РФЯЦ-ВНИИТФ
РОСАТОМ

Масштабируемый ПАК СУБД на основе Postgres Pro Shardman 18 и высокоскоростной сети Альфа с вариантами исполнения на процессорах архитектуры x86-64 и ARMv8 (Байкал-С) для критической информационной инфраструктуры

28 сентября 2026

ФГУП «РФЯЦ–ВНИИТФ им. Академ. Е.И. Забабахина»

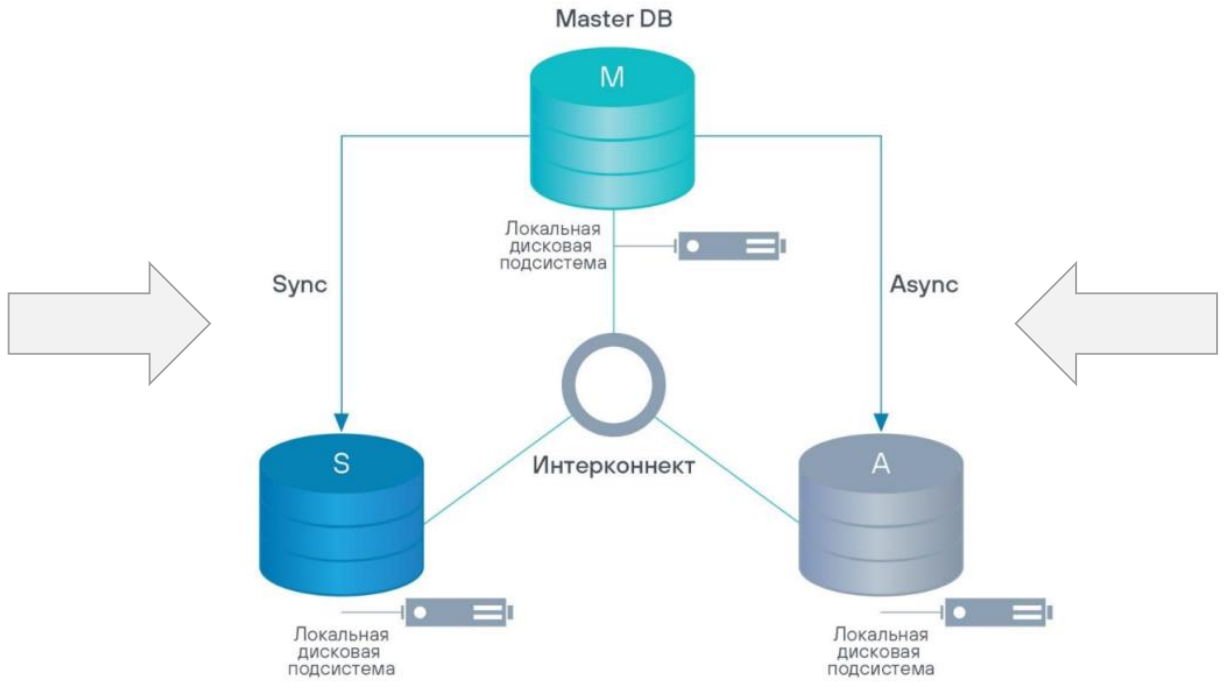
Тимур Фаритович Исмагилов
ведущий инженер-программист, к.ф.-м.н.

Существующие решения

РФЯЦ-ВНИИТФ
РОСАТОМ

Машина баз данных Скала^р МБД.П





Корпоративная машина баз данных Tantor XData



Особенности

Размер монолитной БД более 10 ТБ не рекомендуется
Монолитная БД не масштабируется по производительности и ёмкости
Сеть Ethernet 100G (высоколатентный, без RDMA) ограничивает темп обработки транзакций

