

XIV: Общее описание

Патрушев Александр
Product manager IBM
System p, Storage, System X.
e-mail: apatrushev@landata.ru



Немного о истории создания

Moshe Yanai – главный архитектор EMC Symmetrix/DMX.

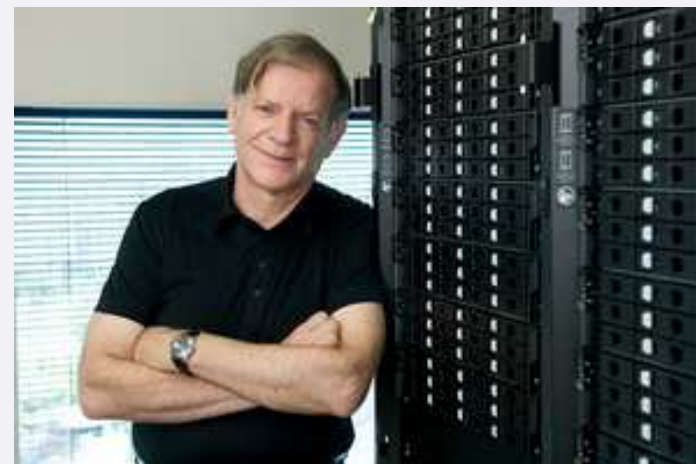
Начало проекта в 2002 году.

В 2005 году вышла первая версия под название Nextra.

В 2007 году IBM приобрела компанию и переименовала продукт в IBM XIV Storage System.

8 сентября 2008 году произошло публичное представление системы.

В настоящее время система продолжает быстро развиваться.



IBM XIV: на что он похож?

- **Качество**
- **Функционал**
- **Удобство**
- **Надежность**



IBM XIV

- Низкая цена владения (TCO)
- Простота администрирования
- Высокая отказоустойчивость и доступность
- Высокая производительность
- Масштабируемость
- Высокий уровень безопасности (LDAP, Роли пользователей, Аудит системы, SSL.)
- Поддержка Thin provisioning, Snapshots, Data migration, Remote mirroring
- Поддержка протоколов FC (до 24 портов) и iSCSI (до 6 портов)
До 120Гб кэш памяти



Физические характеристики

■ Диски

Very High Density Slower Rotation (VHDSR) диски
Скорость вращения 7,200 RPM
Интерфейс SATA
Емкость 1 TB или 2 TB

■ Физические размеры

Высота: 199.1 см
Ширина 60 см
Глубина 114.2 см
Максимальный вес 884 кг

■ Питание

200-240 VAC, at 60A or 30A
возможно однофазное
отказоустойчивое

■ Рабочая среда

Температура 10-35°C
Высота до 2,134 м
Влажность 25-80%





Возможные конфигурации

# Modules	# Disks	Usable capacity* 1 TB/2 TB	Connectivity		Power usage 1 TB/2 TB			Internal (grid) switching (GBps)	# CPUs		Memory (GB)
			FC ports	iSCSI ports	kVA (peak)	kVA (idle)	kBTU		Single	Dual**	
6	72	27/55	8	0	3.4/3.1	2.9/2.8	11.0	30	6	9	48
9	108	43/87	16	4	5.0/4.5	4.2/4.0	16.4	48	9	15	72
10	120	50/102	16	4	5.5/4.9	4.7/4.4	17.8	52	10	16	80
11	132	54/111	16	4	6.1/5.4	5.2/4.7	19.3	56	11	17	88
12	144	61/125	20	6	6.7/5.8	5.7/5.1	20.7	60	12	18	96
13	156	66/134	20	6	7.2/6.2	6.2/5.5	22.2	64	13	19	104
14	168	73/149	24	6	7.8/6.6	6.7/5.9	23.6	68	14	20	112
15	180	79/161	24	6	8.4/7.1	7.2/6.2	25.1	72	15	21	120



Функциональные характеристики

■ Защищенность данных

- Active-active N+1 резервирование системных компонентов
- Обновление микрокодов и оборудования без влияния на систему
- Синхронное и асинхронное зеркалирование
- До 16,000 snapshots
- Мгновенное восстановление после отказа диска или дискового модуля

■ Программное обеспечение

- Data migration
- Storage management and LDAP authentication
- Native multi-path support

■ Подключение хостов

- Fibre Channel rates: 1 Gbps, 2 Gbps, 4 Gbps
- iSCSI over Gigabit Ethernet
- SCSI-3 протокол для FC и iSCSI

■ Гарантия IBM

- 1 и 3 года стандартной гарантии, 24x7 Same Day On-Site Repair
- Установка инженером IBM
- Оповещения по SMS, e-mail, VPN, modem

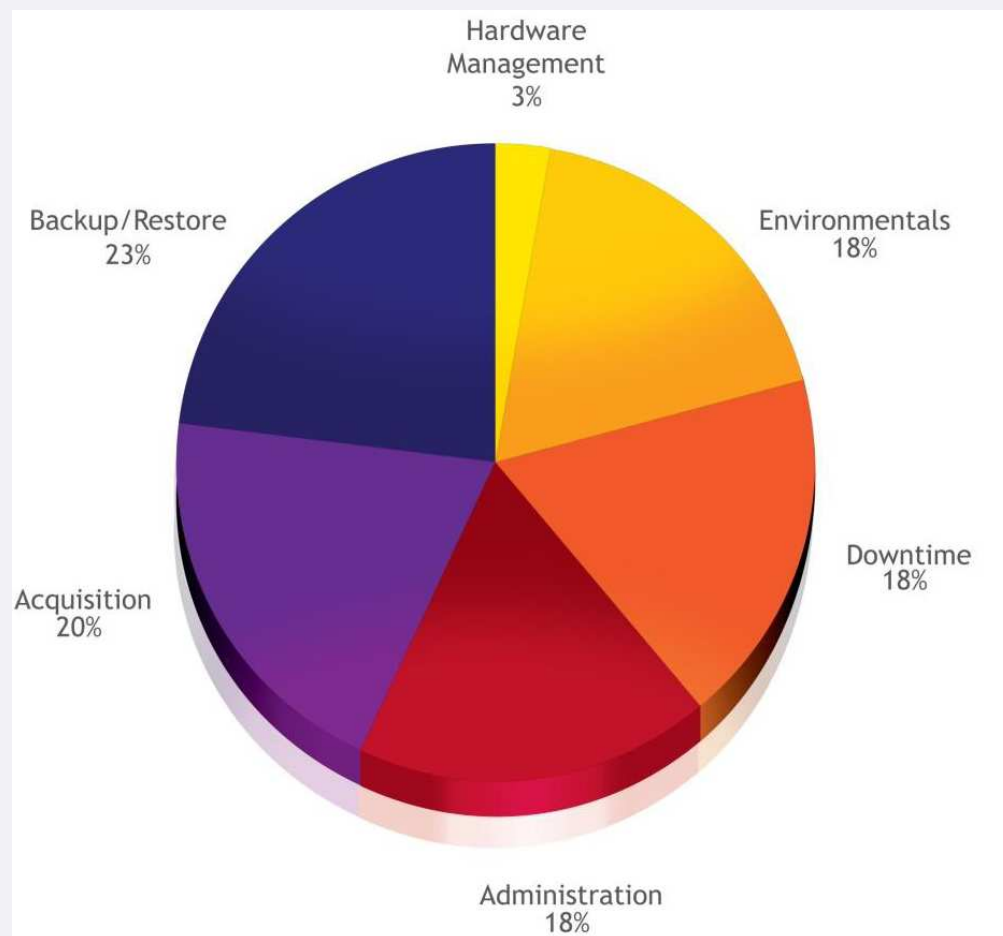


Ожидаемые обновления

- IB Internal Interconnect
- 8GB FC
- SSDs
- 3-way mirroring



Из чего складывается TCO?



