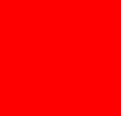




ORACLE®

Exadata X3: новые возможности и решения

Сергей Стеценко (sergey.stetsenko@oracle.com)
Ведущий технический специалист

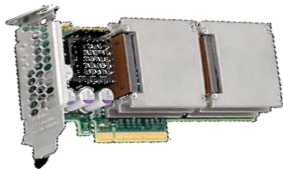
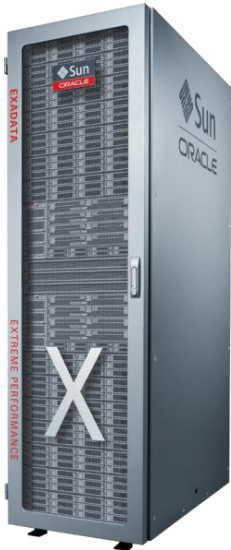


Exadata X3

Новое в оборудовании

Exadata X3 Hardware

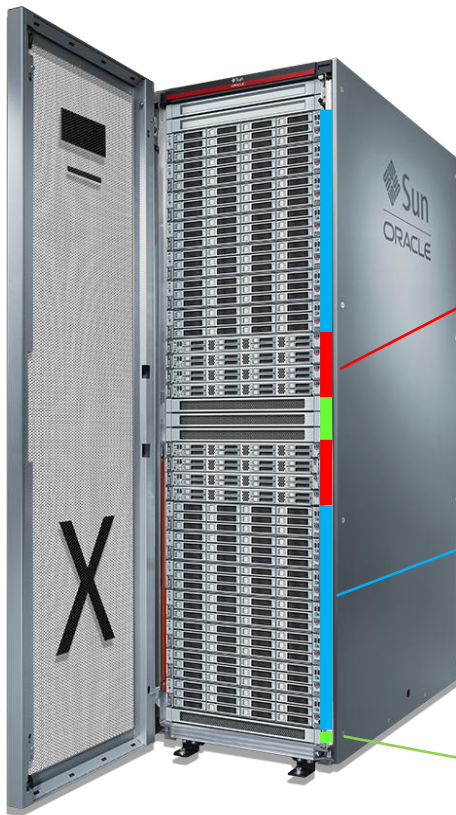
Новое в оборудовании



- Общая архитектура осталась неизменной
 - Нет необходимости в изменениях
 - Имя машины БД изменилось с X2 на X3
 - Например, X3-2 или X3-8
- Exadata X3 имеет
 - Значительно больше быстрой флэш-памяти
 - Увеличенные размеры ОЗУ
 - Более производительные процессоры
 - Расширенные сетевые возможности
 - Пониженное энергопотребление
 - Ту же цену, что и X2
- Новая начальная конфигурация “1/8”

Архитектура Exadata

Регулярная серверная структура



- **Database Grid**

- 8x2- или 2x8-ядерных Xeon-сервера БД
- Oracle Linux или Solaris 11
- **Oracle Database 11g, ASM, RAC**
- 10 Гб и 1Гб Ethernet (к ЦОД)

- **Intelligent Storage Grid**

- 6-ядерные Xeon в серверах хранения
- До 504 ТБ дискового пространства
- 22.4 ТБ флэш-памяти в шкафу
- **Exadata Storage Server Software**

- **InfiniBand Network**

- Внутренние межсоединения (40 Гб/сек)

Изменения в серверах БД Exadata X3-2

Новые 8-ядерные процессоры Intel E5 “SandyBridge”

Сервер БД	X4170 M2 (сейчас)	Sun Server X3-2 (новый)	Улучшение
Процессор	6-ядерный X5675 (3.06 GHz)	8-ядерный E5-2690 (2.9 GHz)	Ускорение от 1.3X до 1.5X
Память	96 ГБ (до 144 ГБ) 8 ГБ DIMMs	128 ГБ (до 256 ГБ) 16 GB DIMMs	1.3X и 1.7X при расширении
Сеть	4 x 1 GbE медь 2 x 10 GbE оптика	4 x 1 или 10 GbE медь 2 x 10 GbE оптика	Больше 10GbE-интерфейсов
Шина PCIe	Gen 2.0	Gen 3.0	Ожидается обновление карт до 3.0