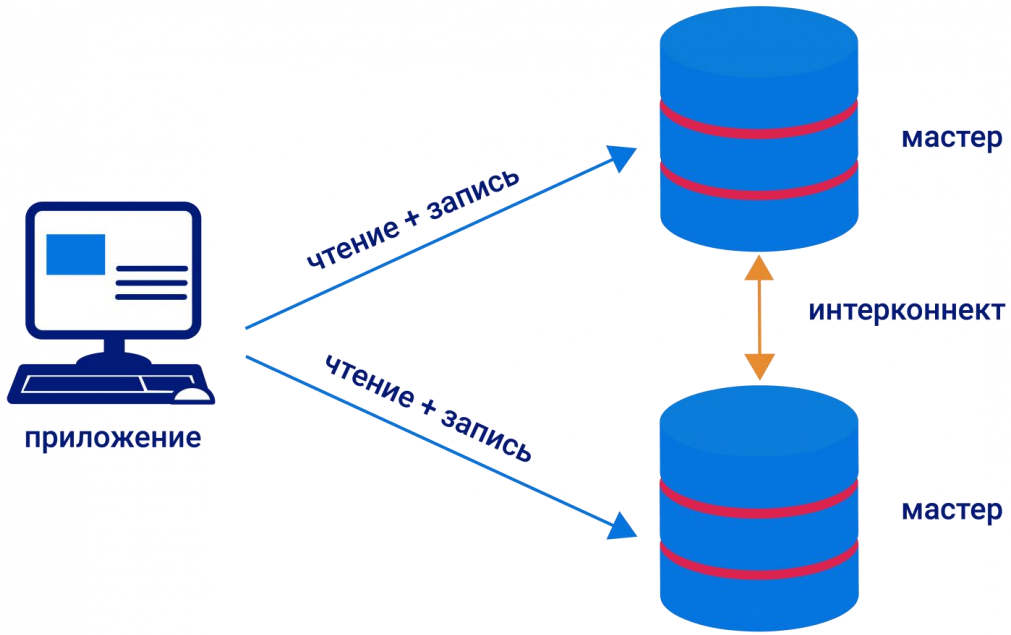
Цель разработки ПАК СУБД

РФЯЦ-ВНИИТФ
РОСАТОМ

Повышение производительности и ёмкости СУБД
на OLTP-нагрузке за счёт:

- **горизонтального масштабирования**
- **доступа** к БД **через любой сервер БД**
- поддержки **распределенных транзакций** с высоким уровнем изоляции

Основа №1
PostgresPro Shardman

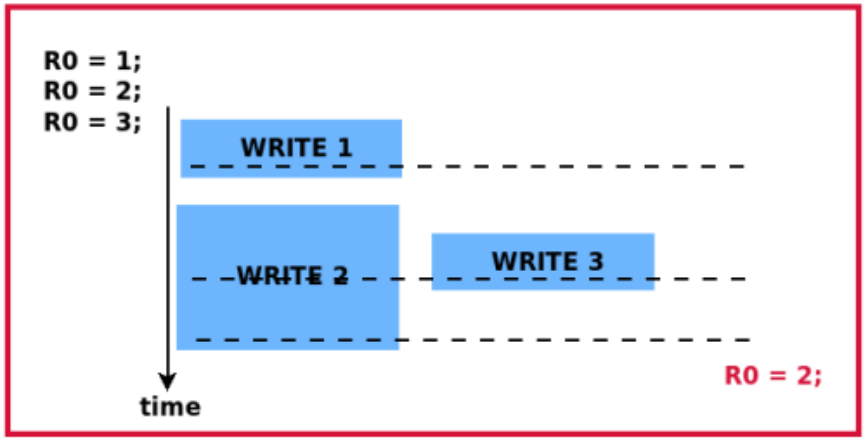


- Новые узлы добавляются без остановки системы и миграций данных
- Архитектура протестирована на объемах до 1 ПБ и сотни узлов
- Shardman сам определяет сбои и переводит трафик на резервные узлы
- Архитектура без единой точки отказа — каждый узел может быть точкой входа
- Соблюдаются свойства ACID и согласованность транзакций
- Оптимизирована под OLTP и real-time-нагрузки