TCO обычной Enterprise системы*

*Information Management Magazine



Как XIV снижает TCO?

Backup/Restore

Snaps без влияния на производительность системы
Поддержка VSS
Простая миграция

Acquisition

Простая архитектура
Низкая цена оборудования
Эффективный thin provisioning
Оптимальное использование емкости
Все ПО включено

Downtime

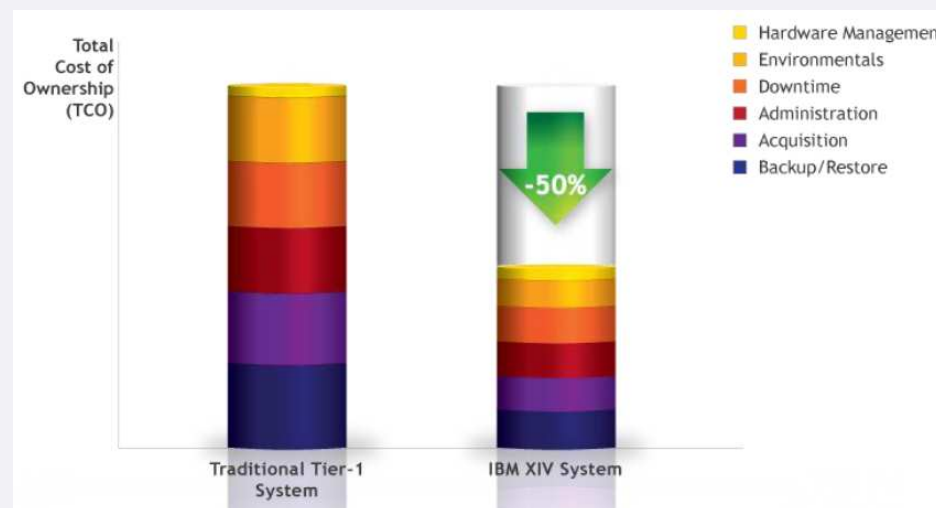
Меньше оборудования которое может отказать
Встроенные UPS
Малое влияние человеческого фактора
Простота мониторинга
Резервирование всех компонент

Окружение

Снижение энергопотребления и охлаждения
Снижение площади оборудования

Administration

Быстрое развертывание
Простое управление
Автоматическая настройка
Увеличение без скрытых затрат





Ситуация на рынке СХД

Block Storage

Transaction Processing & Database

- Tier 1 data – Medium-large scale
- High availability
- Disaster recovery with low RPO/RTO
- **Key storage value:** OLTP performance & data protection

Business Applications

- Tier 1 & 2 data
- High availability
- Selective disaster recovery
- **Key storage value:** Flexibility, ease of management

File & Block Storage

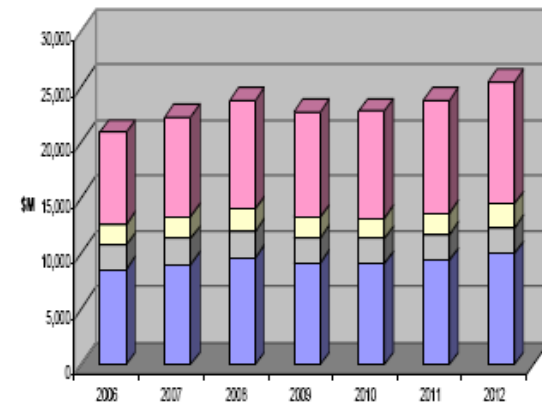
Analytics & HPC

- Extractions of Tier 1 data
- Sensor collection
- **Key storage value:** contribute to system performance

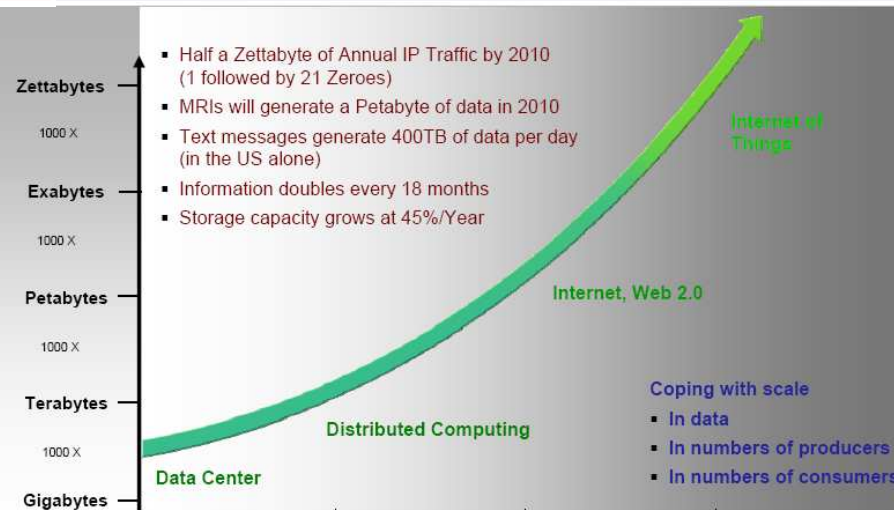
Web, Collaboration and Infrastructure

- Mostly Tier 2 data
- Cost Sensitive, high capacity
- **Key storage value:** Scalability, ease of management

External Storage Spending By Quadrant



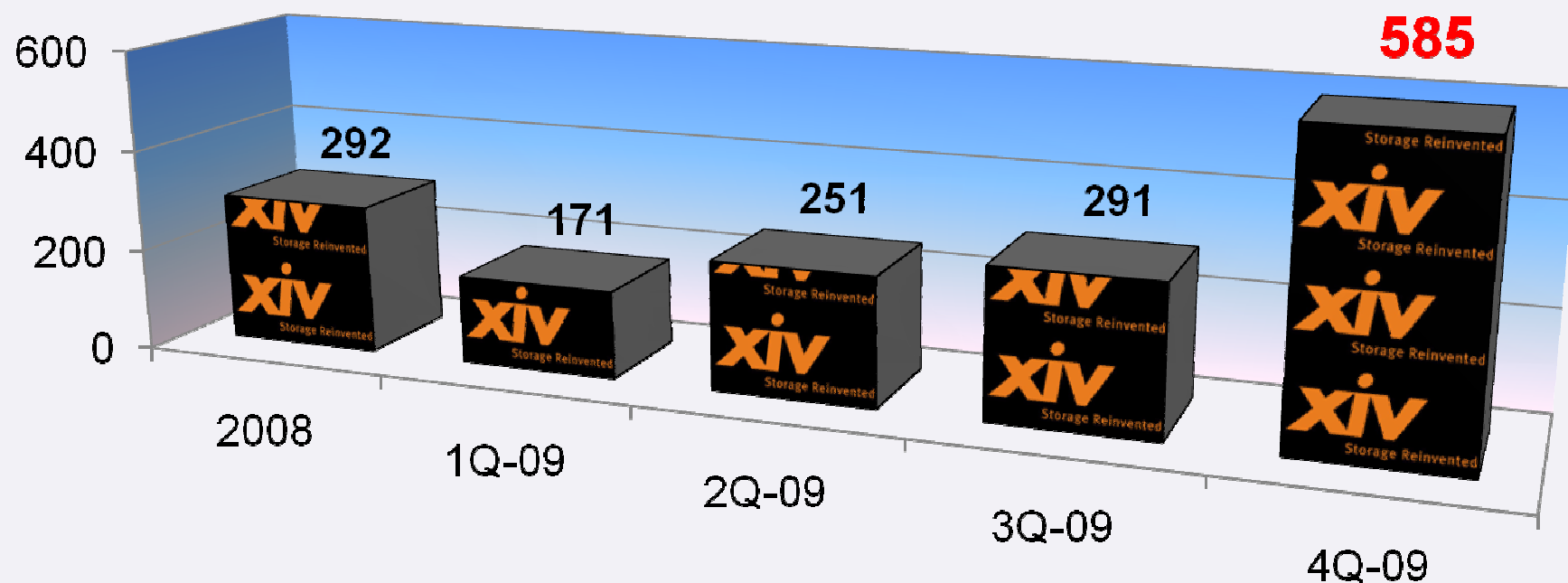
Transaction Processing & Database Business Applications Analytics and HPC Web, Collaboration, Infrastructure