- Внутренняя архитектура Xeon E5 (Sandy Bridge) позволяет достичь большей производительности при меньшей тактовой частоте
- Дисковые контроллеры, диски и инфраструктура InfiniBand не изменились
- В отличие от X2 увеличение объёма ОЗУ не приводит к снижению производительности

Изменения в серверах хранения X3

Четырехкратное увеличение флэш-памяти

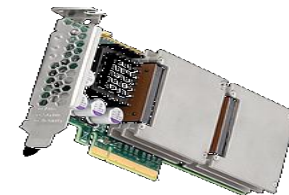
Сервер хранения	X4270 M2 (сейчас)	Sun Server X3-2L (новый)	Улучшение
Флэш	384 ГБ (4 x 96 ГБ)	1600 GB (4 x 400 ГБ)	4X
Диски	12 x 600 ГБ or 12 x 3 ТБ	Без изменений	Без изменений
Процессоры	6-ядерный L5640 (2.26 GHz)	6-ядерный E5-2630L (2.0 GHz)	Без снижения производительности



- Новые процессоры имеют более низкое энергопотребление
- Дисковые контроллеры, диски и инфраструктура InfiniBand не изменились
- Внутренняя память увеличена с 24 до 64 ГБ для поддержки большего флэш-кеша
- Шина PCIe 3.0 реализована, но новые PCI-карты еще не доступны

Новая PCI-карта флэш-памяти F40

Четырехкратная ёмкость и производительность



	Текущая F20	Новая F40	Улучшение
Объем хранения	96 ГБ	400 ГБ	4X
Скорость чтения	1 ГБ/сек	>1.4 ГБ/сек	1.4X

- Технология eMLC (Multi-Level Cell, 2 бита в ячейке)
 - Новый виток в развитии технологии флэш-памяти для промышленных применений
- На 40% быстрее как по чтению, так и по записи
- Реальный показатель IOPs для флэш-памяти намного выше, чем отображено в характеристиках машины баз данных X3
 - Измеряется конечная производительность выполнения SQL
- Упрощенное обслуживание
 - Energy Storage Module (ESM) заменен на другие энергонакопители

Изменения в шкафу X3

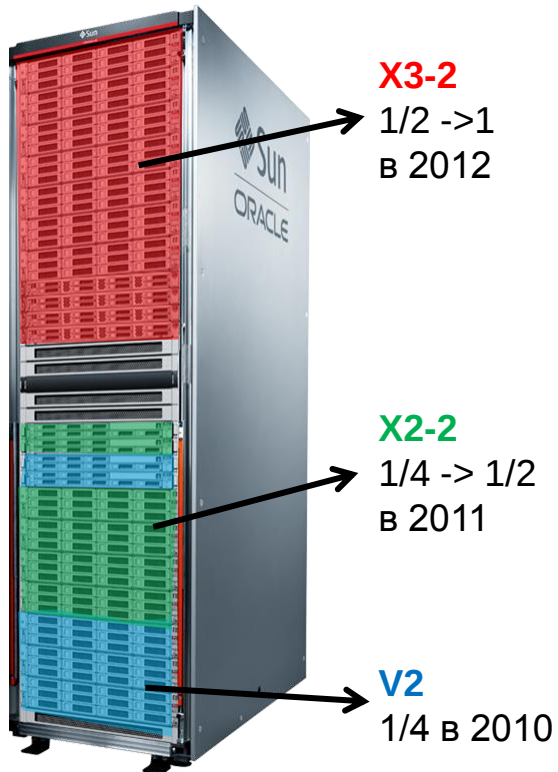
Оптимизация конструкции и улучшения

- Энергопотребление сокращено на 10-30%
 - Энергоэффективные процессоры
 - Оптимизация воздушных потоков
- Отсутствует KVM в X3-2 и шкафах расширения
 - Вместо этого используется ILOM
 - Вверху шкафа доступно 2U для клиентского Ethernet-коммутатора
- Улучшен механический дизайн
 - Слегка уменьшена высота каждого сервера для более легкого выдвижения
- Коммутаторы InfiniBand и 1Gb Ethernet не изменились