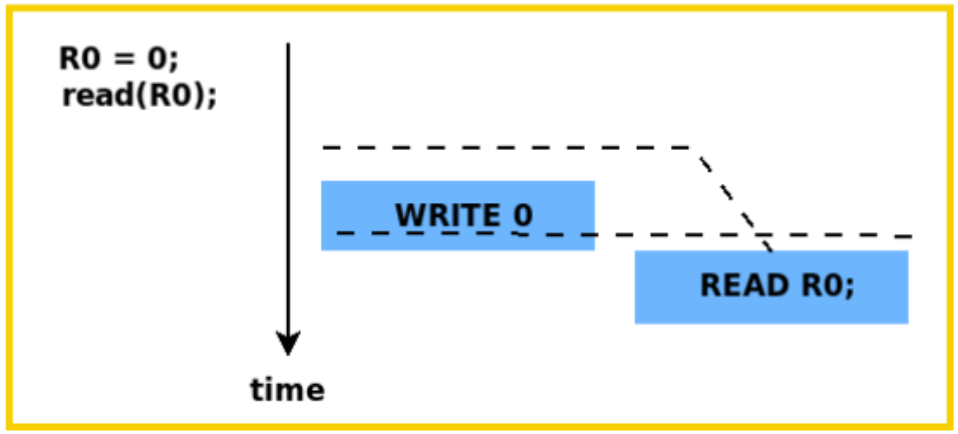
Аномалии данных в распределенных СУБД

РФЯЦ-ВНИИТФ
РОСАТОМ

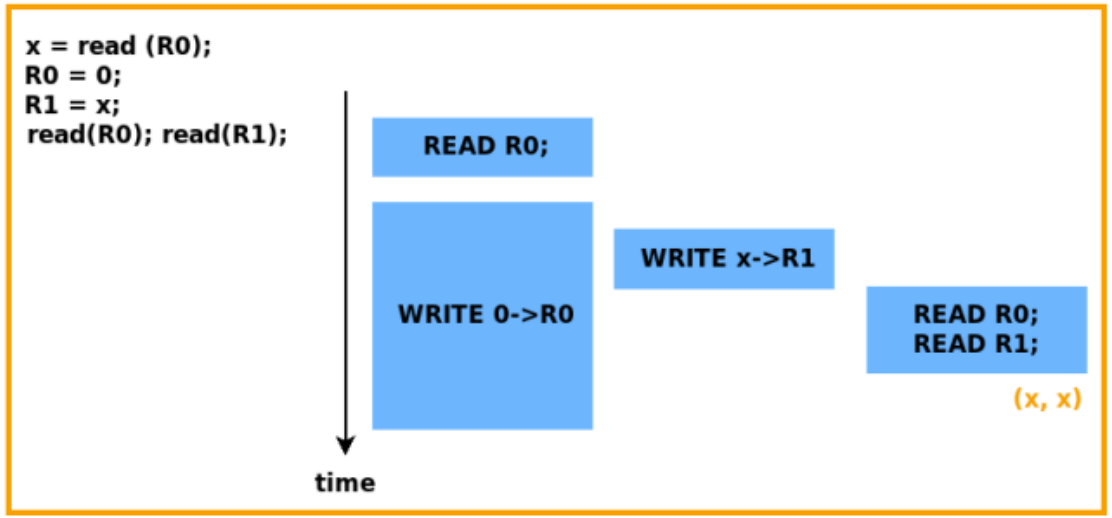
Immortal write



Stale reads



Causal reverse



Иерархия моделей и гарантии консистентности в различных БД

РФЯЦ-ВНИИТФ
РОСАТОМ

Скорость обработки транзакций

	System Guarantee	Dirty read	Non-repeatable read	Phantom Read	Write Skew	Immortal write	Stale read	Causal reverse
Citus DB	READ UNCOMMITTED	POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE
	READ COMMITTED	NOT POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE
	REPEATABLE READ	NOT POSSIBLE	NOT POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE
Arenadata DB	SNAPSHOT ISOLATION	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE
	SERIALIZABLE / ONE COPY SERIALIZABLE / STRONG SESSION SERIALIZABLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	POSSIBLE	POSSIBLE	POSSIBLE
	STRONG WRITE SERIALIZABLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	POSSIBLE	NOT POSSIBLE
Cocroach DB	STRONG PARTITION SERIALIZABLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	POSSIBLE
Postgres Pro Shardman	STRICT SERIALIZABLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE	NOT POSSIBLE