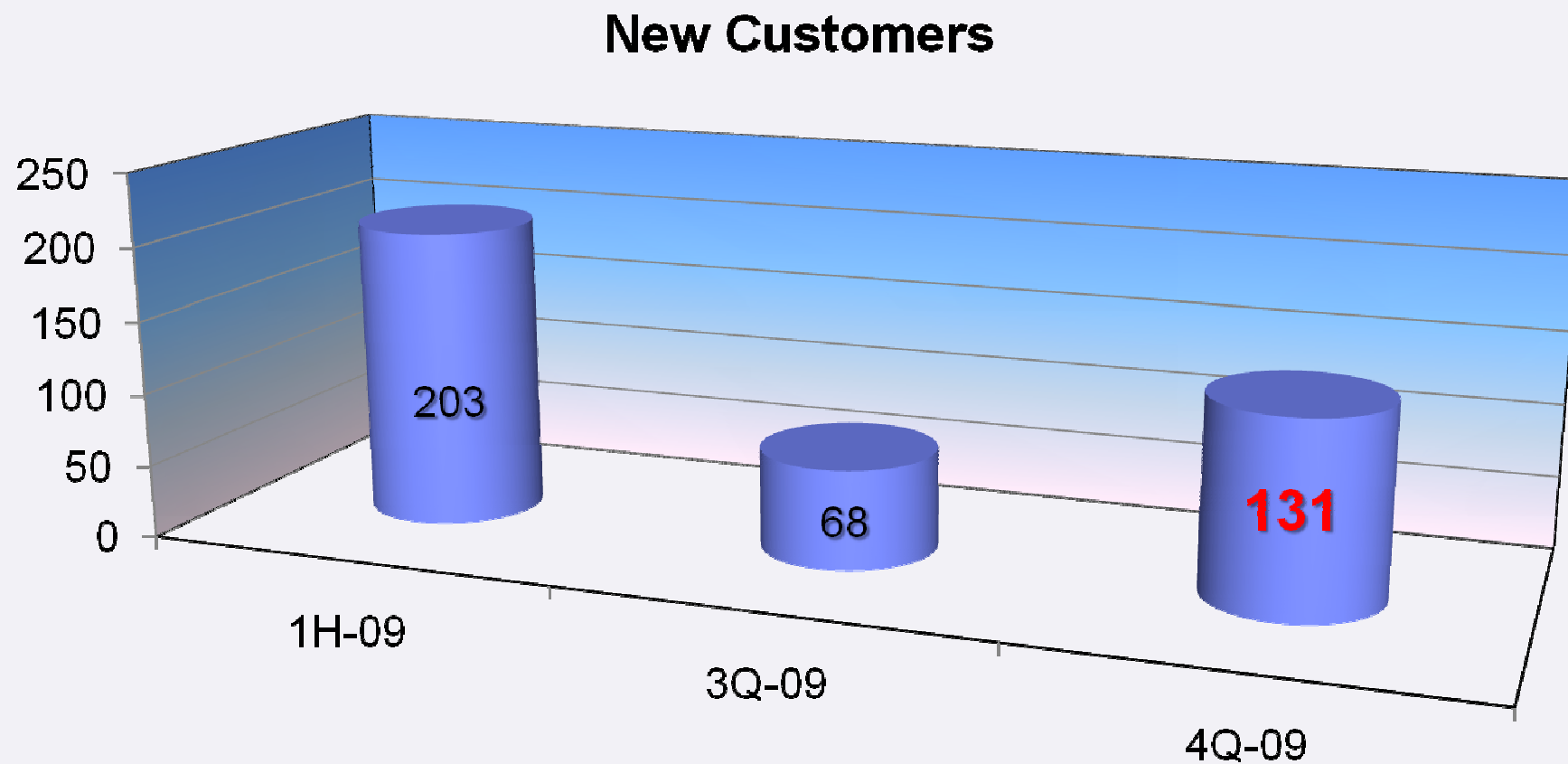


Количество проданных систем

Units Shipped
(1,590 @ EOY 2009)