Обновления X3

Совместимость моделей



- Системы X2 или даже V2 могут быть расширены оборудованием X3
 - В одной машине БД могут работать серверы из разных поколений Exadata
 - Базы данных и кластеры могут одновременно использовать оборудование разных моделей
- Серверы (или их компоненты) **не заменяются**
 - Расширение только путем добавления новых серверов
- X3 требует обновления Storage Server Software до 11.2.3.2.0
 - Обновление баз, ASM, Clusterware не требуется

Exadata X3-2 Storage Expansion Rack



- Шкаф расширения (Storage Expansion Rack) может дополняться серверами хранения X3
- Основным преимуществом является увеличенный объём флэш-памяти
 - 28.8 ТБ флэш-памяти в полном шкафу
 - Дисковые объёмы не изменились (до 648 ТБ “сырого” пространства)
- Может использоваться как с Exadata, так и со SPARC SuperCluster
- Может использоваться с машинами баз данных всех поколений

Ценообразование и лицензирование



- Цена оборудования осталась прежней
 - Но новые процессоры, больше ядер, больше ОЗУ, больше флэш-памяти...
- Цена лицензирования Oracle DB EE для X3-2 увеличилась на треть
 - 8-ядерные процессоры вместо 6-ядерных
 - Никаких изменений в X3-8
- Стоимость ПО для серверов хранения не изменилась
 - Количество дисков в серверах хранения и количество серверов остались прежними

Кратко об оборудовании в Exadata X3

	X3-8 Full Rack	X3-2 Full Rack
Database Servers	2 Sun Server X2-8	8 Sun Server X3-2
Cores (Total)	160 (2.40 GHz)	128 (2.9 GHz)
Memory (Total)	4096 GB	1024 GB (max 2048 GB)
InfiniBand Switches	3	
Exadata Storage Servers	14 Sun Server X3-2L	
Cores (Total)	168	
Flash (Total)	22.4 TB	
Raw Storage (Total)	100 TB or 504 TB	
Disk Data Bandwidth from SQL	25 GB/sec or 18 GB/sec	
Flash Data Bandwidth from SQL	100 GB/s	
Flash IOPS (8k) from SQL	1.5 Million Reads, 1 Million Writes	
Exadata Software	11.2.3.2.0 required for X3 hardware	
Database/ASM Software	All supported 11.2 and higher	

Exadata Eighth Rack (“восьмушка”)

Небольшие и тестовые системы, разработка, DR...



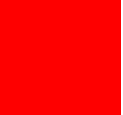
- Оборудование такое же, как в “1/4”
- Половина ресурсов в каждом сервере программным образом отключена
 - Половина ядер в каждом сервере БД и в каждом сервере хранения
 - Половина дисков и половина флэш-модулей (карт) в серверах хранения
 - Все локальные диски на серверах баз данных остаются доступными
 - Оперативная память также не отключается
 - Отказоустойчивость не нарушается
- Обновление 1/8 (Eighth) до 1/4 (Quarter) выполняется специальным скриптом

Ценообразование Exadata Eighth Rack

Почти в два раза дешевле Quarter Rack



- Стоимость Eighth Rack по прайс-листу составляет \$200K
 - \$330K за Quarter Rack
- Лицензии на ПО дешевле в 2 раза
 - Лицензируются половина ядер и дисков (остальные заблокированы)
- Общая цена составляет **чуть больше половины Quarter Rack**
- Апгрейд возможен только с полной 1/8 до полной 1/4
 - Нельзя разблокировать только часть ресурсов (ядер, дисков, флэш-памяти)
 - Новое оборудование при апгрейде не устанавливается