Основа №2 Высокоскоростная сеть Альфа

РФЯЦ-ВНИИТФ
РОСАТОМ

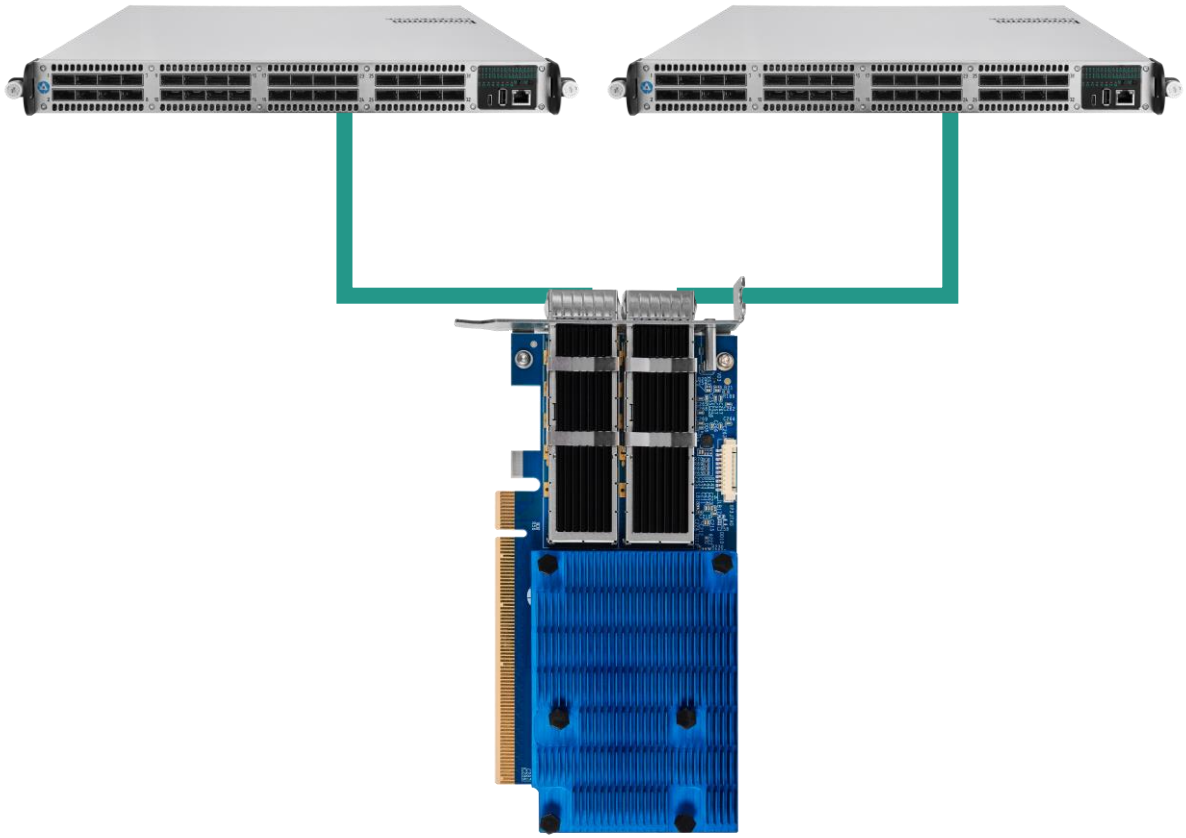
Программно-аппаратная реализация поддержки СУБД в сети Альфа (libsync)