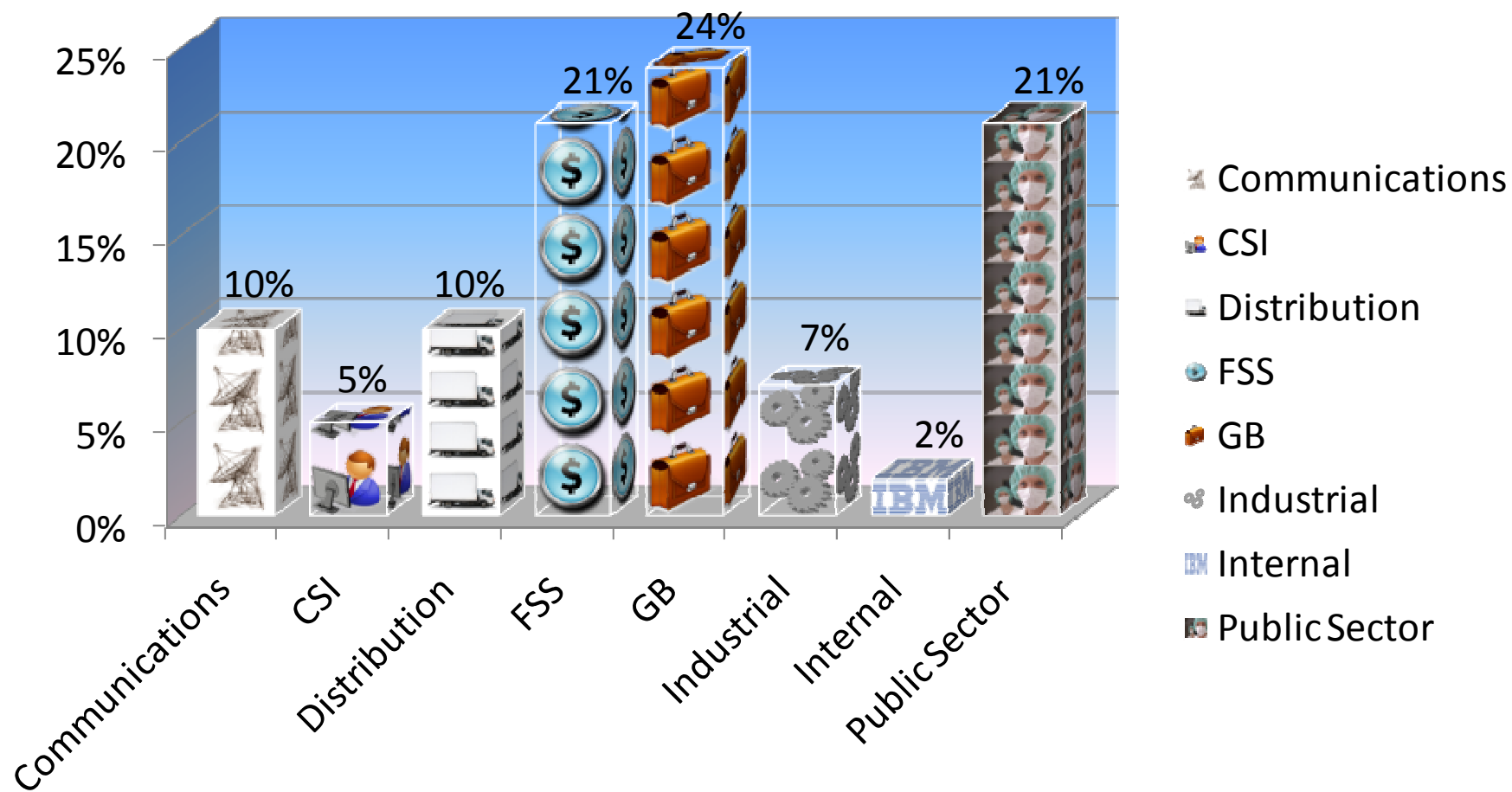


Количество новых заказчиков



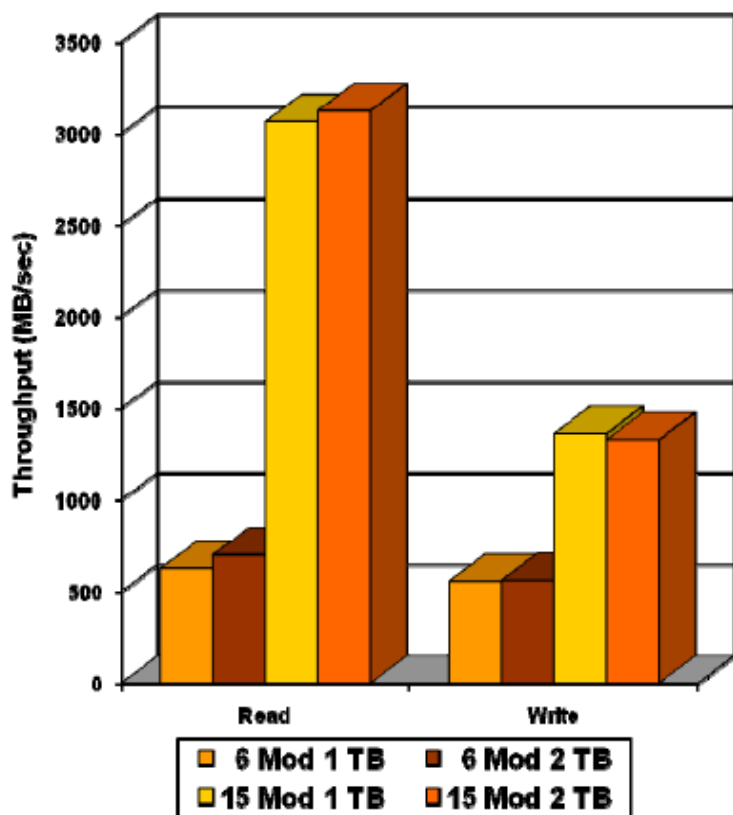
Распределение по индустриям

GB Shipped by Sector – 4Q 2009

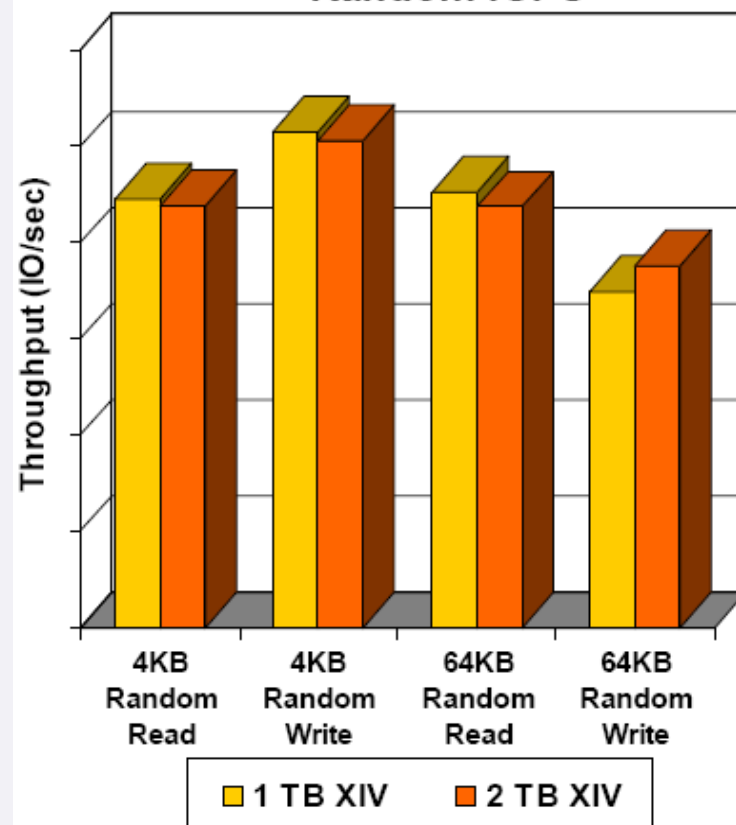


Что лучше 1ТВ или 2ТВ?