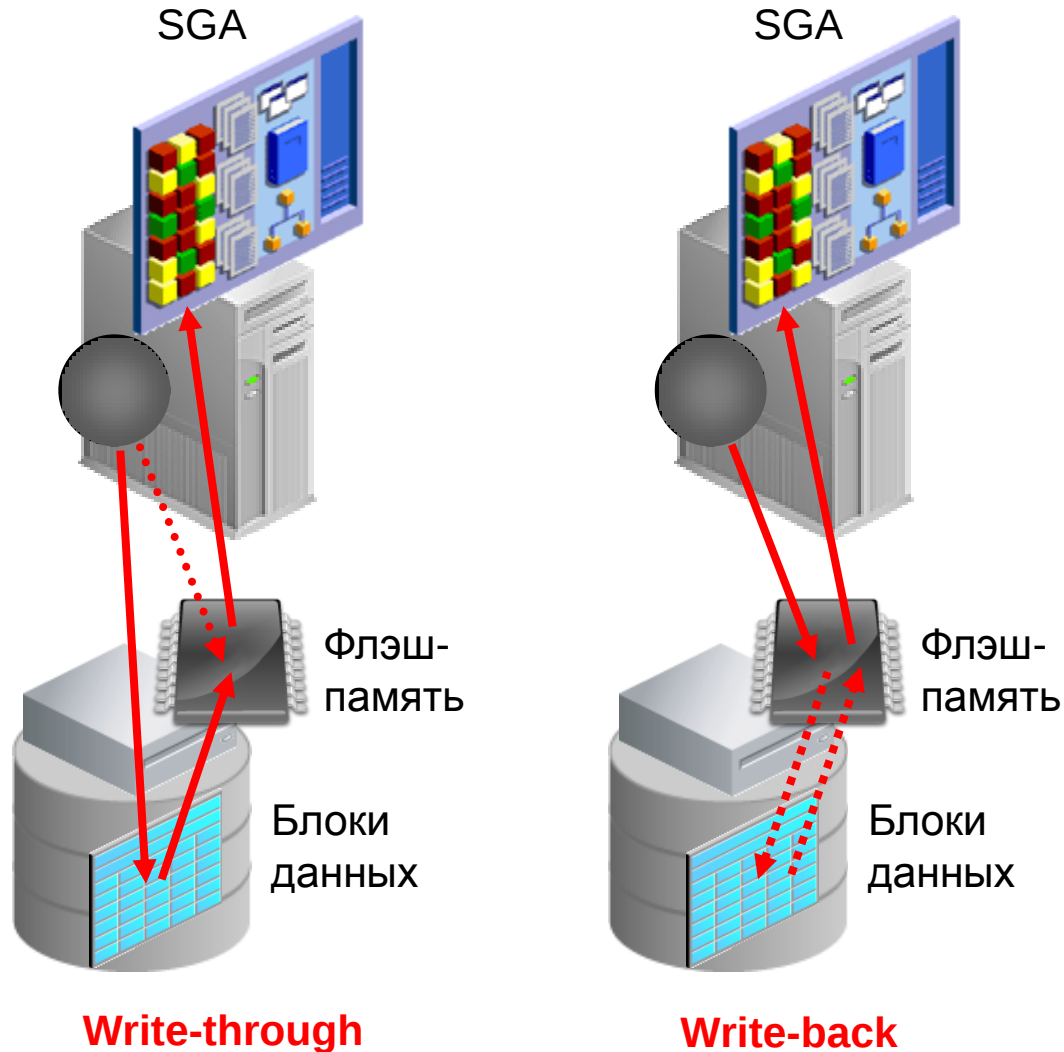


Exadata Software

11.2.3.2.0

Для всех поколений Exadata

Кеширование "write-back" vs "write-through"