Сетевой Адаптер

- Механизмы аппаратного обеспечения изоляции транзакций
- Поддержка RDMA

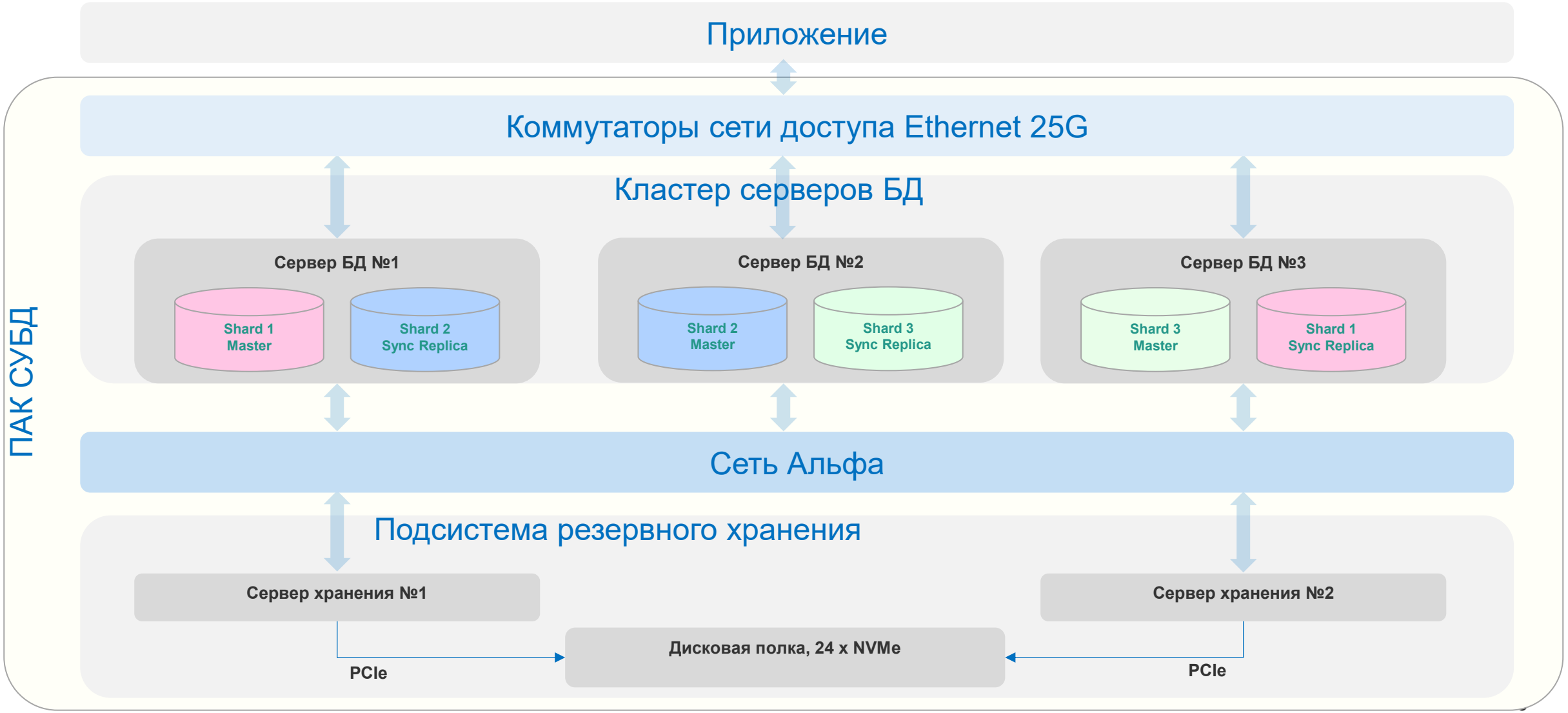
Коммутатор

- Внедрены механизмы отказоустойчивости: режим зеркалирования.