Sequential Bandwidth



Random IOPS





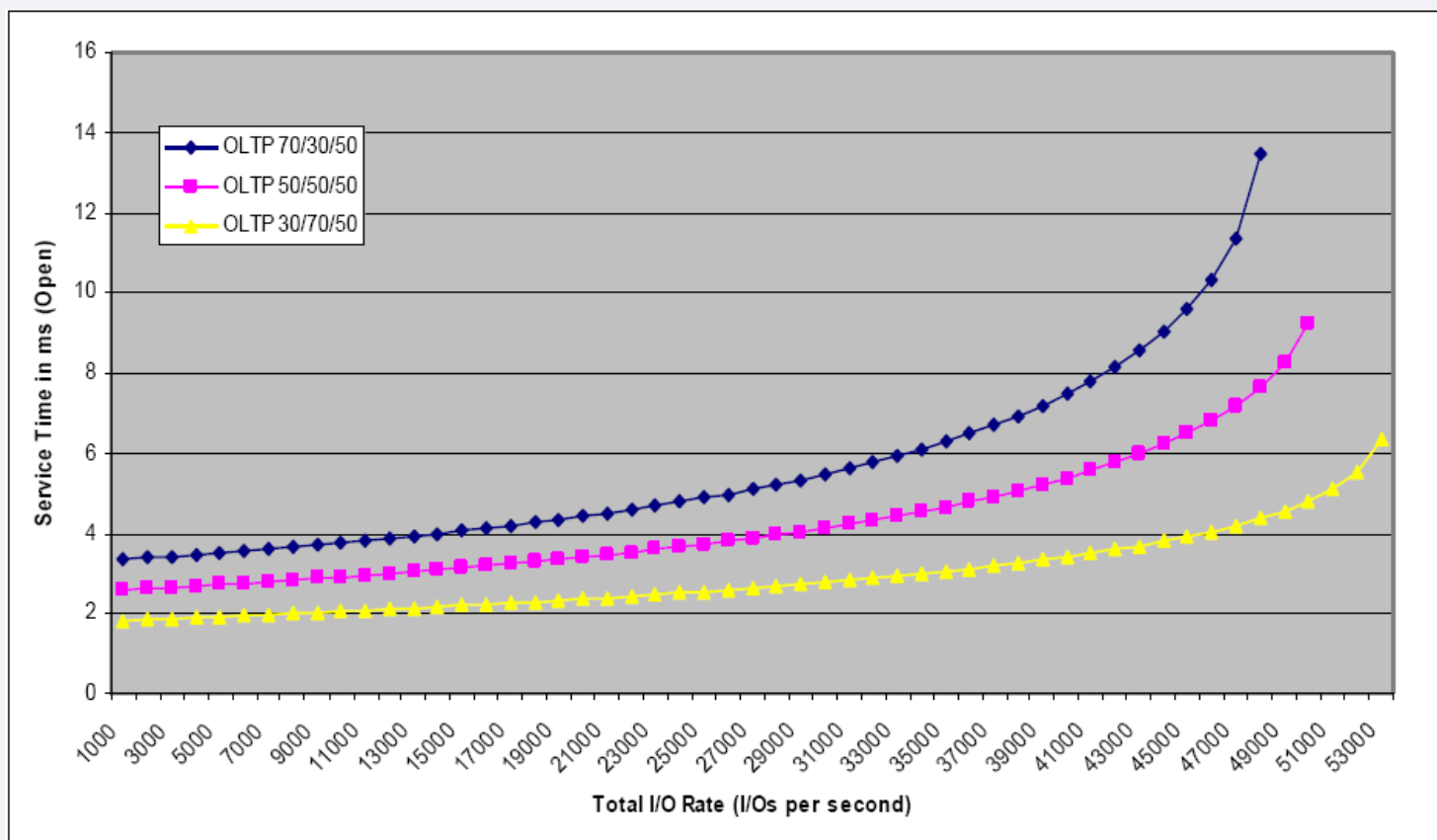
Сайзинг для SAP

- **Системы хранения классифицируются на основе количества SAPS :**
 - ERP (OLTP) 1 SAPS приблизительно равен. 0.4 IO per sec
 - BW/BI (OLAP) 1 SAPS приблизительно равен. 0.6 IO per sec
- **The service time performance constraints of a SAP ERP application:**
 - между 5 and 7 мс - excepted
 - между 10 and 15 мс - considered good
 - более 20 мс - considered as performance bottleneck
 - более 30 мс - system does not behave as expected towards the users