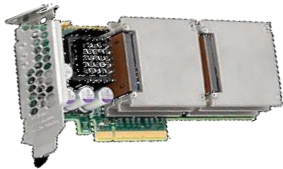


- "Write-through"
 - Запись на диск предваряет запись в кеш
 - Синхронный режим
- "Write-back"
 - Сначала запись в кеш, потом на диск
 - Асинхронный режим
- Способность кеша удерживать данные

Флэш-кеш по записи ("write-back")

20-кратное увеличение IOPS по записи

Writes I/Os



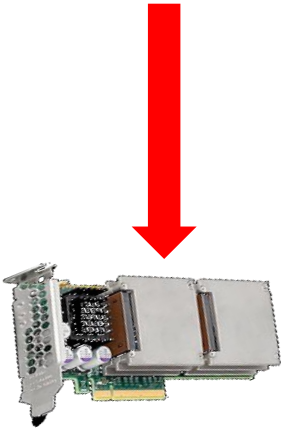
1M Write IOPS
on X3

500K Write IOPS
on X2

- Данные записываются сначала в кеш, а затем асинхронно на диск
 - LRU-дисциплина хранения блоков
 - Блоки могут храниться месяцами
 - При необходимости редко используемые блоки перемещаются на диск
- Значительное ускорение приложений с интенсивной записью (в т.ч. OLTP)
 - Часто обновляемые таблицы и индексы
 - До 20 раз для X3, до 10 раз для X2 и V2
- Кеширование является "разумным"
 - Например, Data Pump не кешируются

Доступность данных во флэш-кеше

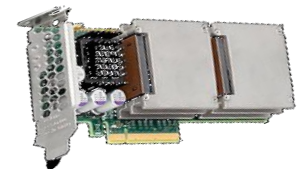
Writes I/Os



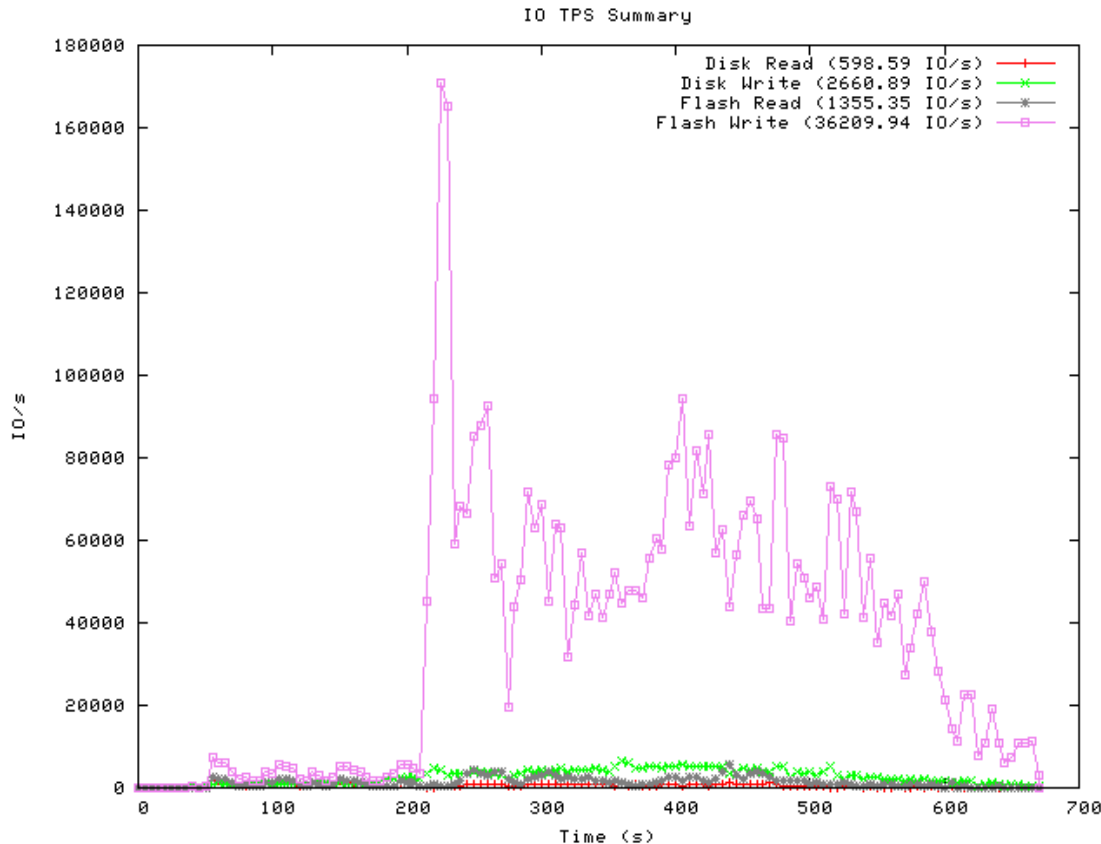
- Содержимое кеша сохраняется при перезагрузках СУБД
 - Не требуется "разогрев" флэш-кеша
 - Ускоряется применение журналов при восстановлении и при работе Data Guard
 - При применении журналов идёт интенсивная запись данных
 - Ускорение процесса восстановления до 3 раз
- Сбой флэш-карты автоматически обрабатывается Exadata и ASM
 - Потерянные данные восстанавливаются из "зеркал" на других серверах хранения
 - Из "зеркального" кеша или с дисков