Архитектура ПАК СУБД в минимальной конфигурации

РФЯЦ-ВНИИТФ
РОСАТОМ

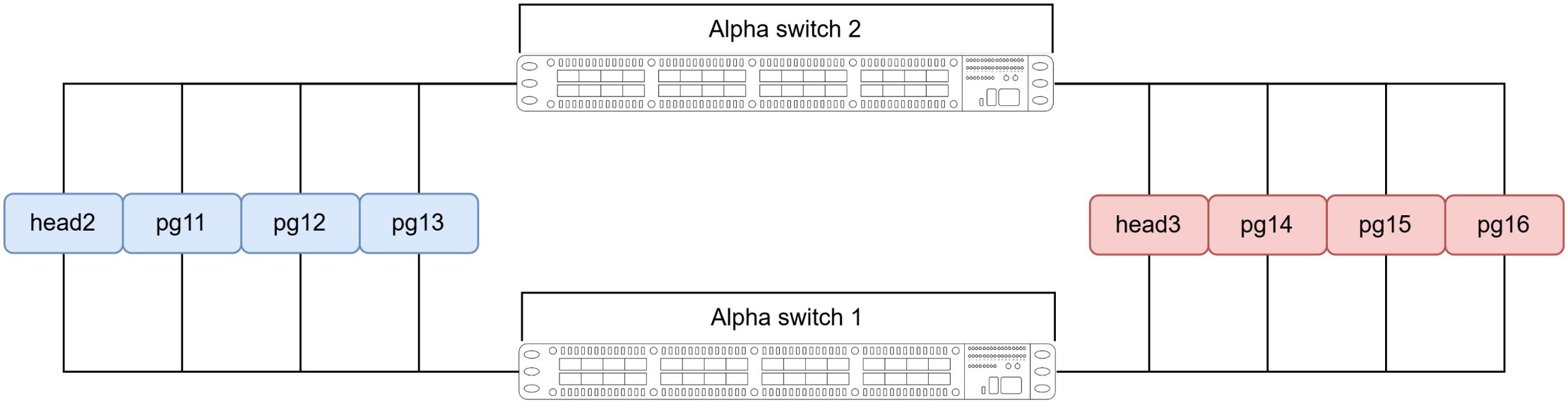


Характеристики серверов БД ПАК СУБД в минимальной конфигурации

РФЯЦ-ВНИИТФ
РОСАТОМ

	1 этап	2 этап
Процессоры	2x Intel Xeon Gold 6442Y	2x Baikal BE-S1000
ОЗУ	512 ГБ DDR4	768 ГБ DDR4
SSD для начальной загрузки	2x 1 ТБ	
SSD NVMe для СУБД	16x 3.2 ТБ	

Описание стенда ПАК СУБД



ПАК СУБД x86

Сервера ФГУП «РФЯЦ – ВНИИТФ им. академ. Е.И. Забабахина»
CPU: Intel Xeon Gold 6442y
Адаптеры Alpha
ПО: Alpha a3eth, libsync
Postgres Pro Shardman 18

ПАК СУБД Baikal

Сервера Спинтроника
CPU: Baikal BE-S1000
Адаптеры Alpha

ПО: Alpha a3eth, libsync
Postgres Pro Shardman 18

Результаты теста pgbench (TPCB)

РФЯЦ-ВНИИТФ
РОСАТОМ

Shardman	CPU	Alpha Sync (strict serializable)	Сеть	TPS
Shardman 17	Intel Xeon 6442Y	off	100GbE	29936
Shardman 17	Intel Xeon 6442Y	on	Альфа	45262
Shardman 18	Intel Xeon 6442Y	on	Альфа	**63680**
Shardman 18	**Baikal BE-S1000**	on	Альфа	**5868** предварительный

Высокоскоростная сеть Альфа
Libfabric (OFI), 2.7.0

РФЯЦ-ВНИИТФ
РОСАТОМ

Libfabric OFI:
единая модель коммуникаций
поверх разных сетей.

MPI · SHMEM · PGAS · Storage

libfabric API

a3net · TCP · verbs · другие провайдеры

ЗАДЕРЖКА · на сообщениях в 8 Б

1,86 мкс
fabtests · fi_rdm_pingpong

1,95 мкс
IMB-MPI1 · pingpong

ПРОПУСКНАЯ СПОСОБНОСТЬ

72,5 Гбит/с
fabtests · fi_rdm_bw

72,4 Гбит/с
IMB-MPI1 · uniband

Стенд: 2 узла
HPE ProLiant DL360 Gen10 Plus
2x Intel Xeon Platinum 8368 CPU 2.40 ГГц, 64 ГБ

Преимущества ПАК СУБД

РФЯЦ-ВНИИТФ
РОСАТОМ

- **Комплексное решение:** Postgres Pro Shardman, масштабируемые модули для обработки и хранения данных, сеть Альфа, подсистема резервного хранения, подсистема управления и мониторинга

- **Postgres Pro Shardman:** прозрачная работа с шардированной базой данных, тот же интерфейс и принципы работы, что и при работе с монолитной базой данных

- **Высокоскоростная сеть Альфа:** сокращение времени выполнения синхронного коммита за счет сверхмалой коммуникационной задержки

- **Горизонтальное масштабирование** путём увеличения размера кластера серверов БД вплоть до 100 серверов



Преимущества в сравнении с монолитными БД:

- **Повышение темпа обработки транзакций** до 30 раз на 100 серверах БД
- **Увеличение** допустимого размера БД до 1 ПБ

Благодарю за внимание!

Тимур Фаритович Исмагилов

tfismagilov@vei.ru