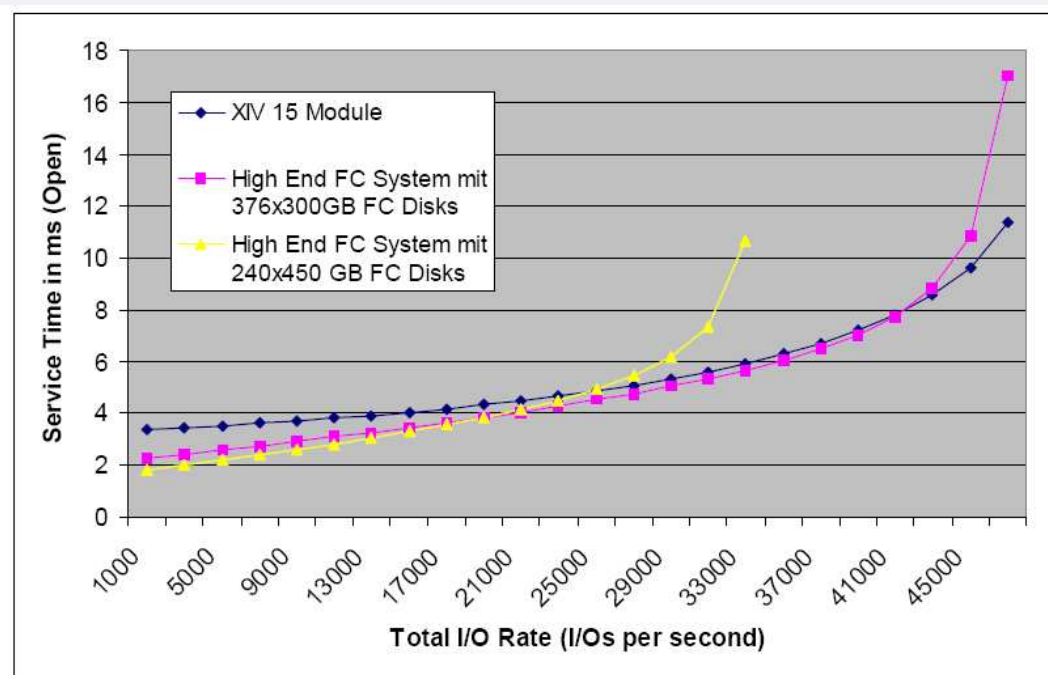


SAP OLTP при XIV

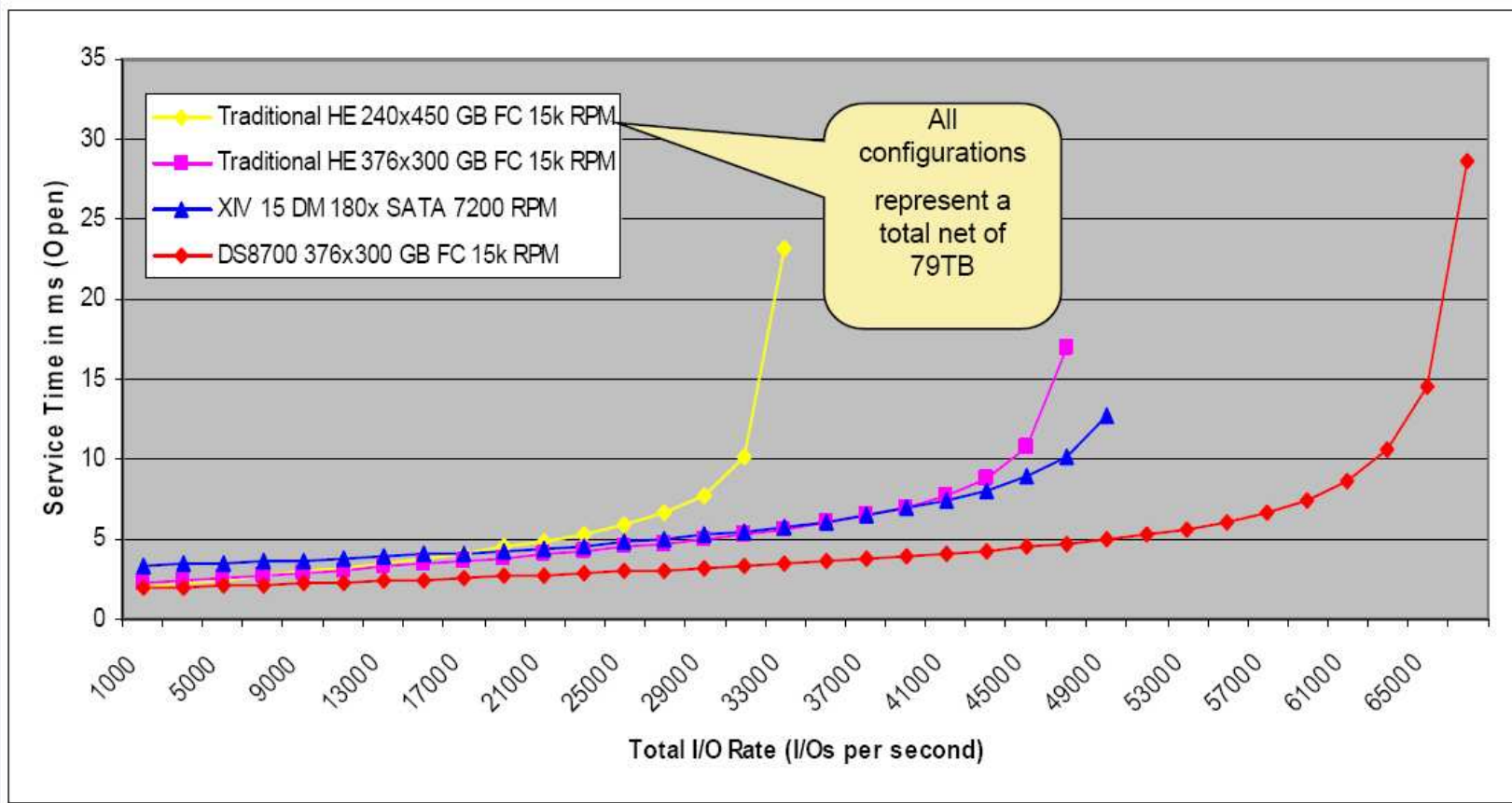
- Типичная OLTP нагрузка and 8k
- XIV показывает 50.000 IOPS при < 20ms



- Обычная OLTP Нагрузка, 70/30/50, Случайный доступ, 8kB
- Традиционным High End дисковым системам необходимо намного больше FC дисков для достижения той же производительности
 - 240x450 GB FC дисков (+64 GB Cache in Dual Controller High End) достигли ~35000 IOPS
 - 376x300GB FC достигли почти той же производительности что и XIV с 180 TB SATA дисками



