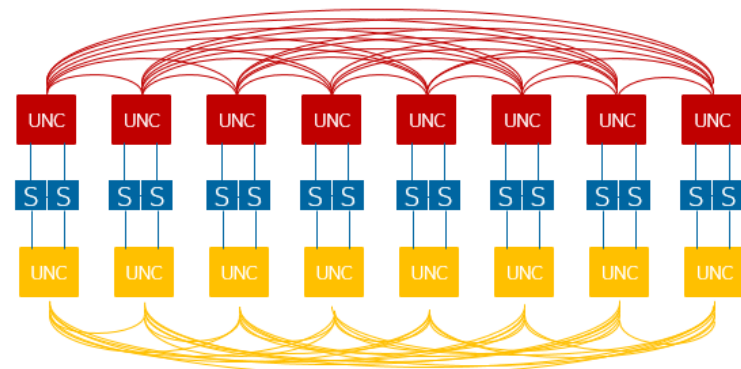
2 типа конфигураций:

- **single rail** (1 контроллер UNC на 2 ЦПУ)
- **dual rail** (2 контроллера UNC на 2 ЦПУ)
 - Вдвое большее количество линков SCI и UPI
 - Обеспечивается избыточность UNC
 - Вдвое большая пропускная способность

Bullion S	BullSequana S
QPI : 20 * 9,6 ГТ/сек = 192 ГБ/сек	UPI : 20 * 10,4 ГТ/сек = 208 ГБ/сек
XQPI : 8*14 ГБ/сек = 112 ГБ/сек	SCI : 8*25 ГБ/сек = 200 ГБ/сек



Топология сервера BullSequana S1600 в режиме single rail



Топология сервера BullSequana S1600 в режиме dual rail

Спасибо за внимание

Для предоставления дополнительной информации:

Алексей Игнатов

М +7 910 472 26 13

a.ignatov@postgrespro.ru

Артем Иванов

М +7 960 116 06 23

artem.ivanov@atos.net



<https://www.facebook.com/BullRussia>

<https://www.facebook.com/PostgresProfessional>

Atos, the Atos logo, Atos Codex, Atos Consulting, Atos Worldgrid, Worldline, BlueKiwi, Bull, Canopy the Open Cloud Company, Unify, Yunano, Zero Email, Zero Email Certified and The Zero Email Company are registered trademarks of the Atos group. May 2016. © 2016 Atos. Confidential information owned by Atos, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval from Atos.



Atos