Включение флэш-кеша по записи

- По умолчанию возможность флэш-кеширования по записи отключена
 - Для включения требуется обновление версии ASM
 - Включается командой
“Alter Cell flashCacheMode=WriteBack”
- Кеширование по записи необходимо включить для приложений с интенсивной записью данных
 - При наличии узких мест в подсистеме I/O по записи
 - При частых событиях ожидания "free buffer waits"

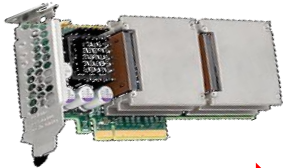


Пример из реальной жизни



- Пакетная обработка данных в Peoplesoft
- X3 Quarter Rack
- Без кеша по записи потребовалось бы **три(!) Full Racks**
- Скорость IO уже ограничивается CPU и самим приложением

Флэш-кеш в хранилищах данных

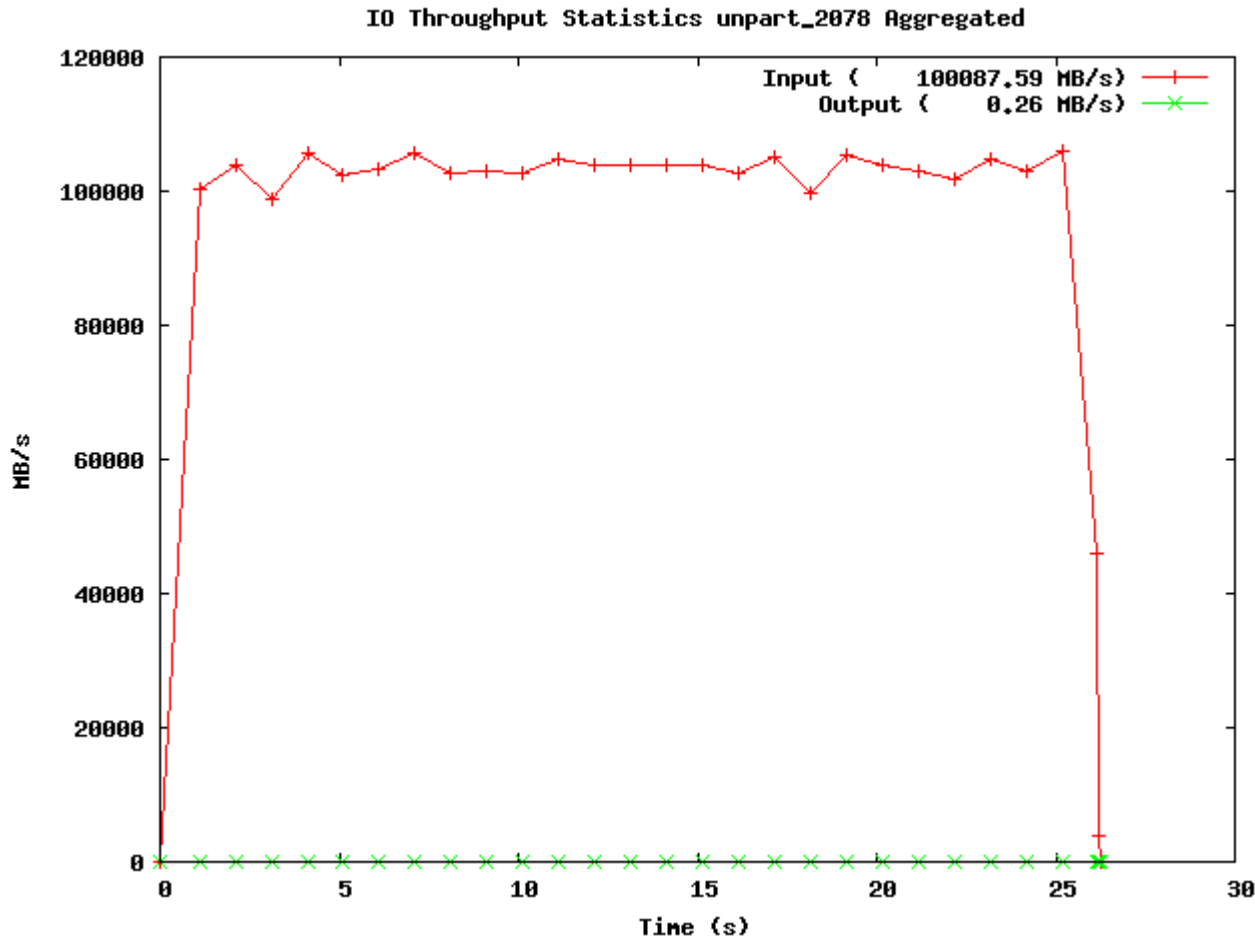


100 GB/sec
Scans

200 TB of
User Data
in Flash

- Новая технология кеширования позволяет ускорить работу в хранилищах данных
 - Скорость сканирования данных в среднем сокращается на 1/3
- Возможно одновременное сканирование данных из кеша и дисков
 - Показатели производительности суммируются
- Экстремальная производительность в сочетании со сжатием данных
 - Например, при 10-кратном гибридном сжатии 200 ТБ данных занимают только 20 ТБ в кеше
 - OLTP-сжатие вносит дополнительные преимущества

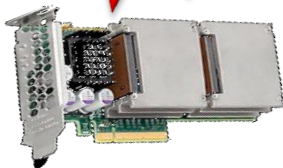
Пример чтения данных из хранилища



- Полное сканирование таблицы
- Full Rack X3-2
- 100 ГБ/сек
- Практически невозможно достичь на обычных массивах

Итак, очень кратко о X3...

Bigger



- Та же архитектура, что и X2
- Новейшие технологии
- **Та же цена**
- Новая конфигурация "1/8"

Faster



4X Larger Flash

22 TB of Flash Memory per Rack

20X More Write Performance

Exadata Smart Flash **Write** Cache

33% More Data Throughput

100 GB/sec running SQL

33% Faster Database CPUs

8-Core Xeon® E5-2690

75 % More Memory

Up to **2 TB** per X3-2 Rack

2X Data Center Bandwidth

40X 10Gb ports per X3-2 Rack

10% to 30% Lower Power

Up to 3 KW Reduction per Rack

ORACLE®