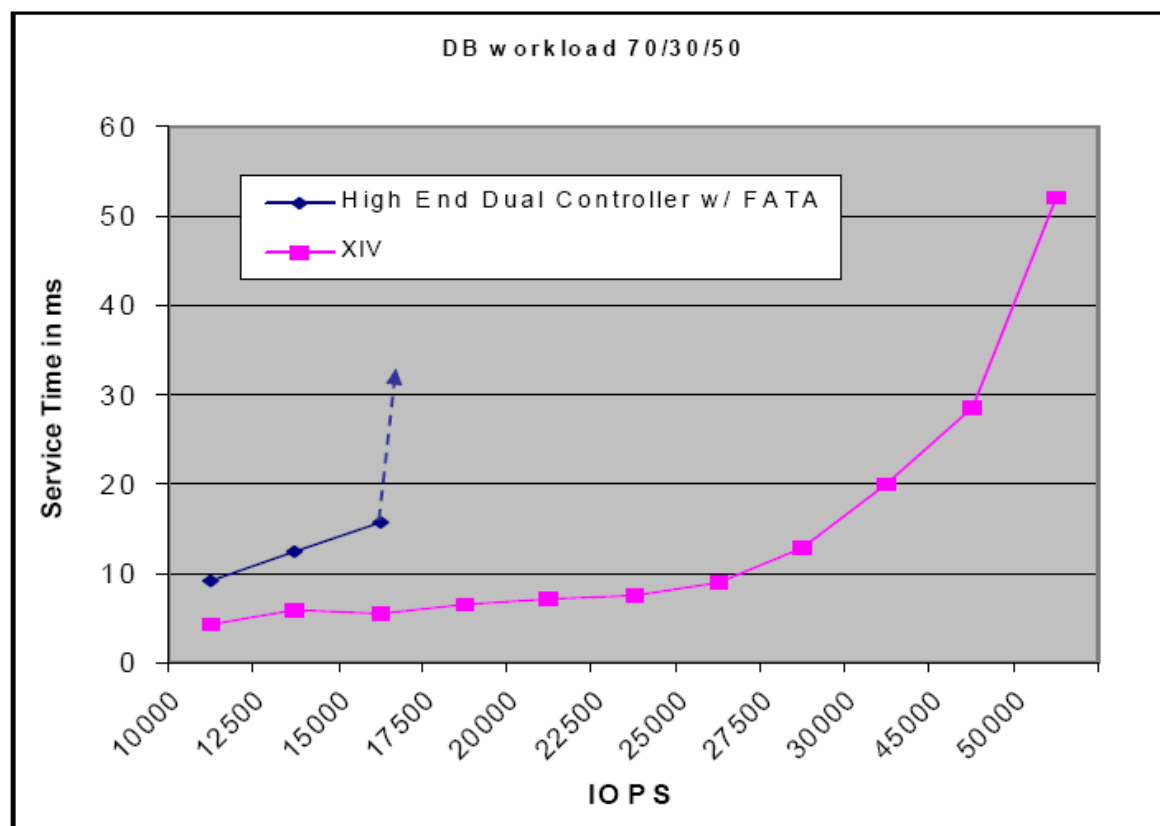
■ Аналогичный тест, но добавляем DS8700





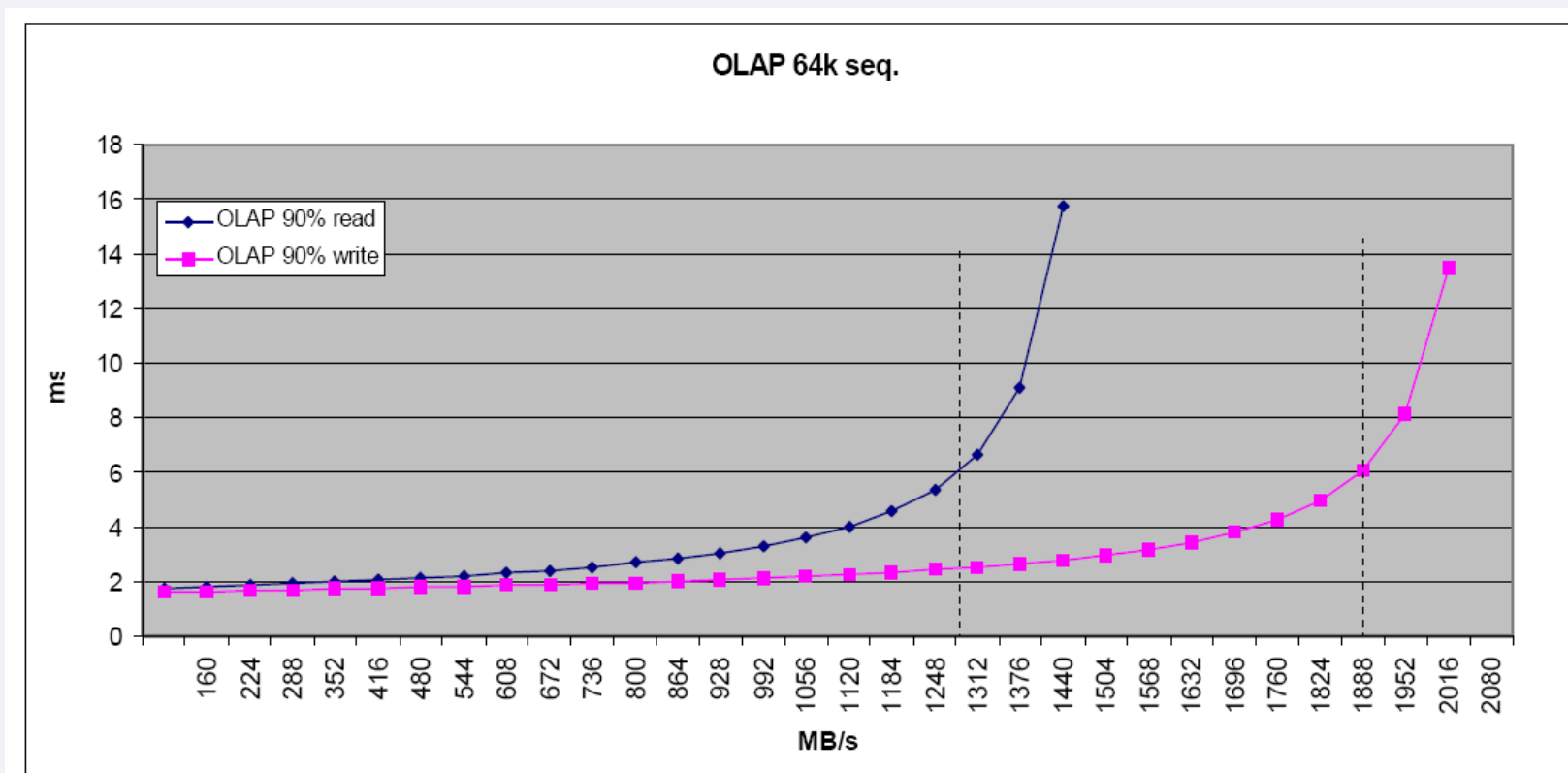
SATA в XIV не тоже самое!

- Рассмотрим что будет если мы поместим SATA диск в традиционную 2 контроллерную систему
Производительность SATA в XIV достигается благодаря GRID архитектуре



OLAP в XIV

90% чтение и запись
90% последовательно
Размер блока 64 kB





Пример сайзинга

Заказчик имеет 40 баз (SAP instances), каждой необходимо 5.000 SAPS

– $40 \times 5.000 \text{ SAPS} = 200.000 \text{ SAPS} \sim 100.000 \text{ IOPS}$

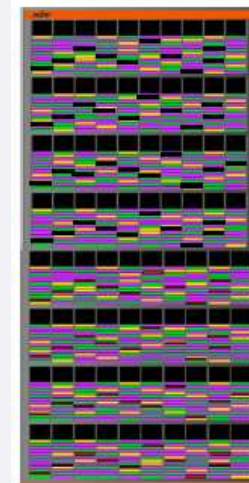
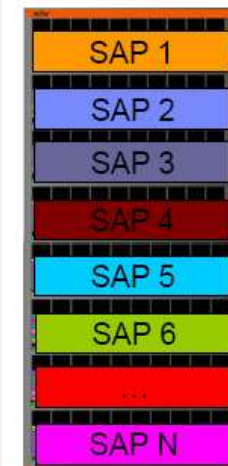
При использовании традиционной СХД нам необходима производительность 100.000 IOPS

□

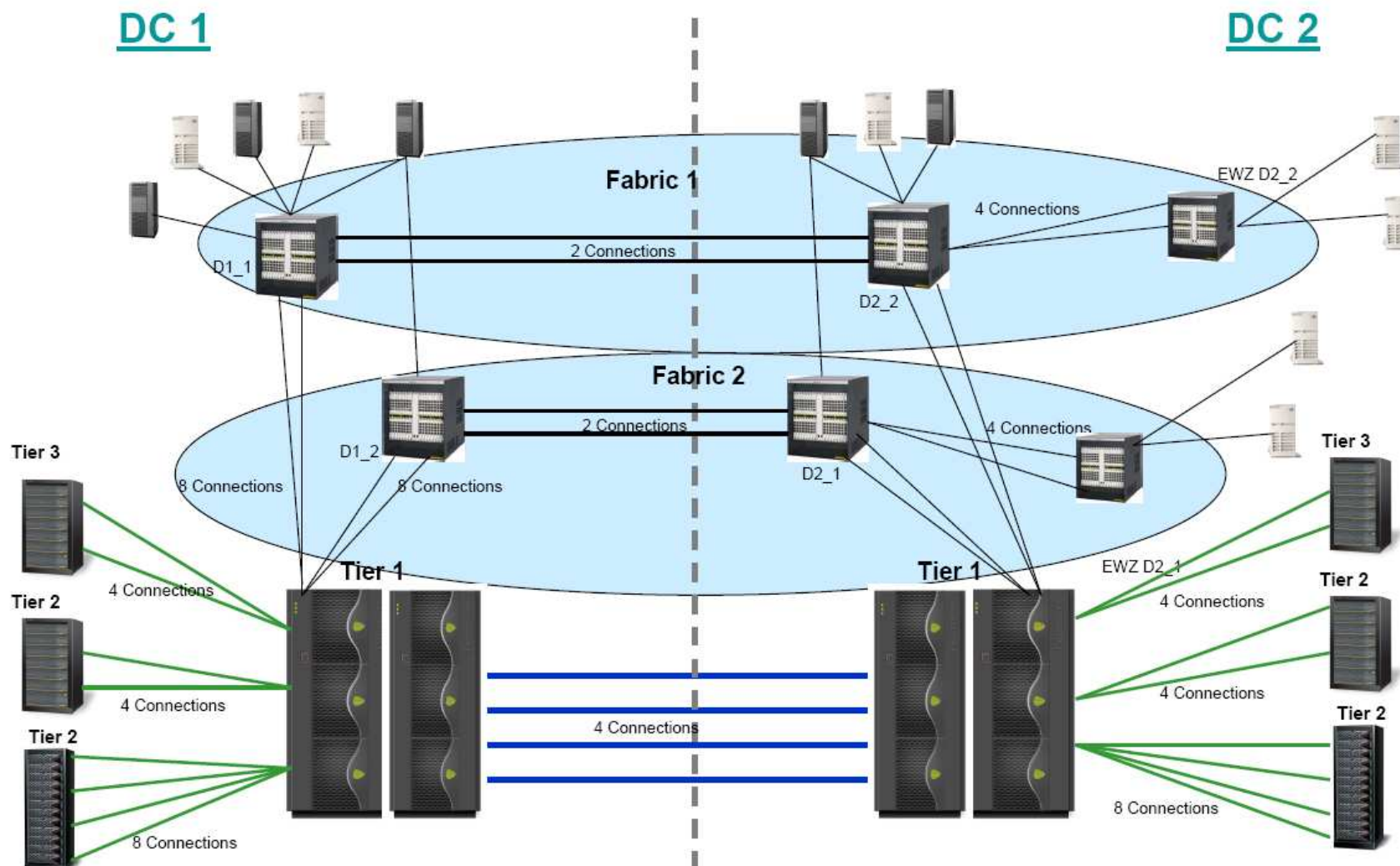
XIV полностью виртуализированна

– Примерно 50% баз (instances) находятся в пике в одно и тоже время

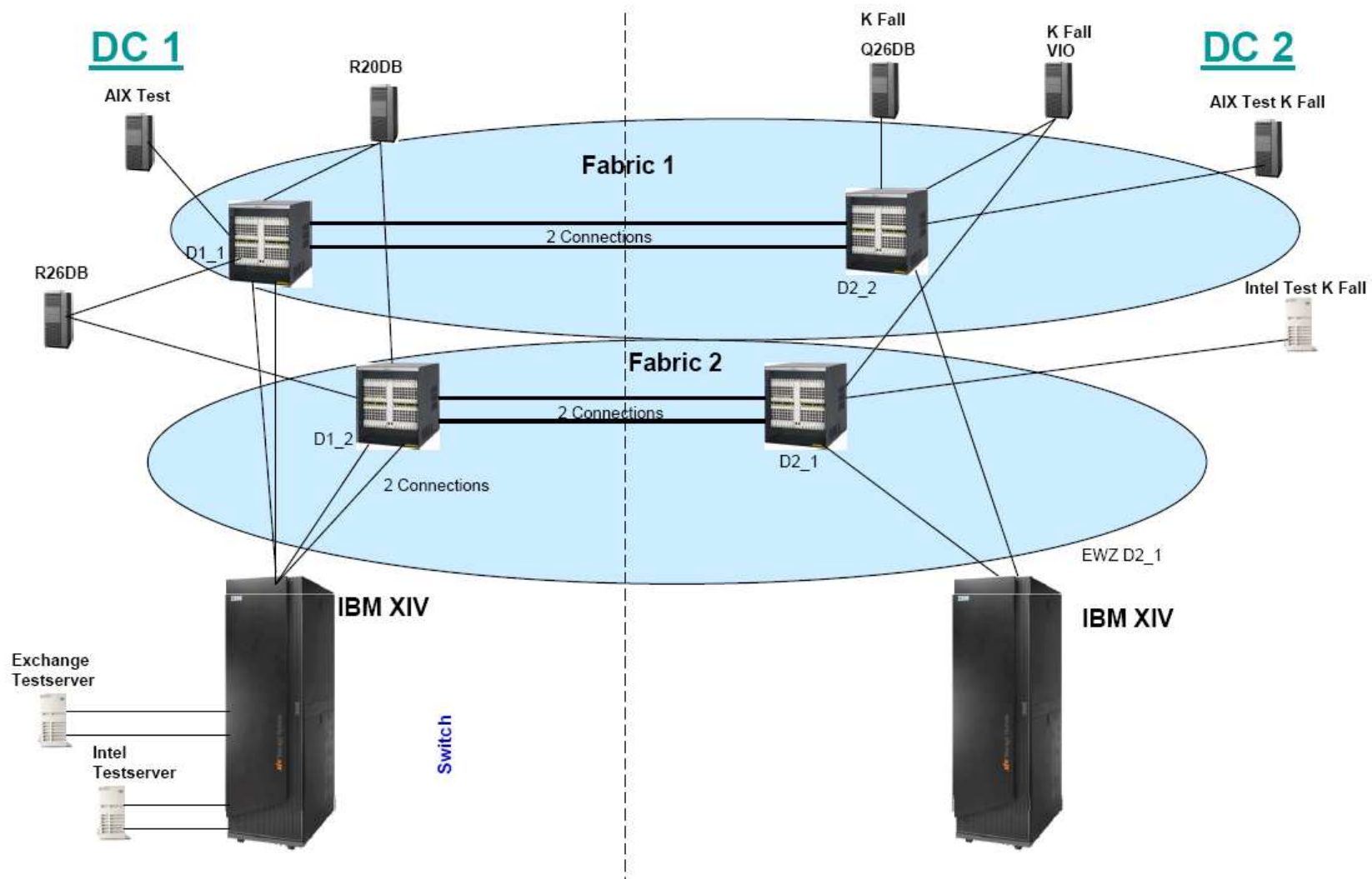
С XIV нам необходимо $100.000 \times 50\% = 50.000 \text{ IOPS}$



Тестовый стенд 1



Тестовый стенд 2





Результаты

Customer Performance Performance Test for Exchange with Jetstress 8.x												
					IBM XIV				Hitachi USPV (Jetstress 8.x/HDLM)			
#User	JetStress Server	Mailbox Size	Threads	Run-time	IOPS	Total #IOs	Database Disk Performance (Write)	Log Disc Performance (Write)	IOPS	Total #IOs	Database Disk Performance (Write)	Log Disc Performance (Write)
1.500	1	150	8	02:00	5867	5867	0,005	0,002	3281	3281	0,007	0,001
1500	2	150	32	02:00	5136	10300	0,025	0,012	3465	7002	0,029	0,012
1500		150		02:00	5164		0,025	0,012	3537		0,028	0,011
1500	4	40	4	02:00	1645	6646	0,006	0,002	1164	4603	0,012	0,001
1500		30	4	02:00	1665		0,006	0,002	1218		0,011	0,001
1000		70	4	02:00	1668		0,005	0,002	1125		0,014	0,002
1000		70	4	02:00	1668		0,005	0,002	1096		0,012	0,002
1500	4	150	8	08:00	3549	12545	0,006	0,002	2091	7949	0,016	0,002
1.500		150	8	08:00	3549		0,006	0,002	2601		0,014	0,002
1.000		100	8	08:00	2725		0,006	0,002	1646		0,012	0,002
1.000		100	8	08:00	2722		0,006	0,002	1611		0,013	0,002
1.500	3	150	8	08:00	3544	10047	0,006	0,002		0		
1.500		150	8	08:00	3569		0,006	0,002				
1.500		150	8	08:00	2934		0,006	0,002				
1.500	2	150	8	08:00	3805	7610	0,006	0,002	2593	5265	0,009	0,001
1.500		150	8	08:00	3805		0,006	0,002	2672		0,008	0,001
1.500	1	100	32	02:00	10979	10979	0,010	0,006	7674	7674	0,013	0,006

Пример внедрения

Банк Leumi



Высокая производительность,
Эффективное использование
логических копий

Задача

- 7ТБ Oracle Database для регистрации транзакций (compliance)
- Чрезвычайно высокие требования производительности
- Безуспешно пытались внедрить DMX
- Сохранение в «горячем режиме» невозможно на существующих системах

Решение

- Система XIV с оптимальной утилизацией драйвов и эффективной способностью логических копий

Преимущества и выгоды

- Достигнута высочайшая скорость обработки транзакций по сравнению с другими Hi-End системами
- Можно делать сохранение в «горячем режиме» без ощутимых потерь производительности
 - Сейчас 4 ежедневных копии для сохранения
 - Копии хранятся в течение недели
 - Можно восстановиться с любой из 28

Landata



**Спасибо за
внимание!**

